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

Davide M. Coluccia Gaia Dossi

This Version: January, 2025

Latest version available here; Online Appendix available here

Abstract

This paper documents that out-migration promotes the diffusion of innovation from the country of destination to the country of origin of migrants. Between 1870 and 1940, nearly four million British immigrants settled in the United States. We construct a novel individual-level dataset linking British immigrants in the US to the UK census, and we digitize the universe of UK patents from 1853 to 1899. Using a triple-differences design, we show that migration ties contribute to technology diffusion from the destination to the origin country. The text analysis of patents reveals that emigration promotes technology transfer and fosters the production of high-impact innovation. Return migration is an important driver of this "return innovation" effect. However, the interactions between emigrants and their origin communities—families and neighbors—promote technology diffusion even in the absence of migrants' physical return.

Keywords: Age of Mass Migration, Innovation, Networks, Out-migration.

JEL Classification: F22, N73, N74, O15, O31, O33.

*Coluccia: School of Economics, University of Bristol, davide.coluccia@bristol.ac.uk; Dossi: Stone Centre, Department of Economics, University College London, g.dossi@ucl.ac.uk.

We are grateful to Mara Squicciarini and Joel Mokyr for their continued guidance and support. For insightful discussions, we also thank Jan Bakker, Enrico Berkes, Maristella Botticini, Jeremiah Dittmar, Joe Ferrie, Sofia Franchini, Carola Frydman, Simon Görlach, Rachel Griffith, Walker Hanlon, Leander Heldring, Matteo Leombroni, Nicola Limodio, Paolo Masella, Filippo Mezzanotti, Alessandro Nuvolari, Laura Ogliari, Sebastian Ottinger, Eleonora Patacchini, Marta Prato, Nancy Qian, Joshua Rosenbloom, Paola Sapienza, Carlo Schwartz, Marco Tabellini, and John Van Reenen. We thank audiences at the EHA (Pittsburgh), EHS (Cambridge), and ES (Berlin) conferences, the 10th CEPR Economic History Symposium, and seminar participants at Bocconi, Bristol, Cologne, Lugano, Milan, Northwestern, Southampton, Tilburg, Torino, and Uppsala. Antonin Bergeaud, Enrico Berkes, and Walker Hanlon kindly shared data with us. We acknowledge financial support from Bocconi, Fondazione Invernizzi, LSE, and POID. All errors are our own.

I Introduction

The diffusion of new technologies across countries is a central driver of productivity growth and economic convergence (e.g., Eaton and Kortum, 1999; Griffith, Harrison and Van Reenen, 2006; Comin and Hobijn, 2010).[1] However, evidence on the mechanisms enabling technology to diffuse across countries is surprisingly scarce. In this paper, we provide novel evidence on the role of out-migration as a source of technology transfer from the country of destination to the country of origin of migrants. We label this phenomenon "return innovation."

The impact of emigration on innovation is *ex-ante* ambiguous. The traditional brain drain hypothesis argues that emigration countries suffer a loss of human capital (e.g., Docquier and Rapoport, 2012). Growth theory, in turn, predicts that this depletion would negatively affect their ability to innovate. On the other hand, recent evidence suggests that exposure to innovation is an important driver of technological change (Akcigit, Caicedo, Miguelez, Stantcheva and Sterzi, 2018; Bell, Chetty, Jaravel, Petkova and Van Reenen, 2019). We argue that, as migrants are exposed to innovation in the areas where they settle, they promote knowledge flows between those areas and their origin country, as documented qualitatively by Saxenian (2006).

We focus on the English and Welsh migration to the United States between 1850 and 1940.[2] During this period, also known as the "Age of Mass Migration," approximately 30 million European migrants settled across the Atlantic. Nearly four million came from England and Wales. Since Rosenberg (1982), economists have interpreted the spread of the Industrial Revolution in terms of waves of technological diffusion originating in Britain. Existing studies document that, as a result, European immigrants contributed to the diffusion of (mainly) British technology in the US (Jeremy, 1981). This pattern reflects the British technological leadership during the first half of the nineteenth century. Since as early as the 1860s, however, 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

[1] Among others, Eaton and Kortum (1999) estimate that 70% of productivity growth in advanced European countries in the 1980s relied on technology developed in the United States and Japan. Several recent endogenous growth models include cross-country technology diffusion dynamics (e.g. Alvarez, Buera and Lucas, 2013; Buera and Oberfield, 2020; Perla, Tonetti and Waugh, 2021).

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

[3] By the 1890s, the American technological primacy was well-established. Nelson and Wright (1992) note that, starting in the 1880s, American technology saw major advancements in textiles, sewing machines, clocks, firearms, boots and shoes, locomotives, bicycles, and cigarettes. From the 1890s, mass production led to innovations in consumer products (canned goods, dairy, and grain products), light machinery (typewriters, cameras), electrical equipment, and industrial machinery, such as boilers, pumps, and printing presses.

networks promoted the diffusion of these technologies back to Britain.

We assemble two novel general-purpose datasets to overcome the limitations of the existing sources. First, since the available data do not contain information on the origin of British immigrants in the US beyond their country of origin, 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, Fitch, Goeken, Hacker, Nelson, Roberts, Schouweiler and Sobek, 2021). This dataset enables us to track individual migration 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 digitize all 300,000 original patent documents issued in England and Wales between 1853 and 1899. This dataset allows us to cover the universe patents granted during the period of our analysis in the UK.

We build a measure of exposure to US innovation that is higher for districts with larger overseas emigrant communities and when the US counties that emigrants settle in are more innovative.[4] Our research design leverages the joint variation in county-level specialization across technology classes and district-county bilateral migration flows between UK districts and US counties.

Identifying the impact of exposure to US innovation on innovation in the UK presents several challenges. The primary reason that would caution against a causal interpretation of the estimates is assortative matching. Namely, British migrants could sort across US counties in a way that is correlated with innovation in their district of origin and in their county of destination. We develop two approaches to deal with this challenge. Our main identification strategy leverages episodes of unusually high patenting activity in the United States—which we label as "innovation shocks"— as shocks to exposure to US innovation. We examine how such shocks diffuse to the UK through emigration ties in double and triple differences designs. In a complementary approach, we build two instrumental variables. Specifically, we exploit variation in the connection timing to the US railway network and shift-share variation in the location of immigrants across US counties to construct predicted county-level immigration shocks (Card, 2001; Sequeira, Nunn and Qian, 2020). We use these shocks to randomize British immigration across counties.

The main finding of this paper, which we label "return innovation," is that innovation in the UK is affected by exposure to US innovation through out-migration linkages. First, we document an increase in total patenting in districts with more US emigrants: moving from the 25th to the 75th percentile of US out-migration yields an 8% increase in the number of patents produced in the United Kingdom.

[4]Our core units of observation are UK registration districts and US counties. In 1901, there were 631 registration districts in England and Wales. Districts were comparable to US counties in population (approximately 40,000). Unlike counties, however, registration districts were statistical entities that did not enjoy political or budgetary autonomy.

Second, districts increase their innovation activity in the technological fields to which emigrants are exposed in the United States. Using an index that leverages joint variation in the location where emigrants settle and the innovation they are exposed to, we find that moving from the 25th to the 75th percentile of the "knowledge exposure" index is associated with an 11% increase in innovation activity. Third, we adopt a text-analysis methodology that computes the textual similarity between UK and US patents to measure the transfer of innovation generated by migration ties. We find that patenting in districts with more US emigrants becomes more similar to US patents: moving from the 25th to the 75th percentile of emigration to the US is associated with UK patents that are twice as similar to American ones. This effect also holds within technologies.

In the double differences regression, innovation in districts exposed to a US innovation shock through migration ties increases by 9%. A back-of-the-envelope calculation shows that migration ties generate a 15% pass-through rate of innovation shocks from the United States into the United Kingdom. In exposed districts, patents become substantially more similar in their textual content to American patents. These results also hold in the triple differences framework, which leverages shocks to US innovation activity across counties and technologies. A causal interpretation of the double and triple differences estimates requires that districts exposed to US innovation shocks and untreated districts would not have displayed differences in patenting activity in the absence of the shocks. Throughout the paper, we estimate event-study specifications and the pre-treatment coefficients are never statistically different from zero, thus supporting the parallel trends assumption.

The effects of exposure to foreign innovation are heterogeneous across sectors. In particular, we uncover an inverted U-shaped relationship between return innovation and the specialization of the UK relative to the US. In response to emigrants' exposure to US innovation, patenting in the UK increases the most in sectors where the US and the UK have similar specialization rates. By contrast, the effect is considerably smaller in areas where the US is more advanced, such as agriculture and metallurgy, and where the UK is more advanced, such as textiles. This pattern is consistent with the theoretical predictions of Van Patten (2023), who argues that the gains from international technology diffusion are highest when a country interacts with partners with higher but similar levels of development.

In the second part of the paper, we exploit the richness of our data to explore the mechanisms behind the return innovation effect. On the one hand, return innovation may require the physical return of migrants. On the other hand, migration ties may promote the diffusion of new technologies irrespective of physical return. We find that the physical return of migrants is an important driver of return innovation, but it does not explain it in full. Namely, return migration flows account for approximately half of the overall return innovation effect. However, the impact of exposure to

US innovation through out-migration remains sizable and significant even in the absence of return migration. This suggests that migration ties contribute to the diffusion of knowledge even without physical return.

First, we study the role of interactions between emigrants and their communities of origin as a driver of the diffusion of innovation. We focus on two factors that could promote such interactions: family ties and geographical proximity. Our analysis builds on a large literature in development economics which links the diffusion of technology to network interactions (e.g. Bandiera and Rasul, 2006; Conley and Udry, 2010; Beaman, BenYishay, Magruder and Mobarak, 2021).

We document that family members of US emigrants increase their patenting activity after their relatives moved to the US. It takes about ten years for a British emigrant to contribute to innovation activity back home. Despite this delay, the magnitude of the effect is large. Importantly, we can distinguish between emigrants who, at some point, return to the UK from those that do not. The impact of return emigrants is larger than those who never return. However, emigrants promote innovation in their families even if they never return. Since return emigrants account for approximately one-third of the entire migrant stock, the magnitudes of these effects are, in aggregate, similar.

As an alternative proxy for local social networks, we leverage the geographical proximity between emigrants and their former neighbors. To estimate its impact on the innovation activity of stayers, we link the UK patent data to the UK full census, and we geo-code information on the universe of the UK population. We find that patenting activity increases for non-migrants after their neighbors migrate to the United States.[5] The estimated effect remains positive and significant when we exclude the neighbors of return migrants, suggesting that cross-country interactions between emigrants and their origin communities are a key driver of return innovation.

Second, we study how migration ties interacted with information technologies to facilitate the flow of knowledge from the United States to the United Kingdom. We study the introduction of the first transatlantic telegraph cable connecting the US and the UK in 1866 as a sudden and sizable reduction in the cost of information transmission (Steinwender, 2018). Using a difference-in-differences framework, we show that districts with higher US emigration rates before the introduction of the transatlantic telegraph cable have higher patenting activity after 1866. Moreover, the gains in patenting activity are larger in those same technologies that districts had been more exposed to through migration ties. These findings suggest that the increased economic integration generated by the telegraph accrued relatively more to districts with pre-existing migration ties with the US market.

[5] In the baseline exercise, two individuals are considered neighbors if they live within five kilometers of each other. The results are very similar when the distance is between 3- and 10-kilometers.

To conclude the analysis, we examine whether information flows promoted by migration ties are specific to innovation. To answer this question, we study the effect of out-migration on newspaper coverage of US-related news in the United Kingdom. We collect data on the coverage of US-related news from a comprehensive repository of historical British newspapers. We find that newspapers in areas with more US emigrants are relatively more likely to cover US-related news. Coverage of a given state or county is higher in districts with more emigrants to that given state or county. This suggests that migration ties generate a flow of information that goes beyond the diffusion of new technologies.

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

Related Literature. This paper is related to four streams of literature. First, we contribute to the literature studying the dynamics and determinants of knowledge flows and technology diffusion across countries (among others, see Jaffe, Trajtenberg and Henderson, 1993; Griffith *et al.*, 2006; Bahar, Hausmann and Hidalgo, 2014). Specifically, we contribute to the papers documenting how human mobility fosters the diffusion of novel knowledge (e.g., Kerr, 2008; Hornung, 2014; Bahar, Hauptmann, Özgüzel and Rapoport, 2019; Bahar, Choudhury, Sappenfield and Signorelli, 2022; Prato, 2022). We contribute to this literature from several perspectives. First, we study the universe of emigrants instead of a selected subgroup of highly skilled individuals. Second, we leverage recent insights by Bell *et al.* (2019) and show that exposure to foreign innovation affects the diffusion of those technologies. Third, we emphasize that the return innovation effect does not exclusively hinge on the physical return of emigrants. Finally, our setting allows us to uncover the long-run effects of emigration and the mechanisms through which it affects innovation in the home country of emigrants.

Second, we contribute to the literature that studies the determinants of the direction of innovation and the allocation of research activity across technological sectors. Pioneering work on directed technical change 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; Moscona and Sastry, 2023; Einiö, Feng and Jaravel, 2023; Dossi, 2024; Truffa and Wong, 2024). We add to this literature by introducing one novel determinant of the direction of innovation, namely, international human mobility, through the return innovation effect.

Third, we contribute to the literature that studies the effects of out-migration on countries sending

migrants. Emigration has been shown to impact wages (e.g., Dustmann, Frattini and Rosso, 2015), attitudes towards democracy (Spilimbergo, 2009; Batista and Vicente, 2011) and political change (Chauvet and Mercier, 2014; Karadja and Prawitz, 2019), technology adoption (Coluccia and Spadavecchia, 2024), entrepreneurship (Anelli, Basso, Ippedico and Peri, 2023), and social norms (Beine, Docquier and Schiff, 2013; Bertoli and Marchetta, 2015; Tuccio and Wahba, 2018). This paper provides new evidence that emigration shapes the rate and direction of innovation because it exposes emigration countries to novel knowledge produced abroad. These findings add to the existing literature, which generally interprets out-migration as a depletion of the stock of human capital.

Finally, by its setting, this paper adds to the literature that studies technical change and diffusion of novel technologies during the Age of Mass Migration. A growing number of papers examines the short-run (e.g., Moser, Voena and Waldinger, 2014; Arkolakis, Lee and Peters, 2020; Moser, Parsa and San, 2020) as well as the long-run (e.g., Akcigit, Grigsby and Nicholas, 2017; Burchardi, Chaney, Hassan, Tarquinio and Terry, 2020) implications of immigration on US innovation. This paper is closest to Andersson, Karadja and Prawitz (2022), who show that mass out-migration in Sweden triggered labor-saving innovation by increasing the relative cost of labor. In this paper, we document an entirely different phenomenon–the "return innovation effect"—whereby emigrants generate knowledge flows from where they settle towards their areas of origin. Our findings show that emigration promotes the cross-border diffusion of knowledge and investigate the underlying mechanisms.

Outline. The rest of the paper is structured as follows. In section II, we describe this study's historical and institutional context. In section III, we describe the data. In section IV, we present the empirical research design. In section V, discuss the main findings. In section VI, we discuss the mechanisms and possible alternative interpretations. In section VII, we conclude.

II Historical Background

In this section, we provide details on the English and Welsh migration to the United States (II.A) and on intellectual property protection in the United Kingdom and in the United States (II.B).

II.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 United States (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).[6] In this section, we provide details on the English and Welsh migration to the US.

II.A.1 Migration Policy in the United Kingdom and the United States

Until 1917, the US applied minor restrictions on European immigration (Abramitzky and Boustan, 2017). This means that the Age of Mass Migration was characterized by virtually no legal constraints to human mobility to the US. Immigrants mostly originated from Northern Europe, particularly the United Kingdom, 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, however, allotted generous quotas to the United Kingdom, which were never filled (Abramitzky and Boustan, 2017).[7]

Like in other European countries, out-migration legislation in the UK encouraged emigration towards the Empire (Baines, 2002).[8] Emigration to the United States was neither subsidized nor discouraged. Attitudes towards out-migration remained positive after World War I. The perceived slowdown of emigrant flows after the War was viewed with concern by policymakers (Leak and Priday, 1933).

II.A.2 The Characteristics of British Emigrants

The British migration to the US presents two main distinctive features compared to the broader European phenomenon.[9] First, unlike most continental countries, Britain was already highly urbanized and industrialized at the inception of the Mass Migration. Starting in the 1880s, urban areas supplied most overseas migrants (Erickson, 1957, 1972; Thomas, 1954).

---

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

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

[8] Out-migration was encouraged in two ways: reduced and subsidized ticket fares and allotment of agricultural lands. Policy efforts were directed towards the Empire, particularly Canada and Australia, through the Committee of the Emigrants' Information Office. In general, however, these policies were not successful. Baines (2002) argues that less than 10% emigrants traveled under government assistance during the entire 1814-1918 period, and Leak and Priday (1933) report similar figures for the post-War era.

[9] Throughout the period, the US was the most frequent destination for English and Welsh migrants. Between 1850 and 1930, more than 40% of total British and Welsh emigrants settled in the US. This compares to 25% in Canada, 20% in Australia, and 15% in other destinations (Baines, 2002).

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, the vast majority were farmers. However, as cities and smaller urban centers gained prominence, migrants were increasingly employed in industrial manufacturing occupations. At the beginning of the 1860s, when the transatlantic migration was taking off, about 15% emigrants were employed in agriculture, and merely five percent were white-collar workers. In the early 1900s, however, this composition had shifted as agriculture workers accounted for a mere five percent of the overall emigrant stock, while those employed in white-collar occupations were 25%.

II.B Intellectual Property Protection in the United Kingdom and in the United States

During the period of our analysis, patents in the United Kingdom and the United States were granted by the respective patent offices. However, throughout the period, the UK and the US did not mutually recognize patents. In this section, we describe patent protection in the two countries, as well as international intellectual patent protection over this period.

II.B.1 Patent System in the United Kingdom

The United Kingdom 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 system during the Industrial Revolution (e.g., Macleod, 1988). The 1852 Patent Law Amendment Act sought to reform this process. The US system inspired the reform effort, which reduced application fees and shortened the bureaucratic process. A subsequent reform in 1883 further reduced fees and allowed applications by mail. This reform also introduced a litigation system and the employment of professional patent examiners. A technical examination of novelty was introduced only in 1902. Until 1907, patents were granted conditional on the invention being produced in Britain (Coulter, 1991).

II.B.2 Patent System in the United States

The first article of the United States Constitution establishes that inventors should be granted exclusive rights over their discoveries. In 1836, US Congress passed the Patent Act, which formally instituted

the US Patent Office (USPTO). The USPTO has been credited as the first modern patent system in the world (Khan and Sokoloff, 2004). Two features distinguished the American patent system from its European counterparts. First, professional examiners carried out a novelty examination to ascertain the originality of patent applications. Second, low application fees ensured access to intellectual property protection was widespread.

II.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 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 United States strongly advocated for reciprocity (Penrose, 1951). Second, upon applying for a patent in a member country under Article 4, 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 established in 1970. Since the UK and the US did not mutually recognize patents, we can use them as a proxy for knowledge flows between the two countries.

III Data

To conduct the analysis, we assemble two novel datasets. The first links British migrants in the United States to the UK census, allowing us to construct a matrix of bilateral migration flows at the UK district and US county level. This also provides individual-level information on the origin of migrants and allows us to build a dataset of return migrants. We describe this dataset in section III.A. The second dataset includes the universe of patents granted in England and Wales between the second half of the 19th century. We digitize these data, which to date were not available in disaggregated

form, and link inventors to the UK census. We describe this dataset in section III.B. In section III.C, we describe the additional data sources we use in the analysis.

III.A A Novel Individual-level Dataset of British Immigrants in the United States

Our empirical analysis requires information on the origin of English and Welsh immigrants *within* the United Kingdom. The available data, however, do not contain this information. Neither the US nor the UK collected disaggregated records on the origin of immigrants and 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 are able to observe an individual in the US and to track him to his UK census record before he emigrated.[10] This is the first dataset that reconstructs migration flows at this granular level of aggregation for a non-Scandinavian European country during the Age of Mass Migration and adds to work by Abramitzky, Boustan and Eriksson (2014) Andersson *et al.* (2022) for Norway and Sweden.[11]

To construct the linked dataset, we leverage confidential individual-level data from population censuses in the United Kingdom (Schurer and Higgs, 2020) and the United States (Ruggles *et al.*, 2021). We extract the universe of British male immigrants from US censuses in 1900, 1910, 1920, and 1930.[12] These records contain the name and surname, birth year, and immigration year of each migrant. Using these variables, we match these individuals to the last British census where the emigrants appeared.[13] We use state-of-the-art census-linking algorithms adapted from pioneering work by Abramitzky, Boustan, Eriksson, Feigenbaum and Pérez (2021), as documented in appendix sections A.III.1 and A.III.2, which discuss in more detail the primary sources and the technical implementation of the algorithm.

In our preferred version of the data, we match approximately 65% of male British immigrants who

[10] Throughout the paper, we use the masculine to refer to individuals in our data because, as we explain in detail later, we can only work with male individuals.

[11] England and Wales were larger in terms of the overall population and US immigrant population. The population of Sweden and Norway in 1890 was approximately 4.7 and 2 million, while it was 27 million in England and Wales.

[12] 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.

[13] For example, we link an individual who immigrated to the US in 1905 to the 1901 UK census. Because no census was taken in 1870, we match those who migrated between 1870 and 1881 to the 1860 census. Moreover, since the last available UK census was in 1911, we match all those who emigrated after 1911 to that one. This implies that we have no information on migrants born after 1911. Since the median age of migrants is 30 and less than 10% of the distribution is younger than 19 in the rest of the sample, this bears little quantitative implications for the matching rate in the later part of the sample.

appear in the United States census, as shown in Appendix Figure A.9. The linking rate is considerably higher than benchmark datasets linking individuals over time (e.g., Abramitzky *et al.*, 2021). There are at least three potential explanations for this difference. First, British immigrants were a positively selected group of urban skilled manufacturing workers. Their census records would, therefore, plausibly be more accurate than for other populations.[14] Second, British immigrants spoke English. Linguistic factors do not threaten our linking algorithm, but they are challenging to resolve in intergenerational linking routines. Finally, the high quality of UK census records facilitates the linking procedure.

This approach nonetheless presents some important caveats.[15] First, it may deliver spurious links if the matching variables are insufficient to restrict the pool of potential matches. Second, the matching probability may be correlated with individual characteristics. To address the first issue, we discard the matches that do not attain a high level of string similarity. Moreover, we weigh each migrant by the inverse of the number of matches he is paired with to minimize the weight placed on false positive matches. In Table A.3, we report the correlation between emigrants' observable characteristics and the linking probability. Overall, this association is small and it is often statistically insignificant.

It is challenging to validate the data with external sources that contain disaggregated information on the origin of British emigrants to the US because, to the best of our knowledge, they do not exist. In Appendix Figure A.11, we correlate our data with county-level estimates of aggregate transatlantic emigration for the period 1880–1910 assembled by Baines (2002) and confirm that the two are positively correlated.[16]

In addition, we construct a dataset of return migrants. To assemble it, we apply the same logic as before. The only difference is that migrants are matched to the UK censuses taken in the decades *after* their immigration year. As an example, someone who migrated to the US in 1895 is matched to censuses in 1901 and 1911. To avoid double counting, if an individual is matched to more than one census, we keep the match(es) in the first.

Figure I reports in grey the number of English and Welsh immigrants in the United States by year of immigration, digitized from official statistics (Willcox, 1928). The blue line on the right y-axis tabulates the number of immigrants in our linked dataset. Our data co-moves with official statistics data. Figure IIa reports the spatial distribution of emigration rates across districts in the final sample

[14] This is in line with evidence by Helgertz, Price, Wellington, Thompson, Ruggles and Fitch (2022) that individuals with relatively higher socio-economic status have higher linking rates in the US census.

[15] Appendix Section A.III.3 presents a more detailed discussion of the internal and external validation exercises we undertake to assess the plausibility of the linked emigrant sample.

[16] See Appendix A.III for a more detailed discussion.

and highlights its cross-sectional spatial heterogeneity. In Appendix Figure A.12, we break down the map by decade and uncover substantial variation in the origin of US emigrants over time. We discuss this in detail in Appendix Section A.III. Table I, Panel A provides descriptive statistics for variables tabulated from the 1880 individual census. Panel B lists the district-level number of emigrants and return migrants in the UK-US-linked migrants sample by decade.

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

We measure innovation activity using patents, as standard in the literature (Griliches, 1998).[17] We use data on the universe of patents granted in the United States digitized by Berkes (2018). Leveraging information on the county of residence of inventors and the technology class (both reported on patent documents), we construct a balanced panel at the county-technology class-grant year level.[18]

We collect all the patents granted in the United Kingdom in 1900-1939 from PATSTAT, which provides bulk access to documents from the European Patent Office. These data contain information on authors and CPC classes but do not report the geographic location of inventors. To retrieve the coordinates of the inventors, we merge them with data by Bergeaud and Verluise (2024) and map them to registration districts at their 1890 borders.

For the period between 1853 and 1898, UK patent data was not available in digitized format. To tackle this data limitation, we digitize the universe of patents granted in England and Wales between 1853 and 1899 from scans of the original documents that we obtained from the UK Patent Office (UK IPO). We apply optical character recognition (OCR) techniques on more than 300,000 original documents to convert the scanned images into machine-readable texts. Then, we use large language models to parse the information contained in the raw text into a structured dataset. Compared to standard text extraction algorithms, this method harnesses artificial intelligence to deal with OCR issues and non-standard formulations of patent texts.[19] We parse information on the title, text, inventors' names,

[17] Patents are not a flawless measure of innovation because non-patented innovation represents a non-negligible share of overall technological progress (Moser, 2019). We nonetheless believe that this is a comparatively minor issue for our analysis. As discussed in section II.B, before our study period, the US and the UK had enacted important reforms that decreased the cost of access to patent protection. These drastically increased the number of patents in both countries, thus ensuring that patents convey an informative picture of the state of technology in both countries. In addition, our results also hold *within* technologies. Hence, if the propensity to patent in a given sector is similar in the UK and in the US, different propensities to patent across sectors should not be a concern for our analysis.

[18] We map patents to counties at 1900 borders. From the three-digit Cooperative Patent Classification (CPC) class, we map patents to a coarser taxonomy of twenty technological sectors.

[19] Appendix A.II provides more details on the implementation procedure of the OCR and parsing algorithms.

geo-referenced addresses, filing and issue dates, assignee status, and whether the patent was filed by a patent agent. To conduct the analysis, we map patents to districts at 1890 borders.

This dataset adds to the one developed by Hanlon (2016) (1855–1883) by adding information on patent titles, texts, and geography and extending the period until 1898, and expands previous work by Nuvolari and Tartari (2011), which covers the period until 1853. Figure A.4 reports the number of patents granted in the UK by year (Panel A.4a). The blue dots report the dataset we assemble, and the red data are tabulated from PATSTAT. The two series are consistent and display a steady growth in the number of patents issued throughout the period. Panel A.4b shows the evolution in the number of patents granted by technological class.

Patents differ by novelty and economic value. In contemporaneous settings, patent forward citations are the standard proxy for patent impact. However, these were not reported in historical documents (Andrews, 2021). To construct a measure of patent "impact," we apply the text-based methodology developed by Kelly, Papanikolaou, Seru and Taddy (2021) to the corpus of British patents. Additionally, we use the same methodology to measure the textual similarity between British and American patents and proxy for technology transfer between the US and the UK. Appendix A.II.6 describes in detail the practical implementation of the algorithm.

Finally, to perform the neighborhood-level analysis, we link the patent data to the census. To perform this linking, we match inventors based on the string similarity between their name and surname and those recorded in the census, conditional on geographic proximity. We describe the precise implementation in Appendix section A.II.7. Figure A.8 displays the geographic distribution of inventors across England and Wales using the geo-coded addresses of the inventors linked to the census. The map highlights that most innovation happens in urban centers.

In Table I, Panels C and D, we report descriptive statistics for this sample. Panel C reports the total number of patents granted across districts by year and breaks it down into the five most common fields in the sample. Panel D reports statistics for the sample of inventors linked to the population census. Figure II, Panel IIb reports the spatial variation in number of patents normalized by district population in 1891.

III.C Additional Variables

To conclude this section, we provide details on the additional variables that we construct from existing sources. Specifically, we collect data on UK district-level statistics, on British local newspapers, and on the telegraph network in the UK.

III.C.1 Data Constructed from the British Population Census

We assemble district-level statistics from UK population censuses at a decade frequency between 1851 and 1911. Districts are the level of observation in most of the analysis. This is because they were statistical units with an average population of 40,000, which makes them broadly comparable to US counties.[20] Districts undergo minor boundary changes during the analysis period. To ensure geographical consistency, we cross-walk all variables to districts in 1890 (see Appendix A.I.6).

III.C.2 Newspapers

We use newspaper coverage of US-related topics as a measure of attention to the United States in public opinion. We collect the data from the British Newspaper Archive. Beach and Hanlon (2023) discuss this dataset in detail. We run three sets of queries. First, we search for the joint mention of the words "United States"; second, we search for mentions of each US state; third, we search for mentions of each US county, jointly with either the state name or "United States". We collect these data at the newspaper level from 1850–1939. Additionally, we know each newspaper's publishing address, which we geo-reference to 1890-border districts. Ultimately, we assemble three datasets at the district, district-state, and district-county levels, each at decade frequency. Figure A.2 reports the distribution of newspapers, and Table A.4 provides descriptive statistics on their activity.

III.C.3 Telegraph Network

We reconstruct the English and Welsh telegraph network from *Zeitschrift des Deutsch-Österreichischen Telegraphen-Vereins, Jahrgang*, volume IX, 1862. This directory lists all telegraph stations outside of London in 1862. To the best of our knowledge, it is the most comprehensive list before the establishment of the transatlantic telegraph cable connecting the UK and the US (1866). We geo-reference all the stations and assign them to 1890-border districts. Since the source does not list stations in the London area, we conflate London urban districts into a single "London," which we assume to be connected to the telegraph network. Figure A.1 reports the distribution of the stations.

IV Empirical Strategy

In this section, we describe our empirical strategy. In subsection IV.A, we provide details on the construction of technology exposure in the United States. In subsection IV.B, we outline our baseline empirical specifications and discuss the potential endogeneity associated with them in subsection

[20] Differently from US counties, UK districts did not have budgetary or administrative authority.

IV.C. In subsection IV.D, we outline our identification strategy and the associated double and triple differences designs.

IV.A Measuring Exposure to American Technology

We construct two measures of exposure to US technology. The first is the number of emigrants to the United States from any given UK district. The underlying hypothesis is that districts with relatively more emigrants to the United States have higher exposure to US technology.

The second measure uses two margins of variation: along technological classes and along districts. First, local specialization across counties measures the knowledge that diffuses from those counties. Second, the number of migrants that leave a given district and settle in a given county measures the intensity of the return knowledge channel. To fix ideas, consider two districts and call them A and B. The same number of emigrants n leaves each district. Emigrants from A settle in county a, which only produces innovation in sector s_a. Emigrants from B settle in county b, which only innovates in sector s_b. Then, we expect district A (resp. B) to innovate comparatively more in sector s_a (resp. s_b).

We define knowledge exposure as:

$$\text{Knowledge Exposure}_{ik,t} \equiv \sum_{j \in J} \left(\frac{\text{Patents}_{jk,t}}{\text{Patents}_{j,t}} \times \text{Emigrants}_{i \to j,t} \right), \tag{1}$$

where i, j, k, and t denote a (UK) district, a (US) county, a technology class, and a decade, respectively.[21] The set J denotes the universe of counties. The knowledge exposure term averages district-level exposure to county-level specialization across technology classes.

The first term in the summation captures specialization, while the second term captures district-level exposure. The relative share of patents may inflate the influence of specialization in counties with a small number of granted patents. While this bias is unlikely to be large as those counties would likely have low district-level exposure, we code an alternative knowledge exposure variable that measures specialization as the raw count of patents in a given technology class. An additional challenge is that districts with larger bilateral linkages are probably larger and, hence, selected. To account for district-level time-varying confounding variables, we control non-parametrically for district-by-time fixed effects. However, we also report results for an alternative knowledge exposure that measures exposure through relative emigrant shares. We discuss these alternative definitions in more detail in the Appendix Table C.I.2.

[21] Throughout the paper, we refer to decade t to mean the ten years before the upper bound t. Hence, the decade indexed by 1890 refers to 1881–1890.

IV.B Empirical Specifications

We estimate two sets of regressions. First, we consider a specification that relates the volume of patenting activity in the UK with the number of transatlantic emigrants:

$$y_{i,t} = \alpha_i + \alpha_t + \beta \times \text{US Emigrants}_{i,t} + X_{i,t}\Gamma + \varepsilon_{i,t}, \tag{2}$$

where i and t denote, respectively, a district and a decade; $y_{i,t}$ is the outcome variable; α_i and α_t are district and decade fixed effects; $\text{US Emigrants}_{i,t}$ is the number of emigrants to the United States from district i between years $t-9$ and t; and $X_{i,t}$ is a set of district-level controls—population, employment shares across sectors, share of males—measured in 1880 and interacted with year dummies. Standard errors are clustered at the district level.

The second specification correlates patenting in the UK across technologies with the measure of exposure to US technology in (1):

$$y_{ik,t} = \alpha_{it} + \alpha_k + \beta \times \text{Knowledge Exposure}_{ik,t} + \varepsilon_{ik,t}, \tag{3}$$

where k denotes technology classes, the term α_{it} denotes district-by-decade fixed effects whose inclusion allows to control non-parametrically for time-varying unobserved heterogeneity at the district level; the term α_k denotes technology-fixed effects and excludes variation arising from factors that are technology-specific but common across areas and over time. Standard errors in this specification are clustered two-way at the district and technology level.

We consider three different sets of outcomes throughout the paper. First, we use the log(1+) number of patents to measure the volume of patenting activity. Second, to account for heterogeneous novelty across patents, we compute the log(1+) number of patents in the top 20% of the impact distribution of the measure constructed by Kelly *et al.* (2021) and applied to the corpus of UK patents. Third, to measure technology flows from the US into the UK, we compute the average text-based similarity between US and UK patents (Appendix A.II.6 provides additional details). In Appendix C.I.1, we present alternative specifications with different transformations and definitions of all the dependent variables and show that results remain qualitatively unchanged.

IV.C Potential Sources of Endogeneity

The main factor that cautions against a causal interpretation of the estimates of regression (2) is that out-migration is not random across districts. In particular, any unobserved variable correlating

with out-migration and the volume of innovation y induces bias in the estimated β coefficient. To address this issue, in (3), we include district-by-time fixed effects to partial out time-varying heterogeneity within districts. This allows us to exclusively leverage cross-technology variation. A remaining potential issue is assortative matching, meaning there may be a—possibly unobserved—variable that correlates with the location where emigrants settle in the United States and the composition of patenting activity across technology classes.

In section II.A, we discussed that the historical and empirical evidence suggests that, over time, emigrants originated from increasingly affluent and urbanized areas. Suppose emigrants also settled in comparatively more urban and affluent counties in the United States, and there was a correlation between patenting activity in specific fields and economic growth. In that case, the selection issue may bias the OLS estimates upward. We note that the bias arises only if (i) the correlation between patenting and the underlying confounding variable is heterogeneous across technology classes and (ii) the correlation is similar in the US and the UK. If (i) does not hold, then the omitted confounding variable would be absorbed by district-by-time fixed effects. If (ii) does not hold, the selection bias would work against our result.

Assortative matching may also arise if pre-existing differences in specialization across technology classes predicted the counties where emigrants chose to settle. For example, suppose that emigrants from a largely textile area, say Lancashire, were comparatively more likely to settle in counties with larger textile sectors. Then, the estimated β from equation (3) would reflect pre-existing innovation similarities between sending and settling areas rather than capture the effect of return innovation. Evidence by Hanlon (2018) and Ottinger (2020), among others, suggest that location decisions may not be random. In Appendix Section B.III, we show that British migrants in this period do not display a similar pattern.[22] However, ultimately, we cannot rule out assortative matching across US counties.

In Online Appendix section C.I.4, we implement two instrumental variable approaches following the migration literature to rule out that assortative matching is generating our results. We exploit conditionally exogenous variation in the location of British emigrants across the United States. This allows us to construct valid instruments for emigration and exposure to US technology through migration ties.[23]

[22] Appendix Table B.3 presents evidence against this pattern.

[23] The first instrument uses the expansion of the railway network in the United States over the course of the 1800s to predict county-level immigration, building on Sequeira *et al.* (2020). The second instrument builds on Card (2001), and constructs predicted county-level immigration by interacting county-specific inflows of British immigrants at baseline with aggregate immigration inflows in the United States.

IV.D Isolating the Effects of Exposure to American Technology: Double and Triple Differences Analysis

To estimate the causal impact of immigrants' exposure to US technology on innovation in the United Kingdom, we leverage shocks to US innovation activity to randomize the volume and content of innovation that emigrants are exposed to.

Suppose we observe a sudden increase in the number of patents granted in some counties in a technology class. Then, the "return innovation" effect would imply that districts with more emigrants to those counties would display increased innovation activity in that class. In other words, innovation shocks in the United States should "reverberate" in the United Kingdom through pre-existing migration linkages.

We label any period of unusually high patenting activity as a "shock" to US innovation. As in the previous analysis, shocks can pool technologies together or be technology-specific. In the former case, to isolate the shocks, we residualize the (log) number of patents against county and year-fixed effects and label as a shock any observation whose residualized patenting activity is in the top 5% of the overall distribution. In the latter case, we residualize patenting against a fully saturated set of county, technology, and year-fixed effects and similarly tag as a shock any observation whose residualized patenting activity is in the top 5% of the overall distribution. From the perspective of US counties, innovation shocks are periods of unusually high patenting activity. In Appendix Table C.12 and Appendix Figure C.12, we report the difference-in-differences estimated increase in patenting associated with the shocks. Patenting increases by 58% (in the pooled case) and by 13% (in the technology-specific instance) after a shock.

Let $\sigma_{j,t}$ be an indicator variable equal to one if county j undergoes an innovation shock in year t. Similarly, let $\sigma_{jk,t}$ be an indicator equal to one if the county-technology pair ik undergoes a shock in year t and zero otherwise. To construct a binary indicator of exposure to these shocks across British districts, we first compute the number of emigrants from district i that are exposed to a shock as $\sum_{j \in J} \left(\text{Emigrants}_{i \to j} \times \sigma_{j,t} \right)$, where the term $\text{Emigrants}_{i \to j}$ is the number of emigrants who moved from district i to county j in the decade 1870–1879, and we residualize the resulting series against district and year fixed effects. Then, we define a district as exposed to an innovation shock if the residual of this regression is in the top 5% of the overall distribution of residualized emigrants exposed to innovation shocks. Let the indicator variable ξ_i be equal to one in the first period when district i is exposed to a US innovation shock and zero otherwise.[24]

[24] The technology-specific case follows the same logic. We compute the number of emigrants from district i that, in year t, are exposed to US innovation shocks in technology k as $\sum_{j \in J} \left(\text{Emigrants}_{i \to j} \times \sigma_{jk,t} \right)$. We residualize the resulting

Given a set of district-level shocks ξ_i and district-technology-level shocks ξ_{ik}, we estimate double and triple differences regressions to study how exposure to US innovation shocks impacts innovation in the UK. In the case of district-level shocks, the regression is as follows:

$$y_{i,t} = \alpha_i + \alpha_t + \sum_{h=-a}^{b} \beta_h \times I\left[D_{i,t} = h\right] + \varepsilon_{i,t}, \tag{4}$$

where t now denotes years and $D_{i,t} \equiv (t - \xi_i)$ is the number of periods since district i is first exposed to a US innovation shock ξ_i and $I[\cdot]$ is an indicator variable which is equal to one when the argument is true and zero otherwise. In the case of technology-specific shocks, the regression is a triple-differences model that reads out as:

$$y_{ik,t} = \alpha_{ik} + \alpha_{it} + \alpha_{kt} + \sum_{h=-a}^{b} \beta_h \times I\left[D_{ik,t} = h\right] + \varepsilon_{ik,t}, \tag{5}$$

where α_{ik}, α_{it}, and α_{kt} denote, respectively, district-by-technology, district-by-year, and technology-by-year fixed effects, and $D_{ik,t} \equiv (t - \xi_{ik})$ denotes the number of years since the district-technology couple ik is first exposed to a US innovation shock ξ_{ik}.

Under a standard parallel trends assumption, in equations (4)–(5), the coefficients $\beta_{h \geq 0}$ estimate the dynamic treatment effects of exposure to US innovation shocks on outcome y. The parallel trends assumption requires that, in the absence of the shocks, the outcome variable y would not have differed across treated and untreated units. While this assumption is not testable, we show that, in all regressions, the estimates of the pre-treatment coefficients β_h for $h < 0$ are never statistically different from zero. This consistent pattern provides evidence in favor of the plausibility of the parallel trends assumption.

The roll-out of shocks across units is staggered: different districts and district-technology pairs may be exposed to US innovation shocks at different moments. Goodman-Bacon (2021) shows that, in these settings, the two-way fixed-effects estimator fails to estimate the average treatment effect when treatment effects are heterogeneous. In robustness check, we employ the estimator proposed by Sun and Abraham (2021) to deal with the staggered nature of the research design and confirm all the baseline results. Moreover, we show that the results are not sensitive to alternative thresholds to define the shocks to US innovation or the district-level exposure to such shocks. We discuss robustness exercises more in detail in Section V.

---

series against district, technology, and year-fixed effects, and we tag an observation as a shock if the residualized value is in the top 5% of the overall distribution of exposure. Then, let ξ_{ik} denote the associated indicator variable equal to one the first year when the district-technology pair ik is exposed to a US innovation shock and zero otherwise.

V Main Results

In this section, we document our main result: the "return innovation" effect. Section V.A provides descriptive evidence on the association between exposure to innovation in the United States through out-migration and innovation in the United Kingdom. In section V.B, we provide causal evidence of the return innovation effect. Finally, in section V.C, we document that the effect is stronger in sectors where the US and the UK have similar relative specialization.

V.A Exposure to Innovation in the United States and the Volume and Direction of Innovation in the United Kingdom

We begin by studying the association between out-migration to the United States and patenting activity in the United Kingdom. In Table II, column (1), we report the estimates of equation (2). The coefficient reports the association between the log(1+) number of patents and the number of US emigrants (in thousands) over time across registration districts. We find a positive and statistically significant association between the two variables: moving from the bottom 25th to the top 75th percentile of the distribution of emigrants is correlated with an 8% increase in patenting activity in the district of origin. The magnitude of the coefficient halves but remains statistically significant when we include district-level controls interacted with time fixed effects (column 2).

In column (3), we provide a first test that the shift in innovation activity may come from increased exposure to US innovation through emigration. The outcome variable is the average text-based similarity between US and British patents issued in a given district and decade. We find that out-migration to the US is associated with a large increase in the average similarity between British and American patents. Moving from the 25th to the 75th percentile of out-migration is linked to a doubling of average similarity. We interpret this result as the first piece of evidence in favor of the initial hypothesis that emigrants foster the transfer of technologies from the US into the UK.

In column (4), we report the results of estimating equation (2) on the log(1+) number of patents in the top 20% of impact. We adopt the methodology introduced by (Kelly *et al.*, 2021), and we apply it to the texts of British patents. This measure flags more "impactful" patents whose text is distant from existing patents at the time of issue but similar to those issued afterward. We use a 5-year window to compute this similarity metric, but the results are not sensitive to alternative bandwidths. Using this measure, we verify that the effect of emigration on innovation is not entirely composed of low-impact innovations. In fact, moving from the 25th to the 75th percentiles of out-migration is associated with an 11% increase in the number of breakthrough patents.

These results may not have a causal interpretation because of omitted factors that are time-varying within districts and correlate with both out-migration and innovation, as discussed in section IV.C. To address this concern, we adopt two instrumental variables approaches: a railway and a leave-out instrument for emigration and innovation exposure. The results, reported in Panels B and C of Table C.2, confirm the OLS estimates.

In columns (5–8) of Table II, we study the link between exposure to US technology and the direction of innovation in the United Kingdom. In this case, district-technology class pairs are observed at a decade frequency. The baseline regression, shown in column (5), includes district-by-time and technology class-fixed effects. The coefficient in equation (3) reports the association between changes in exposure to US innovation by technology and innovation in the United Kingdom. We uncover a positive and statistically significant correlation between innovation exposure and innovation. Moving from the 25th to the 75th percentile of innovation exposure is associated with an 11% increase in patenting activity in that technology class. In column (6), we saturate regression (3) with a full set of fixed effects and confirm that the correlation remains positive and statistically significant.

As with overall innovation, in column (7), we show that exposure to US technologies is associated with an increase in the similarity between British and American patents. Quantitatively, the association is large, as moving from the bottom 25th to the top 75th percentile of the main regressor is associated with a three-fold increase in the average similarity between UK and US patents. In column (8), we confirm that the association between exposure to US innovation increases and patenting remains positive, statistically significant, and quantitatively unchanged when the dependent variable only includes patents in the top 20% of the impact distribution.

As discussed in section IV.C, these estimates may not have a causal interpretation if there is an omitted term that predicts where British immigrants settled across the United States and is correlated with the direction of innovation in their areas of origin. We refer to this issue as "assortative matching." Table C.2 reports the IV estimates for the district-technology sample. These are consistent with the OLS results, suggesting that assortative matching is unlikely to drive the results.

Finally, we explore the long-run association between exposure to US technology through emigration and innovation. Using data on patenting activity between 1940 and 2015, we compute the association between knowledge exposure in the 1930s and innovation in the later twentieth century. The association between knowledge exposure and innovation remains stable until the 1960s. It starts to decline in the 1970s, but the pace accelerates in the 1980s, and it never recovers. We then repeat the exercise but report the time-varying correlation separately for each sector. The impact of exposure to US technology through emigration ties thus retains an enduring impact on innovation dynamics

in England and Wales, which persists well into the second half of the twentieth century. Appendix Section B.II provides a detailed discussion of this analysis.

V.B The Impact of Innovation Shocks in the United States on Innovation in the United Kingdom

In this section, we describe the results of estimating the double and triple differences research design described in Section IV.D. This enables us to draw a causal link between exposure to US innovation via migration ties and the volume and direction of innovation in the UK.

We begin by studying the impact of US innovation shocks on the volume of innovation in the UK. Column (1) of Table III reports the baseline estimate of regression (4), where we aggregate all pre- and post-treatment periods into two bins, and the treatment is equal to one after a given district is exposed to a US innovation shock and zero otherwise. After the district is exposed to a shock, patenting creases by approximately 9%. As shown in Table C.12, in the United States, an innovation shock is associated with a 58% increase in patenting activity in the county where it occurs. Hence, on average, emigration ties generate a 15% pass-through rate of innovation shocks from the US to the UK. Following a US innovation shock, the similarity between patents granted in exposed districts and American patents doubles (column 2). In addition, in column (3), we show that the number of patents in the top 20% of the impact distribution also increases—by 8.2%—in response to US innovation shock, implying that emigrants facilitate the transfer of high-quality patents from the US into the UK.

In columns (4–6), we report the estimates associated with regression (5), where again we conflate pre- and post-treatment periods into two single categories, and the treatment is an indicator that returns one after a district-technology class pair is exposed to a US innovation shock and zero otherwise. In this case, the regression includes the interactive fixed effects. Hence, identification hinges on within-district-technology exposure to US innovation shocks over time. The results confirm the previous findings. Innovation increases by 6% in districts exposed to an innovation shock. In this setting, emigrant ties generate a 50% pass-through rate of the increased patenting activity in the US into the UK. Patents in the same technologies and districts that are exposed to US innovation shocks also become more similar in their textual content to American patents. Finally, in column (6), we confirm that the results are quantitatively similar if we only include highly original patents in the dependent variable, as the number of patents in the top 20% of the impact distribution increases by 5% when units are exposed to the treatment.

Figure III reports the event-study estimates associated to model (4) (Panel IIIa) and (5) (Panel IIIb). In both cases, we find no evidence of statistically significant pre-treatment coefficients. In each Figure,

we test for the joint significance of all pre-treatment coefficients and find no statistical support for the alternative hypothesis that there is at least one significant pre-treatment coefficient. Overall, we interpret these patterns as evidence in favor of the parallel trends assumption stated in IV.D. After the treatment, the number of patents increases. In fact, essentially all post-treatment coefficients are highly statistically—and jointly—significant. The exposure to US innovation shocks has a persistent impact on UK patenting. As shown in Figure C.6, patenting activity in the US increases sharply after a shock, but it remains significantly above the pre-shock mean after the shock period. The persistent effect observed in the UK is consistent with this pattern.[25]

In Figure IV, we re-estimate models (4) (Panel IVa) and (5) (Panel IVb) using the similarity between British and American patents as the outcome variable. As before, we estimate no statistically significant pre-treatment coefficient, except a ten-year lead in IVa. The tests for the joint significance of all pre-treatment coefficients provide further evidence in favor of the validity of the pre-trend assumption. After the shock, British patents in exposed units become more similar to American patents. The effect persists until ten years after the shock occurs. As in the previous case, the effect of the shock exhibits substantial persistence because the shocks are themselves persistent within counties.

V.C Return Innovation and Relative Specialization

To conclude this section, we study how the magnitude of return innovation changes by technology. Anecdotal evidence suggests the US was relatively more specialized than the UK in agricultural and heavy machinery technologies even early in the sample period, whereas the UK was more specialized in chemical and textile innovation (David, 1966; Rosenberg, 1970). To measure relative specialization across technologies, 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 innovative than the US in terms of share of total patents, and below one otherwise (Balassa, 1965).[26] To study the heterogeneous responses to emigrants' exposure to US technology, we estimate equation (3) interacting knowledge exposure with a set of technology dummies. In Panel Va of Figure

[25] It is also plausible—although we do not test this hypothesis—that, after a shock, new technological advances are more salient to British immigrants, thereby amplifying the initial effect of the shock.

[26] In the international trade literature, the 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

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

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.

V, 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 uncover an inverted U-shaped relationship between the size of return innovation and specialization. Exposure to US innovation generates larger gains in technologies where the US and the UK have similar rates of specialization, such as electricity, scientific instruments, and engines. By contrast, the gains are considerably smaller, and often not significant, in areas where the US is more advanced, such as agriculture and metallurgy, and where the UK is at the frontier, specifically textiles. The Figure reports the coefficients of a regression between the coefficients, the RCA, and its squared values and confirms the existence of such a hump-shaped relationship. 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.

In Panel Vb, we repeat this exercise using the triple-differences regression (5). In particular, we interact the treatment with technology-specific dummy variables. Each dot reports, on the y-axis, the estimated treatment effect for one technology class. The coefficients are plotted against the measure of revealed comparative advantage. The resulting plot exhibits a pattern that largely resembles Panel Va. Exposure to US innovation shocks generates larger increases in patenting activity in the UK in fields with an RCA closer to one, namely, those where the UK and the US are equally active. By contrast, the treatment effect is zero—or negative, albeit negative estimates are never statistically significant—in fields where the US is more active than the UK, such as agriculture, and the reverse, such as textiles and chemistry.

V.D Robustness Checks

To conclude, we describe a series of checks that we perform to test the robustness of our results.

In the baseline analysis, our baseline measure of patenting is the log of the number of patents to which we add one to avoid dropping the zeros. In Tables C.1, C.2, and C.10, we show robustness to alternative transformations of the dependent variable for the OLS-2SLS (2)–(3) and the double and triple differences regressions (C.7)–(5) and find that all results remain qualitatively and quantitatively unchanged using the raw count, the inverse hyperbolic sine, or adding 0.1 instead of one to the log of the count. Moreover, Panels B and C replicate all specifications using the two instruments described in Appendix Section C.I.4. In Appendix Table C.5, the baseline equation (2) is estimated using the Poisson quasi-maximum likelihood estimator described in Correia, Guimarães and Zylkin (2020), which confirms the results.[27]

[27] We prefer a log-linear to a Poisson specification as the baseline because it allows us to test for robustness to potential concerns such as spatial autocorrelation, which would be hard to address with a Poisson regression.

In the baseline analysis, the text-based measures—similarity and originality—are computed over a five-year window around the issue year of the patent. In practice, this window implies that, when computing the originality of a given British patent, we compare it with those issued in the preceding and following five years. When computing the similarity between a British patent and other US patents, we consider those issued in the preceding five years. This is an arbitrary choice, hence in Tables C.1, C.2, and C.10, we consider two additional windows—one and ten years—and find that all the baseline results remain qualitatively unchanged.

In all specifications, when we look at district-level outcomes, standard errors are clustered at the district level. When the outcome is at the district-technology field level, we adopt two-way clustering by district and technology. In Figures C.1 and C.5, we use a set of alternative standard errors for both OLS-2SLS and double and triple differences regressions. In particular, we confirm that the results remain statistically significant when accounting for spatial autocorrelation as suggested by Conley (1999).

A possible concern with patents as a proxy for innovation is that they do not adequately reflect economically relevant innovations, i.e., technologies that are concretely adopted by firms. To address this concern, in Table C.3, we replicate the baseline findings of regressions (2)–(3) using as outcome variable only patents with a firm assignee. By law, the inventor of a patent in the UK—as in the US—must be an individual. However, many patents indicate when the inventor is employed in a firm or whether they are the owner. These patents provide an arguably more direct measure of firm innovation. All results hold when applying this sample restriction.

In Table C.4, we show that the baseline association between patenting and exposure to US technology remains qualitatively unchanged when using several alternative definitions of exposure, as detailed in C.I.2.

In the double and triple differences analysis, the roll-out of the innovation shocks across units in the double and triple differences models is staggered because they occur at different moments across districts and technologies. Following the literature, in Figure C.3, we replicate the baseline results obtained using the standard two-way fixed-effects estimator using the framework developed by Sun and Abraham (2021). We obtain very similar results when using the log(1+) number of patents (Panels C.3a–C.3b) and the similarity between British and American patents (Panels C.3c–C.3d) as outcome variables.

In Figure C.4, we replicate the baseline event studies but estimate the specifications (4)–(5) as Poisson quasi-maximum likelihood regressions. The estimates confirm the baseline results: we find no evidence

of statistically significant pre-treatment coefficients, and innovation increases after UK districts are exposed to shocks to US innovation.

Finally, in Section IV.D, we explain how we construct the treatment in the double and triple differences models. In both cases, we define a unit—either a district or a district-technology pair—as exposed to a US innovation shock in a given period if the number of migrants who are exposed to a US innovation shock from that district is in the top 5% of the overall distribution of exposed emigrants. The underlying logic is that a unit is exposed if a "sufficiently large" number of emigrants from that place is in turn exposed to a US innovation shock where they are settled. The 5% threshold is arbitrary. In Table C.11, we adopt two alternative definitions of the shocks (top 10% and top 1%). Intuitively, more restrictive thresholds should deliver larger treatment effects because they require more emigrants to be exposed to a US innovation shock to activate the treatment at the district level. The estimates confirm this insight: while most of the coefficients remain statistically significant irrespective of the threshold, the magnitude of the estimated treatment effect increases with more restrictive thresholds.

VI Mechanisms

After documenting the "return innovation" effect, we study its drivers. In section VI.A, we show that return migration accounts for approximately half of the overall effect. In section VI.B, we document that, even in the absence of return migration, emigrants impact the innovation activity of their social networks in the United Kingdom. In section VI.C, we study the introduction of the transatlantic telegraph cable and document that lower communication costs fostered return innovation. Additionally, we show that out-migration fostered a flow of information beyond the diffusion of new technologies, as local newspapers in the UK were more likely to mention the counties with high out-migration from the given UK district (section VI.D). Finally, in section VI.E, we discuss potential alternative mechanisms that cannot be tested empirically.

VI.A The Role of Return Migration

Return migration is a primary candidate for explaining our findings through two channels. First, return migrants may engage in innovation activities in the fields they were exposed to abroad. Second, return migrants may facilitate access to US knowledge without directly undertaking innovation activities. The literature does not offer conclusive evidence on the effect of return migration on innovation (e.g., Giorcelli, 2019; Ash, Cai, Draka and Liu, 2022).

The baseline linked sample of British emigrants traces them back to the UK census before they migrated. To measure return migration, we instead link them to UK censuses completed after they

had migrated to the US. Then, we aggregate return migration flows at the district-by-county level and at decade frequency. Analogously to equation (1), we compute a measure of "return knowledge exposure" as follows:

$$\text{Return Knowledge Exposure}_{ik,t} \equiv \sum_{j \in J} \left(\frac{\text{Patents}_{jk,t}}{\text{Patents}_{j,t}} \times \text{Return Migrants}_{j \to i,t} \right) \tag{6}$$

where $\text{Return Migrants}_{j \to i,t}$ is the number of migrants that return from county j to district i in decade t. Because UK censuses are available only until 1911, return migration data span 1870–1910.

Table IV reports the results of two classes of regressions which mirror specifications (2)–(3), except that we include returning US emigrants (columns 1–3) and the return knowledge exposure indicator (columns 4–6), at different lags. Throughout the specifications, the coefficients of US emigrants and the baseline knowledge exposure metric remain statistically significant. By contrast, the association between return migrants, return knowledge exposure, and patenting is less clear. It is mostly positive, as one would expect, but it is often not significant. Quantitatively, moving from the 25th to the 75th percentile of out-migration—resp. knowledge exposure—in column (1)—resp. column (4)—is associated with a 5.2%—resp. 7.1%—increase in patenting activity. By comparison, shifting by the same amount along the distribution of return emigration—resp. return knowledge exposure—yields a 3.8%—resp. 3.7%—increase in patenting. Overall, while we document a positive correlation between return migration and patenting, it appears that there remains substantial variation in innovation associated with out-migration that is left unexplained by return flows.

In the rest of the paper, we provide evidence of two additional mechanisms explaining the return innovation effect, which do not rely on return migration. First, we focus on the interactions between the emigrants and local communities in Britain. Second, we explore how information technologies interacted with migration ties to promote the diffusion of knowledge.

VI.B The Role of Social Networks

In this section, we study how emigrants impact the innovation activity of their social network in Britain. We focus on two factors promoting interactions between the emigrants and the non-migrant population: family ties and geographical proximity. This analysis builds heavily on experimental evidence from developing countries, which links the diffusion of technology to interactions over social networks (e.g., Bandiera and Rasul, 2006; Conley and Udry, 2010; Beaman *et al.*, 2021)

VI.B.1 Transatlantic Emigration Within Families

To reconstruct extended families from the census data, which only contains information on the members of the same household, we leverage the geographical distribution of surnames. Specifically, we assume that individuals with the same surname who live in geographical proximity—in the same county—are related. This assumption is reasonable as long as surnames are not too common: in this analysis, we, therefore, drop the top 5% most common surnames. Results are robust to alternative samples excluding the 1% or 10% most common surnames.

We study how US emigration impacts the innovation activity produced by the relatives of the migrants who do not move in a difference-in-differences setting. The treatment leverages variation in the surname of US emigrants by county. Formally, we estimate variations on the following regression:

$$y_{f,t} = \alpha_f + \alpha_t + \sum_{h=-a}^{b} \beta_h \times I[D_{f,t} = h] + \varepsilon_{f,t}, \tag{7}$$

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

First, we report the estimates of (7) using a pre-post indicator for treatment status in Panel A of Table V. Emigration within the family is associated with an increase in patenting (column 1), which remains unchanged if we include county-by-year fixed effects (column 2) and is larger the more migrants from the family move to the United States (column 3). Since patenting is a skewed outcome, in column (4), we use a binary indicator that confirms the results. Importantly, the patents granted to the relatives of those who migrate to the US become more textually similar to US patents (column 5). This provides further evidence pointing at information flows between migrants and their relatives in the UK. Finally, in column (6), using the text-based impact measure, we confirm that the increased patenting activity generated by within-family migration is not restricted to low-impact innovations. Figure VI–Panel VIa provides the associated event-study estimates. We find no evidence of statistically significant pre-treatment coefficients and a large increase in patenting, which builds up over the three years following the emigration of the family member.

We then distinguish between emigrants who return and those who do not. Emigrants could interact with their families upon returning, but they could also maintain ties while abroad. In Panel B, we restrict the treatment to emigrants who never return to the UK. The estimates confirm that even

when emigrants do not return, they nonetheless promote innovation within their families in the UK. Comparing the coefficients in Panel A and B, we find that the treatment effect of non-return migrants is about 70% of the average treatment effect, irrespective of return status. This difference is consistent with the descriptive evidence discussed in section VI.A and confirms that migrants promote innovation in their origin communities even when they do not return.

VI.B.2 Transatlantic Emigration Within Neighborhoods

Geographical 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 migrated to the US.

We leverage the granular nature of our data and perform an individual-level analysis. First, we extract all men aged between 18 and 50 in 1900 that do *not* emigrate from the 1891 census. We create a yearly balanced panel that reports the number of patents obtained by each individual between 1880 and 1900 using the linked inventor-census data described in Appendix A.II.7. We complement this data with information on the geographical proximity between these stayers and migrants to the US. Specifically, we define a variable $D_{p,t}$ that codes the periods since at least one individual living within k kilometers from non-migrant p migrates to the US and zero otherwise. In the baseline analysis, we consider a threshold k of five kilometers.[28]

The regression specification mirrors (7), except that the variable $D_{p,t}$ substitutes $D_{f,t}$. All regressions include individual and year-fixed effects, and in various specifications, we include district-by-time and parish-by-time fixed effects to control for time-varying heterogeneity at fine levels of spatial aggregation. Standard errors are clustered at the parish level because neighborhoods form subdivisions of parishes.

Table VI (Panel A) reports the results. Within-neighborhood migration to the United States has a positive and statistically significant impact on the innovation activity of stayers (column 1). Importantly, the effect remains when including parish-by-time fixed effects (column 3). In this case, the identifying variation consists of within-parish neighborhood-level emigration and ensures that we compare similar areas. In column (4), we show that conditional on having at least one emigrant in

[28] In Figure C.7, we show that the results are robust to alternative thresholds. In particular, the estimated treatment effect remains positive as in the baseline specification. Exceedingly restrictive thresholds, however, exclude a large number of treated individuals, and consequently, the estimated treatment effect is biased toward zero. Slack definitions of neighborhoods, by contrast, introduce measurement error, which reduces the precision of the estimates. We discuss these modeling choices in Appendix C.III.

the neighborhood, having more does not scale up the positive effect of emigration on innovation. In column (5), we use a binary indicator instead of the (log) number of patents, which confirms the baseline results. Importantly, in column (6), we find that those who are exposed to US emigrants through neighborhood ties produce patents that are more similar to those issued in the United States. This pattern confirms that neighborhood migrants plausibly transfer knowledge to their communities of origin. As in the rest of the paper, column (7) shows that neighborhood out-migration also promotes high-impact patenting. Figure VIb replicates these estimates in an event-study framework. We do not find statistically significant pre-treatment coefficients, even though these are not precisely estimated. The effect of neighborhood out-migration is largest four to five years after the treatment is first active and decreases thereafter.

In analogy to the family analysis, in Panel B, we restrict the attention to non-return migrants. The framework mirrors the previous specification, except that we now only include non-return migrants in the definition of the treatment $D_{f,t}$. We find that emigration within the neighborhood increases innovation even if the emigrant never comes back to the United Kingdom. In this case, the estimates in Panels A and B are largely overlapping; hence, the contribution of return migrants appears modest.

In Figure B.2, we estimate the baseline regression and interact the treatment variable with dummies for occupational categories of the inventor. We find substantial heterogeneity in response to neighborhood out-migration: treatment effects are larger for positively selected inventors working as entrepreneurs, utility workers—a catch-all term which comprises independent inventors—, chemistry and metallurgy workers, and engineers. By contrast, inventors employed in low-skilled occupations, such as agriculture and construction, do not react to exposure to within-neighborhood out-migration.

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

VI.C Information Technologies and Knowledge Flows

Next, we explore how migration ties interacted with information technologies to foster the flow of novel knowledge from the United States to the United Kingdom. We focus on the main information technology of the period: the telegraph.

The first transatlantic telegraphic cable connected the US and UK domestic networks in 1866. The telegraph represented a major revolution in communication technology that drove unprecedented mar-

ket integration (Steinwender, 2018). Before 1866, steam mail was the cheapest and fastest way to communicate between the UK and the US. It took seven to fifteen days to transmit information in this way. This delay was reduced to one day overnight between June 27 and 28, 1866. The connection timing was unanticipated and exogenous (Steinwender, 2018).[29]

Our hypothesis is that the telegraph represented a shock to information frictions across the ocean, facilitating the exchange of knowledge between emigrants and their communities of origin. The telegraph could also operate *indirectly* by facilitating the market integration between the US and the UK, thus further fostering knowledge flows (Aleksynska and Peri, 2014). To test these hypotheses, we estimate the following difference-in-differences models:

$$y_{i,t} = \alpha_i + \alpha_t + \sum_{h=-a}^{b} \beta_h \times [\text{US Emigrants}_i \times I(t-1866=h)] + \varepsilon_{i,t}, \tag{8a}$$

$$y_{ik,t} = \alpha_{ik} + \alpha_{it} + \alpha_{kt} + \sum_{h=-a}^{b} \beta_h \times [\text{Knowledge Exposure}_{ik} \times I(t-1866=h)] + \varepsilon_{ik,t}, \tag{8b}$$

where i, k, and t denote a district, technology class, and year, respectively. The terms (US Emigrants_i) and ($\text{Knowledge Exposure}_{ik}$) denote the number of US emigrants and exposure to US knowledge in the 1870s.[30] Finally, the term $I(t-1866=h)$ denotes the number of years since the transatlantic cable was laid down. In equation (8a), the treatment coefficients $\{\beta_{h\geq 0}\}$ quantify the effect of the transatlantic cable by comparing districts by the number of US emigrants; in equation (8b), we also leverage variation across sectors and exposure to US innovation.

We report the static versions that combine pre- and post-treatment years into two periods in column (1) of Table VII. Panel A reports the estimates of equation (8a), while Panel B refers to (8b). We estimate a positive and significant effect of the transatlantic telegraph on innovation. To provide more convincing evidence on the plausibility of this result, we test if the transatlantic cable affects innovation only in districts connected to the British domestic network.[31] We thus reconstruct the entire telegraph network before the introduction of the transatlantic cable. The exact location of each station is displayed in Appendix Figure A.1. In columns (3) and (4), we show that the positive effect

[29] The project for a transatlantic telegraphic cable had been underway for a long time before 1866. Previous attempts in 1857, 1858, and 1865 all failed due to logistic and technical challenges. The 1866 attempt was thus one among many, and its success had not been anticipated.

[30] The cable was laid down in 1866. Our migration data started in 1870. To construct district-level emigration, we can only use emigrants from 1870–1875. This would be problematic if the telegraph fostered out-migration, which, by available historical accounts, was not the case.

[31] We do not claim that there were no cross-district spillover effects even if districts were not connected to the domestic UK network. We nonetheless believe the effect on connected districts would arguably be more significant.

of the telegraph is exclusively driven by districts connected to the domestic UK network. We fail to detect any significant effect on non-connected districts. In column (5), we document that British patents became more similar to US patents after the telegraph connected the two countries. The quantitative effects are modest: moving from the 25th to the 75th percentile of US emigration yields a 2.9% increase in the number of patents produced after 1866. In connected districts, the magnitude amounts to a 3.6% increase. The same shift increases the similarity of British and American patents by approximately 7%.

Because the location of telegraph stations was not random, one may argue that this exercise only reflects pre-existing differences between connected and non-connected districts. Identification in this setting requires that patenting in connected and unconnected districts was on the same trend before the introduction of the telegraph and that it would not have differed had the cable not been laid down. In Figure VII, we report dynamic double-differences estimates of equation (8b), which we estimate separately on connected and unconnected districts. We find that connected and unconnected districts were on the same trend before 1866. We estimate positive and significant treatment effects only for the former and after 1866, whereas the patenting in the latter does not respond to the shock. In 1873 and 1874, the second and third cables became operational.

VI.D Migration Networks and Information Flows Beyond Innovation: The Coverage of American News in British Newspapers

Until this point, we have restricted the focus of the analysis to information flows related to new technologies. In this section, we document that migration ties between the UK and the US generated general-purpose information flows beyond those directly related to innovation.

We rely on the British Newspaper Archive, which contains the digitized contents of thousands of historical British newspapers (for a discussion of the dataset, see Beach and Hanlon, 2023). To proxy for information flows between the US and the UK, we construct a measure of coverage of US-related information flows in UK newspapers by counting the references to US-related words.

We estimate three sets of regression equations:

$$\text{US Mentions}_{i,t} = \alpha_i + \alpha_t + \beta_1 \times \text{US Emigrants}_{i,t} + \varepsilon_{i,t}, \tag{9a}$$

$$\text{US State Mentions}_{is,t} = \alpha_{is} + \alpha_t + \beta_2 \times \text{US Emigrants}_{i \to s,t} + \varepsilon_{is,t}, \tag{9b}$$

$$\text{US County Mentions}_{ij,t} = \alpha_{ij} + \alpha_t + \beta_3 \times \text{US Emigrants}_{i \to j,t} + \varepsilon_{ij,t}, \tag{9c}$$

where i, j, s, and t denote a UK district, a US county, a US state, and a decade, respectively.

Regression (9a) is run at the district level and leverages variation in the number of emigrants to the United States; in regressions (9b) (resp. (9c)), instead, we look at district-by-state (resp. district-by-county) migration flows.

Table VIII reports the results. Columns (1–3), (4–5), and (7–8), respectively display the estimated β_i coefficients of estimated equations (9a), (9b), and (9c). Columns (1), (4), and (7) report the effect on the log(1+) number of mentions of "United States," single states, and single counties. In columns (2), (5), and (8), we exclude from the sample all districts where no single newspaper is based in the sample period. Columns (3), (6), and (9) display the association between emigration and the number of US-related topics normalized by (1+) the number of active newspapers in the district. Appendix Table C.14 shows that the associated estimates obtained using the instrumental variables described in Section C.I.4 are consistent with the OLS estimates.

We find a positive effect of out-migration on newspaper coverage of general-interest US-related news. We interpret this pattern as evidence that out-migration generates information flows between the areas where emigrants settle and where they originate. We cannot disentangle—-and this goes beyond the scope of this paper—the precise underlying mechanism. For example, increased coverage of US-related news may be demand-driven because the local population may demand information covering the areas where their relatives settled. On the other hand, US emigrants could have sponsored local newspapers to cover news about the areas where they were located. In this sense, our estimates may reflect a supply-side factor. Taken together, these results show that out-migration fostered cross-country information flows.

VI.E Alternative Mechanisms

To conclude this section, we discuss additional mechanisms that may explain the return innovation result, which we cannot directly test in the paper.

Temporary Migrations In our study of the mechanisms behind the return innovation effect, we concluded that physical return is an important determinant but not the exclusive driver. It is possible, however, that short-term temporary migrations influence the dynamics of innovation in the UK. Since we construct migration flows from population censuses, which are conducted every ten years, our data cannot identify temporary migrants. For the same reasons, we cannot quantify the importance of industrial espionage.

However, we believe that the quantitative effects of temporary migrations and industrial espionage are modest. First, the notion of a "temporary migrant" in XIX-century transatlantic migration is not

well-defined. Piore (1980) refers to Southern and Eastern European migrants as temporary because they planned to return to their origin countries at some point. This spell could take, however, decades. For example, a one-way cabin travel ticket from New York to Liverpool, at roughly 100$, would cost as much as 20% of the average annual US income. The extent of short-term stays must have been, therefore, relatively limited. Moreover, Piore (1980) notes that "temporary" migrants were relatively low-skilled and, therefore, less likely to operate technology transfer. Industrial espionage does not appear to be quantitatively sufficient to generate the return innovation effect that we estimate. Finally, our research designs speak against the temporary migration and the industrial espionage mechanisms. For temporary migration or industrial espionage to explain the double and triple differences result, in fact, one would need such flows to be correlated with the county-level innovation shocks. This channel seems unlikely, although it cannot be directly tested and refuted due to the absence of data.

Monetary Remittances Along with classical "brain drain" arguments, monetary remittances have been a major subject of investigation in the literature. While the contribution of remittances to economic development appears to be modest, it is possible that this inflow of capital may have sustained increased innovation, perhaps by relaxing constraints in access to credit (Gorodnichenko and Schnitzer, 2013). This capital inflow, however, could explain why out-migration influences the direction of innovation unless knowledge *and* monetary remittances go hand in hand. Disaggregated data on financial remittances, unfortunately, do not exist. Hence, this is a possibility that we cannot explore. It nonetheless highlights that, if anything, financial and "knowledge" remittances are likely to shape innovation in a complementary way rather than in a mutually exclusive one.

VII Conclusions

Previous literature has shown that the diffusion of innovation across countries is a major driver of long-run growth and economic convergence. However, evidence on *how* innovation diffuses across countries remains limited. In this paper, we provide the first causal evidence that out-migration enables the diffusion of innovation from the country of destination toward the country of origin of migrants.

Our analysis focuses on the English and Welsh mass migration to the United States between 1870 and 1940. We link the individual-level census records of British immigrants in the US to the UK population census to measure spatially disaggregated emigration flows. We complement these data with newly digitized patents that, added to existing data, allow us to cover the universe of patents granted in England and Wales in the nineteenth and twentieth centuries. The absence of stringent international intellectual property protection allows us to use patents as a measure of the diffusion of knowledge across countries.

Using a double and triple differences research design, we document that exposure to foreign technology through migration ties contributes to its diffusion in the country of origin of migrants. First, we document that the volume of innovation in the UK increases in response to higher exposure to US innovation through emigration. Second, innovation activity in the UK shifts towards technologies that emigrants are most exposed to in the US. By looking at the textual similarity between UK and US patents, we find that exposure to US knowledge stimulates technology transfer. Additionally, we show that this diffusion involves both low and high-impact UK innovations. We define this phenomenon "return innovation."

We then turn to studying the drivers of return innovation. The physical return of migrants explains approximately half of the effect. However, we find that interactions between emigrants and their communities of origin represent another important channel of technology diffusion, even in the absence of return migration. This is further supported by the fact that information technologies, such as the telegraph, facilitated knowledge flows between emigrants' destination areas in the US and their region of origin in the UK.

While the setting of this paper is historical, our results are likely to have implications in contemporaneous settings. In fact, our findings echo qualitative evidence by Saxenian (1999, 2006), who documents similar return innovation effects in the Taiwanese and Indian emigration to the Silicon Valley area at the end of the 20th century. Taken together, our results document that emigration does not necessarily lead to underdevelopment or stagnation, as suggested by the "brain drain" hypothesis (Docquier and Rapoport, 2012), but it can foster the diffusion of innovation, which is a leading driver of economic growth. The return innovation and brain drain effects are likely to coexist, and which prevails is likely to be a function of the type of migration (skilled or unskilled), as well as the relative position of the origin and destination countries in the technology frontier. Beyond the relative size of these two effects, our results suggest that fostering economic integration between the country of origin of migrants and their country of emigration is likely to promote knowledge exchange, including return innovation.

Finally, given the recent advances in information and communication technologies, a natural question is whether emigrants' physical exposure to foreign technology is still necessary to promote the diffusion of innovation. While we cannot speak directly to the substitutability of in-person interactions, the existing literature shows that physical proximity is still a major driver of the production and diffusion of knowledge.[32]

[32] Among others, Akcigit *et al.* (2018) highlight the importance of interactions in shaping the career trajectories of inventors. Prato (2022) shows that immigrant inventors retain ties with their colleagues in their origin countries, fostering technology transfer and productivity. Boudreau, Brady, Ganguli, Gaule, Guinan, Hollenberg and Lakhani (2017) document that

References

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

Abramitzky, R., L. Boustan, K. Eriksson, J. Feigenbaum and S. Pérez (2021). "Automated Linking of Historical Data." *Journal of Economic Literature*, 59(3): 865–918.

Abramitzky, R., L. P. Boustan and K. Eriksson (2014). "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration." *Journal of Political Economy*, 122(3): 467–506.

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

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

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

Akcigit, U., S. Caicedo, E. Miguelez, S. Stantcheva and V. Sterzi (2018). "Dancing With the Stars: Innovation Through Interactions." *NBER Working Paper*, (No. w24466).

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

Aleksynska, M. and G. Peri (2014). "Isolating the Network Effect of Immigrants on Trade." *The World Economy*, 37(3): 434–455.

Alvarez, F. E., F. J. Buera and R. E. Lucas (2013). "Idea Flows, Economic Growth, and Trade." *NBER Working Paper*, (No. w19667).

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

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

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

Arkolakis, C., S. K. Lee and M. Peters (2020). "European Immigrants and the United States Rise to the Technological Frontier." *Working Paper*.

Ash, E., D. Cai, M. Draka and S. Liu (2022). "Bootstrapping Science? The Impact of a "Return Human Capital" Programme on Chinese Research Productivity." *Working Paper*.

Atkin, D., M. K. Chen and A. Popov (2022). "The Returns to Face-to-Face Interactions: Knowl-

search costs affect the probability of forming scientific collaborations. Atkin, Chen and Popov (2022) provide direct evidence that face-to-face interactions promote the diffusion of knowledge and agglomeration economies.

edge Spillovers in Silicon Valley." *NBER Working Paper*, (No. w30147).

Bahar, D., P. Choudhury, J. Sappenfield and S. Signorelli (2022). "Talent Flows and the Geography of Knowledge Production: Causal Evidence from Multinational Firms." *Working Paper*.

Bahar, D., A. Hauptmann, C. Özgüzel and H. Rapoport (2019). "Migration and Knowledge Diffusion: The Effect of Returning Refugees on Export Performance in the Former Yugoslavia." *The Review of Economics and Statistics*, pp. 1–50.

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

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

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

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

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

Beach, B. and W. W. Hanlon (2023). "Historical Newspaper Data: A Researcher's Guide." *Explorations in Economic History*, 90: 101541.

Beaman, L., A. BenYishay, J. Magruder and A. M. Mobarak (2021). "Can Network Theory-Based Targeting Increase Technology Adoption?" *American Economic Review*, 111(6): 1918–1943.

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

Bell, A., R. Chetty, X. Jaravel, N. Petkova and J. Van Reenen (2019). "Who Becomes an Inventor in America? The Importance of Exposure to Innovation." *The Quarterly Journal of Economics*, 134(2): 647–713.

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

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

Bertoli, S. and F. Marchetta (2015). "Bringing It All Back Home–Return Migration and Fertility Choices." *World Development*, 65: 27–40.

Bottomley, S. (2014). *The British Patent System during the Industrial Revolution 17001852: From Privilege to Property*. Cambridge (UK): Cambridge University Press.

Boudreau, K. J., T. Brady, I. Ganguli, P. Gaule, E. Guinan, A. Hollenberg and K. R.

LAKHANI (2017). "A Field Experiment on Search Costs and the Formation of Scientific Collaborations." *Review of Economics and Statistics*, 99(4): 565–576.

BRYAN, K. A. and J. LEMUS (2017). "The Direction of Innovation." *Journal of Economic Theory*, 172: 247–272.

BUERA, F. J. and E. OBERFIELD (2020). "The Global Diffusion of Ideas." *Econometrica*, 88(1): 83–114.

BURCHARDI, K. B., T. CHANEY, T. A. HASSAN, L. TARQUINIO and S. J. TERRY (2020). "Immigration, Innovation, and Growth." *NBER Working Paper*, (No. w27075).

CARD, D. (2001). "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration." *Journal of Labor Economics*, 19(1): 22–64.

CHAUVET, L. and M. MERCIER (2014). "Do Return Migrants Transfer Political Norms to Their Origin Country? Evidence from Mali." *Journal of Comparative Economics*, 42(3): 630–651.

COLUCCIA, D. M. and L. SPADAVECCHIA (2024). "Emigration Restrictions and Economic Development: Evidence from the Italian Mass Migration to the United States." *Working Paper*.

COMIN, D. and B. HOBIJN (2010). "An Exploration of Technology Diffusion." *American Economic Review*, 100(5): 2031–2059.

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

CONLEY, T. G. and C. R. UDRY (2010). "Learning About a New Technology: Pineapple in Ghana." *American Economic Review*, 100(1): 35–69.

CORREIA, S., P. GUIMARÃES and T. ZYLKIN (2020). "Fast Poisson Estimation with High-Dimensional Fixed Effects." *The Stata Journal*, 20(1): 95–115.

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

DAVID, P. A. (1966). "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.

DOCQUIER, F. and H. RAPOPORT (2012). "Globalization, Brain Drain, and Development." *Journal of Economic Literature*, 50(3): 681–730.

DOSSI, G. (2024). "Race and Science." *Working Paper*.

DUSTMANN, C., T. FRATTINI and A. ROSSO (2015). "The Effect of Emigration from Poland on Polish wages." *The Scandinavian Journal of Economics*, 117(2): 522–564.

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

EATON, J. and S. KORTUM (1999). "International Technology Diffusion: Theory and Measurement." *International Economic Review*, 40(3): 537–570.

Einiö, E., J. Feng and X. Jaravel (2023). "Social Push and the Direction of Innovation." *Working Paper*.

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

——— (1972). *Who were the English and Scots emigrants in the 1880s?*, pp. 87–125. Arnold.

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

Goodman-Bacon, A. (2021). "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics*, 225(2): 254–277.

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

Griffith, R., R. Harrison and J. Van Reenen (2006). "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*, 96(5): 1859–1875.

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

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

——— (2016). "British Patent Technology Classification Database: 1855-1882." Unpublished [data collection].

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

Helgertz, J., J. Price, J. Wellington, K. J. Thompson, S. Ruggles and C. A. Fitch (2022). "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*, 55(1): 12–29.

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

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

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

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

Jeremy, D. J. (1981). *Transatlantic Industrial Revolution: The Diffusion of Textile Technologies Between Britain and America, 1790-1830s*. Cambridge (MA): MIT Press.

KARADJA, M. and E. PRAWITZ (2019). "Exit, Voice, and Political Change: Evidence from Swedish Mass Migration to the United States." *Journal of Political Economy*, 127(4): 1864–1925.

KELLY, B., D. PAPANIKOLAOU, A. SERU and M. TADDY (2021). "Measuring Technological Innovation Over the Long Run." *American Economic Review: Insights*, 3(3): 303–320.

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

KHAN, B. Z. and K. L. SOKOLOFF (2004). *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.

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

MACLEOD, C. (1988). *Inventing the Industrial Revolution.* Cambridge (UK): Cambridge University Press.

MOSCONA, J. (2021). "Environmental Catastrophe and the Direction of Invention: Evidence from the American Dust Bowl." *Working Paper.*

MOSCONA, J. and K. A. SASTRY (2023). "Does Directed Innovation Mitigate Climate Damage? Evidence from US Agriculture." *The Quarterly Journal of Economics*, 138(2): 637–701.

MOSER, P. (2019). "Patents and Innovation in Economic History." In "Research Handbook on the Economics of Intellectual Property Law," pp. 462–481. Edward Elgar Publishing.

MOSER, P., S. PARSA and S. SAN (2020). "Immigration, Science, and Invention. Evidence from the Quota Acts." *Working Paper.*

MOSER, P., A. VOENA and F. WALDINGER (2014). "German Jewish émigrés and US invention." *American Economic Review*, 104(10): 3222–55.

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

NUVOLARI, A. and V. TARTARI (2011). "Bennet Woodcroft and the Value of English Patents, 1617–1841." *Explorations in Economic History*, 48(1): 97–115.

OTTINGER, S. (2020). "Immigrants, Industries, and Path Dependence." *Working Paper.*

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

PERLA, J., C. TONETTI and M. E. WAUGH (2021). "Equilibrium Technology Diffusion, Trade, and Growth." *American Economic Review*, 111(1): 73–128.

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

PRATO, M. (2022). "The Global Race for Talent: Brain Drain, Knowledge Transfer, and Economic Growth." *Working Paper.*

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

——— (1982). "The International Transfer of Technology: Implications for the Industrialized Countries." In "Inside the Black Box: Technology and Economics," New York: Cambridge University Press.

RUGGLES, S., C. FITCH, R. GOEKEN, J. HACKER, M. NELSON, E. ROBERTS, M. SCHOUWEILER and M. SOBEK (2021). "IPUMS ancestry full count data: Version 3.0 [dataset]." *Minneapolis, MN: IPUMS*.

SAXENIAN, A. (1999). *Silicon Valley's New Immigrant Entrepreneurs*. Public Policy Institute of California.

——— (2006). *The New Argonauts: Regional Advantage in a Global Economy*. Harvard University Press.

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

SEQUEIRA, S., N. NUNN and N. QIAN (2020). "Immigrants and the Making of America." *The Review of Economic Studies*, 87(1): 382–419.

SPILIMBERGO, A. (2009). "Democracy and Foreign Education." *American Economic Review*, 99(1): 528–43.

STEINWENDER, C. (2018). "Real Effects of Information Frictions: When the States and the Kingdom Became United." *American Economic Review*, 108(3): 657–96.

SUN, L. and S. ABRAHAM (2021). "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics*, 225(2): 175–199.

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

TRUFFA, F. and A. WONG (2024). "Undergraduate Gender Diversity and the Direction of Scientific Research." *CESifo Working Paper*.

TUCCIO, M. and J. WAHBA (2018). "Return Migration and the Transfer of Gender Norms: Evidence from the Middle East." *Journal of Comparative Economics*, 46(4): 1006–1029.

VAN PATTEN, D. (2023). "International Diffusion of Technology: Accounting for Heterogeneous Learning Abilities." *Working Paper*.

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

Tables

Table I. Descriptive Statistics

	Mean	Std. Dev.	Min	Max	Units	Obs.
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Demographics (in 1880)						
Population$_i$ (1,000s)	39.988	46.183	0.184	310.706	621	621
Share of Men$_i$ (%)	49.238	2.124	38.662	56.616	621	621
Share of Employment$_i$ (%)	61.091	7.240	0.000	74.427	621	621
in Agriculture$_i$ (%)	14.406	6.892	1.989	31.117	621	621
in Construction$_i$ (%)	3.747	1.597	1.420	20.618	621	621
in Engineering$_i$ (%)	3.818	1.483	0.848	11.843	621	621
in Public and Liberal Occupations$_i$ (%)	2.573	1.674	0.819	16.045	621	621
in Other Manufacturing$_i$ (%)	3.289	2.772	0.618	19.684	621	621
in Textiles$_i$ (%)	4.996	5.135	0.736	34.390	621	621
in Trade$_i$ (%)	1.612	1.003	0.372	9.783	621	621
in Transportation$_i$ (%)	2.382	1.340	0.438	13.857	621	621
Panel B. Emigration to the United States						
US Emigrants$_{id}$	105.783	148.023	0.000	1752.157	621	3726
Return US Emigrants$_{id}$	11.845	17.108	0.000	185.591	621	2484
Panel C. Patents						
Total Patents$_{it}$	21.355	68.158	0.000	2088.000	621	31671
Patents by Technology Class:						
in Electricity$_{it}$	1.855	14.724	0.000	1084.000	621	31671
in Instruments$_{it}$	1.777	7.349	0.000	217.000	621	31671
in Lighting and Heating$_{it}$	1.486	5.487	0.000	145.000	621	31671
in Personal Goods$_{it}$	1.810	6.573	0.000	170.000	621	31671
in Transportation$_{it}$	3.499	10.761	0.000	244.000	621	31671
Panel D. Linked Inventor-Census Sample						
Number of Patents$_{it}$	0.121	0.842	0.000	274.000	110532	2210640
Age$_i$	37.104	13.729	9.000	79.000	110532	110532
Employed in:						
Agriculture$_i$	0.133	0.339	0.000	1.000	110532	110532
Construction$_i$	0.140	0.347	0.000	1.000	110532	110532
Engineering$_i$	0.164	0.371	0.000	1.000	110532	110532
Trade$_i$	0.089	0.285	0.000	1.000	110532	110532
Transportation$_i$	0.092	0.289	0.000	1.000	110532	110532
Neighborhood Emigrants$_i$	2.295	4.081	0.000	53.424	110532	110532
Non-Return Neighborhood Emigrants$_i$	2.257	4.012	0.000	52.677	110532	110532

Notes. This Table lists descriptive statistics for the variables used in the main analysis at the district (Panel A), district-decade (Panel B), district-year (Panel C), and individual level (Panel D). Demographic data in Panels A and D are tabulated from the population census. Panel A refers to the 1880 census. Emigration data in Panels A and D are computed from the linked migrant sample. Demographic data are cross-walked to consistent 1891 district borders. The subscripts report the aggregation level: Panel A by district i; Panel B by district-decade id; Panel C by district-year it; Panel D by individuals i. Referenced on page(s) 11, 13.

TABLE II. Exposure to Innovation in the United States and the Volume and Direction of Innovation in the United Kingdom

	Patents by District				Patents by District-Technology			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Number	Number	US Patents Similarity	High Impact	Number	Number	US Patents Similarity	High Impact
Dependent Variable Mean	*3.518*	*3.518*	*162.555*	*2.042*	*1.006*	*1.006*	*74.488*	*0.388*
US Emigrants (1,000s)	0.882***	0.488**	24.081***	1.204***				
	(0.185)	(0.222)	(7.067)	(0.221)				
Knowledge Exposure					0.023***	0.006***	0.231***	0.022***
					(0.002)	(0.001)	(0.058)	(0.001)
District FE	Yes	Yes	Yes	Yes	–	–	–	–
Decade FE	Yes	–	Yes	Yes	–	–	–	–
Controls × Time	No	Yes	No	No	–	–	–	–
County-Year FE	No	Yes	No	No	–	–	–	–
District-Year FE	–	–	–	–	Yes	Yes	Yes	Yes
Technology FE	–	–	–	–	Yes	–	Yes	Yes
District-Technology FE	–	–	–	–	No	Yes	No	No
Technology-Year FE	–	–	–	–	No	Yes	No	No
Number of Districts	621	620	621	621	621	621	621	621
Observations	3,726	3,720	3,726	3,726	70,433	70,433	70,433	70,433
Std. Beta Coef.	0.065	0.036	0.037	0.097	0.185	0.050	0.024	0.292

Notes. This Table reports the association between emigration to the United States and innovation in the United Kingdom. In columns (1–4) (resp. 5–8), the unit of observation is a district (resp. district-technology pair) observed at a decade frequency between 1870 and 1930. In columns (1–2) and (5–6), the dependent variable is the log(1+) number of patents. In columns (3) and (7), the dependent variable is the average text-based similarity between UK and US patents issued in the previous five years. In columns (4) and (8), the dependent variable is the log(1+) of the number of patents in the top 20% of the impact distribution ("Breakthrough" patents). In columns (1–4), "US Emigrants (1,000s)" is the number of migrants to the United States from the given district and decade, in thousands. In columns (5–8), "Knowledge Exposure" is defined in equation (1). Columns (1) and (3–4) include district and decade fixed effects. Column (2) includes district fixed effects, district-level controls measured in 1880 and interacted with decade indicators, and county-by-decade fixed effects. Columns (5) and (7–8), include district-by-year and technology fixed effects. Column (6) includes district-by-year, district-by-technology, and technology-by-year fixed effects. Standard errors, reported in parentheses, are clustered at the district level. Referenced on page(s) 20, 21. *: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

TABLE III. The Effect of Shocks to Innovation in the United States on Innovation in the United Kingdom

	Patents by District			Patents by District-Technology		
	(1) Number	(2) US Patents Similarity	(3) High Impact	(4) Number	(5) US Patents Similarity	(6) High Impact
Dependent Variable Mean	*1.687*	*121.561*	*0.736*	*0.292*	*30.124*	*0.082*
Post × US Innovation Shock	0.091**	18.523***	0.082*	0.064***	8.760***	0.020***
	(0.045)	(3.473)	(0.043)	(0.009)	(1.145)	(0.006)
District FE	Yes	Yes	Yes	–	–	–
Year FE	Yes	Yes	Yes	–	–	–
District-Year FE	–	–	–	Yes	Yes	Yes
Technology-Year FE	–	–	–	Yes	Yes	Yes
District-Technology FE	–	–	–	Yes	Yes	Yes
Number of Districts	621	621	621	621	621	621
Observations	31,671	31,671	31,671	601,749	601,749	601,749

Notes. This Table reports the effect of shocks to US innovation activity on innovation in the United Kingdom. In columns (1–3), the unit of observation is a district observed at a yearly frequency between 1870 and 1930. In columns (4–6), the unit of observation is a district-technology pair observed over the same period. In columns (1) and (4), the dependent variable is the log(1+) number of patents. In columns (2) and (5), the dependent variable is the average text-based similarity between UK patents and US patents issued in the previous five years. In columns (3) and (6), the dependent variable is the log(1+) number of patents in the top 20% of the impact distribution ("Breakthrough" patents). "Post" is an indicator equal to one for all years after the observation unit is exposed to a shock to US innovation activity, equal to zero otherwise. "US Innovation Shock" is defined in Section IV.D. Columns (1–3) include district and year fixed effects. Columns (4–6) include district-by-year, technology-by-year, and district-by-technology fixed effects. Standard errors, reported in parentheses, are clustered by district. Referenced on page(s) 22. *: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

TABLE IV. Return Migration and Innovation in the United Kingdom

	Patents by District			Patents by District-Technology		
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable Mean	*3.553*	*3.898*	*3.738*	*1.011*	*1.170*	*1.090*
US Emigrants$_t$ (1,000s)	0.540***		-0.101			
	(0.161)		(0.284)			
US Emigrants$_{t-1}$ (1,000s)		0.735***	0.479***			
		(0.157)	(0.167)			
Return US Emigrants$_t$ (1,000s)	4.050***		2.723			
	(1.394)		(1.915)			
Return US Emigrants$_{t-1}$ (1,000s)		0.495	2.623			
		(1.177)	(2.057)			
Knowledge Exposure$_t$				0.012**		0.026**
				(0.005)		(0.012)
Knowledge Exposure$_{t-1}$					0.012***	0.003
					(0.004)	(0.003)
Return Knowledge Exposure$_t$				0.671***		0.043
				(0.216)		(0.191)
Return Knowledge Exposure$_{t-1}$					0.594*	-0.390
					(0.285)	(0.471)
District FE	Yes	Yes	Yes	–	–	–
Decade FE	Yes	Yes	Yes	–	–	–
District-Year FE	–	–	–	Yes	Yes	Yes
Technology FE	–	–	–	Yes	Yes	Yes
Number of Districts	621	621	621	621	621	621
Observations	2,484	2,484	1,863	47,196	47,196	35,397

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 decade frequency between 1880 and 1910 (columns 1–2 and 4–5) and between 1880 and 1900 (columns 3 and 6). Different sample periods depend on the lag structure. In all columns, the dependent variable is the log(1+) number of patents. "US Emigrants (1,000)" is the number of emigrants to the US in the current decade or in the previous one $(t-1)$, in thousands. "Return US Emigrants (1,000s)" is the number of emigrants to the US who return to the UK in the current decade or in the previous one $(t-1)$, in thousands. "Knowledge Exposure" is constructed following equation (1) and refers to the current decade or the previous one $(t-1)$. "Return Knowledge Exposure" is defined similarly to Knowledge Exposure, but only look at flows of emigrants who eventually return to the UK. Columns (1–3) include district and decade-fixed effects. Columns (4–6) include district-by-time and technology-fixed effects. Standard errors, reported in parentheses, are clustered by district. Referenced on page(s) 27. *: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

TABLE V. Within-Family Migration to the United States and Innovation in the United Kingdom

	Number of Patents				Text-Based Measures	
	(1)	(2)	(3)	(4)	(5)	(6)
	Number	Number	Number	I(Patents > 0)	US Patents Similarity	High Impact
Panel A. Family Emigration						
Dependent Variable Mean	*4.919*	*4.919*	*4.919*	*5.835*	*6.612*	*1.229*
US Emigrant in Family × Post	1.375***	1.505***	-0.312	1.303***	3.873***	0.452***
	(0.363)	(0.314)	(0.218)	(0.402)	(0.567)	(0.118)
US Emigrant in Family × Post × N. of Emigrants			1.299***			
			(0.316)			
Number of Families	46,864	46,864	46,864	46,864	46,864	46,864
Observations	2,858,704	2,858,704	2,858,704	2,858,704	2,858,704	2,858,704
Panel B. Family Non-Return Emigration						
Dependent Variable Mean	*4.919*	*4.919*	*4.919*	*5.835*	*6.612*	*1.229*
Non-Return US Emigrant in Family × Post	1.030***	1.152***	-1.181**	0.970**	3.065***	0.306***
	(0.333)	(0.281)	(0.448)	(0.373)	(0.497)	(0.092)
Non-Return US Emigrant in Family × Post × N. of Emigrants			1.813***			
			(0.626)			
Number of Families	44,879	44,879	44,879	44,879	44,879	44,879
Observations	2,737,619	2,737,619	2,737,619	2,737,619	2,737,619	2,737,619
Family FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	–	Yes	Yes	Yes	Yes
County-Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This Table reports the impact of emigrants to the United States on the innovation activity of their 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 yearly frequency between 1870 and 1930. In columns (1–3), the dependent variable is the log(1+) number of patents issued to family members in the UK. In column (4), the dependent variable is an indicator equal to one if family members in the UK were issued at least one patent, equal to zero otherwise. In column (5), the dependent variable is the average text-based similarity between UK patents and US patents issued in the previous five years. In column (6), the dependent variable is the log(1+) number of patents in the top 20% impact distribution ("Breakthrough" patents). "US Emigrant in Family × Post" 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. "US Emigrant in Family × Post × N. of Emigrants" is an interaction between "US Emigrant in Family × Post" and the number of emigrants from that family. "Non-Return US Emigrant in Family × Post" 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. "Non-Return US Emigrant in Family × Post × N. of Emigrants" is an interaction between "Non-Return US Emigrant in Family × Post" and the number of non-return emigrants from that family. All columns include family (i.e., surname-county) and year fixed effects. Column (3) also includes county-by-year fixed effects. Standard errors, reported in parentheses, are clustered at the county level. The dependent variables in columns (1–4) are multiplied by 100. Referenced on page(s) 28. *: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

TABLE VI. Within-Neighbohood Emigration to the United States and Innovation in the United Kingdom

	Number of Patents					Text-Based Measures	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Number	Number	Number	Number	I(Patents > 0)	US Patents Similarity	High Impact
Panel A. Neighborhood Emigration							
Dependent Variable Mean	*6.721*	*6.721*	*6.741*	*6.775*	*8.483*	*6.792*	*1.550*
Post × Emigrant in Neighborhood	0.440***	0.236**	0.254**	0.205	0.552***	0.477***	0.116**
	(0.130)	(0.110)	(0.116)	(0.154)	(0.160)	(0.132)	(0.052)
Post × N. Emigrants in Neighborhood				0.210			
				(0.209)			
Number of Individuals	136,880	136,872	134,936	117,649	136,880	136,880	136,880
Observations	2,737,600	2,737,440	2,698,720	2,352,980	2,737,600	2,737,600	2,737,600
Panel B. Neighborhood Non-Return Emigration							
Dependent Variable Mean	*6.721*	*6.721*	*6.741*	*6.775*	*8.483*	*6.792*	*1.550*
Post × Non-Return Emigrant in Neighborhood	0.409***	0.252**	0.263**	0.430***	0.517***	0.453***	0.109**
	(0.130)	(0.113)	(0.120)	(0.163)	(0.159)	(0.133)	(0.054)
Post × N. Non-Return Emigrants in Neighborhood				0.064			
				(0.216)			
Number of Individuals	120,575	120,567	118,628	102,714	120,575	120,575	120,575
Observations	2,411,500	2,411,340	2,372,560	2,054,280	2,411,500	2,411,500	2,411,500
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parish-Year FE	No	No	Yes	No	No	No	No
District-Year FE	No	Yes	–	No	No	No	No
Year FE	Yes	–	–	Yes	Yes	Yes	Yes

Notes. This Table reports the impact of emigrants to the United States on the innovation activity of their former neighbors in the United Kingdom. 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. In columns (1–4), the dependent variable is the log(1+) number of patents granted to members of the emigrant's family. In column (5), the dependent variable is an indicator equal to one if members of the emigrant's family are granted at least one patent, equal to zero otherwise. In column (6), the dependent variable is the average text-based similarity between UK patents and US patents issued in the previous five years. In column (7), the dependent variable is the log(1+) number of patents in the top 20% of the impact distribution. "Post × Emigrant in Neighborhood" is an indicator equal to one in the year the first neighbor of the inventor move to the US and in subsequent years, and equal to zero in the previous years. "Post × N. Emigrants in Neighborhood" is an interaction between "Post × Emigrant in Neighborhood" and the number of emigrants to the US from the neighborhood. "Post × Non-Return Emigrant in Neighborhood" is an indicator equal to one in the year the first neighbor of the inventor move to the US (without returning to the UK) and in subsequent years and equal to zero in the previous years. "Post × N. Non-Return Emigrants in Neighborhood" is an interaction between "Post × Non-Return Emigrant in Neighborhood" and the number of emigrants to the US from the neighborhood. Columns (1) and (4–7) include inventor and year fixed effects. Column (2) includes inventor and district-by-year fixed effects. Column (3) includes inventor and parish-by-year fixed effects. Standard errors, reported in parentheses, are clustered at the parish level. The dependent variables in columns (1–4) are multiplied by 100. Referenced on page(s) 29. *: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

TABLE VII. The Transatlantic Telegraph Cable and Innovation in the United Kingdom

	Number of Patents				Text-Based Measures	
	(1) Number	(2) Number	(3) With Station	(4) Without Station	(5) US Patents Similarity	(6) High Impact
Panel A. Patents by District						
Dependent Variable Mean	*0.803*	*0.803*	*0.914*	*0.546*	*21.026*	*0.311*
Post Telegraph × US Emigrants	0.617***	0.116	0.698***	-0.327	148.198***	0.769***
	(0.201)	(0.266)	(0.237)	(0.328)	(10.637)	(0.160)
Post Telegraph × US Emigrants × N. Stations		0.215**				
		(0.086)				
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Number of Districts	621	621	435	186	621	621
Observations	16,146	16,146	11,310	4,836	16,146	16,146
Panel B. Patents by District-Technology						
Dependent Variable Mean	*0.091*	*0.091*	*0.099*	*0.070*	*3.090*	*0.022*
Post Telegraph × Knowledge Exposure	0.375***	0.195*	0.454***	0.037	59.277***	0.157***
	(0.080)	(0.104)	(0.093)	(0.166)	(4.594)	(0.041)
Post Telegraph × Knowledge Exposure × N. Stations		0.077***				
		(0.023)				
District-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
District-Technology FE	Yes	Yes	Yes	Yes	Yes	Yes
Technology-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Number of Districts	621	621	435	186	621	621
Observations	306,774	306,774	214,890	91,884	306,774	306,774

Notes. This Table reports the effect of the transatlantic telegraph cable that connected the UK and the US in 1866 on innovation activity in the UK. In Panel A, the unit of observation unit is a district observed at yearly frequency between 1853 and 1879. In Panel B, the unit of observation is a district-technology class pair observed at a yearly frequency between 1853 and 1879. The analysis period starts in 1853 due to the availability of the patent data. We include years until 1879 to have a symmetric window around the first treatment period (1866). In columns (1–4), the dependent variable is the log(1+) number of patents. In column (5), the dependent variable is the average text-based similarity between UK and US patents issued in the previous five years. In column (6), the dependent variable is the log(1+) number of patents in the top 20% of the impact distribution. "Post Telegraph" is an indicator equal to one for years between 1866 and 1879, equal to zero otherwise. "US Emigrants" is the number of emigrants from the district to the US in the 1870s. "Knowledge Exposure", constructed following equation (1), refers to exposure to US innovation via out-migration in the 1870s. "N. Stations" is the number of telegraph stations located in the district. In column (3), we include districts with at least one telegraph station in 1862. In column (4), we exclude districts with one or more telegraph stations in 1862. In Panel A, all columns include district and time fixed effects. In Panel B, all columns include district-by-time, district-by-technology, and technology-by-time fixed effects. Standard errors, reported in parentheses, are clustered by district (in Panel A) and two-way by district and technology (in Panel B). Referenced on page(s) 31. *: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

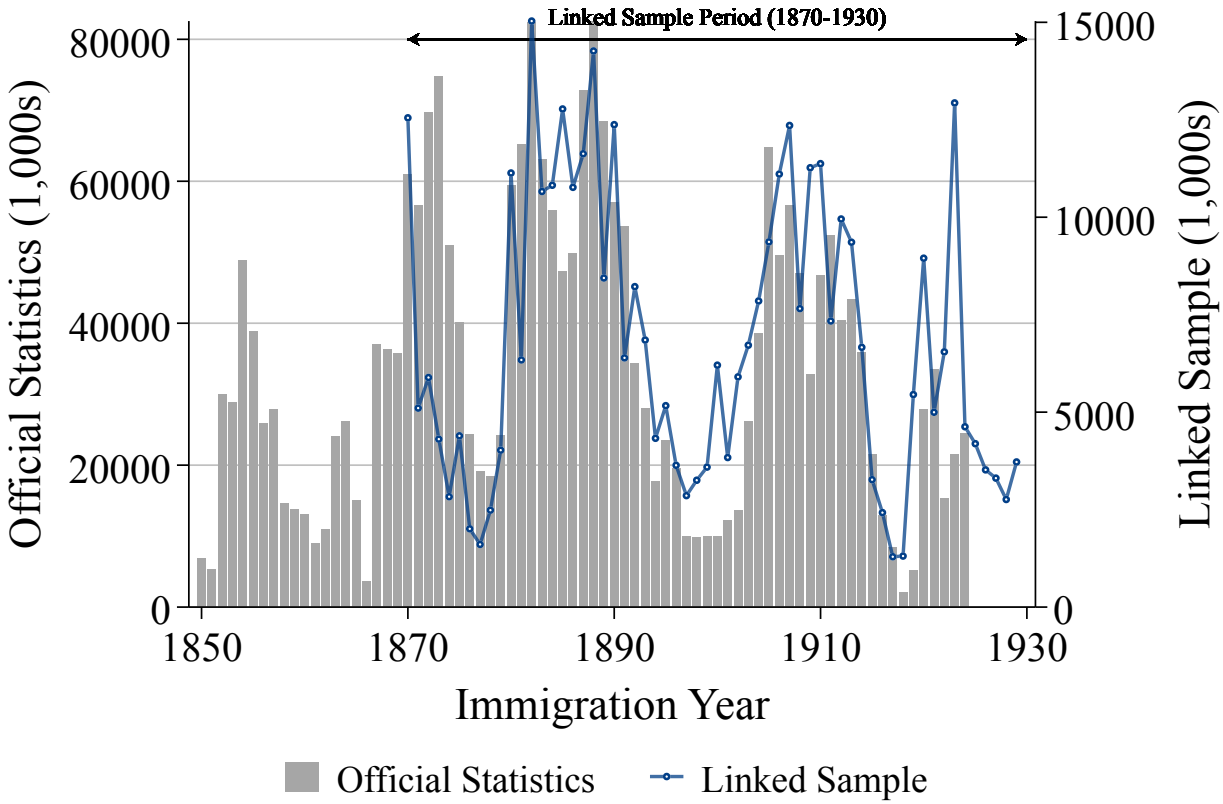
Table VIII. Emigration and Newspaper Coverage of US News

	Mentions of U.S.			Mentions of U.S. States			Mentions of U.S. Counties		
	(1) Total Mentions	(2) With Newspaper(s)	(3) Average Mentions	(4) Total Mentions	(5) With Newspaper(s)	(6) Average Mentions	(7) Total Mentions	(8) With Newspaper(s)	(9) Average Mentions
Dependent Variable Mean	*1.491*	*3.685*	*1.328*	*1.019*	*2.398*	*0.350*	*0.001*	*0.003*	*0.001*
US Emigrants (1,000s)	1.340***	0.654	0.936**	19.539***	10.024***	7.198***	6.710***	9.625***	3.216***
	(0.418)	(0.609)	(0.375)	(3.367)	(3.856)	(1.564)	(1.085)	(1.968)	(0.497)
UK District FE	Yes	Yes	Yes	–	–	–	–	–	–
UK District-US State FE	–	–	–	Yes	Yes	Yes	–	–	–
UK District-US County FE	–	–	–	–	–	–	Yes	Yes	Yes
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Districts	591	307	591	591	316	591	591	302	591
Observations	4,137	1,660	4,137	215,124	91,000	215,124	13,465,344	4,784,640	13,465,344
Std. Beta Coef.	0.071	0.045	0.056	0.042	0.024	0.038	0.014	0.015	0.011

Notes. This Table reports the association between emigration to the United States and newspaper coverage of US-related topics in the United Kingdom. The unit of observation is a district (columns 1–3), district-US state (columns 4–6), and district-US county (columns 7–9). Units are observed at a decade frequency between 1870 and 1930. The dependent variable is the log(1+) number of mentions of "United States" (1), of US states (4), and of US counties (7). In columns (2), (5), and (8), we exclude all districts where no newspaper is based at any point in time. In columns (3), (6), and (9), we use the average number of mentions per newspaper, computed as the ratio between the total number of mentions and the number of newspapers, as the dependent variable. "US Emigrants (1,000s)" is the number of US emigrants in thousands (columns 1–3), the number of emigrants, in thousands, between the district and the state (columns 4–6), and the number of emigrants, in thousands, between the district and the county (columns 7–9). All columns include dyadic fixed effects—i.e., district FEs in columns (1–3), district-by-state FEs in columns (4–6), and district-by-county FEs in columns (7–9)—as well as time fixed effects. Standard errors, reported in parentheses, are clustered at the district level. Referenced on page(s) 33. *: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

Figures

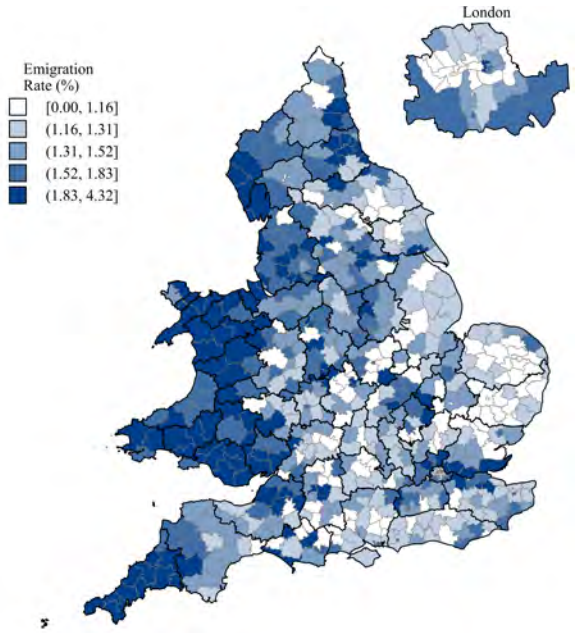
Figure I. British Emigration to the United States: Official Statistics and Linked Sample



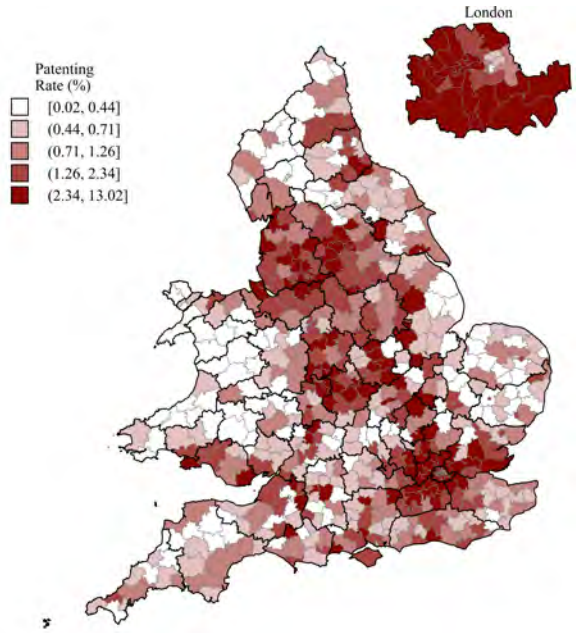
Notes. This Figure compares the total number of English and Welsh immigrants in the United States as recorded in official statistics from Willcox (1928) with the linked emigrants' sample assembled in this paper. The gray bars report, on the left y-axis, the number of English and Welsh immigrants in the official statistics by their recorded immigration year in the United States. The blue line, whose values are reported on the right y-axis, reports 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 one British census. Numbers are in thousand units. The black two-sided arrow marks the period covered by the linked sample, i.e., the years between 1870 and 1930. Referenced on page(s) 11.

FIGURE II. Patenting and Emigration Rates in the United Kingdom

(A) Emigration Rate (%)

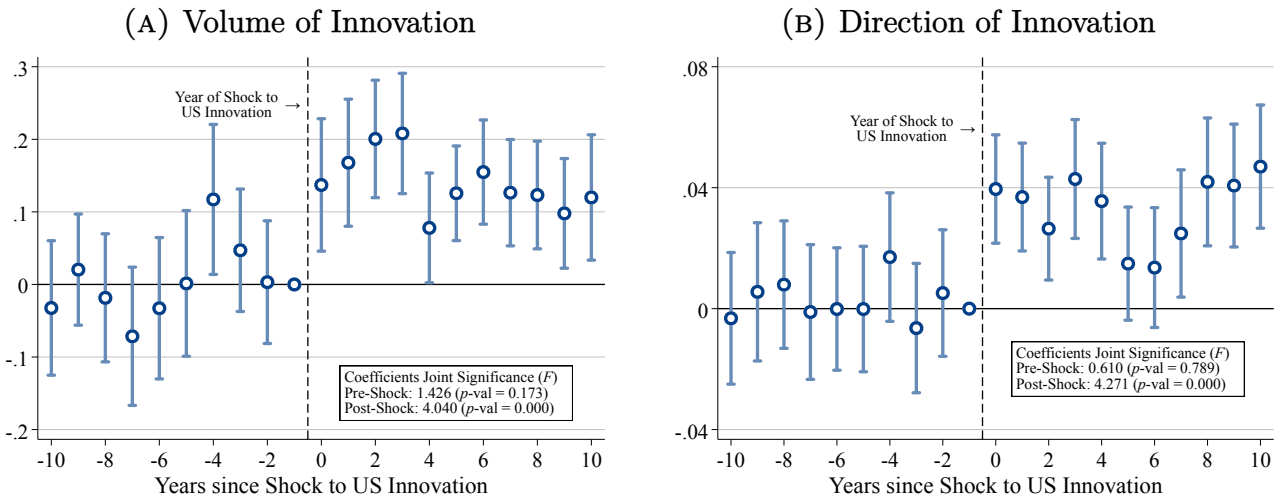


(B) Patenting Rate (%)



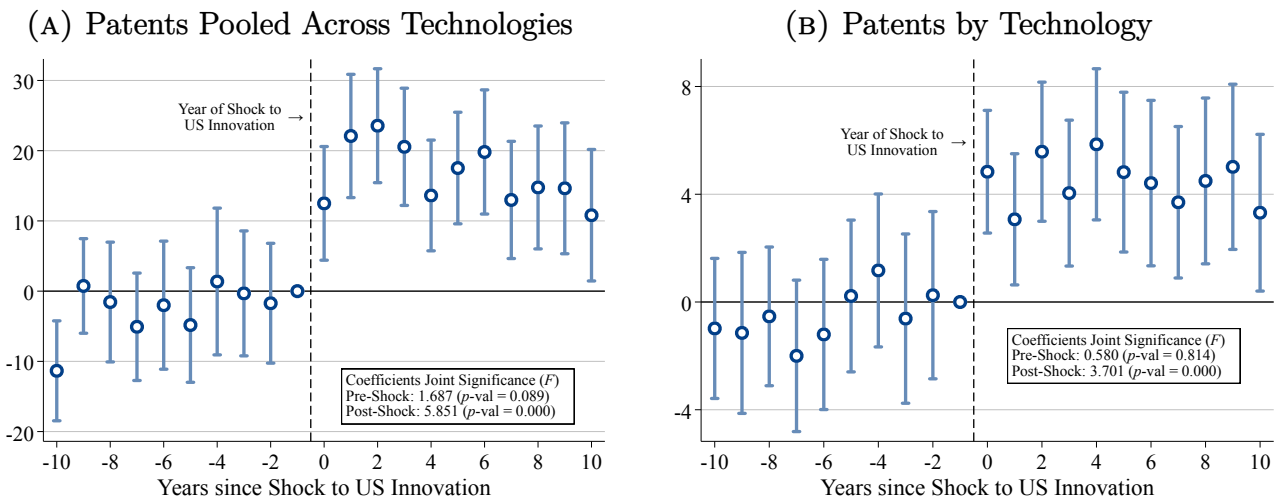
Notes. This Figure reports the spatial distribution of emigrants across English and Welsh districts over the period 1870–1930 (Panel IIa) and the number of patents issued over the same period (Panel IIb). Both series are normalized by district population in 1891 and are reported in percentage points. Districts are displayed at 1891 historical borders, and the emigrant population is cross-walked to consistent borders. Lighter to darker tones of color indicate increasing quantiles of the variable. The London area is displayed separately. Referenced on page(s) 13.

FIGURE III. Dynamic Effect of Exposure to Innovation Shocks in the United States



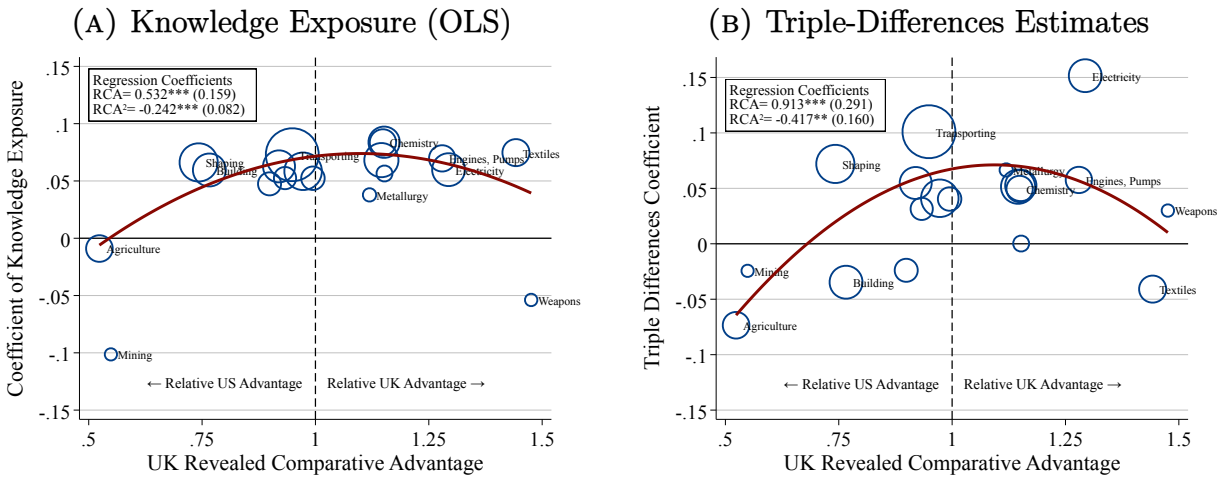
Notes. This Figure reports the impact of shocks to US innovation activity on innovation in the UK. In Panel IIIa, the unit of observation is a district observed at yearly frequency between 1870 and 1930; in Panel IIIb, the unit of observation is a district-technology pair observed over the same period. The dependent variable is the log(1+) number of patents. Each dot reports the coefficient of an indicator variable, which codes the time since the observation unit is exposed to a shock in US innovation activity through emigration ties. The black dashed line indicates the treatment period. The last period before the shock serves as the baseline category. In Panel IIIa, the regression includes district and year fixed effects; in Panel IIIb, the regression includes district-by-year, technology-by-year, and district-by-technology fixed effects. Standard errors are clustered at the district level. The bands report 95% confidence intervals. Each Figure reports separate F-statistics for the joint significance of the pre-and post-treatment coefficients and associated p-values. Referenced on page(s) 22.

FIGURE IV. Dynamic Effect of Exposure to Innovation Shocks in the United States on the Similarity Between British and American Patents



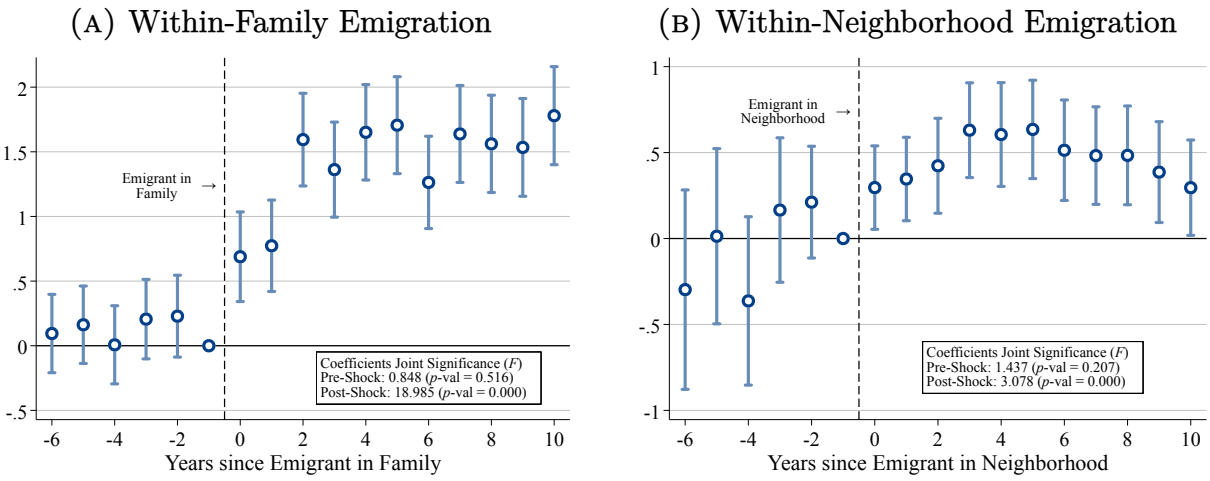
Notes. This Figure reports the impact of shocks to US innovation activity on innovation in the UK. In Panel IVa, the unit of observation is a district observed at yearly frequency between 1870 and 1930; in Panel IVb, the unit of observation is a district-technology pair observed over the same period. The dependent variable is the average text-based similarity of UK patents issued in a given period to US patents issued in the previous five years. Each dot reports the coefficient of an indicator variable, which codes the time since the observation unit is exposed to a shock in US innovation activity through emigration ties. The black dashed line indicates the treatment period. The last period before the shock serves as the baseline category. In Panel IVa, the regression includes district and year fixed effects; in Panel IVb, the regression includes district-by-year, technology-by-year, and district-by-technology fixed effects. Standard errors are clustered at the district level. The bands report 95% confidence intervals. Each Figure reports separate F-statistics for the joint significance of the pre-and post-treatment coefficients and associated p-values. Referenced on page(s) 23.

FIGURE V. Heterogeneous Effects of Exposure to Innovation in the United States Across Technologies



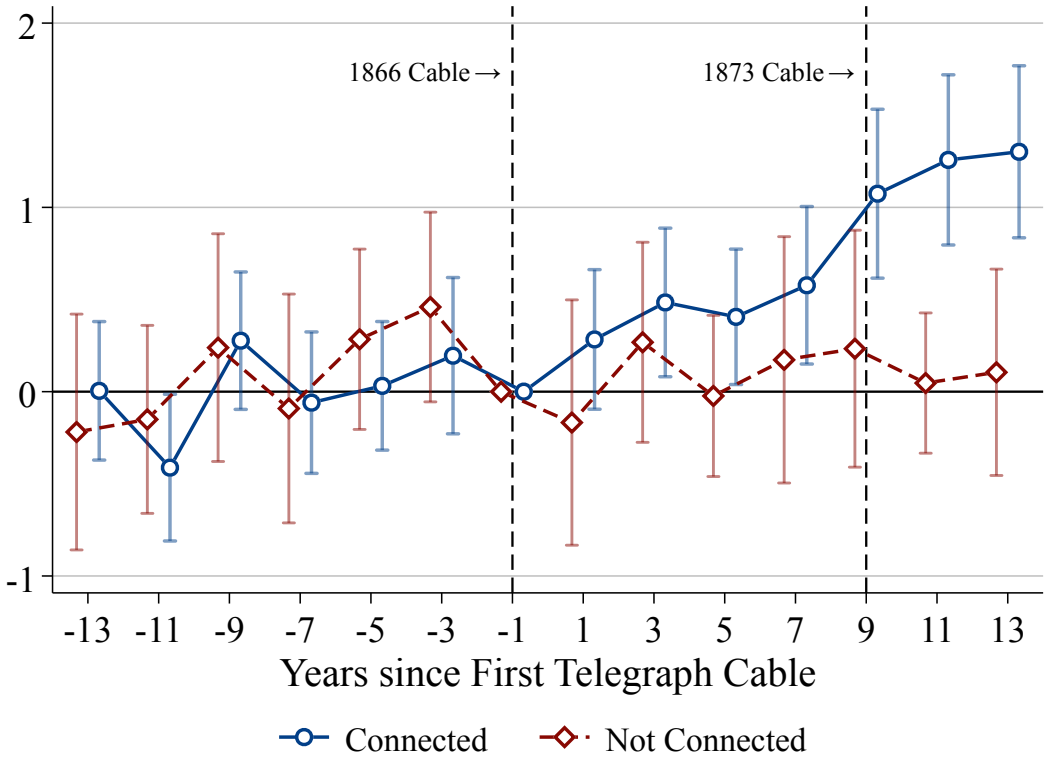
Notes. This Figure reports the heterogeneous effects of exposure to US technology on UK innovation across technology classes. In Panel Va, the unit of observation is a district-technology couple observed at decade frequency between 1870 and 1930; in Panel Vb, the same units are observed at yearly frequency over the same period. In Panel Va, each dot reports the coefficient of an interaction term between exposure to US technology and technology indicator variables; in Panel Vb, each dot reports the coefficient of an interaction term between technology dummies and an indicator variable that is equal to one for periods after the unit is exposed to a US innovation shock for the first time and zero otherwise. In both cases, the dependent variable is the log(1+) number of patents. In Panel Va, the underlying regression includes district-by-time and technology-fixed effects; in Panel Vb, the triple-differences model includes district-by-year, technology-by-year, and district-by-technology fixed effects. The estimated coefficients are plotted against the measure of revealed comparative advantage discussed in the main text. Loosely speaking, the UK is relatively more active than the US in technologies to the right of one, and it is less active than the US otherwise. Both plots superimpose a polynomial of degree two and report the estimated coefficients of the measure of revealed comparative advantage of orders one and two, along with their robust standard errors. The size of each dot reflects the number of patents issued in the UK over the sample period in each technology class. Referenced on page(s) 23.

FIGURE VI. Dynamic Effects of Family and Neighborhood Emigration



Notes. This Figure reports the impact of emigration to the United States on the innovation activity of their relatives in the United Kingdom (Panel VIa) and of their neighbors (Panel VIb). In Panel VIa, the unit of observation is an extended family defined as comprising all those who share the same surname in a given county. Families are observed at yearly frequency between 1870 and 1930. In Panel VIb, the unit of observation is an individual observed at yearly frequency between 1880 and 1900. The dots report the coefficients of an indicator variable that codes the period since the first emigrant in the family (Panel VIa) or in the neighborhood (Panel VIb). The black dashed line indicates the period when the treatment is first active; the last year before the treatment is active serves as the baseline category. The dependent variable is the log(1+) number of patents. All regression specifications control for unit (family or individual) and year fixed effects. Standard errors are clustered at the family (Panel VIa) and district (Panel VIb) level. The bands report 95% confidence intervals. Each Figure reports separate F-statistics for the joint significance of the pre- and post-treatment coefficients and associated p-values. Referenced on page(s) 28.

FIGURE VII. Dynamic Effect of the Transatlantic Telegraph Cable



Notes. This Figure reports the dynamic effects of the transatlantic telegraph cable(s) that connected the UK and the US in 1866 on innovation activity in the UK. The unit of observation is a district-technology class pair observed at a yearly frequency between 1853 and 1879. The analysis period is constrained by the patent data starting in 1853. The year 1879 is chosen to have a symmetric window around the first treatment period (1866). The dependent variable is the log(1+) number of patents. The dots report the coefficient of an interaction term between knowledge exposure in the 1870s and year dummies that code the period since the introduction of the first transatlantic telegraph cable in 1866. The dashed black lines mark the first cable and subsequent ones in 1873 and 1874. Years are grouped in two-year windows; the 1865- 1866 bin is the baseline category. The blue dots refer to the sub-sample of districts part of the British telegraph network. The red dots refer to the unconnected sub-sample. The regressions on both sub-samples include district-by-technology, technology-by-year, and district-by-year fixed effects; standard errors are clustered at the district level. Bands report 95% confidence intervals. Referenced on page(s) 32.

Online Appendix

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

Davide M. Coluccia & Gaia Dossi

January, 2025

Table of Contents

A Data Appendix **A2**

A.I Summary of Data Sources A2

A.II Patent Data A6

A.III Linked Migrants Sample A12

Tables A18

Figures A22

B Additional Results **B34**

B.I Selection of British Emigrants B34

B.II Long-Run Association between Emigration and Innovation B35

B.III Assortative Matching of Emigrants in the United States B36

B.IV Anecdotal Evidence of Return Innovation B37

Tables B39

Figures B42

C Robustness Analyses **C44**

C.I OLS and 2SLS Analysis C44

C.II Double and Triple Differences Analysis C49

C.III Neighborhood Analysis C51

Tables C52

Figures C66

A Data Appendix

This section describes the data sources and the construction of the final sample (section A.I), the construction of the newly digitized patent data and the approach to link patents to the census (section A.II), and the methodology we developed to link the records of British immigrants in the United States to the UK census (section 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 numerous inventors, we weigh each by the inverse of the number of inventors to avoid multiple counting. English and Welsh patents after 1900 are available at 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 section A.II, 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 transporting categories 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 "quality" or "innovativeness" following Kelly *et al.* (2021). This metric flags as influential those patents that introduce terms not used before they were granted and become common after that. We evaluate this metric on 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. Neither US authorities nor the sending ones in the UK collected them. We thus 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 A.III. Ideally, we observe the universe of British emigrants to the United States between 1870 and 1930. For those individuals, we know all the information contained in the US Census and those detailed in the UK one. Most notably, we know where they came from. As we discuss more in detail later, we also link return migrants. Since the last publicly available UK census dates to 1911, we can only construct return migration flows over the period 1870–1910.

A.I.3 Data Constructed from the Population Censuses

The main data sources we leverage are the 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 thus avoid discussing it any further (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 who have resided in the UK since 1850. The first census was run in 1841, but only 1851, 1861, 1881, 1891, 1901, and 1911 are entirely digitized.[1] Data in the census include the name and surname, birth year, division, county, district, parish, precise address of residence, the specific occupation detailed through HISCO codes, and other variables that we do not use in the paper, such as the type of dwelling and fertility information. We augment these variables by geo-coding the universe of addresses that appear in the census to precise geographical coordinates.

A.I.4 Historical Newspapers and Coverage of US-related News

e collect data on newspaper coverage of US-related news from the British Newspaper Archive.[2] Beach and Hanlon (2023) describe this dataset in detail. In this paper, we run a set of three queries. First, we search for the words "United States". Second, we perform 49 searches, one for each state, excluding New York, because we could not distinguish mentions of the state from mentions of the city. Finally, we perform approximately three thousand searches, one for each county. Each search spans the period 1850–1939. We collect the information at the article level. For each entry in the database, we know the

[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] A limited free-tier access to newspaper data is available at `www.britishnewspaperarchive.co.uk`.

journal, day, month, and year of publication, whether it is an article or some other type of content—e.g., an obituary–, the page, and the word count. Importantly, we collect information on the universe of newspapers in the archive. Journal-level data contain the publishing address at the city level, the first and last day, month, and year of activity, and the publication frequency—e.g., quarterly, daily. We then geocoded each newspaper to the coordinates of the city where it was published and mapped those to 1891 registration districts. We can thus construct a measure of newspaper coverage at the district-year level.[3]

In Table A.4, we provide a set of summary statistics on the resulting dataset. We collect information for 2022 newspapers: of these, 1459 are based in England, and 93 are published in Wales. We exclude Scottish and Irish newspapers from the analysis. The average life of a newspaper in this period is 40 years. In Panel B, we report district-level statistics by decade. The number of newspapers decreases over time, as noted by Beach and Hanlon (2023), from an average of 2.3 newspapers per district in the 1870s to 0.7 in the 1930s. It is unclear whether this is due to incomplete coverage in the later period. In Panel C, we report the district-level statistics by division and find that newspapers appear to be quite sparse across the country except for the London division. Figure A.2 displays the spatial distribution of the number of newspapers across districts over the period and confirms the impression that newspapers tend to evenly cover a substantial share of districts. London is a major outlier: we thus perform all exercises dropping London districts and find consistent results.

A.I.5 Miscellaneous Data

To construct the domestic UK telegraph network before the first transatlantic UK–US cable (1866), we digitize the list of telegraph stations reported in the *Zeitschrift des Deutsch-Österreichischen Telegraphen-Vereins, Jahrgang*, volume IX, 1862. This directory lists the universe of telegraph stations outside of London in 1862. To the best of our knowledge, it is the most complete directory before the introduction of the transatlantic cable. We geo-code each station to precise coordinates. The red dots in Figure A.1 report each station. We then label each district with at least one telegraph station as "connected" to the domestic network and as "not connected" otherwise.

We construct US county-level exposure to the Great Influenza pandemic using mortality statistics collected by the US Bureau of Census. These data are available for a subset of counties representing approximately 60% of the US population in 1900.

To compute the railway-based instrument, we construct US-county level immigration shocks following

[3]Unfortunately, for newspapers based in London, we only know their city, i.e., London. In the newspaper analysis, we are thus forced to conflate all urban London districts into a single "London" geographical unit.

the methodology described in Sequeira *et al.* (2020). We use the same data sources. Hence, we defer the interested reader to their paper for a more detailed discussion.

A.I.6 GIS Shapefiles & Boundary Harmonization

Patents and telegraph stations are mapped to 1900 registration district borders using historical GIS files and their coordinates.[4] However, all data from the population censuses appear at historical borders. Registration districts have not undergone major boundary changes over the period that we studied. However, we adapt the method presented by Eckert, Gvirtz, Liang and Peters (2020) to UK districts to ensure that we work with consistent geographical units.

To construct geographical crosswalks using their method, one needs to assume that variables are evenly distributed over the area of geographical units. The crosswalk is then obtained by overlapping geographical units over time. Suppose unit x in decade d is split, and 80% of its territory is assigned to itself, while 20% is assigned to another district y. To construct a cross-walk relative 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 re-assigned in this way. We adopt the same methodology to map counties to their 1900 borders.

A.I.7 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 adopt a different strategy, only analyze the 1901 census, and geo-reference the 1851-1911 censuses. The geo-coded census sample is used in the neighborhood individual-level analysis; all other exercises do not rely on these data.

There are two ways to geo-reference historical addresses. One approach is manually digitizing 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. A second automated approach is to run text-based address matching between historical data sources and address databases that have already

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

been geo-referenced. We follow this latter method since we need to geo-reference 5,464,578 unique addresses.

To perform the actual matching, we first operate a preliminary manual trimming 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. Crucially, 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 ensures that the algorithm matches are not spurious.

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

A.II Patent Data

A.II.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 patenting activity since 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, Tartari and Tranchero, 2021). We fill this gap by constructing the first dataset of English and Welsh patents that spans the period 1853- 1900 and contains detailed information on the text, geographical location, inventors' personal information, and date for the universe of patents.

A.II.2 Data Sources

The UK Intellectual Property Office allowed us access to restricted full-page scans of original patent documents. These are the universe of patents granted in England and Wales between 1617 and 1899. This paper focuses on the period 1853-1899 for two main 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.3: in panel A.3a, we show the patent granted to Henry Bessemer for the eponymous process to produce steel, and in panel A.3b we display 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 comprise (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 immaterial, for only granted patents are included in the sample.

A.II.3 Digitization

We individually 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 state-of-the-art large language models—specifically, `GPT 3.5`—to parse all variables (i–ix). The text of patent grants 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. This exercise resulted in a database of approximately 800,000 patents granted between 1853 and 1899.

A.II.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. In Figure IIb, we report the resulting distribution of patents per capita, whereas Figure A.5 reports the spatial distribution of patent grants across technologies. Reassuringly, these are consistent with underlying population and economic development indicators as well as with historical evidence (e.g., note the substantial clustering of textile patents in the Lancashire districts).

A.II.5 External Validation

To validate our data, we consider the only two series covering a portion of the 1853-1899 years. Hanlon (2016) digitized an index of patents issued between 1855 and 1883. His data list, for each patent, the inventor(s) and their profession(s), a technology class, and the issue year. On top of the longer time coverage, our data thus contain several additional information, including the geographical coordinates. The second dataset we use to compare is the "A Cradle of Invention" (COI) series, published by Finishing Publications (2018). These data, too, were digitized from indices and thus 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 figures after 1883 for the COI series.

In Table A.1, we report the aggregate number of patents issued according to our series (columns 2 and 6), COI (columns 3 and 7), and Hanlon (2016) (columns 4 and 8). Reassuringly, the three series are highly consistent. Our series is closest to Hanlon (2016), but the COI figures are not too far off either. Overall, 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.II.6 Measuring the Similarity Between US and UK Patents and their Quality

This section describes how we construct the patent similarity metric to measure "copying" and "originality" of UK innovation activity. The approach borrows heavily on Kelly *et al.* (2021). We adapt their methodology to our context by leveraging text information in titles only. Even though we do not have access to full US patent texts, the title of a patent is usually very informative about its 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 the later years; titles for US patents are collected from PATSTAT throughout the sample period.

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

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

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 shall have

$$\text{TF}_{wp} \equiv \frac{\sum_{c \in p} 1(c = w)}{\sum_{c \in p} 1(c)} \tag{A.2}$$

where the numerator returns how many times word w appears in patent p, and the denumerator is simply 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:

$$\text{TF-BIDF}_{wp,t} \equiv \text{TF}_{wp} \times \text{BIDF}_{w,t} \tag{A.3}$$

and, thus, 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 the $\text{TF-BIDF}_{p,t}$ vectors for US and UK patents, but the $\text{BIDF}_{w,t}$ 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 sake of the paper's narrative, we define as "copying". Formally we define

$$\text{Backward Similarity}_i^{\tau} \equiv \sum_{j \in \mathcal{F}_i^{-\tau}} \rho_{i,j} \tag{A.4}$$

where the set $\mathcal{F}_i^{-\tau}$ denotes the set of US patents issued within τ 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), who suggest 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

$$\text{Excess Forward Similarity} \equiv \frac{\sum_{j \in \mathcal{F}_j^{+\tau}} \rho_{i,j}}{\sum_{j \in \mathcal{F}_i^{-\tau}} \rho_{i,j}} \tag{A.5}$$

where $\mathcal{F}_i^{+\tau}$ denotes the set of US patents issued within τ 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. In Tables C.1, C.2, and C.10, we report the results using an alternative threshold of ten years. Moreover, in the same Table, we report 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 we do not conflate shifting terminology fashions in the similarity measures.

A.II.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 the census data determines the sample restriction. First, we seek to follow the inventing activity of individuals over time. Hence, we want to link patent records to one single census. When linking inventors to the census, we exploit the location listed on the document to restrict the pool of potential matches in the census. This practice, however, assumes that the inventor did not move between the time when the patent was granted and the closest census year. This assumption is sensible for patents issued not too far from the closest census year. Since the records of the 1871 census are not available, we restrict the attention to patents issued no earlier than ten years from the closest census, i.e., 1891.

Matching Algorithm Given a patent p, define the set of 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 ℓ is a couple 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. Notice that these are progressively coarser units: parishes are contained in districts, which form counties. Unfortunately, we do not know the inventor-location pair. To match the generic A_p, we thus perform the following operations:

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 year_i and t_p respectively denote the birth year of i and the issue date:

$$\mathcal{M}_{A_p}^{\text{parish}} = \left\{ i \in \mathcal{L}_p^{\text{parish}} \middle| \, t_p - \text{year}_i \geq 18 \right\} \tag{A.6}$$

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:

$$\text{Similarity}_i^{A_p} = \alpha \times \text{Name Similarity}_i^{A_p} + (1-\alpha) \times \text{Surname Similarity}_i^{A_p} \tag{A.7}$$

for some $\alpha \in [0,1]$. In our baseline setting, we pick $\alpha = .3$ to assign a larger weight to the surname.

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

$$\overline{\mathcal{M}}_{A_p}^{\text{parish}} = \left\{ i \in \mathcal{M}_{A_p}^{\text{parish}} \middle| \text{Similarity}_i^{A_p} = \max_{i' \in \mathcal{M}_{A_p}^{\text{parish}}} \text{Similarity}_{i'}^{A_p} \right\} \tag{A.8}$$

and define Similarity^{A_p} as 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 τ such that if $\text{Similarity}^{A_p} < \tau$, $\overline{\mathcal{M}}_{A_p}^{\text{parish}} = \emptyset$, otherwise pass.

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 that we have geographical information on the location of inventors. Inventors are mobile, however, and there may be a considerable time between the moment the patent is granted 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.6, we report the matching rate of this exercise. We match approximately 75% of the inventors in the sample. The share remains constant throughout the period. Almost 50% of the matches are attained at the parish level—the smallest geographical layer we consider. About half of the remaining matches are obtained within the 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 we consider.

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, hence, 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 we may need to go national to obtain a match, which would most likely be dropped because of the multiple-match issue. These hypotheses are difficult to test. In Figure A.7, however, we report the overall distribution of the number of matches in the sample. Approximately 60% of the inventors listed in the patent grants are linked to one single census record, and the share of inventors linked to more than 10 records is negligible. In fact, we can restrict the

analysis to single matches and obtain very similar results. In Table A.2, we compute the correlation between the number of matches in our sample and a set of individual observed characteristics. In Panel A, we have the age and the residence divisions while in Panel B we list 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, exclude them from the sample, and confirm that the results hold.

A.III Linked Migrants Sample

A.III.1 Data Sources

We rely on two sources of externally compiled data.[5] For the US, we have access to the IPUMS full-count non-anonymized census (Ruggles *et al.*, 2021). A census was taken in the US every ten years starting in 1790, except for 1890. Until 1840, the census was run at the household level. From 1850 on, instead, we have detailed *individual* information on the universe of the US population.[6] For confidentiality, these data are available up until 1940. Our dataset, therefore, contains snapshots of the entire US population at any given decade between 1850 and 1940, although for the sake of this paper, we restrict to the years 1870-1930. Crucially, we have access to the non-anonymized version of the IPUMS data. Hence, we also know each individual's recorded name and surname besides publicly available information.

In the UK, the I-CeM data mirrors the IPUMS (Schurer and Higgs, 2020) content. More precisely, it contains information on the universe of people living in England, Scotland, and Wales. Similarly to the US and virtually every other census, it was run at a decade frequency from 1851 until 1911. No census was taken in 1871. As with the IPUMS data, we can access the full-count, non-anonymized version of the dataset. Besides publicly available information, this contains full names and addresses of the universe of individuals living in the UK at any given decade.

A.III.2 Linking Algorithm

Our methodology relies on Abramitzky *et al.* (2021). This dataset tackles the problem that neither the US nor the UK—nor any other European countries—recorded where British immigrants came from *within* the UK. Thus, we try to match British immigrants residing in the US with their entry in the

[5] 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.

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

UK census, which records where they come from at a granular geographical level.[7] More precisely, we take the stock of British 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.[8] 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 sample we link consists of individuals who report, in the US census, either England or Wales—or analogous denominations, such as Great Britain—as their country of origin. In the 1900 census, we take all those who immigrated between 1870 and 1899. In the subsequent censuses, until 1930, we retrieve 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 this sequence of operations:

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 as A—call the resulting sample $\mathcal{M}^A = \{m_1^A, \ldots, m_N^A\}$.

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[9]

$$\text{Similarity}_i^A = \alpha \times \text{Name Similarity}_i^A + (1-\alpha) \times \text{Surname Similarity}_i^A \tag{A.9}$$

   for some $\alpha \in [0,1]$. In our baseline setting, we set $\alpha = 0.3$ to give higher weight to the surname.

4. The set of matches is defined as

$$\overline{\mathcal{M}}^A = \left\{ m_i^A \in \mathcal{M}^A \,\middle|\, \text{Similarity}_i^A = \max_{m_{i'}^A \in \mathcal{M}^A} \text{Similarity}_{i'}^A \right\} \tag{A.10}$$

   which means that we restrict the set of possible matches to include only those whose similarity score with the entry in the US census A is the largest.

5. Finally, for a given threshold $\tau > 0$, we select only the possible matches whose similarity score

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

[8] Since no census was taken 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.

[9] We cannot simply match on exact same name and surname because coding errors are commonplace in historical census data (Abramitzky *et al.*, 2021).

is above τ. The set of effective matches thus boils down to:

$$\widetilde{\mathcal{M}}_{\tau}^{A} = \left\{ m_i^A \in \overline{\mathcal{M}}^A \,\middle|\, \text{Similarity}_i^A \geq \tau \right\} \tag{A.11}$$

Clearly, $\widetilde{\mathcal{M}}^A$ can ideally be empty, meaning that A has no effective matches. It can have one element, in which case we refer to it as a "perfect match," or it can have multiple matches. In our baseline exercise, we set $\tau = 0.7$ as we see a clear elbow in the distribution of similarities there.

We evaluate the distance between two strings i and j in terms of their Jaro-Winkler similarity d_{ij}:

$$d_{ij} \equiv \widehat{d_{ij}} + \ell p(1 - \widehat{d_{ij}}) \tag{A.12}$$

where

$$\widehat{d_{ij}} \equiv \begin{cases} 0 & \text{if } m = 0 \\ \frac{1}{3}\left(\frac{m}{|i|} + \frac{m}{|j|} + \frac{m-t}{m}\right) & \text{else} \end{cases} \tag{A.13}$$

where m is the number of matching characters, $|i|$ is the length of string i, t is half the number of transpositions, ℓ is the length of common an eventual common prefix no longer than four characters between i and j, and $p = 0.1$ is a constant scaling factor. Two characters are matching only if they are the same and are not farther than $\left\lfloor \frac{\max(|i|,|j|)}{2} \right\rfloor - 1$. Half the number of matching characters in different sequence order is the number of transpositions.[10]

The Jaro-Winker distance has been shown to perform well in linking routines (Abramitzky *et al.*, 2021). In our particular case, however, this metric outperforms more standard string dissimilarity metrics, such as the cosine or the Levenshtein distances, because the Jaro-Winkler assigns a "bonus" score to strings starting with closer initial substrings. In addition, coding errors are far more frequent at the end of names and surnames than at the beginning. A manual assessment confirmed that the Jaro-Winkler metric outperforms other measures in our setting.

A.III.3 Internal and External Validation

Matching Rate In Figure A.9, we report the key matching rate statistics of the linked sample. In particular, Figure A.9a 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

[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)

link British immigrants born after the last census before they migrated (i.e., those who migrate when younger than ten), we distinguish between the full sample in blue and the "matchable" sample in red. The matching rate in the matchable sample is approximately 65%. It declines from 75% in the 1870s to 60-65% 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 asked whether matched and un-matched immigrants are balanced on observables. Panel A.9b replicates the previous plot but reports the absolute number of emigrants in the sample and those linked instead of the percentage shares.

In Panel A.9c, we display the distribution of the number of matches for linked immigrants. A large fraction ($\approx 45\%$) of immigrants 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 9.4. To ensure that multiple matches do not confound our analysis, we always weigh emigrants by the inverse of the number of matches. In unreported analyses, however, we confirm that all our results remain qualitatively unchanged even if we exclude instances with more than two matches from the analysis sample.

In Figure A.10, 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 (.9). 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 90% of the sample lies right to the .9 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 0.9.

Balance of Linked Emigrants In Table A.3, we report 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 variables 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. As one may expect, the linking probability does not correlate with either literacy—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, and as sales workers. These differences, however, are small in magnitude and affect a small portion of the population. Manufacturing workers (skilled "craftsmen" and unskilled "operatives"), instead, account for almost 60% of the population and do not appear to be selected in

terms of their linking probability. In terms of their region of residence, emigrants living in the North-east are more likely to be linked, although the difference is minimal in magnitude (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 are primarily not selected on observable characteristics and, importantly, the statistically significant differences are minor in magnitude.

Validation of the Data It is challenging to validate our sample with external data because, as we note in the main text, we do not have primary geographically disaggregated data on the origin of British emigrants. 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 issues. First, they are based on interpolations from population data tabulated from the population census. First, the author tries to account for internal migration, which remains an important confounding factor. Second, the estimates refer to aggregate emigration outflows, whereas our data precisely captures emigration to the United States. Third, the estimates are computed at the county level. Counties are coarse geographical units. Hence, this dataset does not warrant any modern econometric exercise.

Keeping these shortcomings in mind, in Figure A.11, we aggregate our data at the county-decade level and plot it against the estimates produced by Baines (2002). Both series are taken in log terms to reduce the influence of extreme values. Moreover, we weight the data by county population. We do so to ensure that the resulting correlation reflects the actual sizable differences in population across counties. We estimate a positive and statistically significant correlation between our data and the Baines series. Our figures are generally lower than those provided by Baines, which plausibly reflects that our data do not cover emigration towards countries other than the United States. Overall, we view this graph as providing supporting evidence of the plausibility of our dataset.

Finally, in unreported results, we construct an intergenerational linked sample from the English and Welsh population census. This sample allows us to follow individuals over two consecutive censuses between 1851 and 1911. Using this sample, we find that individuals recorded in census t that also appear as emigrants between census t and $t+10$ are 60% less likely to be linked to the census in $t+10$ than those that do not appear in the linked migrant sample. Moreover, eliminating from the emigration data those linked to the census in $t+10$ yields qualitatively and quantitatively similar results to those shown in the paper. The intergenerational linking algorithm presents important issues, as described in Abramitzky *et al.* (2021), but provides additional evidence supporting the informativeness of our intergenerational linking exercise.

A.III.4 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 a minor variation on the algorithm described previously. Since the last UK census is the 1911 one, 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.

Previous research suggests that return migration rates during the Age of Mass Migration were substantial (Bandiera, Rasul and Viarengo, 2013), although probably less so in the UK than in second-wave countries such as Italy. Using our linked sample methodology, we find an approximately 30% return migration rate, broadly consistent with previous estimates.

TABLE A.1. External Validation of Newly Digitized Patent Data

Years 1853–1876				Years 1877–1899			
Year	New Data	Hanlon	COI	Year	New Data	Hanlon	COI
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(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(s) A8.

TABLE A.2. Balance of Linked Inventor Sample

	Unconditional		Year FE	
	Mean	Std. Err.	Mean	Std. Err.
	(1)	(2)	(3)	(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 page(s) A11, C51.

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

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	0.885	0.965	0.079	(0.069)
Income	3.147	3.141	-0.006	(0.011)
Panel B. Occupation				
Professional	0.053	0.041	-0.012***	(0.002)
Farmer	0.085	0.091	0.006	(0.010)
Manager	0.066	0.059	-0.006	(0.004)
Clerical	0.058	0.045	-0.013***	(0.004)
Sales	0.062	0.047	-0.014***	(0.001)
Craftsman	0.260	0.281	0.020*	(0.010)
Operative	0.310	0.330	0.020	(0.014)
Service	0.050	0.049	-0.001	(0.003)
Laborer	0.057	0.056	-0.000	(0.004)
Panel C. Region of Residence				
North East	0.519	0.539	0.019***	(0.005)
Midwest	0.277	0.271	-0.006	(0.005)
South	0.057	0.046	-0.011***	(0.001)
West	0.146	0.145	-0.002	(0.008)

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 page(s) 11, A15, A15.

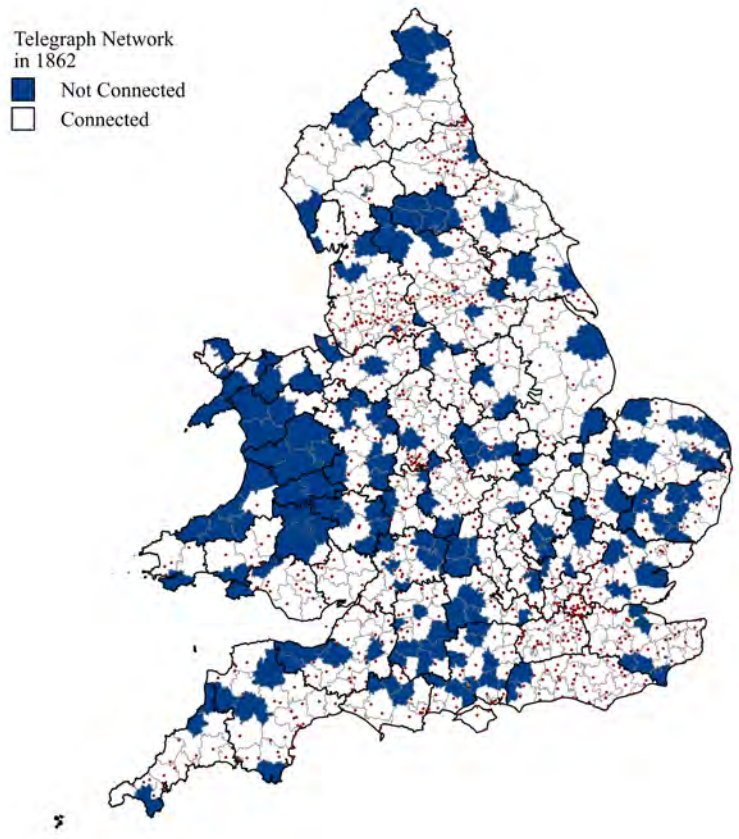
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

TABLE A.4. Newspaper Summary Statistics

	(1) Mean	(2) Std. Dev.	(3) Min.	(4) Max.	(5) Observations
Panel A. Journal-Level Statistics					
Number of Issues	2795.843	4959.740	1	46163	2022
First Publication Year	1869.746	44.171	1699	1996	2094
Last Publication Year	1910.692	49.470	1699	2009	2094
Publication Lifespan	40.946	40.490	0	273	2094
Publication Lifespan if English	40.993	41.921	0	273	1459
Publication Lifespan if Welsh	38.161	36.920	0	178	93
Publication Lifespan if Scottish	45.144	41.107	0	251	229
Publication Lifespan if Irish	41.336	34.809	0	170	241
Panel B. District-Level Statistics, by Decade					
1870s	2.309	14.860	0	285	637
1880s	1.885	11.610	0	233	636
1890s	1.494	8.587	0	160	634
1900s	1.166	5.893	0	114	634
1910s	0.942	3.845	0	83	633
1920s	0.809	2.381	0	50	633
1930s	0.714	1.274	0	24	633
Panel C. District-Level Statistics, by Division					
East	1.631	1.272	1	8	111
East Midlands	2.349	2.409	1	14	43
London	18.767	97.312	1	534	30
North East	2.079	1.761	1	8	38
North West	3.600	3.477	1	17	40
South East	1.800	1.271	1	6	100
South West	1.747	1.382	1	8	79
Wales	2.327	2.391	1	10	52
West Midlands	2.342	2.722	1	18	79
Yorkshire	2.186	2.201	1	10	59

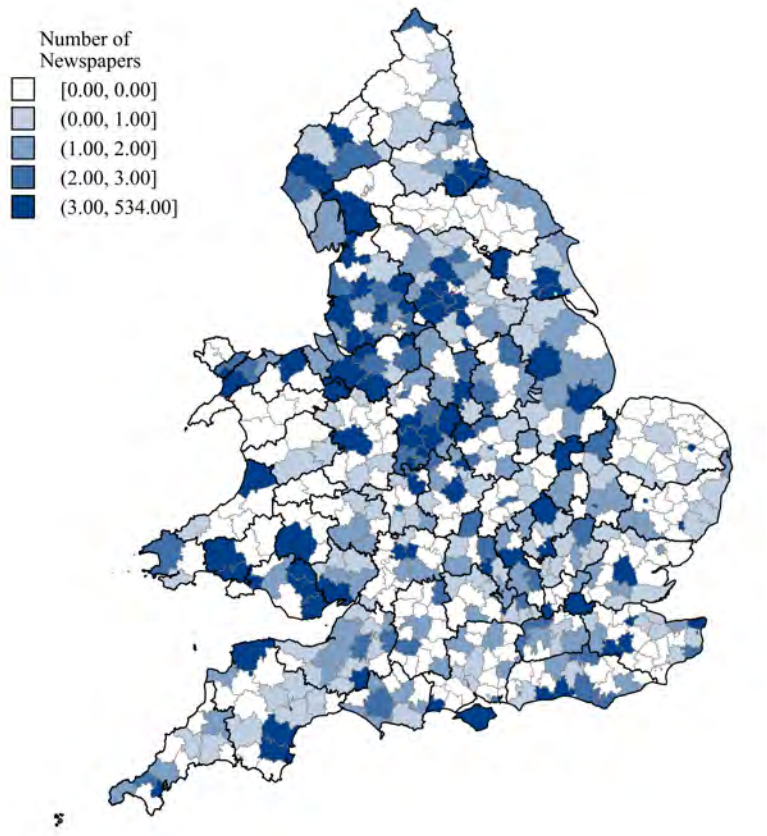
Notes. This Table reports descriptive statistics on newspapers active in the UK between 1880 and 1940. In Panel A, figures are computed at the newspaper level; Panel B computes district-level statistics on the number of newspapers by decade; Panel C computes district-level statistics on the number of newspapers by division. Panels B and C only restrict the observation sample to English and Welsh districts. Newspapers were geo-coded to their publishing address and assigned to districts based on their borders in 1891. Referenced on page(s) 14, A4.

FIGURE A.1. Geographic Distribution of Telegraph Stations (1862)



Notes. This Figure reports the spatial distribution of telegraph stations across districts in 1862. Red markers display the location of telegraph stations. Districts without any telegraph station are displayed in dark blue. To retrieve the coordinates of each telegraph station, we geo-reference the city where it is located. The list of telegraph stations is taken from the *Zeitschrift des Deutsch-Österreichischen Telegraphen-Vereins, Jahrgang*, volume IX, 1862. This source does not list telegraph stations in London. We thus dissolve urban districts in the London area into a single "London" unit and assume that this unit is connected to the domestic telegraph network. Referenced on page(s) 14, 31, A4.

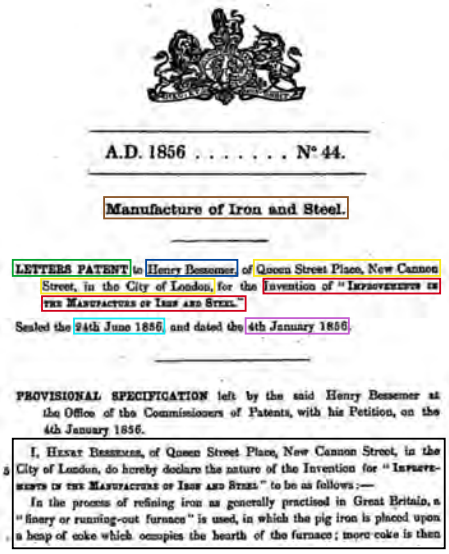
FIGURE A.2. Geographic Distribution of Active Newspapers



Notes. This Figure reports the spatial distribution of active newspapers across districts from 1880 to 1940. A publication must be active for at least one year between 1880 and 1940 to be included in the data. To retrieve the location of each journal, we geo-reference its publishing address and overlay historical district boundaries to assign it to consistent 1891 districts. The publishing address only lists the city. Hence, we cannot distinguish between the eleven London urban districts. We consequently dissolve these districts into a single "London" unit. Referenced on page(s) 14, A4.

FIGURE A.3. Sample Annotated Patent Documents: the Bessemer Process and the First Modern Safety Bicycle

(A) Henry Bessemer's 1856 Patent



A.D. 1856 . . . . . N° 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."

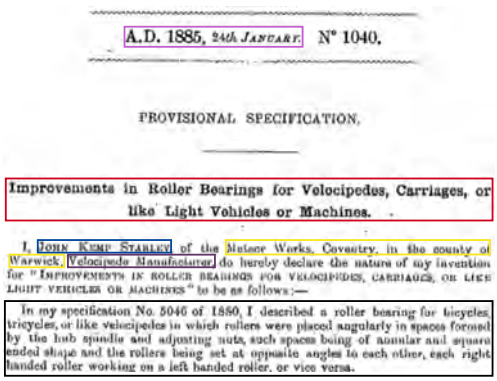
Sealed the 24th June 1856, and dated the 4th January 1856.

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. N° 1040.

PROVISIONAL SPECIFICATION.

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

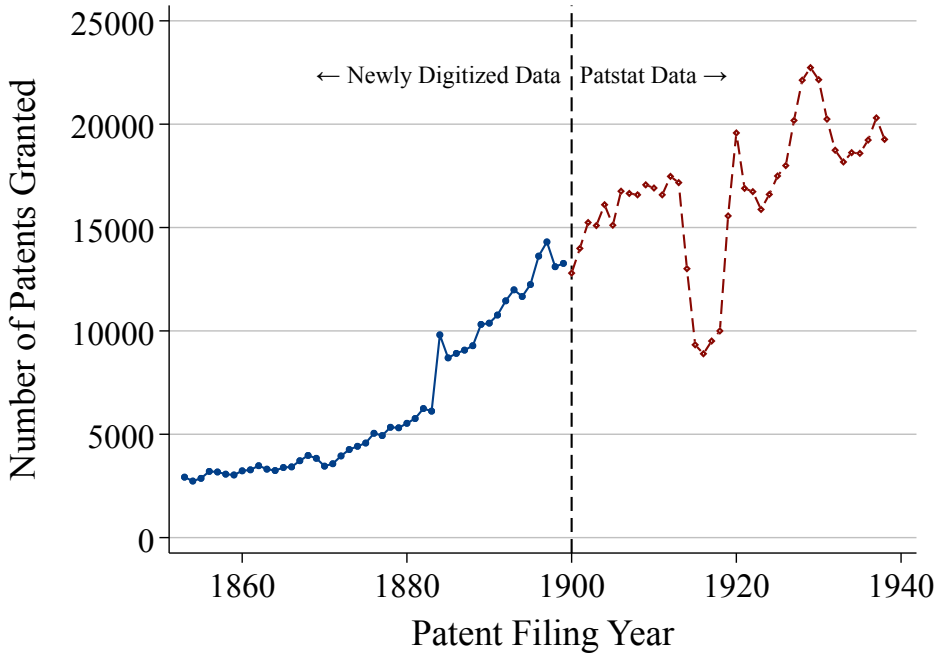
I, JOHN KEMP STARLEY of the Meteor Works, Coventry, in the county of Warwick, Velocipede Manufacturer, do hereby declare the nature of my invention for "IMPROVEMENTS IN ROLLER BEARINGS FOR VELOCIPEDES, CARRIAGES, OR LIKE LIGHT VEHICLES OR MACHINES" to be as follows:—

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.

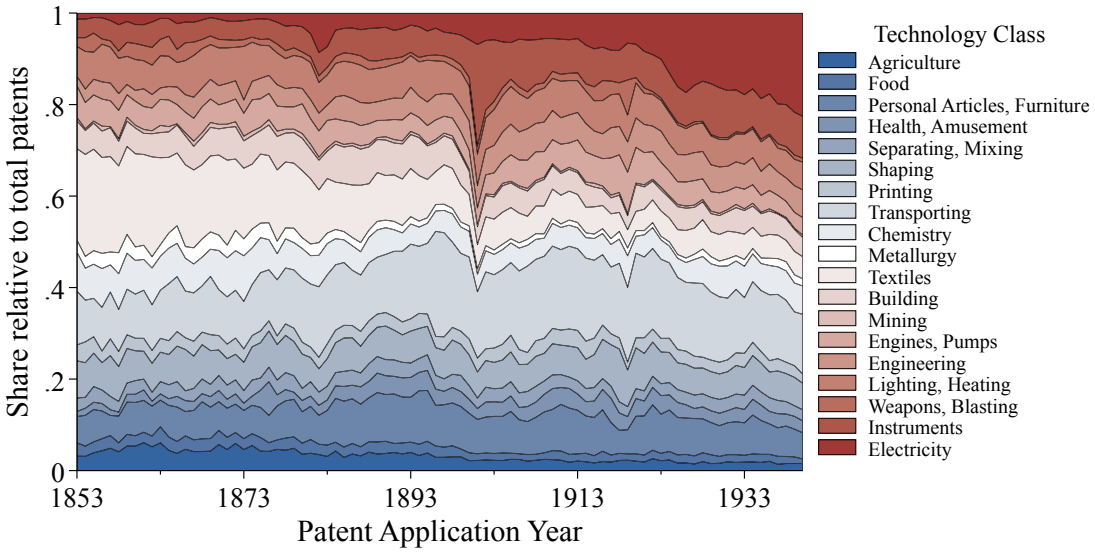
Notes. This figure displays two sample patent documents in our dataset. Panel A.3a was granted to Henry Bessemer in 1856 to invent the famous eponymous process for the mass production of steel from molten pig iron. Panel A.3b was granted to John Starley in 1885 to invent the first modern bicycle, which 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(s) A6.

FIGURE A.4. Number and Composition of Patents Over Time

(A) Time Series of Patents

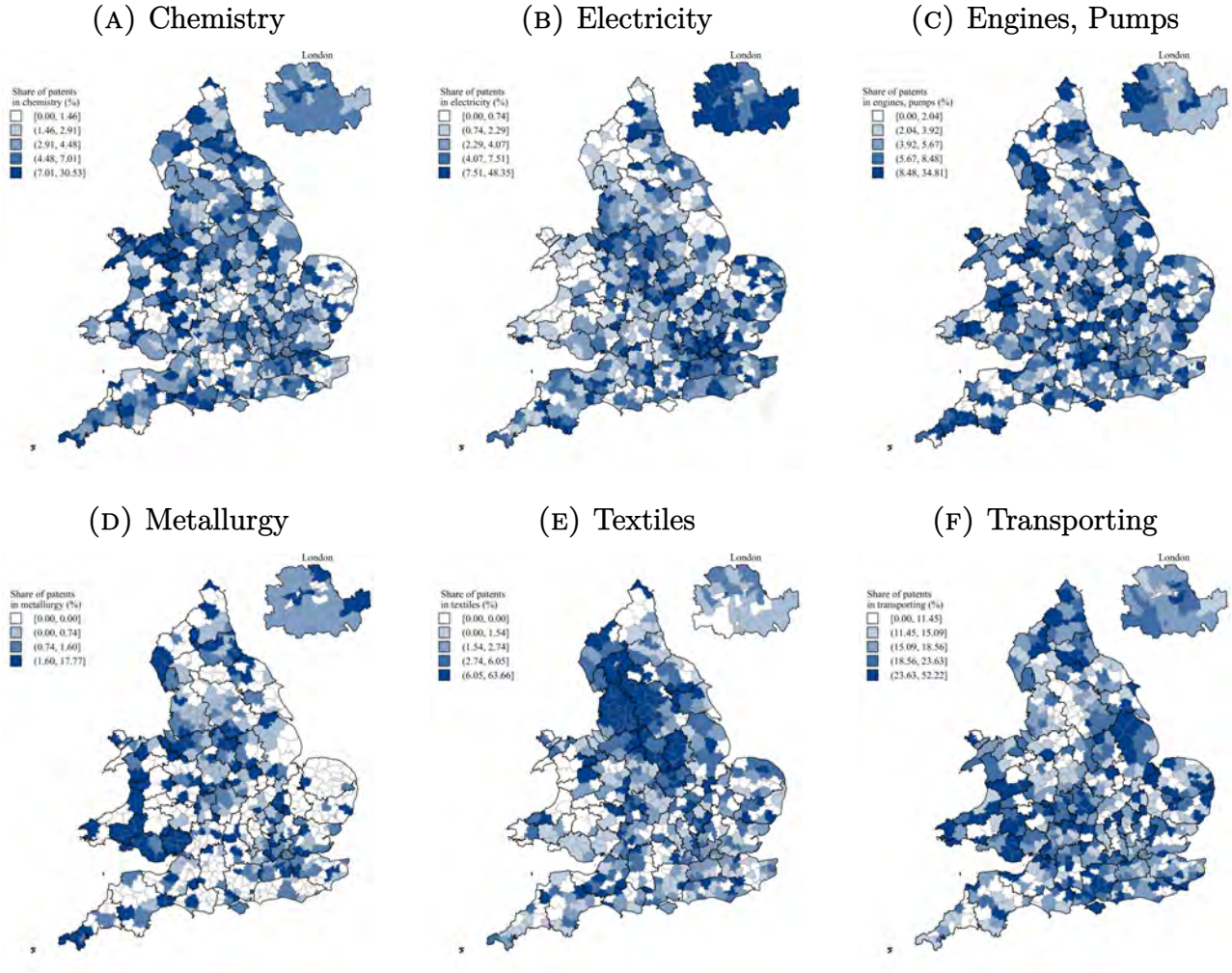


(B) Composition of Patents Across Technologies



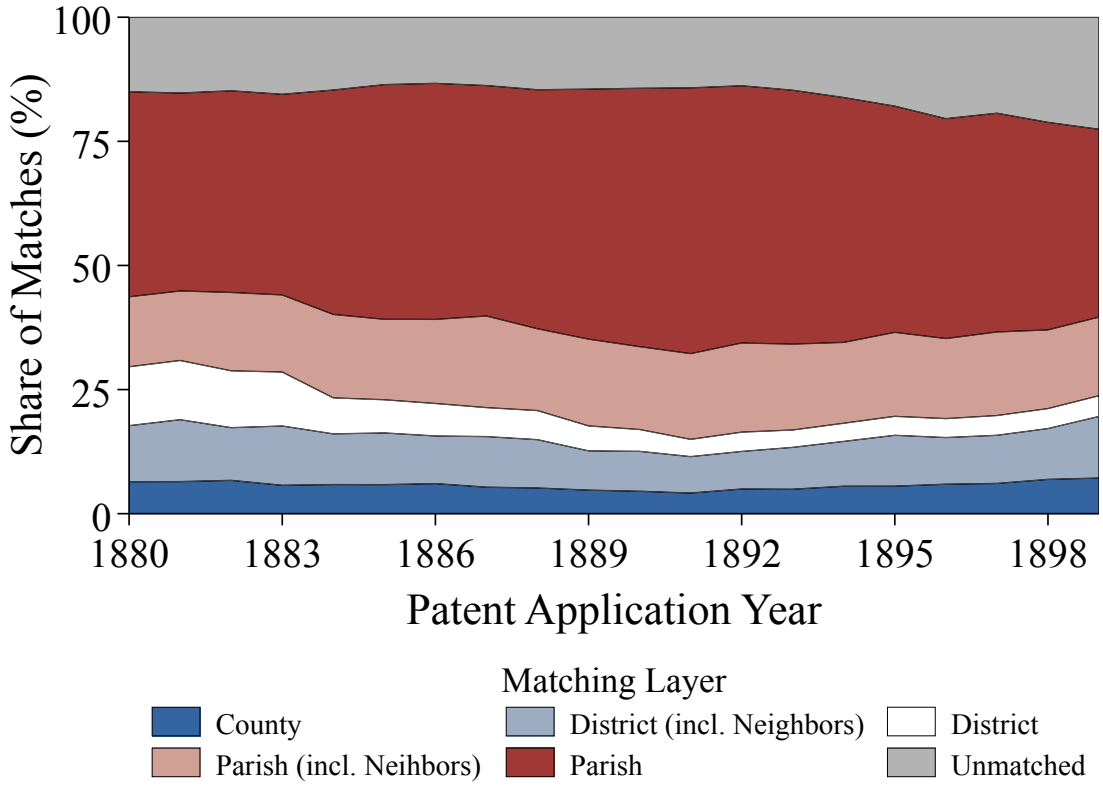
Notes. This Figure reports time-series information on the innovation activity in Britain between 1853 and 1939. In Panel A.4a, we report the total number of patents granted in the UK over the period. The blue dots report the newly digitized data that we assembled for this paper; the red dots report tabulations from the Patstat repository. In Panel A.4b, we plot the share of patents granted over time across technology classes. Referenced on page(s) 13.

FIGURE A.5. 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(s) A7.

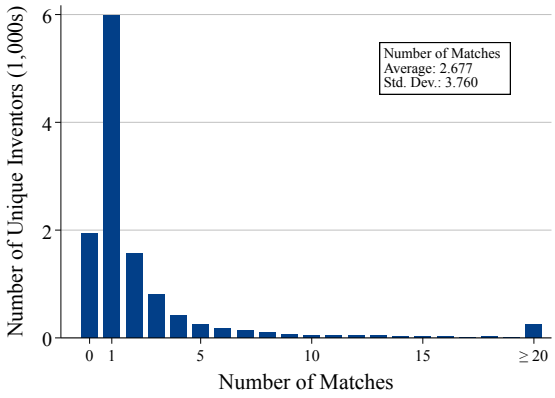
FIGURE A.6. Matching Rate of Linked Inventors Sample



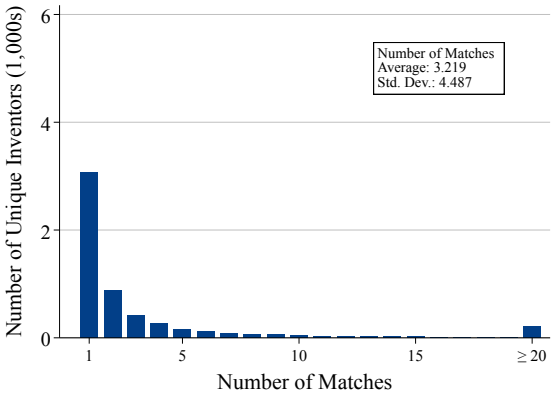
Notes. This Figure reports the matching rate of the linked sample of inventors. The records of the inventors who obtained a patent between 1880 and 1899 are linked to the 1891 population census, as detailed in the main text. The matching rate, i.e., the share of inventors successfully matched to the 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(s) A11.

FIGURE A.7. Number of Matches in Linked Inventor Sample

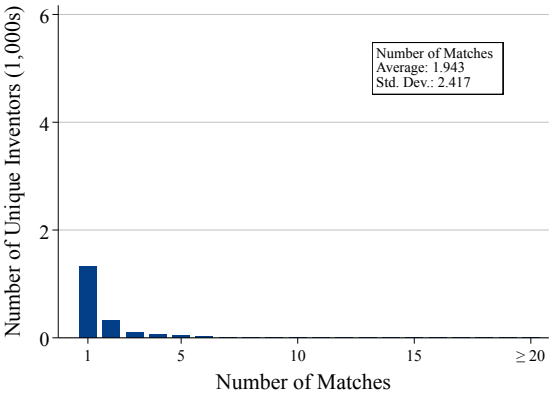
(A) Pooled Data



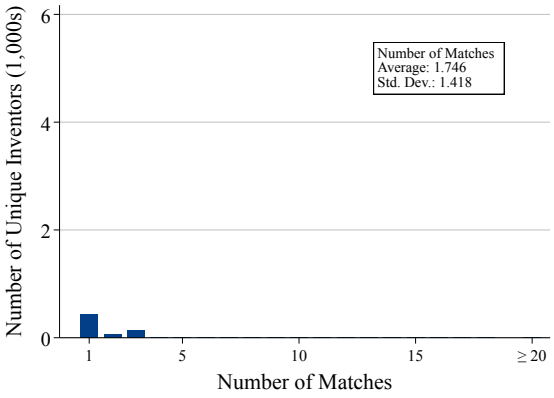
(B) Parish



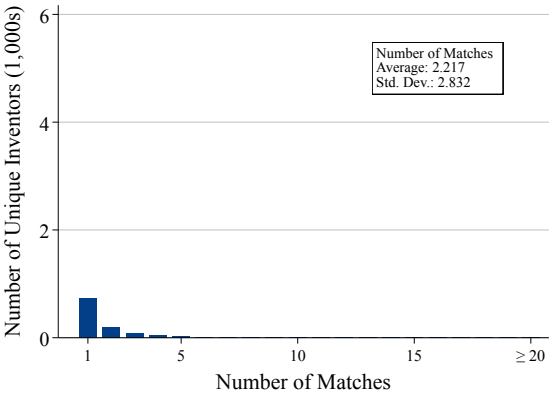
(C) Parish (incl. Neighbors)



(D) District

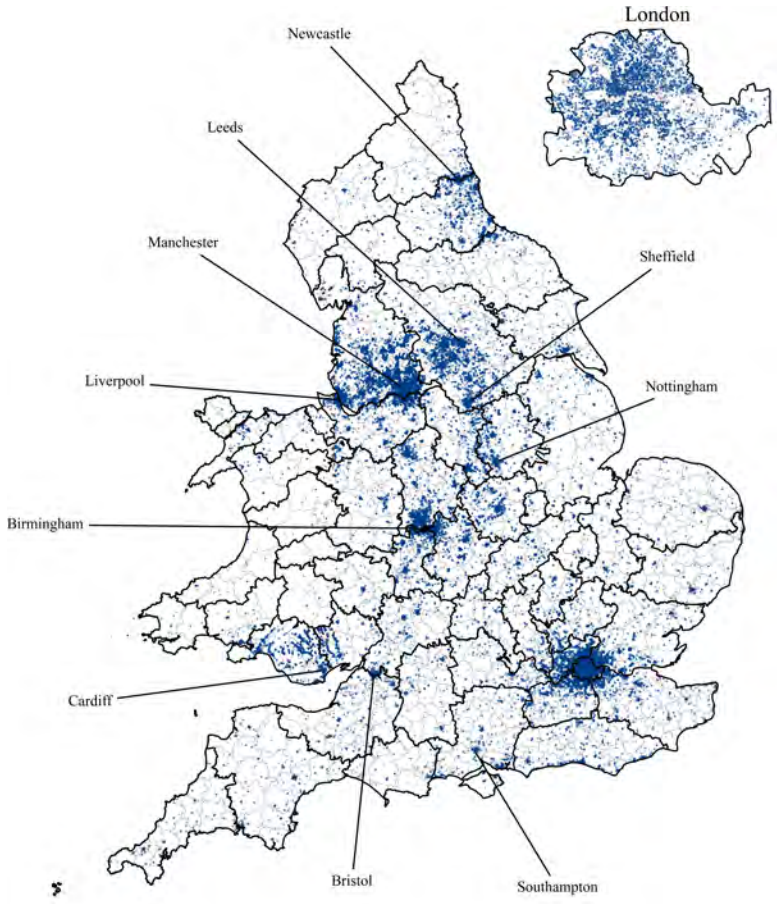


(E) District (incl. Neighbors)



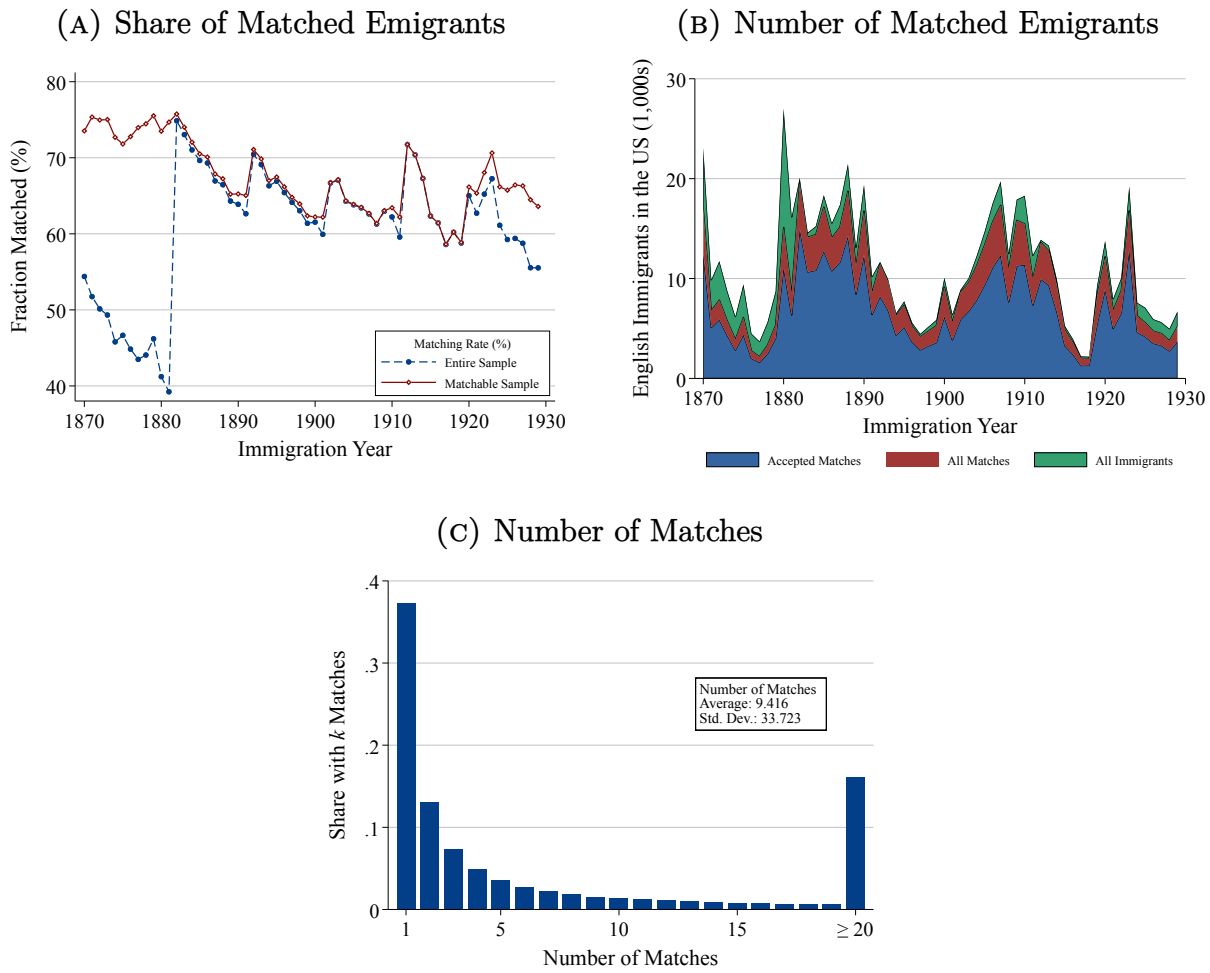
Notes. This Figure reports the number of census entries each inventor in the linked sample is matched to. Panel A.7a reports the overall distribution, while Panels A.7b–A.7e report the distributions broken down by geographic layers. In each graph, we separately report the average number of matches and its standard deviation. Referenced on page(s) A11.

FIGURE A.8. 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(s) 13.

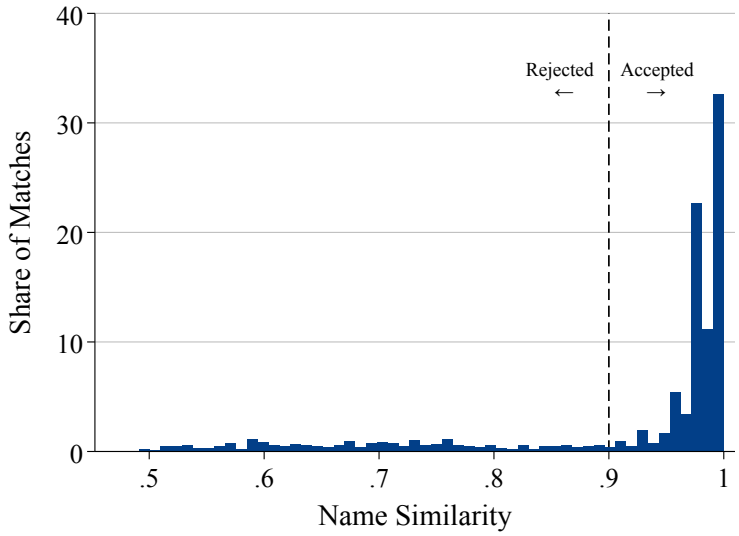
FIGURE A.9. Matching Rate and Number of Matches



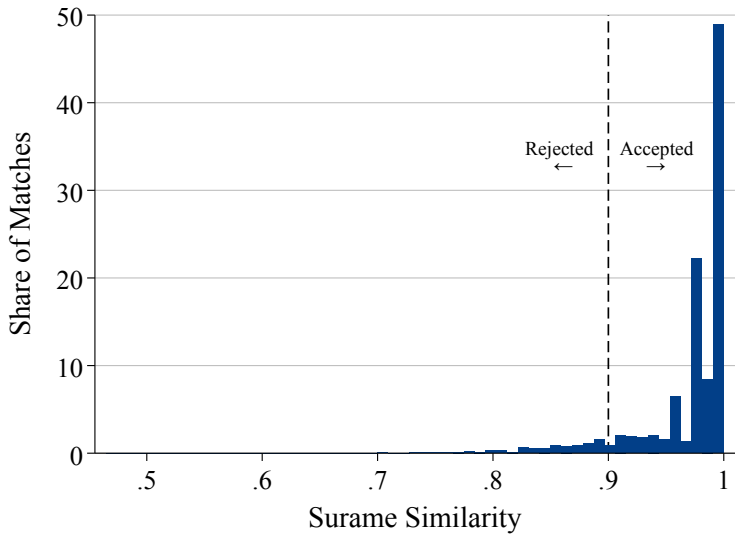
Notes. This Figure reports information on the matching performance of the emigrants' linked sample. In Panel A.9a, we display the matching rate, i.e., the share of emigrants recorded in the US census that are successfully linked to at least one entry in the UK census. The blue dots refer to the overall 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.9b, 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.9c reports the distribution of the number of matches, where the last bin collects all those with more than 20 matches. Referenced on page(s) 10, A14.

FIGURE A.10. Quality of Matches

(A) Name Similarity

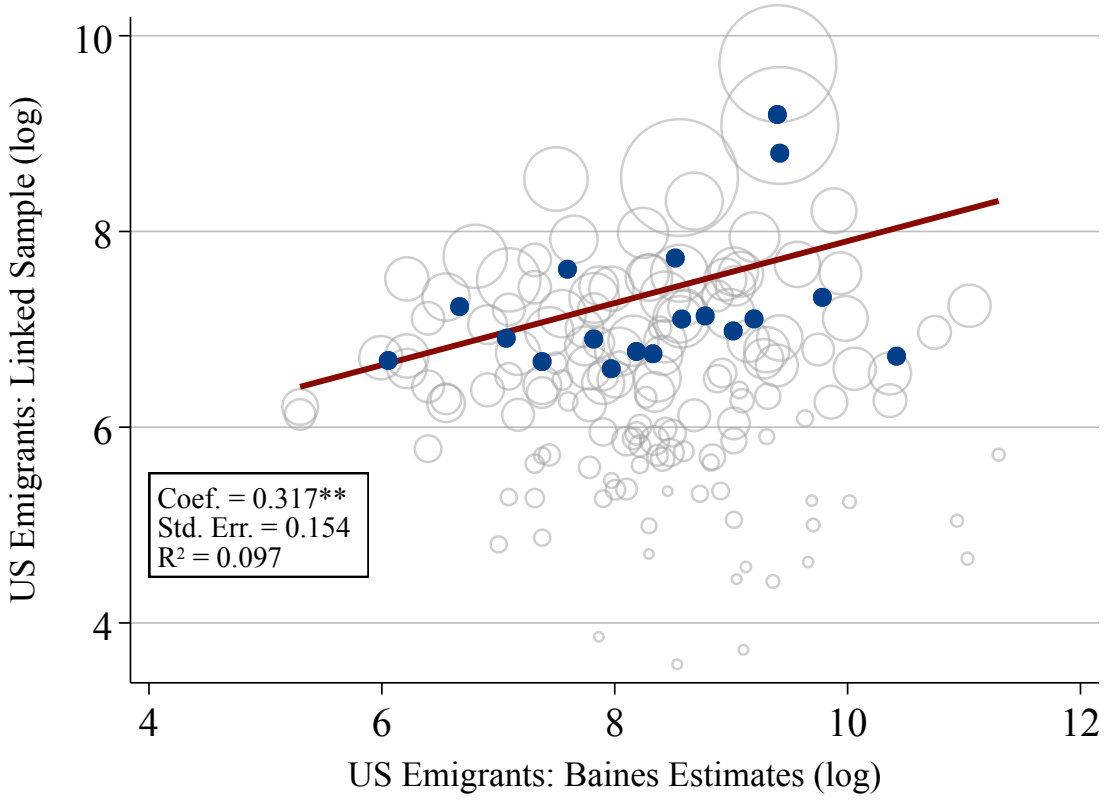


(B) Surname Similarity



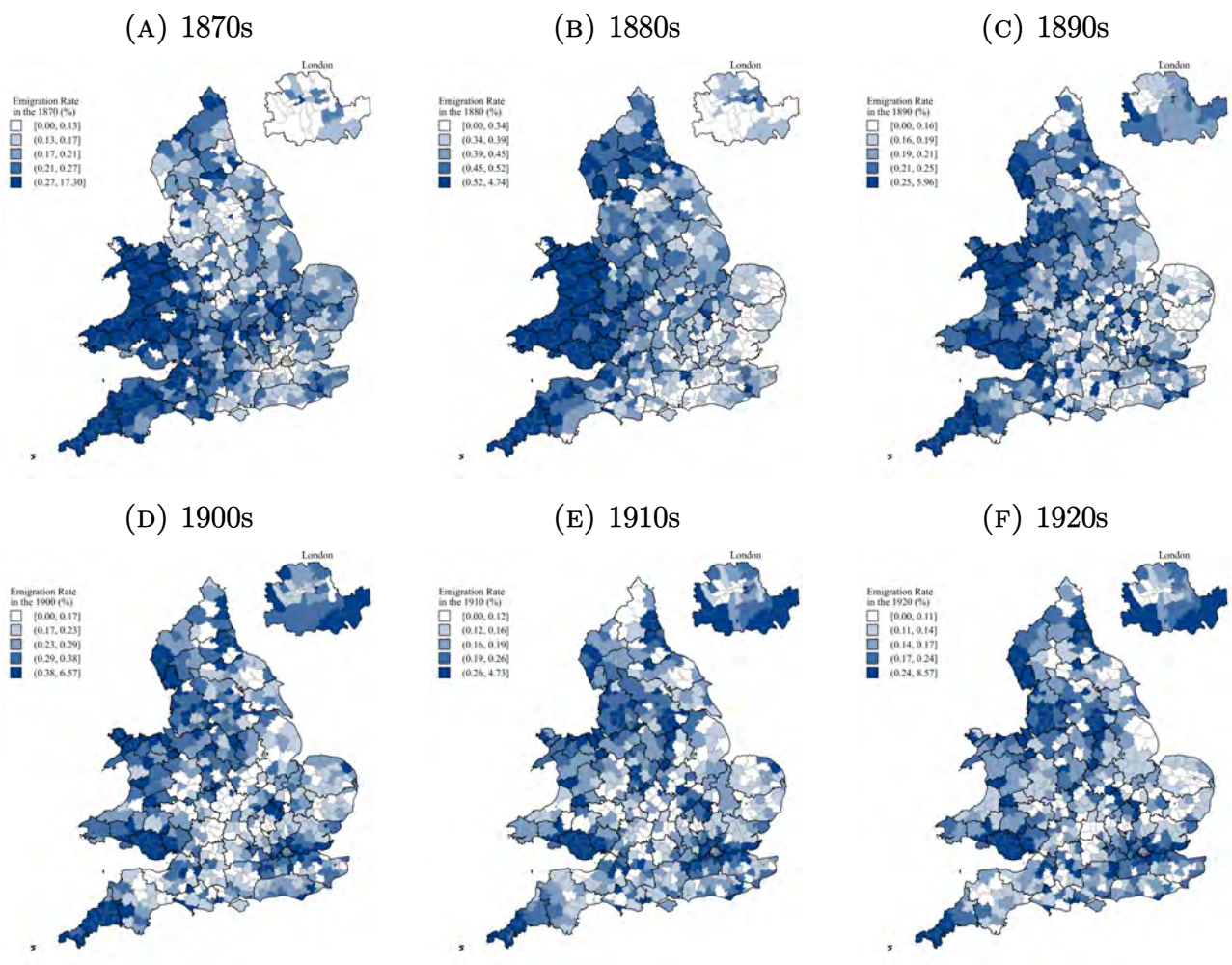
Notes. This Figure reports the distribution of the match quality in terms of name and surname similarity for the set of records with no more than two matches in the baseline sample. The similarity measure we use to construct the links is the Jaro-Winkler. This string metric measures the edit distance between the name and surname of the British immigrant recorded in the US census and their match(es) in the UK census. Panel A.10a reports the distribution of the name similarity; Panel A.10b 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 page(s) A15.

FIGURE A.11. 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 page(s) 11, A16.

FIGURE A.12. 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(s) 11.

B Additional Results

This section presents in some detail several additional results that are mentioned in passing in the main text: the selection of British emigrants (Section B.I), the long-run association between emigration and innovation (Section B.II), and the possibility that immigrants sort into counties that are similar to their origin area (Section B.III).

B.I Selection of British Emigrants

The historical scholarship argues that the English and Welsh mass migration to the US starkly differed from that of other countries (Berthoff, 1953; Baines, 2002). Unlike other European countries, such as Germany, Sweden, or Italy, UK emigration to the US in the second half of the nineteenth century was not a low-skilled rural phenomenon. Especially after the 1880s, people started to leave urban, industrial areas. Importantly, emigrants did not represent the bottom of the human capital distribution, as was the case in Italy (Spitzer and Zimran, 2018) or Norway (Abramitzky *et al.*, 2014). This is crucial 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 was the case, it would be equally unlikely that those rural areas would have the ability to reproduce US patents. While these considerations are helpful for our analysis, they largely rely on anecdotal evidence or analyses of incomplete census sources. In this section, we present evidence on the selection of English emigrants to the United States and returning migrants to the United Kingdom relative to the staying population in the UK. To construct these data, we exploit the linked US-UK migrants sample and the English population census.

Table B.1 presents. Column (1) refers to non-migrants, and columns (2) and (5) refer to emigrants and return migrants, respectively. In columns (3) and (6), we compute 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. Overall, 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 lesser-skilled sector compared to emergent industrial jobs, and are more likely to work as professionals, public officers, in transports—which, in this period, would mainly comprise railway workers—and as utility workers. According to this sketched first inquiry, it is likely that the decision to return to the United Kingdom was relatively more common among those who were less successful

in their American enterprise.

Individuals who migrated to the United States are more likely to originate from the North West—including the industrial Lancashire districts—South West—especially the rural areas of Cornwall and Devon—and Wales. As mentioned in the main text, the origin of emigrants shifts over time from the mainly rural areas in Southern England to the industrial regions in the North and the Midlands. It appears that the probability of returning was not homogeneous across sending regions. Return migrants are less likely to reside in the East, in Wales, and in the West Midlands, while they are more common in the London area and, as are the emigrants, in South West. The different geographic distribution of emigrants and return migrants is crucial when disentangling their contributions to innovation activity in the UK.

B.II Long-Run Association between Emigration and Innovation

We now investigate the persistence of the effect of exposure to foreign knowledge through migration ties on the direction of patenting activity. While this exercise cannot be tasked with any claim of causality, it nonetheless suggests the possible far-reaching effects of out-migration on innovation.

We estimate the following regression:

$$\text{Patents}_{ik,t} = \alpha_{i\times k} + \alpha_t + \sum_{\tau\in\mathcal{T}} \beta^\tau \left[\text{Knowledge Exposure}_{ik} \times 1\left(t = \tau \,|\, t = \tau + 1\right)\right] + \varepsilon_{ik,t} \tag{B.1}$$

where i, k, and t denote a district, technology class, and year, respectively. In this setting, we have $t \in [1940, 2015]$. The term $\text{Knowledge Exposure}_{ik}$ refers to knowledge exposure in the years 1900–1930, i.e., before the sample period. To reduce noise in the estimated β^τ coefficients, we conflate years in $\mathcal{T}$ in biennial windows. The estimated set of β^τ expresses the conditional correlation between historical exposure to knowledge exposure and innovation activity in the two-year window indexed by τ.

In Figure B.1, we report the set of estimated β^τ over time. The correlation between historical knowledge exposure and patenting activity remained positive and significant until the early 1980s, although it—reassuringly— decreased over time. We interpret this as 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 thus only include district and decade-fixed effects. Columns report the estimated β^τ by decade. 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 down 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 $\mathbf{P}_{j,t} = \{p_{1j,t}, \ldots, p_{Nj,t}\}$ 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 $\mathbf{P}_{i,t}$ be the portfolio of district i. We define a metric of innovation similarity as follows:

$$\text{Innovation Similarity}_{ij,t} \equiv \frac{\mathbf{P}_{i,t}^{\intercal}\mathbf{P}_{j,t}}{\|\mathbf{P}_{i,t}\| \cdot \|\mathbf{P}_{j,t}\|} = \frac{\sum_k p_{ki,t} p_{kj,t}}{\sqrt{\sum_k p_{ki,t}^2}\sqrt{\sum_k p_{kj,t}^2}} \leq 1 \tag{B.2}$$

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

We then estimate variations on the following simple linear probability model:

$$\text{Emigrants}_{i \to j,t} = \alpha_{i \times j} + \alpha_t + \beta \times \text{Innovation Similarity}_{ij,t} + X_{ij,t}\Gamma + \varepsilon_{ij,t} \tag{B.3}$$

where the dependent variable is the flow of emigrants from district it to county j in decade d, and $\alpha_{i \times j}$ denotes county-by-district fixed effects. The coefficient β thus yields the correlation between the similarity of innovation activity and migration flows. The dependent variable is measured in logs, and standard errors are two-way clustered by district and county. Under sorting, one would expect $\hat{\beta} > 0$.

We test this prediction in Table B.3. We find limited evidence that the innovation similarity between origin and destination areas explains migration patterns. The regression coefficients are positive but they are seldom significant both contemporaneously (column 1) as well as lagging them by one (column 2) or two (column 3) decades. 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 similar. We do not wish to over-emphasize these results. Our measure of innovation similarity, while intuitive, 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. This insight is consistent with the fact that the OLS and the 2SLS estimates presented in the text are quantitatively similar.

B.IV Anecdotal Evidence of Return Innovation

Who were the immigrants that contributed to the diffusion of US technology in Britain? History is rife with examples of skilled artisans, entrepreneurs, and factory workers who were exposed to some novel technology where they settled and promoted its diffusion, or in some cases appropriated it, in the UK.

In this section, we provide three instructive examples. All three are cases of return migration. Historical records typically focus on successful migrants who, upon returning, bring their technology to their origin areas and promote economic development there. The statistical analysis that we present later, 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 Kelly-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.[11] 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 inventions. Second, technology transfer did not necessitate the very upper tail of the human capital distribution. Skilled workers, such as puddlers, could be the 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.

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

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 United States, first to New York and then to Connecticut. There, he took on apprenticeships in engineering and metal-working firms. He obtained several engineering patents—chiefly 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. A wealthy man respected for his philanthropic endeavors, he was elected mayor of Leeds in 1873. 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.

TABLE B.1. Comparison between English Emigrants and Stayers

	Non-Migrants	Emigrants			Return Migrants		
	Mean	Mean	Difference	Std. Err.	Mean	Difference	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Employment by Sector (Dependent variable = 100 if individual employed in:)							
Agriculture	29.349	28.995	-0.100	(0.072)	26.255	-1.369***	(0.244)
Chemicals	0.949	0.916	-0.057***	(0.015)	1.006	-0.035	(0.055)
Construction	15.686	16.014	0.068	(0.059)	17.029	0.444**	(0.210)
Engineering	14.501	14.612	0.225***	(0.056)	14.280	0.185	(0.195)
Entrepreneur	0.007	0.003	-0.003***	(0.001)	0.004	-0.001	(0.004)
Liberal Professions	3.928	3.184	-0.792***	(0.029)	4.516	0.247**	(0.118)
Metallurgy	3.110	3.704	0.514***	(0.030)	3.247	-0.068	(0.097)
Public Administration	3.497	3.437	-0.124***	(0.029)	4.007	0.220**	(0.110)
Textiles	9.488	10.002	0.658***	(0.049)	8.135	-0.226	(0.154)
Trade	7.909	7.852	-0.068	(0.045)	8.534	-0.038	(0.158)
Transports	10.771	10.475	-0.317***	(0.048)	11.759	0.482***	(0.180)
Utilities	0.805	0.807	-0.003	(0.014)	1.226	0.159***	(0.061)
Panel B. Region of Origin (Dependent variable = 100 if individual resides in:)							
East	10.494	8.436	-2.061***	(0.033)	10.017	-0.328**	(0.149)
East Midlands	6.384	5.883	-0.519***	(0.028)	6.212	-0.062	(0.119)
Greater London	13.639	12.059	-1.712***	(0.040)	15.358	1.733***	(0.182)
North East	6.580	7.107	0.562***	(0.031)	6.436	-0.210*	(0.121)
North West	17.431	19.598	2.521***	(0.047)	17.287	-0.407**	(0.184)
South East	12.408	10.443	-1.918***	(0.037)	12.899	-0.139	(0.167)
South West	6.444	7.377	0.923***	(0.033)	6.521	0.932***	(0.126)
Wales	6.450	8.365	1.918***	(0.030)	6.379	-1.204***	(0.115)
West Midlands	11.189	11.396	0.182***	(0.038)	10.413	-0.565***	(0.149)
Yorkshire	8.596	8.900	0.057	(0.035)	7.874	0.098	(0.133)

Notes. This Table displays the selection of emigrants and returning migrants to and from the United States 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. The associated robust standard errors are displayed in parentheses in columns (4) and (7). Referenced on page(s) B34.

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

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

	(log) Number of Patents by Technology Class						
	1940s	1950s	1960s	1970s	1980s	1990s	2000s
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Agriculture	0.078***	0.044***	0.029***	0.029***	0.017*	0.016**	0.020***
	(0.012)	(0.009)	(0.010)	(0.010)	(0.009)	(0.008)	(0.007)
Building	0.186***	0.158***	0.129***	0.112***	0.055***	0.025**	0.013
	(0.017)	(0.022)	(0.016)	(0.017)	(0.011)	(0.012)	(0.010)
Chemistry	0.040***	0.037***	0.033***	0.019***	0.014***	0.007***	0.006***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Electricity	0.046***	0.045***	0.036***	0.032***	0.022***	0.007	0.005
	(0.007)	(0.008)	(0.007)	(0.007)	(0.005)	(0.005)	(0.004)
Engineering	0.112***	0.101***	0.110***	0.092***	0.062***	0.018	0.031***
	(0.020)	(0.016)	(0.020)	(0.019)	(0.016)	(0.011)	(0.008)
Engines, Pumps	0.053***	0.048***	0.050***	0.043***	0.023***	0.014***	0.002
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
Food	0.098***	0.098***	0.098***	0.082***	0.057***	0.058***	0.046***
	(0.012)	(0.011)	(0.012)	(0.010)	(0.011)	(0.005)	(0.005)
Health, Amusement	0.040***	0.037***	0.036***	0.029***	0.016***	0.006***	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Instruments	0.109***	0.089***	0.106***	0.090***	0.071***	0.054***	0.033***
	(0.009)	(0.011)	(0.008)	(0.011)	(0.008)	(0.006)	(0.006)
Lighting, Heating	0.259***	0.191***	0.253***	0.180***	0.157***	0.060***	0.020
	(0.037)	(0.040)	(0.031)	(0.029)	(0.038)	(0.021)	(0.013)
Metallurgy	0.107***	0.105***	0.093***	0.087***	0.059***	0.030***	0.021***
	(0.008)	(0.007)	(0.009)	(0.007)	(0.007)	(0.006)	(0.005)
Mining	0.047***	0.039***	0.039***	0.030***	0.019***	0.016***	0.006*
	(0.005)	(0.005)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
Personal Articles, Furniture	0.108***	0.124***	0.112***	0.112***	0.132***	0.011	-0.020
	(0.029)	(0.028)	(0.027)	(0.028)	(0.035)	(0.019)	(0.018)
Printing	0.063***	0.054***	0.050***	0.036***	0.020***	0.001	0.004
	(0.007)	(0.006)	(0.005)	(0.006)	(0.006)	(0.005)	(0.004)
Separating, Mixing	0.064***	0.061***	0.061***	0.048***	0.031***	0.010***	0.002
	(0.006)	(0.005)	(0.005)	(0.006)	(0.004)	(0.004)	(0.004)
Shaping	0.072***	0.060***	0.061***	0.040***	0.028***	0.018***	0.012***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)
Textiles	0.344***	0.125***	0.121***	0.103***	0.142***	0.084***	0.052**
	(0.079)	(0.028)	(0.033)	(0.024)	(0.031)	(0.027)	(0.023)
Transporting	0.047***	0.045***	0.041***	0.032***	0.019***	0.005	0.000
	(0.005)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
Weapons, Blasting	0.061***	0.046***	0.041***	0.033***	0.017***	-0.005*	-0.004
	(0.004)	(0.005)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)

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 decade frequency between 1940 and 2010. 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-time, 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(s) B35.

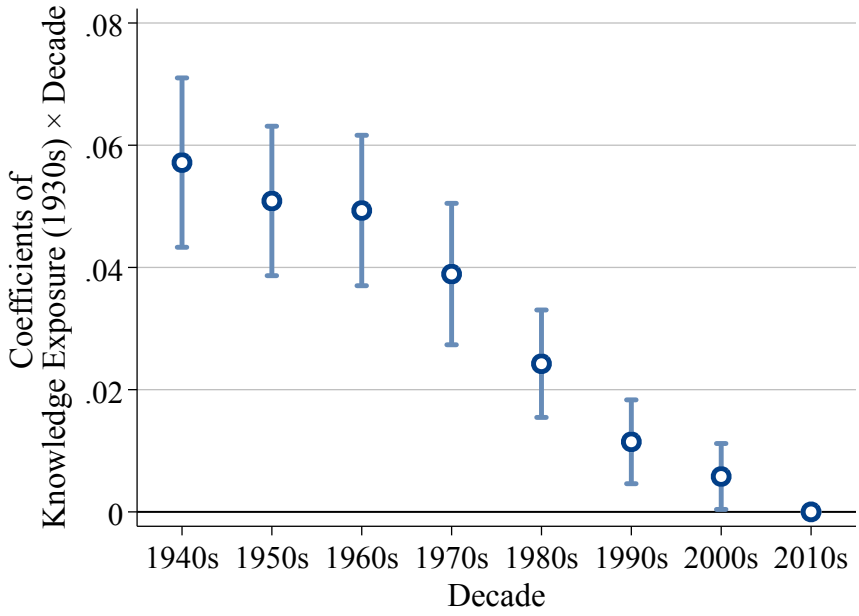
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

TABLE B.3. Assortative Matching of British Immigrants in the United States

	Dep. Var.: Number of Emigrants					
	Full Sample			Positive Migration Flows		
	(1)	(2)	(3)	(4)	(5)	(6)
Innovation Similarity	2.804			99.151*		
	(2.054)			(52.485)		
Innovation Similarity$_{t-10}$		4.876*			96.833*	
		(2.539)			(58.643)	
Innovation Similarity$_{t-20}$			0.888			-1.719
			(2.262)			(79.780)
County-District FE	Yes	Yes	Yes	Yes	Yes	Yes
County-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
District-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
Number of Counties	1,767,996	1,737,939	1,643,487	229,326	183,993	138,281
Observations	8,783,947	7,082,276	5,304,087	769,818	570,248	383,520
Mean Dep. Var.	0.043	0.045	0.043	0.457	0.526	0.544

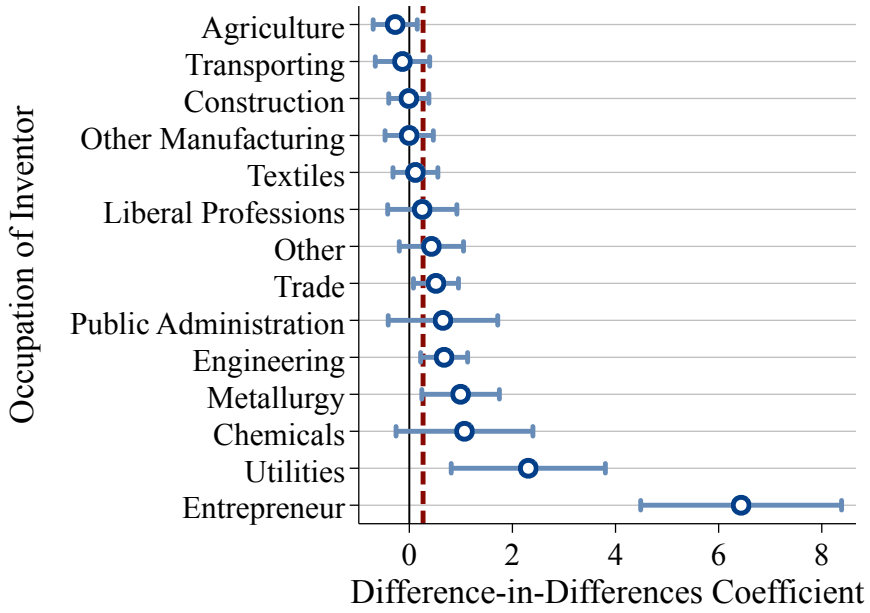
Notes. This Table tests the hypothesis that British immigrants settled in US counties that innovated in the same fields of their district of origin. The unit of observation is a district-county pair. Units are observed at a decade frequency between 1870 and 1930. The dependent variable is the number 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. All regressions include county-by-district, county-by-time, and district-by-time fixed effects. Standard errors are clustered two-way by district and county and are displayed in parentheses. Referenced on page(s) 17, B36. *: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

FIGURE B.1. Long-Run Correlation between Exposure to US Innovation Activity and Innovation in the United Kingdom: Pooled Regression



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-technology pair observed at a decade frequency between 1940 and 2010. The dependent variable is the (log) number of patents. The independent variable is an interaction term between exposure to US technology over the 1930s and decade dummies. The 2010 decade dummy serves as the baseline category. Each dot reports the coefficient of an interaction term by decade. The regression includes district-by-time, district-by-technology, and technology-by-time fixed effects. Standard errors are clustered two-way by district and technology. Bands report 95% confidence intervals. Referenced on page(s) B35.

FIGURE B.2. Heterogeneous Effects of Within-Neighborhood Emigration on Innovation by Occupation of the Stayers



Notes. This Figure reports how emigrants to the United States 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 (log) number of patents produced by the members of the family. 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 county level. The bands report 95% confidence intervals. Referenced on page(s) 30.

C Robustness Analyses

In this section, we describe in detail the exercises we perform to assess the robustness of the results presented in the paper on the OLS and 2SLS analysis (Section C.I), the double and triple differences regressions (Section C.II), and the neighborhood analysis (Section C.III).

C.I OLS and 2SLS Analysis

C.I.1 Alternative Dependent Variables

In the principal analysis, we use the (log 1+) number of patents at varying levels of aggregation as the dependent variable. Following Chen and Roth (2024), who note that this transformation makes the estimates of the average treatment effect scale-dependent, in Table C.1–C.2 (columns 1–5), we show that the results are robust using a battery of alternative transformations.

In the baseline analysis, we measure the originality of British patents and their similarity to US patents using a five-year window, following (Kelly *et al.*, 2021), as detailed in Appendix A.II. In columns (6–8) and (9–11), however, we show that the results remain robust when adopting different time windows to compute these text-based measures.

In addition, in Table C.3, we restrict the dependent variable to comprise only patents with at least one firm assignee. This restriction generates a series of patenting that can potentially reflect technology adoption and diffusion within firms rather than the activity of independent inventors. We confirm that the baseline results are confirmed using this stricter definition of patenting.

C.I.2 Alternative Definitions of Knowledge Exposure

In Table C.4, we employ four alternative measures of knowledge exposure. First, we take the log of the baseline. Second, we construct a measure that fixes bilateral emigrant flows:

$$\text{Knowledge Exposure}^{2}_{ik,t} = \sum_{j} \left(\frac{\text{Patents}_{jk,t}}{\text{Patents}_{j,t}} \times \text{Emigrants}_{i \to j, 1880} \right) \tag{C.1}$$

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

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

Compared to the main measure, this ensures that knowledge exposure does not conflate variation in patenting activity across counties determined or influenced by English immigrants. Finally, we define an alternative measure that leverages the *stock*, instead of the *flow* of patents issued:

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

The idea behind (C.3) is that specialization can be defined in terms of the cumulative number of patents filed before the given period. Finally, we construct the baseline measure but using the level of patents instead of the share:

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

This last metric accounts for the fact that, in small counties with little patenting activity, the share of patents would misrepresent the actual composition of innovation relative to large counties with diversified innovation portfolios. In practice, however, the measure in (C.4) is not vastly different from the baseline metric because few immigrants settled in those small counties to begin with. In Table C.4, we show that all these measures yield quantitatively similar results.

C.I.3 Alternative Standard Errors

In Figure C.1, we estimate the baseline regressions using several estimators for the standard errors. In particular, we employ the estimator developed by Conley (1999) to account for the potential spatial autocorrelation in both emigration rates and exposure to US innovation across technology classes. All the results remain statistically significant at standard confidence levels using these alternative estimators.

C.I.4 Instrumental Variable Strategy

We construct two instruments for out-migration to the United States. The logic of both instruments is to break the assortative matching dynamics by randomizing British immigration across US counties. The first instrument uses the expansion of the railway network in the United States over the course of the 1800s to predict county-level immigration along the lines of Sequeira *et al.* (2020). The second instrument builds on Card (2001) and constructs county-level immigration by interacting county-specific inflows of British immigrants at baseline with aggregate immigration inflows in the United States.

Railway Instrument The first instrument—which we refer to as "railway instrument"—leverages the expansion of the railway network to predict immigration across counties. Specifically, we predict the county-level immigrant share, which is not specific to British immigrants, from a regression between the actual immigrant shares and an interaction between the timing of connection to the railway network and the aggregate inflow of immigrants. We control for county-level unobserved time-invariant heterogeneity and several other potential confounding variables at the county level. This strategy closely mimics the instrument developed by Sequeira *et al.* (2020) to estimate the long-run effect of immigration in the US.

To construct such shocks, we follow a two-step procedure. We first estimate the following zero-stage equation:

$$\begin{aligned}\text{Immigrant Share}_{j,t} = \alpha_j + \alpha_t + \beta \text{Immigrant Share}_{j,t-1} + \gamma I^{\text{Rail}}_{j,t-1} + \\ + \delta \left(I^{\text{Rail}}_{j,t-1} \times \text{Immigrant Flow}_{t-1} \right) + \zeta \left(\text{Industrialization}_{t-1} \times I^{\text{Rail}}_{j,t-1} \right) + \\ + \eta \left(\text{GDP Growth}_{t-1} \times I^{\text{Rail}}_{j,t-1} \right) + X'_{j,t-1} \Theta + \varepsilon_{j,t}\end{aligned} \tag{C.5}$$

where (Immigrant Share) is the share of foreign-born individuals, $I^{\text{Rail}}_{j,t}$ is a dummy variable returning value one if county j is connected to the railway network in decade t, and zero otherwise, (Immigrant Flow) is the aggregate immigration inflow computed from Willcox (1928), (Industrialization) is an index of industrial production computed by Davis (2004), and annual average GDP growth is obtained from Maddison (2007) data. The other terms control for confounding factors and non-random connections to the railway network. The term X includes log-population density, lagged urbanization, and an interaction between lagged urbanization and lagged aggregate immigrant flow. The core of the identification strategy that we borrow from Sequeira *et al.* (2020) is to exploit variation generated by the interaction between aggregate immigration inflows and connection to the railway network (δ). The underlying idea is that connection to the railway only induces a larger immigrant inflow if it occurs during a period of high immigration. If this reasoning holds, the estimate of β should be close to zero, and that of δ should be positive. We confirm these predictions in Appendix Table C.6.

Leave-out Instrument An alternative approach—which, for brevity, we refer to as "leave-out instrument"—to randomize immigration across counties is constructing an instrument along the lines of Card (2001) and Tabellini (2020). The instrument predicts the county-level immigrant share by interacting the number of immigrants by country of origin that had settled in the county before the sample period—i.e., in the 1860s—with the subsequent aggregate inflow of immigrants over time. We exclude British

immigrants from the calculation.[12]

The leave-out instrument that borrows heavily on the literature that uses shift-share instruments to estimate the effects of immigration (e.g. Card, 2001; Tabellini, 2020). The rationale that underlies this approach is that if assortative matching across counties by British immigrants is the main threat to identification in the baseline regression, then it is possible to leverage the distribution of immigrants from *other* countries to construct county-level immigration shocks that yield consistent estimates because they do not reflect such assortative matching effects.

In practice, let ω_j^M be the share of immigrants from country M that settle in county j in the period 1860-1870, i.e., before the beginning of the analysis years. We then compute the aggregate inflow of immigrants from country M in each subsequent decade and construct the predicted immigrant inflows as

$$\widehat{\text{Immigrant Share}}_{j,t} = \frac{1}{\text{Population}_{j,t}} \sum_{\substack{M \neq \text{UK} \\ M \in \mathcal{M}}} \left(\omega_j^M \times \text{ Immigrant Inflow}_t^M \right) \tag{C.6}$$

where $\mathcal{M}$ is a set of origin countries. Both (C.5) and (C.6) yield a set of county-specific immigration shocks that do not conflate the immigration patterns of the British.

We allow multiple sets of origin countries $\mathcal{M}$ to account for possible correlation between British immigrants and those from other nationalities. The results are displayed in Table C.7, which collects all the various leave-out instruments. In particular, we drop all countries in Northern Europe (column 2), which may have been more similar to England and Wales. Moreover, in column (6), we only include non-European countries and show that results hold nonetheless. The coefficients remain relatively stable across all specifications, indicating the possibility that assortative matching may be a quantitatively mild issue.

Construction of the Instruments Both approaches yield a series of predicted immigrant shares at the county level. Let ω_{jt} denote such predicted shares. Then, we construct district-level out-migration as follows:

$$\widehat{\text{Emigrants}}_{i,t} = \sum_{j \in J} \omega_{j,t} \times \text{US Emigrants}_{i \to j, 1880}, \tag{C.7}$$

where $\text{US Emigrants}_{i \to j, 1880}$ denotes the number of emigrants that leave district i and settle in county j at the beginning of the sample, i.e., between 1870 and 1879. Equation (C.7) generates an instrument for the number of US emigrants across districts in the estimating equation (C.6). To construct an

[12] In Appendix Table C.7, we show that the results remain qualitatively unchanged when we vary the set of countries included in the construction of the instrument. In the baseline case, we exclude immigrants from England, Wales, and Scotland.

instrument for knowledge exposure, we plug the predicted district-county flows from (C.7) into the baseline measure of knowledge exposure (1).

The validity of both instruments hinges on the quasi-random assignment of immigration shocks across counties. In the case of the railway instrument, shocks are conditionally randomly assigned if the expansion of the railway network is not correlated with the settlement decisions of British immigrants. In the case of the leave-out instrument, the validity of the design requires that the initial location of non-British immigrants does not predict the settlement decisions of British immigrants. Following Borusyak, Hull and Jaravel (2022), we show evidence that supports both presumptions. Appendix Table C.8 shows that while out-migration correlates with district-level observable characteristics (columns 1–2), predicted immigration shares do not (columns 3–4 for the railway instrument, and 5–6 for the leave-out instrument). Similarly, in Appendix Table C.9, we confirm that while immigration correlated with most county-level variables (columns 1–2), both instrument displays smaller and insignificant correlations with the same variables (columns 3–6). When we employ the two-stage least squares estimator, we also report the companion F-statistic to ascertain the relevance of the first stage.

First-Stage Correlations Figure C.2 reports the first-stage correlation between out-migration and the railway (Panel C.2a) and the leave-out (Panel C.2b) instruments and between exposure to US technology and the railway (Panel C.2c) and the leave-out (Panel C.2d) instruments. For consistency with the main analysis, the two top panels include district and year-fixed effects, whereas the two bottom panels include district-by-year and technology-fixed effects. The gray dots indicate single observations, whereas the blue dots report binned means.

We find a positive and statistically significant correlation between the instruments and the observed out-migration and knowledge exposure data. Importantly, the instruments explain a large portion of the variability of the endogenous variables as well, as one can see by looking at the R^2, which we report after residualizing both dependent and independent variables against the included fixed effects.

Falsification of the Instruments The validity of the shift-share instrument for knowledge exposure that we construct hinges on the exogeneity of the shocks constructed using either (C.5) or (C.6), following Borusyak *et al.* (2022). In practice, they advise conducting two types of falsification tests. First, shocks should be orthogonal to observed county-level characteristics. Second, the instrument should not be systematically correlated with district-level observable variables. The first test provides evidence of the exogeneity of the shocks, while the second should support the exclusion restriction that underlies the instrument.

We perform the results of the first exercise in Table C.9. Columns (1–2) display the correlation of the observed immigration shares with county-level observable characteristics. As expected, immigration is not random as it tends to be concentrated in larger counties, which also display higher patenting activity. In columns (3–4) and (5–6), we report the correlation between the predicted immigrant shares using the railway-based and the leave-out approaches, respectively. We fail to detect a statistically significant correlation between the so-constructed immigrant shares and the large majority of county-level observable variables.[13] This provides reassuring evidence in favor of the validity of the instruments.

We report the second exercise in Table C.8. Columns (1–2) display the correlations between district-level variables and observed emigration; columns (3–4) and (5–6) display the correlations with the railway and the leave-out instruments, respectively. Unsurprisingly, districts featuring higher emigration flows are larger, produce more patents, and have a larger share of the population working in agriculture and textiles. On the other hand, synthetic out-migration, whether constructed using the railway or the leave-out shocks, is not correlated with any such variables. Once more, we interpret these results as evidence supporting the validity of the shift-share research design.

C.II Double and Triple Differences Analysis

C.II.1 Alternative Dependent Variables

As in the previous analysis, the baseline results use the (log 1+) as the main dependent variable. In columns (1–4) of Table C.10, we adopt different transformations of the dependent variables and confirm that our results hold throughout. In the baseline analysis, we also adopt a five-year window to compute the text-based measures (originality and similarity). In columns (5–7) and (8–10), we show that the qualitative nature of the results remains unchanged when using different thresholds.

C.II.2 Alternative Definitions of the Shocks

In the baseline analysis, we consider a district in the UK to be exposed to an innovation shock in the United States if the number of emigrants from that district that are exposed to an innovation shock is in the top 5% of the overall distribution of exposed emigrants, net of district, year, and, when applicable, technology class fixed effects. In Table C.11, we show that results remain qualitatively unchanged when using the top 1% and 10% as alternative thresholds. Unsurprisingly, the magnitude of the estimates of the average treatment effect increases in the restrictiveness of the threshold. As

[13] Even when the correlation remains significant, the standardized beta coefficient is substantially lower than in the benchmark column (1) and the statistical significance is marginal.

we move from the top 10% to the top 1% we require that increasingly more emigrants are exposed to a shock in the United States. According to the logic explained in the main text, this shift would translate into a relatively more intense exposure to the shock from the perspective of the district.

C.II.3 Alternative Standard Errors

In Figure C.5, we show that the effect of US innovation shocks on the volume of patents, either pooled (Panel C.5a) or broken down by technology (Panel C.5b), remains statistically significant when using alternativeC.5a estimators for the standard errors. In Panel C.5a, we find that the statistical significance of the effect of the pooled shocks is diminished when adjusting for spatial autocorrelation. The estimates remain, however, significant at the 10% level and, importantly, adjusting for spatial autocorrelation does not alter the statistical significance of the triple differences estimates shown in Panel C.5b.

C.II.4 Alternative Estimator

The roll-out of US innovation shocks across districts is staggered, because different districts—or district-technology class pairs—can be exposed to a shock to US innovation at different times. Goodman-Bacon (2021) shows that, in this case, the standard two-way fixed-effects (TWFE) may fail to estimate the average treatment effect if the effect is heterogeneous across units and/or over time. In Figure C.3, we show that all the results shown in the main text remain unchanged when employing the estimator developed by Sun and Abraham (2021). In particular, we estimate similar pre- and post-treatment coefficients for the volume of patents (Panels C.3a–C.3b) and the similarity between US and UK patents (Panels C.3c–C.3d). We estimate standard TWFE models because they allow for simpler deviations from the baseline estimation strategy for the standard errors and the heterogeneity analyses, as previously discussed.

C.II.5 Innovation Shocks in the United States

The double and triple differences analyses implicitly rely on the fact that our methodology to flag innovation shocks in US counties can reliably isolate periods of intense patenting activity. We test this assumption in Table C.12. We consider the universe of US counties (in columns 1–2) and county-technology pairs (columns 3–4). The dependent variable is the (log) number of patents (in columns 1 and 3) and the (log) number of original patents (in columns 2 and 4), and the treatment returns a value of one for treated units, i.e. counties or county-technology pairs after the innovation shocks occur, and zero otherwise. The regressions are saturated with fixed effects. These double and triple

differences regressions do not capture a causal effect but rather indicate the actual intensity of the shock to US innovation activity.

An innovation shock is associated with a 50% increase in the number of patents (column 1). In the triple-differences setting, patenting increases by 10% on average after the shock. In Figure C.6, we repeat the estimation in a flexible setting, which uncovers substantial heterogeneity over time. When counties undergo an innovation shock, they produce 125% more patents. The effect persists over time, but the spike is short-lived. Within technologies, a US innovation shock is associated with an 80% shock to innovation activity, which reverts to the pre-shock mean very rapidly. We thus conclude that our methodology successfully isolates sharp and large shocks to innovation activity in the United States. The salience of the shock is, in both cases, relatively short-lived, albeit substantial.

C.III Neighborhood Analysis

We consider two very simple 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.2, the number of matches in the sample of inventors linked to the census is larger for inventors residing in Wales. By excluding them from the estimation sample, we thus 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, before migrating, he lived within five kilometers of the inventor. In Figure C.7, we consider five alternative threshold distances: one, two, three, ten, and twenty kilometers. We then estimate the baseline regression for these various thresholds and separately report the estimated average treatment effects. The coefficients remain positive for all thresholds, but they are largest in magnitude—and statistically significant at the 1% level—for three- and five-kilometer neighborhoods. The estimate for the ten-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. This notwithstanding, it is reassuring that the coefficients remain positive throughout.

TABLE C.1. Effect of US Emigration on Innovation: Alternative Dependent Variables

	Number of Patents					k-Originality			k-Similarity with US Patents		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Number	$\ln(1+\cdot)$	$\ln(\varepsilon+\cdot)$	$\ln(\cdot)$	Arcsinh	1 yr.	5 yrs.	10 yrs.	1 yr.	5 yrs.	10 yrs.
Dependent Variable Mean	*161.499*	*3.512*	*3.254*	*3.568*	*4.105*	*2.083*	*2.037*	*2.021*	*33.833*	*162.046*	*305.021*
Panel A. OLS Estimates											
US Emigrants (1,000s)	403.841**	0.686***	0.572*	0.685***	0.657***	0.712***	0.987***	0.915***	3.783**	15.014*	23.412
	(166.325)	(0.226)	(0.310)	(0.236)	(0.238)	(0.252)	(0.262)	(0.255)	(1.807)	(8.577)	(15.728)
Panel B. 2SLS Estimates: Railway Instrument											
US Emigrants (1,000s)	-1844.586**	1.496**	2.159***	1.638**	1.772***	1.344	1.402*	0.729	8.300	31.282	31.282
	(733.182)	(0.624)	(0.776)	(0.657)	(0.648)	(0.837)	(0.827)	(0.794)	(9.260)	(45.726)	(45.726)
Kleibergen-Paap F	53.884	53.884	53.884	51.958	53.884	53.884	53.884	53.884	53.884	53.884	53.884
Panel C. 2SLS Estimates: Leave-out Instrument											
US Emigrants (1,000s)	-3163.833**	2.186**	5.226***	2.336**	2.942***	0.650	1.748	0.826	1.409	2.309	2.309
	(1328.979)	(1.021)	(1.355)	(1.048)	(1.024)	(1.368)	(1.320)	(1.358)	(12.689)	(61.550)	(61.550)
Kleibergen-Paap F	40.452	40.452	40.452	39.438	40.452	40.452	40.452	40.452	40.452	40.452	40.452
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Districts	620	620	620	619	620	620	620	620	620	620	620
Observations	3,720	3,720	3,720	3,575	3,720	3,720	3,720	3,720	3,720	3,720	3,720
Mean Dep. Var.	161.499	3.512	3.254	3.568	4.105	2.083	2.037	2.021	33.833	162.046	305.021

Notes. This Table displays the association between emigration and innovation in the United Kingdom. The unit of observation is a district observed at a decade frequency between 1870 and 1930. The dependent variables are: in column (1), the number of patents; in column (2), the log(1+) number of patents; in column (3), the log(0.1+) number of patents; in column (4), the log number of patents, which excludes zeros; in column (5), the inverse hyperbolic sine of the number of patents; in columns (6–8), patents in the top 20% of the novelty distribution in the previous 1, 5, and 10 years; in columns (9–11), the dependent variable is the similarity of British patents with American patents issued in the previous 1, 5, and 10 years. The independent variable is the number of US emigrants. Panel A reports the ordinary least squares estimates; Panel B reports the two-stage least squares estimates obtained using the railway-based instrument; Panel C reports the two-stage least squares estimates obtained using the leave-out instrument. All regressions include district and year-fixed effects. Standard errors in parentheses are clustered at the district level. Referenced on page(s) 24, 24, A9, C44.

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

TABLE C.2. Effect of Exposure to US Knowledge on Innovation: Alternative Dependent Variables

	Number of Patents					k-Originality			k-Similarity with US Patents		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Number	$\ln(1+\cdot)$	$\ln(\varepsilon+\cdot)$	$\ln(\cdot)$	Arcsinh	1 yr.	5 yrs.	10 yrs.	1 yr.	5 yrs.	10 yrs.
Dependent Variable Mean	*0.079*	*1.000*	*-1.389*	*1.678*	*1.236*	*0.403*	*0.382*	*0.376*	*15.432*	*73.914*	*139.440*
Panel A. OLS Estimates											
Knowledge Exposure	0.014***	0.040***	0.027***	0.034***	0.042***	0.035***	0.034***	0.034***	0.137***	0.640***	1.201***
	(0.001)	(0.001)	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.017)	(0.083)	(0.156)
Panel B. 2SLS Estimates: Railway Instrument											
Knowledge Exposure	0.017***	0.043***	0.026***	0.041***	0.045***	0.036***	0.035***	0.036***	0.094***	0.420***	0.420***
	(0.001)	(0.002)	(0.005)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.027)	(0.129)	(0.129)
Kleibergen-Paap F	67.476	67.476	67.476	47.777	67.476	150.084	150.084	150.084	150.084	150.084	150.084
Panel C. 2SLS Estimates: Leave-out Instrument											
Knowledge Exposure	0.017***	0.043***	0.026***	0.040***	0.045***	0.040***	0.040***	0.040***	0.103***	0.481***	0.481***
	(0.001)	(0.002)	(0.005)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.027)	(0.131)	(0.131)
Kleibergen-Paap F	135.149	135.149	135.149	95.316	135.149	135.149	135.149	135.149	135.149	135.149	135.149
District-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Technology-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Districts	621	621	621	620	621	621	621	621	621	621	621
Observations	70,433	70,433	70,433	35,924	70,433	70,433	70,433	70,433	70,433	70,433	70,433

Notes. This Table displays the association between exposure to US technology through out-migration ties and innovation in the United Kingdom. The unit of observation is a district-technology class pair observed at a decade frequency between 1870 and 1930. The dependent variables are: in column (1), the number of patents; in column (2), the log(1+) number of patents; in column (3), the log(0.1+) number of patents; in column (4), the log number of patents, which excludes zeros; in column (5), the inverse hyperbolic sine of the number of patents; in columns (6–8), patents in the top 20% of the novelty distribution in the previous 1, 5, and 10 years; in columns (9–11), the dependent variable is the similarity of British patents with American patents issued in the previous 1, 5, and 10 years. The independent variable is the knowledge exposure metric described in the main text. Panel A reports the ordinary least squares estimates; Panel B reports the two-stage least squares estimates obtained using the railway-based instrument; Panel C reports the two-stage least squares estimates obtained using the leave-out instrument. All regressions include district-by-time and technology-fixed effects. Standard errors in parentheses are clustered two-way at the district and technology class level. Referenced on page(s) 20, 21, 24, 24, A9, C44.

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

TABLE C.3. Emigration to the United States and Innovation: Patents with Firm Assignee

	Patents by District			Patents by District-Technology		
	(1) Number	(2) High-Impact	(3) US Similarity	(4) Number	(5) High-Impact	(6) US Similarity
Dependent Variable Mean	*2.686*	*1.426*	*107.436*	*0.619*	*0.216*	*36.724*
US Emigrants (1,000s)	0.728***	0.765***	41.363***			
	(0.189)	(0.210)	(10.374)			
Knowledge Exposure				0.038***	0.024***	1.225***
				(0.001)	(0.001)	(0.061)
District FE	Yes	Yes	Yes	–	–	–
Decade FE	Yes	Yes	Yes	–	–	–
District-Decade FE	–	–	–	Yes	Yes	Yes
Technology FE	–	–	–	Yes	Yes	Yes
Number of Districts	621	621	621	621	621	621
Observations	3,726	3,726	3,726	70,433	70,433	70,433

Notes. This Table displays the association between emigration and innovation in the United Kingdom. In columns (1–3) (resp. 4–6), the unit of observation is a district (resp. district-technology pair) observed at a decade frequency between 1870 and 1930. In columns (1) and (4), the dependent variable is the (log) number of patents; in columns (2) and (5), we include patents in the top 20% of the novelty distribution; in columns (3) and (6), the dependent variable is the average similarity between UK and US patents. We only include patents with a firm assignee. In columns (1–3) (resp. 4–6), the independent variable is the number of migrants (resp. exposure to US technology). In columns (1–3), regressions include district and decade-fixed effects; in columns (4–6), regressions include district-by-time and technology-fixed effects. Standard errors in parentheses are clustered at the district level. Referenced on page(s) 25, C44.

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

TABLE C.4. Correlation between Exposure to US Technology and Innovation: Alternative Measures of Knowledge Exposure

	log(1 + Number of Patents)					
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable Mean	*0.999*	*0.999*	*0.999*	*0.999*	*0.999*	*0.999*
Knowledge Exposure	0.373***					
	(0.012)					
Log(1+Knowledge Exposure)		6.944***				
		(0.275)				
Fixed-Emigration Knowledge Exposure			0.655***			
			(0.034)			
Fixed-Innovation Knowledge Exposure				0.297***		
				(0.012)		
Cumulative Knowledge Exposure					0.070***	
					(0.007)	
Level of Patents Knowledge Exposure						0.165***
						(0.004)
District-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Technology FE	Yes	Yes	Yes	Yes	Yes	Yes
Number of Districts	621	621	621	621	621	621
Observations	70,794	70,794	70,794	70,794	70,794	70,794

Notes. This Table displays the association between exposure to US technology through migration ties and innovation in the United Kingdom. The unit of observation is a district-technology pair observed at a decade frequency between 1870 and 1930. The dependent variable is the (log) number of patents. The independent variable is: in column (1), the baseline metric of knowledge exposure; in column (2), the log of the baseline metric; in column (3), the baseline measure but keeping migration ties between districts and counties fixed in 1880; in column (4), the baseline metric but keeping the share of patents across classes fixed in 1880; in column (5), the baseline metric but considering patents as a stock instead of a flow and taking the cumulative number of patents over time; in column (6), the exposure metric interacts migration ties with the number of patents produced in each technology by the county, instead of the share. All regressions include district-by-year and technology-fixed effects. Standard errors in parentheses are clustered at the district level. Referenced on page(s) 25, C44, C45.

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

TABLE C.5. Emigration, Exposure to US Knowledge, and Innovation in the United Kingdom: Poisson Quasi-Maximum Likelihood Estimates

	Patents by District				Patents by District-Technology			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Number	Number	US Patents Similarity	High Impact	Number	Number	US Patents Similarity	High Impact
Dependent Variable Mean	*182.270*	*182.484*	*41.365*	*38.521*	*10.006*	*11.816*	*2.275*	*2.506*
US Emigrants (1,000s)	0.867***	0.385*	0.846***	0.917***				
	(0.212)	(0.230)	(0.214)	(0.282)				
Knowledge Exposure					0.437***	0.033	0.423***	0.754***
					(0.091)	(0.097)	(0.094)	(0.132)
District FE	Yes	Yes	Yes	Yes	–	–	–	–
Decade FE	Yes	–	Yes	Yes	–	–	–	–
Controls × Time	No	Yes	No	No	–	–	–	–
County-Year FE	No	Yes	No	No	–	–	–	–
District-Year FE	–	–	–	–	Yes	Yes	Yes	Yes
Technology FE	–	–	–	–	Yes	–	Yes	Yes
District-Technology FE	–	–	–	–	No	Yes	No	No
Technology-Year FE	–	–	–	–	No	Yes	No	No
Number of Districts	621	620	621	610	621	620	621	610
Observations	3,726	3,720	3,726	3,660	67,754	57,371	67,697	56,164
Std. Beta Coef.	0.000	0.000	0.001	0.001	0.001	0.000	0.003	0.005

Notes. This Table displays the association between emigration and innovation in the United Kingdom. In columns (1–4) (resp. 5–8), the unit of observation is a district (resp. district-technology pair) observed at a decade frequency between 1870 and 1930. Each regression is estimated using the Poisson quasi-maximum likelihood estimator described in Correia *et al.* (2020). The dependent variable is: in columns (1–2) and (5–6), the number of patents; in (3) and (7), the text-based similarity between British and American patents issued in the previous five years; in (4) and (8), the number of patents in the top 20% of the novelty distribution. In columns (1–4) (resp. 5–8), the independent variable is the number of migrants (resp. exposure to US technology). In columns (1) and (3–4), the model includes district and decade-fixed effects; column (2) includes district-level controls measured in 1880 and interacted with decade indicators and county-by-decade fixed effects. In columns (1) and (7–8), regressions include district-by-year and technology-fixed effects; column (6) also includes district-by-technology and technology-by-year fixed effects. Standard errors in parentheses are clustered at the district level. Referenced on page(s) 24.

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

TABLE C.6. Zero-Stage Regressions of the Railway Instrument

	Dep. Var.: Immigrant Share (%)				
	Baseline	Excluding States in...			
	(1)	(2) Northeast	(3) Midwest	(4) South	(5) West
Dependent Variable Mean	*7.831*	*7.282*	*5.400*	*13.077*	*7.556*
$I^{\text{Rail}}_{t-1} \times \text{Immigrant Flow}_{t-1}$	0.034***	0.028***	0.021**	0.028*	0.033***
	(0.008)	(0.009)	(0.008)	(0.015)	(0.008)
I^{Rail}_{t-1}	0.276	0.318	0.169	-0.954	0.368
	(0.314)	(0.313)	(0.316)	(0.716)	(0.316)
County FE	Yes	Yes	Yes	Yes	Yes
Decade FE	Yes	Yes	Yes	Yes	Yes
Number of Counties	2,761	2,545	1,743	1,514	2,606
Observations	17,892	16,380	11,383	9,457	17,218

Notes. This Table reports the results of the zero-stage regressions that we estimate to construct the railway-based county-level immigration shocks. This table largely replicates Sequeira *et al.* (2020). The unit of observation is a county observed at a decade frequency between 1870 and 1930. The dependent variable is the share of the foreign-born population. The main dependent variable is an interaction between the one-decade-lagged national inflow of immigrants and an indicator variable that returns value one if the county was connected to the national railway network in the previous decade and zero otherwise. The regressions also control for the railway indicator, the lagged share of foreign-borns, an interaction between lagged national industrial production and the railway indicator, an interaction between lagged GDP and the railway indicator, population density, the share of the population living in urban centers, and an interaction between the share of the urban population and the national inflow of immigrants. The parameter restriction imposed by the instrument's logic requires that the railway indicator's coefficient be non-positive. In column (1), the sample is the universe of counties; in columns (2), (3), (4), and (5), we drop states in, respectively, the North-East, Midwest, South, and West Census Bureau regions. Each regression includes county and decade-fixed effects. Standard errors, clustered at the county level, are displayed in parentheses. Referenced on page(s) C46.
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

TABLE C.7. Alternative Leave-Out Instrumental Variables

	Leaveout IV excluding immigrants from...					
	(1) UK	(2) Northern Europe	(3) Southern Europe	(4) Eastern Europe	(5) Central Europe	(6) Europe
Panel A. Volume of Emigrants						
US Emigrants (1,000s)	1.866**	2.805***	3.639***	3.676***	2.654***	2.136***
	(0.754)	(0.936)	(1.095)	(1.163)	(0.843)	(0.691)
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,726	3,726	3,726	3,726	3,726	3,726
Kleibergen-Paap's F	50.017	35.147	26.285	22.701	30.429	82.562
Mean Dep. Var.	3.513	3.513	3.513	3.513	3.513	3.513
Panel B. Exposure to US Knowledge						
Knowledge Exposure	0.041***	0.040***	0.037***	0.036***	0.042***	0.043***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
District-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
Technology FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	70,433	70,433	70,433	70,433	70,433	70,433
Kleibergen-Paap's F	142.882	149.732	160.376	163.395	147.070	158.961
Mean Dep. Var.	1.000	1.000	1.000	1.000	1.000	1.000

Notes. This Table reports the association between emigration, exposure to US technology, and innovation in the United Kingdom using alternative leave-out strategies to construct the instrumental variable. The unit of observation in Panel A (resp. B) is a district (resp. district-technology class pair) at a yearly frequency between 1870 and 1930. The dependent variable is the (log) number of patents. In Panel A (resp. B), the independent variable is the number of US emigrants (resp. exposure to US technology through emigration ties). The coefficients report the two-stage least squares estimates using the leave-out instrument constructed by including different groups of immigrants in the shift term: in column (1), we exclude UK immigrants (the baseline scenario); in column (2), we exclude immigrants from Sweden, Denmark, Norway, and Finland; column (3) excludes immigrants from Spain, Portugal, Italy, France, Greece; column (4) excludes immigrants from Poland, Romania, Russian Empire, Bulgaria; column (5) excludes immigrants from Germany, Austria-Hungary, Switzerland, Netherlands, and Belgium. In Panel A, all regressions include district and year-fixed effects; in Panel B, all regressions include district-by-time and technology-fixed effects. Standard errors are shown in parentheses and are clustered by district level. Each column and Panel reports the first-stage F-statistic. Referenced on page(s) C46.

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

TABLE C.8. Falsification of the Instruments: UK Analysis

	Immigration		Railway IV		Leaveout IV	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Demographics						
Population	0.070***	(0.011)	0.026	(0.017)	0.085	(0.087)
Share of Men	0.010	(0.011)	0.017	(0.021)	0.113	(0.144)
Panel B. Employment						
Employment Share	-0.077***	(0.018)	0.022	(0.066)	-0.216	(0.297)
Agriculture	0.037***	(0.009)	0.006	(0.025)	0.237	(0.183)
Chemistry	0.025**	(0.012)	-0.045	(0.027)	-0.168	(0.176)
Construction	0.044***	(0.014)	0.005	(0.043)	-0.190	(0.188)
Engineering	0.063***	(0.018)	0.047	(0.039)	0.285	(0.281)
Metallurgy	0.014	(0.009)	-0.097***	(0.028)	-0.515	(0.345)
Textiles	0.066***	(0.011)	0.037	(0.027)	0.262	(0.198)
Trade	0.053***	(0.013)	-0.062*	(0.033)	-0.274	(0.230)
Transports	0.009	(0.012)	-0.059	(0.039)	0.119	(0.183)
Panel C. Patents						
Total Patents	0.220***	(0.033)	0.029	(0.048)	0.166***	(0.055)
Agriculture	-0.022	(0.017)	-0.011	(0.044)	-0.029	(0.077)
Chemistry	-0.064***	(0.021)	-0.033	(0.066)	0.076	(0.113)
Electricity	0.055***	(0.020)	-0.037	(0.097)	-0.107	(0.185)
Instruments	0.003	(0.018)	0.107*	(0.062)	0.093	(0.105)
Lighting, Heating	-0.008	(0.020)	0.058	(0.053)	0.161*	(0.092)
Metallurgy	-0.013	(0.024)	0.073	(0.046)	0.343***	(0.115)
Personal Articles, Furniture	-0.001	(0.016)	0.003	(0.051)	-0.136	(0.094)
Textiles	0.014	(0.021)	0.029	(0.040)	0.037	(0.071)
Transporting	-0.017	(0.019)	0.050	(0.051)	0.225**	(0.100)

Notes. This Table reports the correlation between district-level variables and observed out-migration (columns 1–2), the railway instrument (columns 3–4), and the leave-out instrument (columns 5–6). Columns (3–6) report two-stage least squares estimates. The unit of observation is a district at a decade frequency between 1870 and 1930. In each row, we report the correlation between one-decade lagged values of the dependent variable, shown on the rows, and the three independent variables, shown on the columns. Each regression includes district and decade-fixed effects. Standard errors are shown in parentheses and are clustered at the district level. Patents are expressed as shares. Both outcome and explanatory variables are taken in log terms and standardized for comparability. Referenced on page(s) C48, C49.

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

TABLE C.9. Falsification of the Instruments: US Analysis

	Immigration		Railway IV		Leaveout IV	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Demographics						
Population	0.083***	(0.021)	-0.093	(0.073)	-0.035	(0.150)
Age	-0.008	(0.007)	-0.023	(0.017)	0.010	(0.034)
Share of Whites	-0.004***	(0.002)	0.045*	(0.026)	0.071	(0.072)
Share Urban	-0.000	(0.007)	0.136	(0.084)	0.385	(0.422)
Income per Capita	0.027***	(0.010)	0.143	(0.096)	0.095	(0.103)
Panel B. Employment						
Employment Share	0.008	(0.006)	-0.022	(0.033)	-0.192	(0.192)
Agriculture	-0.072***	(0.025)	0.088	(0.081)	-0.094	(0.143)
Chemistry	-0.001	(0.010)	0.221	(0.160)	0.216	(0.259)
Construction	0.011	(0.011)	0.231	(0.147)	0.384	(0.374)
Engineering	0.036***	(0.008)	0.209	(0.140)	0.005	(0.070)
Metallurgy	0.001	(0.016)	0.114	(0.084)	0.141	(0.169)
Textiles	-0.007	(0.013)	0.109	(0.068)	0.031	(0.079)
Trade	-0.014*	(0.008)	0.197	(0.121)	0.379	(0.398)
Transports	-0.016***	(0.004)	0.007	(0.017)	0.166	(0.197)
Panel C. Patents						
Total Patents	0.183	(0.113)	-0.872*	(0.499)	-3.915	(3.670)
Agriculture	-0.010*	(0.005)	0.009	(0.017)	-0.062	(0.063)
Chemistry	0.011	(0.012)	0.049	(0.039)	0.155	(0.118)
Electricity	0.018	(0.018)	-0.031	(0.069)	-0.121	(0.154)
Instruments	-0.012*	(0.006)	0.012	(0.026)	-0.045	(0.068)
Lighting, Heating	-0.001	(0.008)	-0.014	(0.025)	0.139	(0.110)
Metallurgy	0.023***	(0.009)	0.033	(0.024)	0.107	(0.078)
Personal Articles, Furniture	-0.015*	(0.008)	-0.025	(0.032)	0.030	(0.073)
Textiles	0.001	(0.007)	0.004	(0.022)	0.059	(0.076)
Transporting	-0.009	(0.009)	-0.031	(0.021)	0.012	(0.050)

Notes. This Table reports the correlation between county-level variables and observed British immigration (columns 1–2), the railway instrument (columns 3–4), and the leave-out instrument (columns 5–6). Columns (3–6) report two-stage least squares estimates. The unit of observation is a county at a decade frequency between 1870 and 1930. In each row, we report the correlation between one-decade lagged values of the dependent variable, shown on the rows, and the three independent variables, shown on the columns. Each regression includes county and decade-fixed effects. Standard errors are shown in parentheses and are clustered at the county level. Patents are expressed as shares. Both outcome and explanatory variables are taken in log terms and standardized for comparability. Referenced on page(s) C48, C48.

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

TABLE C.10. Double and Triple Differences: Alternative Dependent Variables

	Number of Patents				k-Originality			k-Similarity with US Patents		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Number	$\ln(1+\cdot)$	$\ln(\varepsilon+\cdot)$	Arcsinh	1 yr.	5 yrs.	10 yrs.	1 yr.	5 yrs.	10 yrs.
Panel A. Innovation by District										
Dependent Variable Mean	*1.687*	*21.355*	*0.206*	*2.055*	*0.759*	*0.736*	*0.728*	*2.549*	*12.156*	*23.010*
Post × US Innovation Shock	0.091**	20.777***	0.077	0.066	0.138***	0.082*	0.089*	0.364***	1.852***	3.569***
	(0.045)	(5.847)	(0.083)	(0.049)	(0.041)	(0.043)	(0.047)	(0.071)	(0.347)	(0.672)
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	31,671	31,671	31,671	31,671	31,671	31,671	31,671	31,671	31,671	31,671
Panel B. Innovation by District-Technology										
Dependent Variable Mean	*0.292*	*1.124*	*-3.456*	*0.371*	*0.297*	*0.304*	*0.313*	*0.629*	*3.012*	*5.708*
Post × US Innovation Shock	0.064***	0.994**	0.093**	0.073***	0.057**	0.052**	0.047**	0.172***	0.876***	1.712***
	(0.019)	(0.361)	(0.042)	(0.023)	(0.020)	(0.020)	(0.021)	(0.047)	(0.234)	(0.465)
District-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Technology-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Technology FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	601,749	601,749	601,749	601,749	601,749	601,749	601,749	601,749	601,749	601,749

Notes. This Table displays the effect of shocks to US innovation activity on innovation produced in the United Kingdom. In Panel A, the unit of observation is a district observed at a yearly frequency between 1870 and 1930; in Panel B, the unit of observation is a district-technology pair observed over the same period. The dependent variables are: in column (1), the number of patents; in column (2), the log(1+) number of patents; in column (3), the log(0.1+) number of patents; in column (4), the inverse hyperbolic sine of the number of patents; in columns (5–7), the number of patents in the top 20% of distribution of originality in the previous 1, 5, and 10 years; in columns (8–10), the dependent variable is the similarity of British patents with American patents issued in the previous 1, 5, and 10 years. The independent variable is equal to one for all years after the observation unit is exposed to a shock to US innovation activity. The definition of exposure is provided in the main text. All regressions in Panel A include district and year fixed effects; all regressions in Panel B include district-by-year, technology-by-year, and district-by-technology fixed effects. Standard errors are clustered by district and are displayed in parentheses. Referenced on page(s) 24, 24, A9, C49.

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

TABLE C.11. Double and Triple Differences: Alternative Definitions of the Shocks

	Number of Patents			High-Impact Patents			Similarity with US Patents		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. Innovation by District									
Dependent Variable Mean	*1.687*	*1.687*	*1.687*	*1.738*	*1.738*	*1.738*	*12.156*	*12.156*	*12.156*
Post × US Innovation Shock (Top 10%)	0.081*			0.060			1.587***		
	(0.043)			(0.043)			(0.341)		
Post × US Innovation Shock (Top 5%)		0.091**			0.074*			1.852***	
		(0.045)			(0.045)			(0.347)	
Post × US Innovation Shock (Top 1%)			0.111*			0.070			2.869***
			(0.063)			(0.063)			(0.307)
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	31,671	31,671	31,671	31,671	31,671	31,671	31,671	31,671	31,671
Panel B. Innovation by District-Technology									
Dependent Variable Mean	*0.292*	*0.292*	*0.292*	*0.304*	*0.304*	*0.304*	*0.302*	*0.302*	*0.302*
Post × US Innovation Shock (Top 10%)	0.035***			0.027**			0.040**		
	(0.011)			(0.011)			(0.014)		
Post × US Innovation Shock (Top 5%)		0.062***			0.050**			0.087***	
		(0.019)			(0.020)			(0.023)	
Post × US Innovation Shock (Top 1%)			0.093***			0.078**			0.144***
			(0.028)			(0.030)			(0.037)
District-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Technology-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Technology FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	612,408	612,408	612,408	612,408	612,408	612,408	612,408	612,408	612,408

Notes. This Table displays the effect of shocks to US innovation activity on innovation produced in the United Kingdom. In Panel A, the unit of observation is a district observed at a yearly frequency between 1870 and 1930; in Panel B, the unit of observation is a district-technology pair observed over the same period. The dependent variable is in columns (1–3), the (log) number of patents; in columns (4–6), the number of patents in the top 20% of the originality distribution; in columns (7–9), the average text-based similarity between British patents and American patents produced in the previous five years. The independent variable is equal to one for all years after the observation unit is exposed to a shock to US innovation activity. A unit is exposed to a shock if the number of emigrants from that unit that are exposed to a US innovation shock is in the k-percentile of the distribution. We consider three alternative values for the k threshold, 10%, 5% (the baseline), and 1%. All regressions in Panel A include district and year fixed effects; all regressions in Panel B include district-by-year, technology-by-year, and district-by-technology fixed effects. Standard errors are clustered by district and are displayed in parentheses. Referenced on page(s) 26, C49.

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

TABLE C.12. Shocks to United States Innovation Activity

	Double Diff.		Triple Diff.	
	(1) Number	(2) High-Impact	(3) Number	(4) High-Impact
Dependent Variable Mean	*0.954*	*0.165*	*0.135*	*0.016*
Post × Innovation Shock	0.585***	0.057***	0.129***	0.030***
	(0.018)	(0.009)	(0.010)	(0.003)
County FE	Yes	Yes	–	–
Year FE	Yes	Yes	–	–
County-Year FE	–	–	Yes	Yes
Technology-Year FE	–	–	Yes	Yes
County-Technology FE	–	–	Yes	Yes
Number of Counties	2,848	2,848	2,848	2,848
Observations	202,208	196,953	3,841,952	3,832,217

Notes. This Table reports the "first stage" of the shocks to US innovation, i.e., how much innovation in the US increases in the period following what we define as a shock. In columns (1–2) (resp. 3–4), the unit of observation is a county (resp. county-technology pair) observed at a yearly frequency between 1870 and 1930. In columns (1) and (3), the dependent variable is the (log) number of patents; in columns (2) and (4), it is the (log) number of patents in the top 20% of the novelty distribution. The independent variable is an indicator equal to one in the years after the observation units undergo an innovation shock. Regressions in columns (1–2) include county and year-fixed effects; regressions in columns (3–4) include county-by-year, technology-by-year, and county-by-technology fixed effects. Standard errors are shown in parentheses and are clustered at the county level. Referenced on page(s) 18, 18, 22, C50.

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

TABLE C.13. Within Neighborhood US Emigration and Innovation in the United Kingdom: Excluding Wales

	Number of Patents					Text-Based Measures	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
					I(Patents > 0)	US Patents Similarity	High Impact
Panel A. Neighborhood Emigration							
Dependent Variable Mean	*6.818*	*6.818*	*6.839*	*6.855*	*8.579*	*6.891*	*1.585*
Post × Emigrant in Neighborhood	0.428***	0.220**	0.238**	0.176	0.540***	0.473***	0.111**
	(0.133)	(0.112)	(0.118)	(0.157)	(0.164)	(0.135)	(0.054)
Post × N. Emigrants in Neighborhood				0.261			
				(0.214)			
Panel B. Neighborhood Non-Return Emigration							
Dependent Variable Mean	*6.818*	*6.818*	*6.839*	*6.855*	*8.579*	*6.891*	*1.585*
Post × Non-Return Emigrant in Neighborhood	0.394***	0.236**	0.248**	0.409**	0.502***	0.448***	0.101*
	(0.133)	(0.116)	(0.123)	(0.166)	(0.163)	(0.136)	(0.055)
Post × N. Non-Return Emigrants in Neighborhood				0.108			
				(0.222)			
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parish-Year FE	No	No	Yes	No	No	No	No
District-Year FE	No	Yes	–	No	No	No	No
Year FE	Yes	–	–	Yes	Yes	Yes	Yes
Number of Individuals	124,329	124,321	122,528	108,319	124,329	124,329	124,329
Observations	2,486,580	2,486,420	2,450,560	2,166,380	2,486,580	2,486,580	2,486,580

Notes. This Table reports how emigrants to the United States 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, excluding inventors residing in Wales. The dependent variable is in columns (1–4), the (log) number of patents produced by the members of the family; in column (5), an indicator equal to one for the same variable; in column (6), the average text-based similarity between British patents and American patents produced in the previous five years; in column (7), the number of patents in the top 20% of the novelty distribution. In Panel A, the baseline treatment is an indicator equal to one after the first neighbor of the inventor moves to the US, and zero otherwise; in Panel B, we restrict the treatment to non-return emigrants. In column (4), we further include an interaction with the number of emigrants. All regressions include inventor and year fixed effects; in columns (2) and (3) we include, respectively, district-by-year and parish-by-year fixed effects. Standard errors are clustered at the county level and are displayed in parentheses. Referenced on page(s) C51.

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

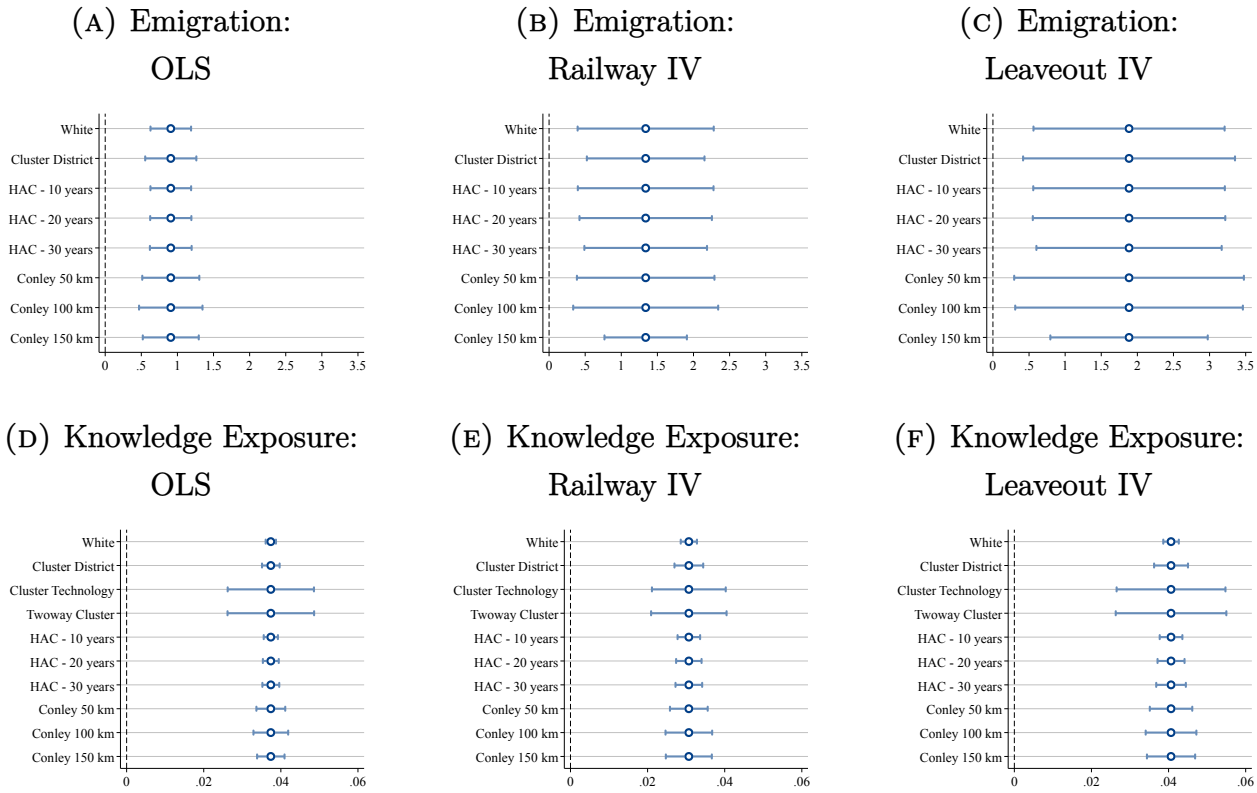
TABLE C.14. Emigration and Newspaper Coverage of US News: Instrumental Variable Estimates

	Mentions of U.S.			Mentions of U.S. States			Mentions of U.S. Counties		
	(1) Railway	(2) Leave-out	(3) Overid.	(4) Railway	(5) Leave-out	(6) Overid.	(7) Railway	(8) Leave-out	(9) Overid.
Dependent Variable Mean	*1.491*	*1.491*	*1.491*	*1.019*	*1.019*	*1.019*	*0.001*	*0.001*	*0.001*
US Emigrants (1,000s)	2.942***	3.377***	3.085***	72.081***	99.984***	77.308***	29.887***	23.229***	27.211***
	(0.989)	(1.213)	(1.031)	(9.739)	(10.193)	(9.710)	(4.095)	(3.254)	(3.725)
UK District FE	Yes	Yes	Yes	–	–	–	–	–	–
UK District-US State FE	–	–	–	Yes	Yes	Yes	–	–	–
UK District-US County FE	–	–	–	–	–	–	Yes	Yes	Yes
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Districts	591	591	591	591	591	591	591	591	591
Observations	4,137	4,137	4,137	215,124	215,124	215,124	13,465,344	13,465,344	13,465,344
Kleibergen-Paap's F	468.129	193.701	184.530	4969.583	2779.663	2182.919	9941.251	7824.292	4635.378
Std. Beta Coef.	0.156	0.179	0.164	0.154	0.213	0.165	0.064	0.049	0.058

Notes. This Table reports the association between emigration to the United States and newspaper coverage of US-related topics in the United Kingdom. The unit of observation is a district (columns 1–3), district-US state (columns 4–6), and district-US county (columns 7–9). Units are observed at a decade frequency between 1870 and 1930. The dependent variable is the (log) number of mentions of "United States" (1–3), the state (4–6), and the county (7–9). The independent variable is the number of US emigrants (columns 1–3), the number of emigrants between the district and the state (columns 4–6), and the number of emigrants between the district and the county (columns 7–9). In columns (1), (4), and (7), US out-migration is instrumented with the railway-based instrumental variable; in columns (2), (5), and (8), we report the estimates obtained using the leave-out instruments; columns (3), (6), and (9), report the two-stage least squares of the over-identified model that employs both instruments. Regressions include the dyadic fixed effects—i.e., district FEs in columns (1–3), district-by-state FEs in columns(4–6), and district-by-county FEs in columns (7–9)—as well as time fixed effects. Standard errors in parentheses are clustered at the district level. Referenced on page(s) 33.
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

Figures

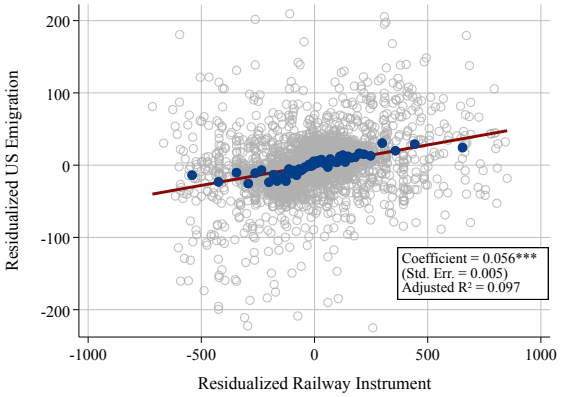
FIGURE C.1. Emigration, Exposure to US Knowledge, and Innovation: Alternative Standard Errors



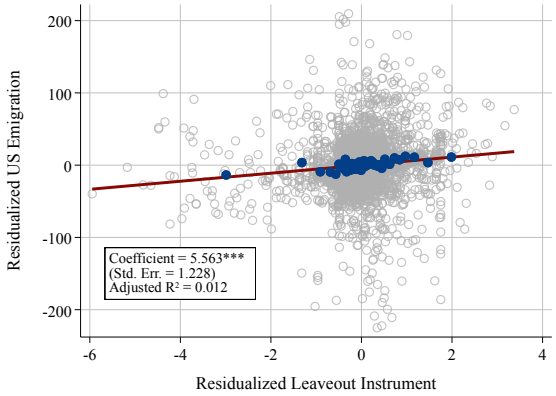
Notes. This Figure displays the association between emigration and innovation in the United Kingdom using various estimators for the standard errors. In Panels C.1a–C.1c (resp. C.1d–C.1f), the unit of observation is a district (resp. district-technology pair) observed at a decade frequency between 1870 and 1930. Each dot reports the coefficient of a regression between the (log) number of patents and an independent variable which, in Panels C.1a–C.1c (resp. C.1d–C.1f) is the number of migrants (resp. exposure to US technology). In Panels C.1b–C.1c and C.1e–C.1f, we report two-stage least squares estimates using the railway and leave-out instruments described in the main text. In Panels C.1a–C.1c (resp. C.1d–C.1f), regressions include district and decade-fixed effects (resp. district-by-technology and year-fixed effects). We consider various estimators for the standard errors: robust to heteroskedasticity; clustered by district, technology, and two-way by district and technology; robust to heteroskedasticity and autocorrelation at various bandwidths; and robust to spatial autocorrelation following (Conley, 1999) at various bandwidths. All confidence bands report 95% confidence intervals. Referenced on page(s) 25, C45.

FIGURE C.2. First-Stage Correlations

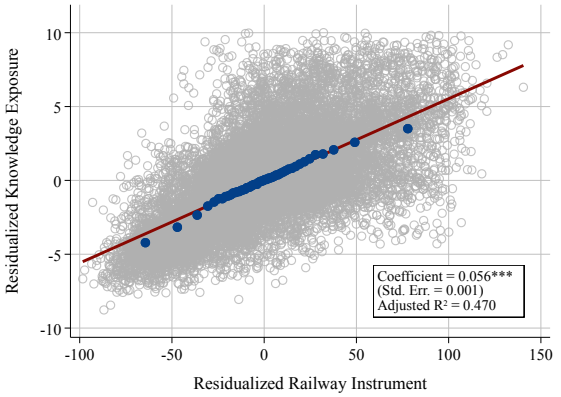
(A) Emigration: Railway IV



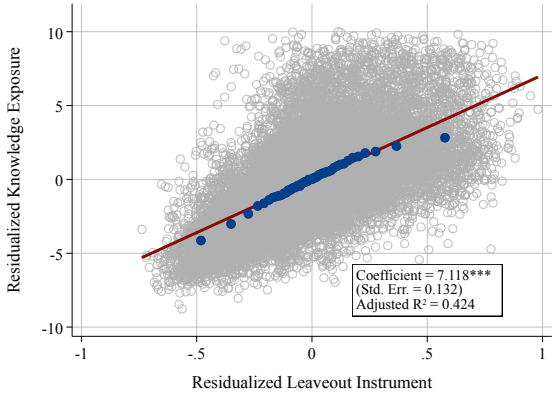
(B) Emigration: Leave-out IV



(C) Knowledge Exposure: Railway IV

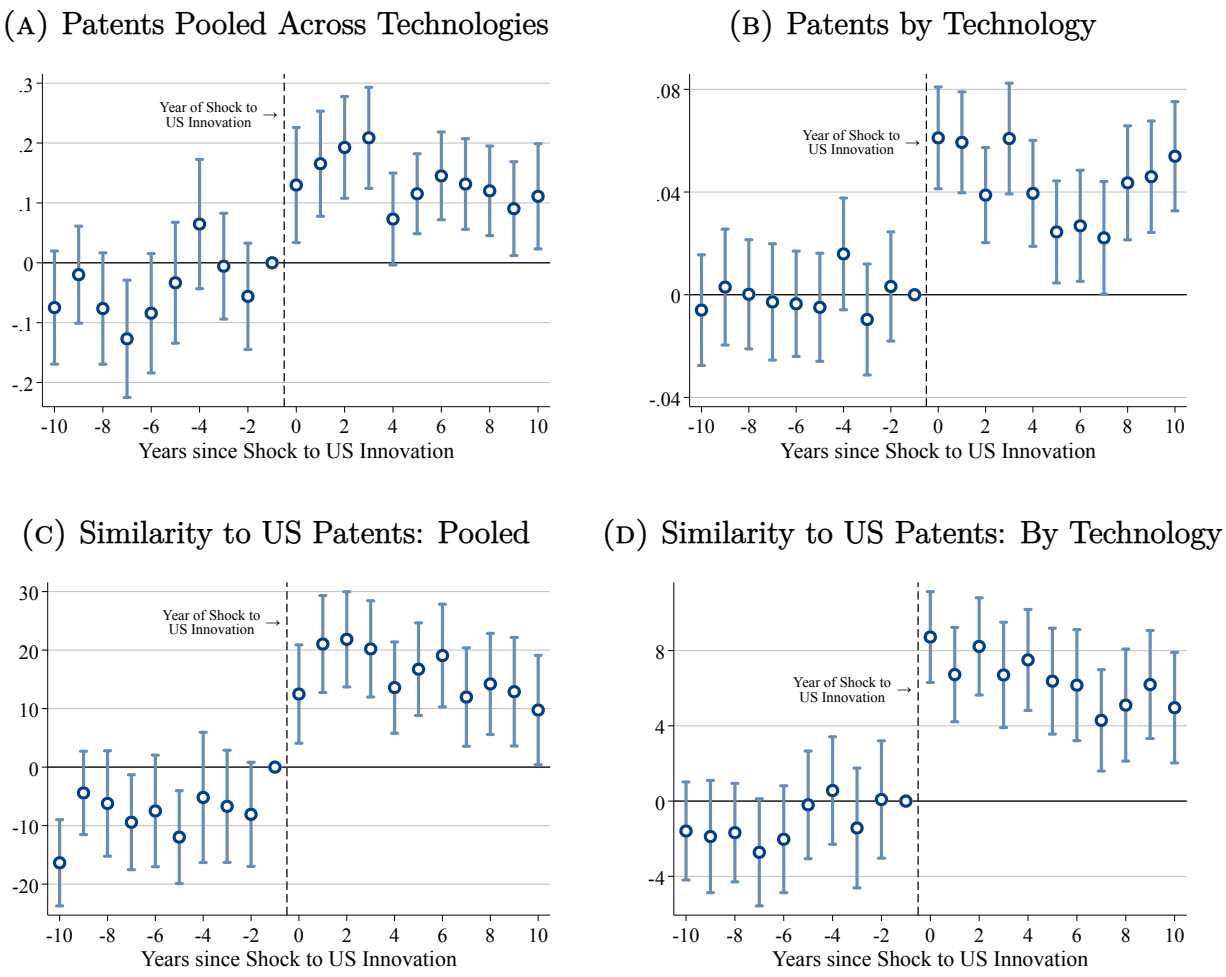


(D) Knowledge Exposure: Leave-out IV



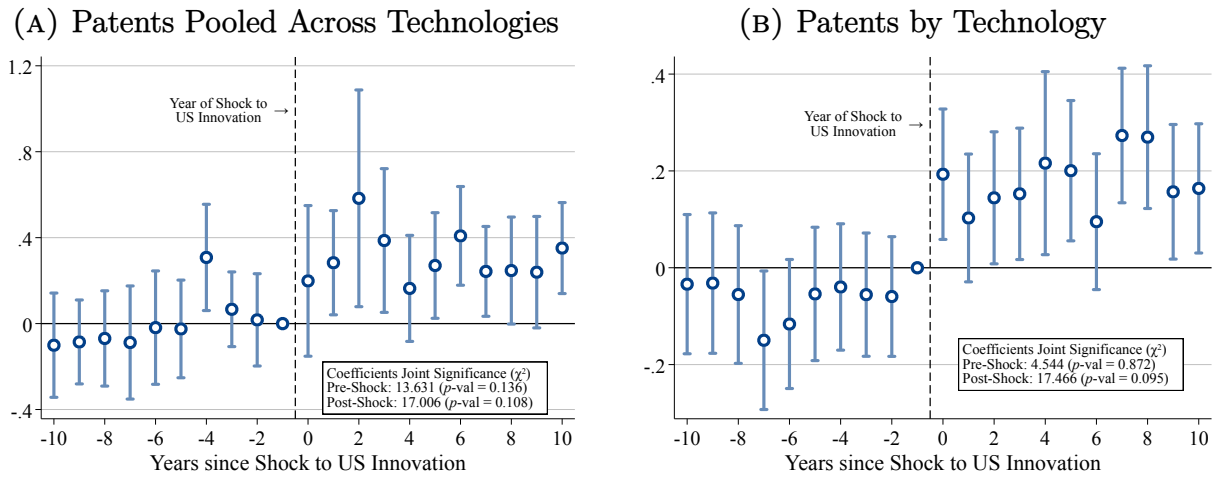
Notes. This Figure reports the first-stage association between out-migration (Panels C.2a–C.2b), exposure to US technology (Panels C.2c–C.2d), and the railway and leave-out instrumental variables described in the main text. In Panels C.2a–C.2b (resp. C.2c–C.2d), the unit of observation is a district (district-technology class pair) at a decade frequency between 1870 and 1920. Panels C.2a–C.2b (resp. C.2c–C.2d) plot the residuals of the observed variable and the instruments against district and year (resp. district-by-technology and year) fixed effects. The blue dots report the binned means. The red line overlays a linear fit between the two variables. Each graph reports the slope of the line along with the associated clustered standard error and the R^2 of each residualized regression. Referenced on page(s) C48.

FIGURE C.3. The Dynamic Effect of Shocks to US Innovation on Innovation in the UK: Alternative Estimator



Notes. This Figure displays how shocks to US innovation activity impact innovation in the UK. In Panels C.3a and C.3c, the unit of observation is a district observed at a yearly frequency between 1870 and 1930; in Panels C.3b and C.3d, the unit of observation is a district-technology pair, observed over the same time. The dependent variable is, in Panels C.3a–C.3b, the (log) number of patents, and in Panels C.3c–C.3d, the average text-based similarity of the UK patents issued in a given period to US patents issued in the previous five years. Each dot reports the coefficient of an indicator variable which codes the time since the observation unit is exposed to a shock to US innovation activity through emigration ties. We employ the estimator developed by Sun and Abraham (2021) to account for the staggered roll-out of the shocks across observation units. The black dashed line indicates the treatment period. The last period before the shock serves as the baseline category. In Panel C.3a, the regression includes district and year-fixed effects, and standard errors are clustered at the district level; in Panel C.3b, the regression includes district-by-year, technology-by-year, and district-by-technology fixed effects, and standard errors are clustered at the district level. The bands report 95% confidence intervals. Referenced on page(s) 25, C50.

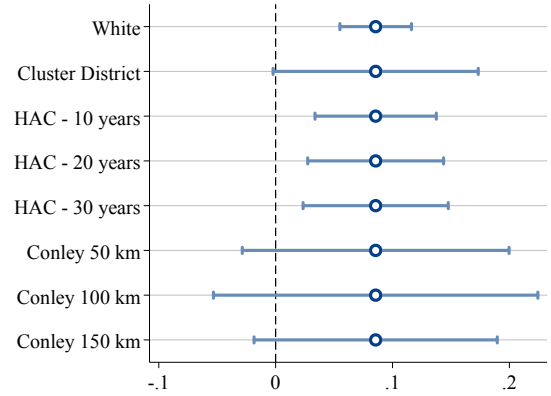
FIGURE C.4. The Dynamic Effect of Shocks to US Innovation on Innovation in the UK: Poisson Quasi-Maximum Likelihood Regressions



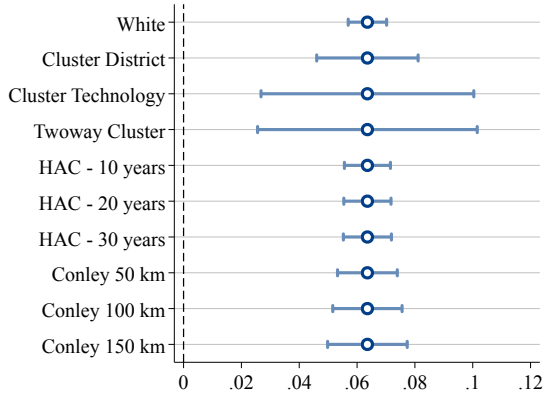
Notes. This Figure displays how shocks to US innovation activity impact innovation in the UK. In Panel C.4a, the unit of observation is a district observed at a yearly frequency between 1870 and 1930; in Panel C.4b, the unit of observation is a district-technology pair observed over the same time. The dependent variable is the number of patents normalized by the number of patents issued in each observation unit before the treatment period. Each regression is estimated using the Poisson quasi-maximum likelihood estimator described in Correia *et al.* (2020). Each dot reports the coefficient of an indicator variable, which codes the time since the observation unit is exposed to a shock in US innovation activity through emigration ties. The black dashed line indicates the treatment period. The last period before the shock serves as the baseline category. In Panel C.4a, the regression includes district and year-fixed effects; in Panel C.4a, the regression includes district-by-year, technology-by-year, and district-by-technology fixed effects. Standard errors are clustered at the district level. The bands report 90% confidence intervals. Each Figure reports separate χ^2-statistics for the joint significance of the pre-and post-treatment coefficients and associated p-values. Referenced on page(s) 25.

FIGURE C.5. The Effect of Shocks to US Innovation on UK Innovation: Alternative Standard Errors

(A) Patents Pooled Across Technologies

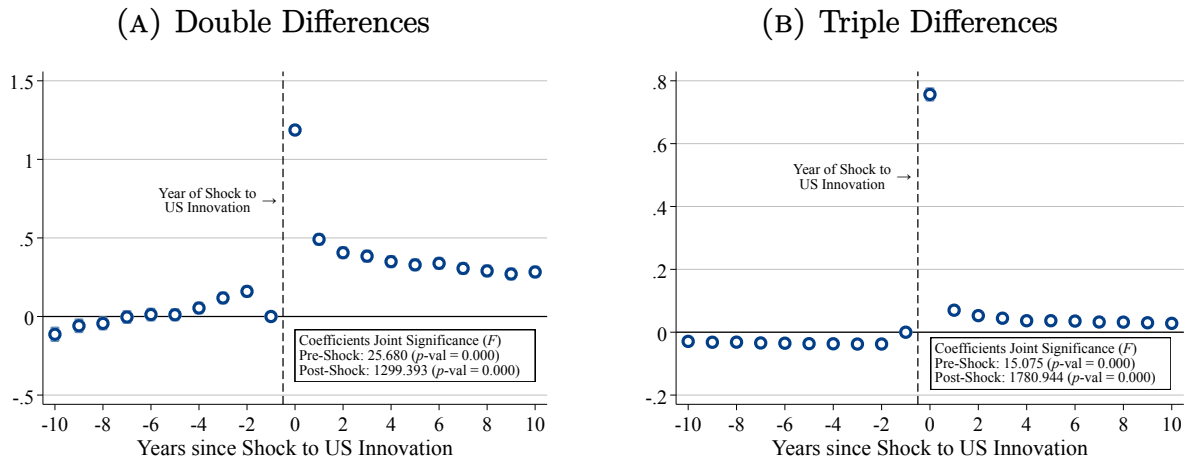


(B) Patents by Technology



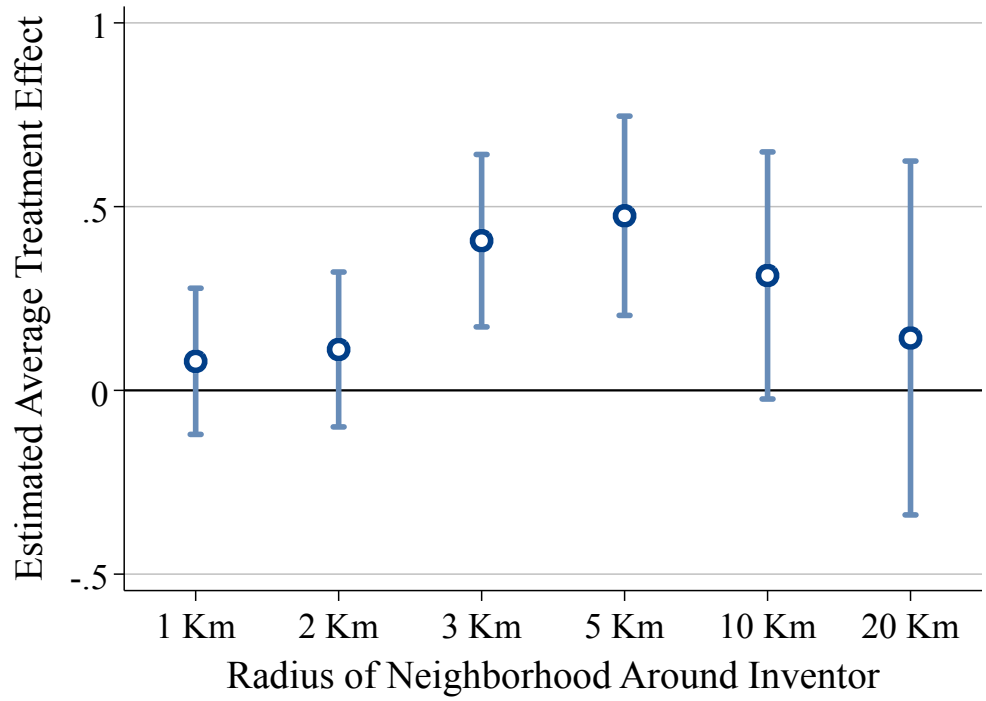
Notes. This Figure displays the effect of shocks to US innovation activity on innovation produced in the United Kingdom using various estimators for the standard errors. In Panel C.5a, the unit of observation is a district observed at a yearly frequency between 1870 and 1930; in Panel C.5b, the unit of observation is a district-technology pair observed over the same period. The dependent variable is the (log) number of patents. Each dot reports the coefficient of the treatment variable, which is equal to one for all years after the observation unit is exposed to a shock to US innovation activity. The definition of exposure is provided in the main text. All regressions in Panel C.5a include district and year fixed effects; all regressions in Panel C.5b include district-by-year, technology-by-year, and district-by-technology fixed effects. We consider various estimators for the standard errors: robust to heteroskedasticity; clustered by district, technology, and two-way by district and technology; robust to heteroskedasticity and autocorrelation at various bandwidths; and robust to spatial autocorrelation following (Conley, 1999) at various bandwidths. All confidence bands report 95% confidence intervals. Referenced on page(s) 25, C50.

FIGURE C.6. Dynamic Shocks to US Innovation Activity



Notes. This Figure reports the "first stage" of the shocks to US innovation, i.e., how much innovation in the US increases in the period following what we define as a shock. In Panel C.6a (resp. C.6b), the unit of observation is a county (resp. county-technology pair) observed at a yearly frequency between 1870 and 1930. The dependent variable is the (log) number of patents. Each dot reports the coefficient of an indicator variable that codes the number of periods since the observation units undergo an innovation shock. Regressions in Panel C.6a (resp. C.6b) include county and year-fixed effects (resp. county-by-year, technology-by-year, and county-by-technology fixed effects). Standard errors are clustered at the county level. Bands report 95% confidence intervals. Each graph reports separately the F-statistics of joint significance of the pre-and post-treatment coefficients, along with their p-values. Referenced on page(s) 22, C51.

FIGURE C.7. Within-Neighborhood US Emigration: Alternative Neighborhood Threshold Values



Notes. This Figure reports how emigrants to the United States 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 (log) number of patents produced by the members of the family. 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 1, 2, 3, 5, 10, and 20 kilometers. The baseline case reported in the main text assumes $k = 5$. All regressions include inventor and year-fixed effects. Standard errors are clustered at the county level. Bands report 95% confidence intervals. Referenced on page(s) 29, C51.

Appendix References

Abramitzky, R., L. Boustan, K. Eriksson, J. Feigenbaum and S. Pérez (2021). "Automated Linking of Historical Data." *Journal of Economic Literature*, 59(3): 865–918.

Abramitzky, R., L. P. Boustan and K. Eriksson (2014). "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration." *Journal of Political Economy*, 122(3): 467–506.

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

Bandiera, O., I. Rasul and M. Viarengo (2013). "The Making of Modern America: Migratory Flows in the Age of Mass Migration." *Journal of Development Economics*, 102: 23–47.

Beach, B. and W. W. Hanlon (2023). "Historical Newspaper Data: A Researcher's Guide." *Explorations in Economic History*, 90: 101541.

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

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

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

Borusyak, K., P. Hull and X. Jaravel (2022). "Quasi-Experimental Shift-Share Research Designs." *The Review of Economic Studies*, 89(1): 181–213.

Card, D. (2001). "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration." *Journal of Labor Economics*, 19(1): 22–64.

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

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

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

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

Davis, J. H. (2004). "An Annual Index of US Industrial Production, 1790–1915." *The Quarterly Journal of Economics*, 119(4): 1177–1215.

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

Finishing Publications, L. (2018). "Early British Patents. A Cradle of Inventions. British Patents

1617 to 1895." MFIS [data collection].

GOODMAN-BACON, A. (2021). "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics*, 225(2): 254–277.

HANLON, W. W. (2016). "British Patent Technology Classification Database: 1855-1882." Unpublished [data collection].

KELLY, B., D. PAPANIKOLAOU, A. SERU and M. TADDY (2021). "Measuring Technological Innovation Over the Long Run." *American Economic Review: Insights*, 3(3): 303–320.

LAN, T. and P. LONGLEY (2019). "Geo-Referencing and Mapping 1901 Census Addresses for England and Wales." *ISPRS International Journal of Geo-Information*, 8(8): 320.

MADDISON, A. (2007). *Contours of the World Economy 1-2030 AD: Essays in Macro-Economic History*. Oxford (UK): Oxford University Press.

NICHOLAS, T. (2014). "Technology, Innovation, and Economic Growth in Britain since 1870." In R. Floud, J. Humphries and P. Johnson, eds., "The Cambridge Economic History of Modern Britain," pp. 181–204. Cambridge (UK): Cambridge University Press.

NUVOLARI, A. and V. TARTARI (2011). "Bennet Woodcroft and the Value of English Patents, 1617–1841." *Explorations in Economic History*, 48(1): 97–115.

NUVOLARI, A., V. TARTARI and M. TRANCHERO (2021). "Patterns of Innovation During the Industrial Revolution: A Reappraisal Using a Composite Indicator of Patent Quality." *Explorations in Economic History*, 82: 101419.

ROSENBERG, N. and M. TRAJTENBERG (2004). "A General-Purpose Technology at Work: The Corliss Steam Engine in the Late-Nineteenth-Century United States." *The Journal of Economic History*, 64(1): 61–99.

RUGGLES, S., C. FITCH, R. GOEKEN, J. HACKER, M. NELSON, E. ROBERTS, M. SCHOUWEILER and M. SOBEK (2021). "IPUMS ancestry full count data: Version 3.0 [dataset]." *Minneapolis, MN: IPUMS*.

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

SEQUEIRA, S., N. NUNN and N. QIAN (2020). "Immigrants and the Making of America." *The Review of Economic Studies*, 87(1): 382–419.

SPITZER, Y. and A. ZIMRAN (2018). "Migrant Self-Selection: Anthropometric Evidence from the Mass Migration of Italians to the United States, 1907–1925." *Journal of Development Economics*, 134: 226–247.

SUN, L. and S. ABRAHAM (2021). "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics*, 225(2): 175–199.

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

WAGNER, D. (2008). *Science and Civilisation in China.* Cambridge University Press (UK).

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