Return Innovation

The Knowledge Spillovers of the British Migration to the US, 1870–1940*

Davide M. Coluccia† (Job Market Paper)

Gaia Dossi‡

This Version: December, 2023 – Click here to download the latest version

Abstract

How does innovation diffuse across countries? In this paper, we document 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 over 1853–1899. Using a new shift-share instrument for bilateral migration and a triple-differences design, we document that migration ties contribute to technology diffusion from the US to the UK. Through high-dimensional text analysis, we find that emigrants promote technology transfer, but they also nurture the production of original innovation. Physical return migration is an important driver of this "return innovation" effect. However, we find that the interactions between emigrants and their origin communities promote technology diffusion, even absent return migration. Additionally, we show that migration ties propel knowledge flows by fostering cross-border market integration.

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

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

---

*We are extremely grateful to Mara Squicciarini and Joel Mokyr for their continued guidance and support. For insightful comments and discussions, we also thank Jan Bakker, Enrico Berkes, Vivek Bhattacharya, Maristella Botticini, José Espin-Sánchez, Joe Ferrie, Carola Frydman, Simon Görlach, Walker Hanlon, Leander Heldring, Matteo Leombroni, Nicola Limodio, Paolo Masella, Filippo Mezzanotti, Alessandro Nuvolari, Laura Ogliari, Sebastian Ottinger, Eleonora Patacchini, Marta Prato, Nancy Qian, Lukas Rosenberger, Joshua Rosenbloom, Paola Sapienza, Carlo Schwartz, Marco Tabellini, Edoardo Teso, and John Van Reenen. We thank audiences at the EHA (Pittsburgh), EHS (Cambridge), and ES (Berlin) conferences, at the 10th CEPR Economic History Symposium, and seminar participants at Bocconi, Milan, Northwestern, Cologne, and Uppsala. Antonin Bergeaud, Enrico Berkes, and Walker Hanlon kindly shared data with us. We are grateful to Bocconi, Fondazione Invernizzi, LSE, MapTiler AG, and POID for financial support. All errors are our own.

†Department of Economics, Northwestern University. Email: davide.coluccia@northwestern.edu. Website: dcoluccia.github.io.

‡Department of Economics, LSE. Email: g.g.dossi@lse.ac.uk. Website: sites.google.com/view/gaiadossi.

1 Introduction

Technological progress and, thus, economic growth hinge on the diffusion of knowledge across countries (Griffith *et al.*, 2006; Comin and Hobijn, 2011). Eaton and Kortum (1999), for instance, estimate that in the 1980s, approximately 70% of productivity growth in advanced European countries relied on technology developed in the United States and Japan.[1] Recent models emphasize that exposure to foreign technology is crucial for the cross-country diffusion of innovation (Alvarez *et al.*, 2013; Buera and Oberfield, 2020). Empirically, however, estimating the impact of exposure to foreign technology on domestic innovation is challenging because it requires observing joint variation in the intensity and composition of exposure across observation units—e.g., firms or regions—and technologies.

In this paper, we overcome this challenge by studying out-migration as a novel source of knowledge diffusion from the country of destination to the country of origin of migrants. Drawing on the British mass migration to the United States (1870–1940), we observe that, within Britain, migration ties impact exposure to US technology along two margins. First, exposure in Britain is more intense in regions with higher US emigration rates. Second, emigrants are exposed to different technologies depending on where they settle across the US. By combining these two components, out-migration offers an ideal test to estimate the impact of exposure to foreign knowledge on innovation.

Leveraging this insight, we present novel, causal evidence that exposure to foreign technology through out-migration linkages contributes to the diffusion of innovation to emigration countries.[2] To the best of our knowledge, this is the first paper to document this phenomenon, which we label the "return innovation" effect. Compared to the influential "brain drain" hypothesis (Docquier and Rapoport, 2012), this paper thus introduces a new and competing perspective on the effects of out-migration on the economic development of emigration countries. We find that physical return migration is an important driver of the return innovation effect. However, we present evidence that the interactions between the emigrants and their origin communities—families and former neighbors—further promote technology diffusion in the absence of physical return migration. Moreover, migration ties foster cross-country market integration, which facilitates innovation diffusion into the emigration country.

The impact of emigration on innovation is ambiguous. Traditional "brain drain" arguments suggest

[1] Economic historians have long argued that the diffusion of knowledge is a key driver of productivity growth and catching up (e.g., see Gerschenkron, 1962; Rosenberg, 1982). However, endogenous growth models featuring cross-country diffusion dynamics have emerged only recently (Benhabib *et al.*, 2021; Perla *et al.*, 2021; Van Patten, 2023).

[2] A vast scholarship documents that immigrants actively contribute to several dimensions of economic development in their destination countries spanning entrepreneurship (Kerr and Kerr, 2020; Azoulay *et al.*, 2022), innovation (Ganguli, 2015; Bahar *et al.*, 2019; Burchardi *et al.*, 2020; Bernstein *et al.*, 2022) and science (Moser *et al.*, 2014, 2020), local specialization (Ottinger, 2020), and the formation of political preferences (Giuliano and Tabellini, 2020). As pointed out by Clemens (2011), emigration has generally generated far less attention than immigration.

that emigration countries suffer from a loss of human capital (for a review, see Gibson and McKenzie, 2011). Growth theory, in turn, predicts that this depletion would negatively hamper their ability to innovate (e.g., Jones, 1995). On the other hand, recent scholarship suggests that exposure to innovation is crucial for innovation activity (Akcigit *et al.*, 2018; Bell *et al.*, 2019). Therefore, 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). Since this "return innovation" effect and the "brain drain" channel operate in opposing directions, empirical evidence is necessary to assess the impact of out-migration on innovation in emigration countries.

We examine this question in the context of the English and Welsh migration to the United States between 1850 and 1940, when approximately 30 million European migrants settled across the Atlantic. Nearly four million came from Britain.[3] Since Rosenberg (1982), economists have interpreted the spread of the Industrial Revolution in terms of waves of technological diffusion originating in Britain. Hence, existing studies document that 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 approached the technology frontier in many industries, from interchangeable parts and machine tools to engines and agricultural machinery (David, 1966; Rosenberg, 1970; Rosenberg and Trajtenberg, 2004).[4] It is therefore plausible, although unexplored, that migration ties promoted the diffusion of these technologies back to Britain, which, throughout this period, increasingly lagged behind the newly industrialized countries.

Besides its historical importance, this setting allows us to overcome three limitations of contemporary scenarios that hindered previous attempts to study this question. First, our novel individual-level dataset allows us to look at the entire population of transatlantic migrants. Second, we measure international knowledge flows using historical patent data. This approach would be infeasible with contemporary data due to international intellectual property protection in force since 1945.[5] Finally, the near-complete absence of migration regulations targeting British migrants ensures that possibly endogenous policy interventions do not confound the analysis.

[3]This figure does not include the Irish. In the paper, we focus on the English and Welsh migration. We use, for the sake of brevity, the terms "British" and "English" as shortcuts to collectively refer to England and Wales, thus excluding Scotland.

[4]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. Starting in 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.

[5]At the time, the same invention could be patented by different people in the US and in the UK with no legal penalty. Additionally, the British patent office required that at least one applicant—usually a patent agent for foreign inventors—be a permanent resident in the UK.

To estimate the effect of out-migration on innovation, we observe that districts in the UK would be exposed to different technologies depending on the US county where emigrants from those areas would settle.[6] Our research design thus leverages the joint variation in county-level specialization across technology classes and district-county bilateral migration flows. Consider two English districts, say Staffordshire and Camden, and two US counties, say San Diego and Cook. Assume hypothetically that Staffordshire and Camden are observationally identical, but all emigrants from Staffordshire settle in San Diego County, whereas all those originating in Camden move to Cook County. Suppose that San Diego County specializes in shipbuilding, whereas Cook County specializes in chemistry. Then, Staffordshire will be exposed to shipbuilding technology, whereas Camden will be exposed to innovations in chemistry.

We assemble two novel, detailed, 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 within the UK, we leverage confidential individual-level UK and US census data to link individual records of British immigrants in the US to the UK census (Schurer and Higgs, 2020; Ruggles *et al.*, 2021). The resulting novel dataset allows us to track individual out-migration and return migration between the US and the UK. Second, to reconstruct the geography of innovation in the UK in the second half of the nineteenth century, we digitize the universe of 300,000 original patent documents issued in England and Wales between 1853 and 1899. We thus assemble the first comprehensive dataset covering patented innovation during the Second Industrial Revolution in the United Kingdom.

The granularity of our data allows us to deal with the potential endogeneity of exposure to US knowledge. The primary reason that would caution against a causal interpretation of the estimates is assortative matching, namely the possibility that British migrants sort across US counties depending on where they came from. Suppose, drawing on the previous example, that Staffordshire specializes in shipbuilding. Then, Staffordshire would be exposed to shipbuilding technology because emigrants from that district settle in San Diego County, which also specializes in shipbuilding, but the coefficient of a naïve regression between knowledge exposure and innovation would conflate pre-existing specialization patterns into the treatment effect of exposure to US technology.[7]

We develop two approaches to deal with this potential source of endogeneity. First, we build a shift-share instrumental variable that exploits conditional variation in the connection timing to the US railway network to construct county-level immigration shocks (Sequeira *et al.*, 2020). These shocks allow us to randomize British immigration across counties and avoid the assortative matching issue (Borusyak

[6] In most of the analysis, the 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 terms of population (approximately 40,000). Unlike counties, however, registration districts were statistical entities that did not enjoy political or budgetary autonomy.

[7] This example serves illustrative purposes, but our baseline research design non-parametrically rules out the possibility that differences in initial specialization across UK districts drive our results.

et al., 2022). Second, we note that the return innovation effect would imply that shocks to innovation activity in the United States—defined as unusually large deviations from the average yearly number of patents by technology class—would diffuse to UK districts whose emigrants had settled in the areas where these shocks manifest. To test this hypothesis, we implement a triple differences analysis that compares districts and technology classes by exposure to innovation shocks in the US.

Our main result is that innovation in the UK shifts in response to exposure to US innovation through migration linkages. The instrumental variable design confirms the existence of a causal link between exposure to US knowledge and innovation in the UK. The triple differences analysis provides evidence that innovation shocks in the US diffuse into the United Kingdom through migration ties. We estimate that exposure to an innovation shock in the United States—which, on average, is associated with twenty more patents in a given county-technology class pair—results in two more patents produced in the UK. This implies a 10% pass-through rate of US innovation shocks to Britain through migration ties. This figure is sizable, as it accounts for approximately one-third of the average annual number of patents by district technology class in Britain. This result, which we label the "return innovation" effect, is larger in industries in which the UK was relatively more specialized than the US. Exposure to foreign knowledge through migration ties thus appears to nurture existing industries rather than creating new ones.

We then ask whether the knowledge flows generated by migration linkages stimulate technology transfer between the US and Britain or if they propel original innovation in the UK. To do so, we adopt a text-based approach that quantifies (i) the similarity between UK patents and previous US patents and (ii) the "originality" of the former with respect to the latter. We find that areas more exposed to US knowledge produce patents that are more similar to those granted in the United States. We also estimate that those areas produce more innovative patents compared to the existing stock of US knowledge. These results are not contradictory. In the immediate aftermath of a US innovation shock, the similarity of newly produced UK patents with previous US patents increases. However, in later periods, original patents take over the bulk of the increased innovation activity. Taken together, these results indicate that return innovation conflates two margins: a technology transfer catching-up effect *à la* Gerschenkron (1962) and a positive spillover channel that stimulates the production of novel knowledge.

In the second part of the paper, we exploit the richness of our data to explore the mechanisms that underlie the return innovation effect. On the one hand, return innovation may require the physical return of migrants. On the other, however, migration ties may promote the diffusion of technology irrespective of physical return. We find that physical return is an important but not the exclusive driver of return innovation. Interactions between emigrants and their communities of origin further promote technology diffusion even absent physical return. Moreover, we provide indirect evidence that migration ties foster cross-border market integration, further facilitating innovation diffusion into the UK.

Since our data allow us to observe return migration at a high level of spatial granularity, we can measure exposure to US technology through return migrant flows and compare it with the effect of outward emigration flows. We find that return migration accounts for approximately half of the overall return innovation effect. Importantly, however, the effect of exposure to US technology through out-migration ties remains sizable and significant even if we control for return inflows. This result suggests that return migration is an important determinant of return innovation, but it also indicates that migration ties contribute to the diffusion of knowledge even absent physical return.

We first explore the role of interactions between emigrants and their communities of origin as a driver of technology diffusion. We focus on two factors that could promote such interactions: family ties and geographical proximity. Our results connect to 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 *et al.*, 2021). Then, we study the channels through which migration ties impact innovation in emigration countries without directly relying on personal relationships between the emigrants and their origin communities.

We find that the family members of US emigrants display increased patenting activity after their relative moves 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 substantial. 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 considerably larger than that of those who never return. This further confirms that return migration is a major driver of return innovation. At the same time, however, emigrants promote innovation in their families even if they never return. Overall, since return emigrants account for approximately one-third of the entire migrant stock, the magnitude of these effects is, in aggregate, similar.

The geographical proximity between emigrants and their former neighbors can be interpreted as an alternative proxy for local social networks. To estimate its impact on the innovation activity of stayers, we leverage the individual-level nature of our migration and patenting data. Using a linked patent-census sample and geo-coded information on the universe of the UK population, we find that patenting activity increases for non-migrants after their neighbor(s) migrate(s) to the United States.[8] Moreover, the estimated effect remains positive and significant when restricting the treatment to include only US migrants who never return. These results strongly suggest that cross-country interactions between emigrants and their origin communities are a key driver of return innovation, even absent physical return.

Building on previous literature, we provide indirect evidence that migration ties contribute to cross-border market integration, thus promoting knowledge flows. We leverage the introduction of the first

[8] In the baseline exercise, two individuals are considered as neighbors if they live in the same street. However, in robustness regressions, we define neighborhoods as areas of a 100-meter radius centered around each individual in the sample.

transatlantic telegraph cable connecting the US and the UK in 1866 as a sudden and sizable increase in the integration of the British and the American markets.[9] In a difference-in-differences setting, we show that districts with higher US emigration rates before the introduction of the transatlantic telegraph cable display higher patenting activity after 1866. Moreover, innovation does not increase evenly across technology classes. The gains in patenting activity manifest in those same technologies that districts had been more exposed to through migration ties. This suggests that the increased economic integration generated by the telegraph accrued relatively more to districts that had pre-existing migration ties with the US market.

To provide additional evidence that migration ties foster market integration, we study trade disruptions arising from the Smoot-Hawley Act (1930), which severely increased US import duties. Trade is commonly interpreted as a measure of cross-border market integration. Importantly, the tariff increase was not homogeneous across goods categories. Leveraging this cross-industry variation, we find that patenting in the UK decreases in districts more exposed—through migration ties—to technologies that the Act targeted more heavily. This result suggests that migration ties promote market integration, which facilitates cross-border knowledge diffusion.

Finally, we investigate whether the information flows generated by migration ties are restricted to innovation or if they encompass a broader set of subjects. 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. Newspaper coverage of a given state (resp. county) is broader in districts with more emigrants to that given state (resp. county). This exercise suggests that the scope of information flows generated by migration ties is not limited to innovation and encompasses a broader set of topics.

This paper provides new evidence on how knowledge diffuses across countries. More specifically, we find that exposure to foreign technology through migration ties contributes to the diffusion of innovation from the country of destination to the country of origin of migrants. Our results imply that out-migration can promote innovation and, thus, long-term growth by fostering the diffusion of knowledge into emigration countries. Despite cautions on external validity, ever-increasing international human mobility and advancements in communication technology suggest that our results bear relevant policy implications for economic growth, especially in developed countries.

Related Literature. This paper is related to four streams of literature. First, we contribute to the literature that studies the determinants of the direction of innovation and the allocation of research ac-

[9] The telegraph represented a fundamental development in information and communication technology. Steinwender (2018) documents that the transatlantic cable allowed information to flow more rapidly and efficiently across the Atlantic Ocean, thus enabling trade and reducing international arbitrage opportunities.

tivity across technological sectors. Pioneering work on directed technical change by Hicks (1932) and Habakkuk (1962) was formalized by Acemoglu (2002, 2010). More recently, this question has been studied both theoretically (Bryan and Lemus, 2017; Hopenhayn and Squintani, 2021; Acemoglu, 2023) as well as empirically (Hanlon, 2015; Aghion *et al.*, 2016; Moscona, 2021; Moscona and Sastry, 2022; Einiö *et al.*, 2022; Gross and Sampat, 2022). We inform this literature by introducing one novel determinant of the direction of innovation, namely, international human mobility, through the return innovation effect.[10]

Second, 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 *et al.*, 2015), attitudes towards democracy and voting (Spilimbergo, 2009; Batista and Vicente, 2011; Ottinger and Rosenberger, 2023) and political change (Chauvet and Mercier, 2014; Kapur, 2014; Karadja and Prawitz, 2019), technology adoption (Coluccia and Spadavecchia, 2022), entrepreneurship (Anelli *et al.*, 2023), and social norms (Beine *et al.*, 2013; Bertoli and Marchetta, 2015; Tuccio and Wahba, 2018). This paper provides new evidence that emigration shapes the dynamics and the direction of innovation because it exposes sending countries to novel knowledge produced abroad. This enriches the traditional narrative that reduces out-migration to a mere depletion of the human capital stock.

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 (Arkolakis *et al.*, 2020; Moser *et al.*, 2020) as well as the long-run (Akcigit *et al.*, 2017; Burchardi *et al.*, 2020; Sequeira *et al.*, 2020) implications of immigration on US innovation. Ottinger (2020) shows that European immigration influenced US industry specialization. This paper is closest to Andersson *et al.* (2022). They show that mass out-migration in Sweden triggered labor-saving innovation by increasing the relative cost of labor. Instead, we look at the diffusion of technology from the areas where migrants settle to those they originate from. We are thus able to dissect the impact of out-migration on technology diffusion from the US to Britain. We show that migration ties facilitate the cross-border diffusion of technologies and find that information flows, rather than physical return migration, is the main underlying channel of this "return innovation" effect.

Finally, we relate to the literature studying the dynamics and determinants of knowledge flows and technology diffusion across countries (among others, see Jaffe *et al.*, 1993; Griffith *et al.*, 2006; Bahar *et al.*, 2014; Pauly and Stipanicic, 2021). Specifically, we contribute to the papers documenting how human mobility fosters the diffusion of novel knowledge (Kerr, 2008; Hornung, 2014; Bahar *et al.*, 2019, 2022a; Prato, 2021). We contribute to this literature from several perspectives. First, we enlarge the observation sample to include the universe of emigrants instead of a selected subgroup of highly skilled individuals.

[10] A related literature highlights that the direction of innovation bears relevant consequences in terms of subsequent technical change because it can lead to technology lock-ins (Dosi, 1982; Arthur, 1989; Acemoglu and Lensman, 2023).

Second, we leverage recent insights by Akcigit *et al.* (2018) and Bell *et al.* (2019) and show that exposure to foreign technology is a major driver of technology transfers. 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.

Outline. The rest of the paper is structured as follows. In section 2, we describe this study's historical and institutional context. Section 3 introduces the novel datasets we construct. We present the empirical research design in section 4 and discuss the main findings in section 5. Section 6 uncovers the possible mechanisms underlying the results and discusses possible alternative interpretations. Section 7 concludes.

2 Historical and Institutional Background

This section offers a concise overview of the historical and institutional features of our study setting. Throughout it, we highlight key aspects and details that were relevant to the empirical investigation. We conclude by presenting three examples of technology transmission to the UK operated by British immigrants in the US.

2.1 The English and Welsh Migration to the United States

Between 1850 and 1940—during the so-called Age of Mass Migration—more than 30 million Europeans migrated to the United States (Abramitzky and Boustan, 2017). Migrants from Great Britain—England and Wales in particular—accounted for approximately 10% of this flow (Willcox, 1928). Emigration rates in Britain were among the highest in Europe, except for the years 1890–1900. They steadily increased throughout the period (Baines, 2002).[11]

2.1.1 Migration Policy in the United Kingdom and the United States

The virtual absence of legal constraints to human mobility represents a major appealing feature of the Age of Mass Migration for economic research. Until 1917, the US applied minor restrictions on European immigration (Abramitzky and Boustan, 2017).[12] Immigrants mostly originated from Northern Europe, particularly the United Kingdom, Ireland, Germany, Sweden, and Norway. This positive attitude towards

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

[12] Immigration from China had been severely restricted since as early as 1882. Restrictions on European immigration before 1917 targeted selected groups, such as convicts and disabled persons. In 1917, Congress passed an act that sanctioned legal immigrants' detention and deportation if they committed a crime within five years of their arrival. The act also imposed literacy tests, which, however, did not significantly impact immigration from European countries (Goldin, 1994).

immigration ceased as flows from Eastern and Southern Europe increased in the 1890s (Goldin, 1994). The restrictive immigration policies of the 1920s, however, allotted generous quotas to the United Kingdom, which were never filled (Abramitzky and Boustan, 2017).[13]

Like in other European countries, out-migration legislation in the UK sought to help emigrants, if not explicitly to foster emigration (Baines, 2002, p. 72). 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. Emigration to the United States was neither subsidized nor discouraged. Attitudes towards out-migration remained positive after the First World War. The perceived slowdown of emigrant flows after the War was viewed with concern by policymakers (Leak and Priday, 1933).

This overview suggests that institutional constraints to US out- and immigration were largely absent for English and Welsh migrants throughout the XIX and early XX century. Compared to contemporary scenarios, this historical setting thus allows us to abstract from confounding factors arising from endogenous migration legislation.

2.1.2 English and Welsh Emigrants: The Perspective of Great Britain

Compared to the broader European phenomenon, the British migration to the US presents two main distinctive features.[14] First, unlike continental countries, Britain was already highly urbanized and industrialized at the inception of the Mass Migration. Erickson (1957, 1972) and Thomas (1954) highlight the centrality of urban areas which, starting in the 1880s, supplied the majority of overseas migrants. Baines (2002) provides some estimates on the origin of migrants based on birth certificates over the years 1850–1900. Emigration ratios were highest in Northern and South-Western England and lowest in Lancashire and neighboring areas. Second, the selection of British migrants radically differed from that in continental countries (Erickson, 1957; Abramitzky *et al.*, 2020). 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

[13]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).

[14]Throughout the period, the US was the most relevant destination for English and Welsh migrants. Between 1850 and 1930, more than 40% emigrants settled in the US. This compares to 25% in Canada, 20% in Australia, and 15% in other destinations (Baines, 2002, p. 63).

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 (Baines, 2002). 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%.

Our newly constructed migration database allows us to assess the historical evidence quantitatively. In Appendix Table D.4, we compare individual-level characteristics of emigrants with the staying population. On average, emigrants were more likely to come from North West and South East England. Moreover, they were less likely to be farmers. By contrast, emigrants' share of high and low-skilled manufacturing workers is substantially larger than among stayers. Similar—although less marked—patterns were observed for return migrants. Appendix Figure C.6 displays the origin of emigrants over time at the district level. The data vividly show that rural areas in central and south-western England, which initially feature the highest emigration rates, were gradually replaced by urban industrial districts in the North and South. Taken together, this evidence confirms the qualitative historical knowledge.

2.1.3 English and Welsh Immigrants: The Perspective of the United States

British immigrants have been central throughout the economic and political history of the United States (Berthoff, 1953; Fischer, 1989). Several features distinguish the English from the continental transatlantic migrations. First, English and Welsh immigrants were, especially after the 1880s, artisans and manufacturing workers, who settled where their skills were in highest demand (Berthoff, 1953).[15] Textile workers from Manchester typically settled in Massachusetts, whereas coal miners from Southern Wales mostly settled in the Midwest and Pennsylvania. In 1890, 63% British-born were employed in manufacturing (Thistlethwaite, 1958). Second, English immigrants—unlike the Welsh—did not form ethnic clusters (Furer, 1972). Instead, they tended to be scattered around settlement areas in highly diverse ethnic communities. Finally, British immigrants were economically successful and assimilated relatively easily with the US-born population (Abramitzky *et al.*, 2020).

We quantitatively evaluate these observations in Table D.5. First, we compare individual-level characteristics observed in the US census between the US-born and the British immigrants. The analysis suggests that British immigrants were substantially different from the average native. For example, they were richer, more literate, and more likely to live in urban centers. Consequently, they were less likely to be farmers and more likely to be employed in manufacturing occupations with high or low-skill content.

[15] Thistlethwaite (1958) presents one instructive example. The pottery industry, a highly skilled and labor-intensive sector, was concentrated in the Five Towns of Staffordshire. As transatlantic migration ensued, ceramic workers located in just two localities: Trenton, New Jersey, and East Liverpool, Ohio.

In addition, English immigrants were comparatively more concentrated in North Atlantic states and the West and less in Southern states. Similar patterns emerge for return migrants.

These results, coupled with Table D.4, identify British immigrants in the US as part of an urban industrial class of skilled and semi-skilled workers. This is crucial in our analysis: it would have been much more difficult for illiterate farmers to facilitate knowledge flows across the Atlantic Ocean.

2.2 Intellectual Property Protection in the US and the UK

We measure innovation and knowledge flows using patent data. In this section, we briefly present the key features of the British and American patent systems and discuss the state of international intellectual property protection in the XIX and early XX centuries.

2.2.1 National Patent Systems

Britain established the world's oldest continuously operating patent system in 1623-1624. Until 1850, access to intellectual property protection was, however, difficult (Gomme, 1948; Bottomley, 2014). Fees amounted to approximately four times the average income in 1860, and the application process was lengthy and rife with uncertainty (Dutton, 1984). A large literature documents the poor performance of this system during the Industrial Revolution (Macleod, 1988; Moser, 2012). The 1852 Patent Law Amendment Act sought to reform this process. The US system inspired the reform effort, which reduced application fees and attempted to streamline bureaucratic procedures. One subsequent reform in 1883 further reduced fees, allowed applications by mail, designed a litigation system and provided for the employment of professional patent examiners (Nicholas, 2011). 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).

The first article of the United States Constitution establishes that inventors be granted exclusive rights over their discoveries. In 1836 the 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, an examination of novelty was carried out by professional examiners to ascertain the originality of patent applications. Second, low application fees ensured that access to intellectual property protection was widespread (Sokoloff and Khan, 1990). Several scholars documented how effectively the US patent system fostered innovation well into the 20$^{\text{th}}$ century (Khan, 2020).

2.2.2 International Intellectual Property Protection

As national patent systems spread across Europe and the US during the 19$^{\text{th}}$ 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).

The Paris Convention emerged out of a decade of multilateral confrontations spurred by World Exhibitions in Vienna (1873) and Paris (1878). The 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 had vigorously demanded reciprocity (Penrose, 1951). Second, upon applying for a patent in one 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. The provisions contained in Article 4 were central within the broader legal apparatus (Penrose, 1951). 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. This represented a substantial bureaucratic and financial burden. While the Paris Convention—and its numerous amendments—are still in operation today, international patents were established only in 1970. The UK joined the Convention in 1884, while the US waited until 1887.

The state of international intellectual property protection during our period is a major advantage of this historical setting. Since the UK and the US did not mutually recognize patents, we can use them as an informative proxy of knowledge flows between the two countries. This approach would be impracticable in modern settings.

2.3 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.

2.3.1 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.[16] 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.

2.3.2 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.

2.3.3 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.

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

3 Data

This section presents our primary data sources and discusses the key methodology we adopt to assemble the final datasets. We provide a more detailed description of the data in Appendix sections A, B, and C. Table 1 lists the main variables and provides descriptive statistics.

3.1 Migration Data

To conduct our analysis, we need information on the origin of English and Welsh immigrants in the United States *within* the United Kingdom. Currently available data, however, do not contain this information. Neither the US nor the UK collected disaggregated data on, respectively, the origin of immigrants and the destination of emigrants. We tackle this limitation of the data by developing a new dataset that links British immigrants in the US to the UK census. This allows us to observe an individual in the UK and to track him to his US census record after he emigrated.[17] This is the first dataset that reconstructs migration flows at this granular level of aggregation for a major European country in this period.[18]

To construct our linked dataset, we leverage non-anonymized individual-level data from the population censuses in the United Kingdom (Schurer and Higgs, 2020) and the United States (Ruggles *et al.*, 2021). We first extracted the universe of British immigrants from the US censuses in 1900, 1910, 1920, and 1930.[19] These list, among other variables, the name and surname, birth year, and immigration year of each migrant. We then match these records to the closest census when they appear. Hence, for example, we try to link an individual who immigrated to the US in 1905 to the 1901 UK census.[20] The matching variables we consider are the name, surname, and reported birth year. We use state-of-the-art census-linking algorithms adapted from pioneering work by Abramitzky *et al.* (2021). Appendix C.1 lists in more detail the primary sources and the technical implementation of the algorithm. This class of linking algorithms relies on the observation that a simple exact matching routine would artificially discard many

[17] 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.

[18] Data assembled by Abramitzky *et al.* (2014) and Andersson *et al.* (2022) serve a similar purpose for, respectively, Norway and Sweden. England and Wales, however, were substantially larger in terms of the overall population and the US immigrant population. The population of Sweden and Norway in 1890 was approximately 4.7 and 2 million. In the same year, the population in England and Wales stood at 27 million.

[19] We cannot use information contained in the 1870 and 1880 censuses because the immigration year was not recorded. Individual-level data from the 1890 census have not survived.

[20] 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.

plausible links between the two sources because of minor coding errors by the census enumerators. Since human hand-checking is unfeasible, we implement an algorithm that returns a match whenever the string similarity between the US and the UK records is above a certain threshold, conditional on the birth year.

This approach presents some important caveats (Bailey *et al.*, 2020). 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. This would be the case if, for instance, the likelihood that names and surnames were correctly enumerated in the censuses correlated with education. We discard the matches that do not attain a high level of string similarity to address the first concern. Moreover, we only keep immigrants matched with up to two records in the UK census. This ensures that we minimize the rate of false positives as much as possible. We provide evidence against the second issue in Table C.1, which shows that the correlation between the number of matches and individual-level observable characteristics is seldom significant, and always very small in magnitude.

We perform one additional exercise to assess the plausibility of the linked migrant sample. Following Abramitzky *et al.* (2021), we construct an intergenerational linked sample that identifies individuals in census decade t in the subsequent census in decade $t+10$. The underlying rationale is that the matching rate in this intergenerational linked sample should be lower for US emigrants than for the non-migrant population. We discuss this approach in more detail in Appendix section C.2.3. Figure C.4 reports the results of this exercise. We link approximately 40% non-migrants to an individual in the following decade. As expected, this figure decreases to 20% for US emigrants. This exercise thus provides reassuring evidence that the UK-US linked sample can confidently identify migrants. Moreover, Figure C.5 confirms that the migratory flows in the baseline sample are highly consistent with those that are obtained by repeating the UK-US linking but excluding individuals that are matched in the intergenerational linked sample.

Finally, we construct a dataset of return migrants. To assemble it, we apply the exact previous logic, except that migrants are matched to the UK censuses taken in the decades *after* their immigration year. Hence, as an example, someone who migrated to the US in 1895 is matched to censuses in 1901 and 1911. To avoid double counting, if a migrant is matched to more than one census, we keep the match(es) in the first. Data on return migrants are generally scant historically and with modern data (Dustmann and Görlach, 2016). This exercise is thus a valuable feature of our methodology.

In Figure 1, we report in gray 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. We attain a matching rate of about 60% after dropping multiple matches and links with below-threshold matching quality. Note that we are forced to discard women whose surname was likely to change after marriage. The matching rate aligns with the literature on census linking (Abramitzky *et al.*, 2021).[21] Moreover, reassuringly, our data co-moves with official statistics data. Figure 2 reports the spatial distribution of emigration rates across districts in the final sample and highlights its cross-sectional spatial heterogeneity. In Appendix Figure C.6, we break down the map by decade and uncover substantial variation in the origin of US emigrants over time.

3.2 Patent Data

We measure innovation activity using patents, as is standard in the literature (Griliches, 1998).[22] Patents for the United States have been digitized from original documents by Berkes (2018). The data contain, among others, information on the authors' addresses, the filing date, and the CPC patent classification. We use these to construct a balanced panel dataset at the county-technology class-year level.[23]

Patents for the United Kingdom for the period 1895-1939 are collected from PATSTAT, which in turn provides bulk access to data stored at 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 (2022) and map them to registration districts at their 1890 borders. Patent data for previous years, unfortunately, are not currently available. To tackle this data limitation, we digitize the universe of patents granted in England and Wales between 1853 and 1895. As a result, we assemble a unique patent-level database that leverages textual information from approximately 300,000 original patent documents.[24] We have information on the title, text, inventors' geo-references addresses, filing and issue date, and other variables not used in this paper. Next, we map patents to districts at 1890 borders. We then employ a simple machine learning classification algorithm, discussed in Appendix section B.1, to assign technology classes using information contained in the titles.

This newly developed dataset is the first with geographical and textual information on the universe of patents granted in England and Wales during the second half of the nineteenth century. Data by Han-

[21] In Appendix section C.2, we provide a more detailed discussion of the algorithm's performance.

[22] Previous research shows that 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 2.2, before our study period, the US and the UK had enacted important reforms that decreased the cost of access to patent protection (Gomme, 1948). 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.

[23] We map patents to counties at 1900 borders using the inventors' coordinates. From the three-digit CPC class, we map patents to a coarser taxonomy of twenty sectors. Appendix A.1 provides additional details.

[24] Appendix section B.1 describes the primary sources and methodology we develop to extract and structure the data from the original documents. In section B.2, we compare our series with two existing series and find that the three are highly consistent for the period of common support.

lon (2016), for instance, do not list titles or texts and do not report geographic information. This dataset thus expands previous work by Nuvolari and Tartari (2011) and Nuvolari *et al.* (2021) and provides the first comprehensive assessment of innovation in Britain during the Second Industrial Revolution.

In some empirical applications, we link patent data to the census. This allows us to assign a unique, consistent identifier to single inventors appearing in multiple patents and to observe individual-level characteristics recorded in 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.3.

3.3 Other Variables

In this section, we provide a brief description of the additional data that we assemble. Appendix section A.1 discusses each more diffusely.

3.3.1 UK Census Data

We assemble district-level statistics from 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 neither budgetary nor administrative authority. The average population was 40,000, which makes them roughly comparable to US counties. Districts undergo minor boundary changes during the analysis period. However, to ensure geographical consistency, we cross-walk all variables to districts in 1890 using the method described in Eckert *et al.* (2020). In particular, the census allows reconstructing the employment shares across sectors and other demographic information.

3.3.2 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 (2022) 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.3 reports the distribution of newspapers.

3.3.3 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, however, the source does not list stations in the London area, in the sample of the telegraph analysis, we conflate London urban districts into a single "London" unit, which we assume to be connected to the telegraph network. Figure A.4 reports the distribution of the stations.

4 Empirical Strategy

This section describes our baseline empirical strategy. We discuss the potential caveats that hinder a causal interpretation of the resulting estimates. Then, we discuss two strategies to address these concerns.

4.1 Baseline Methodology

The central hypothesis of this paper is that exposure to foreign—in this case, American—knowledge through migrant linkages shapes the direction of innovation of the country of origin of the emigrants. We thus develop a simple measure of exposure to US knowledge that leverages two sources of variation. 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).

To implement this intuition, we define knowledge exposure as follows:

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

where i, j, k, and t denote a (UK) district, a (US) county, a technology class, and a decade, respectively.[25] The set J denotes the universe of counties. The knowledge exposure term thus averages district-level exposure to county-level specialization across technology classes. The first term in the summation captures specialization, while the second term codes district-level exposure. One may argue, however, that the relative share of patents may inflate the influence of specialization in counties with a small number of granted patents. While this is unlikely to significantly bias our results as those countries would likely have low district-level exposure, we code an alternative knowledge exposure variable that measures

[25] 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.

specialization as the raw count of patents in a given technology class. One further challenge to measure (1) 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 E.2.

We estimate variants of the following regression model:

$$\text{Patents}_{ik,t} = \alpha_{i\times t} + \alpha_{i\times k} + \beta \times \text{Knowledge Exposure}_{ik,t} + \varepsilon_{ik,t} \tag{2}$$

where the coefficient of interest (β) quantifies the correlation between innovation activity and exposure to foreign knowledge. The term $\alpha_{i\times t}$ denotes district-by-decade fixed effects whose inclusion allows to control non-parametrically for time-varying unobserved heterogeneity at the district level; the term $\alpha_{i\times k}$ denotes district-by-technology fixed effects and excludes variation arising, for example, from the possibility that district-level technology specialization and immigration location decisions may be correlated. We comment more on this second point in the next section. The error term is the $\varepsilon_{ik,t}$. Standard errors in this specification are clustered at the district level. We mainly estimate model (2) through ordinary least squares. Since the dependent variable presents a non-negligible share of zeros, we also report the estimates of the Poisson regression associated with the baseline model.[26]

4.2 Threats to Identification

The main factor that cautions against a causal interpretation of the estimates of model (2) is assortative matching, meaning that 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 2.1, we discussed that the historical and quantitative 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 the same in the US and the UK. If (i) does not hold, then the omitted confounding variable would be ab-

[26] In the innovation literature, it is common practice to apply a log transformation to the dependent variable. We do not follow this practice because Chen and Roth (2022) show that average treatment effects for transformations of the dependent variable defined in zero are arbitrarily scale-dependent. In Appendix section D.4, we present alternative specifications with multiple transformations of the dependent variable.

sorbed by district-by-time fixed effects. If (ii) does not hold, the selection bias would be working against our result.

Assortative matching also arises 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 of model (2) 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 non-random location decisions may represent a severe threat in this context. We attempt to quantify this issue in Appendix section D.4.2. We measure the similarity of innovation portfolios between districts and counties and check whether this measure of specialization proximity correlates with observed bilateral migration flows. Table D.2 reports the results. We find no significant association between innovation similarity and migration choices. This suggests that assortative matching is a plausibly minor concern for our analysis. Moreover, in the baseline estimation equation (2), we include district-by-technology fixed effects. Hence, for assortative matching to bias our estimates, the underlying confounding variable would need to vary over time across district-technology pairs.

While we present evidence against the presence of assortative matching, we ultimately cannot rule it out. We thus develop two strategies that, we argue, ameliorate residual endogeneity concerns.

4.3 Shift-Share Instrumental Variable Strategy

We design a shift-share instrument that leverages recent advancements in the econometric literature to deal with selection and assortative matching. Identification critically hinges on the observation that instrument validity can be obtained from the quasi-random assignment of shocks (Borusyak *et al.*, 2022). We construct county-specific immigration shocks by interacting aggregate immigration flows in the US with the gradual expansion of the railway network along the lines of Sequeira *et al.* (2020). These generate exogenous shocks to county-level immigration in a quasi-experimental shift-share design à la Borusyak *et al.* (2022).

To construct the shocks, 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. Importantly, we control for county-level unobserved time-invariant heterogeneity and several other potential confounding variables at the county level.[27] In our context, shocks are conditionally exogenous if the settlement decisions of British immigrants did not influence the direction of the enlargement of the US railway net-

[27] In Appendix section E.2, we describe in more detail the practical computation of the immigration shocks.

work. In other words, instrument validity requires that shocks randomly assign British emigrants across counties. Under this assumption, the instrument breaks concerns of assortative matching. This may fail if, for instance, immigrants settled in counties more similar to their area of origin among the counties connected to the network in a given period. Since county-level shocks yield the overall predicted immigrant shares—and not those of the British only—we believe this is a relatively minor concern to rule out by assumption. Following Borusyak *et al.* (2022), we show that shocks are uncorrelated with county-level confounding variables and that the instrument does not systematically predict district-level characteristics. Appendix Figure E.3 shows that while immigrant shares correlate with district-level observable characteristics (Panel A), predicted immigration shares do not (Panel B). Similarly, in Appendix Figure E.4, we confirm that while out-migration correlated with most district variables, the instrument displays smaller and insignificant correlations with the same variables. These exercises provide evidence in favor of the validity of our research design.

Let $\omega_{j,t}$ be the immigrant share in county j in decade t, and let $\hat{\omega}_{j,t}$ be its prediction. We thus define the instrument as

$$\widehat{\text{Emigrants}}_{i\to j,t} \equiv \hat{\omega}_{j,t} \times \sum_{j\in J} \left(\hat{\omega}_{j,t} \times \text{Emigrants}_{i\to j,1880}\right) \tag{3}$$

where $\text{Emigrants}_{i\to j,1880}$ denotes the number of emigrants leaving district i and settling in county j at the beginning of the sample period. Importantly, this exposure term is allowed to be endogenous by design. Identification stems from the quasi-exogeneity of the shocks $\{\hat{\omega}_{j,t}\}$. Given a predicted set of bilateral flows, we construct the instrument for knowledge exposure as in (1), except that the predicted flows replace the observed ones.

Even though we present evidence suggesting the opposite, the conditional exogeneity of the timing of railway connection is ultimately an untestable assumption. To validate the results obtained with the instrument (3), we construct an additional series of county-level shocks $\{\hat{\omega}_{j,t}\}$ that leverages a different source of variation. Specifically, we compute "leave-out" predicted county-level immigrant shares by interacting start-of-period immigrant shares with aggregate inflows by nationality. Importantly, we exclude British immigrants when calculating these shocks. This ensures that the "leave-out" shares do not reflect the settling decisions of the British. We describe the procedure in more detail in Appendix E.2.2. This alternative instrument yields results that are highly consistent with the railway-based approach.

4.4 Shock Propagation Difference-in-Differences Strategy

The shift-share instrumental variable relies on identifying variation across counties that become connected to the US railway network. Therefore, the associated estimates deliver a local average treatment effect for a complying group of individuals who settle in counties that become connected to the railway network during this period. The literature suggests, however, that it is plausible that these "frontier" migrants would display a relatively higher probability to undertake innovation activity, perhaps due to entrepreneurial attitudes (Bazzi *et al.*, 2020). Under this interpretation, the IV estimates would yield an upper bound to the effect of overall out-migration on British innovation. In addition, because they rely on the specific group of counties that become connected to the railway network, they do not reflect the overall composition of US innovation across technology classes.

To provide additional causal evidence and circumvent these limitations, we devise a research design that leverages geographically clustered innovation shocks in the United States in a triple-differences setting. We start by observing a logical corollary of the return innovation argument. Suppose we observe a sudden increase in the number of patents granted in some counties in some technology classes. Then, one would expect that districts whose emigrants had settled more extensively in those counties would display increased innovation activity in those classes. In other words, innovation shocks in the United States should "reverberate" in the United Kingdom through pre-existing migration linkages.

We test this prediction using two sets of innovation shocks. First, as we describe in more detail in Appendix E.3.1, we construct a set of county-technology class synthetic innovation shocks at yearly frequency. The intuition behind these shocks is that we seek to isolate periods of unusual patenting in a given county-technology class-year, controlling for the average volume of patents produced in that county-class cell. We thus regress the number of patents against fixed effects to obtain the residualized innovation activity. Then, we flag an innovation shock $\xi_{jk,t}$ whenever the residualized number of patents in a given county j, technology class k, and year t is in the top 0.1% of the overall distribution.[28] Appendix Table D.8 documents that shocks are relevant, as one such shock is associated with an average of more thirty patents in the given county. Second, we leverage recent evidence by Berkes *et al.* (2023), who document that the Great Influenza pandemic (1918–1919) significantly and positively affected pharmaceutical innovation in counties that were more exposed to the pandemic. We thus claim that districts that were comparatively more exposed to affected counties should feature increased pharmaceutical innovation. We provide additional details on the construction of county-level exposure to the pandemic in Appendix E.3.2.[29] We code county-level exposure to the pandemic as a dummy φ_j that returns value one if the ratio between deaths during the pandemic (1918–1919) and deaths in the preceding three years (1915–1917) is in the top 25%, and zero otherwise.

We measure district-level exposure to the county-level shocks in terms of the emigrants that had left the given district to settle in the given county *before* the period of analysis.[30] Formally, we compute

[28] In Appendix Table E.11, we show that the results remain consistent when imposing different values to flag innovation shocks.

[29] Since the technology taxonomy used in this paper is different from Berkes *et al.* (2023), in Appendix Table D.7 we confirm that their result holds in our data. Figure D.2 reports the associated flexible triple differences specification. Moreover, in Figure E.6a, we confirm that the pandemic affected innovation activity only in the pharmaceutical sector.

[30] This part of the analysis restricts the outcome variable to 1900–1930, so we can leverage migrant flows in the preceding decade

exposure to synthetic shocks in technology class k as

$$\text{Synthetic Shock Emigrants}_{ik,t} = \sum_{j \in J} \left(\text{Emigrants}_{i \to j,1900} \times \xi_{jk,t} \right) \tag{4}$$

and analogously, we define exposure to counties affected by the pandemic as

$$\text{Influenza Emigrants}_{i} = \sum_{j \in J} \left(\text{Emigrants}_{i \to j,1900} \times \varphi_{j} \right) \tag{5}$$

To avoid issues of continuous treatment described by Callaway *et al.* (2021), we recast each exposure metric in terms of a dummy variable that returns value one if the associated continuous measure is in the top 25%, and zero otherwise.[31]

To estimate the effect of US synthetic shocks on UK innovation activity, we estimate the following triple differences specification:

$$\text{Patents}_{ik,t} = \alpha_{i\times k} + \alpha_{k\times t} + \alpha_{i\times t} + \sum_{h=-a}^{b} \beta^{h} \times \text{I}\left[D_{ik,t} = h\right] + \varepsilon_{ik,t} \tag{6}$$

where $\alpha_{i\times k}$, $\alpha_{k\times t}$, and $\alpha_{i\times t}$ denote, respectively, district-by-technology class, technology class-by-year, and district-by-year fixed effects.[32] The term $\left(D_{ik,t} \equiv t - \text{I}\left[\text{Synthetic Shock Emigrants}_{ik,t}\right]\right)$ denotes the number of years since the district-technology class ik was exposed to a synthetic innovation shock ξ. The roll-out of the treatment is staggered across units. Different district-class pairs may be exposed to the exposure treatment at different points in time.[33] Goodman-Bacon (2021) shows that the standard two-way fixed effects estimator shown in (6) fails to estimate the average treatment effect when treatment effects are heterogeneous, either over time or across groups. Several estimators have been proposed to deal with this difficulty. In the main results, we report estimates obtained using the imputation procedure presented in Borusyak *et al.* (2021). Other estimators yield qualitatively similar results, as shown in Appendix Figure E.7.

We follow a similar approach to estimate the effect of US exposure to the Great Influenza pandemic on UK innovation. In particular, the model is entirely similar to (6), except that the treatment variable is defined as $\left(D_{ik,t} \equiv t - I\left[\text{Influenza Emigrants}_{i}\right]\right)$ as it codes the number of years since the influenza, and it is interacted with a dummy variable returning value one for the pharmaceutical technology class, and

(1890–1899) to construct fixed exposure shares.

[31] In Appendix Table E.11 we consider alternative thresholds to code the exposure variable (4). In Appendix Table E.12, we report the results using the continuous measure (5).

[32] When we estimate regression (6) using variation in exposure to the pandemic shock, we normalize the dependent variable by the average number of patents granted before the pandemic to ensure that the estimated coefficients' size are comparable.

[33] Notice that the treatment is also potentially repeated, for the same unit can be treated multiple times. This is, however, not the case in the baseline case, where we define synthetic shocks in the top 0.1% of the overall residualized innovation shock distribution.

zero otherwise.[34]

The primary estimation strategy in this setting is thus a triple difference estimator (Olden and Møen, 2022). A causal interpretation of the resulting estimates requires that the difference between the within-group differences are not statistically different from zero before the treatment. Several papers highlight that, compared to the standard difference-in-differences estimator, the parallel trends assumption in this setting is relatively weak because it only requires that no contemporaneous shock affects the relative outcome of the treatment and the control group (Gruber, 1994). Throughout the paper, we present flexible triple difference estimates to provide evidence supporting the parallel trends assumption.

5 Empirical Results

In this section, we present the main return innovation result. Then, we document that shocks to US innovation diffuse into the UK through migration ties. We interpret these results as evidence that migration flows contribute to the diffusion of innovative knowledge to countries sending migrants.

5.1 Exposure to US Innovation Shapes Innovation in the UK

The primary finding of this paper is that exposure to foreign technology through migration ties shapes the dynamics and direction of innovation in the emigrants' country of origin.[35] We label this novel finding "return innovation". We first estimate regression (2) through a simple OLS linear probability model to document it. We report the results in columns (1–3) of panel A of Table 2. There is a positive, significant, and quantitatively large correlation between the baseline measure of exposure to foreign knowledge and the number of patents at the district-technology class level. Moreover, the correlation persists over time, as the estimates remain statistically significant after two decades. In columns (1–3) of panel B we repeat this exercise, but we normalize the number of patents by the district-level population at the beginning of the sample (1880). We confirm the positive association between knowledge exposure and per-capita patents. Importantly, all regressions include district-by-decade fixed effects to account for unobserved time-varying heterogeneity at the district level. Moreover, we control for district-by-technology class fixed effects to partial out spurious variation arising from initial district-level specialization across classes.

As discussed in section 4.2, at least two factors hinder a causal interpretation of the estimates pre-

[34] This specification focuses on the ATE on pharmaceuticals compared to other technology classes. In Appendix Table E.12, we report the double differences estimates associated with model (6). Then, in Figure E.6b, we show that, as in the United States, the influenza had a major effect on pharmaceutical innovation only.

[35] A recent literature produced compelling evidence that exposure to innovation is a key determinant of subsequent innovation activity (Akcigit *et al.*, 2018; Bell *et al.*, 2019). Our results can thus be interpreted as additional new evidence in favor of this thesis.

sented in panel A. First, out-migration is not random across districts. Second, there may be some latent determinant of the settlement location decisions of emigrants that correlates with innovation activity in their origin areas. To ensure that our estimates do not reflect spurious correlation arising from omitted variable bias issues, we estimate model (2) using the instrument (E.6). In columns (4–6) of panels A and B, we report the reduced-form association between the instrument and the dependent variable. Figure 3 visually compares the OLS and the reduced-form IV regressions. We confirm the positive and statistically significant effect of knowledge exposure on innovation. The effect persists until one decade, as opposed to two from the OLS estimates.[36] Columns (7–9) report the two-stage least-squares (TSLS) estimation results. First, the instrument is relevant.[37] Second, the TSLS estimates confirm knowledge exposure's positive, large, and statistically significant effect on innovation. The magnitude of the TSLS estimates is roughly similar to the OLS, although the latter appears to be slightly upward biased. The OLS estimates possibly reflect the upward bias introduced by assortative matching across district-county pairs.

The evidence in Table 2 is at the district-technology level. To explore the heterogeneity of the return innovation effect across industries, however, we estimate model (2) at the district level separately for each technology class. We report the resulting reduced-form coefficients of the knowledge exposure instrument—one for each regression—in Figure 4. We estimate the largest treatment effects for industries like electricity and chemistry that were at the forefront of the Second Industrial Revolution (e.g., Mokyr, 1998). We employ the UK-revealed comparative advantage to measure the relative sector-level innovation specialization.[38] We find that the return innovation effect is larger in sectors where the UK retained an advantage at the beginning of the period (the 1880s). Rather than igniting the emergence of entirely new sectors, our results suggest that exposure to US knowledge through migration ties nurtured already-existing industries.

The setting of this study allows for gauging the persistence of the association between exposure to

[36] Bilateral migration flows are known to be highly persistent over time—a phenomenon known as "chain migration". This would inflate the OLS association between lagged knowledge exposure and innovation. At the same time, this would explain why the associated TSLS estimates are not significant. The instrument, in fact, effectively breaks the persistence of migratory flows using plausibly exogenous county-level immigration shocks.

[37] We report the complete first-stage estimates in Appendix Table E.7. The instruments are always relevant and capture a substantial share of the variation of the endogenous variables.

[38] In the international trade literature, the revealed comparative advantage is a widely-employed 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

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

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

foreign knowledge and innovation.[39] In Appendix Figure D.3, we report the coefficients of a regression between the number of patents and an interaction term between knowledge exposure in the period 1900–1930 and biennial time dummies from 1940 to 2014. The estimates suggest that the positive effect of knowledge exposure on innovation persists for almost four decades, albeit the magnitude decreases over time. Starting in the mid-1970s, the association gradually becomes small and statistically insignificant. In Appendix Table D.9, we repeat the exercise by technology class and find consistent results across sectors. Migration ties thus generate enduring knowledge flows that shape innovation activity over the long run.

The analysis presented thus far focuses on how out-migration shaped the *direction* of innovation.[40] A natural question is, however, whether it also impacted the *volume* of patents. Our data are not especially well-suited to answer this question because we lack disaggregated data on outright emigration. Nevertheless, if emigration to countries other than the United States correlated with US emigration, we can present some suggestive evidence. In Table D.6, we estimate the effect of out-migration on innovation, measured as the number of patents granted. The OLS and TSLS estimates show that out-migration has a negative short-term impact on innovation, but this reverses in the medium run (after one decade). Our findings thus appear to reconcile evidence of "brain drain," which views out-migration as a depletion of human capital that hampers innovation, with "brain gain" arguments suggesting that emigrants may be conducive to economic growth via, for instance, monetary remittances (Docquier and Rapoport, 2012). The results suggest that the former hypothesis is predominant in the short-run, while the brain-gain perspective materializes in the medium-to-long term. The effect of out-migration on the volume of innovation has been the focus of many of the existing studies (Agrawal *et al.*, 2011; Andersson *et al.*, 2022). This paper, instead, provides evidence that emigration is a fundamental driver of the direction of innovation.[41] From this perspective, our results thus inform the recent literature studying the determinants of the direction of innovation (Bell *et al.*, 2019; Einiö *et al.*, 2022).

We perform several robustness exercises to gauge the robustness of our results. These are reported in the Appendix and discussed in section E.1. First, we consider alternative dependent variable transformations in Table E.1. Second, Table E.2 reports the results using five different definitions of knowledge exposure that hold fixed various margins of variation. The baseline specification of model (2) includes district-by-decade and technology class-by-decade fixed effects. In Table E.3, we show that the results are robust to alternative, demanding specifications. The standard errors are clustered at the district level in the baseline specification. In Figure E.1, we adopt various estimators and confirm that they all preserve

[39] We discuss the technical details of the long-run analysis in Appendix section D.3.

[40] Appendix section D.4.1 explores this aspect in more detail and provides the technical details of the analysis.

[41] Our results resonate with evidence by Fackler *et al.* (2020). While their study essentially leverages cross-country variation in emigration destinations, our analysis is based on within-country disaggregated data on the origin and destination of migrants. This allows us to credibly estimate the causal effect of out-migration and investigate possible underlying mechanisms.

the statistical significance of the main results. Another concern is that the return innovation effect concentrates on low-quality innovation and thus bears little relevance in terms of aggregate productivity growth. In Table E.4 we thus report the results of the baseline and instrumental variable regressions, accounting for patent "quality".[42] The results confirm that the number of high-quality patents increases in districts exposed to US knowledge. In fact, the magnitudes are larger than using the raw patent count. This may indicate that the return innovation effect is more intense for marginally more valuable patents. Analogously, in Table E.5 we compute knowledge exposure weighting US patents by their quality using different thresholds and definitions. The results qualitatively confirm the baseline estimates. We do not have information on the actual adoption of technology by the firms. However, in Table E.6, we restrict the outcome variable to include only patents that list at least one firm as an assignee.[43] These plausibly reflect actual economic activity carried out by British firms. Here, too, the results remain qualitatively similar to the baseline estimates. The instrument used in Table 2 leverages variation in the connection timing to the railway network to randomize immigration across counties. In Table E.8, we report the results using an alternative "leave-out" instrument, described in section E.2. Importantly, we can also use both instruments simultaneously and provide over-identification tests. In Table E.9, we confirm that the leave-out instrument results are robust to various alternative definitions of the county-level shocks.

5.2 Innovation Shocks in the US Diffuse to the UK

The return innovation result indicates that migration ties shape the direction of innovation in the origin areas of emigrants. We claim that this finding implies that fluctuations in patenting activity in the United States would reverberate in the United Kingdom through migration linkages. We estimate model (6) using two different sources of such fluctuations—which we label innovation shocks—to test this hypothesis.

Table 3 reports the results of this exercise. Columns (1–4) refer to the synthetic shocks series we construct by residualizing the observed patenting activity against fixed effects and flagging large increases in the resulting series as "innovation shocks". As a preliminary robustness test, we report the full-sample estimate in column (1), while columns (2–4) exclude districts in the top three areas in terms of patents granted. We estimate a positive, large, and statistically significant effect of US synthetic innovation shocks on innovation activity in the UK. We estimate an average of 0.4 patents per year in the treated technology class after the shock in exposed districts. This is a quantitatively sizable effect since

[42] Following Kelly *et al.* (2021), we define a text-based measure of quality which flags as influential those patents that introduce words that did not appear before they were published, and become used thereafter. Because we have full texts for the period 1880–1899 and abstracts only between 1900 and 1939, in this exercise we restrict the sample period to the latter years.

[43] Unfortunately, data on firm assignees is only available for the sub-sample 1870–1900, when we have the full text of the patent specifications form which this information is extracted.

the average number of patents per district-class pair is 1.3. Moreover, the relative size of the effect remains consistent throughout the regression samples. Next, we explore heterogeneous treatment effects over time in Figure 5a. Reassuringly, the figure provides evidence that supports the parallel trends assumption. The effect of the innovation shock is the largest and most significant after two years since the shock initially manifested in the United States. This time lag seems plausible, especially since our data shows an average of 1.1 years delay between the application and issue date at the UK patent office. The effect persists up until six years following the synthetic shock. We estimate the effect of synthetic shocks sector by sector in the appendix Figure E.5. As in 4, we find the largest treatment effect for electricity.

Next, we investigate how exposure to the Great Influenza pandemic across US counties impacted UK innovation. The logic behind this exercise is that exposure to the pandemic fostered innovation in the pharmaceutical sector (see Table D.7 and Berkes *et al.*, 2023). We thus expect districts whose emigrants had settled in counties more exposed to the pandemic to display higher patenting rates in pharmaceuticals. We report our findings in columns (5–8) of Table 3. We estimate the pandemic shock's effect on British innovation to be positive and sizable. On average, two patents per year are granted in the pharmaceutical sector in districts more exposed to counties severely affected by the influenza. We estimate the associated dynamic treatment effects in Figure 5b. We find only one marginally significant and very small coefficient in the pre-treatment period. By comparison, the post-treatment coefficients are large and highly significant. The effect of the pandemic materialized six-seven years after the shock in the United States. As noted before, this delay is partly due to the shift between patent application and issue by the patent office, except that we now have to compound delays at the US and UK offices. Moreover, the effect of the pandemic shock on US innovation in pharmaceuticals was not immediate, as shown in Appendix Figure D.2. Taken together, it is plausible that the propagation of the innovation shock into the UK is observed with some delay. We estimate statistically significant treatment effect coefficients for more than a decade thereafter.

The pandemic shock only impacted innovation in pharmaceuticals in the US (Figure E.6a). We thus expect to retrieve a similar effect in Britain. Figure E.6b shows that, although the point estimates are not as sharp as in the US case, the pharmaceutical sector is the one that benefits the most from the influenza shock. The point estimate for pharmaceuticals is nearly three times larger than the second-largest estimate. The estimated effect in some sectors may be negative because of crowding-out out of those fields into pharmaceuticals, although we cannot entirely disentangle the underlying reason. We interpret this exercise as a falsification check: Figure E.6 provides convincing evidence that the pandemic shock affected the same sector in the US and the UK.

We assess the robustness of these results through several robustness checks. First, we consider alternative thresholds to (i) flag synthetic shocks and (ii) flag district exposure to synthetic shocks. In

Table E.11, we estimate larger treatment effects for smaller thresholds. This is reasonable since smaller thresholds impute, on average, larger innovation shocks. The synthetic shock triple differences model is a staggered design since shocks generally occur in different periods across technology classes and districts. The baseline estimates are obtained from the imputation estimator developed by Borusyak *et al.* (2021). In Figure E.7, the estimated treatment effect remains consistent across various estimators. In particular, the one developed by De Chaisemartin and D'Haultfœuille (2022) allows repeated treatments and yields consistent results. In Table E.12, we report several specifications to gauge the robustness of the pandemic shock results. First, in columns (1–2), we report the double differences estimates that compare pharmaceutical innovation across districts by exposure to counties affected by the pandemic. Then, in columns (3–7), we report various triple differences specifications that exclude districts in areas with very high patenting activity. The results remain consistent throughout.

5.3 Technology Transfer and Spillovers: A Text-Based Approach

It is natural, at this point, to ask to what extent the return innovation effect manifests because the *same* patents that are granted in the United States are, at some point, issued in Britain. In other words, is the return innovation effect about UK areas with higher exposure to US knowledge "copying" US-developed innovation?

To quantify the extent of copying and, on the other hand, the spillovers in terms of "original" innovation that would be generated by migrant linkages, we develop two text-based similarity measures.[44] First, we compute the backward similarity between patents granted in the UK and previous patents granted in the US. Leveraging textual information contained in patent titles, this approach allows us to measure whether patents produced in areas with larger exposure to US knowledge become more similar to US inventions. Second, we compute a measure of patent "originality" by comparing patents granted in the UK to previous and subsequent US patents. Specifically, we deem a given patent as more innovative if it is more similar to subsequent US patents relative to previous ones. This approach mirrors the methodology of Kelly *et al.* (2021). Both indices are computed at the patent level and aggregated up at the district-technology class level at a decade frequency.

In Figure 6, we report the effect of exposure to US synthetic innovation shocks on the similarity between patents produced in the UK and those issued in the US. In Panel 6a, we estimate a positive and significant effect of exposure to US knowledge on the backward similarity between UK and US patents. This suggests that, to some extent, knowledge flows generated by migration ties stimulate emulation and technology transfer between the United States and Britain. In Panel 6b, however, we estimate the same model using the "originality" of UK patents as the dependent variable. Here, too, we find that areas with

[44] In Appendix section B.3, we present the analytical derivation of the similarity metrics that we abridge from Kelly *et al.* (2021).

more intense exposure to foreign innovation produced more innovative patents (compared to previous US innovations). These results thus suggest that the return innovation effect conflates two distinct margins through which exposure to foreign knowledge affects the production of innovation. First, migrant linkages fostered the technology transfer of already-existing inventions in the UK. Second, they also propelled the development of novel, original inventions. Interestingly, we estimate that a synthetic shock triggers a sudden increase in the backward similarity of innovation, while the effect on original patents manifests more slowly.

In Appendix Table E.14, we tabulate the associated estimates and report the results for the Great Influenza pandemic shock as well. We confirm that following an innovation shock in the US—either a synthetic one or the Great Influenza pandemic—the similarity between UK and US patents increases (Panel A). At the same time, areas more exposed to the shock start producing more original patents (Panel B).

In Appendix Table E.13 we report the OLS and instrumental variable estimates of regression (2), which confirm the baseline results obtained using the triple differences estimator. We gauge the robustness of these results using three alternative measures of similarity, all displayed in Appendix Table E.10. First, we use the "raw" similarity measure between titles, which does not take the log of the cosine similarity between patent titles (columns 1–3). Second, we net out the year and technology fixed effects from the patent-level originality and backward similarity measures to ensure that our estimates do not conflate time-varying terminology and fashion trends (columns 4–6). Finally, while in the baseline analysis, we compute the similarity metrics over a ten-year window around each patent, in columns (7–9), we restrict it to five years. The results remain qualitatively unchanged throughout.

6 Potential Mechanisms and Discussion

Several concurrent, not necessarily mutually exclusive mechanisms can explain the return innovation result. In this section, we present our analysis to disentangle some. First, we establish whether return innovation is solely a consequence of return migration. Then, we discuss some complementary and possibly quantitatively more substantial channels.

6.1 Is Return Innovation Return Migration?

Return migration is a primary candidate to explain 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. On the one hand, several studies estimate modest effects for recruiting programs of high-skilled nationals working abroad (Ash *et al.*, 2022; Shi *et al.*, 2023). On the other, Giorcelli (2019) shows, although from a different perspective, that those exposed to (managerial) foreign knowledge change their behavior once back in their origin country.[45] In this section, we quantify the relative importance of return migration in generating return innovation.

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 and compute a measure of "return knowledge exposure" which is analogous to (1):

$$\text{Return Knowledge Exposure}_{ik,t} \equiv \sum_{j \in J} \left(\frac{\text{Patents}_{jk,t}}{\text{Patents}_{j,t}} \times \text{Return Migrants}_{j \to i,t} \right) \tag{7}$$

where Return $\text{Migrants}_{j \to i,t}$ 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 the period 1870–1910.

As a first step, we estimate model (2) controlling for return knowledge exposure. Table 4 reports the results: in columns (1–3), we present specifications with various levels of fixed effects; columns (4) and (5) display the coefficients of lagged values of the independent variables; in column (6) we report the full lag model. Throughout the specifications, the coefficients of baseline and return knowledge exposure remain comparable in size—by looking at the respective standardized beta coefficients—and statistically significant. These results suggest that return migration is an important driver of return innovation. In our data, approximately 30% emigrants return, and these account for approximately 50% of the total return innovation effect. At the same time, however, a substantial proportion of return innovation is not explained by return migration.

Absent physical return, in the rest of the paper, we provide evidence of two additional mechanisms that underlie the return innovation effect. First, we focus on interactions between the emigrants and local communities in Britain. Second, we explore how migration ties facilitate cross-country market integration, thereby promoting knowledge flows.

6.2 Return Innovation Through Interactions

In this section, we explore if and, in case, how emigrants interact with local communities who remained in Britain. We distinguish between two cases. On the one hand, interactions could require physical return. On the other, emigrants may exchange information while abroad. When studying the interactions between emigrants and stayers, one needs to delimit the set of stayers with whom emigrants could plausibly interact. We focus on two factors that could promote social interactions between the emigrants and their communities of origin: family ties and pre-migration geographical proximity (neighbors).

[45] Choudhury (2016) shows that R&D firms with returnee managers are disproportionately more likely to file patents in the United States. Bahar *et al.* (2022b) show that return migration can influence trade.

6.2.1 *Interactions Between the Emigrants and their Family*

Our data do not contain exhaustive information on the families of emigrants. At best, we know those living in the same household. This would be an exceedingly restrictive definition because, in most cases, it would exclude brothers or parents. We thus adopt a less conservative approach. In particular, we assume that it is likely that individuals with the same surname who live in geographical proximity—in the same UK county—are relatives. This assumption is reasonable as long as surnames are not too common: in this analysis, we thus the top 5% most common surnames. Results are robust to alternative sample cuts.

We implement a triple-differences estimation. The outcome variable is the total number of patents granted to inventors with a given surname who are recorded living in a given county in the UK. The treatment leverages variation in the surname of US emigrants by county. Under the previous assumption, this model quantifies how the emigration of family members impacts the patenting activity of those who remain in the UK. Formally, it is

$$\text{Patents}_{sc,t} = \alpha_{s\times c} + \alpha_{c\times t} + \alpha_{s\times t} + \beta \times \text{US Emigrant}_{sc,t} + \varepsilon_{sc,t} \tag{8}$$

where s, c, and t denote, respectively, surnames, counties, and years. The treatment $(\text{US Emigrant}_{sc,t})$ is a variable equal to one after an individual with surname s from county c emigrates to the US, and zero otherwise. Therefore, under the standard parallel trends assumption, β estimates the impact of emigration on patenting activity carried by the relatives of the emigrant. To deal with the sharp left skewness of the outcome variable, for each estimate, we report an analogous model that features as the outcome variable a categorical indicator that returns a value of one if the number of patents is strictly positive and zero otherwise.[46]

Columns (1) and (5) of Table 5 report the baseline estimates. Emigration to the US has a positive effect on patenting activity by the relatives of emigrants who remain in the UK. The effect is quantitatively sizable. In Appendix figure E.8 we report the associated flexible triple differences model. The estimates provide evidence in support of the parallel trends assumption. Moreover, they show that it takes, on average, ten years before a British emigrant to the US contributes to the innovation activity of his family in the UK. This delayed effect is plausible inasmuch as it would take time for emigrants to settle in the US and be exposed to technology that they could diffuse back into the UK.

In the second part of the analysis, we distinguish between emigrants who return from those who do not. Emigrants could interact with their families upon returning, but they could also maintain ties while abroad. In columns (2) and (6), we thus restrict to emigrants that, at some point, return to the UK. In columns (3) and (7), by contrast, the treatment includes only those that never return. The impact of

[46] The results remain qualitatively unchanged estimating model (8) as a Poisson regression.

return emigrants is four to five times as large as that of those who never return. This difference suggests that return migration is, as we argued in the previous section, a major driver of return innovation. At the same time, however, emigrants promote innovation by their relatives even if they never return. When we compare the two coefficients in columns (4) and (8), we confirm that both return and non-return emigration contribute to innovation in the UK. The relative magnitudes remain unaltered. Because return migration only accounts for approximately 30% of the overall emigration, this analysis suggests that, in quantitative terms, interactions between stayers and emigrants who never return to the UK account for approximately half of the overall return innovation effect.

6.2.2 *Interactions Between the Emigrants and their Neighborhood*

In this section, we interpret geographical proximity between emigrants and stayers as an alternative proxy for local social networks. The rationale is that, before leaving the UK, emigrants plausibly maintained social ties to those who lived in their neighborhood. We thus hypothesize that stayers could interact with their former neighbors who migrated to the US. Here, too, we further distinguish emigrants who never return to quantify the relative importance of physical return migration on interactions with origin communities.

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 1911 census. We then create a yearly balanced panel dataset that reports the number of patents obtained by each individual between 1900 and 1920. To do so, we leverage the linked inventor-census data described in Appendix A.3. Next, each individual is geo-referenced to precise coordinates as described in Appendix A.2. We complement this with information on the geographical proximity between these "stayers" and migrants. More specifically, we define a dummy variable $(\text{Neighborhood Migrant}^{k}_{p,t})$ that returns value one in all periods after the first time at least one individual living within k meters from individual p migrates to the US, and zero otherwise. In the baseline analysis, we consider $k = 0$, meaning that we only consider emigrants in the same street as the observed individuals, and recast $(\text{Neighborhood Migrant}^{100}_{p,t})$ as simply $(\text{Neighborhood Migrant}_{p,t})$ for brevity. We label this variable an indicator of "neighborhood migration". To estimate the effect of neighborhood migration on the probability of patenting, we thus estimate the following double difference regression:[47]

$$\text{Patents}_{p,t} = \alpha_p + \alpha_t + \beta \times \text{Neighborhood Migrant}_{p,t} + \varepsilon_{p,t} \tag{9}$$

where p and t denote, respectively, individuals and years, and α_p and α_t are the associated fixed effects. The term β yields, under a standard parallel trends assumption, the estimated causal effect of neighborhood migration on innovation.

The logic beneath equation (9) builds on Bell *et al.* (2019), who document the importance of geographical proximity to inventors as a driver of subsequent innovation activity. A positive and significant estimate of β would be evidence that emigrants promote the innovation activity of their neighbors. Then, we define a (Non-Return Neighborhood Migrant$^{k}_{pt}$) dummy entirely analogous to the previous treatment, except that we condition the neighborhood emigrant to not return to the UK. In this case, a positive estimate of β would suggest that neighbors benefit from interactions with the emigrants even if those never return.

We report the estimates of equation (9) in Table 6. The dependent variable is the yearly number of patents. In columns (1–4), the sample includes individuals from all districts; in columns (5–7), we exclude individuals in the top three-producing patents areas (London, Lancashire, and the South-West). In panel A, the treatment is activated by any US neighborhood emigrant. In panel B, we restrict to neighborhood emigrants that never return in the sample period. We estimate a positive effect of neighborhood emigration on innovation by non-migrants. The effects hold in the baseline specification (columns 1 and 5), as well as including parish-by-time fixed effects (columns 2 and 6) and applying coarsened exact matching (CEM, columns 3 and 4).[48] Importantly, the estimated coefficient remains if we restrict the sample to exclude all non-inventors, thus reducing the sample size considerably (columns 4 and 8). Panels A and B show that overall and non-return neighborhood migration has a positive statistically significant effect on the probability of inventing regardless of the dependent variable, the fixed effects, and the matching scheme. In Figure E.10, we report the associated flexible difference-in-differences estimates, which indicate the absence of statistically significant pre-trends. In Appendix Table D.10 we explore the hetero-

[47] To avoid an excessive computational burden, we estimate model (9) on a 10% random sample of the population. Moreover, the model is a staggered difference-in-differences design with (potentially) repeated treatments. We thus estimate regression (9) using the estimator proposed by Borusyak *et al.* (2021). In Appendix Figure E.9 we show that the estimated coefficient remains stable using several different staggered difference-in-differences estimators. In Appendix Figure E.15, we show that results hold if the neighborhood-migrant treatment is activated whenever emigrants within 100 meters from the individual in the sample migrate.

[48] Parishes are very small geographical units with a population of approximately 2,500. Coarsened exact matching weights are calculated to balance individuals in terms of age, parish of residence, and occupation. Appendix Figure E.11 reports the correlation between treatment status and pre-treatment individual-level observable characteristics for the baseline sample (panel A) and the CEM weighted sample (panel B).

geneous response to neighborhood out-migration across the age and occupation of the stayer individual. In particular, we find that the gains are larger for relatively young individuals (column 1) and accrue to those employed in skilled occupations (columns 2–4).

Evidence presented in Table 6 provides additional evidence that emigrants promote innovation in the communities they come from. Those who never migrate but who were in close geographical proximity with the emigrants before they left, a proxy for local social networks, benefit from interactions with the emigrants. This channel operates even if the former neighbor never returns. These results highlight the importance of cross-country interactions between the emigrant population and their origin communities. Our findings confirm experimental evidence from developing countries linking technology diffusion with network interactions (e.g. Bandiera and Rasul, 2006; Conley and Udry, 2010; Beaman *et al.*, 2021). In the rest of the paper, we investigate more aggregate effects of migration ties on innovation activity.

6.3 Return Innovation Through Economic Integration

In this section, we explore whether migration ties promote the cross-border diffusion of innovation irrespective of direct interactions between the emigrants and their origin communities. Building on previous literature, we will argue that migration ties foster the integration between markets. This, in turn, facilitates the diffusion of information and thus fosters knowledge flows.

6.3.1 The Transatlantic Telegraph Increased Innovation In Emigration Districts

We exploit one historically relevant event to provide evidence that migration ties foster cross-border economic integration: the first transatlantic telegraphic cable that connected the US and UK domestic networks (1866). The telegraph represented a major revolution in communication technology that ushered unprecedented market 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).[49]

We claim that if migration ties fostered cross-border market integration between the British and the American market, then districts with relatively higher US emigration rates would be more exposed to the telegraph shock. Under this interpretation, we thus expect that districts with higher US emigration rates after 1866 would display (i) increased innovation activity in districts with more US emigrants and (ii) increased innovation activity in the fields emigrants were exposed to in the US. To test these hypotheses,

[49] 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.

we estimate the following difference-in-differences models:

$$\text{Patents}_{i,t} = \alpha_i + \alpha_t + \sum_{h=-a}^{b} \beta^h \left[\text{US Emigrants}_i \times \text{I}(t - 1866 = h)\right] + \varepsilon_{i,t} \tag{10a}$$

$$\text{Patents}_{ik,t} = \alpha_{i\times k} + \alpha_t + \sum_{h=-a}^{b} \beta^h \left[\text{Knowledge Exposure}_{ik} \times \text{I}(t - 1866 = h)\right] + \varepsilon_{ik,t} \tag{10b}$$

where i, k, and t denote a district, technology class, and yea, respectively. The term (US Emigrants_i) and $(\text{Knowledge Exposure}_{ik})$ code the number of US emigrants and exposure to US knowledge.[50] Lastly, the variable $\text{I}(t - 1866 = h)$ denotes the number of years since the transatlantic cable was laid down. In equation (10a), the treatment coefficients $\{\beta^h\}$ quantify the effect of the transatlantic cable by comparing districts by the number of US emigrants; in equation (10b), we also leverage variation across sectors and exposure to US innovation.

We report the static versions that conflate pre- and post-treatment years in two periods in columns (1) and (4) of Table 7. 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 expect the transatlantic cable to affect innovation only in districts that were connected to the British domestic network.[51] 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.4. We refer to districts with at least one station as "connected". In columns (2) and (5), we show the estimated effect of the telegraph on connected districts. By comparison, columns (3) and (6) report the estimates for non-connected districts. The results of this exercise are sharp. We estimate a positive effect of the transatlantic cable only for districts connected to the domestic UK network, as expected. We fail to detect any significant effect on non-connected districts.

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. However, 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 7, we thus report the flexible double-differences estimates of model (10a), 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. Our estimates suggest positive treatment

[50] 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.

[51] 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.

effects for those.

Building on Steinwender (2018), we interpret these results as evidence that the increased economic integration between the UK and the US ushered by the transatlantic telegraph was relatively more intense in districts that had previous migration ties with the US. Thus, migration ties facilitate market integration and, indirectly, the diffusion of knowledge across countries. This is in line with evidence by Aleksynska and Peri (2014), who document that migrants promote trade between their origin and their destination countries. We provide additional evidence in this direction in section 6.3.3 and discuss potential additional mechanisms in section 6.4.

6.3.2 Newspaper Mentions of United States Topics in Emigration Districts

Thus far, we have restricted the focus of the analysis to information flows that pertain to innovative knowledge (patents). This section provides evidence that migration ties between the UK and the US generated more general-purpose information flows that did not directly concern innovation. We exploit the vast British Newspaper Archive that contains the digitized contents of thousands of historical British newspapers (for a detailed description of the data, see Appendix section 3.3.2 and Beach and Hanlon, 2022). Ideally, we would like to measure the intensity of US-related information flows into the United Kingdom. We tackle the absence of direct hard data by measuring how frequently US-related news appeared in historical newspapers.

We estimate three sets of regressions:

$$\text{US Mentions}_{i,t} = \alpha_i + \alpha_t + \beta^1 \times \text{US Emigrants}_{i,t} + \varepsilon_{i,t} \tag{11a}$$

$$\text{US State Mentions}_{is,t} = \alpha_i + \alpha_{s\times t} + \beta^2 \times \text{US Emigrants}_{i\to s,t} + \varepsilon_{is,t} \tag{11b}$$

$$\text{US County Mentions}_{ij,t} = \alpha_i + \alpha_{j\times t} + \beta^3 \times \text{US Emigrants}_{i\to j,t} + \varepsilon_{ij,t} \tag{11c}$$

where i, j, s, and t denote a UK district, a US county, a US state, and a decade, respectively. Regression (11a) is run at the district level and leverages the variation of the overall US emigration rate; in regressions (11b) (resp. (11c)), instead, we look at district-by-state (resp. district-by-county) migration flows. We estimate regressions (11) using actual out-migration and the shift-share instrument described in section 4.3.

Table 8 reports the results. Panels A, B, and C respectively display the estimated β^i coefficients of models (11a), (11b), and (11c). In columns (1–3), we report the correlation between measured out-migration flows and newspaper coverage; columns (4–5) report the OLS reduced-form association with the instrument; columns (7–9) display the two-stage least-square estimates. In columns (3), (6), and (9), we restrict the sample to districts with at least one newspaper. We find a strong and positive effect of out-migration on newspaper coverage of general-interest US-related news. Importantly, we always control for time-varying confounding factors at the level of the receiving place, whether it be the country, single states, or single counties. This ensures that the estimates do not reflect shocks in those areas.

We interpret this result as evidence that out-migration generates general—not only innovation—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 loved ones settled. On the other hand, US emigrants could have sponsored local newspapers to cover news of the areas where they had located. In this sense, our estimates may reflect a supply-side factor. What is crucial for this paper is that, notwithstanding the precise underlying mechanism, out-migration ignites cross-country information flows. The return innovation effect is thus one of the possibly many effects of out-migration on countries sending migrants.[52]

6.3.3 Trade-Induced Technology Transfer

The telegraph analysis suggests that market integration, fostered by migration ties, is a major driver of the return innovation result. Here, we provide one additional piece of evidence to support this interpretation.[53] The scope of this exercise is to explore the possible proximate determinants of information diffusion. First, it may be that emigrants themselves are exposed to novel knowledge, which they contribute to spreading. Alternatively, migration ties may facilitate the establishment of trade linkages, which in turn foster cross-border knowledge flows (Aleksynska and Peri, 2014; Ottaviano *et al.*, 2018).

To study this second effect, we explore one specific historical example. In 1930, the United States passed a tariff—the Smoot-Hawley Act —which sharply increased import duties and hampered trade (Eichengreen, 1986). We leverage variation in the tariff increase across technology classes in a difference-in-differences setting.[54] We find that patenting decreases in districts more exposed to technologies that the Act more heavily targeted. This result suggests that migration ties may facilitate international trade, thus contributing to market integration and nurturing the diffusion of novel knowledge. At the same time, however, the magnitude of the estimated effect is modest, especially given the large increase in

[52] A recent literature documents the disparate effects of out-migration on attitudes towards democracy (Spilimbergo, 2009), demand for political change (Karadja and Prawitz, 2019), wages (Dustmann *et al.*, 2015), technology adoption and innovation (Andersson *et al.*, 2022; Coluccia and Spadavecchia, 2022), social norms (Tuccio and Wahba, 2018). Our analysis confirms that migration ties nurture the exchange of information. These flows prompt the cross-border diffusion of novel knowledge but their influence extends well beyond.

[53] We discuss the literature and the technical implementation of the empirical analysis in Appendix section D.1.

[54] We thus exploit between-class variation in exposure to the tariff increase to estimate the effect of trade on knowledge flows. The underlying intuition is that trade in industries that were more heavily targeted by the act suffered relatively more. Aggregate trade volumes support this interpretation.

tariff duties sanctioned by the Act. We thus view trade-induced knowledge diffusion as one, but likely not the only, determinant of information flows.

The literature identifies several margins along which trade can impact innovation. Since the tariff reform was one-sided, it is unlikely that depressed import competition or access to intermediate inputs drive this result (e.g., see Bloom *et al.*, 2016; Autor *et al.*, 2020). We are unable to conclusively disentangle the impact of export opportunities (e.g., see Atkin *et al.*, 2017) from the information access effect of migration ties (Aleksynska and Peri, 2014). The research design, however, leverages cross-district variation in previous US emigration rates. A purely export-driven effect would not reflect variation along this margin and would thus be unlikely to drive the results.

6.4 Potential Additional Mechanisms

In this section, we discuss some potential additional mechanisms that may explain the return innovation result. It is worth stressing that these may operate on top, and not instead, of return migration, cross-border interactions, and economic integration.

6.4.1 Temporary Migrations

When disentangling the possible mechanisms behind the return innovation effect, we contrasted those requiring physical return migration with those that do not. We concluded that physical return is an important determinant of return innovation, but we provide evidence that other mechanisms operate on top of it that do not require physical return. It may be possible that (unobserved) short-term temporary migrations influence the dynamics of innovation in the UK. We cannot observe temporary migrants because we construct migration flows from census data. Censuses are, in turn, only administered to the residing population every ten years. Our data would thus fail to reflect such temporary migration movements. For the reasons above, we cannot quantify the importance of industrial espionage. Episodes of industrial espionage were relatively common during the Industrial Revolution (Harris, 1998).

Temporary migrations and industrial espionage would confound our estimates if such migrations were correlated with observed migration patterns. We believe that it is unlikely that this factor bears relevant quantitative implications. First, the notion of a "temporary migrant" in XIX-century transatlantic migration is unclear. Piore (1980) refers to Southern and Eastern European migrants as temporary because they planned to return to their origin countries at some point. This 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 (Dupont *et al.*, 2017). This suggests that the extent of short-term stays must have been relatively limited. Moreover, Piore (1980) notes that "temporary" migrants were relatively low-skilled and, thus, less likely to operate technology transfer. Industrial espionage, in turn, does not appear to be quantitatively sufficient to generate the return innovation effect that we estimate.

Furthermore, our research designs speak against the temporary migration and the industrial espionage mechanisms. First, the instrumental variable research design largely rules these mechanisms out. Suppose that measured out-migration and unobserved temporary migrations or industrial espionage were correlated across origin districts and destination counties. Our pull instrumental variable randomizes county-level immigration shocks leveraging (conditional) variation in the decade counties were connected to the railway network. While we show that the resulting instrument predicts actual out-migration, it is likely that the source of pull variation is not as active for temporary "business" migrants or spies. Second, for temporary migration or industrial espionage to explain the double and triple differences result, 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.

6.4.2 Monetary Remittances

Along with classical "brain drain" arguments, monetary remittances have been a major subject of empirical investigations in the migration literature (Clemens, 2011). Remittances have been found to contribute only modestly to the economic development of emigration countries. This notwithstanding, it is possible that the inflow of capital through remittances may have sustained increased innovation activity, perhaps by relaxing financial constraints or access to credit (Gorodnichenko and Schnitzer, 2013). It would be more difficult, however, that it would have impacted the *direction* of innovation and, most importantly, that this effect would have been correlated with variation in knowledge exposure.

Disaggregated data on financial remittances, unfortunately, do not exist. We thus remain silent on the possibility that the documented positive effect of out-migration on innovation depends on financial remittances. This capital inflow, however, cannot explain why out-migration influences the direction of innovation unless knowledge *and* monetary remittances go hand in hand. This is a possibility that we cannot explore. It nonetheless highlights that financial and innovation remittances shape innovation in a complementary, rather than mutually exclusive, fashion.

6.5 Discussion

Our results bear potentially far-reaching implications for policy-makers. We show that emigration does not necessarily further underdevelopment or stagnation, as the "brain drain" literature seems to suggest (Docquier and Rapoport, 2012). Instead, out-migration can foster innovation, technology adoption, and diffusion and thus empower long-run economic growth. Rather than focusing on blocking the emigration of skilled individuals, our central recommendation to policy-makers in emigration countries would be to foster cooperation and exchanges between them and the population remaining in the home country. Our results and more recent albeit narrative evidence by Saxenian (1999, 2006) suggest that this approach can yield important and lasting benefits on the economic development of emigration countries.

Consider, as an example, the case of the ongoing diaspora of the Italian scientific community. Italy is often described as an archetypal instance of brain drain (Anelli *et al.*, 2023). In 2020, fifty-five Italian researchers were awarded a European Research Council (ERC) starting grant, possibly the most prestigious award for early-career scholars working in the European Union. Only nineteen ($\approx$ 35%) of them worked in Italian institutions. Italy is the country with the lowest share of ERC-winning researchers working in home institutions among those for which data are reported. This paper sheds new light on the economic contribution of the remaining thirty-six ($\approx$ 65%) on science, innovation, and, ultimately, growth in Italy.[55] Several other countries, however, have been witnessing important episodes of out-migration, ranging from European countries to India, China, or Pakistan. In fact, Saxenian (1999) specifically analyses the Indian and Taiwanese emigration to Silicon Valley. More generally, we study how the entire stock of emigrants influences the dynamics of innovation in their sending countries. In doing so, we enrich our understanding of the consequences of emigration compared to studies that focus on sub-samples of super-skilled emigrants (e.g., see Prato, 2021).

Concerns over the external validity of these results are natural, given the setting we analyze. We nonetheless think that History can inform the scholarly debate and policy-making for two main reasons. First, as previously mentioned, Saxenian (1999, 2006) qualitatively documents similar return innovation effects with respect to the Taiwanese and Indian emigration to the Silicon Valley area. Second, we provide evidence that the UK emigration to the US in the XIX century largely resembles, *mutatis mutandis*, migration between European countries and the United States during the XXI. Compared to the rest of the English population, migrants were positively selected. They were similarly more likely to be employed in skilled occupations than the average native and to live in urban centers. These patterns suggest that a cautious comparison between historical and contemporary migration episodes can yield important insights for policy-makers and scholars.

7 Conclusions

The diffusion of innovation across countries is a major factor shaping long-run economic development. In this paper, we argue that international migrations generate knowledge flows that contribute to the diffusion of innovation into emigration countries. This result—which we label "return innovation"—offers a more nuanced view of the effect of emigration on innovation compared to the traditional "brain drain" hypothesis, which interprets it as a depletion of the human capital of countries sending migrants.

[55] These figures are the result of authors' calculations over data released by the ERC, available at this link. We mention the Italian case because Italy underwent a major loss of skilled human capital in recent years. Between 2008 and 2016, more than 500,000 Italians emigrated. Comparable high-skilled emigration, however, concerns several other developed economies (United Nations and OECD, 2013).

To study this question, this paper explores the English and Welsh mass migration to the United States between 1870 and 1940. To construct a granular measure of transatlantic migration flows, we link census records of British immigrants in the US to the individual-level UK population census. The resulting dataset allows us to observe the universe of (male) British immigrants after they migrate to the US and before leaving the UK. We complement this with newly digitized patent data covering the universe of patents in England and Wales. On top of these unique, high-quality data, the absence of stringent international intellectual property protection and active migration policies provide two prominent appealing features of this historical setting compared to contemporary scenarios.

We provide novel, causal evidence that exposure to US innovation through migration ties contributes to the diffusion of US technology in Britain. Innovation activity in the UK shifts to sectors that emigrants are most exposed to in the US. To address endogeneity concerns arising from the assortative matching of British immigrants in the US, we develop a new shift-share instrument that exploits the conditional timing of connection to the railway network to randomize emigration across counties. Moreover, we implement a triple-differences research design that leverages variation across counties and technology classes. We can thus document a causal link between exposure to foreign knowledge through migration ties and innovation activity. By looking at the textual similarity between UK and US patents, we document that exposure to US knowledge stimulates cross-border technology transfer while also nurturing original innovation in Britain.

What drives the return innovation effect? We find that the physical return of migrants is a crucial driver as it explains approximately half of the return innovation effect. Exploiting the granular nature of our data, we provide evidence that social interactions between the emigrants and their communities of origin represent another important channel through which technologies diffuse into the emigration country, even if emigrants do not return. Additionally, we document that migration ties promote the diffusion of knowledge by fostering information flows and cross-border market integration. Leveraging a comprehensive repository of historical newspapers, we show that migration linkages further promote general-purpose information flows by increasing attention to US-related news in UK media outlets.

The historical evidence suggests that the British mass migration to the United States may be comparable to present-day cross-border movements between developed countries. As the number of international migrants has been steadily rising over the past decades, the role of human mobility as a driver of knowledge and information diffusion across countries in a globalized world economy bears quantitatively relevant implications. History can thus inform the scholarly literature and policymakers on the complex relationship between out-migration, innovation, and, ultimately, long-run economic growth.

References

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

ABRAMITZKY, R., L. BOUSTAN and K. ERIKSSON (2020). "Do Immigrants Assimilate More Slowly Today Than in the Past?" *American Economic Review: Insights*, 2(1): 125–41.

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?" In "AEA Papers and Proceedings," volume 113, pp. 1–28.

ACEMOGLU, D. and T. LENSMAN (2023). "Technology Paradigms, Lock-in, and Economic Growth." *Mimeo*.

AGHION, P., A. DECHEZLEPRÊTRE, D. HEMOUS, R. MARTIN and J. VAN REENEN (2016). "Carbon Taxes, Path Dependency, and Directed Technical Change: Evidence from the Auto Industry." *Journal of Political Economy*, 124(1): 1–51.

AGRAWAL, A., D. KAPUR, J. MCHALE and A. OETTL (2011). "Brain Drain or Brain Bank? The Impact of Skilled Emigration on Poor-Country Innovation." *Journal of Urban Economics*, 69(1): 43–55.

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*.

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*.

ARTHUR, W. B. (1989). "Competing Technologies, Increasing Returns, and Lock-in by Historical Events." *The Economic Journal*, 99(394): 116–131.

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., A. K. KHANDELWAL and A. OSMAN (2017). "Exporting and Firm Performance: Evidence from a Randomized Experiment." *The Quarterly Journal of Economics*, 132(2): 551–615.

AUTOR, D., D. DORN, G. H. HANSON, G. PISANO and P. SHU (2020). "Foreign Competition and Domestic Innovation: Evidence from US Patents." *American Economic Review: Insights*, 2(3): 357–74.

AZOULAY, P., B. F. JONES, J. D. KIM and J. MIRANDA (2022). "Immigration and Entrepreneurship in the United States." *American Economic Review: Insights*, 4(1): 71–88.

BAHAR, D., P. CHOUDHURY, J. SAPPENFIELD and S. SIGNORELLI (2022a). "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.

——— (2022b). "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.

BAILEY, M. J., C. COLE, M. HENDERSON and C. MASSEY (2020). "How Well Do Automated Linking Methods Perform? Lessons from US Historical Data." *Journal of Economic Literature*, 58(4): 997–1044.

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.

BAZZI, S., M. FISZBEIN and M. GEBRESILASSE (2020). "Frontier Culture: The Roots and Persistence of "Rugged Individualism" in the United States." *Econometrica*, 88(6): 2329–2368.

BEACH, B. and W. W. HANLON (2022). "Historical Newspaper Data: A Researcher's Guide and Toolkit." *NBER Working Paper*, (No. w30135).

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.

Benhabib, J., J. Perla and C. Tonetti (2021). “Reconciling Models of Diffusion and Innovation: A Theory of the Productivity Distribution and Technology Frontier.” *Econometrica*, 89(5): 2261–2301.

Bergeaud, A. and C. Verluise (2022). “A New Dataset to Study a Century of Innovation in Europe and in the US.” *Working Paper*.

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

Berkes, E., D. M. Coluccia, G. Dossi and M. P. Squicciarini (2023). “Dealing with Adversity: Religiosity or Science? Evidence from the Great Influenza Pandemic.” *Working Paper*.

Bernstein, S., R. Diamond, A. Jiranaphawiboon, T. McQuade and B. Pousada (2022). “The Contribution of High-Skilled Immigrants to Innovation in the United States.” *NBER Working Paper*, (No. w30797).

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

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

Bloom, N., M. Draca and J. Van Reenen (2016). “Trade Induced Technical Change? The Impact of Chinese Imports on Innovation, IT and Productivity.” *The Review of Economic Studies*, 83(1): 87–117.

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

Borusyak, K., X. Jaravel and J. Spiess (2021). “Revisiting Event Study Designs: Robust and Efficient Estimation.” *Working Paper*.

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

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).

Callaway, B., A. Goodman-Bacon and P. H. Sant’Anna (2021). “Difference-in-differences with a continuous treatment.”

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.

Chen, J. and J. Roth (2022). “Log-like? ATEs Defined with Zero Outcomes are (Arbitrarily) Scale-dependent.” *Working Paper*.

Choudhury, P. (2016). “Return Migration and Geography of Innovation in MNEs: A Natural Experiment of Knowledge Production by Local Workers Reporting to Return Migrants.” *Journal of Economic Geography*, 16(3): 585–610.

Clemens, M. A. (2011). “Economics and Emigration: Trillion-dollar Bills on the Sidewalk?” *Journal of Economic Perspectives*, 25(3): 83–106.

Coluccia, D. M. and L. Spadavecchia (2022). “Emigration Restrictions and Economic Development: Evidence from the Italian Mass Migration to the United States." *Working Paper*.

COMIN, D. and B. HOBIJN (2011). "Technology Diffusion and Postwar Growth." *NBER Macroeconomics Annual*, 25(1): 209–246.

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

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

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

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.

DE CHAISEMARTIN, C. and X. D'HAULTFŒUILLE (2022). "Difference-in-Differences Estimators of Intertemporal Treatment Effects." *NBER Working Paper*, (No. w29873).

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

DOSI, G. (1982). "Technological Paradigms and Technological Trajectories: A Suggested Interpretation of the Determinants and Directions of Technical Change." *Research Policy*, 11(3): 147–162.

DUPONT, B., D. KEELING and T. WEISS (2017). "First Cabin Fares from New York to the British Isles, 1826–1914." In "Research in Economic History," Emerald Publishing Limited.

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.

DUSTMANN, C. and J.-S. GÖRLACH (2016). "The Economics of Temporary Migrations." *Journal of Economic Literature*, 54(1): 98–136.

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.

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).

EICHENGREEN, B. (1986). "The Political Economy of the Smoot-Hawley Tariff." *NBER Working Paper*, (No. w2001).

EINIÖ, E., J. FENG and X. JARAVEL (2022). "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.

FACKLER, T. A., Y. GIESING and N. LAURENTSYEVA (2020). "Knowledge Remittances: Does Emigration Foster Innovation?" *Research Policy*, 49(9): 103863.

FISCHER, D. H. (1989). *Albion's seed: Four British folkways in America.* Oxford University Press.

FURER, H. B. (1972). *The British in America: 1578-1970*. Oceana Publications.

GANGULI, I. (2015). "Immigration and Ideas: What Did Russian Scientists "Bring" to the United States?" *Journal of Labor Economics*, 33(S1): S257–S288.

GERSCHENKRON, A. (1962). *Economic Backwardness in Historical Perspective: A Book of Essays*. Cambridge (MA): Harvard University Press.

GIBSON, J. and D. MCKENZIE (2011). "Eight Questions About Brain Drain." *Journal of Economic Perspectives*, 25(3): 107–28.

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

GIULIANO, P. and M. TABELLINI (2020). "The Seeds of Ideology: Historical Immigration and Political Preferences in the United States." *NBER Working Paper*, (No. w27238).

GOLDIN, C. (1994). "The Political Economy of Immigration Restriction in the United States, 1890 to 1921." In "The Regulated Economy: A Historical Approach to Political Economy," pp. 223–258. University of Chicago Press.

GOMME, A. A. (1948). *Patents of invention: origin and growth of the patent system in Britain*. British council.

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.

GROSS, D. P. and B. N. SAMPAT (2022). "Crisis Innovation Policy From World War II to COVID-19." *Entrepreneurship and Innovation Policy and the Economy*, 1(1): 135–181.

GRUBER, J. (1994). "The Incidence of Mandated Maternity Benefits." *American Economic Review*, 84(3): 622–641.

HABAKKUK, H. J. (1962). *American and British Technology in the Nineteenth Century: the Search for Labour Saving Inventions*. Cambridge (MA): Cambridge University 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.

Harris, J. (1998). *Industrial Espionage and Technology Transfer: Britain and France in the Eighteenth Century*. Ashgate Publishing Limited, Aldershot.

Hicks, J. (1932). *The Theory of Wages*. Macmillan, London (UK).

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.

Jones, C. I. (1995). "R&D-based Models of Economic Growth." *Journal of Political Economy*, 103(4): 759–784.

Kapur, D. (2014). "Political Effects of International Migration." *Annual Review of Political Science*, 17: 479–502.

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, S. P. and W. Kerr (2020). "Immigrant Entrepreneurship in America: Evidence from the Survey of Business Owners 2007 & 2012." *Research Policy*, 49(3): 103918.

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

Khan, B. Z. (2020). *Inventing Ideas: Patents, Prizes, and the Knoweldge Economy*. Oxford University Press, USA.

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.

Mokyr, J. (1998). "The Second Industrial Revolution, 1870-1914." In V. Castronovo, ed., "Storia dell'Economia Mondiale," pp. 219–245. Rome (Italy): Laterza.

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

Moscona, J. and K. A. Sastry (2022). “Does Directed Innovation Mitigate Climate Damage? Evidence from US Agriculture.” *The Quarterly Journal of Economics*, Forthcoming.

Moser, P. (2012). “Innovation without patents: Evidence from World’s Fairs.” *The Journal of Law and Economics*, 55(1): 43–74.

——— (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.

Nicholas, T. (2011). “Cheaper patents.” *Research Policy*, 40(2): 325–339.

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.

Olden, A. and J. Møen (2022). “The Triple Difference Estimator.” *The Econometrics Journal*, 25(3): 531–553.

Ottaviano, G. I. P., G. Peri and G. C. Wright (2018). “Immigration, Trade and Productivity in Services: Evidence from UK Firms.” *Journal of International Economics*, 112: 88–108.

Ottinger, S. (2020). “Immigrants, Industries, and Path Dependence.” *Working Paper*.

Ottinger, S. and L. Rosenberger (2023). “The American Origin of the French Revolution.” *Working Paper*.

Pauly, S. and F. Stipanicic (2021). “The Creation and Diffusion of Knowledge: Evidence from the Jet Age.” *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. (2021). “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.

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*.

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.

SHI, D., W. LIU and Y. WANG (2023). “Has China’s Young Thousand Talents program been successful in recruiting and nurturing top-caliber scientists?” *Science*, 379(6627): 62–65.

SOKOLOFF, K. L. and B. Z. KHAN (1990). “The Democratization of Invention During Early Industrialization: Evidence from the United States, 1790–1846.” *The Journal of Economic History*, 50(2): 363–378.

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.

THISTLETHWAITE, F. (1958). “The Atlantic Migration of the Pottery Industry.” *The Economic History Review*, 11(2): 264–278.

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

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.

UNITED NATIONS and OECD (2013). “World Migration in Figures. A Joint Contribution by UN-DESA and the OECD to the United Nations High-Level Dialogue on Migration and Development.”

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

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.

Tables

	(1) Observations	(2) Mean	(3) Std. Dev.	(4) Min.	(5) Max.
Panel A. Innovation					
Total Patents	5489	225.826	826.432	1	19789
Electricity Patents	5489	23.565	200.755	0	9430
Instruments Patents	5489	17.69	80.2	0	1850
Personal Articles & Furniture	5489	20.136	73.812	0	1548
Ships & Aeronautics	5489	16.463	55.062	0	1152
Transportation	5489	20.024	74.479	0	1923
Panel B. Emigration					
N. of US Emigrants	3779	61.765	91.36	0.303	1073.998
N. of Return US Emigrants	2494	35.342	52.202	0.064	730
Panel C. Census Tracts					
Population (1,000s)	3773	42.165	54.973	0.092	703.559
Share of Males (%)	3773	47.645	2.586	36.112	62.686
Share of Manufacture Empl. (%)	3773	13.213	6.306	2.569	42.723
Share of Agriculture Empl. (%)	3773	14.43	6.889	1.454	32.914
Share of Transportation Empl. (%)	3773	2.578	1.272	0	13.857
Share of Liberal Professions (%)	3773	1.679	0.65	0.43	6.873
Share of Public Servants (%)	3773	0.897	1.427	0	24.498
Panel D. Individual-Level Panel					
Share of Inventors	471013	0.009	0.094	0	1
N. of Patents	471013	0.018	0.356	0	87
N. of Patents if Inventor	4210	1.993	3.205	1	87
N. of Neighborhood Emigrants	471013	13.62	43.338	0	756
N. of Non-Return Neighborhood Emigrants	471013	12.979	40.888	0	512

TABLE 1. DESCRIPTIVE STATISTICS OF SELECTED VARIABLES

Notes. This table displays summary descriptive statistics for a subset of the variables in the dataset. In Panels A, B, and C, variables are observed at the district level and at a decade frequency. In Panel D, the statistics are computed for individuals observed for twenty years around the 1891 and 1911 census years. An individual is labeled an inventor if they obtain at least one patent over this period. Panel A reports statistics for the top five most frequent technological classes. In Panels B and C, the underlying data are cross-walked to 1900 district borders.

◄ Back: Data

	Ordinary Least Squares			Reduced Form			Two-Stages Least-Squares		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. Dependent variable: Number of patents									
Knowledge Exposure$_t$	1.342***			0.037***			1.224***		
	(0.143)			(0.007)			(0.195)		
Knowledge Exposure$_{t-1}$		0.909***			0.015***			0.488**	
		(0.145)			(0.005)			(0.190)	
Knowledge Exposure$_{t-2}$			0.379***			-0.012			-0.398
			(0.112)			(0.014)			(0.478)
Mean Dep. Var.	10.392	13.345	15.256	8.706	12.045	15.314	8.708	12.049	15.319
Std. Beta Coef.	0.299	0.148	0.050	0.075	0.022	-0.013	0.296	0.088	-0.053
K-P F-stat							109.826	109.826	109.826
Panel B. Dependent Variable: Patents per capita ($\times$ 10,000)									
Knowledge Exposure$_t$	0.178***			0.004***			0.146***		
	(0.020)			(0.001)			(0.027)		
Knowledge Exposure$_{t-1}$		0.092***			0.002***			0.078***	
		(0.018)			(0.001)			(0.024)	
Knowledge Exposure$_{t-2}$			0.049***			0.000			0.001
			(0.015)			(0.001)			(0.043)
Mean Dep. Var.	2.066	2.629	2.973	1.748	2.345	2.980	1.747	2.346	2.980
Std. Beta Coef.	0.124	0.054	0.023	0.023	0.011	0.000	0.093	0.046	0.000
K-P F-stat							107.825	107.825	107.825
District-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Technology Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. of District-Class	11268	11268	11268	11214	11214	11214	11214	11214	11214
N. of Observations	67549	67549	56295	56070	56070	56070	56047	56047	56047

TABLE 2. EFFECT OF EXPOSURE TO US TECHNOLOGY ON INNOVATION IN GREAT BRITAIN

Notes. This table displays the association between innovation and exposure to US knowledge. The unit of observation is a district-technology class pair, observed at a decade frequency between 1880 and 1939. The main explanatory variable is knowledge exposure. In Panel A, the dependent variable is the number of patents; in Panel B, the dependent variable is the number of patents normalized by district-level population in 1880 and multiplied by 10,000 for readability. In columns (1–3), we estimate the OLS correlation with the observed measure of knowledge exposure; in columns (4–6), we estimate the reduced-form association with the railway-based instrument of knowledge exposure through OLS; columns (7–9) report the two-stage least-squares estimate. Each model includes district-by-decade and district-by-technology class fixed effects. Standard errors are reported in parentheses and are clustered at the district level.

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

◄ Back: Results

	Synthetic Shocks				Great Influenza Pandemic Shock			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Full Sample	No London	No Lancs	No S/W	Full Sample	No London	No Lancs	No S/W
Synth. Shock × Post × Emigrants	0.434***	0.277***	0.578***	0.420***				
	(0.121)	(0.082)	(0.125)	(0.127)				
Pharma × Post × Emigrants					0.613***	0.417***	0.678***	0.461***
					(0.164)	(0.140)	(0.172)	(0.156)
District-by-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-by-Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class-by-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Units	10029	9697	8760	9547	10727	10217	9384	10047
Number of Observations	393046	382153	343450	375850	429080	408680	375360	401880
Mean Dep. Var.	1.361	1.029	1.263	1.276	0.725	0.532	0.682	0.668

TABLE 3. EFFECT OF EXPOSURE TO US INNOVATION SHOCKS ON UK INNOVATION

Notes. This table displays the effect of US innovation shocks on innovation activity in the UK. The unit of observation is a district-technology class pair observed at a yearly frequency between 1900 and 1939. The dependent variable is the number of patents. In columns (1–4), the independent variable is an indicator that, for a given district–technology, returns value one after a synthetic innovation shock in that technology class is observed in at least one county where the district has above-average out-migration. A synthetic innovation shock is observed whenever the residualized number of patents observed in the country is in the top 0.5% of the overall distribution. In columns (5–8), the independent variable is an indicator that returns value one for pharmaceutical patents only and only if emigration from the observed district to counties in the top quartile of the influenza mortality distribution is in the top quartile across districts. Both models should thus be interpreted as triple-difference designs. Since models in columns (1–4) are staggered designs, we estimate them using the imputation estimator developed by Borusyak *et al.* (2021). In columns (2) and (6), we drop districts in the London area; in columns (3) and (7), we exclude districts in the Lancashire area; in columns (4) and (8), we drop districts in the South-West area. Excluded regions are the first three in terms of patents granted. All models include district-by-year, district-by-technology class, and technology class-by-year fixed effects; standard errors, clustered two-way by district and technology class, are shown in parentheses.

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

◄ Back: Results

	Dep. Var.: Number of Patents					
	(1)	(2)	(3)	(4)	(5)	(6)
Knowledge Exposure$_t$	1.385***	1.347***	1.301***			0.697***
	(0.306)	(0.389)	(0.319)			(0.092)
Return Knowledge Exposure$_t$	0.306***	0.254***	0.311***			0.273**
	(0.075)	(0.084)	(0.084)			(0.108)
Knowledge Exposure$_{t-1}$				1.833***		0.060
				(0.489)		(0.444)
Return Knowledge Exposure$_{t-1}$				0.156***		0.170*
				(0.028)		(0.084)
Knowledge Exposure$_{t-2}$					0.416	0.139
					(0.243)	(0.436)
Return Knowledge Exposure$_{t-2}$					0.218	0.192*
					(0.131)	(0.099)
District-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
Technology Class FE	Yes	–	–	Yes	Yes	Yes
Class-Decade FE	No	Yes	Yes	No	No	No
District-Class FE	No	No	Yes	No	No	No
N. of District-Class	11376	11376	11376	11375	11371	11340
N. of Observations	45464	45464	45464	34114	22742	22680
Mean Dep. Var.	9.869	9.869	9.869	12.175	15.871	15.907
Std. Beta (KE)	0.211	0.205	0.198	0.200	0.041	
Std. Beta (Return KE)	0.287	0.238	0.291	0.136	0.054	

TABLE 4. ASSOCIATION BETWEEN UK INNOVATION AND EXPOSURE TO US TECHNOLOGY THROUGH OVERALL AND RETURN EMIGRATION

Notes. This table reports the association between innovation and the baseline measure of knowledge exposure, accounting for return knowledge exposure. The unit of observation is a district-technology class pair, observed at a decade frequency between 1880 and 1920. The dependent variable is the number of patents by district-technology decade. Return knowledge exposure is constructed by interacting county-level specialization with district-county return migration flows analogously to the baseline knowledge exposure measure. In columns (1) and (4–6), we include district-by-decade and technology class fixed effects. In column (2), we add technology-by-decade fixed effects; the specification in column (3) is saturated. Standard errors, clustered at the district level, are displayed in parentheses. The Table reports the standardized beta coefficient of both the baseline knowledge exposure term and the return knowledge exposure term.

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

◄ Back: Mechanisms

	N. of Patents				I(N. of Patents > 0)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post × Relative US Emigrant	0.059***				0.035***			
	(0.006)				(0.003)			
Post × Relative Return US Emigrant		0.335***		0.332***		0.126***		0.124***
		(0.068)		(0.068)		(0.031)		(0.031)
Post × Relative Non-Return US Emigrant			0.060***	0.059***			0.035***	0.034***
			(0.006)	(0.006)			(0.003)	(0.003)
County-Surname FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Surname-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. of Surname-County	416735	416735	416735	416735	416735	416735	416735	416735
N. of Observations	25004100	25004100	25004100	25004100	25004100	25004100	25004100	25004100
Mean Dep. Var.	0.083	0.083	0.083	0.083	0.054	0.054	0.054	0.054

Table 5. Effect of Family Member Emigration on Innovation Produced by Relatives of the Emigrant in the UK

Notes. This table reports the effect of transatlantic emigration on innovation by inventors with the same surname as the emigrant. The unit of observation is a surname-county couple, observed at a year frequency between 1870 and 1929. The dependent variable in columns (1–4) is the number of patents granted to inventors with a given surname in a given county and year; the dependent variable in columns (5–8) is a categorical variable that takes value one if the number of patents is strictly positive. In columns (1) and (5), the treatment takes value one after at least one individual from a given county and with a given surname emigrates to the US, and zero otherwise; in columns (2) and (6), we restrict to emigrants that at some point return; in columns (3) and (7), we restrict to emigrants that never return; in columns (4) and (8) we horse-race the two latter treatments. Each regression includes county-by-surname, surname-by-year, and county-by-year fixed effects. Standard errors, clustered at the surname level, are displayed in parentheses.

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

◄ Back: Mechanisms

	Baseline Sample				Dropping Individuals in...		
	(1)	(2)	(3)	(4)	(5) London	(6) Lancashire	(7) South-West
Panel A. All Emigrants							
Neighborhood Emigrant × Post	0.167***	0.180***	0.170***	16.208***	0.130**	0.180***	0.198***
	(0.053)	(0.055)	(0.056)	(5.758)	(0.061)	(0.055)	(0.061)
Std. Beta Coef.	0.022	0.024	0.023	0.211	0.018	0.025	0.025
Panel B. Only Non-Return Emigrants							
Non-Return Neighborhood Emigrant × Post	0.165***	0.189***	0.167***	15.749***	0.108*	0.183***	0.211***
	(0.054)	(0.056)	(0.057)	(5.987)	(0.060)	(0.056)	(0.062)
Std. Beta Coef.	0.021	0.024	0.021	0.196	0.014	0.024	0.026
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	–	Yes	Yes	Yes	Yes	Yes
Parish × Year FE	No	Yes	No	No	No	No	No
Matching	No	No	Yes	No	No	No	No
Sample	Full	Full	Full	Inventors	Full	Full	Full
N. of Individuals	473112	473112	469585	4224	410327	422230	352064
N. of Observations	9462240	9419787	9391700	84480	8206540	8444600	7041280
Mean Dep. Var.	0.890	0.892	0.893	99.716	0.794	0.836	0.893
S.D. Dep. Var.	40.291	40.324	40.351	414.695	37.439	39.126	41.333

TABLE 6. EFFECT OF EMIGRATION ON INNOVATION PRODUCED BY FORMER NEIGHBORS OF THE EMIGRANT IN THE UK

Notes. This table reports the effect of neighborhood out-migration on innovation. The units of observation are individuals who are observed yearly between 1900 and 1920. In columns (1–3) and (5–7), the sample consists of the universe of males who did not emigrate over the period and that were at least 18 years old in 1900; in columns (4) and (8), we restrict the sample to inventors. The dependent variable is the number of patents obtained annually. In columns (1–4), the sample consists of individuals residing in all England and Wales divisions; in columns (5–7), we exclude the top tree-patents producing areas: London, Lancashire, and the South-West. In Panel A, the independent variable is an indicator that, for a given individual, returns value one after at least one person that was living in the same neighborhood as the individual migrates to the United States; in Panel B, we restrict to emigrants that never return in the period of observation. In this context, "neighborhood" refers to the same street, square, or similar. We explore an alternative distance-based definition in Appendix Table E.15. Each model includes individual and—at least—year fixed effects; in column (2), we include parish-by-year fixed effects; in column (3), individuals are weighted by their coarsened exact matching weight. The estimates are obtained using the method discussed in Borusyak *et al.* (2021) to account for the staggered roll-out of the treatment across individuals. Standard errors, clustered at the district level, are reported in parentheses.

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

◄ Back: Mechanisms

	Double Differences			Triple Differences		
	(1)	(2)	(3)	(4)	(5)	(6)
	All	Connected	Not Connected	All	Connected	Not Connected
Post × Emigrants	1.345***	1.639***	-0.083			
	(0.451)	(0.559)	(0.097)			
Post × Knowledge Exposure				0.027**	0.027**	-0.003
				(0.010)	(0.011)	(0.005)
District FE	Yes	Yes	Yes	–	–	–
Class FE	Yes	Yes	Yes	–	–	–
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
District-Class FE	–	–	–	Yes	Yes	Yes
N. of District-Class	631	463	168	631	463	168
N. of Observations	10096	7408	2688	181728	133344	48384
Mean Dep. Var.	5.610	7.241	1.115	0.312	0.402	0.062
Std. Beta Coef.	0.114	0.125	-0.039	0.035	0.034	-0.007

TABLE 7. THE TRANSATLANTIC TELEGRAPH CABLE AND INNOVATION IN THE UK

Notes. This table displays the estimated effect of the connection of the US and UK telegraph lines on innovation in the UK. The units of observation are districts in columns (1–3) and district-technology class pairs in columns (4–6). Units are observed yearly between 1860 and 1875. The dependent variable is the total number of patents granted. In columns (1–3), the independent variable is an interaction between the—time-invariant—number of US emigrants in the 1870s and an indicator variable that returns value one after the transatlantic cable successfully connected the UK and the US in 1866, and zero otherwise; in columns (4–6) the treatment interacts—time-invariant—knowledge exposure in the 1870s with the same posttreatment indicator. In columns (1) and (4), the sample includes all districts; in columns (2) and (5) (resp. 3 and 6), we restrict the estimation to districts that were (resp. were not) connected to the domestic UK telegraph system. Models (3) and (6) should be interpreted as placebo exercises. Regressions include fixed effects for district and year in columns (1–3) and district-by-class and year in columns (4–5). Standard errors, clustered at the district level, are reported in parentheses.

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

◄ Back: Mechanisms

	Dependent Variable: Number of Newspaper Mentions								
	OLS			Reduced Form			2SLS		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. US-Wide Coverage									
US Emigrants	6.753***	6.632***	7.207***				24.396***	24.228***	25.061***
	(0.958)	(1.006)	(0.600)				(1.570)	(1.611)	(0.912)
US $\widehat{\text{Emigrants}}$				1.451***	1.440***	1.501***			
				(0.121)	(0.124)	(0.078)			
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean Dep. Var.	1017.635	1017.635	2276.989	1017.635	1017.635	2276.989	1017.635	1017.635	2276.989
Panel B. State-Wide Coverage									
US Emigrants	1.050***	1.049***	1.103***				10.060***	10.061***	10.091***
	(0.095)	(0.096)	(0.052)				(0.428)	(0.430)	(0.458)
US $\widehat{\text{Emigrants}}$				0.038***	0.038***	0.039***			
				(0.001)	(0.001)	(0.001)			
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean Dep. Var.	57.486	57.486	127.984	64.672	64.672	136.660	64.672	64.672	136.660
Panel C. County-Wide Coverage									
US Emigrants	1.120***	1.120***	1.217***				4.861***	4.863***	5.130***
	(0.148)	(0.148)	(0.079)				(0.471)	(0.460)	(0.291)
US $\widehat{\text{Emigrants}}$				0.055***	0.055***	0.058***			
				(0.006)	(0.005)	(0.004)			
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean Dep. Var.	0.003	0.003	0.007	0.004	0.004	0.008	0.004	0.004	0.008
N. of Newspapers	No	Yes	No	No	Yes	No	No	Yes	No
Districts in Sample	All	All	w/News	All	All	w/News	All	All	w/News
N. of Districts	602	602	321	602	602	321	602	602	321

TABLE 8. EFFECT OF US EMIGRATION ON NEWSPAPER COVERAGE OF US-RELATED NEWS

Notes. This table displays the effect of out-migration on newspaper coverage of emigrants' destinations. The observation unit is: in Panel A, a district; in Panel B, a district-US state pair; in Panel C, a district-US county pair. Units are observed at a decade frequency between 1880 and 1930. The dependent variable is the number of articles mentioned: in Panel A, "United States"; in Panel B, US states; in Panel C, US counties. The independent variable is: in Panel A, the number of US emigrants; in Panel B, the district-state emigrants; in Panel C, the district-county emigrants. Models (1–3) estimate the model through OLS; models (4–5) report the reduced-form association between mentions and the out-migration instrument; models (7–9) report the two-stage least squares estimates. Regressions include district fixed effects and: in Panel A, decade fixed effects; in Panel B: state-by-decade fixed effects; in Panel C: county-by-decade fixed effects. Standard errors, clustered at the district level, are reported in parentheses.

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

◄ Back: Mechanisms

Figures

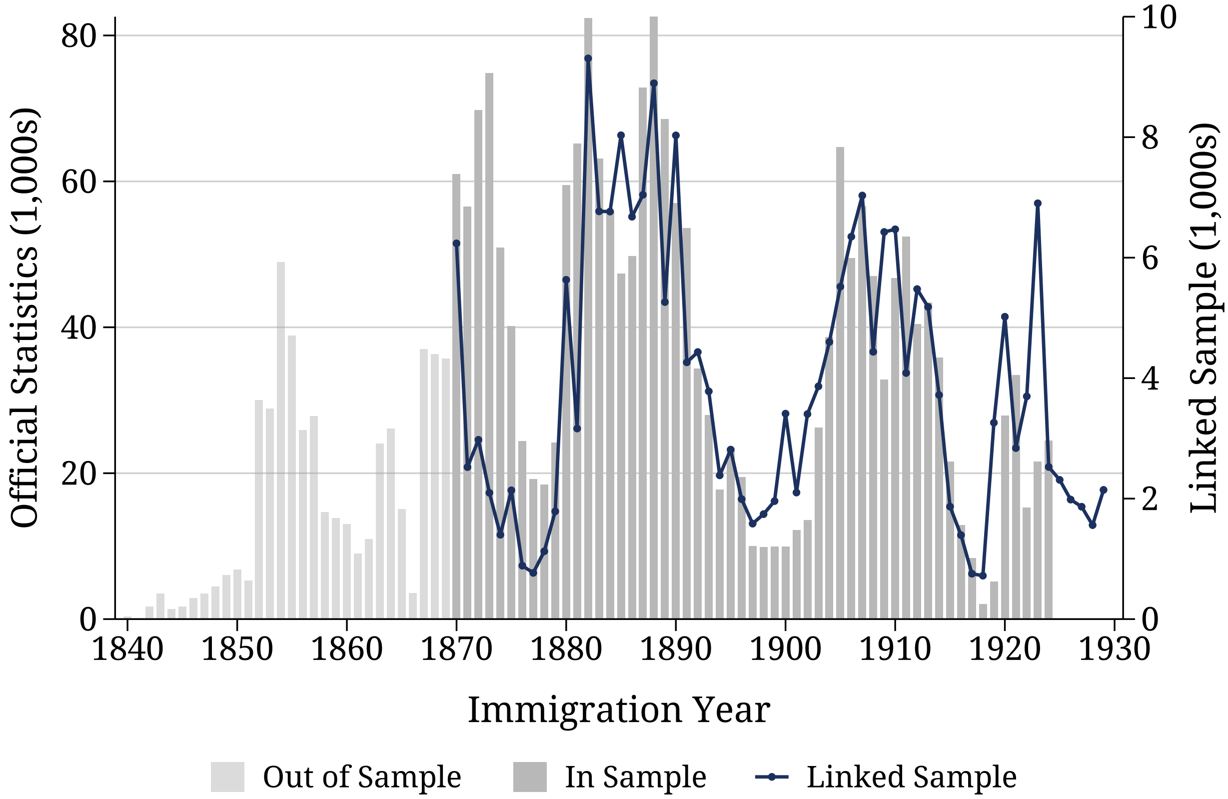


FIGURE 1. NUMBER OF US IMMIGRANTS FROM THE UK AND LINKED US-UK MIGRANTS, 1840–1930

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 developed in this paper. The light gray bars display the total inflow of English and Welsh immigrants in the US over the period 1840–1870, i.e., out of the period we study. The darker gray bars report the same figure for 1870–1924. 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. Figures are in thousand units.

◄ Back: Data

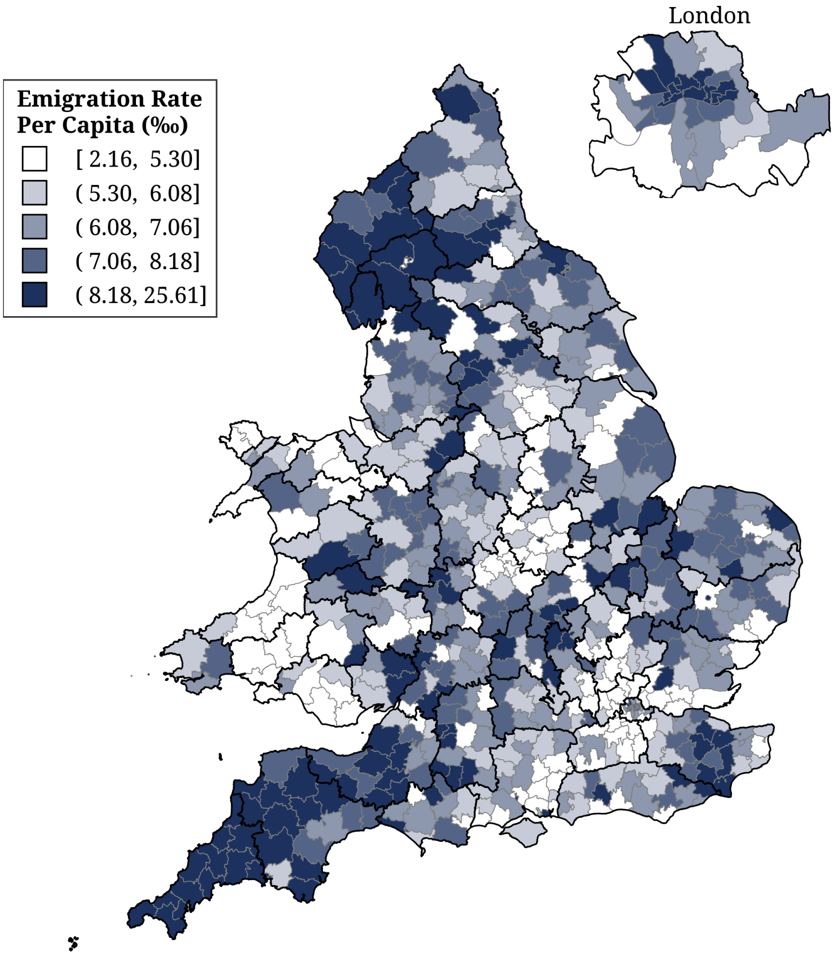
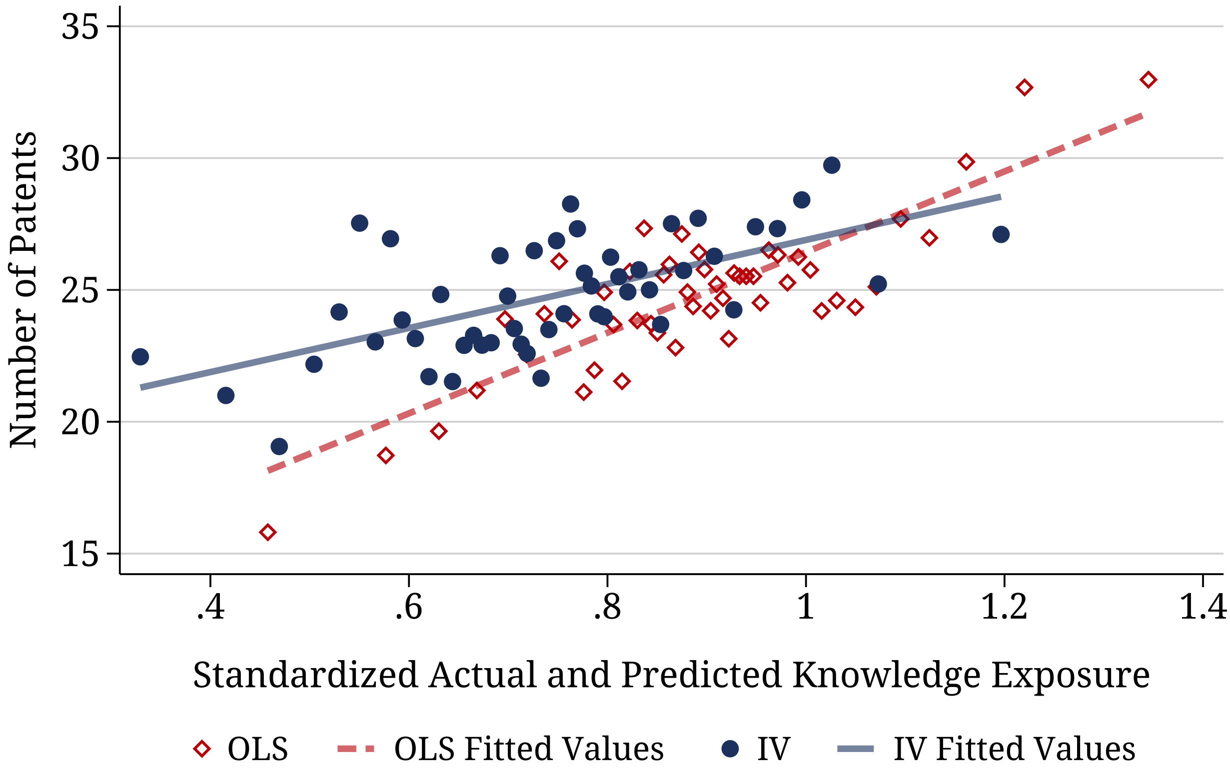


FIGURE 2. SPATIAL DISTRIBUTION OF US MIGRANTS ACROSS BRITISH DISTRICTS

Notes. This figure reports the spatial distribution of emigrants across English and Welsh districts over the period 1870–1930. Data are from the matched emigrants' sample. The total number of emigrants over the period is normalized by district population in 1900 and is reported in ‰ units. Districts are displayed at 1900 historical borders, and the emigrant population is cross-walked to consistent borders as described in A.1. Lighter to darker blues indicate higher emigration rates.

◄ Back: Data



OLS Coefficient = 10.675 (Clust. Std. Err. = 1.138). R^2 = 0.707.
IV Coefficient = 2.174 (Clust. Std. Err. = 0.395). R^2 = 0.731.

FIGURE 3. OLS AND REDUCED-FORM ASSOCIATION BETWEEN KNOWLEDGE EXPOSURE AND INNOVATION

Notes. This figure is a binned scatter plot of the OLS (in red) and the IV reduced-form (in blue) association between knowledge exposure and innovation. The unit of observation is a district-technology class observed at a yearly frequency between 1880 and 1939. The graph partials out district-by-decade and district-by-technology class fixed effects. The red dots report the correlation between actual knowledge exposure and the number of patents; the blue dots report the reduced-form association between the instrument for knowledge exposure and the number of patents. We report in note the regression coefficient for both the OLS and the IV regression with their standard errors clustered at the district level and the R^2.

◄ Back: Results

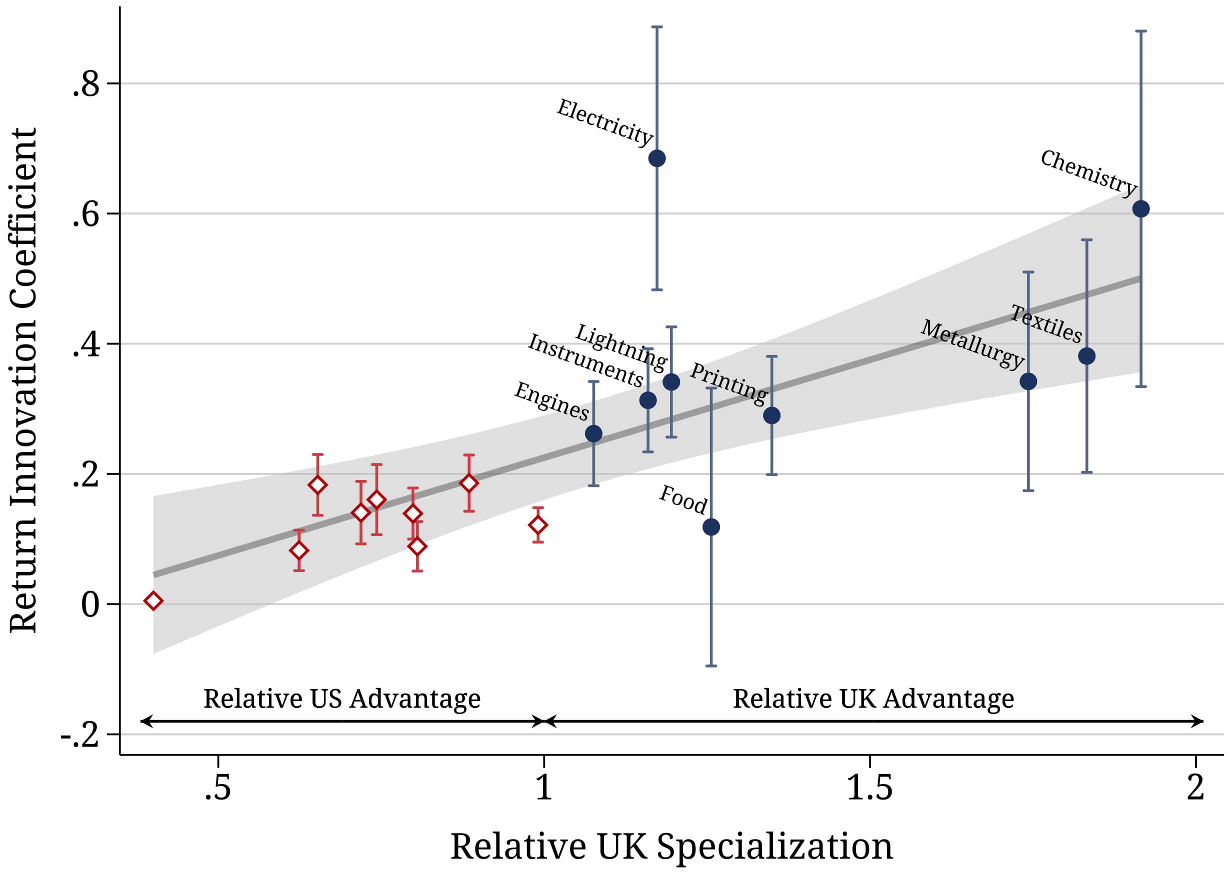
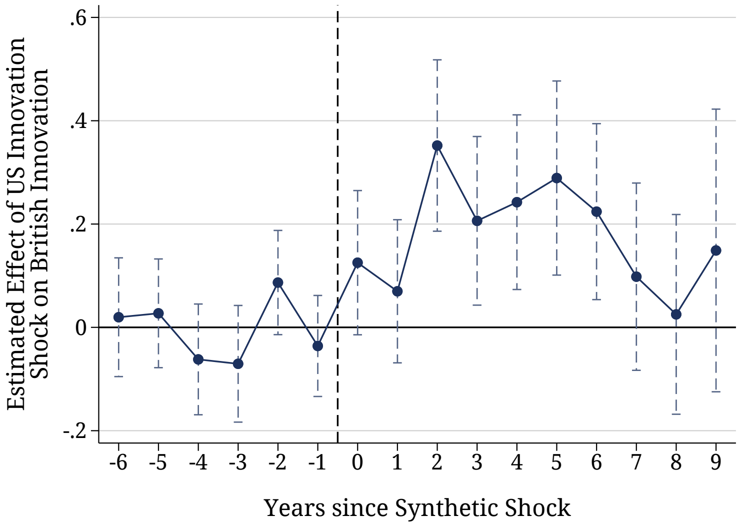


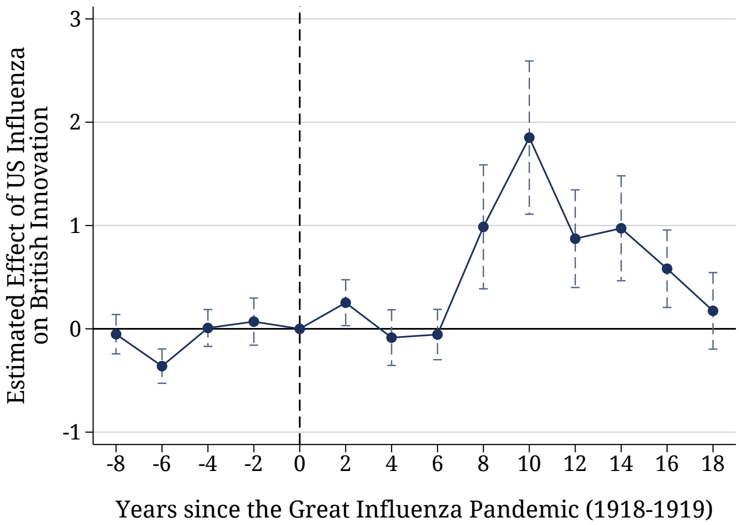
FIGURE 4. HETEROGENEOUS EFFECTS OF RETURN INNOVATION ACROSS TECHNOLOGY CLASSES

Notes. This figure reports the reduced-form effect of instrumented knowledge exposure on innovation by technology class. Each dot reports the coefficient of a regression between the total number of patents and the instrument for knowledge exposure in a given technology class. The unit of observation in each regression is a district, observed at a decade frequency between 1880 and 1939. Regressions include district and decade fixed effects. Bands report 95% confidence intervals. Standard errors are clustered at the district level. The grey line plots the correlation between the revealed UK specialization on the x-axis and the return innovation coefficient for each sector on the y-axis. Red (resp. blue) dots display the regression coefficients for the US (resp UK) revealed comparative advantage sectors.

◄ Back: Results



(A) Synthetic Shocks

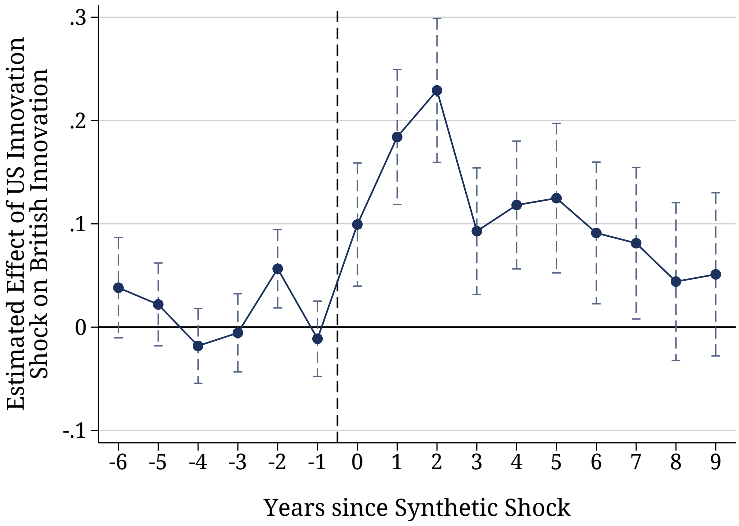


(B) Great Influenza Pandemic Shock

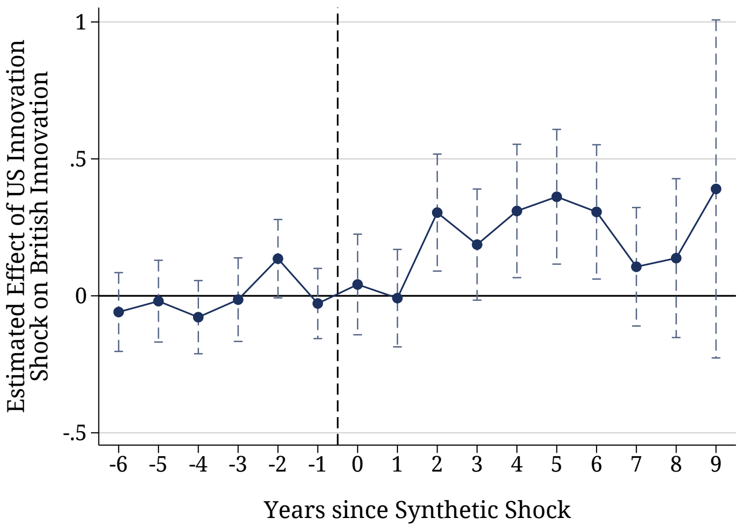
FIGURE 5. EFFECT OF EXPOSURE TO US INNOVATION SHOCKS ON UK INNOVATION

Notes. These figures report the dynamic treatment effects of synthetic shocks (Panel 5a) and the Great Influenza Pandemic shock (Panel 5b) on innovation. The units of observation are district-technology class pairs; units are observed at a yearly frequency in Panel 5a and at a biennial frequency in Panel 5b between 1900 and 1939. The dependent variable is the number of patents. The treatment is an indicator equal to one if: in Panel 5a, a synthetic shock is observed in a given technology in at least one county where the district has above-median out-migration; in Panel 5b, for pharmaceutical patents, emigration from a given district to counties in the top quartile of the mortality distribution is in the top quartile across districts. The black dashed line indicates the timing of the treatment. Standard errors are two-way clustered by district and technology class; bands report 95% confidence intervals.

◄ Back: Results



(A) Backward Similarity ("Copying")



(B) Excess Forward Similarity ("Originality")

FIGURE 6. EFFECT OF EXPOSURE TO US SHOCKS ON THE UK-US PATENT SIMILARITY

Notes. These figures report the dynamic treatment effects of synthetic shocks on backward (Panel 6a) and excess forward (Panel 6b) similarity between UK and US patents. The units of observation are district-technology class pairs observed at a yearly frequency in 1900–1939. In Panel 6a, the dependent variable is the text similarity between UK patents and US patents issued five years before ("copying"); in Panel 6b, the dependent variable is the similarity of UK patents with US patents granted in the subsequent five years, over the similarity of UK patents with US patents granted in the preceding five years ("originality"). The similarity measure is akin to Kelly *et al.* (2021). The treatment is an indicator equal to one if a synthetic shock is observed in a given technology in at least one county where the district has above-median out-migration. Regressions include district-by-year, technology class-by-year, and district-by-technology class fixed effects. The black dashed line indicates the timing of the treatment. Standard errors are two-way clustered by district and technology class; bands report 95% confidence intervals.

◄ Back: Results

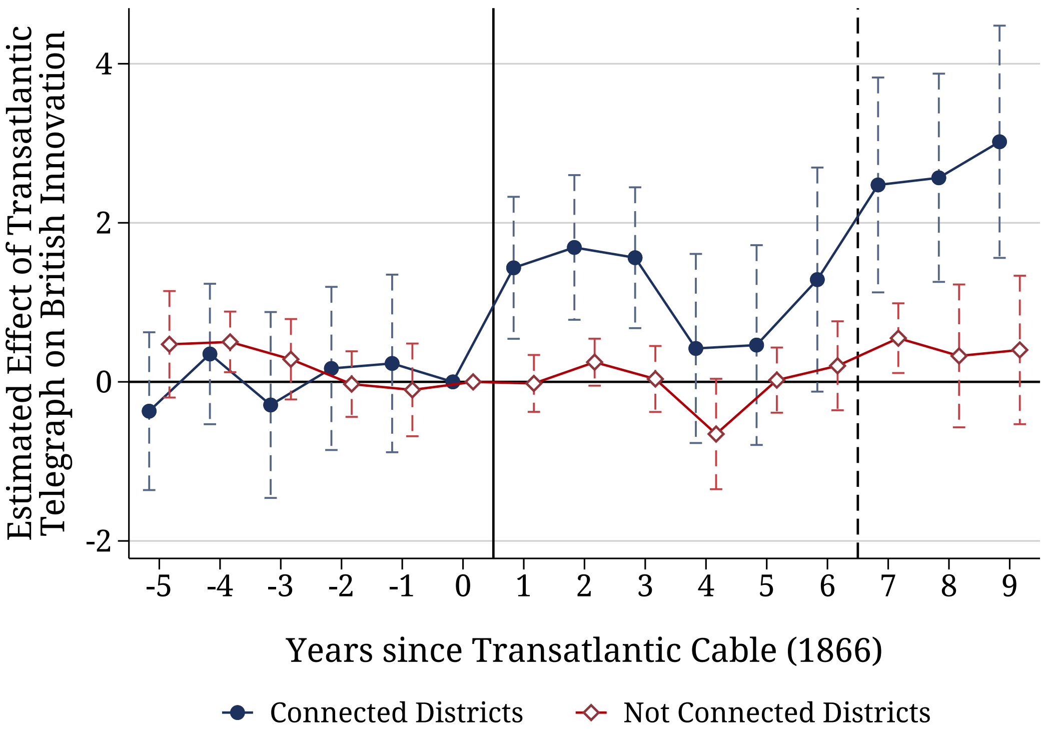


FIGURE 7. EFFECT OF THE TRANSATLANTIC TELEGRAPH CABLE ON INNOVATION

Notes. The figure displays the estimated dynamic treatment effect of the connection of the US and UK telegraph lines on innovation in the UK. The units of observation are districts observed at a yearly frequency between 1860 and 1875. The dependent variable is the number of patents. The independent variable is an interaction between the—time-invariant—number of emigrants in the 1870s and a posttreatment indicator that equals one after the transatlantic telegraph cable. Blue dots report the dynamic treatment effects on the sample of districts connected to the domestic UK telegraph network in 1862; red dots report those for the districts not connected to the network. The black solid vertical bar indicates the year the first cable was laid down (1866); the dashed black vertical line flags the year when the second and third cables were laid (1873-1874). Regressions include district and year fixed effects. Standard errors are clustered at the district level; bands report 95% confidence bands.

◄ Back: Mechanisms

Online Appendix

Return Innovation

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

Davide M. Coluccia and Gaia Dossi

December, 2023

Appendix

A Data Sources and Methods

This section describes the sources and methods we adopted to assemble and merge the various datasets that underlie the empirical analysis. We defer a more detailed discussion of the novel patent data that we digitize, and the linked international migrants sample to sections B and C, respectively.

A.1 Summary of Data Sources

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 multiple 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, 2022). Data before 1900 are not available. In section B, we describe how we digitize the universe of patent documents issued over the period 1853–1900 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 that were not used before they were granted, and become common thereafter. We evaluate this metric on the abstracts of patents granted after 1900. We apply this sample restriction for consistency: in the period 1880–1899, in fact, we observe the full text of patents instead of their abstract.

Migration Data Disaggregated data on the origin of English and Welsh immigrants—and, more generally, of all other nationalities—do not exist. These we collected neither by receiving US authorities nor by sending UK offices. 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 C. Ideally, we observe the universe of British emigrants to the United States between 1870 and 1930. For those individuals, we know all 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, however, we can only construct return migration flows over the period 1870–1910.

Population Census 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 completely digitized.[56] 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, as detailed in section A.2.

Newspapers We collect data on newspaper coverage of US-related news from the British Newspaper Archive.[57] Beach and Hanlon (2022) 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 fifty searches, one for each state. 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 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 geocode each newspaper to the coordinates of the city where it was published and map those to 1891 registration districts. We can thus construct a measure of newspaper coverage at the district-year level.[58] In Table A.1, we provide a set of summary statistics on the resulting dataset. We collect information for a total of 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 (2022), from an average of 2.3 newspapers per district

[56] 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.

[57] A limited free-tier access to newspaper data is available at the following `link`.

[58] 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.

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, except for the London division, newspapers appear to be quite sparse across the country. Figure A.3 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 stands as a major outlier: we thus perform all exercises dropping London districts and find consistent results.

Miscellaneous To construct the domestic UK telegraph network prior to 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 prior to the introduction of the transatlantic cable. We geo-code each station to precise coordinates. The red dots in Figure A.4 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 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.

We digitize import and export yearly data from the 1935 edition of the *Statistical Abstract of the United States*.[59] In particular, we collect the yearly tariff rates applied between 1925 and 1929, i.e., before the Smoot-Hawley Act, and between 1930 and 1935, i.e., after the Act. Tariff rates are available by sector. We then map each industry to a technology class, as listed in Table A.3. In the baseline analysis, we consider an industry protected if its tariff rate increases by more than 50% between 1925-1929 and 1930-1935. We consider alternative thresholds as robustness checks.

GIS Shapefiles & Boundary Harmonization Patents and telegraph stations are mapped to 1900 registration district borders using historical GIS files and their coordinates.[60] However, all data from the population censuses appear at historical borders. Registration districts do not undergo major boundary changes over the period that we study. However, we adapt the method presented by Eckert *et al.* (2020) to UK districts to ensure that we work with consistent geographical units. To construct geograph-

[59] This publication is freely available at the following `link`.

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

ical 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.2 Geo-referenced Census Records

A notable feature of the UK census is that it contains precise information on the residential address of the universe of 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 section, we discuss the methodology that we apply to assign geographical coordinates to textual addresses. This dataset expands earlier work by Lan and Longley (2019), who adopt a different strategy and only analyze the 1901 census, whereas we geo-reference the entire 1851-1911 censuses. Furthermore, the geo-coded census sample is used in the individual-level analysis only. All other exercises do not rely on these data.

A.2.1 Methodology

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

To implement the latter approach Lan and Longley (2019) exploit open-source address data from `OpenStreetMaps`. In this paper, instead, we take advantage of the commercial geo-referenced database developed by `MapTiler AG`. This has three key benefits compared to `OpenStreetMaps`-powered engines. First, the data has some historical "depth", meaning that historical names of locations are sometimes recorded. Second, `MapTiler AG` provides a flexible address-correction engine that matches the query to the closest address available in their dataset. Finally, this commercial database has better coverage than `OpenStreetMaps` in rural areas.

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.

A.2.2 Matching Performance

In Figure A.2, we report the distribution of the share of geo-referenced addresses by district and census decade. The blue bars refer to the simple matching rate, defined as the share of geo-referenced addresses. The black-contoured bars, instead, adjust for the number of residents recorded in each address. In each figure, we report the average matching rates and their respective standard deviations. The average matching rate ranges between 76% in 1851 and 86% in 1911. All distributions display substantial right-skewness, meaning there are very few districts with a matching rate lower than 50%. The matching rate increases over time for two reasons. First, the quality of recorded addresses increases in more recent censuses. Second, the urban geography in 1911 is more similar to that in the `MapTiler AG` database than in 1851. This is due to street re-labeling and urban agglomerates' growth and consolidation. Figure A.1 displays the spatial distribution of the average geo-referencing rates across censuses. Figure A.1a reports the crude rate, whereas Figure A.1b, we adjust by address-population. Except for Wales and some rural districts at the center of England, the geo-referencing rates are above 80% everywhere. It is particularly high—above 90%—in North-Western and South-Eastern England. More urbanized areas generally tend to feature larger geo-referencing rates because addresses tend to be more informative. This notwithstanding, differences are quantitatively small as the matching rate is remarkably homogeneous across registration districts. Wales is the single most relevant exception. The geo-referencing rate there is very low because addresses in the census until 1901-1911 tend to be reported in Welsh, especially in Western areas.

Taken together, the results of the geo-referencing algorithms are satisfactory. More than 80% addresses are successfully matched to precise geographical coordinates. This ratio is even higher in areas outside Wales, where innovation and migration activity are more intense.

A.3 Linked Inventor Sample

This section presents the methodology we use to link patents to census records. The linked inventors-census sample is used in the individual-level analysis only. All other exercises do not rely on these data.

A.3.1 Methodology

We follow the logic of Berkes (2018), who links patents to census records in the US. We link patents between 1881 and 1899 to the 1891 census and those between 1901 and 1920 to the 1911 census. Relative to our baseline sample, we thus drop patents issued after 1920 because we cannot observe individuals born after 1911. While this is probably a minor issue for patents granted until 1930, it may induce some selection of linked inventors for later patents. Patent data contain the name and surname of inventors, their residence, and the issue year.

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 $\mathcal{L}_p^{\text{district}}$ and $\mathcal{L}_p^{\text{county}}$ be the set of, respectively, 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:

$$\mathcal{M}_{A_p}^{\text{parish}} = \left\{ i \in \mathcal{L}_p^{\text{parish}} \middle| t_p - \text{year}_i \geq 18 \right\} \tag{A.1}$$

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:

$$\text{Similarity}_i^{A_p} = \alpha \times \text{Name Similarity}_i^{A_p} + (1 - \alpha) \times \text{Surname Similarity}_i^{A_p} \tag{A.2}$$

   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:

$$\overline{\mathcal{M}}_{A_p}^{\text{parish}} = \left\{ i \in \mathcal{M}_{A_p}^{\text{parish}} \middle| \text{Similarity}_i^{A_p} = \max_{i' \in \mathcal{M}_{A_p}^{\text{parish}}} \text{Similarity}_{i'}^{A_p} \right\} \tag{A.3}$$

   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 it is empty, repeat steps 2–4 conditioning on records in $\mathcal{L}_p^{\text{district}}$. If $\overline{\mathcal{M}}_{A_p}^{\text{district}}$ is empty, repeat steps 2–4 conditioning on records in $\mathcal{L}_p^{\text{county}}$. If $\overline{\mathcal{M}}_{A_p}^{\text{county}}$ is empty, repeat steps 2–4 without imposing geographical conditions on records i. In the baseline setting, we only accept county-level and country-level matches if the name and surname of the match(es) exactly match A_p's.

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 1911 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. Parishes are small, as their average population is less than 10,000. When a match at the parish level is feasible, it is usually unique. We then progressively expand the set of records by coarsening their geographic location. Districts are larger than parishes, and counties are, in turn, larger than districts. If we cannot find one match at the county level, we look for one within the entire population of England and Wales. Unlike the migrants sample, we do not have information on the birth year. To ensure that county- and country-level matches are reliable, we require that their name and surname are verbatim those recorded in the patent document.

A.3.2 Matching Statistics

In Figure A.5, we report the matching rate of this exercise. We focus on two matching rates: the gross rate is the share of inventors that have at least one match, relative to the overall set of inventors; the net rate is the share of inventors with at least one *acceptable* match, relative to the overall group of inventors. In the analysis, a match is acceptable if (i) the similarity between name and surname is above 0.95 and (ii) a given inventor has no more than five matches. Panel A.5a reports both margins over time. The gross matching rate remains consistently above 80% throughout the period. The net matching rate, however, rejects approximately 20% of the matches. This is mainly due to inventors linked to more than five census records. This notwithstanding, the share of acceptable matches is approximately constant and above 60% each year. Our algorithm delivers satisfactory performance compared to standard linking rates in the literature. In panel A.5b, we break down the number of matches by the geographical unit where the match is attained. Blue, red, green, and yellow bars report the matching rates at the parish, district, county, and national levels. The share of inventors matched with more than 20 census records is larger at the national level; there, we look for possible matches with no information on the residence. Multiple matches are somewhat common at the parish level as well. This is because we first try to match inventors at the parish level. Hence parish matches represent the large majority of the linked sample, while district-level matches are residual and, thus, more accurate. Figure A.6 displays the spatial distribution of inventors, who are plotted using the geo-coded census coordinates described in the previous section.

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. While these hypotheses are ultimately challenging to test, 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 age; in Panel B, we list the set of occupational categories; in Panel C, we list the residence divisions. We find no clear association between the number of matches and these variables in the overall sample (column 1) and across matches selected by geographical layer (columns 2–5). Overall, we interpret the Table as conveying reassuring evidence that the selection of inventors into the linked sample does not appear to systematically favor particular groups.

A.4 Tables

	(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

TABLE A.1. DESCRIPTIVE STATISTICS ON NEWSPAPERS AND NEWSPAPER COVERAGE IN THE UK

Notes. This table reports descriptive statistics on newspapers active in the UK between 1850 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 1900.

◄ Back: Appendix 1 – Data

	Overall Sample	Parish Matches	District Matches	County Matches	Nationwide Matches
	(1)	(2)	(3)	(4)	(5)
Panel A. Demographics					
Age	0.005	-0.018	-0.005	-0.014*	0.026***
	(0.010)	(0.021)	(0.012)	(0.007)	(0.005)
Dependent Variable – Dummy = 1 if Matched Inventor is in:					
Panel B. Occupation					
Agriculture	0.123*	0.272**	0.073***	0.000	0.028**
	(0.073)	(0.129)	(0.019)	(0.016)	(0.011)
Chemicals	-0.010***	-0.019***	-0.012*	-0.011***	-0.005***
	(0.004)	(0.005)	(0.006)	(0.004)	(0.001)
Construction	-0.015	-0.032	0.006	0.010	-0.001
	(0.016)	(0.031)	(0.008)	(0.017)	(0.002)
Engineering	-0.018	-0.042*	-0.017**	-0.007	0.004
	(0.012)	(0.025)	(0.007)	(0.007)	(0.003)
Liberal Professions	-0.014***	-0.016***	-0.003	-0.018***	-0.019***
	(0.005)	(0.006)	(0.010)	(0.002)	(0.005)
Metallurgy	-0.020	-0.031	0.018	-0.015	0.004**
	(0.013)	(0.019)	(0.019)	(0.019)	(0.002)
Other Manufacturing	-0.024*	-0.042**	-0.027	0.005	-0.007**
	(0.012)	(0.018)	(0.016)	(0.007)	(0.003)
Public Administration	-0.009	-0.017	-0.014	-0.015	-0.008**
	(0.008)	(0.014)	(0.010)	(0.011)	(0.003)
Textiles	-0.013	-0.042*	0.002	0.058***	0.003
	(0.012)	(0.026)	(0.024)	(0.013)	(0.005)
Trade	-0.031***	-0.044***	-0.038***	-0.021*	-0.025***
	(0.011)	(0.016)	(0.007)	(0.011)	(0.005)
Transport	-0.008	-0.025	0.001	0.006	0.003
	(0.014)	(0.026)	(0.008)	(0.011)	(0.003)
Utilities	-0.013***	-0.020***	-0.031***	-0.016***	-0.007*
	(0.004)	(0.005)	(0.005)	(0.006)	(0.004)
Panel C. Division of Residence					
East	-0.004	-0.059	-0.061	-0.074	-0.010
	(0.015)	(0.067)	(0.065)	(0.089)	(0.011)
East Midlands	0.004	-0.061	0.070	-0.055	0.012
	(0.012)	(0.070)	(0.102)	(0.066)	(0.013)
London	-0.048	-0.039	-0.053	-0.008	-0.025
	(0.069)	(0.173)	(0.053)	(0.079)	(0.022)
North East	0.028	-0.045	-0.041	-0.060	0.016
	(0.031)	(0.052)	(0.049)	(0.072)	(0.016)
North West	-0.057	-0.165	-0.069	0.195***	0.012
	(0.050)	(0.167)	(0.080)	(0.056)	(0.013)
South East	-0.024	-0.050	-0.106	-0.118	-0.028
	(0.030)	(0.057)	(0.106)	(0.136)	(0.027)
South West	0.001	-0.025	-0.049	-0.051	-0.019
	(0.008)	(0.029)	(0.054)	(0.062)	(0.020)
Wales	0.233	0.469**	0.540***	-0.026	0.046
	(0.187)	(0.212)	(0.137)	(0.032)	(0.046)
West Midlands	-0.049	-0.102	-0.104	-0.130	0.004
	(0.050)	(0.112)	(0.103)	(0.148)	(0.007)
Yorkshire	0.005	-0.021	-0.029	0.008	-0.003
	(0.013)	(0.025)	(0.032)	(0.021)	(0.007)
Decade FE	Yes	Yes	Yes	Yes	Yes

Table A.2. Correlation Between Inventors' Characteristics and Number of Matches

Notes. This table reports the correlation between inventor-level variables observed in the UK census and the number of matches in the linked sample. In column (1), the sample is the entire linked dataset. We restrict to matches at the parish (column 2), district (column 3), county (column 4), and national level (column 5). The Table reports standardized beta coefficients for comparability. Regressions include decade fixed effects. Standard errors are clustered at the division level and are reported in parentheses.

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

◄ Back: Appendix 1 – Data

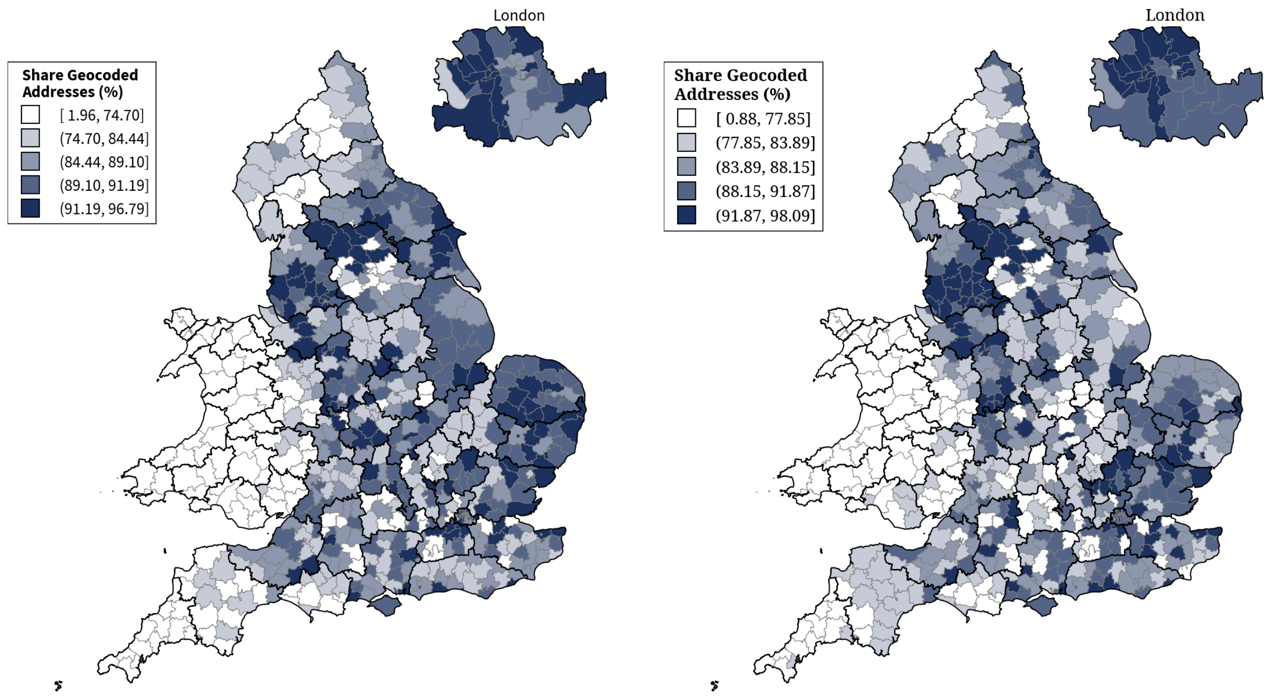
Sector	Technology Class	Tariff Rate Before S-H	Tariff Rate After S-H	Change in Tariff	Treated
(1)	(2)	(3)	(4)	(5)	(6)
Agricultural products and provisions	Agriculture	23.059	40.204	74.352	Yes
Chemicals, oils, and paints	Chemistry	29.577	40.195	35.900	No
Cotton Manufactures	Textiles	34.876	44.764	28.352	No
Earths, earthenware, and glassware	Personal Articles, Furniture	47.321	53.049	12.106	No
Flax, hemp, and jute, and manufacture thereof	Textiles	18.948	26.104	37.766	No
Metals, and manufacture thereof	Metallurgy	34.534	36.803	6.572	No
Pulp, paper, and books	Printing	25.652	25.591	-0.239	No
Silk and silk goods	Textiles	55.768	58.115	4.208	No
Spirits, wines, and other beverages	Food	37.298	59.007	58.226	Yes
Sugar, molasses, and manufactures thereof	Food	68.971	110.022	59.519	Yes
Sundries	Personal Articles, Furniture	38.149	36.587	-4.096	No
Tobacco, and manufactures thereof	Agriculture	58.176	81.636	40.326	No
Wood, and manufactures thereof	Building	23.727	20.672	-12.875	No
Wool, and manufactures thereof	Textiles	49.344	78.255	58.591	Yes

TABLE A.3. LIST OF INDUSTRIES BY TARIFF RATE, 1925–1935

Notes. This table reports the US tariff rate applied to the categories listed in the *Statistical Abstracts of the United States*. Column (1) reports the listed sector; column (2) maps the sector to technology classes in our baseline taxonomy; columns (3) and (4) report the tariff rate applied, respectively, before and after the Smoot-Hawley Act (1930). Tariff rates before the Act are averages in the five years before the reform (1925–1929); tariff rates after the Act are averages in the five years posterior to the reform (1930–1935). Column (5) computes the change in the tariff rates. In column (6), we list the technology classes we considered targeted by the Act, namely, those whose tariff rate increase exceeded 50%. Data are digitized from the 1935 *Statistical Abstracts of the United States*.

◄ Back: Appendix 1 – Data ◄ Back: Appendix 4 – Additional Results

A.5 Figures

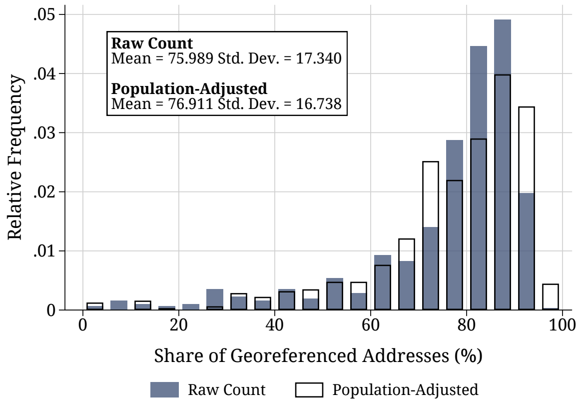


(A) Unweighted Geo-coding Ratio (B) Population-Weighted Geo-coding Ratio

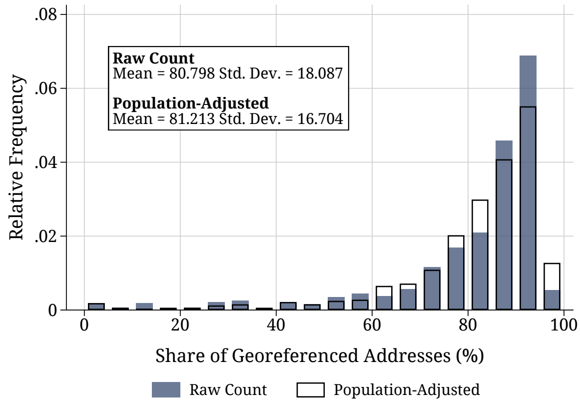
FIGURE A.1. SPATIAL DISTRIBUTION OF THE SHARE OF GEO-CODED ADDRESSES IN THE UK POPULATION CENSUSES, 1851–1911

Notes. These figures report the spatial distribution of the share of geo-referenced addresses from the UK censuses, 1851–1911. For each census, we obtain a list of more than five million addresses by fine geographical unit (i.e., parishes). We then geo-reference these addresses to precise geographical coordinates. Panel A.1a reports the district-level share of successfully geocoded addresses. In Panel A.1b, we weigh each address by the number of people reported to live in that address. The performance of the geo-referencing algorithm is relatively poor in Wales because addresses there are often reported in Welsh.

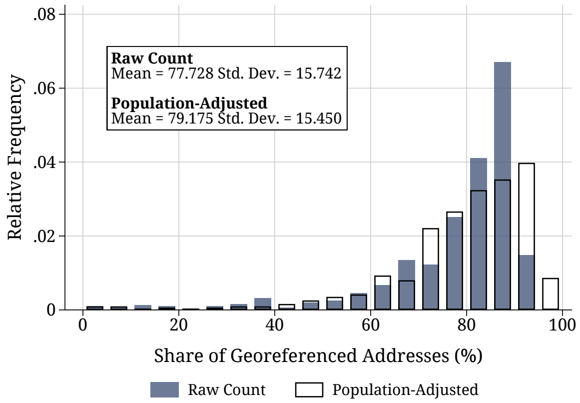
◄ Back: Appendix 1 – Data



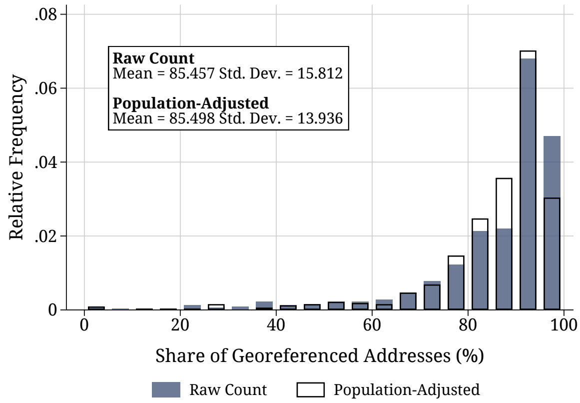
(A) 1851 Census



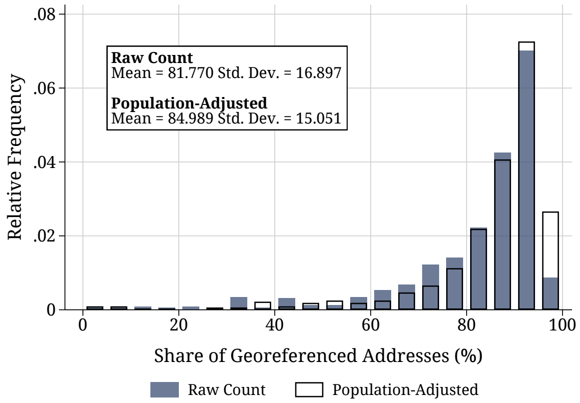
(B) 1861 Census



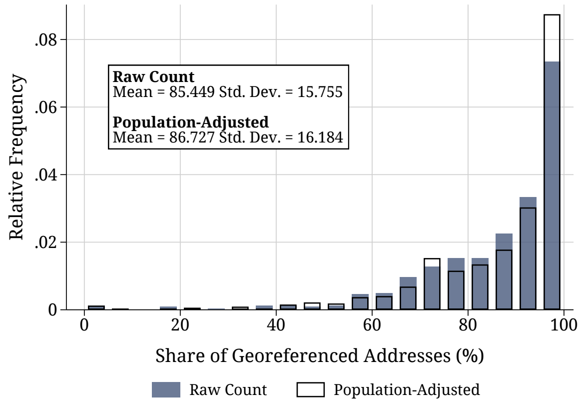
(C) 1881 Census



(D) 1891 Census



(E) 1901 Census



(F) 1911 Census

FIGURE A.2. DISTRIBUTION OF THE SHARE OF GEO-CODED ADDRESSES BY CENSUS

Notes. These figures display the district-level distribution of the share of geo-coded addresses from the UK censuses (1851–1911) by decade. For each census, we obtain a list of more than five million addresses by fine geographical unit (i.e., parishes). We then geo-reference these addresses to precise geographical coordinates. The black-contoured bars report the crude geo-coding rate; the blue bars report the population-adjusted geo-coding rate. Each figure reports the average and standard deviation of the two distributions.

◄ Back: Appendix 1 – Data

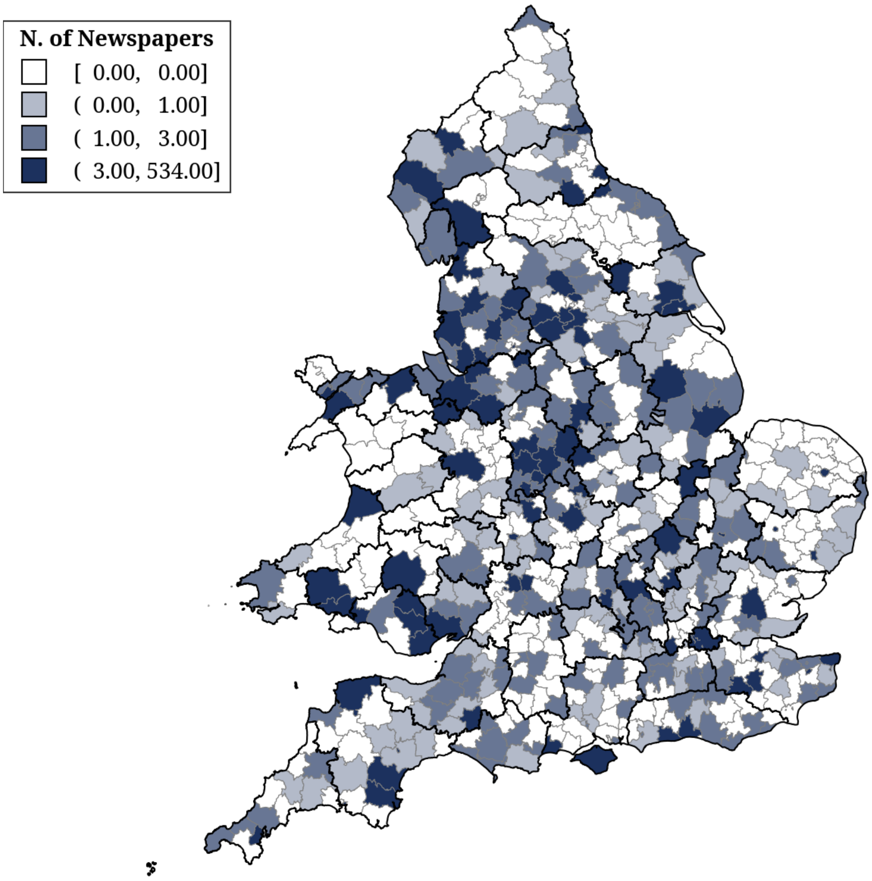


FIGURE A.3. NUMBER OF ACTIVE NEWSPAPERS OVER THE PERIOD 1880–1940, BY DISTRICT

Notes. This figure reports the spatial distribution of the number of active newspapers across districts over the period 1880–1940. To be included in the data, a publication must be active for at least one year between 1880 and 1940. To retrieve the location of each journal, we geo-reference its publishing address and overlay historical district boundaries to assign it to consistent 1900 districts. The publishing address only lists the city. Hence we cannot distinguish across the eleven London urban districts. We consequently dissolve these districts into a single "London" unit.

◄ Back: Data ◄ Back: Appendix 1 – Data

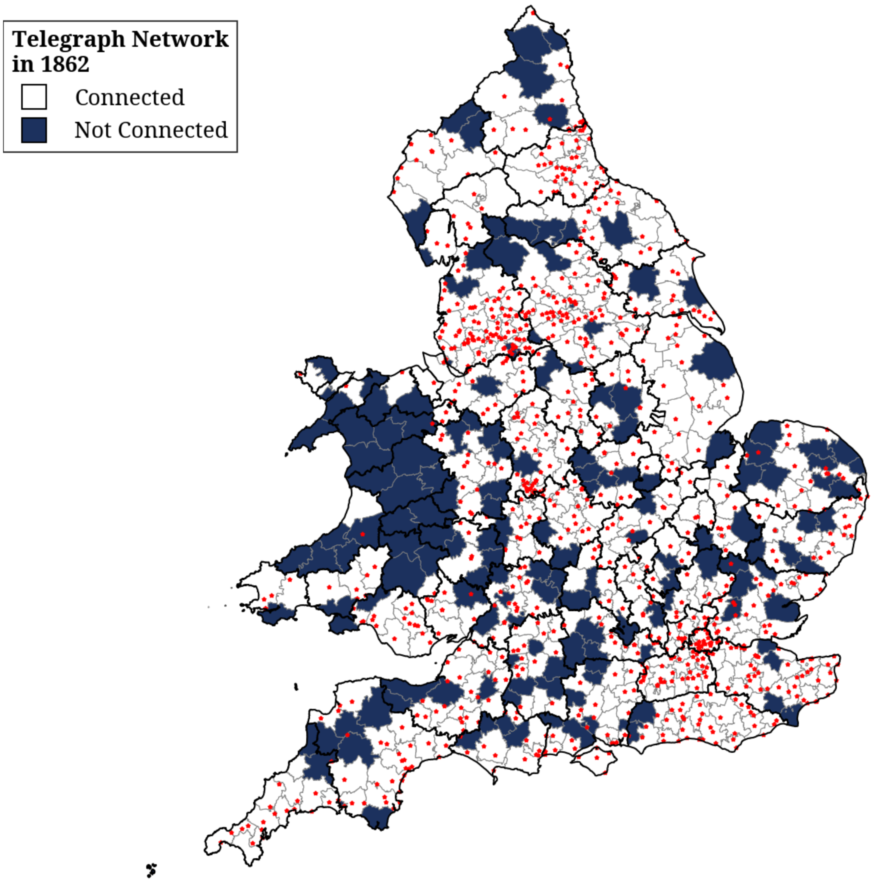
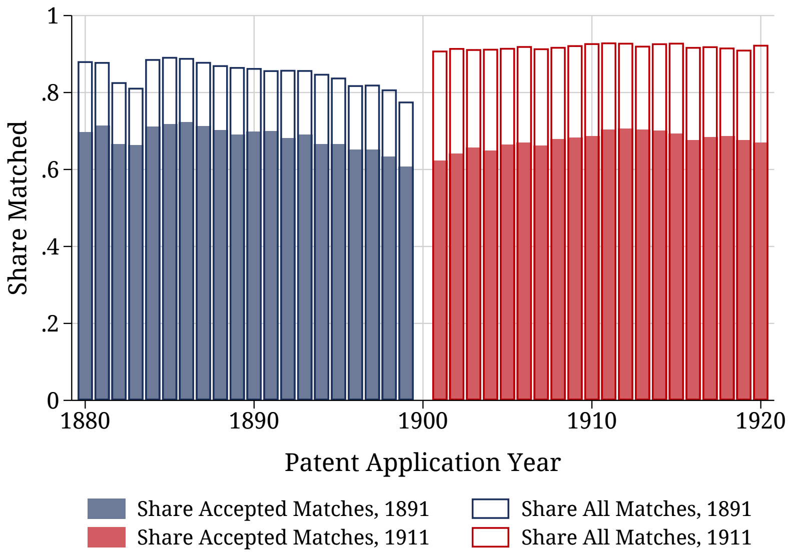


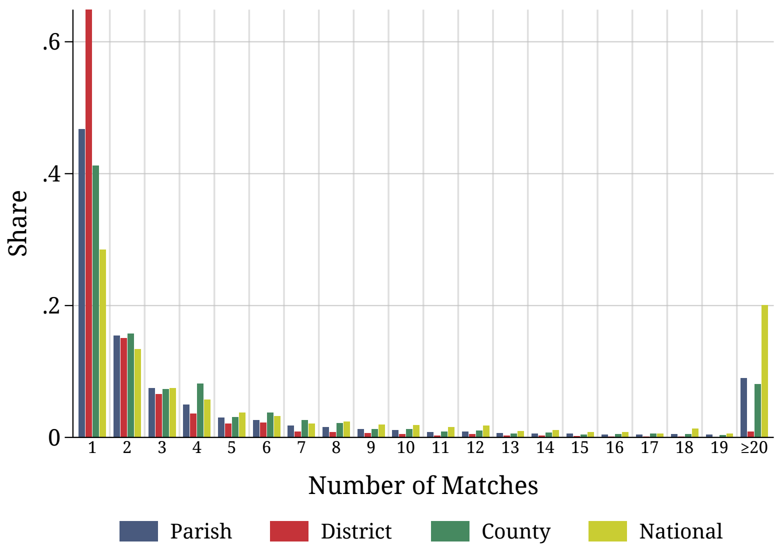
FIGURE A.4. DISTRIBUTION OF DISTRICTS CONNECTED TO THE UK TELEGRAPH NETWORK IN 1862

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.

◄ Back: Data ◄ Back: Appendix 1 – Data



(A) Matching Rate Over Time



(B) Matching Rate by Geographic Layer

FIGURE A.5. MATCHING RATE OF THE LINKED INVENTORS-CENSUS SAMPLE, 1881–1911

Notes. These figures report the matching rate for the linked inventor-census sample. Panel A.5a reports the matching rate over time for the 1881–1900 sample (blue bars) and the 1901–1920 sample (red bars). Color-contoured bars report the share of records with at least one match; color-filled bars report the share of acceptable linked matches. A record match is acceptable if it has no more than five multiple matches. Panel A.5b reports the share of matches by the number of matches, broken down by geographical layers. In Panel A.5a, we do not show the few matches with quality below .95. In Panel A.5b, the sample is restricted to records with at least one match.

◄ Back: Appendix 1 – Data

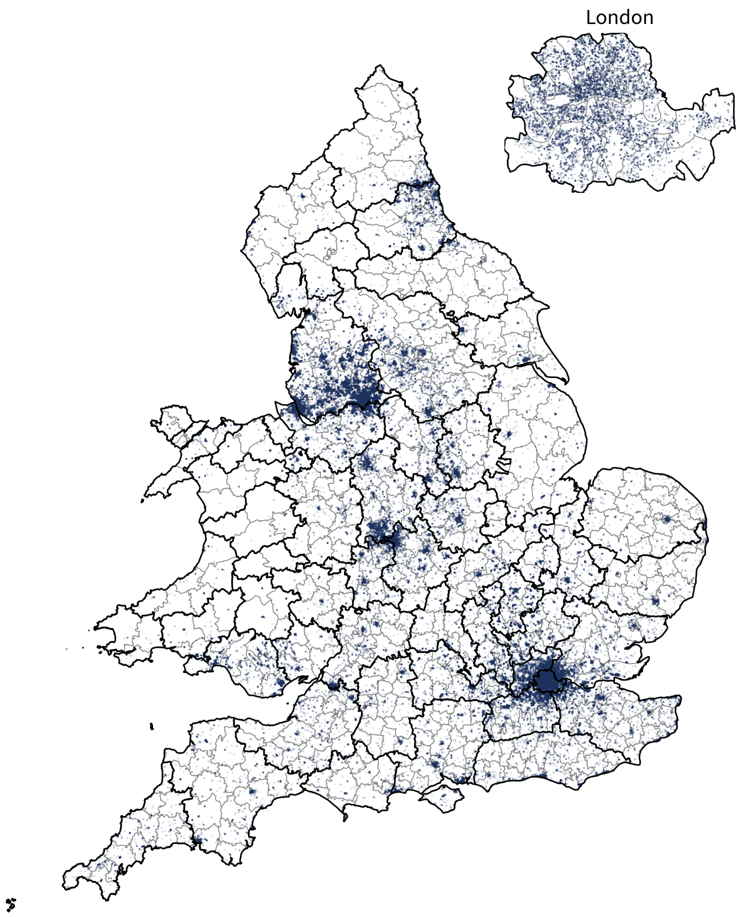


FIGURE A.6. DISTRIBUTION OF INVENTORS ACROSS UK DISTRICTS, 1881–1911

Notes. This figure displays the spatial distribution of inventors across districts between 1881 and 1911. 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 districts at historical borders in 1900.

◄ Back: Appendix 1 – Data

B Novel Patent Data

B.1 Sources and Digitization

This section presents the motivation for developing a new patent dataset for England and Wales that spans the second half of the XIX century. Then, we describe the sources we use and how we structure the textual data they contain into a machine-usable dataset. Finally, we describe two data-augmentation routines that we perform to geocode the patents and assign them a modern technology class.

B.1.1 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 *et al.*, 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.

B.1.2 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 B.1: in panel B.1a we show the patent granted to Henry Bessemer for the eponymous process to produce steel, and in panel B.1b 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.

B.1.3 Digitization

We perform optical character recognition (OCR) on each patent individually 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, which is 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 implement a script that leverages regular expressions to identify the variables (i–ix). Fortunately, the text of each patent is fairly standardized; hence this routine yields detailed and high-quality results for all variables except (v), which is not used in this paper. This exercise results in a database of approximately 800,000 patents granted between 1853 and 1899.

B.1.4 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 B.4, we report the resulting distribution of patents (panel B.4a) and patents per capita (panel B.4b). Reassuringly, these are consistent with underlying population and economic development indicators.

B.1.5 Technology Class Assignment

Naturally, historical patent documents do not list CPC classes. Yet, the technological classification is a key variable in our empirical exercise. To reconstruct the class, we adopt a supervised machine-learning approach. We conjecture, following Xu (2018), that titles are informative of technological classes. We split the PATSTAT data, which covers the years 1900–1939 and for which we observe both titles and classes, in a train and a test set, with a proportion of 4:1. We apply a term frequency-inverse document frequency vectorization algorithm to the titles of both datasets. Then, we estimate a linear support vector machine (LSVC) on the train set. An LSVC is a non-probabilistic classifier that assigns class labels to maximize the width of the gap between classes. Formally, consider a set of points $(\mathbf{x}_i, y_i)_{i=1}^N$ where $\mathbf{x} \in \Re^N$ represent the features—in our case, words—and y is the class. For simplicity, assume $y \in \mathcal{Y} = \{-1, 1\}$. An LSVC solves for the hyperplane $\mathcal{W} = \{\mathbf{w} \in \Re^N \text{ s. t. } \mathbf{w}^\intercal \mathbf{x}_i - \ell = 0\}$ that maximizes the distance between the group i such that $y_i = 1$, and the group where $y_i = -1$. The distance that is most commonly used that allows for non-linearly separable data is the hinge loss, which is defined as $d_i = \max\{0, 1 - y_i(\mathbf{w}^\intercal \mathbf{x}_i - \ell)\}$. In our case, however, we allow for multiple classes, that is, the cardinality of $\mathcal{Y}$ be more than two. We employ an LSVC because the literature notes that it yields particularly robust results. However, the classification outcome would remain fairly unchanged using different algorithms.[61]

On the training set, the LSVC yields a 95% accuracy, measured as the share of patents with a correctly imputed class relative to the total number of patents. This decreases to 85% on the test set, which is not used to train the algorithm. Given that state-of-the-art models trained on modern US data achieve approximately a test 90% accuracy, we interpret these results as rather encouraging (Li *et al.*, 2018). We report the confusion matrix on the test set in Figure B.2. For a given cell, the row label is the true technology class, and the column label is the imputed class. A perfect classifier would thus yield a diagonal confusion matrix. Overall, we find that misclassification errors are evenly distributed, in relative terms, across classes. Hence, even though the classifier is not perfect, there does not seem to be any systematic measurement error in class imputation.

B.2 External Validation

To validate our data, we consider the only two series that cover—a portion of—the years 1853–1899. 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 that we use as a comparison 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 are also listed. Absent a way to identify granted patents, we do not report figures after 1883 for the COI series.

In Table B.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.

B.3 Measuring Pairwise Similarity Between US and UK Patents

In this section, we describe in detail how we construct the patent similarity metric we adopt 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 contained in titles only. Even

[61] In particular, we tested the Naïve Bayes classifier, several Boosting algorithms (e.g., AdaBoost, XGBoost), a random forest classifier, and a simple convolutional neural network. All the above yield similar classification results but slightly lower accuracy than the LSVC. Additionally, we explored alternative vectorization algorithms using transformers (e.g., BERT and RoBERTa) with no significant performance gains.

though we do not have access to full US patent texts, the title of a patent is usually very informative about its content. In fact, 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 the period 1870–1899 and are collected from PATSTAT for the later years; titles for US patents are collected from PATSTAT throughout the sample period.

We start by defining 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

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

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 both the index of the patent and the set of words it contains. We shall have

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

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:

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

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

We compute the TF-BIDF$_{p,t}$ vectors for US and UK patents, but the 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 narrative of the paper, we define as "copying". Formally we define

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

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

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

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 Table E.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 that we do not conflate shifting terminology fashions in the similarity measures.

B.4 Summary Statistics and Stylized Facts

We conclude this section by presenting some stylized statistics and facts our new data allow us to uncover. First, as noted in Table B.1, the number of patents granted generally grows over time, although at a somewhat stagnating path. There is, however, a sizable discontinuity between 1883 and 1884, when the number of patents jumps from 6074 to 9873. In 1883 the Patents Act reduced application fees by 83%, as noted by Nicholas (2014). It seems plausible to attribute the discontinuity to this reform.

Second, in Figure B.3, we report the composition of patenting activity by technology class. In each year, we compute the share of patents in a given sector with respect to the total number of patents issued that year. We report such shares over time between 1853 and 1939. The composition of innovation exhibits two clear patterns. First, the share of textiles patents, which in the 1850s represented nearly 20% of the total, shrinks considerably, and in 1939 it accounts for less than 5%. This is consistent with the historical preeminence of textiles during the First Industrial Revolution and their subsequent loss of importance. Second, electricity-related innovation grows considerably in the later part of the period. In 1939, it represented more than 20% of the total number of patents issued. Once more, this echoes historical, anecdotal evidence highlighting the centrality of electricity during the later stages of the Second Industrial Revolution and beyond (David, 1990; Mokyr, 1998).

Finally, a crucially novel component of our dataset is that it allows studying the geographical dimension of the innovation process. Thus, in Figure B.4, we report the spatial distribution of the number of patents in absolute number (panel B.4a) and normalized by population (panel B.4b). These maps attest to the importance of duly considering the geography of innovation. The patenting activity appears to be widely dispersed across England and Wales. Heavily industrial areas, such as Lancashire, the Midlands, the Tyne, and South Wales, all feature prominently in terms of issued patents. Similarly, the London area is also a major innovation hub. By contrast, Northern Wales, Anglia, Cornwall, and Cumbria perform poorly. In Figure B.5, we repeat this exercise, but we break down the number of patents by selected technology classes: chemistry (panel B.5a), electricity (panel B.5b), engineering (panel B.5c), engines and pumps (panel B.5d), metallurgy (panel B.5e), and textiles (panel (panel B.5f). While innovation centers remain roughly similar across sectors, some differences emerge. For example, the metallurgy industry was particularly deep-rooted in the Midlands, where we note the largest concentration of metallurgy patenting. Similarly, textile innovation centers in the Lancashire area, the historic “cotton districts”. Our database allows studying a novel, thus far largely unexplored dimension of the innovation and patenting activity. Therefore, the analysis carried out in this paper is one of many that may take advantage of this contribution.

B.5 Tables

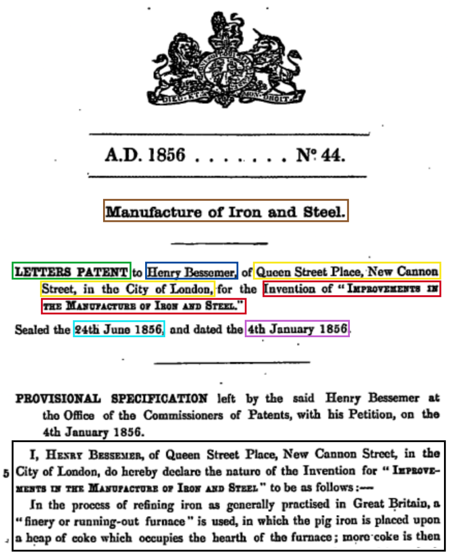
Years 1853-1876				Years 1877-1899			
(1)	(2)	(3	(4)	(5)	(6)	(7)	(8)
Year	Our Series	COI	Hanlon	Year	Our Series	COI	Hanlon
1853	3042	3016		1877	4943	4928	4940
1854	2759	2690		1878	5336	5143	5333
1855	2960	2866	2955	1879	5332	5305	5325
1856	3107	2967	3102	1880	5499	5132	5509
1857	3206	3092	3197	1881	5744	5620	5745
1858	3023	2954	2999	1882	6159	6150	6233
1859	3048	2989	2998	1883	6074	6006	5981
1860	3192	3139	3190	1884	9873		
1861	3261	3269	3272	1885	8783		
1862	3482	3459	3486	1886	8999		
1863	3301	3299	3308	1887	9218		
1864	3256	3225	3257	1888	9331		
1865	3378	3364	3378	1889	10325		
1866	3451	3408	3452	1890	10355		
1867	3724	3692	3720	1891	10686		
1868	4008	3908	3984	1892	11429		
1869	3832	3741	3781	1893	11985		
1870	3407	3288	3405	1894	11648		
1871	3525	3479	3525	1895	12198		
1872	3969	3940	3967	1896	13597		
1873	4276	4281	4282	1897	14249		
1874	4494	4516	4491	1898	13100		
1875	4557	4451	4557	1899	13172		
1876	5049	5012	5064				

Table B.1. Total Number of Patents Granted in the UK: Comparison Across Three Datasets

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.

◄ Back: Appendix 2 – Patents

B.6 Figures



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."

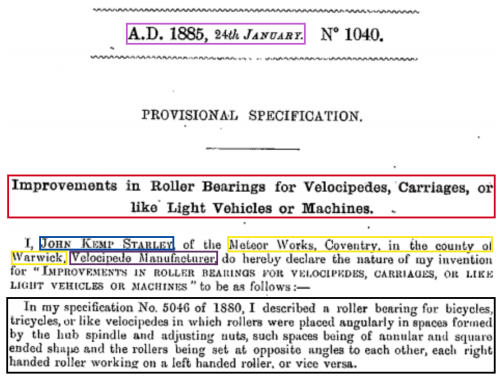
Sealed the 24th June 1856, and dated the 4th January 1856.

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

(A) Henry Bessemer's 1856 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.

(B) John K. Starley's 1885 Bicycle Patent

FIGURE B.1. SAMPLE ANNOTATED PATENT DOCUMENTS: THE BESSEMER PROCESS AND THE FIRST MODERN SAFETY BICYCLE

Notes. This figure displays two sample patent documents in our dataset. Panel B.1a was granted to Henry Bessemer in 1856 for the invention of the famous eponymous process for the mass production of steel from the molten pig iron. Panel B.1b was granted to John Starley in 1885 for the invention of the first modern bicycle, which would soon revolutionize mobility in Europe and in 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.

◄ Back: Appendix 2 – Patents

True Technology Class \ Predicted Technology Class	Agriculture	Building	Chemistry	Electricity	Engineering	Engines, Pumps	Food	Health, Amusement	Instruments	Lightning, Heating	Metallurgy	Personal Articles, Furniture	Printing	Separating, Mixing	Shaping	Ships, Aeronautics	Textiles	Transporting
Agriculture	951	10	23	2	2	4	24	8	5	6	0	45	2	21	34	21	6	12
Building	9	2310	42	11	58	6	0	14	25	23	4	65	3	12	63	58	7	79
Chemistry	7	56	4704	34	18	21	21	21	19	78	62	39	31	89	133	31	71	9
Electricity	1	9	25	5146	30	43	1	14	163	53	21	20	12	8	34	40	14	63
Engineering	1	41	15	55	1975	105	0	11	37	36	6	62	4	3	100	62	19	151
Engines, Pumps	1	5	13	43	80	2113	0	2	31	47	2	5	0	10	10	40	1	59
Food	11	2	60	1	5	1	936	9	4	32	0	27	2	22	44	46	3	1
Health, Amusement	6	21	57	29	20	7	5	1525	64	24	7	101	5	15	31	54	17	20
Instruments	3	19	28	304	76	43	11	48	4148	69	23	93	78	14	40	115	17	92
Lightning, Heating	10	22	53	65	35	51	18	7	51	2739	31	50	5	40	24	30	16	40
Metallurgy	2	5	85	18	14	6	0	9	26	39	1565	18	8	17	58	37	5	14
Personal Articles, Furniture	12	67	43	15	26	3	26	47	72	40	3	3847	44	13	90	127	60	26
Printing	1	12	49	16	2	3	4	4	80	13	4	49	1567	11	77	58	45	5
Separating, Mixing	12	8	145	10	19	9	13	16	27	44	33	18	17	1057	54	44	18	1
Shaping	20	61	119	47	111	13	11	19	51	35	52	121	55	26	3737	143	57	70
Ships, Aeronautics	16	51	24	38	73	36	37	25	92	24	21	102	60	22	116	3736	100	97
Textiles	4	10	102	11	18	5	4	5	11	6	4	82	27	12	54	40	2898	4
Transporting	14	59	4	88	149	42	0	5	50	59	3	55	1	1	36	60	4	3863

FIGURE B.2. CONFUSION MATRIX OF THE TECHNOLOGY SECTOR CLASSIFIER

Notes. This figure displays the confusion matrix of the patent technology classifier. The algorithm assigns to each patent an imputed technology class using information contained in the title. Titles undergo pre-processing and term frequency-inverse document frequency (tf-idf) vectorization. The classifier is trained on an 80% sub-sample of the universe of British patents granted over the period 1900–1940. The figure reports the classifier's performance on the remaining 20% test set, which is not used in training. The y-axis reports the true patent class; the x-axis reports the class imputed by the classifier. A perfect classifier would yield a diagonal confusion matrix. The accuracy in the training (resp. test) set is ≈98% (resp. ≈85%). Lighter to darker blue indicates an increasing number of patents in the cell.

◄ Back: Appendix 2 – Patents

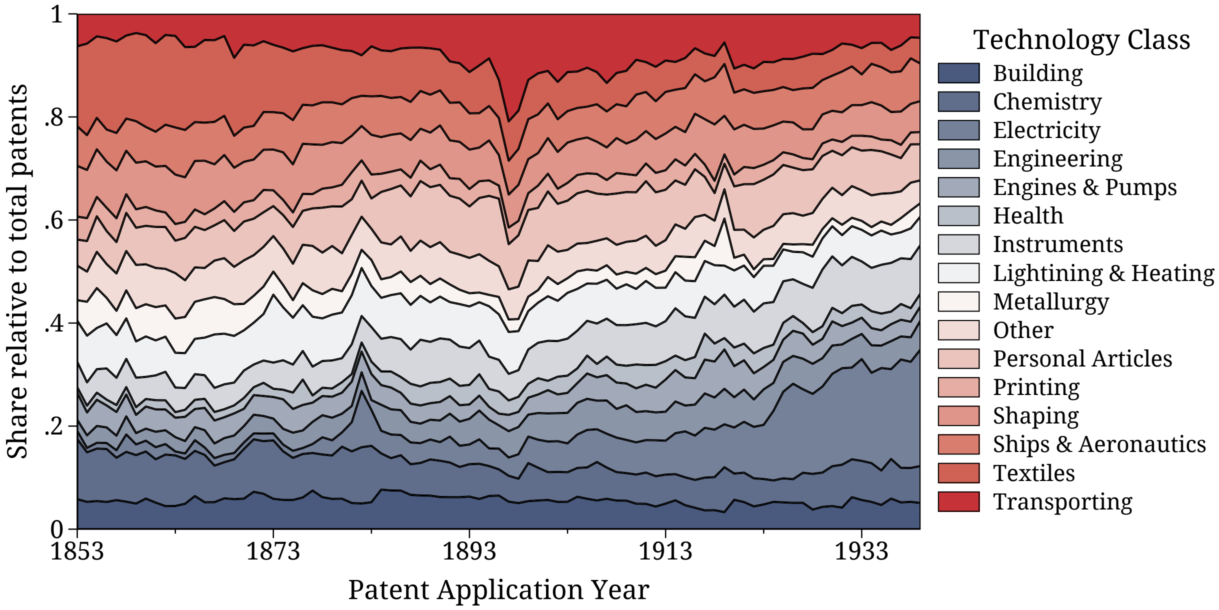
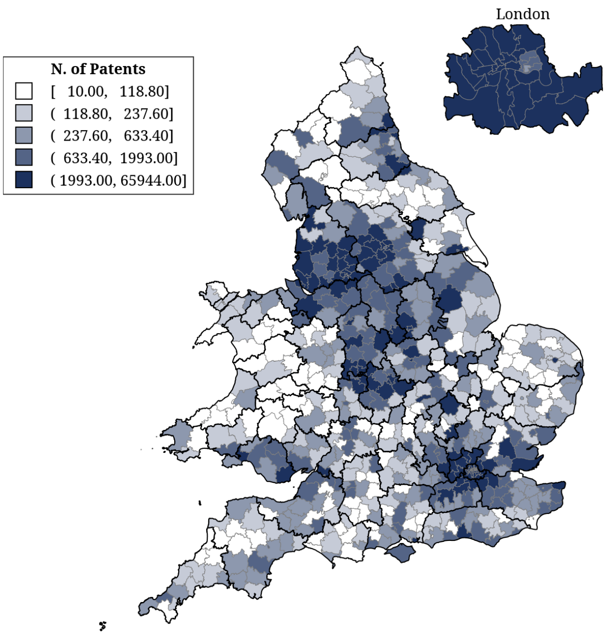


FIGURE B.3. COMPOSITION OF TOTAL PATENTS GRANTED IN THE UNITED KINGDOM ACROSS TECHNOLOGY CLASSES, 1880–1939

Notes. This figure displays the evolution of innovation in Britain across technology classes from 1853–1939. For each year, we compare the share of patents in each class in our database relative to the total number of patents granted in that year. Data for the period 1853–1899 are from the newly digitized universe of patents; data for the period 1900–1939 are made available by the European Patent Office repository.

◄ Back: Appendix 2 – Patents



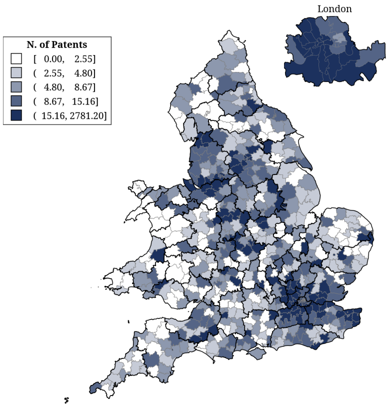
(A) Number of Patents

(B) Patents per Person

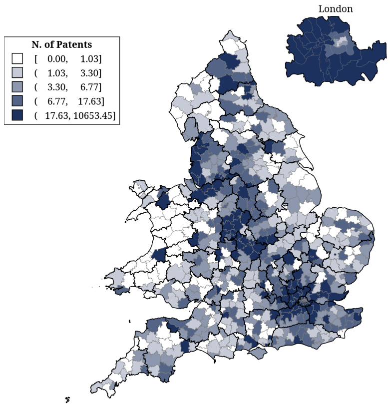
FIGURE B.4. DISTRIBUTION OF PATENTS AND PATENTS PER CAPITA ACROSS DISTRICTS, 1880–1939

Notes. These figures report the intensity of patenting activity across districts over the period 1880–1939. Panel B.4a reports the total number of patents granted; Panel B.4b normalizes this by district population in 1900 and expresses the resulting rate in ‰ units. Districts are displayed at 1900 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.

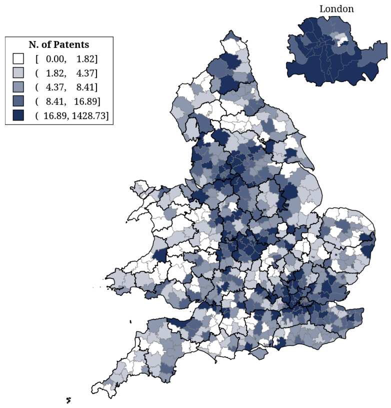
◄ Back: Appendix 2 – Patents



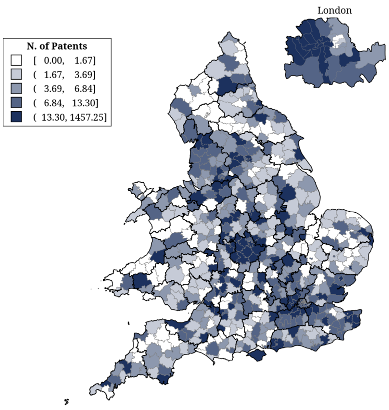
(A) Chemistry



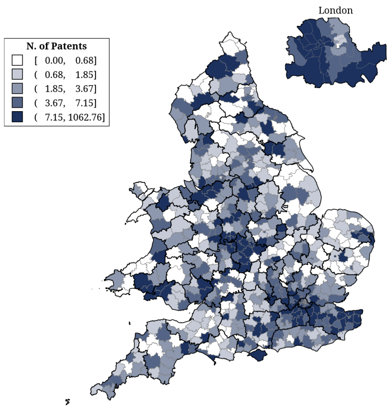
(B) Electricity



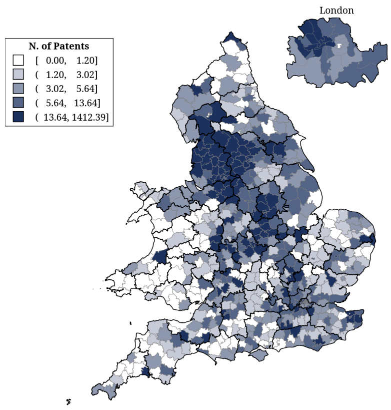
(C) Engineering



(D) Engines, Pumps



(E) Metallurgy



(F) Textiles

FIGURE B.5. DISTRIBUTION OF THE TOTAL NUMBER OF PATENTS GRANTED ACROSS DISTRICTS AND SELECTED TECHNOLOGY CLASSES, 1880–1939

Notes. These figures report the intensity of patenting activity across districts over 1880–1939 for selected technology classes. Districts are displayed at 1900 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.

◄ Back: Appendix 2 – Patents

C Linked International Migrants Sample

This section discusses our methodology to link English and Welsh immigrants in the US to the UK census and presents key statistics on the resulting dataset.

C.1 Sources and Linking Algorithm

We rely on two sources of externally compiled data.[62] 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.[63] 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, besides publicly available information, we also know each individual's recorded name and surname.

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 decade frequency starting in 1851 and 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.

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. We thus try to match British immigrants residing in the US with their entry in the UK census, which records where they come from at a granular geographical level.[64] 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.[65] 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

[62] 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.

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

[64] Since women usually change their name upon marriage, we are unable to match them. This is a common problem in linking algorithms (Abramitzky *et al.*, 2021).

[65] 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.

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[66]

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

   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

$$\overline{\mathcal{M}}^A = \left\{ m_i^A \in \mathcal{M}^A \,\middle|\, \text{Similarity}_i^A = \max_{m_{i'}^A \in \mathcal{M}^A} \text{Similarity}_{i'}^A \right\} \tag{C.2}$$

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

$$\widetilde{\mathcal{M}}_\tau^A = \left\{ m_i^A \in \overline{\mathcal{M}}^A \,\middle|\, \text{Similarity}_i^A \geq \tau \right\} \tag{C.3}$$

   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}:

$$d_{ij} \equiv \widehat{d_{ij}} + \ell p (1 - \widehat{d_{ij}}) \tag{C.4}$$

where

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

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

where m is the number of matching characters, $|i|$ is the length of string i, and 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.[67]

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.

C.2 Internal and External Validation

C.2.1 Matching Statistics

We now present key statistics on the dataset that we assemble. In Figure C.1, we report the matching rate by the number of matches (panel C.1a) and over time (panel C.1b). The matching rate is the ratio between the number of matched individuals and the number of English and Welsh immigrants in the US census. We break down the matching rate by the number of matches every immigrant is associated with. About 40% of the overall immigrant population is matched to one single record in the UK census. Another 10% is matched to two records, and the remaining 50% is matched to three or more records in the UK census. By construction, we can never match someone not appearing in the UK census. This is possible if a child born in, say, 1895 emigrates before 1901, which is the closest subsequent census. In Figure C.1a, we report a corrected matching rate whose denumerator removes these "unmatchable" observations. Overall, 55% of the total number of English and Welsh immigrants is matched to no more than two records in the UK census. This constitutes the baseline sample that we analyze. A 55% matching rate is consistent with standard historical linking algorithms (Abramitzky *et al.*, 2021), although a more precise quantitative assessment is complex because the benchmark statistics refer to intergenerational census linking exercises.

In panel C.1b, we report the matching rate by immigration year. In blue, we report the total number of immigrants; those paired with at least one match are shown in red; the green area reports our baseline sample, which is composed of all those immigrants with no more than two matches. We also impose a quality threshold on names and surnames. Suppose an immigrant is matched to someone born in the

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

same year. In that case, we require both the name and the surname to have a similarity above .85. If an immigrant is matched to someone born either one year before or one year after, we impose a stricter threshold of .9 on both name and surname. We set high thresholds because we are concerned about false positive matches. Following Abramitzky *et al.* (2021), we are thus willing to give up on power to maximize accuracy. In Figure C.2, we report the overall distribution of name (panel C.2a) and surname (panel C.2b) match quality. The solid and dashed red lines superimpose the aforementioned coarse and strict thresholds. The quality distribution is substantially skewed to the right: most matches are of excellent quality. Dropping low-quality ones is, therefore, quantitatively second-order.

Since we match immigrants to the UK census before their migration year, the matching rate decreases over a decade. This is clear from the black line in Figure C.1b, which jumps up at the turn of each decade until 1911. The matching rate before 1881 is relatively low. This is because no census was taken in the UK in 1871. Therefore, we match all those who migrated to the US between 1870 and 1881 to the 1861 census. This mechanically reduces the matching rate, for we cannot match all those born between 1862 and 1881 who migrate during this period. Similarly, the matching rate decreases after 1911. This is because censuses after 1911 are protected by British privacy law. We thus match all those who migrate after 1911 to that census. However, this implies that we cannot match all those who migrated after 1911 and were born after that year. To ensure that our results are not driven by these asymmetries at the edges of the sample, in robustness analyses, we show that restricting the period to the years 1880-1920 does not affect our main findings.

C.2.2 Number of Matches and Observable Characteristics

One plausible concern is that instances of migrants with multiple matches in the UK census are not randomly distributed. This may be due to various reasons (Bailey *et al.*, 2020). First, educated individuals are more likely to report their name and surname in full, with consistent spelling over time. This would generate non-classical measurement error because the matching rate would be higher for a selected subsample of the population. This issue does not seem to be relevant in this case, as the matching rate—*i.e.* the share of immigrants that are *eventually* matched, irrespective of the number of matches—approaches the universe of the observations. Second, the number of matches may not be orthogonal to individual characteristics. This may be the case if wealthier individuals give relatively uncommon names, as documented by Olivetti *et al.* (2020). To assess the severity of this concern, we regress the number of matches on a set of individual-level observable variables observed in the US and UK censuses. Under classical measurement error, we would expect no statistically significant correlation between the number of matches and observable characteristics. Table C.1 reports the estimates thus obtained. We find minimal and marginally significant correlations between the number of matches and individual-level characteristics observed in the US census. The number of matches correlates positively with agriculture and low-skilled employment. However, these correlations are very small: one more match is associated with a .01% increase in the probability of being employed in agriculture. This association is marginally larger for low-skilled manufacturing employment (0.03%). These very low magnitudes are unlikely to affect the results we document in this paper quantitatively. Moreover, notice that most correlations are not statistically significant. Most importantly, we do not find any significant association between the number of matches and the location of English immigrants. This is reassuring because our identification assumption crucially hinges on the variation arising from settlement decisions. We believe this is solid evidence of our linking algorithm and the novel database we assemble.

C.2.3 Plausibility Checks

Official statistics do not contain disaggregated data on emigration outflows. We thus rely on data compiled by Baines (2002) to attempt a validation of our series. These are not, however, based on official reports. The author tabulates emigration figures estimating the "missing population" from enumeration tables published by the census. This methodology yields necessarily approximate results. Moreover, and more crucially for our analysis, Baines (2002) is only able to construct data by counties, a much coarser level of aggregation than registration districts, and report figures on the overall number of overseas emigrants. These include outflows toward Scotland, Ireland, as well as the US and other overseas destinations. Lastly, the data only cover the last three decades of the nineteenth century. These caveats imply that we do not expect this validation exercise to yield unambiguously conclusive results. This notwithstanding, since the US was a major destination for British emigrants, this comparison is useful to gauge the plausibility of our estimates. Figure C.3 reports the correlation between the two datasets. We find a positive and statistically significant correlation between overall out-migration and US emigration, both unconditionally (C.3a), as well as conditioning on county fixed effects (C.3b) and county fixed effects and a time trend (C.3c). Overall, this exercise indicates that our linked dataset is consistent with the previous historical literature.

We now describe an exercise to evaluate the plausibility of the linking algorithm. Building on Abramitzky *et al.* (2021), we construct an intergenerational linked sample of English and Welsh individuals from the population censuses in 1881, 1891, and 1901.[68] The algorithm is very standard: for any given individual in census t, we look at individuals with the same name, surname, and birth year—with a one-year tolerance—who were recorded living in the same parish at year $t+10$. If at least one record is found, we link that individual to that record(s). Otherwise, we look for potential matches in the same district. If no match is found, we leave that individual unmatched. The idea of the exercise is the probability to link

[68] We do not use the 1861 because no census was taken in 1881. This would force us to link individuals in the 1861 census to the 1881 one. However, this imbalance may bias the linking rate between 1861 and 1881. We thus prefer to focus on the censuses for which we have the follow-up taken one decade after.

migrants to the census after they migrate to the US should be lower than for the rest of the (non-migrant) population.

We compare the linking rate across migrants and non-migrants in Figure C.4.[69] The blue bars report the linking rate in the intergenerational sample for migrants; the red bars, instead, refer to non-migrants. We find that non-migrants are more than two times more likely than migrants to be linked to the follow-up census. The matching rate of the intergenerational sample is 42% for non-migrants, but it is only 21% for US migrants. The most conservative interpretation of this result is that it provides an upper bound to the share of false positive matches of the international migrant sample. Suppose that *all* matches in the intergenerational linked sample were true positives. Then, the share of false positive links in the migrant sample would be 40%. In other words, even in this "worst-case" scenario, approximately 60% of the linked migrant matches would be true positives. It should be noted, however, that this represents a somewhat unlikely limit case for intergenerational linkage techniques display substantial type-I error rates (Bailey *et al.*, 2020). Overall, we view this exercise as evidence in favor of the plausibility of the international migrant sample.

To further assess the robustness of the UK-US linkage, we perform one additional linking exercise that excludes individuals that are matched in the intergenerational linked sample from the pool of entries which we attempt to link US migrants with. In other words, we exclude individuals that we would identify as plausibly living in the UK ten years after a given census is taken. In Figure C.5 we compare this linked migrant sample with the baseline dataset that does not apply this trimming to the set of potential matches. These two exercises yield extremely consistent migration flows.

C.3 Return Migration Data

Following the logic explained in section C.1, we construct a linked sample of return migrants. This identifies English and Welsh immigrants in the US in decade d and looks for possible matches in the UK census in decade $d + 1$, using a minor variation on the algorithm described previously. Since the last UK census that we have 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 *et al.*, 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.

[69] To avoid differential attrition due to mortality across migrants and non-migrants, we restrict the sample to individuals that were no older than 40 in the starting census year.

C.4 Summary Statistics and Stylized Facts

The newly developed dataset we construct presents some key novelties compared to available data. It is the first dataset that allows retrieving the origin of US immigrants from England and Wales at a fine level of geographical aggregation during a period of massive international migrations (1880–1930).[70] The dataset's granular—individual—structure allows us to observe several individual characteristics of immigrants at home and in the US. This section briefly discusses key stylized facts that our new data allow us to document.

In Figure C.6, we explore the origin of English and Welsh emigrants to the US over time. Each figure reports the emigration rate normalized by population in 1900, in thousand units. Two patterns emerge. First, substantial cross-sectional heterogeneity exists in the intensity of out-migration across districts throughout the sample period. Second, we find that the intensity of US emigration flows is initially larger in rural districts, especially in the South West and East regions, but this shifts over time toward industrial and urban areas. By the 1910s, the industrialized Lancashire districts featured as a prominent area of emigration. This finding provides a sound quantitative validation of historical—largely anecdotal—evidence (Erickson, 1972; Baines, 2002).

Additionally, we can study the selection patterns of English and Welsh emigrants along two margins. Specifically, we can compare them to (i) the native US population in the areas where they settled and (ii) the non-migrant population in England and Wales who lived in their origin areas. These exercises extend seminal historical work by Baines (2002), who performed a similar exercise using incomplete information from the population censuses. We defer a discussion of selection patterns to the main text. Here, we only note that our dataset is well-suited to study the selection of British emigrants because it identifies individuals before they migrate, thus conveying a complete picture of selection issues during the period.

[70] Similar data-sets have been produced for Norwegian Abramitzky *et al.* (2014) and Swedish (Andersson *et al.*, 2022) immigrants. Our is the first such effort for a major European country: in 1890, the population in England and Wales stood at more than 27 million inhabitants. This compares to approximately 2 million Norwegians and 4.7 million Swedes.

C.5 Tables

	Dep. Var.: Number of Matches		
	(1)	(2)	(3)
Panel A. Occupations			
Agriculture	0.003	0.004*	0.003**
	(0.001)	(0.001)	(0.001)
Low-Skilled Manufacture	0.013**	0.011**	0.008**
	(0.003)	(0.003)	(0.002)
High-Skilled Manufacture	0.006	0.007	0.008*
	(0.003)	(0.003)	(0.003)
Professionals	–0.001	–0.001	–0.001
	(0.001)	(0.001)	(0.001)
Public Administration	–0.004**	–0.003*	–0.002
	(0.001)	(0.001)	(0.001)
Manager	–0.000	–0.000	0.000
	(0.000)	(0.000)	(0.000)
Service Worker	0.001	0.001	0.002
	(0.001)	(0.001)	(0.001)
Panel B. Origin			
Northeast	–0.004		
	(0.006)		
Midwest	0.004		
	(0.004)		
South	–0.002		
	(0.001)		
West	0.001		
	(0.002)		
State FE	No	Yes	No
County FE	No	No	Yes
Year FE	No	Yes	Yes

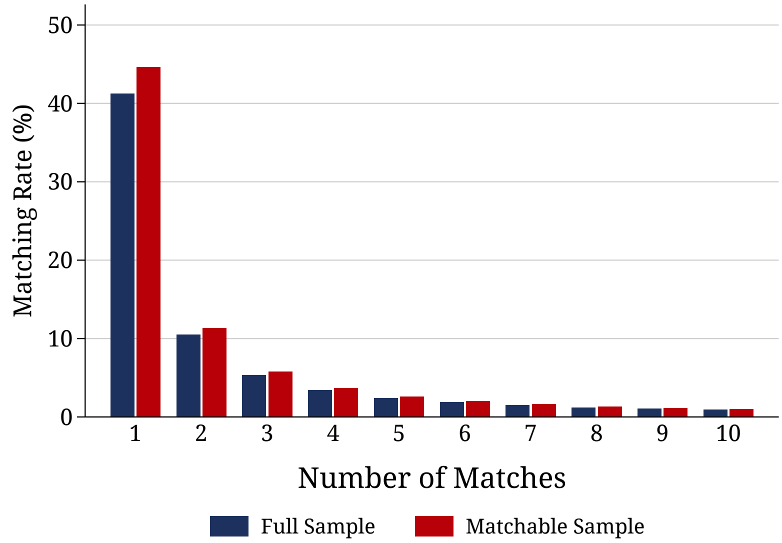
TABLE C.1. CORRELATION BETWEEN NUMBER OF MATCHES AND OBSERVABLE CHARACTERISTICS

Notes. This table reports the correlation between observable characteristics of British immigrants in the US census and the number of matches in the linked sample database. In each row, the table displays the correlation between the number of matches and an indicator equal to one if for immigrants that correspond to the row variable and zero otherwise. The sample is restricted to the set of matches we effectively use in the analysis. Column (1) reports unconditional correlations; column (2) includes state and census decade fixed effects; column (3) adds county fixed effects. In Panel A, the characteristics are the occupations; in Panel B, the variables are the Census Bureau region of residence. Standard errors, clustered at the county level, are shown in parentheses.

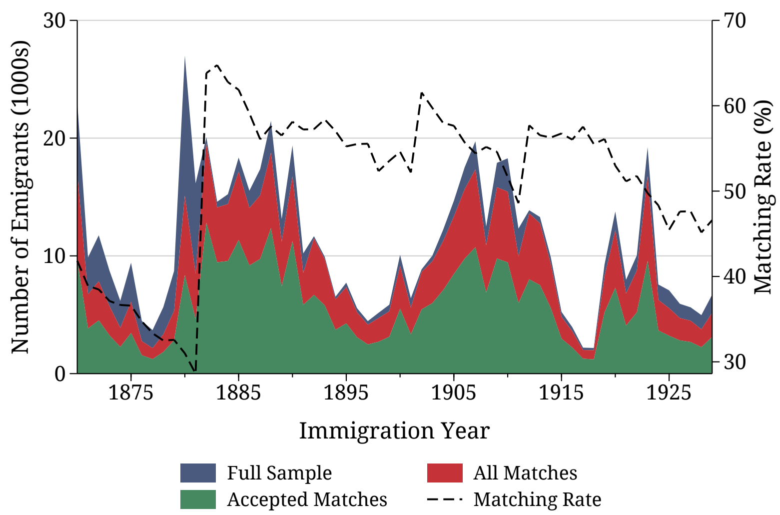
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$.

◄ Back: Data ◄ Back: Appendix 3 – Migrants

C.6 Figures



(A) Matching Rate by Number of Matches

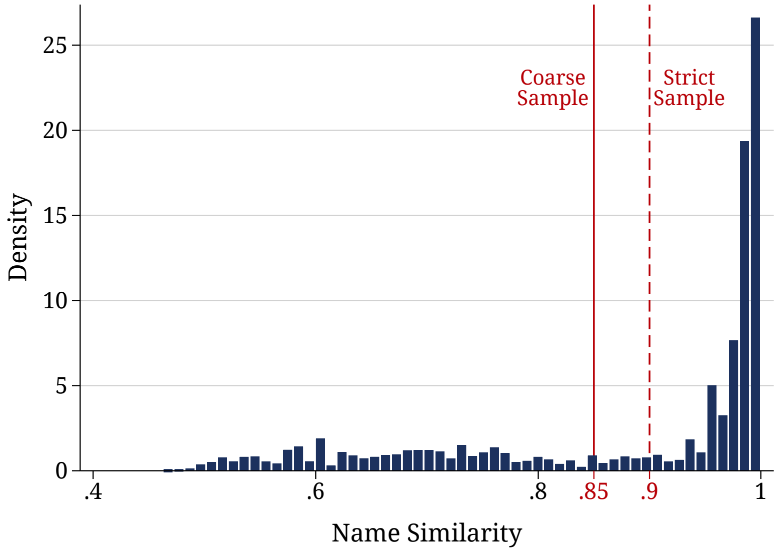


(B) Matching Rate Over Time

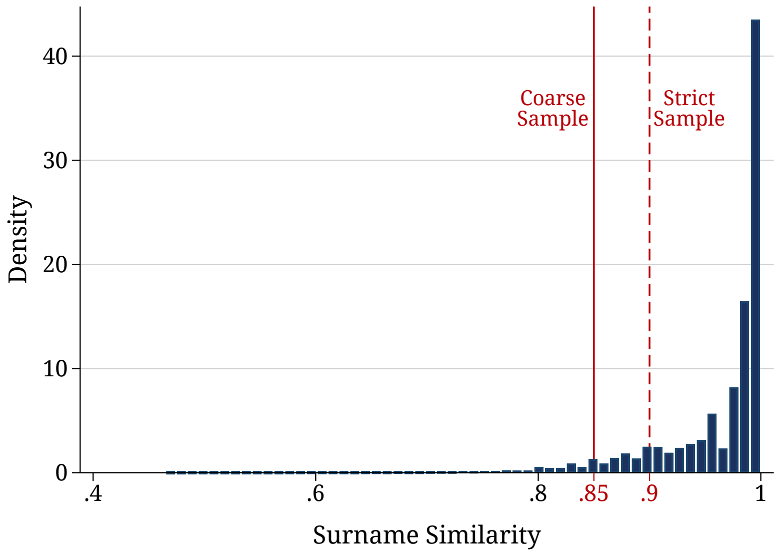
FIGURE C.1. SHARE OF BRITISH IMMIGRANTS IN THE US CENSUS MATCHED TO THE UK CENSUS

Notes. These figures report the share of English and Welsh immigrants recorded in the US census that we match to the UK census. Panel C.1a plots the share of records that we match to the UK census and whose match quality is such that we retain it in the linked sample, broken down by the number of matches. In the baseline sample, we keep records with no more than two matches. Blue bars report ratios relative to the entire number of immigrants, and red bars restrict the set of immigrants to those we can match. Panel C.1b reports the matching rate over time. The blue area reports the total number of US immigrants, the red area reports the entire number of matches we obtain, and the green area reports the matches that eventually enter our baseline linked sample. The black dashed line on the right y-axis is the ratio between the green and the blue areas.

◄ Back: Appendix 3 – Migrants



(A) Name Match Quality

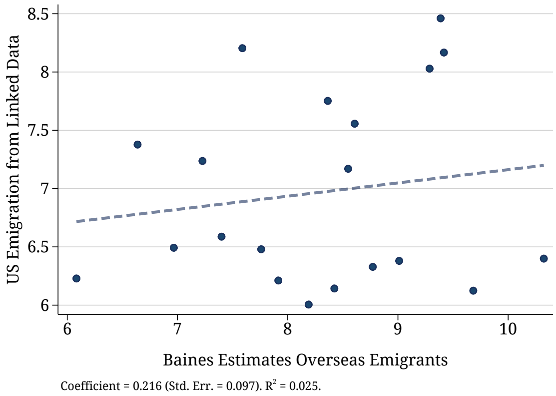


(B) Surname Match Quality

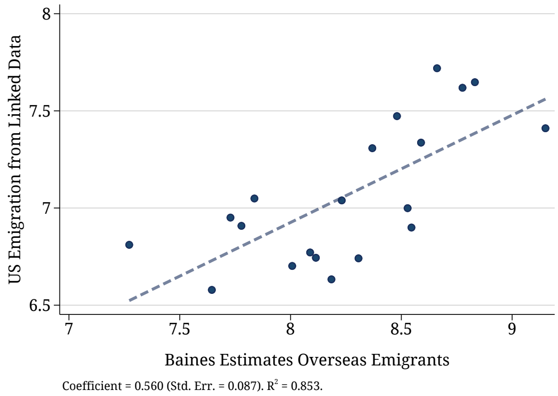
FIGURE C.2. QUALITY OF MATCHES IN THE COMPLETE LINKED SAMPLE: NAMES AND SURNAMES

Notes. The figures report 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 C.2a reports the distribution of the name similarity; Panel C.2b refers to surnames. The vertical lines mark the quality thresholds we impose for a match to be part of the final linked sample.

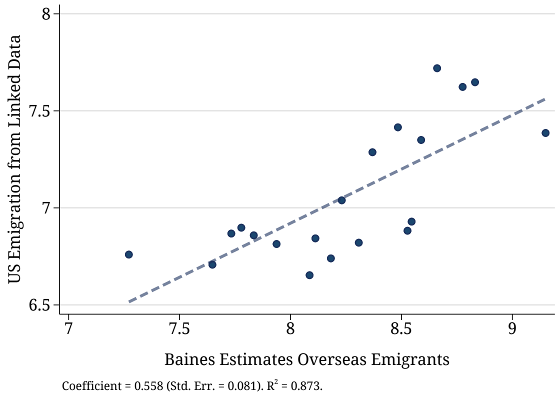
◄ Back: Appendix 3 – Migrants



(A) Unconditional Correlation



(B) County Fixed Effects



(C) County Fixed Effects and Decade Trend

FIGURE C.3. COMPARISON BETWEEN LINKED DATA AND ESTIMATES FROM BAINES

Notes. These figures report the correlation between county-level out-migration measured using our linked emigrant sample and data from (Appendix 1 Baines, 2002). The dataset listed by the author is at the county level at a decade time frequency between 1870 and 1900 and reports the overall number of overseas emigrants. Thus, it conflates emigration to Scotland, Ireland, European, and trans-oceanic out-flows. In panel C.3a we correlate the two series; in panels C.3b, we control for county fixed effects; in panel C.3c, we include a decade time trend. Observations are weighted by county-level population in 1880. Each graph reports in note the regression coefficient, along with its standard error, and the coefficient of determination of each regression.

◄ Back: Appendix 3 – Migrants

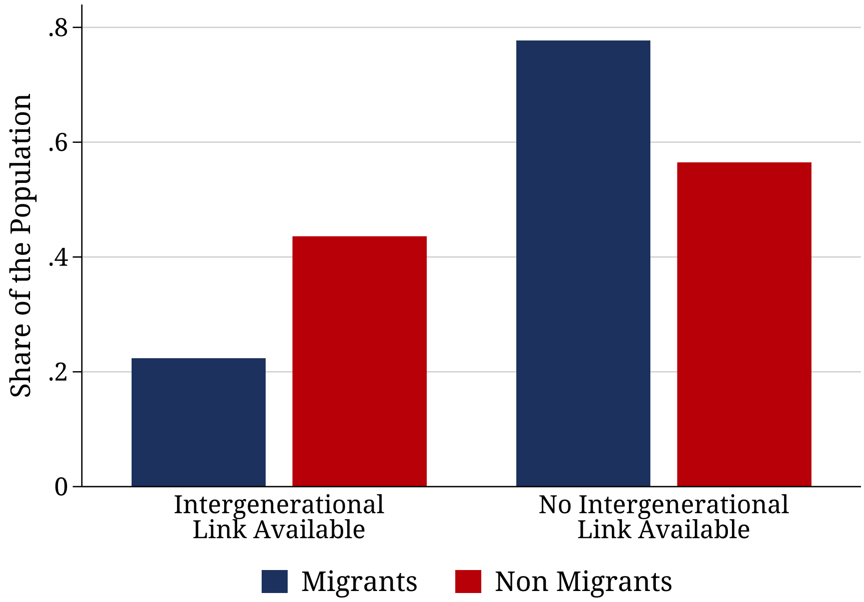
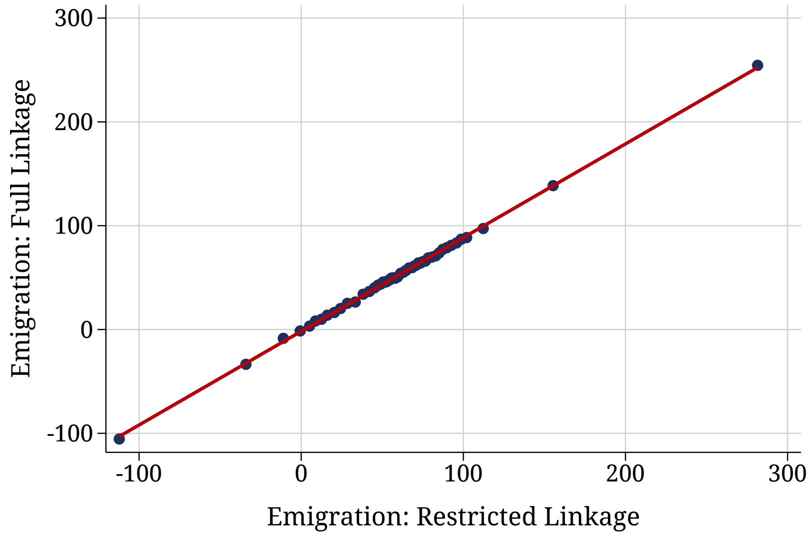


FIGURE C.4. FALSIFICATION EXERCISE OF THE INTERGENERATIONAL LINKED SAMPLE

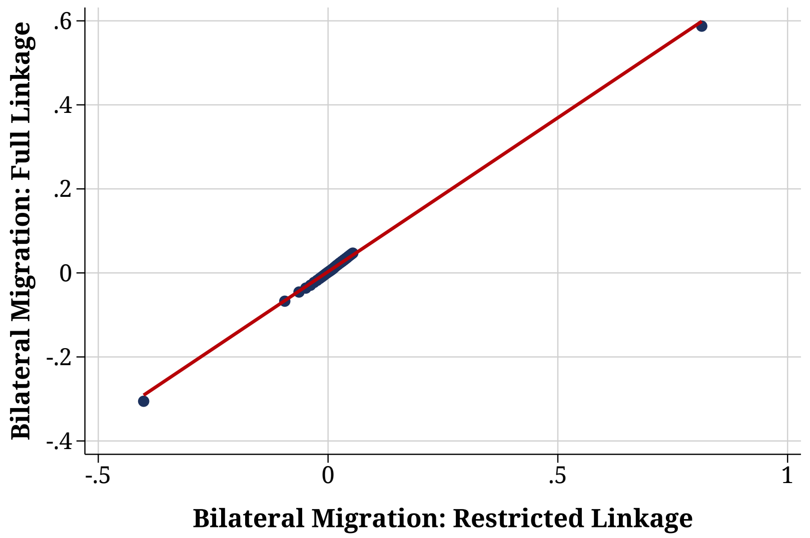
Notes. This figure reports the matching rate in the intergenerational linked sample. The blue bars report the matching rate among individuals that are identified as US migrants in the UK census. The red bars refer to individuals that are not identified as US migrants in the UK census. We exclude the 1861 census because there is no 1871 census in the UK and the intergenerational sample is unbalanced over the years 1861–1881. The intergenerational linked sample includes individuals with one single match obtained at the parish or at the district level. We exclude individuals over forty years old because differential mortality across the life cycle may impact the linking rate of the intergenerational sample.

◄ Back: Data ◄ Back: Appendix 3 – Migrants



Notes. Coefficient = 0.906 (Clust. Std. Err. = 0.011). R^2 = 0.991.

(A) District-Level Migration Flows



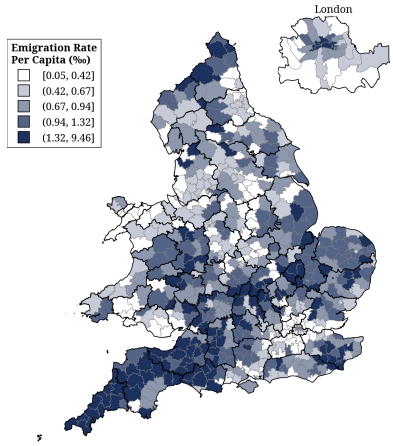
Notes. Coefficient = 0.866 (Clust. Std. Err. = 0.007). R^2 = 0.872.

(B) District-County Migration Flows

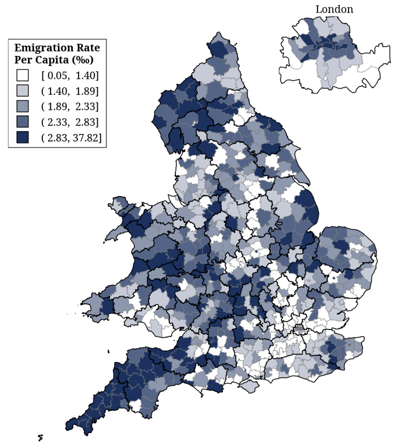
FIGURE C.5. COMPARISON BETWEEN LINKED SAMPLES: FULL AND RESTRICTED

Notes. This figure compares the migration flows of the baseline UK-US linked sample, on the *y*-axis, with those obtained by restricting the pool of potential matches to those entries in the UK census that are not matched to any individual in the following census decade, on the *x*-axis. Panel C.5a reports the district-level US emigration; the unit of observation is a district, observed at a decade frequency between 1870 and 1900. The figure includes district and decade fixed effects. The reported standard error is clustered at the district level. Panel C.5b reports the district-county-level US migration flows; the unit of observation is a district-county pair, observed at a decade frequency between 1870 and 1900. The figure includes district, county, and decade fixed effects. The reported standard error is clustered at the district-by-county level.

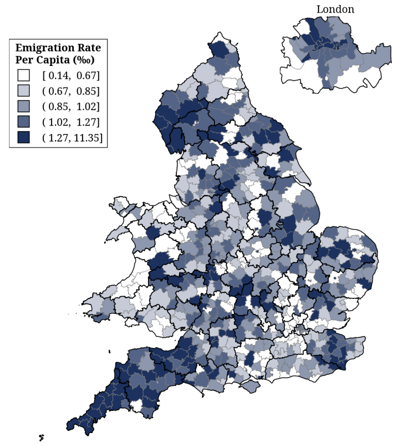
◄ Back: Data ◄ Back: Appendix 3 – Migrants



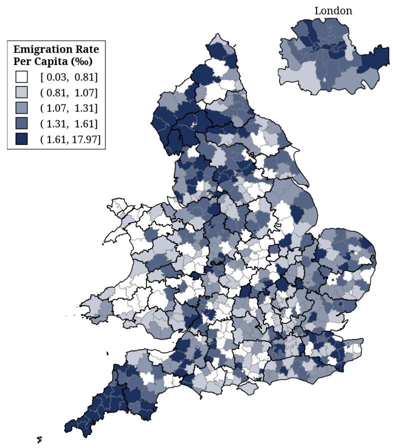
(A) 1871-1880



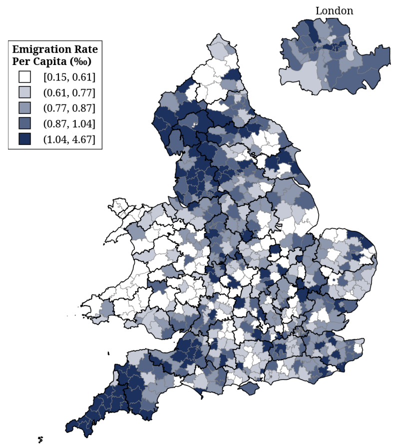
(B) 1881-1890



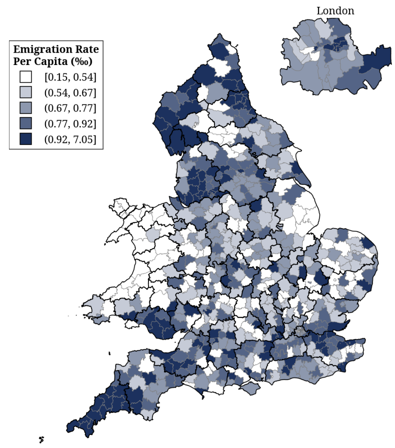
(C) 1891-1900



(D) 1901-1910



(E) 1911-1920



(F) 1921-1930

FIGURE C.6. DISTRIBUTION OF THE EMIGRATION RATE ACROSS DISTRICTS, 1871–1930

Notes. These figures report the distribution of US emigrants across districts in England and Wales over the period 1871–1939 by decade. Data are from the matched emigrants' sample. The number of emigrants in each decade is normalized by population in 1900 and is expressed in ‰ units. Districts are displayed at their 1900 borders. Out-migration is also cross-walked to consistent historical borders. Lighter to darker blues indicate higher emigration rates.

◄ Back: Historical Background ◄ Back: Data ◄ Back: Appendix 3 – Migrants

D Additional Results

This section presents in some detail several additional results that are mentioned in passing in the main text.

D.1 Trade-Induced Technology Transfer

Our favored explanation of the return innovation result is that migrants facilitate the flow of knowledge between the areas where they settle and those they originate from. We argue that those flows are fostered by the diffusion of information and by market integration. This section presents one more piece of evidence in this direction. We focus on international trade as a measure of bilateral market integration. Previous research documents that trade fosters innovation, either because of increased import competition (Bloom *et al.*, 2016; Autor *et al.*, 2020), export opportunities (Bustos, 2011; Atkin *et al.*, 2017; Aghion *et al.*, 2018), access to intermediate inputs (Juhász and Steinwender, 2018), and increased market size (Coelli *et al.*, 2022).[71] In our analysis, we interpret trade as a means of facilitating technology transfer between the UK and the US, following Aleksynska and Peri (2014) and Ottaviano *et al.* (2018).

We consider a major shock to trade flows between the US and the UK: the 1930 Smoot-Hawley Act. The Act was a major trade policy reform enacted in response to the Great Depression (Eichengreen, 1986; Crucini, 1994). Importantly for our setting, the Act did not establish a uniform tariff rate. Instead, as we report in Table A.3, tariffs vastly differed across industries before and after the shock. We leverage this variation, interacted with the before-Act knowledge exposure in a difference-in-differences setting.[72] The key idea that underlies this approach is that if migration linkages generate return innovation flows through international trade, then an increase in trade costs is expected to reduce patenting in the UK in the sectors that (i) districts were more exposed to, through migrations, and (ii) were targeted by the tariff increase.

We thus estimate the following double differences model separately for protected and non-protected industries:

$$\text{Patents}_{ik,t} = \alpha_{i\times k} + \alpha_t + \sum_{h=-a}^{b} \beta_h \times \left[1\,(t = 1931 + h) \times \text{Knowledge Exposure}_{ik}\right] + \varepsilon_{ik,t} \qquad \text{(D.1)}$$

Where i, k, and t respectively denote a district, technology class, and year, and Knowledge Exposure$_{ik}$ is the average sector-level knowledge exposure in the decade before the Act (1920–1930). In the baseline analysis, an industry is protected if its tariff rate increases by more than 50 p.p. between 1925–1930 and

[71] Shu and Steinwender (2019) provide a critical assessment of the literature studying the effect of international trade on innovation.

[72] We first map sectors defined in the Act to technology classes. We then assign one class to the treatment group if its average *ad valorem* import duty changes by more than fifty percentage points between 1925–1930 and 1931–1936. Yearly tariff rates have been digitized from the *Statistical Abstract of the United States*.

1931–1935. Then, we estimate the triple-differences specification that compares treated and non-treated industries:

$$\begin{aligned}\text{Patents}_{ik,t} &= \alpha_{i\times k} + \alpha_{i\times t} + \alpha_{k\times t} + \\ &+ \sum_{h=-a}^{b} \beta_h \times \left[\text{Tariff}_k \times 1\,(t = 1931 + h) \times \text{Knowledge Exposure}_{ik}\right] + \varepsilon_{ik,t}\end{aligned} \tag{D.2}$$

where Tariff_k is an indicator returning value one for protected industries and zero otherwise.

In columns (1–2) of Table D.3 we report the results of model (D.1). Column (1) presents the estimated coefficient for non-protected industries, while column (2) focuses on protected ones. We find no effect for the former and a negative effect for the latter. This is confirmed when looking at the associated flexible specification, reported in Figure D.1. This also provides evidence supporting the parallel trends assumption for the two groups of technology classes. In columns (3–5), we report the estimates of the triple differences model (D.2). We consider three possible threshold values of the increase in the tariff rate after the Act to define a protected sector (10%, 30%, and 50%). All yield quantitatively similar estimates. Note, however, that the estimated ATE reassuringly increases in absolute magnitudes for larger tariff increases.

The analysis suggests that trade—which we interpret as a proxy for market integration—is a relevant channel through which migration ties generate knowledge flows and technology transfer. However, it is worth noting that the magnitude of the estimated treatment effects of the tariff reform on UK innovation is modest, despite the large increase in tariff duties. We thus interpret trade as one additional, although plausibly not the pivotal, factor driving return innovation.

D.2 Selection of British Migrants

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 US and their integration into the US. To construct these statistics, we leverage the novel linked sample that allows us to observe individual-level characteristics before emigrants left—in the UK census—and after they settled—in the US census.

Table D.4 compares UK emigrants with the non-migrant population. 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. Migrants are less likely to work in agriculture and as professionals. They are, however, more likely to be employed in industrial sectors, such as textiles and metallurgy. This overall confirms the historical analysis of Baines (2002). Emigrants mainly originated from the North West, including Lancashire, and South West, chiefly, Devon and Cornwall. Similar patterns emerge when looking at return migrants, who are even less likely to be employed as agricultural workers. Return rates in high-emigration areas of the South West appear low compared to the rest of the country, while they are very high in the London area.

In Table D.5, we compare English and Welsh immigrants to the rest of the US population. Column (1) refers to natives, and columns (2) and (5) refer to emigrants and return migrants, respectively. In columns (3) and (6), we compute the difference between natives and emigrants and natives and return migrants, respectively. UK immigrants differ substantially from the rest of the US population: they are less likely to work in agriculture and as civil servants. By comparison, they are more likely to be employed in metallurgy, textiles, and trade. This aligns well with evidence by Erickson (1972), who argues that English immigrants in the US tended to specialize in industries where they had a comparative advantage. Similar patterns emerge for return migrants. Regarding their geographical distribution, UK immigrants settled most commonly in the New England and Mid-Atlantic regions.

D.3 Long-Run Effect of Return 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:

$$\text{Patents}_{ik,t} = \alpha_{i\times k} + \alpha_t + \sum_{\tau\in\mathcal{T}} \beta^\tau \left[\text{Knowledge Exposure}_{ik} \times 1\,(t = \tau \,|\, t = \tau + 1)\right] + \varepsilon_{ik,t} \tag{D.3}$$

where i, k, and t denote a district, technology class, and year, respectively. In this setting, we have $t \in [1940, 2015]$. The term 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 D.3, 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 D.9 we re-estimate model (D.3), sector-by-sector, by decade. Compared to (D.3), 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 almost all sectors and only a few display significant coefficients after the 1980s.

D.4 Further Additional Results

D.4.1 Out-Migration and the Volume of Innovation

The main analysis concentrates on the effect of knowledge exposure on the direction of innovation. Knowledge exposure leverages variation in specialization across US counties and bilateral flows between UK districts and US counties. In this section, we briefly comment on the effect of out-migration on the *volume* of innovation.

We estimate variations on the following model:

$$\text{Patents}_{i,t} = \alpha_i + \alpha_t + \beta \times \text{US Emigrants}_{i,t} + \varepsilon_{i,t} \tag{D.4}$$

where US Emigrants$_{i,t}$ is the total number of emigrants from district i in decade t. As in the main text, we instrument total out-migration flows with the shift-share instruments constructed using railway-based and leave-out immigration shocks. Compared to the model estimated in the main text, endogeneity concerns in (D.4) are severe. However, if the instruments are valid, then the estimated β coefficients measured the causal effect of out-migration on patenting. A perhaps more crucial concern in regression (D.4) is that we do not have information on emigration to countries other than the US. Suppose emigration rates to, say, Australia or Canada (the second and third most common destinations) were correlated with US out-migration. In that case, we may fail to single out the effect of out-migration.

With these caveats in mind, in Table D.6, we report the estimates of regression (D.4). In panel A, columns (1–3), we report the correlation between measured out-migration and patenting, while columns (4–6) and (7–9) display the reduced form association with, respectively, the railway-based and the leave-out instruments. In panel B we report the 2SLS estimates. We find that the contemporaneous effect of out-migration on innovation is negative. This is reasonable given that out-migration entails a loss of human capital, which, in the light of the selection analysis, was probably relatively skilled and is consistent with the "brain drain" literature. Once we lag emigration by one decade, however, we find a positive effect. This sign reversal is robust across the two instruments in the reduced form and the two-stage least-square estimates. It is tempting to interpret it as evidence of "brain gain", that is, that return innovation increases the volume of innovation (Docquier and Rapoport, 2012). While the results are consistent with this interpretation, they are not conclusive because of the caveats that underlie this exercise.

D.4.2 Assortative Matching

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:

$$\text{Innovation Similarity}_{ij,t} \equiv \frac{\mathbf{P}_{i,t}^{\intercal}\mathbf{P}_{j,t}}{\|\mathbf{P}_{i,t}\| \cdot \|\mathbf{P}_{j,t}\|} = \frac{\sum_k p_{ki,t} p_{kj,t}}{\sqrt{\sum_k p_{ki,t}^2}\sqrt{\sum_k p_{kj,t}^2}} \leq 1 \tag{D.5}$$

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:

$$\text{Emigrants}_{i\to j,t} = \alpha_{i\times j} + \alpha_t + \beta \times \text{Innovation Similarity}_{ij,t} + X_{ij,t}^{\intercal}\Gamma + \varepsilon_{ij,t} \tag{D.6}$$

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 D.2. We find no correlation between the similarity of innovation portfolios across county districts and the migration flow between them. This holds irrespective of whether we take the contemporaneous similarity (columns 1–2) or if we lag by one (columns 3–4) or two (columns 5–6) decades. Notably, the standardized beta coefficient of the innovation similarity term is always minimal in magnitude. This suggests that assortative matching based on innovation similarity between origin and destination places is probably not a significant threat to a causal interpretation of our estimates. This notwithstanding, since the similarity of innovation portfolios is measured with error, we do not claim that we can exclude it *tout court*.

D.5 Tables

	Baseline	Excluding States in...			
	(1)	(2)	(3)	(4)	(5)
		Northeast	Midwest	South	West
$I^{\text{Rail}}_{t-1} \times$ Immigrant Flow$_{t-1}$	0.372***	0.399***	0.198**	0.461**	0.252**
	(0.102)	(0.111)	(0.097)	(0.198)	(0.101)
I^{Rail}_{t-1}	0.845	4.014	-4.288	-17.024***	3.374
	(2.765)	(2.775)	(2.798)	(5.918)	(2.775)
County FE	Yes	Yes	Yes	Yes	Yes
Decade FE	Yes	Yes	Yes	Yes	Yes
N. of Counties	2759	2543	1742	1513	2479
N. of Observations	17308	15803	10919	9222	15980
R^2	0.905	0.903	0.921	0.880	0.915
Mean Dep. Var.	79.842	74.284	55.019	132.174	72.101

TABLE D.1. ZERO-STAGE REGRESSIONS BETWEEN IMMIGRANT SHARES AND RAILWAY ACCESS

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 brackets.

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

◄ Back: Appendix 5 – Robustness

	Contemporaneous		10 Years Lag		20 Years Lag	
	(1)	(2)	(3)	(4)	(5)	(6)
Innovation Similarity	0.083	201.876				
	(2.847)	(155.968)				
Innovation Similarity$_{t-1}$			0.419	333.940		
			(2.800)	(205.476)		
Innovation Similarity$_{t-2}$					1.370	-172.428
					(2.485)	(218.951)
District-County FE	Yes	Yes	Yes	Yes	Yes	Yes
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
N. of District-Counties	1743283	32176	1665941	31383	1505060	29948
N. of Observations	9029476	88636	7266084	86789	5476833	83220
Sample	All	Non-Zero	All	Non-Zero	All	Non-Zero
R^2	0.473	0.675	0.553	0.675	0.662	0.676
Mean Dep. Var.	0.022	1.617	0.027	1.635	0.034	1.670
Std. Beta Coef.	0.000	0.010	0.000	0.018	0.001	-0.010

TABLE D.2. BRITISH IMMIGRANTS ASSORTATIVE MATCHING ACROSS US COUNTIES

Notes. This table reports the association between the similarity of innovation activity and migration flows between US counties-UK districts pairs. The unit of observation is a county-district pair, observed at a decade frequency between 1870 and 1920. The dependent variable is the number of emigrants that leave the given district and settle in the given county. The independent variable is the similarity of the innovation portfolios between the county and the district. The innovation similarity is computed as the cosine distance of the respective patent portfolios over the decade. Columns (1), (3), and (5) report results for the universe of county-district pairs; columns (2), (4), and (6) restrict to pairs with non-zero migration flows. Columns (1) and (2) estimate the contemporaneous correlation; in columns (3) and (4), innovation similarity appears with a one-decade lag; in columns (5) and (6), it is included with a two-decade lag. Regressions include district-by-county and decade fixed effects. Standard errors, clustered at the district level, are reported in parentheses.

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

◄ Back: Results ◄ Back: Appendix 4 – Additional Results

	Double Differences		Triple Differences		
	(1)	(2)	(3)	(4)	(5)
	Not Protected	Protected	+10%	+30%	+50%
Knowledge Exposure × Post	-0.040	-0.463**			
	(0.063)	(0.219)			
Knowledge Exposure × Post × Protected (+10%)			-0.469***		
			(0.078)		
Knowledge Exposure × Post × Protected (+30%)				-0.478*	
				(0.269)	
Knowledge Exposure × Post × Protected (+50%)					-0.685***
					(0.206)
Year FE	Yes	Yes	–	–	–
District-Year FE	No	No	Yes	Yes	Yes
District-Class FE	Yes	Yes	Yes	Yes	Yes
Class-Year FE	No	No	Yes	Yes	Yes
N. of District-Class	632	632	632	632	632
N. of Observations	63200	37920	101120	101120	101120
R^2	0.653	0.563	0.713	0.713	0.713
Mean Dep. Var.	2.125	1.260	1.801	1.801	1.801
Std. Beta Coef.	0.000	0.000	0.000	0.000	0.000

TABLE D.3. DOUBLE AND TRIPLE DIFFERENCES EFFECT OF THE SMOOT-HAWLEY ACT ON INNOVATION

Notes. This table reports the estimated effect of an increase in the US tariff rate on innovation in Britain. The unit of observation is a district-technology class pair observed at a yearly frequency between 1920 and 1939. The dependent variable is the number of patents by district technology class. In columns (1–2), the independent variable is the interaction between knowledge exposure over 1910–1920 and a post-reform (1930) indicator variable. The regression in column (1) is estimated over technology classes not targeted by the Act; in column (2), we focus on classes that the Act targets. We define a class as "targeted" if its average tariff rate increases by more than 50% after the Smoot-Hawley Act. In columns (3), (4), and (5), the treatment interacts the previous one with an indicator that returns value one for technology classes whose tariff rates increases by more than, respectively, 10%, 30%, and 50% after 1930. Regressions (1–2) are thus double-difference designs; regressions (3–5) are triple-difference designs. Consequently, in columns (1–2), we include district-by-class and year fixed effects, while in columns (3–5), we add district-by-year and technology-by-year fixed effects. Standard errors, clustered at the district level, are reported in parentheses.

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

◄ Back: Appendix 4 – Additional Results

	Non Migrants	Emigrants			Return Migrants		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Mean	Mean	Difference	Std. Err.	Mean	Difference	Std. Err.
Panel A. Employment (Dependent variable = 1 if individual employed in:)							
Agriculture	0.281	0.271	-0.009***	(0.001)	0.252	-0.028***	(0.002)
Chemicals	0.008	0.008	-0.001**	(0.000)	0.009	0.001**	(0.000)
Construction	0.141	0.142	0.001	(0.001)	0.145	0.004***	(0.002)
Engineering	0.138	0.138	-0.000	(0.001)	0.148	0.010***	(0.002)
Liberal Profession	0.035	0.032	-0.002***	(0.000)	0.036	0.002*	(0.001)
Metallurgy	0.029	0.034	0.005***	(0.001)	0.032	0.003***	(0.001)
Other Manufacturing	0.074	0.074	-0.000	(0.001)	0.074	-0.001	(0.001)
Public Administration	0.030	0.028	-0.001***	(0.000)	0.032	0.002***	(0.001)
Textiles	0.090	0.099	0.009***	(0.001)	0.082	-0.008***	(0.001)
Trade	0.072	0.079	0.007***	(0.001)	0.078	0.007***	(0.001)
Transport	0.097	0.090	-0.007***	(0.001)	0.103	0.006***	(0.001)
Utilities	0.007	0.006	-0.000	(0.000)	0.009	0.002***	(0.000)
Panel B. Region of Residence (Dependent variable = 1 if individual lives in:)							
East	0.102	0.086	-0.016***	(0.001)	0.089	-0.014***	(0.001)
East Midlands	0.065	0.057	-0.007***	(0.000)	0.058	-0.006***	(0.001)
London	0.132	0.129	-0.003***	(0.001)	0.139	0.006***	(0.001)
North East	0.067	0.070	0.003***	(0.000)	0.070	0.003***	(0.001)
North West	0.179	0.194	0.015***	(0.001)	0.199	0.020***	(0.001)
South East	0.120	0.110	-0.009***	(0.001)	0.117	-0.003***	(0.001)
South West	0.063	0.085	0.022***	(0.001)	0.065	0.002***	(0.001)
Wales	0.070	0.064	-0.006***	(0.000)	0.069	-0.001	(0.001)
West Midlands	0.114	0.110	-0.004***	(0.001)	0.108	-0.006***	(0.001)
Yorkshire	0.088	0.094	0.006***	(0.001)	0.087	-0.001*	(0.001)

TABLE D.4. SELECTION OF US EMIGRANTS COMPARED TO THE REST OF THE BRITISH POPULATION

Notes. This table compares observable individual characteristics of US emigrants with the rest of the British population. In each row, we define a dummy variable equal to one for individuals in the given employed in the given sector (Panel A) or residing in the given division (Panel B) and compute the average for non-migrants (column 1), emigrants (column 2), and return migrants (column 5). Columns (3) and (6) report the difference between columns (1) and, respectively, columns (2) and (5). Robust standard errors are reported in columns (4) and (7).

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

◄ Back: Historical Background ◄ Back: Appendix 4 – Additional Results

	US Population	Immigrants			Return Migrants		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Mean	Mean	Difference	Std. Err.	Mean	Difference	Std. Err.
Panel A. Employment (Dependent variable = 1 if individual employed in:)							
Agriculture	0.215	0.121	-0.094***	(0.001)	0.129	-0.086***	(0.001)
Chemicals	0.006	0.010	0.004***	(0.000)	0.006	-0.001***	(0.000)
Construction	0.044	0.096	0.052***	(0.001)	0.087	0.042***	(0.001)
Engineering	0.434	0.199	-0.235***	(0.001)	0.262	-0.172***	(0.002)
Liberal Profession	0.042	0.078	0.036***	(0.001)	0.063	0.021***	(0.001)
Other Manufacturing	0.076	0.159	0.083***	(0.001)	0.148	0.072***	(0.001)
Public Administration	0.014	0.009	-0.005***	(0.000)	0.008	-0.006***	(0.000)
Textiles	0.015	0.076	0.061***	(0.001)	0.080	0.066***	(0.001)
Trade	0.069	0.104	0.035***	(0.001)	0.092	0.023***	(0.001)
Transport	0.056	0.087	0.031***	(0.001)	0.085	0.029***	(0.001)
Utilities	0.028	0.059	0.031***	(0.001)	0.041	0.013***	(0.001)
Panel B. Region of Residence (Dependent variable = 1 if individual lives in:)							
East North Central	0.205	0.210	0.005***	(0.001)	0.192	-0.014***	(0.001)
East South Central	0.087	0.008	-0.079***	(0.000)	0.008	-0.078***	(0.000)
Mid Atlantic	0.208	0.350	0.143***	(0.001)	0.365	0.157***	(0.002)
Mountain	0.030	0.058	0.028***	(0.000)	0.062	0.032***	(0.001)
New England	0.068	0.165	0.097***	(0.001)	0.187	0.120***	(0.001)
Pacific	0.054	0.101	0.047***	(0.001)	0.072	0.018***	(0.001)
South Atlantic	0.130	0.026	-0.104***	(0.000)	0.023	-0.107***	(0.000)
West North Central	0.123	0.067	-0.055***	(0.001)	0.077	-0.045***	(0.001)
West South Central	0.095	0.014	-0.081***	(0.000)	0.013	-0.082***	(0.000)

TABLE D.5. SELECTION OF BRITISH IMMIGRANTS COMPARED TO THE REST OF THE US POPULATION

Notes. This table compares observable individual characteristics of British immigrants with the rest of the US population. In each row, we define a dummy variable equal to one for individuals in the given employed in the given sector (Panel A) or residing in the given division (Panel B) and compute the average for non-migrants (column 1), immigrants (column 2), and return migrants (column 5). Columns (3) and (6) report the difference between columns (1) and, respectively, columns (2) and (5). Robust standard errors are reported in columns (4) and (7).

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

◄ Back: Historical Background ◄ Back: Appendix 4 – Additional Results

	Dependent Variable: Number of Patents								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. OLS Estimates									
	Measured US Emigration			Railway Instrument			Leave-Out Instrument		
US Emigrants$_t$	-1.453***								
	(0.303)								
US Emigrants$_{t-1}$		0.951***							
		(0.258)							
US Emigrants$_{t-2}$			-0.702						
			(0.646)						
Railway-Predicted Emigrants$_t$				-0.141***					
				(0.038)					
Railway-Predicted Emigrants$_{t-1}$					0.283***				
					(0.074)				
Railway-Predicted Emigrants$_{t-2}$						0.212**			
						(0.107)			
Leaveout-Predicted Emigrants$_t$							-0.307***		
							(0.093)		
Leaveout-Predicted Emigrants$_{t-1}$								0.473***	
								(0.126)	
Leaveout-Predicted Emigrants$_{t-2}$									0.162
									(0.134)
Std. Beta Coef.	-0.228	0.125	-0.086	-0.185	0.284	0.196	-0.095	0.106	0.034
Panel B. Two-Stage Least-Square Estimates									
	Railway Instrument			Leave-Out Instrument			Overidentified 2SLS		
US Emigrants$_t$	-1.479***			-2.351***			-1.433***		
	(0.372)			(0.396)			(0.389)		
US Emigrants$_{t-1}$		1.670***			1.424***			1.662***	
		(0.458)			(0.366)			(0.455)	
US Emigrants$_{t-2}$			-28.128			59.484			-18.895
			(30.883)			(377.778)			(15.412)
Std. Beta Coef.	-0.232	0.219	-3.441	-0.368	0.187	7.276	-0.224	0.218	-2.311
K-P F-stat	71.165	215.018	0.805	10.054	70.998	0.026	43.918	106.788	0.933
Sargan-Hansen J							3.473	2.735	0.397
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. of District	620	620	618	618	618	618	618	618	618
N. of Observations	2474	1858	1236	2472	1854	1236	2472	1854	1236
Mean Dep. Var.	179.519	221.154	288.871	179.672	221.617	288.871	179.672	221.617	288.871

TABLE D.6. ASSOCIATION BETWEEN OUT-MIGRATION AND THE VOLUME OF INNOVATION

Notes. This table reports the association between US out-migration and the number of patents. The unit of observation is a district, at a decade frequency between 1880 and 1939. The dependent variable is the number of patents. In Panel A, we report the association with measures out-migration (columns 1–3), the reduced-form railway instrument (columns 4–6), and the reduced-form leave-out instrument (columns 7–9). In Panel B, we report the two-stage least-square estimates of the railway (columns 1–3), leave-out (columns 4–6), and combined (columns 7–9) instruments. All regressions include district and decade fixed effects; standard errors are clustered at the district level and are displayed in parentheses.

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

◄ Back: Results ◄ Back: Appendix 4 – Additional Results

	Double Differences		Triple Differences			
	(1)	(2)	(3)	(4)	(5)	(6)
	Level	Level	Level	Level	Share	Share
Excess Deaths × Post	3.120***					
	(0.751)					
1(Q. of Excess Deaths > 75) × Post		1.474***				
		(0.504)				
Excess Deaths × Post × Pharma			2.641***		0.063**	
			(0.678)		(0.032)	
1(Q. of Excess Deaths > 75) × Post × Pharma				1.311***		0.042***
				(0.451)		(0.016)
County FE	Yes	Yes	–	–	–	–
Year FE	Yes	Yes	–	–	–	–
County-Year FE	–	–	Yes	Yes	Yes	Yes
County-Class FE	–	–	Yes	Yes	Yes	Yes
Class-Year FE	–	–	Yes	Yes	Yes	Yes
N. of County-Class	1272	1272	21624	21624	21624	21624
N. of Observations	50880	50880	864960	864960	864960	864960
Classes in Sample	Pharma	Pharma	All	All	All	All
R^2	0.405	0.405	0.683	0.683	0.114	0.114
Mean Dep. Var.	0.991	0.991	0.534	0.534	0.077	0.077
Std. Beta Coef.	0.296	0.083	0.191	0.041	0.032	0.009

TABLE D.7. ESTIMATED EFFECT OF THE INFLUENZA PANDEMIC ON US INNOVATION

Notes. This table reports the effect of exposure to the Great Influenza Pandemic (1918–1919) on innovation in the US. The units of observation are counties (columns 1–2) and county-class pairs (columns 3–6). Units are observed at a yearly frequency between 1900 and 1939. In columns (1–4), the dependent variable is the number of patents granted; in columns (5–6), the dependent variable is the number of pharmaceutical patents divided by the total number of patents granted. In column (1), a post-influenza indicator is interacted with a measure of excess mortality, namely, the ratio between the average number of deaths during the pandemic (1918–1919) and the three previous years. In column (2), the treatment interacts the post-influenza indicator with a dummy variable equal to one for counties in the top quartile of the excess deaths distribution. In columns (3) and (5), the treatment interacts the excess deaths measure with a post-influenza dummy and an indicator variable for pharmaceutical patents; in columns (4) and (6), the excess deaths variable is coded as binary, and returns value one for counties in the top quartile of the excess mortality distribution. In columns (1–2), regressions include county and year fixed effects; in columns (3–6), regressions include county-by-year, county-by-technology class, and technology class-by-year fixed effects. Standard errors, clustered at the district level, are reported in parentheses.

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

◄ Back: Empirical Strategy ◄ Back: Results

	1940s	1950s	1960s	1970s	1980s	1990s	2000s
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent Variable: Number of Patents in:							
Agriculture	3.183***	2.654***	1.914***	1.909**	1.750**	0.647	0.516
	(0.751)	(0.617)	(0.602)	(0.790)	(0.823)	(0.609)	(0.641)
Building	6.065***	4.748***	5.312***	3.918***	4.540***	3.265***	1.804*
	(1.181)	(1.085)	(1.093)	(0.942)	(1.031)	(1.010)	(1.045)
Chemistry	23.553***	15.352***	23.228***	29.817***	11.953**	5.042	2.337
	(7.640)	(5.371)	(6.691)	(12.031)	(5.354)	(6.310)	(6.625)
Electricity	32.018***	31.842**	22.787***	16.405**	8.087	2.124	2.665
	(8.010)	(13.812)	(7.136)	(8.273)	(6.442)	(7.260)	(7.732)
Engineering	9.870***	7.969***	9.549***	8.374***	4.229***	1.167	0.702
	(1.705)	(1.563)	(1.754)	(2.008)	(1.406)	(1.666)	(1.720)
Engines, Pumps	6.551***	7.387***	5.926***	8.172*	4.480*	0.200	0.784
	(2.090)	(2.559)	(1.960)	(4.461)	(2.299)	(2.228)	(2.263)
Food	10.948***	9.570***	8.089***	10.390***	4.173**	0.495	-0.291
	(1.995)	(2.262)	(2.486)	(3.452)	(1.830)	(2.340)	(2.344)
Health, Amusement	4.430***	4.959***	3.988***	7.074***	6.700***	4.318***	5.210***
	(1.307)	(1.405)	(1.419)	(2.111)	(1.536)	(1.526)	(1.708)
Instruments	14.172***	14.338***	15.127***	14.236***	10.658***	4.819	4.079
	(2.937)	(3.244)	(3.480)	(4.101)	(2.905)	(3.072)	(3.599)
Lightning, Heating	11.553***	8.118***	8.113***	5.359***	3.534***	1.581	0.513
	(2.154)	(1.447)	(1.774)	(1.397)	(1.270)	(1.528)	(1.476)
Metallurgy	18.803***	9.443***	13.905***	10.698***	6.346**	1.834	0.849
	(3.888)	(2.573)	(3.541)	(3.493)	(2.722)	(3.110)	(3.236)
Personal Articles, Furniture	6.810***	6.784***	5.250***	2.813***	1.599**	1.367*	1.674**
	(1.014)	(1.175)	(0.811)	(0.755)	(0.803)	(0.811)	(0.821)
Printing	6.914***	7.830***	8.202***	6.030***	3.245***	1.984*	1.455
	(1.226)	(1.341)	(1.573)	(1.205)	(1.045)	(1.190)	(1.313)
Separating, Mixing	7.892***	7.493***	7.633***	8.032***	5.290***	1.602	1.166
	(1.707)	(1.508)	(1.681)	(1.922)	(1.458)	(1.557)	(1.696)
Shaping	9.833***	7.901***	8.591***	7.795***	3.555***	1.156	0.426
	(1.584)	(1.377)	(1.520)	(1.629)	(1.214)	(1.421)	(1.491)
Ships, Aeronautics	8.433***	8.800***	9.757***	6.624***	3.441***	1.319	0.905
	(1.032)	(1.156)	(1.379)	(1.193)	(0.946)	(1.081)	(1.161)
Textiles	14.865***	12.841***	11.039***	10.100***	3.649***	0.752	0.475
	(2.044)	(1.653)	(1.760)	(2.000)	(1.263)	(1.464)	(1.496)
Transporting	5.102***	4.368***	3.974***	2.988**	1.704*	0.471	0.147
	(1.157)	(1.194)	(1.231)	(1.391)	(0.934)	(1.123)	(1.167)
Number of Districts	632	632	632	632	632	632	632

TABLE D.9. LONG-RUN SECTOR CORRELATION BETWEEN KNOWLEDGE EXPOSURE AND INNOVATION

Notes. This table reports the correlation between knowledge exposure in the years 1900–1930 and subsequent patenting activity by sector. For each class displayed in the rows, we estimate a model that interacts knowledge exposure with decade dummies, and we report the coefficients for each decade in the respective column. The 2010s decade serves as the baseline category. All regressions include district and decade fixed effects. Robust standard errors are displayed in parentheses.

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

◄ Back: Results ◄ Back: Appendix 4 – Additional Results

	By Age	By Occupation						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Emigrant × Post	0.105*	0.348***	0.595***	0.172*	-0.040	-0.209	-0.030	0.069
	(0.057)	(0.134)	(0.222)	(0.103)	(0.110)	(0.202)	(0.212)	(0.051)
Age ∈ [18, 30) × Emigrant × Post	0.195**							
	(0.092)							
Age ∈ [30, 40) × Emigrant × Post	0.006							
	(0.068)							
Age ∈ [50, 60) × Emigrant × Post	-0.033							
	(0.083)							
Age ≥ 60 × Emigrant × Post	0.040							
	(0.182)							
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Full	Engineering	Metallurgy	Construction	Textiles	Trade	Pub. Adm.	Agriculture
N. of Individuals	469250	62716	12875	65013	31144	40576	15420	102463
N. of Observations	13608250	1818764	373375	1885377	903176	1176704	447180	2971427
R^2	0.135	0.120	0.097	0.148	0.080	0.097	0.295	0.103
Mean Dep. Var.	0.616	0.871	0.672	0.564	0.548	0.988	0.745	0.253
Std. Beta Coef.	0.002	0.004	0.010	0.003	-0.001	-0.003	-0.000	0.002

TABLE D.10. HETEROGENEITY ANALYSIS OF THE EFFECT OF NEIGHBORHOOD OUT-MIGRATION ON INNOVATION

Notes. This table reports some heterogeneity analysis on the individual-level effect of neighborhood migration on patenting activity. The units of observation are individuals who are observed yearly between 1900 and 1920. The baseline treatment is an indicator that, for a given individual, returns value one after at least one person that was living in the same neighborhood as the individual migrates to the United States. In column (1), we interact this treatment with age category dummies and normalize the dummy for the age range 40–50 as the baseline category. In columns (2–8), we estimate the baseline double differences model by recorded occupations. Hence, in column (2), we estimate the model only for individuals employed in engineering occupations. All models include individual and year fixed effects. Standard errors are clustered at the district level and are reported in parentheses.

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

◄ Back: Mechanisms

	Baseline	Excluding States in ...				Innovation Shock Treshold		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Northeast	Midwest	South	West	0.1%	1%	10%
Innovation Shock × Post	32.727***	40.194***	26.937***	32.514***	33.336***	94.663***	18.948***	3.958***
	(2.610)	(4.018)	(2.613)	(2.633)	(2.802)	(8.056)	(1.384)	(0.207)
District-by-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-by-Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class-by-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Counties	2101783	1942416	1346809	1177807	1892929	2101824	2101250	2093047
Number of Observations	51263	47376	32849	28727	46169	51264	51250	51097
Mean Dep. Var.	0.772	0.511	0.730	1.241	0.765	0.772	0.752	0.615

TABLE D.8. DOUBLE AND TRIPLE DIFFERENCES EFFECT OF SYNTHETIC INNOVATION SHOCKS ON SUBSEQUENT US INNOVATION

Notes. This table reports the effect of synthetic innovation shock on US innovation. These coefficients are not interpreted as causal but as evidence that synthetic shocks capture relevant variation in county-technology-specific patenting activity. The unit of observation is a county-technology class pair observed at a yearly frequency between 1900 and 1939. The baseline treatment is an interaction between an innovation shock and a post-shock indicator. An innovation shock occurs when the residualized patenting activity in a given county technology is in the top 0.5% of the overall distribution of residualized values. Because the setting is staggered, all regressions are estimated using the methodology of Borusyak *et al.* (2021). Column (1) reports the estimate for the entire panel of counties; in columns (2), (3), (4), and (5), we exclude counties in, respectively, the North-East, Midwest, South, and West Census Bureau regions. In columns (6), (7), and (8), instead, we consider different thresholds for the definition of innovation shocks at the top 0.1%, 1%, and 10% of the overall distribution of residualized patents, respectively. All regressions include county-by-year, county-by-technology class, and technology class-by-year fixed effects. Standard errors, clustered at the county level, are reported in parentheses.

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

◄ Back: Empirical Strategy ◄ Back: Appendix 5 – Robustness

D.6 Figures

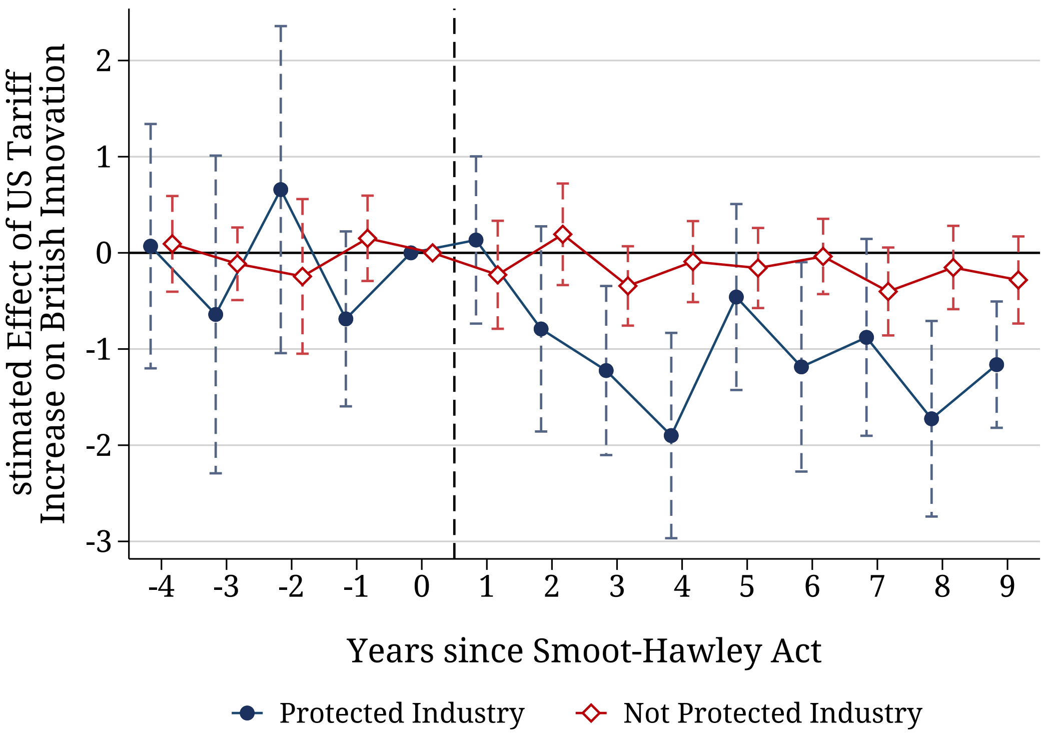


FIGURE D.1. FLEXIBLE DIFFERENCE-IN-DIFFERENCES ESTIMATED EFFECT OF TARIFF REFORM ON INNOVATION

Notes. This figure reports the estimated dynamic treatment effects of increased US tariff rates on innovation in Britain. The unit of observation is a district-technology class pair observed at a yearly frequency between 1920 and 1939. The dependent variable is the number of patents. The independent variable is the interaction between knowledge exposure over 1910–1920 and year dummies. The last year before the Reform, 1929, is the baseline category. The blue dots report the estimated treatment effects for technology classes targeted by the Act; the red dots restrict the sample to non-treated technology classes. We define a class as "targeted" if its average tariff rate increases by more than 50% after the Smoot-Hawley Act. Regressions include district-by-class and year fixed effects. Standard errors, clustered at the district level, are reported in parentheses.

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

◀Back: Appendix 4 – Additional Results

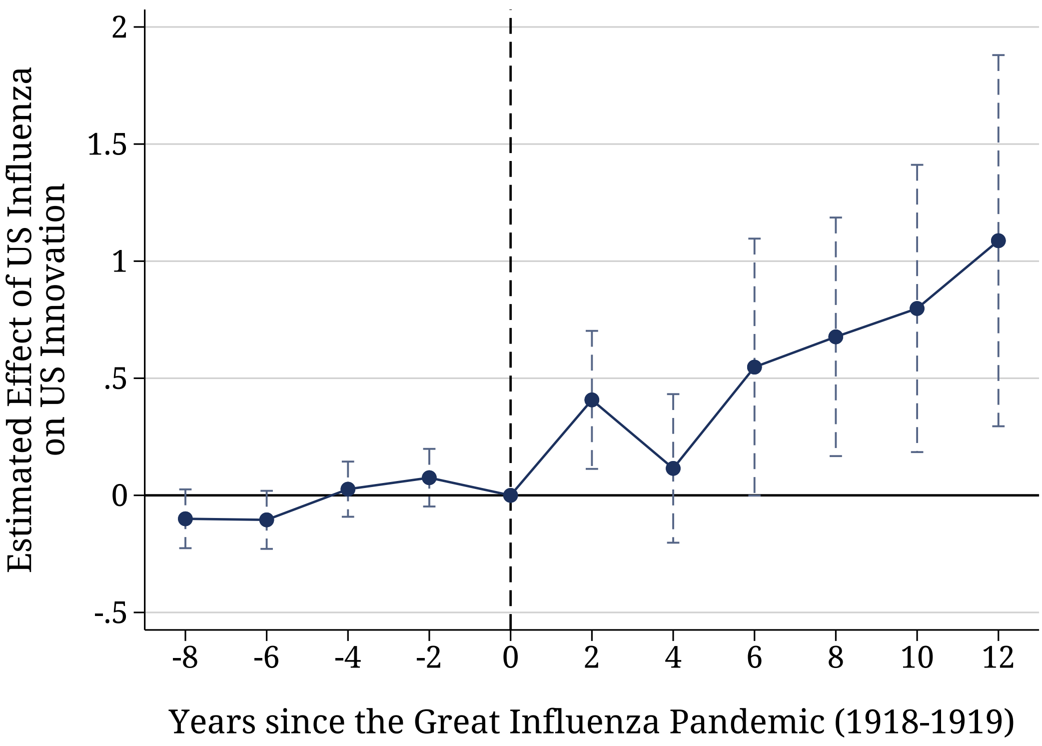


FIGURE D.2. FLEXIBLE TRIPLE DIFFERENCES ESTIMATED EFFECT OF THE INFLUENZA PANDEMIC ON US INNOVATION

Notes. These figures report the dynamic treatment effects of exposure to the Great Influenza Pandemic on innovation in the US. The units of observation are county-technology class pairs; units are observed at a yearly frequency between 1900 and 1939. The dependent variable is the number of patents. The treatment is an indicator equal to one for pharmaceutical patents and districts in the top quartile of the excess mortality distribution. The graph displays the interaction coefficients between the treatment and biennial time dummies, where the last dummy before the pandemic—1916–1917—serves as the baseline category. Excess mortality is computed as the average number of deaths during the pandemic over the average number of deaths in the three years before the pandemic. The black dashed line indicates the timing of the treatment. The regression includes county-by-technology class, technology class-by-biennial, and county-by-biennial fixed effects. Standard errors are two-way clustered by district and technology class; bands report 95% confidence intervals.

◄ Back: Empirical Strategy ◄ Back: Results ◄ Back: Appendix 5 – Robustness

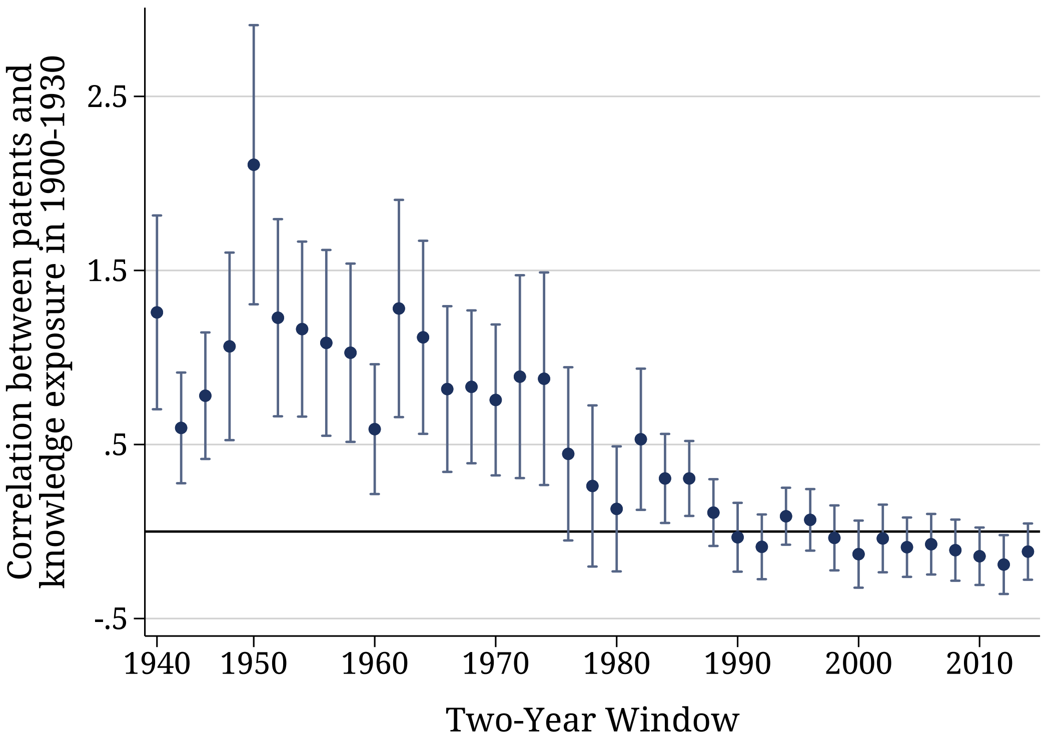


FIGURE D.3. LONG-RUN ASSOCIATION BETWEEN KNOWLEDGE EXPOSURE AND SUBSEQUENT INNOVATION ACTIVITY, 1940–2020

Notes. This figure reports the correlation between knowledge exposure in the period 1900–1930 and subsequent innovation activity. The unit of observation is a district-technology class pair. Units are observed at a biennial frequency between 1940 and 2015. Each dots report the coefficient of an interaction term between—time-invariant—knowledge exposure and biennial time dummies. The last biennial, 2014–2015, serves as the baseline category. The model includes district-by-technology class and decade fixed effects. Standard errors are clustered at the district level. Bands report 95% confidence intervals.

◄ Back: Results ◄ Back: Appendix 4 – Additional Results

E Robustness Analysis

This section provides details on the technical implementation of the analyses discussed in the main text and briefly describes the exercises we perform to ensure the results' robustness.

E.1 Alternative Baseline Specifications

In this section, we list and comment on the alternative specifications of equation (2) that we estimate in the main text.

E.1.1 Alternative Dependent Variables

In the principal analysis, we use the raw number of patents at varying levels of aggregation as the dependent variable. We thus follow Chen and Roth (2022), who note that under transformations of the dependent variable defined at zero—as would be our case, to avoid dropping zero-patents observations—, the estimates of the average treatment effect are scale-dependent. Since it is common practice in the innovation literature to take the log-transformation, in Table E.1, we show that the results are robust using a battery of alternative transformations.

E.1.2 Alternative Definitions of Knowledge Exposure

In Table E.2, 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:

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

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:

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

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:

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

The idea behind (E.3) is that specialization can be defined in terms of the cumulative number of patents filed before the given period. In Table E.2, we show that all these measures yield quantitatively similar results.

E.1.3 Alternative Fixed Effects

In the main text, we report the results for a specification that includes district-by-time and technology class-by-time fixed effects. These are intended to capture time-varying unobserved heterogeneity at the district technology levels that we do not observe. In Table E.3, we show that the—OLS and 2SLS—results are robust when including a wide array of alternative fixed effects. First, in columns (1) and (6), we report the unconditional correlation between innovation and knowledge exposure. This documents that knowledge exposure alone explains a sizable (30%) share of the variation in patenting activity. Then, in columns (2–5) and (7–10), we incrementally include additional fixed effects and show that the significance and magnitude of the coefficients remain very stable. In particular, in columns (5) and (10), we saturate the model with all couples of fixed effects to non-parametrically control for heterogeneity at the district-time, technology-time, and district-technology levels. The results are confirmed even in this demanding specification.

E.2 Instrumental Variable Strategy

This section discusses how we construct the county-level shocks necessary to compute the predicted bilateral flows, as described in section 4. We first present the strategy to construct the shocks for the main railway-based instrument. Then, we explain how we compute the shocks for the additional, leave-out instrument.

E.2.1 Railway-Based Instrument

The baseline instrument leverages county-level immigration shocks obtained from variations in the conditional timing when each county was connected to the US railway network. 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:

$$\begin{aligned} \text{Immigrant Share}_{j,t} = \alpha_j + \alpha_t + \beta \text{Immigrant Share}_{j,t-1} + \gamma I^{\text{Rail}}_{j,t-1} + \\ + \delta \left(I^{\text{Rail}}_{j,t-1} \times \text{Immigrant Flow}_{t-1} \right) + \zeta \left(\text{Industrialization}_{t-1} \times I^{\text{Rail}}_{j,t-1} \right) + \\ + \eta \left(\text{GDP Growth}_{t-1} \times I^{\text{Rail}}_{j,t-1} \right) + X^{\intercal}_{j,t-1} \Theta + \varepsilon_{j,t} \end{aligned} \tag{E.4}$$

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 D.1.

We construct a synthetic series of county-level time-varying immigration shocks from equation (E.4) as follows:

$$\widehat{\text{Immigrant Share}}_{j,t} = \hat{\delta}\left(I^{\text{Rail}}_{j,t-1} \times \text{Immigrant Flow}_{t-1}\right) \tag{E.5}$$

where $\hat{\delta}$ is simply the OLS estimates from the previous model. We thus generate a set of county-level immigration shocks that are orthogonal to economic development and other characteristics that may induce sorting into the US. Variation, in other words, is solely due to the timing when a county is connected to the railway network.

E.2.2 Alternative Instrumental Variable

As further robustness to the railway instrument, we develop a simple 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

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

where $\mathcal{M}$ is a set of origin countries. Both (E.5) and (E.6) yield a set of county-specific immigration shocks that do not conflate the immigration patterns of the British. They leverage very different sources of variation, though, which enables us to use the resulting instruments jointly and perform over-identification tests.

We allow multiple sets of origin countries $\mathcal{M}$. The baseline exercise, whose first-stage relevance is shown in Table E.7 and results are displayed in Table E.8, collates all countries except for the UK.[73] To account for possible correlation between British immigrants and those from other nationalities, however, we vary the set of included countries in Table E.9. In particular, we drop all countries in Northern Europe (column 3), 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.

E.2.3 Tests on Instrument Validity

The validity of the shift-share instrument for knowledge exposure that we construct hinges on the exogeneity of the shocks constructed using either (E.5) or (E.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 first exercise in Figure E.3. Panel (A) displays 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 panels (B) and (C), 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.[74] This provides reassuring evidence in favor of the validity of the instruments.

We report the second exercise in Figure E.4. Each dot displays the correlation between district-level observable variables and actual, railway-based, and leave-out emigration in panels (A), (B), and (C), 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.

[73] In Figure E.2 we report binned scatter plots of the association between predicted and actual knowledge exposure using the two instruments.

[74] Even when the correlation remains significant, the standardized beta coefficient is substantially lower than in the benchmark panel (A).

E.3 Shock Propagation

This section describes the technical definition of the synthetic innovation shocks and exposure to the influenza pandemic, along with two falsification exercises and several sensitivity analyses.

E.3.1 Details on the Construction of the Synthetic Shocks

We define a synthetic innovation shock as an unusual deviation from the number of patents granted in a given county, technology class, and year. Formally, we estimate the following fixed-effects regression:

$$\text{Patents}_{jk,t} = \alpha_{j\times k} + \alpha_{k\times t} + \alpha_{j\times t} + \varepsilon_{jk,t} \tag{E.7}$$

where j, k, and t denote a county, technology class, and year respectively, and α is the associated fixed effect. In particular, we include county-by-year fixed effects to remove fluctuations in patenting activity due to, for instance, economic growth. We remove technology-by-year fixed effects to ensure that the shocks do not reflect aggregate changes in the propensity to patent in a given class. Finally, we average out county-by-class fixed effects to remove asymmetries due to initial specialization. We then construct a series of residualized innovation activity from the residuals of (E.7).

In the baseline analysis, we define an innovation shock as an observed residualized patenting activity in the top 0.1% of the overall distribution. Let $\Gamma(\cdot)$ be the cumulative distribution of the residuals of regression (E.7). Then, the set of shocks $\xi(\tau)$, for $\tau = 0.001$, is given by the set $\xi(\tau) = \{\xi \in \text{supp}(\Gamma) \mid \Gamma(\xi) - \Gamma(\tau) \geq 0\}$. In Table E.11, we use two other threshold values of τ (1% and 0.5%). We find that the average treatment effect decreases as τ increases. This is compelling since larger τ's flag smaller residualized patenting activity as instances of treatment. In Table D.8, we show the "effect" of synthetic shocks on US innovation. This is not a causal effect but rather a measure of the relevance of such shocks. There is a strong and positive increase in the number of patents after the shock is observed, and this also holds excluding specific areas (columns 2–5). In columns (6–8), we show that larger levels of τ are associated with a lower increase in patenting.

E.3.2 Details on the Construction of the Influenza Shock

To construct exposure to the influenza across counties, we follow Berkes *et al.* (2023). From the mortality statistics collected by the Bureau of Census, we define a metric of excess deaths as the ratio between average deaths during the pandemic (1918–1919) relative to the average in the preceding three years.[75] Formally, we have

$$\text{Excess Deaths}_j = \frac{\frac{1}{2}\sum_{t=1918}^{1919}\text{Deaths}_{j,t}}{\frac{1}{3}\sum_{t'=1915}^{1917}\text{Deaths}_{j,t'}} \tag{E.8}$$

[75] Due to data limitations, this is the pre-pandemic period that maximizes the sample size. Mortality statistics thus allow covering 60% of the US population.

We then recast it as a binary variable equal to one if county j is in the top 25% of the excess deaths distribution to avoid issues of continuous treatment (Callaway and Sant'Anna, 2021).

The baseline estimation equation for US counties is then

$$\text{Patents}_{jk,t} = \alpha_{j\times k} + \alpha_{k\times t} + \alpha_{j\times k} + \delta\,(\text{Excess Deaths}_c \times \text{Pharma}_k \times \text{Post}_t) + \varepsilon_{jk,t} \tag{E.9}$$

where Pharma_k is an indicator variable returning value one if k is pharmaceutical patents, and zero otherwise, and Post_t is an indicator variable returning value one for years after 1918, and zero otherwise. Figure D.2 reports the associated flexible triple differences estimates, which, with no evidence of statistically significant pre-treatment coefficients, suggests that the influenza had a strong, positive, and significant effect on pharmaceutical innovation in the US.

E.3.3 Robustness of the Synthetic Shock Analysis

We perform two main exercises to ensure that the results using the synthetic shocks are robust. First, in Figure E.7, we check that the estimated effect of US innovation shocks on UK innovation remains significant and is quantitatively consistent under different estimators that allow for staggered roll-out of the treatment across units. The estimated ATE remains significant, and its magnitude is preserved under various estimators.

Second, in Table E.11, we vary two margins along which a district is considered to be treated. First, as previously discussed, we consider different thresholds τ (1%, 0.5%, and the baseline 0.1%) above which we flag synthetic innovation shocks. Reassuringly, larger levels τ, which require a lower marginal increase in patenting to flag a synthetic shock, lead to smaller ATEs. This is consistent with the idea that larger innovation shocks in the US should lead to larger innovation shocks in the UK. Second, we vary the threshold of emigration that we impose for a district to be considered exposed to the innovation shock. In our main analysis, we consider a district exposed to the innovation shock in a given county if it is in the top quartile of the distribution of emigration to that county. We consider two additional thresholds (top 50% and top 90%). We find that the baseline result is qualitatively robust to all such thresholds. Moreover, we confirm that larger exposure thresholds lead to larger estimated ATEs. This suggests that the more intense the previous migration tie between a county and a district, the stronger the diffusion effect of county-level shocks on district-level innovation.

E.3.4 Robustness of the Influenza Shock Analysis

In Table E.12, we perform several exercises to gauge the robustness of the effect of UK exposure to excess mortality in the US during the Great Influenza pandemic. In columns (1) and (2), we report the double-difference estimates that compare innovation in pharmaceuticals in districts with high and low exposure to counties with high excess mortality. Compared to the baseline triple-difference model, these estimates do not include non-pharmaceutical innovation in the control group. The results of this exercise are quantitatively comparable with the baseline model. In column (3), we report the result of a triple-difference model that does not discretize district-level exposure to US counties. Columns (4–7), instead, restrict the sample by excluding the top-patenting areas. The results remain consistent throughout these specifications.

E.3.5 Shock Falsification Checks

The rationale for the analysis discussed in the main text (table 3 and Figure 5) and thus far is that the influenza only impacted patenting in pharmaceutical patents in the US. If that is the case, then this would ignite an innovation shock that was localized in areas that were more exposed to the influenza, and that could reverberate in the UK to districts whose emigrants had settled in such areas.

We test this assumption in Figure E.6a. Each dot reports an estimated δ coefficient of equation (E.9), except that the treated technology is reported in each row. Thus, the exclusion restriction would require that each coefficient was not statistically different from zero, except for pharmaceuticals. This assumption is confirmed in the data. The ATE for pharmaceuticals is the only one that is positive, significant, and quantitatively large. Figure E.6a thus implies that we expect to observe an increase in pharmaceutical patents only, and only in districts whose emigrants had settled in areas that were more severely exposed to the pandemic.

We test this in Figure E.6b, in which we estimate the baseline triple-difference specification of the main text, except that the treated technology is reported in each row, as before. While estimates are noisier here, we confirm that the estimated ATE for pharmaceuticals is the largest and statistically significant across classes, as expected. Overall, Figure E.6 thus provides convincing evidence that (i) the influenza fostered innovation in pharmaceuticals only in the US, and (ii) that districts whose emigrants had settled in areas that were more severely exposed to the influenza display higher innovation activity in pharmaceuticals.

E.4 Tables

	Level of Patents					Share of Patents				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Baseline	$\ln(\cdot)$	$\ln(1+\cdot)$	$\ln(\varepsilon+\cdot)$	$\text{arcsinh}(\cdot)$	Share	$\ln(\cdot)$	$\ln(1+\cdot)$	$\ln(\varepsilon+\cdot)$	$\text{arcsinh}(\cdot)$
Panel A. OLS Estimates										
Knowledge Exposure	1.342***	0.015***	0.067***	0.142***	0.082***	0.005***	0.015***	0.004***	0.169***	0.005***
	(0.143)	(0.002)	(0.007)	(0.016)	(0.009)	(0.001)	(0.002)	(0.000)	(0.020)	(0.001)
R^2	0.772	0.802	0.824	0.766	0.813	0.330	0.625	0.344	0.523	0.334
Std. Beta Coef.	0.299	0.101	0.396	0.495	0.407	0.411	0.139	0.439	0.629	0.418
Panel B. Reduced-Form Estimates										
Knowledge Exposure	0.037***	0.001***	0.001***	0.001***	0.001***	0.000***	0.001***	0.000***	0.001***	0.000***
	(0.007)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.800	0.811	0.816	0.752	0.805	0.347	0.651	0.358	0.510	0.350
Std. Beta Coef.	0.075	0.032	0.043	0.027	0.039	0.046	0.043	0.046	0.022	0.046
Panel C. Two-Stage Least Square Estimates										
Knowledge Exposure	1.224***	0.018***	0.031***	0.034***	0.034***	0.002***	0.018***	0.002***	0.027***	0.002***
	(0.195)	(0.005)	(0.005)	(0.007)	(0.006)	(0.000)	(0.005)	(0.000)	(0.008)	(0.000)
K-P F-stat	109.826	83.266	109.826	109.826	109.826	109.826	83.266	109.826	109.826	109.826
Std. Beta Coef.	0.296	0.116	0.171	0.108	0.153	0.181	0.158	0.181	0.087	0.181
District-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Technology Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. of District-Class	11268	8475	11268	11268	11268	11268	8475	11268	11268	11268
N. of Observations	67549	36290	67549	67549	67549	67549	36290	67549	67549	67549
Mean Dep. Var.	10.392	1.795	1.137	-0.005	1.400	0.051	-2.946	0.046	-4.636	0.050

TABLE E.1. KNOWLEDGE EXPOSURE AND INNOVATION: ALTERNATIVE DEPENDENT VARIABLES

Notes. This table displays the association between innovation and exposure to US knowledge using alternative transformations of the dependent variable. The unit of observation is a district-technology class pair, observed at a decade frequency between 1880 and 1939. In columns (1–5), the dependent variable is the number of patents; in columns (6–10), the dependent variable is the share of patents in a given technology, normalized by the total number of patents. In columns (1) and (6), we do not transform the dependent variable; in columns (2) and (7), we take the log; columns (3) and (8) report the estimates using log(1+), which avoids dropping zeroes; in columns (4) and (9) we take log(0.01+) of the dependent variable; columns (5) and (10) report the estimates using the inverse hyperbolic sine. The main explanatory variable is knowledge exposure. In Panel A, we estimate the correlation through OLS; in Panel B, we report the reduced-form association between the instrument for knowledge exposure and innovation; in Panel C, we display the two-stage least-squares estimates. Each model includes district-by-decade and district-by-technology class fixed effects. Standard errors are reported in parentheses and are clustered at the district level.

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

◄ Back: Results ◄ Back: Appendix 5 – Robustness

	Dependent Variable: N. of Patents				
	(1)	(2)	(3)	(4)	(5)
Knowledge Exposure	1.342***				
	(0.143)				
ln(1 + Knowledge Exposure)		4.175***			
		(0.228)			
Fixed-Emigrants Knowledge Exposure			2.610***		
			(0.300)		
Fixed-Patents Knowledge Exposure				0.063***	
				(0.015)	
Cumulative Knowledge Exposure					0.136**
					(0.067)
District-Decade FE	Yes	Yes	Yes	Yes	Yes
District-Technology Class FE	Yes	Yes	Yes	Yes	Yes
N. of District-Class	11268	11268	11268	11268	11268
N. of Observations	67549	67549	67547	67555	67549
R^2	0.772	0.766	0.770	0.765	0.764
Mean Dep. Var.	10.392	10.392	10.393	10.392	10.392
Std. Beta Coef.	0.299	0.119	0.249	0.104	0.017

TABLE E.2. KNOWLEDGE EXPOSURE AND INNOVATION: ALTERNATIVE MEASURES OF KNOWLEDGE EXPOSURE

Notes. This table displays the association between innovation and exposure to US knowledge, using alternative transformations of knowledge exposure. The unit of observation is a district-technology class pair, observed at a decade frequency between 1880 and 1939. In column (1), we report the baseline estimate. In column (2), we take knowledge exposure in log terms, adding one to avoid dropping zeros since the baseline measure is defined as non-negative. In column (3), we fix bilateral district-county bilateral exposure shares as the number of emigrants from the given district to the given county in the decade 1870-1880. In column (3), instead, we fix county-level specialization as the share of patents in a given field granted in the decade 1870-1880 only. In column (5), for a given decade, we measure specialization as the sum of patents obtained until the end of that decade relative to the total number of patents obtained until the end of that decade. The measure used in column (5) thus considers the cumulative patent stock instead of its decade-on-decade flow. The main explanatory variable is knowledge exposure. Each model includes district-by-decade and district-by-technology class fixed effects. Standard errors are reported in parentheses and are clustered at the district level.

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

◄ Back: Empirical Strategy ◄ Back: Results ◄ Back: Appendix 5 – Robustness

	Dependent Variable: N. of Patents									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	10)
Panel A. Correlational Estimates										
	OLS					Poisson				
Knowledge Exposure	2.393***	1.947***	1.936***	1.942***	1.241***	0.038***	0.011***	0.014***	0.014***	0.010***
	(0.212)	(0.194)	(0.184)	(0.191)	(0.149)	(0.004)	(0.002)	(0.003)	(0.003)	(0.002)
R^2	0.284	0.431	0.539	0.547	0.781	0.226	0.718	0.749	0.760	0.864
Std. Beta Coef.	0.533	0.433	0.431	0.432	0.276	0.303	0.084	0.112	0.109	0.080
Panel C. Instrumental Variable Estimates										
	Reduced Form					Two-Stage Least Squares				
$\widehat{\text{Knowlegde Exposure}}$	0.158***	0.080***	0.041***	0.041***	0.033***					
	(0.012)	(0.009)	(0.010)	(0.012)	(0.009)					
Knowledge Exposure						2.053***	1.490***	0.867***	0.779***	1.097***
						(0.157)	(0.167)	(0.192)	(0.208)	(0.271)
R^2	0.104	0.385	0.501	0.509	0.808	0.293	0.111	0.065	0.059	0.028
K-P F-stat						184.700	96.871	169.304	132.033	46.312
Std. Beta Coef.	0.322	0.164	0.083	0.083	0.068	4.196	3.045	1.772	1.591	2.243
District FE	No	Yes	–	–	–	No	Yes	–	–	–
Decade FE	No	Yes	–	–	–	No	Yes	–	–	–
Class FE	No	Yes	Yes	–	–	No	Yes	Yes	–	–
District-Decade FE	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Class-Decade FE	No	No	No	Yes	Yes	No	No	No	Yes	Yes
District-Class FE	No	No	No	No	Yes	No	No	No	No	Yes
N. of District-Class	11268	11268	11268	11268	11268	11268	11250	11250	11250	10081
N. of Observations	67549	67549	67549	67549	67549	67549	67474	65946	65946	59703
Mean Dep. Var.	10.392	10.392	10.392	10.392	10.392	10.392	10.404	10.645	10.645	11.758

TABLE E.3. KNOWLEDGE EXPOSURE AND INNOVATION: ALTERNATIVE SETS OF FIXED EFFECTS

Notes. This table displays the association between innovation and exposure to US knowledge. The unit of observation is a district-technology class pair, observed at a decade frequency between 1880 and 1939. The dependent variable is the number of patents. The main explanatory variable is knowledge exposure. In Panel A, in columns (1–5), we estimate the correlation through OLS; in columns (6–10), we estimate the model as a Poisson regression to account for the many zeros in the data; columns (1–5) in Panel B report the reduced-form association between the instrument for knowledge exposure and innovation; columns (6–10) report the two-stages least square estimates. Columns (1) and (6) reports the unconditional regressions; in columns (2) and (7), we include district, decade, and technology class fixed effects; columns (3) and (8) add district-by-decade fixed effects; in columns (4) and (9) we include district-by-decade and class-by-decade fixed effects; models in columns (5) and (10) are saturated with all couples of fixed effects. Standard errors are reported in parentheses and are clustered at the district level.

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

◄ Back: Results

	Quality Weight		Breakthrough 20%		Breakthrough 10%		Breakthrough 5%	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Panel A. Unadjusted Quality Indicator								
Knowledge Exposure$_t$	1.433***	3.812***	0.706***	1.330***	0.387***	0.697***	0.228***	0.328***
	(0.254)	(1.206)	(0.134)	(0.342)	(0.078)	(0.194)	(0.048)	(0.100)
R^2	0.822	-0.038	0.660	-0.049	0.585	-0.033	0.519	-0.037
Mean Dep. Var.	16.327	12.647	3.842	2.103	1.994	1.027	1.003	0.409
Std. Beta Coef.	0.265	0.908	0.408	1.491	0.366	1.443	0.324	1.270
Panel B. Adjusted Quality Indicator (Net of Class-Year FE)								
Knowledge Exposure$_t$	0.051***	0.119***	0.389***	0.808***	0.229***	0.449***	0.110***	0.436***
	(0.017)	(0.043)	(0.083)	(0.186)	(0.059)	(0.126)	(0.033)	(0.100)
R^2	0.464	-0.055	0.649	-0.106	0.580	-0.108	0.519	-0.228
Mean Dep. Var.	0.284	0.168	2.698	1.688	1.621	0.903	0.850	0.460
Std. Beta Coef.	0.180	0.991	0.321	1.261	0.261	1.209	0.188	1.966
District-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Technology Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. of District-Class	11263	11198	11263	11198	11263	11198	11263	11198
N. of Observations	33764	22396	33764	22396	33764	22396	33764	22396
K-P F-stat		44.652		44.652		44.652		44.652

TABLE E.4. RETURN INNOVATION ACCOUNTING FOR PATENT QUALITY

Notes. This table displays the association between innovation and exposure to US knowledge. The unit of observation is a district-technology class pair, observed at a decade frequency between 1900 and 1939. The main explanatory variable is knowledge exposure. The dependent variables capture the "quality" of patents. To measure quality, we adapt the text-based indicator developed by Kelly *et al.* (2021). The sample excludes the years 1880–1899 because, for the subsequent years, we only have abstracts. The quality measure is thus evaluated on abstracts. In columns (1–2), the dependent variable is the number of patents, weighted by their quality. In columns (3–8), following Kelly *et al.* (2021), we only count patents in the top 20%, 10%, and 5% of the overall quality distribution. Odd columns display the OLS correlation between knowledge exposure and the dependent variables; Even columns report the associated two-stage least-square estimates. In panel (A), the quality measure is not adjusted; in panel (B), we remove class-by-year fixed effects in the quality measure following Kelly *et al.* (2021) to control for fashion effects in language. All regressions include district-by-decade and district-by-technology class fixed effects. Standard errors are reported in parentheses and are clustered at the district level.

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

◄ Back: Results

	Quality Weight		Breakthrough 20%		Breakthrough 10%		Breakthrough 5%	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Panel A. Unadjusted Quality–Weighted Knowledge Exposure								
Knowledge Exposure (Weighted)	0.003**	0.233						
	(0.001)	(0.179)						
Knowledge Exposure (top 20%)			0.532***	0.278***				
			(0.048)	(0.081)				
Knowledge Exposure (top 10%)					0.500***	0.189**		
					(0.047)	(0.079)		
Knowledge Exposure (top 5%)							0.446***	-0.037
							(0.047)	(0.072)
Panel B. Adjusted Quality-Weighted Knowledge Exposure (Net of Class-Year FE)								
Knowledge Exposure (Weighted)	0.002	0.205***						
	(0.002)	(0.070)						
Knowledge Exposure (top 20%)			0.531***	0.332***				
			(0.052)	(0.084)				
Knowledge Exposure (top 10%)					0.495***	0.265***		
					(0.049)	(0.077)		
Knowledge Exposure (top 5%)							0.486***	0.093
							(0.050)	(0.075)
District-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Technology Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. of District-Class	11160	11106	11160	11106	11160	11106	11160	11106
N. of Observations	66901	55507	66901	55507	66901	55507	66901	55507
K-P F-stat		1.579		117.459		136.236		156.128

TABLE E.5. RETURN INNOVATION ACCOUNTING FOR PATENT QUALITY OF US PATENTS

Notes. This table displays the association between innovation and exposure to US knowledge. The unit of observation is a district-technology class pair, observed at a decade frequency between 1870 and 1939. The main explanatory variable is knowledge exposure, weighted by the quality of US patents. The dependent variable is the number of patents, in logs. To measure quality, we adapt the text-based indicator developed by Kelly *et al.* (2021). The first row in each panel weights patents by their quality; the second, third, and fourth rows only count patents in the top 20%, 10%, and 5% of the overall quality distribution, respectively. In panel (A), the quality measure is not adjusted; in panel (B), we remove class-by-year fixed effects in the quality measure following Kelly *et al.* (2021) to control for fashion effects in language. All regressions include district-by-decade and district-by-technology class fixed effects. Standard errors are reported in parentheses and are clustered at the district level.

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

◄ Back: Results

	OLS		Reduced Form		Two-Stage Least-Squares	
	(1)	(2)	(3)	(4)	(5)	(6)
Knowledge Exposure$_t$	0.170***				0.068***	
	(0.012)				(0.021)	
Knowledge Exposure$_{t-1}$		0.176***				0.327***
		(0.021)				(0.046)
$\widehat{\text{Knowledge Exposure}}_t$			0.004***			
			(0.001)			
$\widehat{\text{Knowledge Exposure}}_{t-1}$				0.023***		
				(0.003)		
District-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
District-Technology Class FE	Yes	Yes	Yes	Yes	Yes	Yes
N. of District-Class	11268	11266	11214	11214	11214	11214
N. of Observations	33798	22532	33642	22428	33640	22428
K-P F-stat					134.004	154.011
R^2	0.841	0.914	0.835	0.912	0.021	0.014
Mean Dep. Var.	1.250	1.759	1.256	1.767	1.256	1.767
Std. Beta Coef.	0.179	0.165	0.026	0.133	0.071	0.306

TABLE E.6. EFFECT OF EXPOSURE TO US TECHNOLOGY ON INNOVATION IN GREAT BRITAIN: PATENTS WITH FIRM ASSIGNEE

Notes. This table displays the association between innovation and exposure to US knowledge. The unit of observation is a district-technology class pair, observed at a decade frequency between 1880 and 1900. The main explanatory variable is knowledge exposure. The dependent variable is the number of patents with at least one firm listed as an assignee. In columns (1–2), we estimate the OLS correlation with the observed measure of knowledge exposure; in columns (3–4), we estimate the reduced-form association with the railway-based instrument of knowledge exposure through OLS; columns (5–6) report the two-stage least-squares estimate. Each model includes district-by-decade and district-by-technology class fixed effects. Standard errors are reported in parentheses and are clustered at the district level.

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

◄ Back: Results

	Railway-Based (SNQ) Instrument				Leaveout Instrument			
	Baseline	Dropping Districts in...			Baseline	Dropping Districts in...		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		London	Lancs	S-W		London	Lancs	S-W
Panel A. Dependent Variable: Bilateral Flows								
SNQ Migrants	0.007***	0.008***	0.007***	0.007***				
	(0.001)	(0.001)	(0.001)	(0.001)				
Leaveout Migrants					0.006***	0.005**	0.005***	0.005***
					(0.002)	(0.002)	(0.002)	(0.002)
District-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. of District-Counties	1736040	1653240	1518000	1625640	1786360	1701160	1562000	1672760
N. of Observations	8399666	7999046	7344700	7865506	10403031	9906861	9096450	9741471
Panel B. Dependent Variable: Knowledge Exposure								
SNQ Knowledge Exposure	0.005***	0.005***	0.005***	0.005***				
	(0.000)	(0.000)	(0.000)	(0.000)				
Leaveout Knowledge Exposure					0.169***	0.157***	0.159***	0.145***
					(0.034)	(0.036)	(0.034)	(0.032)
District-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. of District-Classes	11322	10782	9900	10602	11304	10764	9882	10584
N. of Observations	56587	53887	49488	52987	67801	64561	59280	63481

TABLE E.7. FIRST STAGE OF THE INSTRUMENTAL VARIABLE ESTIMATION

Notes. This table reports the first-stage estimates of the two shift-share instruments we propose. In Panel A, the observation units are district-county pairs, observed at a decade frequency between 1870 and 1920 (columns 1–4) and 1930 (columns 5–8). In Panel B, the observation units are district-technology classes, at decade frequency between 1870 and 1920 (columns 1–4) and 1930 (columns 5–8). In columns (1–4), the predicted number of emigrants constructed using the railway-based instrument that leverages shocks *à la* Sequeira *et al.* (2020); in columns (5–8), predicted emigrants are constructed using the leave-out instrument. Columns (1) and (5) report the full-sample estimates; in columns (2) and (6), we exclude districts in the London area; columns (3) and (7) exclude districts in the Lancashire area; in columns (4) and (8) we drop districts in the South-West. In Panel A, all models include district-by-decade and county-by-decade fixed effects; in Panel B, regressions include district-by-decade and technology class-by-decade fixed effects. Standard errors, clustered at the district level, are reported in parentheses.

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

◄ Back: Results ◄ Back: Appendix 5 – Robustness

	Reduced Form			TSLS			Overidentified TSLS		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\widehat{\text{Knowledge Exposure}}$	0.007*								
	(0.004)								
$\widehat{\text{Knowledge Exposure}}_{t-1}$		0.018***							
		(0.006)							
$\widehat{\text{Knowledge Exposure}}_{t-2}$			0.029**						
			(0.012)						
Knowledge Exposure				0.093*			0.322***		
				(0.051)			(0.052)		
Knowledge Exposure$_{t-1}$					0.180***			0.082*	
					(0.065)			(0.044)	
Knowledge Exposure$_{t-2}$						0.103**			0.032
						(0.041)			(0.038)
District-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Technology Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. of District-Class	11196	11196	11196	11196	11196	11196	11196	11196	11196
N. of Observations	55980	44784	33588	55957	44761	33586	55957	44761	33586
R^2	0.816	0.831	0.850	0.018	-0.006	-0.002	0.054	-0.001	-0.000
Mean Dep. Var.	1.079	1.202	1.312	1.079	1.203	1.312	1.079	1.203	1.312
Std. Beta Coef.	0.007	0.017	0.015	0.051	0.103	0.055	0.175	0.047	0.017
K-P F-stat				27.303	23.391	248.916	62.737	59.305	198.950
Sargan-Hansen J							24.274	6.131	31.636

TABLE E.8. RETURN INNOVATION RESULT USING THE LEAVEOUT INSTRUMENT

Notes. This table reports the estimated return innovation effect estimated using the leave-out instrument. The unit of observation is a district-technology class pair, observed at a decade frequency between 1880 and 1939. The dependent variable is the number of patents. The main explanatory variable is knowledge exposure. In columns (1–3), we report the reduced-form association between knowledge exposure constructed using predicted emigration flows using the leave-out instrument and the dependent variable; in columns (4–6), we report the associated two-stage least-squares estimates. In columns (7–9), instead, we exploit the railway and the leave-out instruments to estimate an over-identified instrumental variable regression. This allows us to report the associated Sargan-Hansen J-statistic to test the validity of the over-identifying restrictions. The Sargan-Hansen test does not refute the null that the instruments are valid. Each model includes district-by-decade and district-by-technology class fixed effects. Standard errors are reported in parentheses and are clustered at the district level.

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

◄ Back: Results ◄ Back: Appendix 5 – Robustness

	Baseline	Excluding Immigrants from...				
	(1)	(2)	(3)	(4)	(5)	(6)
		UK	UK + North Eu.	UK + South Eu.	UK + East Eu.	UK + Europe
Panel A. Second-Stage Estimates						
Knowledge Exposure	1.849***	0.454***	0.589***	0.428***	0.315**	0.487***
	(0.174)	(0.125)	(0.170)	(0.096)	(0.160)	(0.139)
N. of Observations	78876	78876	78876	78876	78876	78876
Mean Dep. Var.	11.768	11.768	11.768	11.768	11.768	11.768
K-P F-statistic		39.267	30.074	346.557	14.672	52.663
Panel B. First-Stage Estimates						
Knowledge Exposure (No Northern Europe + UK)			0.215***			
			(0.039)			
Knowledge Exposure (No Southern Europe + UK)				0.889***		
				(0.048)		
Knowledge Exposure (No Eastern Europe + UK)					0.481***	
					(0.126)	
Knowledge Exposure (No Europe + UK)						3.103***
						(0.428)
N. of Observations		78876	78876	78876	78876	78876
Mean Dep. Var.		3.063	3.063	3.063	3.063	3.063
District-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
Technology Class FE	Yes	Yes	Yes	Yes	Yes	Yes

TABLE E.9. RETURN INNOVATION RESULT USING THE MODIFIED LEAVEOUT INSTRUMENTS

Notes. This table reports the instrumental variable estimates of the effect of knowledge exposure on innovation using modified versions of the leave-out instrument. The unit of observation is a district-technology class pair, observed at a decade frequency between 1880 and 1939. The dependent variable is the number of patents. The explanatory variable is knowledge exposure. In column (1), we report the OLS correlation. In columns (2–6), we construct predicted bilateral emigrant flows using county-level immigration shocks that exclude immigrants from different parts of the world: in (2), we exclude only immigrants from UK nations; in (3), we exclude the UK immigrants along with those from other Northern Europe countries; in (4), we exclude immigrants from the UK and Southern Europe; in (5), UK and Eastern Europe immigrants are excluded; in (6), we exclude all European immigrants. Panel A reports the second-stage estimates; Panel B reports the associated first-stage estimates. All regressions include district-by-decade and technology class fixed effects. Standard errors, clustered at the district level, are displayed in parentheses.

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

◄ Back: Results ◄ Back: Appendix 5 – Robustness

	10-Year Similarity			log(Similarity)			Net of Year-Technology FE		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	OLS	RF	TSLS	OLS	RF	TSLS	OLS	RF	TSLS
Panel A. Dependent variable: "Copying" (Similarity with Previous US Patents)									
Knowledge Exposure$_t$	0.296***		0.122	0.953***		0.935***	0.340***		0.334***
	(0.045)		(0.096)	(0.106)		(0.151)	(0.042)		(0.057)
$\widehat{\text{Knowledge Exposure}}_t$		0.036			0.276***			0.099***	
		(0.029)			(0.051)			(0.019)	
R^2	0.733	0.767	0.003	0.752	0.795	0.046	0.681	0.692	0.023
Mean Dep. Var.	36.221	32.577	32.585	62.691	49.809	49.818	22.720	17.962	17.964
Std. Beta Coef.	0.165	0.016	0.065	0.340	0.101	0.399	0.303	0.089	0.352
Panel B. Dependent Variable: "Originality" (Similarity with Subsequent US Patents w.r.t. Previous US Patents)									
Knowledge Exposure$_t$	0.195***		0.270***	0.146***		0.158***	0.053***		0.076***
	(0.022)		(0.038)	(0.016)		(0.025)	(0.006)		(0.011)
$\widehat{\text{Knowledge Exposure}}_t$		0.080***			0.046***			0.022***	
		(0.013)			(0.008)			(0.004)	
R^2	0.684	0.748	0.034	0.752	0.791	0.049	0.571	0.522	0.011
Mean Dep. Var.	11.709	9.049	9.051	9.572	7.628	7.630	3.305	2.635	2.635
Std. Beta Coef.	0.338	0.145	0.583	0.344	0.109	0.439	0.285	0.105	0.422
District-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Technology Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. of District-Class	11268	11214	11214	11268	11214	11214	11268	11214	11214
N. of Observations	67553	56070	56050	67553	56070	56050	67553	56070	56050
K-P F-stat			103.344			103.344			103.344

TABLE E.10. ROBUSTNESS ANALYSIS ON THE EFFECT OF EXPOSURE TO US TECHNOLOGY ON THE SIMILARITY AND ORIGINALITY OF INNOVATION IN GREAT BRITAIN

Notes. This table displays the association between innovation and exposure to US knowledge. The unit of observation is a district-technology class pair, observed at a decade frequency between 1880 and 1939. In Panel A, the dependent variable is the text similarity between UK patents and US patents issued before ("copying"); in Panel B, the dependent variable is the similarity of UK patents with US patents granted in the subsequent years, over the similarity of UK patents with US patents granted in the preceding years ("originality"). The similarity measure is akin to Kelly *et al.* (2021). In columns (1), (4), and (7) we report the OLS regressions; columns (2), (5), and (8) report the reduced-form regressions; columns (3), (6), and (9) display the two-stage least-squares coefficients. In columns (1–3), the dependent variable is the baseline, except that we compute similarities over a ten-year window compared to the baseline five; in columns (4–6), we take the log of the patent-level similarity measure; in columns (7–9), remove year-by-technology class fixed effects from the patent-level similarity metrics. Each model includes district-by-decade and district-by-technology class fixed effects. Standard errors are reported in parentheses and are clustered at the district level.

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

◄ Back: Results ◄ Back: Appendix 2 – Patents

	Top 1% Synthetic Shocks			Top 0.5% Synthetic Shocks			Top 0.1% Synthetic Shocks		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Top 50%	Top 75%	Top 90%	Top 50%	Top 75%	Top 90%	Top 50%	Top 75%	Top 90%
Innovation Shock (Above 50%) × Post	0.187**			0.299***			0.617***		
	(0.083)			(0.098)			(0.126)		
Innovation Shock (Above 75%) × Post		0.224***			0.377***			0.617***	
		(0.080)			(0.081)			(0.126)	
Innovation Shock (Above 90%) × Post			0.326			0.825***			0.532***
			(0.269)			(0.229)			(0.175)
District-by-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
District-by-Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Class-by-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Number of Counties	189586	362024	426147	217975	381438	434687	431467	431467	445120
Number of Observations	5247	9187	10671	6762	9786	10900	10834	10834	11128
Mean Dep. Var.	1.022	1.106	1.586	1.410	1.3	1.686	2.064	2.064	2.020

TABLE E.11. TRIPLE DIFFERENCES ESTIMATED EFFECT OF US SYNTHETIC SHOCKS ON UK INNOVATION: ALTERNATIVE THRESHOLDS

Notes. This table displays the effect of US innovation shocks on innovation activity in the UK. The unit of observation is a district-technology class pair observed at a yearly frequency between 1900 and 1939. The dependent variable is the number of patents. The treatment variable is equal to one for district-technology class pairs after a synthetic innovation shock in a given technology class is observed in counties where the district has above k-percentile emigrants. We consider three different thresholds for k: above the median, above the top 25%, and above the top 10%. A synthetic shock is observed whenever the residualized patenting activity in a given county-technology class pair is in the top ℓ-percentile of the residualized patenting activity distribution. We consider three such *ell*: top 1%, in columns (1–3), top 0.5%, in columns (4–6), and top 0.1%, in columns (7–9). Since the treatment timing is staggered, we estimate the models using the imputation estimator developed by Borusyak *et al.* (2021). All models include district-by-year, district-by-technology class, and technology class-by-year fixed effects; standard errors, clustered two-way by district and technology class, are shown in parentheses.

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

◄ Back: Empirical Strategy ◄ Back: Results ◄ Back: Appendix 5 – Robustness

	Double Differences		Triple Differences				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
					No London	No Lancs	No S/W
Influenza Emigration × Post	0.008*						
	(0.004)						
1(Q. of Influenza Emigration > 75) × Post		0.980**					
		(0.463)					
Influenza Emigration × Post × Pharma			0.004**				
			(0.002)				
1(Q. of Influenza Emigration > 75) × Post × Pharma				0.584***	0.396**	0.671***	0.423**
				(0.163)	(0.139)	(0.173)	(0.156)
District FE	Yes	Yes	–	–	–	–	–
Year FE	Yes	Yes	–	–	–	–	–
District-Year FE	–	–	Yes	Yes	Yes	Yes	Yes
District-Class FE	–	–	Yes	Yes	Yes	Yes	Yes
Class-Year FE	–	–	Yes	Yes	Yes	Yes	Yes
N. of District-Class	631	631	10727	10727	10217	9384	10047
N. of Observations	18930	18930	321810	321810	306510	281520	301410
Classes in Sample	Pharma	Pharma	All	All	All	All	All
R^2	0.544	0.544	0.668	0.668	0.616	0.653	0.679
Mean Dep. Var.	0.927	0.927	0.763	0.763	0.559	0.721	0.706
Std. Beta Coef.	0.082	0.082	0.014	0.016	0.015	0.018	0.010

TABLE E.12. DOUBLE AND TRIPLE DIFFERENCES ESTIMATED EFFECT OF THE GREAT INFLUENZA PANDEMIC IN THE US ON INNOVATION IN THE UK: ROBUSTNESS ANALYSIS

Notes. This table displays the effect of the Great Influenza Pandemic shock on innovation activity in the UK. In columns (1–2), the observation unit is a district; in columns (3–7), the observation unit is a pair district-technology class; units are observed at a yearly frequency between 1900 and 1939. The dependent variable is the number of patents. In column (1), the treatment variable is an interaction between an influenza exposure term equal to the share of emigrants to counties in the top 25% of the flu-related excess mortality distribution and a post-Influenza indicator; in column (2), we code exposure as a binary variable equal to one for districts in the top 25% of the continuous exposure distribution. In columns (3) and (5–7), the treatment term in column (1) is interacted with an indicator variable for pharmaceutical patents; in column (4), we interact the treatment term in column (2) with the same pharmaceutical indicator. Regressions in (1–4) report full-sample estimates; in columns (5), (6), and (7), instead, we drop districts in the London, Lancashire, and South-West areas, respectively. Regressions in columns (1–2) include district and year fixed effects; regressions in columns (3–7) include district-by-year, technology class-by-year, and district-by-technology class fixed effects. Standard errors, reported in parentheses, are clustered by district in columns (1–2) and two-way by district and technology class in (3–7).

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

◄ Back: Empirical Strategy ◄ Back: Results ◄ Back: Appendix 5 – Robustness

	Ordinary Least Squares			Reduced Form			Two-Stages Least-Squares		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. Dependent variable: "Copying" (Similarity with Previous US Patents)									
Knowledge Exposure$_t$	0.117***			0.029*			0.098*		
	(0.024)			(0.017)			(0.053)		
Knowledge Exposure$_{t-1}$		0.161***			0.029***			0.098***	
		(0.026)			(0.006)			(0.023)	
Knowledge Exposure$_{t-2}$			0.196***			0.013			0.044
			(0.027)			(0.015)			(0.050)
Mean Dep. Var.	17.043	17.981	19.842	14.906	14.278	19.919	14.910	14.283	19.925
Std. Beta Coef.	0.126	0.140	0.139	0.025	0.030	0.008	0.101	0.121	0.031
R^2	0.694	0.569	0.601	0.712	0.712	0.599	-0.002	0.007	0.002
Panel B. Dependent Variable: "Originality" (Similarity with Subsequent US Patents w.r.t. Previous US Patents)									
Knowledge Exposure$_t$	0.175***			0.065***			0.219***		
	(0.019)			(0.011)			(0.032)		
Knowledge Exposure$_{t-1}$		0.084***			0.015*			0.050*	
		(0.014)			(0.008)			(0.027)	
Knowledge Exposure$_{t-2}$			0.097***			-0.029			-0.099
			(0.017)			(0.025)			(0.087)
Mean Dep. Var.	10.885	13.815	15.851	8.584	12.687	15.913	8.586	12.691	15.918
Std. Beta Coef.	0.341	0.127	0.119	0.132	0.020	-0.030	0.531	0.079	-0.121
R^2	0.699	0.672	0.710	0.775	0.718	0.708	0.047	0.006	-0.013
District-Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-Technology Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. of District-Class	11268	11268	11268	11214	11214	11214	11214	11214	11214
N. of Observations	67553	67553	56299	56070	56070	56070	56050	56050	56050
K-P F-stat							103.344	103.344	103.344

TABLE E.13. EFFECT OF EXPOSURE TO US TECHNOLOGY ON THE SIMILARITY AND ORIGINALITY OF INNOVATION IN GREAT BRITAIN

Notes. This table displays the association between innovation and exposure to US knowledge. The unit of observation is a district-technology class pair, observed at a decade frequency between 1880 and 1939. In Panel A, the dependent variable is the text similarity between UK patents and US patents issued five years before ("copying"); in Panel B, the dependent variable is the similarity of UK patents with US patents granted in the subsequent five years, over the similarity of UK patents with US patents granted in the preceding five years ("originality"). The similarity measure is akin to Kelly *et al.* (2021). In columns (1–3), we estimate the OLS correlation with the observed measure of knowledge exposure; in columns (4–6), we estimate the reduced-form association with the railway-based instrument of knowledge exposure through OLS; columns (7–9) report the two-stage least-squares estimate. Each model includes district-by-decade and district-by-technology class fixed effects. Standard errors are reported in parentheses and are clustered at the district level.

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

◄ Back: Results

	Synthetic Shocks				Great Influenza Pandemic Shock			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Full Sample	No London	No Lancs	No S/W	Full Sample	No London	No Lancs	No S/W
Panel A. Copying (Similarity with Previous US Patents)								
Synth. Shock × Post × Emigrants	0.266***	0.169***	0.328***	0.259***				
	(0.063)	(0.039)	(0.066)	(0.066)				
Pharma × Post × Emigrants					0.763***	0.450***	0.884***	0.664***
					(0.210)	(0.179)	(0.199)	(0.206)
Panel B. Originality (Similarity with Subsequent US Patents w.r.t. Similarity with Previous US Patents)								
Synth. Shock × Post × Emigrants	0.416***	0.238***	0.580***	0.375***				
	(0.140)	(0.094)	(0.148)	(0.144)				
Pharma × Post × Emigrants					5.249***	3.827***	5.538***	4.009***
					(0.990)	(0.868)	(1.068)	(0.925)
District-by-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District-by-Class FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class-by-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Units	393046	382153	343450	375850	429080	408680	375360	401880
Number of Observations	10029	9697	8760	9547	10727	10217	9384	10047
Mean Dep. Var.	0.498	0.376	0.460	0.472	0.791	0.577	0.737	0.733

Table E.14. Triple Differences Effect of Exposure to US Shocks on the Similarity Between UK and US Innovation

Notes. This table displays the effect of US innovation shocks on the similarity and originality of innovation activity in the UK compared to US patents. The unit of observation is a district-technology class pair observed at a yearly frequency between 1900 and 1939. In Panel A, the dependent variable is the text similarity between UK patents and US patents issued five years before ("copying"); in Panel B, the dependent variable is the similarity of UK patents with US patents granted in the subsequent five years, over the similarity of UK patents with US patents granted in the preceding five years ("originality"). The similarity measure is akin to Kelly *et al.* (2021). In columns (1–4), the independent variable is an indicator that, for a given district–technology, returns value one after a synthetic innovation shock in that technology class is observed in at least one county where the district has above-average out-migration. A synthetic innovation shock is observed whenever the residualized number of patents observed in the country is in the top 0.5% of the overall distribution. In columns (5–8), the independent variable is an indicator that returns value one for pharmaceutical patents only and only if emigration from the observed district to counties in the top quartile of the influenza mortality distribution is in the top quartile across districts. Both models are triple-difference designs. Models in columns (1–4) are staggered designs and are estimated using the imputation estimator by Borusyak *et al.* (2021). In columns (2) and (6), we drop districts in the London area; in columns (3) and (7), we exclude districts in the Lancashire area; in columns (4) and (8), we drop districts in the South-West area. All models include district-by-year, district-by-technology class, and technology class-by-year fixed effects; standard errors, clustered two-way by district and technology class, are shown in parentheses.

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

◄ Back: Results

	Baseline Sample				Dropping Individuals in...		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
					London	Lancashire	South-West
Panel A. All Emigrants							
Neighborhood Emigrant × Post	0.120**	0.146**	0.133**	11.846*	0.079	0.142**	0.167***
	(0.059)	(0.068)	(0.059)	(6.144)	(0.065)	(0.062)	(0.061)
Std. Beta Coef.	0.016	0.019	0.018	0.155	0.011	0.020	0.022
Panel B. Only Non-Return Emigrants							
Non-Return Neighborhood Emigrant × Post	0.148***	0.199***	0.160***	14.694**	0.061	0.172***	0.226***
	(0.056)	(0.062)	(0.058)	(6.293)	(0.061)	(0.059)	(0.058)
Std. Beta Coef.	0.019	0.025	0.020	0.186	0.008	0.023	0.028
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	–	Yes	Yes	Yes	Yes	Yes
Parish × Year FE	No	Yes	No	No	No	No	No
Matching	No	No	Yes	No	No	No	No
Sample	Full	Full	Full	Inventors	Full	Full	Full
N. of Individuals	473112	473112	469585	4224	410327	422230	352064
N. of Observations	9462240	9412502	9391700	84480	8206540	8444600	7041280
Mean Dep. Var.	0.890	0.892	0.893	99.716	0.794	0.836	0.893
S.D. Dep. Var.	40.291	40.337	40.351	414.695	37.439	39.126	41.333

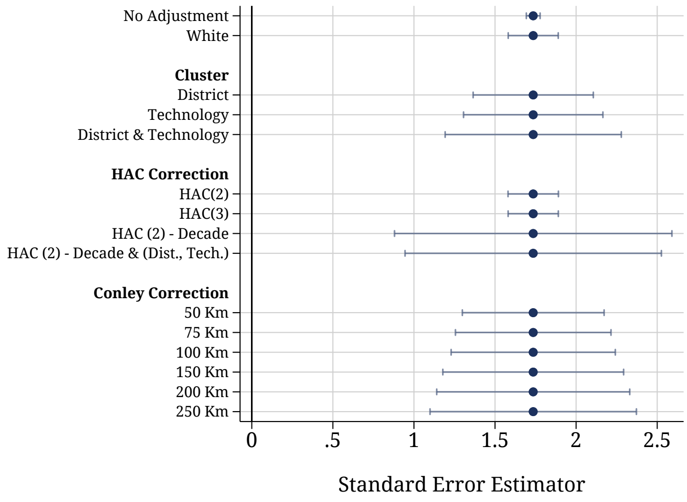
TABLE E.15. DIFFERENCE-IN-DIFFERENCES EFFECT OF NEIGHBORHOOD-LEVEL OUT-MIGRATION ON INNOVATION: ALTERNATIVE PROXIMITY THRESHOLD

Notes. This table reports the effect of neighborhood out-migration on innovation. The units of observation are individuals who are observed yearly between 1900 and 1920. In columns (1–3) and (5–7), the sample consists of the universe of males who did not emigrate over the period and that were at least 18 years old in 1900; in columns (4) and (8), we restrict the sample to inventors. The dependent variable is the number of patents obtained annually. In columns (1–4), the sample consists of individuals residing in all England and Wales divisions; in columns (5–7), we exclude the top tree-patents producing areas: London, Lancashire, and the South-West. In Panel A, the independent variable is an indicator that, for a given individual, returns value one after at least one person that was living in the same neighborhood as the individual migrates to the United States; in Panel B, we restrict to emigrants that never return in the period of observation. In this context, "neighborhood" refers to emigrants within a range of 100 meters from the individual in the sample. Each model includes individual and—at least—year fixed effects; in column (2), we include parish-by-year fixed effects; in column (3), individuals are weighted by their coarsened exact matching weight. The estimates are obtained using the method discussed in Borusyak *et al.* (2021) to account for the staggered roll-out of the treatment across individuals. Standard errors, clustered at the district level, are reported in parentheses.

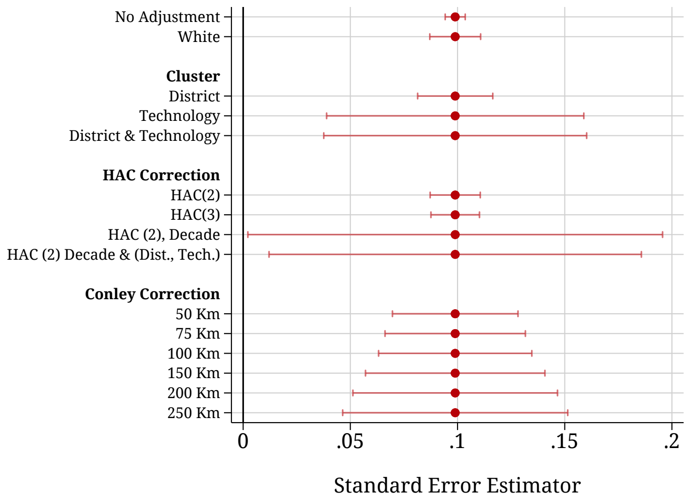
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$.

◄ Back: Mechanisms

E.5 Figures



(A) Measured Knowledge Exposure

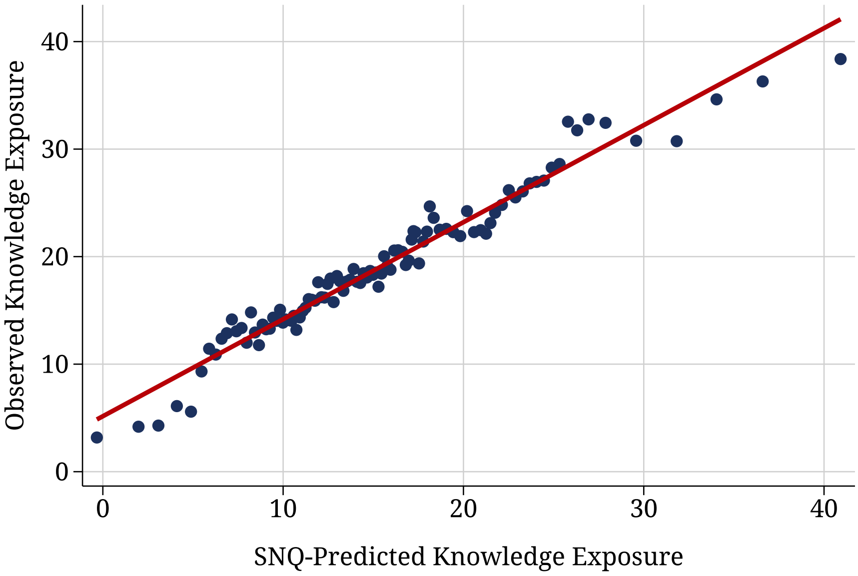


(B) Railway-Based Knowledge Exposure

FIGURE E.1. ALTERNATIVE STANDARD ERRORS ESTIMATORS OF THE RETURN INNOVATION RESULT

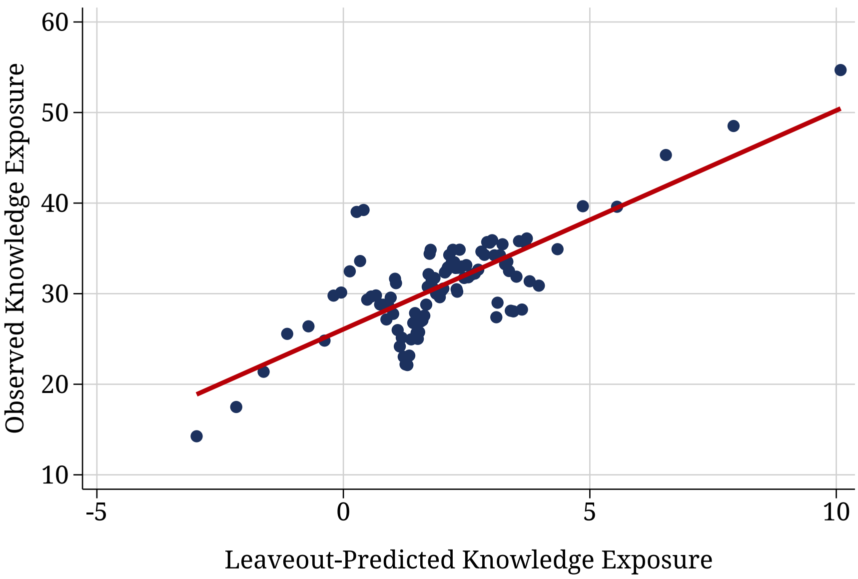
Notes. These figures report alternative estimates for the standard errors (SEs) of the regression between the number of patents and knowledge exposure. The unit of observation is a district-technology pair, observed at a decade frequency between 1880 and 1930. Models include district-by-technology and decade fixed effects. In Panel E.1a, the independent variable is measured knowledge exposure; Panel E.1b reports the estimated reduced-form coefficient between patents and the railway-based instrument. We report unadjusted SEs, robust to heteroskedasticity (White); clustered at the district, technology class, and two-way by district and technology class; robust to heteroskedasticity and autocorrelation of order 2 (HAC (2)), order 3 (HAC (3)); robust to heteroskedasticity and autocorrelation, and clustered by decade (HAC (2) - Decade) and two-way by decade and district-by-technology class (HAC (2) - Decade & (Dist., Tech.). Finally, we also report SEs that account for spatial autocorrelation at various orders (between 50 and 250 kilometers) following Conley (1999). Bands report 95% confidence intervals.

◄ Back: Results



Notes. Coefficient = 0.453 (Clust. Std. Err. = 0.034). R^2 = 0.806.

(A) Railway-Based Instrument



Notes. Coefficient = 1.529 (Clust. Std. Err. = 0.256). R^2 = 0.773.

(B) UK Innovation

FIGURE E.2. FIRST STAGE BINNED SCATTER PLOT

Notes. These figures are binned scatter plots of the association between actual and predicted knowledge exposure obtained using the railway-based instrument (Panel E.2a) and the leave-out instrument (Panel E.2b). The unit of observation is a district-technology class pair, at a decade frequency between 1880 and 1920. Graphs partial out district-by-decade and technology class fixed effects. We report the associated regression coefficients and standard errors, clustered at the district level, below each graph.

◄ Back: Appendix 5 – Robustness

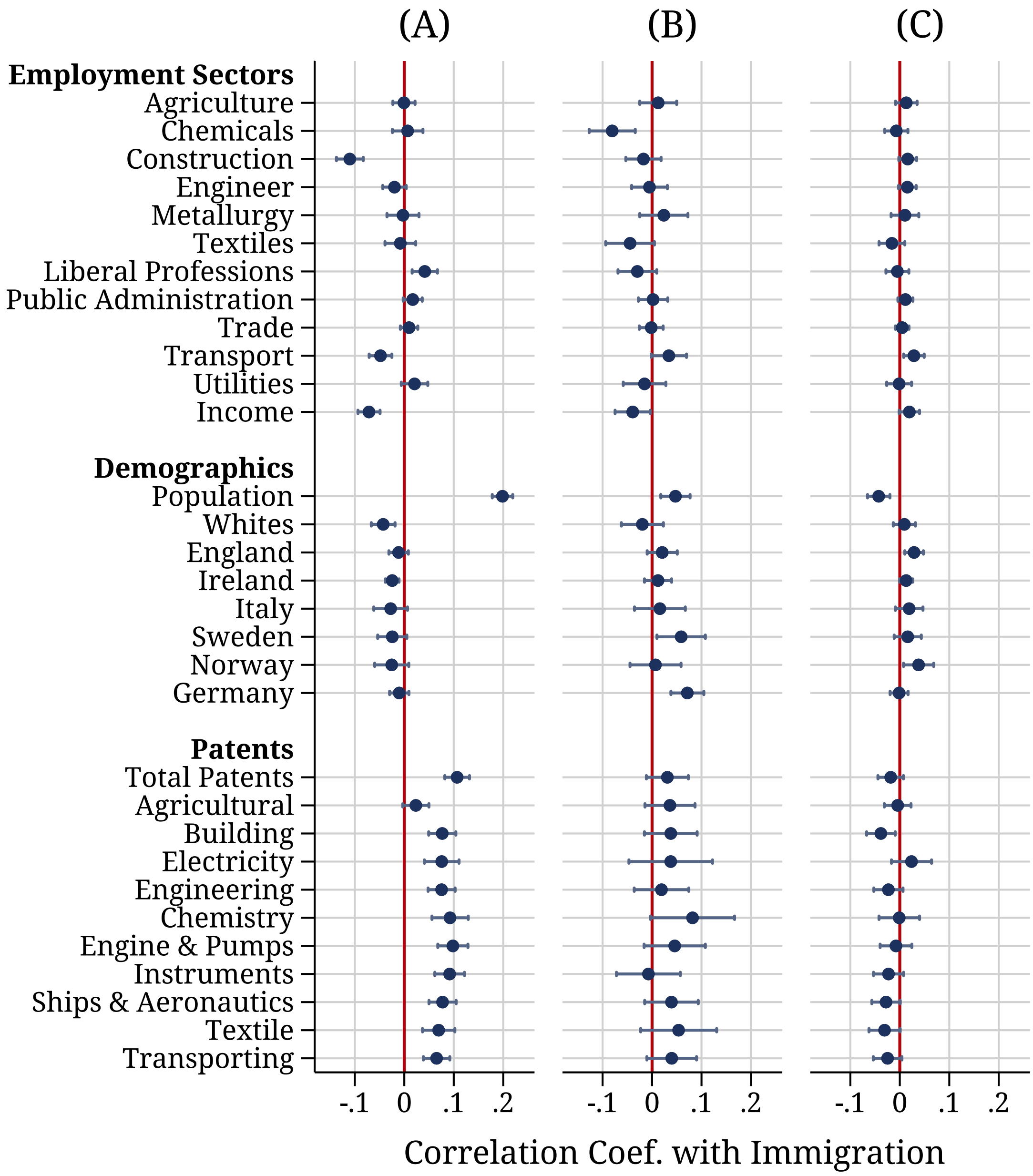


FIGURE E.3. SHOCK-LEVEL BALANCE TESTS FOR INSTRUMENTAL VARIABLE VALIDITY

Notes. This figure reports the correlation between county-level observable characteristics and the (predicted) immigrant share. The unit of observation is a county observed at a decade frequency between 1870 and 1920. Panel (A) refers to the observed immigrant share; Panel (B) refers to the immigrant share predicted from the railway-based shock constructed from the zero-stage estimates *à la* Sequeira *et al.* (2020); Panel (C) refers to the leave-out shocks used to construct the alternative leave-out instrument. Each dot reports the correlation between the row variable and the immigrant share, lagged by one decade. Variables are standardized for the sake of readability. Each model includes county and state-by-decade fixed effects. Standard errors are clustered at the county level. Bands report 95% confidence intervals.

◄ Back: Empirical Strategy ◄ Back: Appendix 5 – Robustness

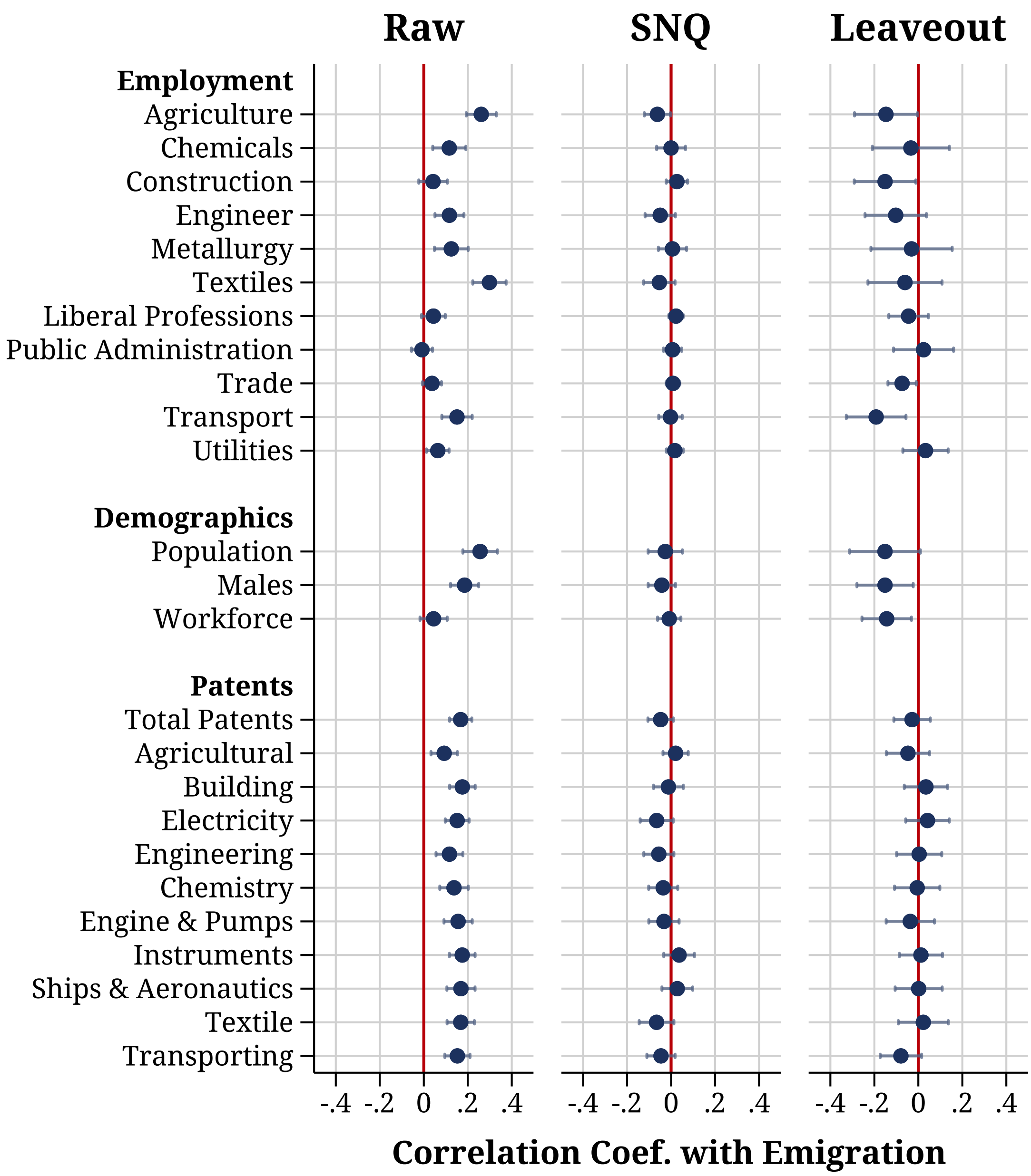


FIGURE E.4. DISTRICT-LEVEL BALANCE TESTS FOR INSTRUMENTAL VARIABLE VALIDITY

Notes. This figure reports the correlation between district-level observable characteristics and the (predicted) number of emigrants. The unit of observation is a district observed at a decade frequency between 1870 and 1920. Panel (A) refers to the observed number of emigrants; Panel (B) refers to the predicted emigrant outflow obtained from the railway-based instrument; Panel (C) refers to the leave-out instrument. Each dot reports the correlation between the row variable and out-migration, lagged by one decade. Variables are standardized for the sake of readability. Each model includes district and decade fixed effects. Standard errors are clustered at the county level. Bands report 95% confidence intervals.

◄ Back: Empirical Strategy ◄ Back: Appendix 5 – Robustness

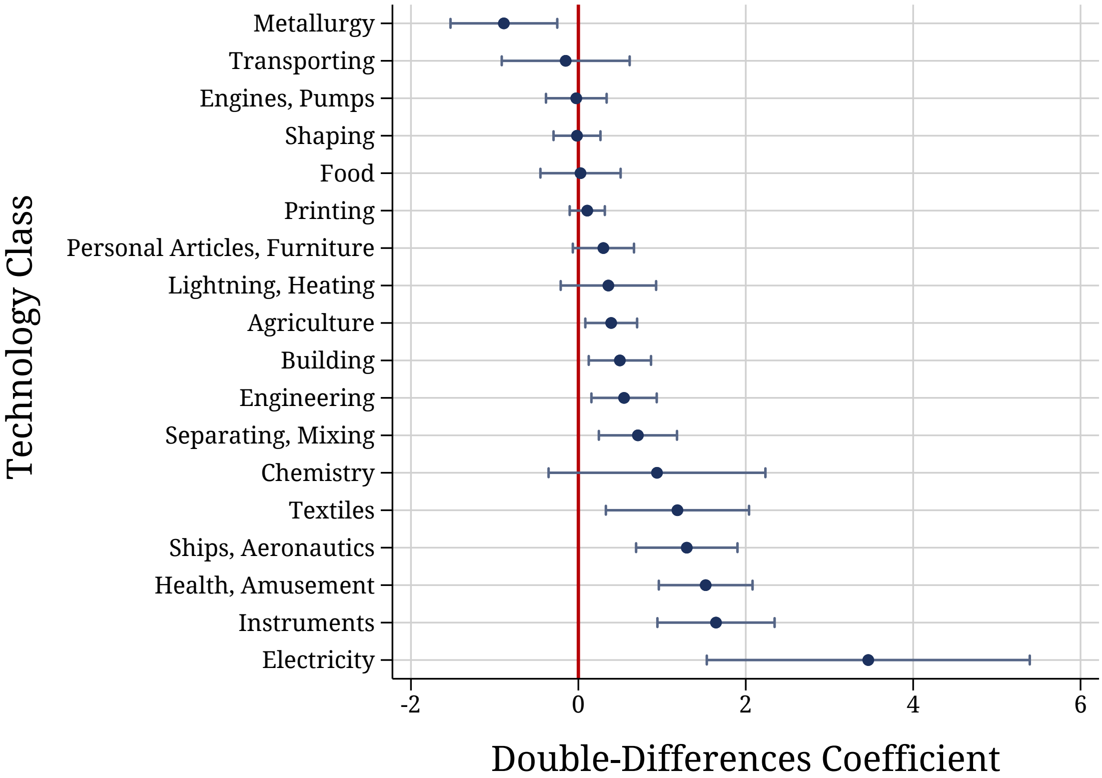
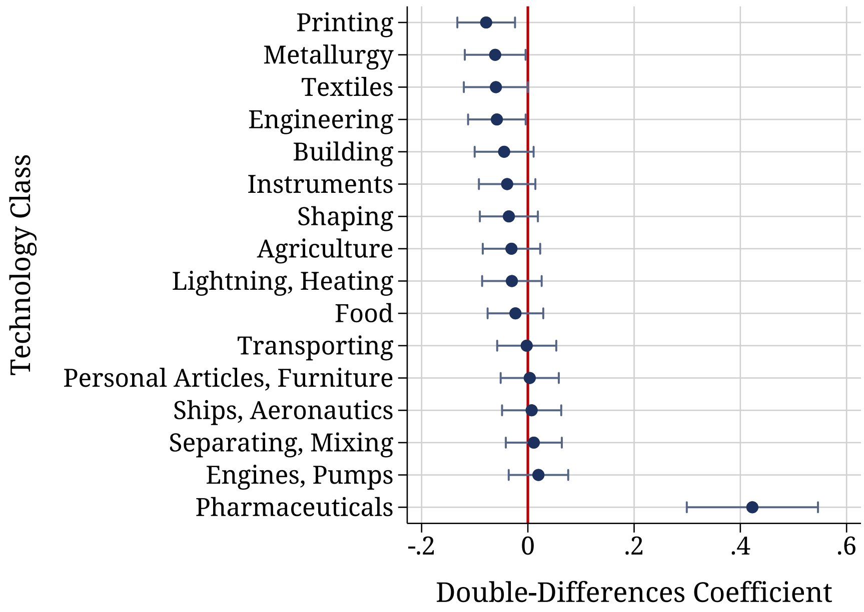


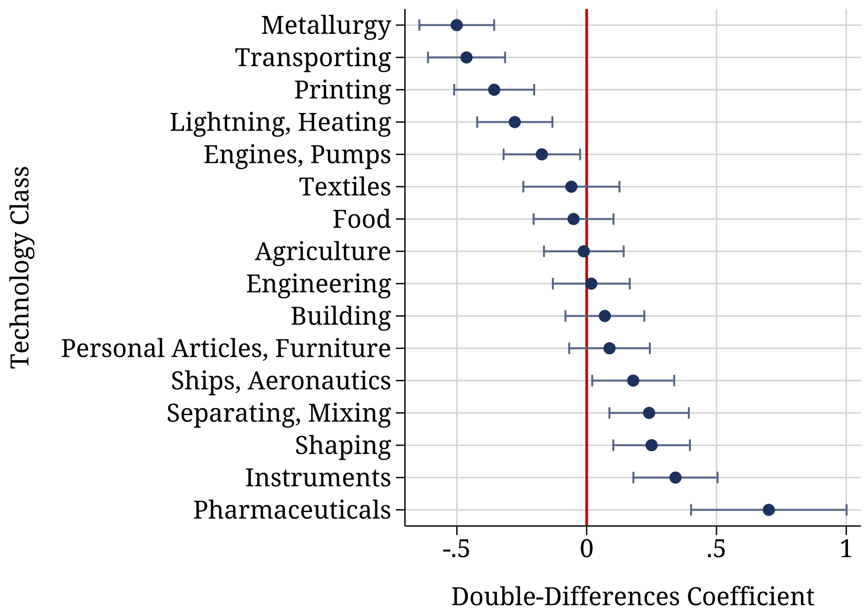
FIGURE E.5. EFFECT OF SYNTHETIC INNOVATION SHOCKS ACROSS TECHNOLOGY CLASSES

Notes. This figure reports the effect of synthetic innovation shocks on innovation in the UK by technology class. Each dot reports one double-differences estimated effect of the baseline exposure treatment with innovation; in each row, the treatment is activated whenever a district has above-median. The unit of observation is thus a district, observed at a yearly frequency between 1900 and 1993. Regressions include district and year fixed effects, and standard errors are clustered at the district level. Bands report 95% confidence intervals.

◄ Back: Results



(A) US Innovation



(B) UK Innovation

FIGURE E.6. EFFECT OF THE INFLUENZA SHOCK ACROSS TECHNOLOGY CLASSES

Notes. This figure reports the effect of the Influenza shock on innovation, by technology classes, in the US (Panel E.6a) and in the UK (Panel E.6b). Each dot reports one triple-differences estimated effect of the baseline exposure treatment with innovation; in each row, exposure is interacted with a sector-specific dummy variable. If the shock only impacted innovation in pharmaceuticals, we would expect each coefficient but the pharmaceutical one to be statistically equal to zero. Regressions are saturated with fixed effects; standard errors are two-way clustered at the technology class and county (Panel E.6a) or district (Panel E.6b) level. Bands report 95% confidence intervals.

◄ Back: Empirical Strategy ◄ Back: Results ◄ Back: Appendix 5 – Robustness

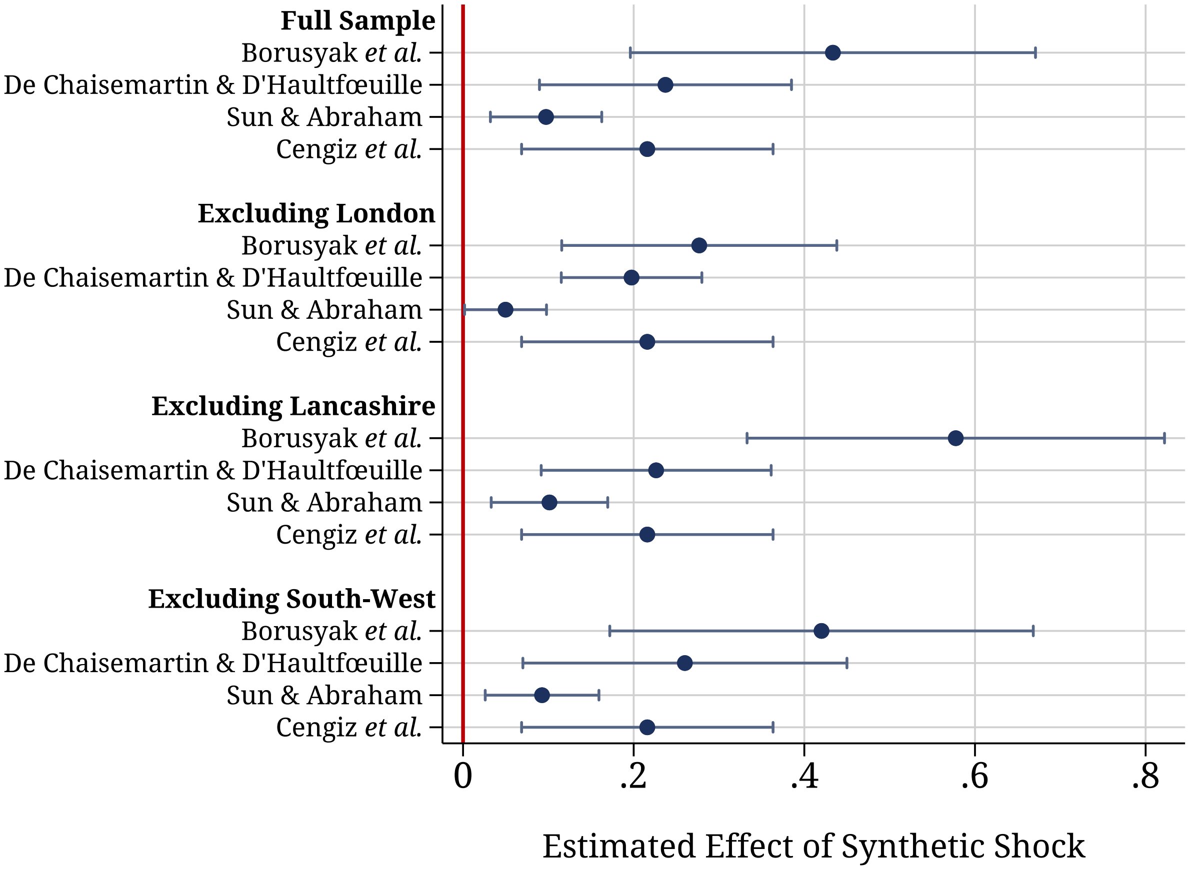


FIGURE E.7. ALTERNATIVE STAGGERED TRIPLE DIFFERENCES ESTIMATORS FOR THE EFFECT OF US SYNTHETIC SHOCKS ON INNOVATION

Notes. This figure reports the estimated effect of synthetic innovation shocks in US counties on innovation activity in the UK, using alternative estimators that explicitly allow for the staggered treatment roll-out design. The unit of observation is a district-technology class pair observed at a yearly frequency between 1900 and 1939. The dependent variable is the number of patents. The treatment variable is an indicator that, for a given district-technology, returns value one after a synthetic innovation shock in that technology class is observed in at least one county where the district has above-average out-migration. A synthetic innovation shock is observed whenever the residualized number of patents observed in the country is in the top 0.5% of the overall distribution. We estimate the models on the full sample of districts, as well as excluding the top three areas in terms of patents granted: London, Lancashire, and the South-West. We report the estimates obtained using four estimators that allow for the inclusion of all the triple differences interactions of the fixed effects: Borusyak *et al.* (2021), De Chaisemartin and D'Haultfœuille (2022), Cengiz *et al.* (2022), and Sun and Abraham (2021). Standard errors are clustered at the district and technology class levels. Bands report 95% confidence intervals.

◄ Back: Empirical Strategy ◄ Back: Results ◄ Back: Appendix 5 – Robustness

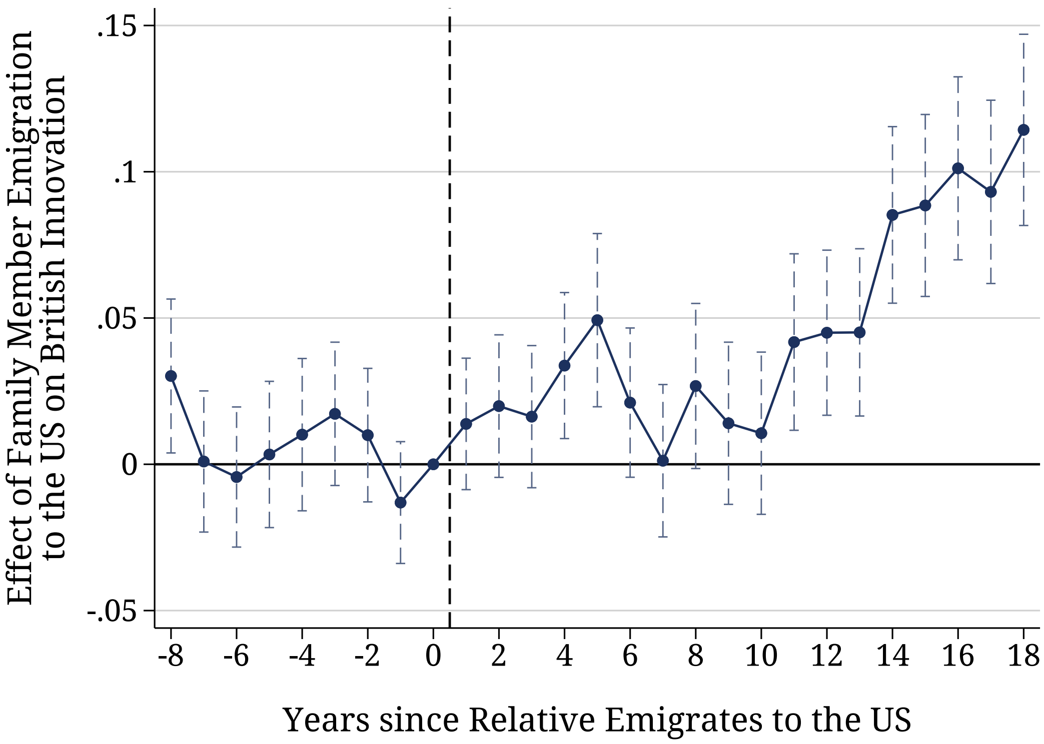


FIGURE E.8. FLEXIBLE TRIPLE DIFFERENCES EFFECT OF FAMILY MEMBER OUT-MIGRATION ON INNOVATION

Notes. This figure reports the effect of transatlantic emigration on innovation by inventors with the same surname as the emigrant. The unit of observation is a surname-county couple, observed at a year frequency between 1870 and 1929. The dependent variable is the number of patents granted to inventors with a given surname in a given county and year. The treatment is an interaction between year dummies and a variable that takes a value of one the first time at least one individual from a given county and with a given surname emigrates to the US, and zero otherwise. Each regression includes county-by-surname, surname-by-year, and county-by-year fixed effects. Standard errors are clustered at the surname level. Bands report 90% confidence intervals.

◄ Back: Mechanisms

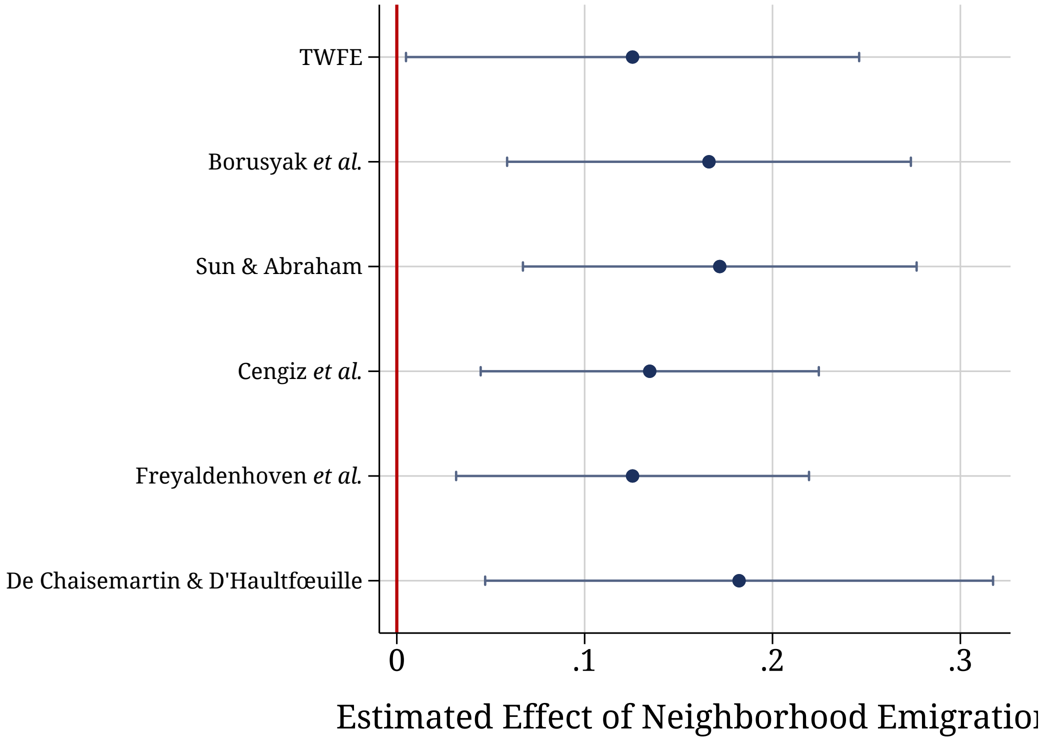
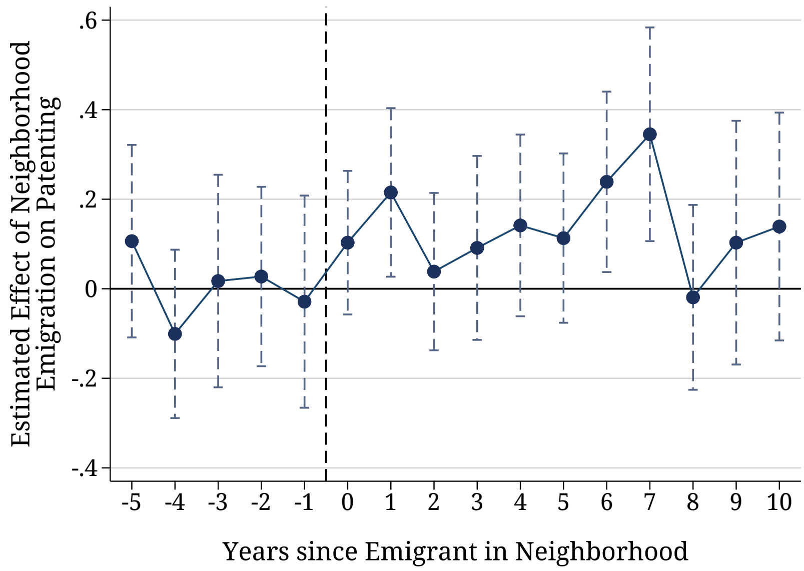


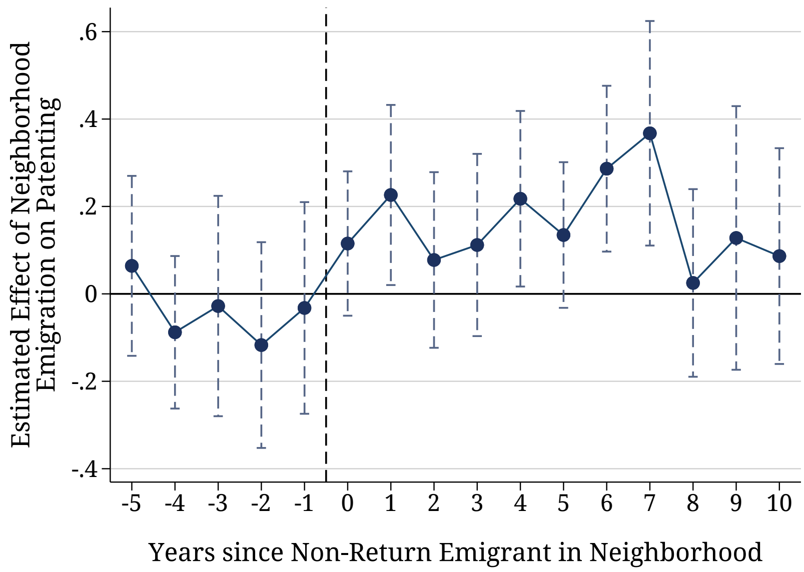
FIGURE E.9. ALTERNATIVE STAGGERED DOUBLE DIFFERENCES ESTIMATORS FOR THE EFFECT OF NEIGHBORHOOD OUT-MIGRATION ON INNOVATION

Notes. These figures report the effect of neighborhood out-migration on innovation. The units of observation are individuals observed at a yearly frequency between 1900 and 1920. The sample consists of all males who did not emigrate over the period and aged at least 18 in 1900. The dependent variable is the number of patents obtained every year. The treatment variable is an indicator that returns value one after at least one person living in the same neighborhood as the individual migrates to the United States. We report the estimates obtained using six estimators that allow staggered roll-out of treatment assignment: the baseline two-way fixed effects (TWFE) estimator, Borusyak *et al.* (2021), Sun and Abraham (2021), Cengiz *et al.* (2022), Freyaldenhoven *et al.* (2019), and De Chaisemartin and D'Haultfœuille (2022). Standard errors are clustered at the district level. Bands report 90% confidence intervals.

◄ Back: Mechanisms



(A) All Emigrants



(B) Non-Return Migrants

FIGURE E.10. FLEXIBLE DIFFERENCE-IN-DIFFERENCES EFFECT OF NEIGHBORHOOD-LEVEL OUT-MIGRATION ON INNOVATION

Notes. These figures report the effect of neighborhood out-migration on innovation. The units of observation are individuals observed at a yearly frequency between 1900 and 1920. The sample consists of all males who did not emigrate over the period and aged at least 18 in 1900. The dependent variable is the number of patents obtained every year. In Panel E.10a, the treatment variable is an indicator that returns value one after at least one person that was living in the same neighborhood as the individual migrates to the United States; in Panel E.10b, we restrict to emigrants that never return in the period of observation. Each model includes individual and parish-by-year fixed effects. Standard errors are clustered at the district level. The estimates are obtained using the estimator discussed in Borusyak *et al.* (2021). Bands report 95% confidence intervals.

◄ Back: Mechanisms

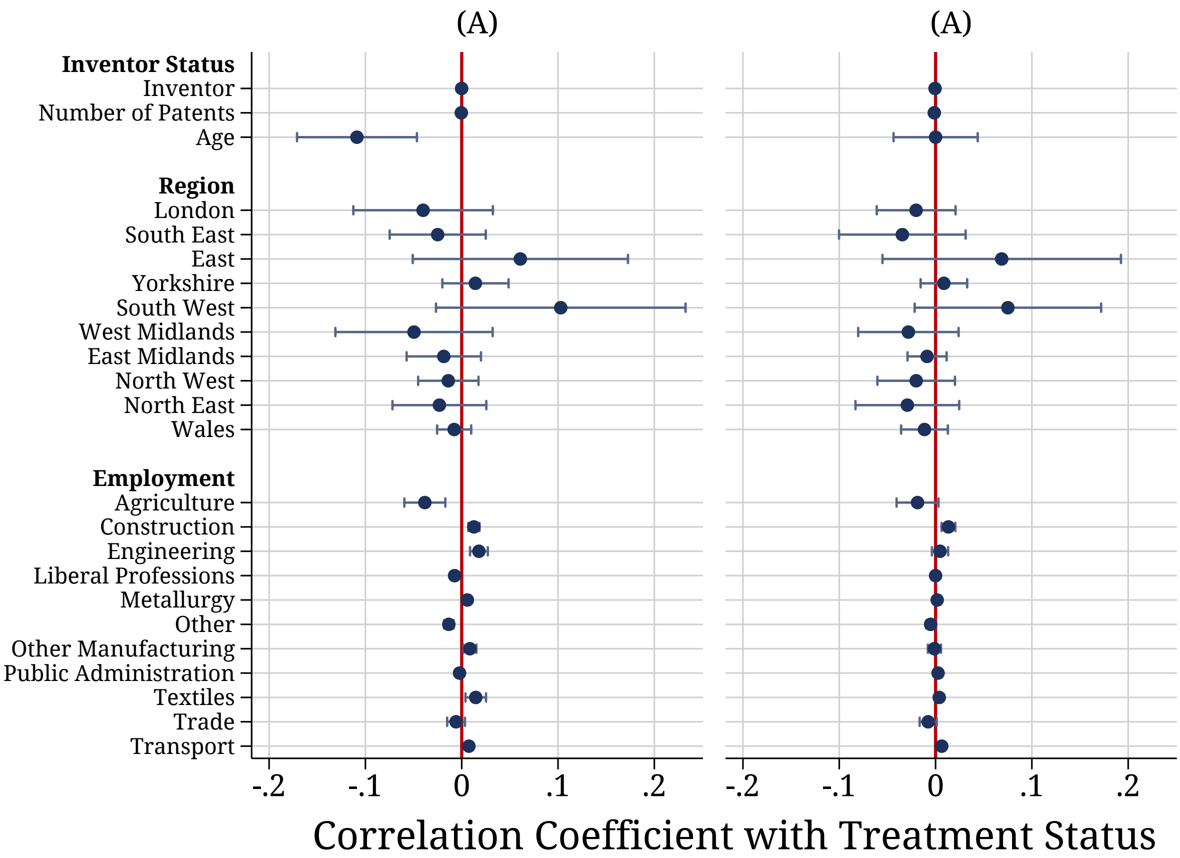


FIGURE E.11. CO-VARIATE BALANCE FOR INDIVIDUAL-LEVEL DESIGN

Notes. These figures report the correlations between individual-level observable characteristics and treatment status in the individual-level analysis. The units of observation are individuals observed at a yearly frequency between 1900 and 1920. The sample consists of all males who did not emigrate over the period and aged at least 18 in 1900. Variables are observed in the 1911 census. Hence some of them are not pre-determined when the treatment initiates. Each dot reports the correlation between the row variable and a dummy variable equal to one if the individual is treated in the observation period and zero otherwise. Variables are standardized for readability. Panel (A) reports the unweighted correlation; in Panel (B), individuals are weighted by their CEM weights. Standard errors are clustered by division. Bands report 95% confidence intervals.

◄ Back: Mechanisms

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.

AGHION, P., A. BERGEAUD, M. LEQUIEN and M. J. MELITZ (2018). "The Impact of Exports on Innovation: Theory and Evidence." *NBER Working Paper*, (No. w24600).

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

ANDERSSON, D. E., M. KARADJA and E. PRAWITZ (2022). "Mass Migration and Technological Change." *Journal of the European Economic Association*.

ATKIN, D., A. K. KHANDELWAL and A. OSMAN (2017). "Exporting and Firm Performance: Evidence from a Randomized Experiment." *The Quarterly Journal of Economics*, 132(2): 551–615.

AUTOR, D., D. DORN, G. H. HANSON, G. PISANO and P. SHU (2020). "Foreign Competition and Domestic Innovation: Evidence from US Patents." *American Economic Review: Insights*, 2(3): 357–74.

BAILEY, M. J., C. COLE, M. HENDERSON and C. MASSEY (2020). "How Well Do Automated Linking Methods Perform? Lessons from US Historical Data." *Journal of Economic Literature*, 58(4): 997–1044.

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 (2022). "Historical Newspaper Data: A Researcher's Guide and Toolkit." *NBER Working Paper*, (No. w30135).

BERGEAUD, A. and C. VERLUISE (2022). "A New Dataset to Study a Century of Innovation in Europe and in the US." *Working Paper*.

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

BERKES, E., D. M. COLUCCIA, G. DOSSI and M. P. SQUICCIARINI (2023). "Dealing with Adversity: Religiosity or Science? Evidence from the Great Influenza Pandemic." *Working Paper*.

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

BLOOM, N., M. DRACA and J. VAN REENEN (2016). "Trade Induced Technical Change? The Impact of Chinese Imports on Innovation, IT and Productivity." *The Review of Economic Studies*, 83(1): 87–117.

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

BORUSYAK, K., X. JARAVEL and J. SPIESS (2021). "Revisiting Event Study Designs: Robust and Efficient Estimation." *Working Paper*.

BUSTOS, P. (2011). "Trade Liberalization, Exports, and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinian Firms." *American Economic Review*, 101(1): 304–40.

CALLAWAY, B. and P. H. SANT'ANNA (2021). "Difference-in-Differences with multiple time periods." *Journal of Econometrics*, 225(2): 200–230.

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

CENGIZ, D., A. DUBE, A. LINDNER and D. ZENTLER-MUNRO (2022). "Seeing Beyond the Trees: Using Machine Learning to Estimate the Impact of Minimum Wages on Labor Market Outcomes." *Journal of Labor Economics*, 40(S1): S203–S247.

CHEN, J. and J. ROTH (2022). "Log-like? ATEs Defined with Zero Outcomes are (Arbitrarily) Scale-dependent." *Working Paper*.

COELLI, F., A. MOXNES and K. H. ULLTVEIT-MOE (2022). "Better, Faster, Stronger: Global Innovation and Trade Liberalization." *The Review of Economics and Statistics*, 104(2): 205–216.

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

CRUCINI, M. J. (1994). "Sources of Variation in Real Tariff Rates: The United States, 1900–1940." *The American Economic Review*, 84(3): 732–743.

DAVID, P. A. (1990). "The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox." *American Economic Review*, 80(2): 355–361.

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

DE CHAISEMARTIN, C. and X. D'HAULTFŒUILLE (2022). "Difference-in-Differences Estimators of Intertemporal Treatment Effects." *NBER Working Paper*, (No. w29873).

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

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).

EICHENGREEN, B. (1986). "The Political Economy of the Smoot-Hawley Tariff." *NBER Working Paper*, (No. w2001).

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

FINISHING PUBLICATIONS, L. (2018). "Early British Patents. A Cradle of Inventions. British Patents 1617 to 1895." MFIS [data collection].

FREYALDENHOVEN, S., C. HANSEN and J. M. SHAPIRO (2019). "Pre-event Trends in the Panel Event-Study Design." *American Economic Review*, 109(9): 3307–38.

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

JUHÁSZ, R. and C. STEINWENDER (2018). "Spinning the Web: The Impact of ICT on Trade in Intermediates and Technology Diffusion." *NBER Working Paper*, (No. w24590).

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.

LI, S., J. HU, Y. CUI and J. HU (2018). "DeepPatent: patent classification with convolutional neural networks and word embedding." *Scientometrics*, 117(2): 721–744.

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

MOKYR, J. (1998). "The Second Industrial Revolution, 1870-1914." In V. Castronovo, ed., "Storia dell'Economia Mondiale," pp. 219–245. Rome (Italy): Laterza.

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.

OLIVETTI, C., M. D. PASERMAN, L. SALISBURY and E. A. WEBER (2020). "Who Married,(To) Whom, and Where? Trends in Marriage in the United States, 1850-1940." *NBER Working Paper*, (No. 28033).

OTTAVIANO, G. I. P., G. PERI and G. C. WRIGHT (2018). "Immigration, Trade and Productivity in Services: Evidence from UK Firms." *Journal of International Economics*, 112: 88–108.

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.

SHU, P. and C. STEINWENDER (2019). "The Impact of Trade Liberalization on Firm Productivity and Innovation." *Innovation Policy and the Economy*, 19(1): 39–68.

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.

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

XU, S. (2018). "Bayesian Naïve Bayes classifiers to text classification." *Journal of Information Science*, 44(1): 48–59.