Online Appendix

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

Davide M. Coluccia & Gaia Dossi

January, 2025

Table of Contents

A Data Appendix

This section describes the data sources and the construction of the final sample (section A.I), the construction of the newly digitized patent data and the approach to link patents to the census (section A.II), and the methodology we developed to link the records of British immigrants in the United States to the UK census (section A.III).

A.I Summary of Data Sources

A.I.1 Patent Data

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

Importantly, we map 3-digit CPC classes to a coarser taxonomy of classes. To do that, we reduce them to functional units using the CPC classification scheme. The scheme is publicly available at the following `link`. To accommodate the historical context, we divide the transporting categories into two classes: ”Transporting”, which includes carriages, railways, and cars, and “Ships and Aeronautics”. Moreover, we conflate the “Weapons and Blasting” and the “Mining” classes into the “Metallurgy” category because few patents were observed in those industries. We further augment patent data by defining a measure of “quality” or “innovativeness” following Kelly *et al.* (2021). This metric flags as influential those patents that introduce terms not used before they were granted and become common after that. We evaluate this metric on the abstracts of patents granted after 1900. We apply this sample restriction for consistency: in 1853–1899, we observe the full text of patents instead of their abstract.

A.I.2 Migration Data

Disaggregated data on the origin of English and Welsh immigrants—and, more generally, of all other nationalities—do not exist. Neither US authorities nor the sending ones in the UK collected them. We thus lack precise information on where British immigrants in the US came from *within* the UK. We fill this gap and link the individual-level UK and US censuses, as described in A.III. Ideally, we observe the universe of British emigrants to the United States between 1870 and 1930. For those individuals, we know all the information contained in the US Census and those detailed in the UK one. Most notably, we know where they came from. As we discuss more in detail later, we also link return migrants. Since the last publicly available UK census dates to 1911, we can only construct return migration flows over the period 1870–1910.

A.I.3 Data Constructed from the Population Censuses

The main data sources we leverage are the individual-level, non-anonymized UK and US population censuses. The US census features prominently in the economic history literature as a major source of detailed microdata, and we thus avoid discussing it any further (Ruggles *et al.*, 2021). The UK census is relatively less well-known (Schurer and Higgs, 2020). Although not as rich as its US counterpart, the UK population census covers individuals who have resided in the UK since 1850. The first census was run in 1841, but only 1851, 1861, 1881, 1891, 1901, and 1911 are entirely digitized.[1] Data in the census include the name and surname, birth year, division, county, district, parish, precise address of residence, the specific occupation detailed through HISCO codes, and other variables that we do not use in the paper, such as the type of dwelling and fertility information. We augment these variables by geo-coding the universe of addresses that appear in the census to precise geographical coordinates.

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

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

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

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

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

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

A.I.5 Miscellaneous Data

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

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

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

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

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

A.I.6 GIS Shapefiles & Boundary Harmonization

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

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

A.I.7 Geo-referencing the Historical British Census

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

There are two ways to geo-reference historical addresses. One approach is manually digitizing historical locations, either streets or enumeration units, from historical maps. However, this method does not scale up and becomes rapidly unfeasible as the data grows. A second automated approach is to run text-based address matching between historical data sources and address databases that have already

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

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

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

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

A.II Patent Data

A.II.1 Motivation

Despite its historical significance, we lack comprehensive patent data for the Second Industrial Revolution period (1850–1900) in the United Kingdom. In particular, it is impossible to reconstruct the geographical distribution of innovation activity during this period. This data limitation sharply contrasts the effort undertaken to document patenting activity since the inception of the English patent law in 1617 up until the end of the First Industrial Revolution in the 1840s (Nuvolari and Tartari, 2011; Nuvolari, Tartari and Tranchero, 2021). We fill this gap by constructing the first dataset of English and Welsh patents that spans the period 1853- 1900 and contains detailed information on the text, geographical location, inventors' personal information, and date for the universe of patents.

A.II.2 Data Sources

The UK Intellectual Property Office allowed us access to restricted full-page scans of original patent documents. These are the universe of patents granted in England and Wales between 1617 and 1899. This paper focuses on the period 1853-1899 for two main reasons. First, Nuvolari and Tartari (2011) already digitized patents before 1853 from Bennet Woodcroft's index, although patent documents contain additional information compared to the index. Second, in 1853, a reform dramatically lowered patent application prices. This makes it challenging to compare patents before and after the reform. Patent documents contain a wealth of unstructured information. We provide two examples in Figure A.3: in panel A.3a, we show the patent granted to Henry Bessemer for the eponymous process to produce steel, and in panel A.3b we display the patent granted to John Starley for the first modern safety bicycle. Both patents are in our dataset. The rectangles identify the location of the textual data that we extract. These comprise (i) a short title, (ii) a long title, (iii) the author(s)'s name(s), (iv) the author(s)'s address(es), (v) the author(s)'s profession(s), (vi) the filing date, (vii) the issue date, (viii) the type of protection, (ix) an indicator of whether the application was filed by an agent on behalf of someone living abroad, and (x) the full text of the patent. Not all (i-x) are available throughout the sample. In particular, (i), (vi), and (viii) are available only until 1873. After that date, a short title is no longer reported, the filing date is reported only sporadically, and the type of protection becomes immaterial, for only granted patents are included in the sample.

A.II.3 Digitization

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

A.II.4 Geo-Coding

To retrieve each patent's location, we geocode each inventor's listed address using the commercial geocoding engine provided by MapTiler AG. To geocode an address, if a coarse geographical unit is listed on the patent (e.g., the county), we condition the outcome coordinates to lie within that unit. In Figure IIb, we report the resulting distribution of patents per capita, whereas Figure A.5 reports the spatial distribution of patent grants across technologies. Reassuringly, these are consistent with underlying population and economic development indicators as well as with historical evidence (e.g., note the substantial clustering of textile patents in the Lancashire districts).

A.II.5 External Validation

To validate our data, we consider the only two series covering a portion of the 1853-1899 years. Hanlon (2016) digitized an index of patents issued between 1855 and 1883. His data list, for each patent, the inventor(s) and their profession(s), a technology class, and the issue year. On top of the longer time coverage, our data thus contain several additional information, including the geographical coordinates. The second dataset we use to compare is the "A Cradle of Invention" (COI) series, published by Finishing Publications (2018). These data, too, were digitized from indices and thus only list authors, issue year, and, often, titles. In principle, this series spans the years 1617–1895. However, after 1883, patent applications that were eventually denied protection were also listed. Absent a way to identify granted patents, we do not report figures after 1883 for the COI series.

In Table A.1, we report the aggregate number of patents issued according to our series (columns 2 and 6), COI (columns 3 and 7), and Hanlon (2016) (columns 4 and 8). Reassuringly, the three series are highly consistent. Our series is closest to Hanlon (2016), but the COI figures are not too far off either. Overall, the Table strongly suggests that our series is as complete as the Hanlon (2016) database. We cannot, however, externally validate it for the later part of the period because there is no data available.

A.II.6 Measuring the Similarity Between US and UK Patents and their Quality

This section describes how we construct the patent similarity metric to measure "copying" and "originality" of UK innovation activity. The approach borrows heavily on Kelly *et al.* (2021). We adapt their methodology to our context by leveraging text information in titles only. Even though we do not have access to full US patent texts, the title of a patent is usually very informative about its content. We previously showed that a title-based machine learning algorithm predicts the technological classification of the patent with nearly 90% accuracy. Titles for UK patents are embedded in the digitized text for 1870–1899 and collected from PATSTAT for the later years; titles for US patents are collected from PATSTAT throughout the sample period.

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

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

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

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

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

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

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

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

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

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

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

where $\mathcal{F}_i^{+\tau}$ denotes the set of US patents issued within τ years after the issue year of patent i. In the baseline analysis, we set a symmetric window of $\tau = 5$ years around each patent's issue date. In Tables C.1, C.2, and C.10, we report the results using an alternative threshold of ten years. Moreover, in the same Table, we report the results obtained by netting out year and technology class fixed effects at the patent level. As noted by Kelly *et al.* (2021), this ensures we do not conflate shifting terminology fashions in the similarity measures.

A.II.7 Census Linking

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

Matching Algorithm Given a patent p, define the set of inventors as $\mathcal{A}_p = \{A_1, \ldots, A_{n_p}\}$. Most patents are solo-authored in this period, meaning $|\mathcal{A}_p| = 1$. Call $\mathcal{L}_p = \{\ell_1, \ldots, \ell_{m_p}\}$ the set of locations patent p is associated to. Each ℓ is a couple of latitude-longitude coordinates. Let $\mathcal{L}_p^{\text{parish}}$ be the set of parishes associated with each coordinate. Analogously, let $\mathcal{L}_p^{\text{neighbor parishes}}$, $\mathcal{L}_p^{\text{district}}$, $\mathcal{L}_p^{\text{neighbor districts}}$, and $\mathcal{L}_p^{\text{county}}$ be the set of, respectively, neighboring parishes, districts, neighboring districts, and counties where each coordinate locates. Notice that these are progressively coarser units: parishes are contained in districts, which form counties. Unfortunately, we do not know the inventor-location pair. To match the generic A_p, we thus perform the following operations:

1. With a slight abuse of notation, let $\mathcal{L}_p^{\text{parish}}$—and, analogously, $\mathcal{L}_p^{\text{district}}$ and $\mathcal{L}_p^{\text{county}}$—denote the set of census records in each parish, district, and county within the respective sets.
2. Take all entries i within the set of parishes $\mathcal{L}_p^{\text{parish}}$ that are at least 18 when the patent p is filed. Let year_i and t_p respectively denote the birth year of i and the issue date:

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

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

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

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

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

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

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

5. Set a threshold τ such that if $\text{Similarity}^{A_p} < \tau$, $\overline{\mathcal{M}}_{A_p}^{\text{parish}} = \emptyset$, otherwise pass.

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

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

Evaluation of the Matching In Figure A.6, we report the matching rate of this exercise. We match approximately 75% of the inventors in the sample. The share remains constant throughout the period. Almost 50% of the matches are attained at the parish level—the smallest geographical layer we consider. About half of the remaining matches are obtained within the neighboring parishes of the location indicated in the patent document. These figures indicate that mobility remained relatively low within the relatively narrow time frame around the census we consider.

A plausible concern is that the probability of obtaining a link is not random. This may be the case if, for instance, more successful inventors were more educated and, hence, more likely to report their names correctly in the census. On the other hand, if successful inventors were relatively more mobile, we may fail at linking them because we may need to go national to obtain a match, which would most likely be dropped because of the multiple-match issue. These hypotheses are difficult to test. In Figure A.7, however, we report the overall distribution of the number of matches in the sample. Approximately 60% of the inventors listed in the patent grants are linked to one single census record, and the share of inventors linked to more than 10 records is negligible. In fact, we can restrict the analysis to single matches and obtain very similar results. In Table A.2, we compute the correlation between the number of matches in our sample and a set of individual observed characteristics. In Panel A, we have the age and the residence divisions while in Panel B we list the occupation reported in the census. The number of matches is correlated with some of these characteristics, but the magnitude of this association is small, except for the fact that inventors living in Wales have substantially more matches in the sample. To ensure that Wales does not drive our results, we replicate the main results, exclude them from the sample, and confirm that the results hold.

A.III Linked Migrants Sample

A.III.1 Data Sources

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

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

A.III.2 Linking Algorithm

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

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

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

UK census, which records where they come from at a granular geographical level.[7] More precisely, we take the stock of British residing in the US in a given census year—say, 1900—and match them with their entry in the preceding UK census—in this case, 1891.[8] This implies that we measure the *flow* of British immigrants over time rather than their stock.

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

1. Take the census that precedes the immigration year of A. Hence, for instance, we match all those who immigrated in 1896 to the 1891 census.
2. Select all records in that census with the same reported birth year as A—call the resulting sample $\mathcal{M}^A = \{m_1^A, \ldots, m_N^A\}$.
3. Compute a string-similarity measure between the name and surname of A and that of all elements of $\mathcal{M}^A$. In other words, for every $m_i^A \in \mathcal{M}^A$, compute[9]

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

   for some $\alpha \in [0,1]$. In our baseline setting, we set $\alpha = 0.3$ to give higher weight to the surname.
4. The set of matches is defined as

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

   which means that we restrict the set of possible matches to include only those whose similarity score with the entry in the US census A is the largest.
5. Finally, for a given threshold $\tau > 0$, we select only the possible matches whose similarity score

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

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

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

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

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

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

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

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

where

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

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

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

A.III.3 Internal and External Validation

Matching Rate In Figure A.9, we report the key matching rate statistics of the linked sample. In particular, Figure A.9a displays the crude matching rate over time, i.e., the share of English immigrants recorded in the US census linked to at least one record in the UK census. Since it is impossible to

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

link British immigrants born after the last census before they migrated (i.e., those who migrate when younger than ten), we distinguish between the full sample in blue and the "matchable" sample in red. The matching rate in the matchable sample is approximately 65%. It declines from 75% in the 1870s to 60-65% and remains constant after that. The large jump in 1880 in the crude matching rate is because the 1871 census is unavailable; hence, between 1870 and 1881, we cannot match anyone younger than 20—as opposed to 10 in the rest of the sample—when they migrated. The matching rate is relatively high compared to other algorithms, such as Abramitzky *et al.* (2021). In the next section, we thus asked whether matched and un-matched immigrants are balanced on observables. Panel A.9b replicates the previous plot but reports the absolute number of emigrants in the sample and those linked instead of the percentage shares.

In Panel A.9c, we display the distribution of the number of matches for linked immigrants. A large fraction ($\approx 45\%$) of immigrants are paired with a single entry in the UK census. Approximately 60% are matched to one or two entries. Beyond this threshold, multiple matches are more common, and the average number of matches is 9.4. To ensure that multiple matches do not confound our analysis, we always weigh emigrants by the inverse of the number of matches. In unreported analyses, however, we confirm that all our results remain qualitatively unchanged even if we exclude instances with more than two matches from the analysis sample.

In Figure A.10, we report the distribution of the name and surname similarities between the records of British immigrants in the US census and their links in the UK census. The dashed black lines plot the threshold below which we reject the link (.9). In the sample; we find that if an immigrant is linked to at least one record in the UK census, then the quality of the match is high, as more than 90% of the sample lies right to the .9 line. However, it is important to note that for a link to be accepted in the final sample, we require that *both* the name and the surname similarities be above 0.9.

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

Validation of the Data It is challenging to validate our sample with external data because, as we note in the main text, we do not have primary geographically disaggregated data on the origin of British emigrants. Baines (2002) attempts to estimate decade-level data for British counties between 1880 and 1910. To the best of our knowledge, these estimates are the only alternative source we can use to gauge the plausibility of our data. They nonetheless present important issues. First, they are based on interpolations from population data tabulated from the population census. First, the author tries to account for internal migration, which remains an important confounding factor. Second, the estimates refer to aggregate emigration outflows, whereas our data precisely captures emigration to the United States. Third, the estimates are computed at the county level. Counties are coarse geographical units. Hence, this dataset does not warrant any modern econometric exercise.

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

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

A.III.4 Return Migration Data

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

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

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

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

Notes. This Table reports the total number of patents in England and Wales between 1853 and 1899. Columns (2) and (6) report the series constructed from our novel dataset; columns (3) and (7) tabulate data from *A Cradle of Inventions* (Finishing Publications, 2018); columns (4) and (8) report data from Hanlon (2016). The *A Cradle of Inventions* series potentially stretches until 1899. However, after 1883 there is no way to distinguish between patents granted and applications. Hence we do not report figures for these later years (Nicholas, 2014). Data from Hanlon (2016) only cover the years 1855–1883. Referenced on page(s) A8.

TABLE A.2. Balance of Linked Inventor Sample

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

Notes. This Table reports the correlation between the number of matches in the linked inventor sample and a set of individual-level co-variates observed in the census. In columns (1–2), we display the unconditional correlation and the associated standard error. In columns (3–4), we repeat the exercise but control for the issue year of the patent. Standard errors are clustered at the county level and are displayed in parentheses. Referenced on page(s) A11, C51.
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

TABLE A.3. Balance of Linked Emigrants Sample

	Correlation with Matching Status			
	= 1 if Not matched	= 1 if Matched	Diff.	Std. Err.
	(1)	(2)	(3)	(4)
Panel A. Individual Characteristics				
Literacy	0.885	0.965	0.079	(0.069)
Income	3.147	3.141	-0.006	(0.011)
Panel B. Occupation				
Professional	0.053	0.041	-0.012***	(0.002)
Farmer	0.085	0.091	0.006	(0.010)
Manager	0.066	0.059	-0.006	(0.004)
Clerical	0.058	0.045	-0.013***	(0.004)
Sales	0.062	0.047	-0.014***	(0.001)
Craftsman	0.260	0.281	0.020*	(0.010)
Operative	0.310	0.330	0.020	(0.014)
Service	0.050	0.049	-0.001	(0.003)
Laborer	0.057	0.056	-0.000	(0.004)
Panel C. Region of Residence				
North East	0.519	0.539	0.019***	(0.005)
Midwest	0.277	0.271	-0.006	(0.005)
South	0.057	0.046	-0.011***	(0.001)
West	0.146	0.145	-0.002	(0.008)

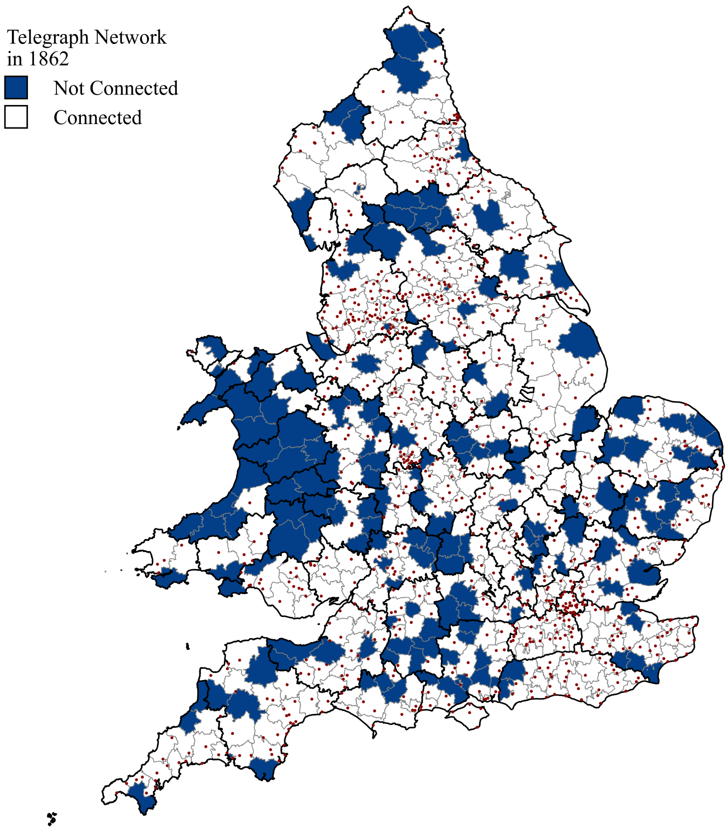
Notes. This Table reports the correlation between the matching probability in the migrants-linked sample and a set of individual-level characteristics observed in the US census. The dependent variable equals one if the immigrant is linked and zero otherwise. We report the average value of the row variable for unmatched (column 1) and matched immigrants (column 2), as well as the difference between the two groups (column 3) along with its standard error clustered at the census year level (column 4). All row variables are indicators except income, which is the log of the occupational income score. Standard errors are clustered at the census year level and are displayed in parentheses. Referenced on page(s) 11, A15, A15.
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

TABLE A.4. Newspaper Summary Statistics

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

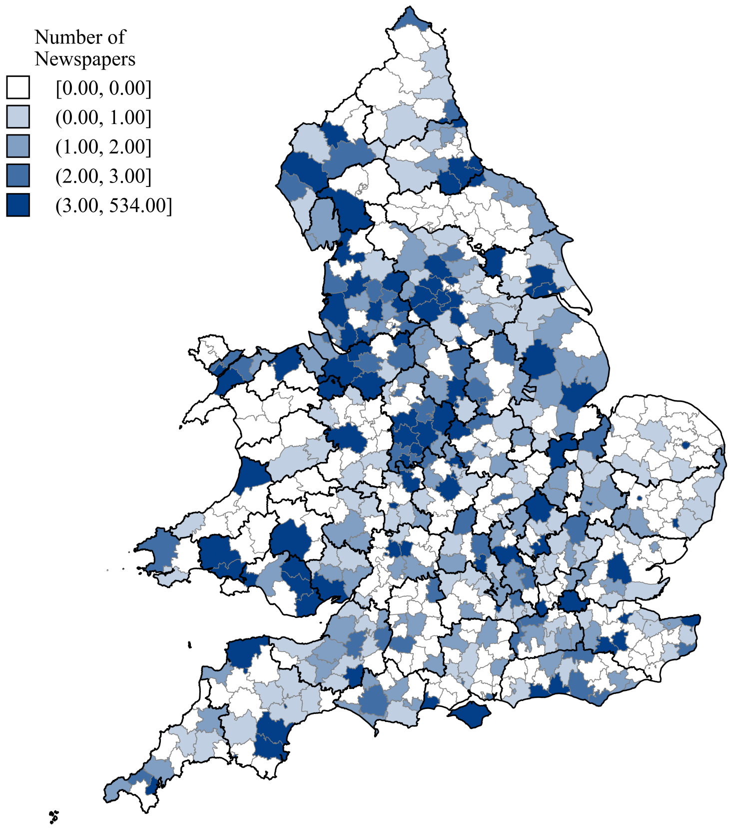
Notes. This Table reports descriptive statistics on newspapers active in the UK between 1880 and 1940. In Panel A, figures are computed at the newspaper level; Panel B computes district-level statistics on the number of newspapers by decade; Panel C computes district-level statistics on the number of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

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



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

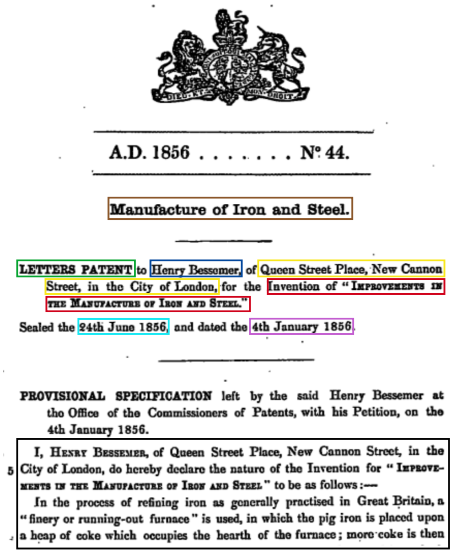
Figure A.2. Geographic Distribution of Active Newspapers



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

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

(a) Henry Bessemer's 1856 Patent



A.D. 1856 . . . . N° 44.

Manufacture of Iron and Steel.

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

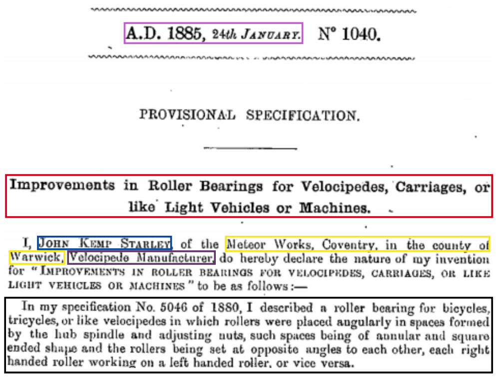
Sealed the 24th June 1856, and dated the 4th January 1856.

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

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

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

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



A.D. 1885, 24th JANUARY. N° 1040.

PROVISIONAL SPECIFICATION.

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

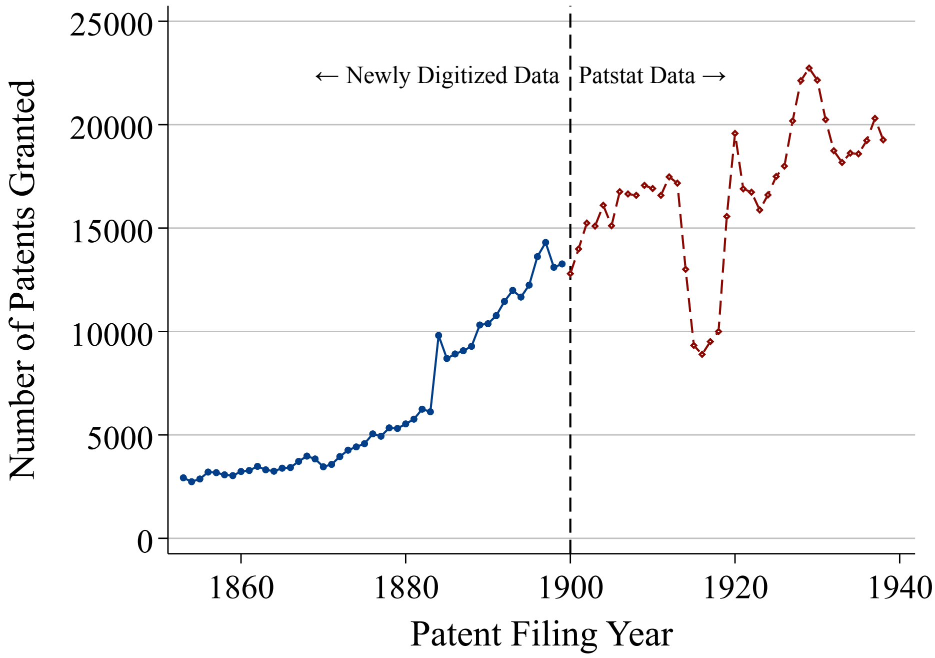
I, JOHN KEMP STARLEY, of the Meteor Works, Coventry, in the county of Warwick, Velocipede Manufacturer, do hereby declare the nature of my invention for "IMPROVEMENTS IN ROLLER BEARINGS FOR VELOCIPEDES, CARRIAGES, OR LIKE LIGHT VEHICLES OR MACHINES" to be as follows:—

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

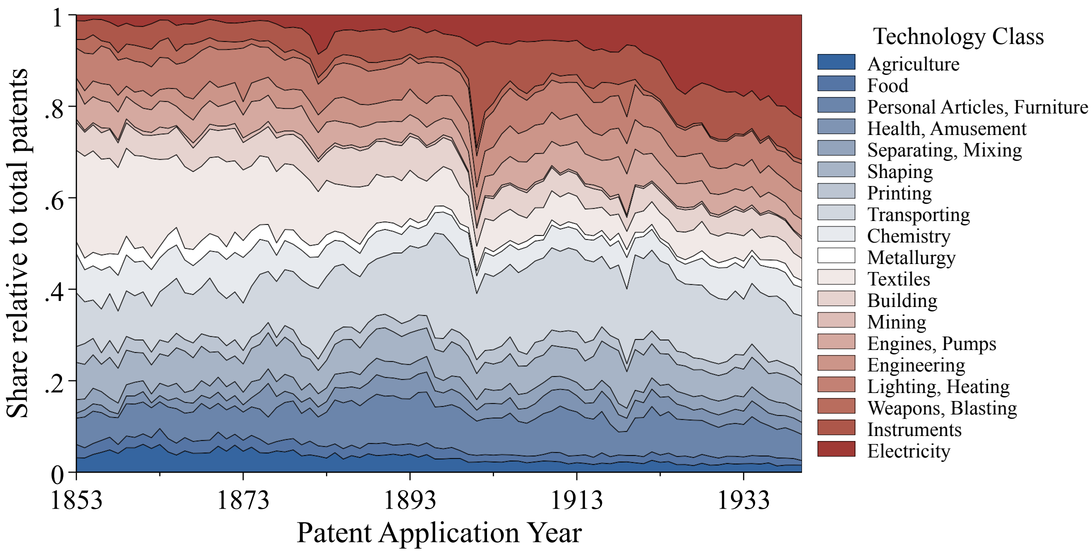
Notes. This figure displays two sample patent documents in our dataset. Panel A.3a was granted to Henry Bessemer in 1856 to invent the famous eponymous process for the mass production of steel from molten pig iron. Panel A.3b was granted to John Starley in 1885 to invent the first modern bicycle, which would soon revolutionize mobility in Europe and the US. Colors mark different variables that we structure in the dataset: (i) in brown, the short title; (ii) in red, the complete title (iii) in green, the type of protection granted; (iv) in blue, the author(s) name(s); (v) in yellow, the author(s)'s address(es); (vi) in light blue, the application date; (vii) in purple, the issue date; (viii) in black, the patent text that continues in the rest of the patent document; (ix) in dark purple, the author(s) profession(s). Not all (i–ix) data are available on every patent and in each year. Referenced on page(s) A6.

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

(A) Time Series of Patents

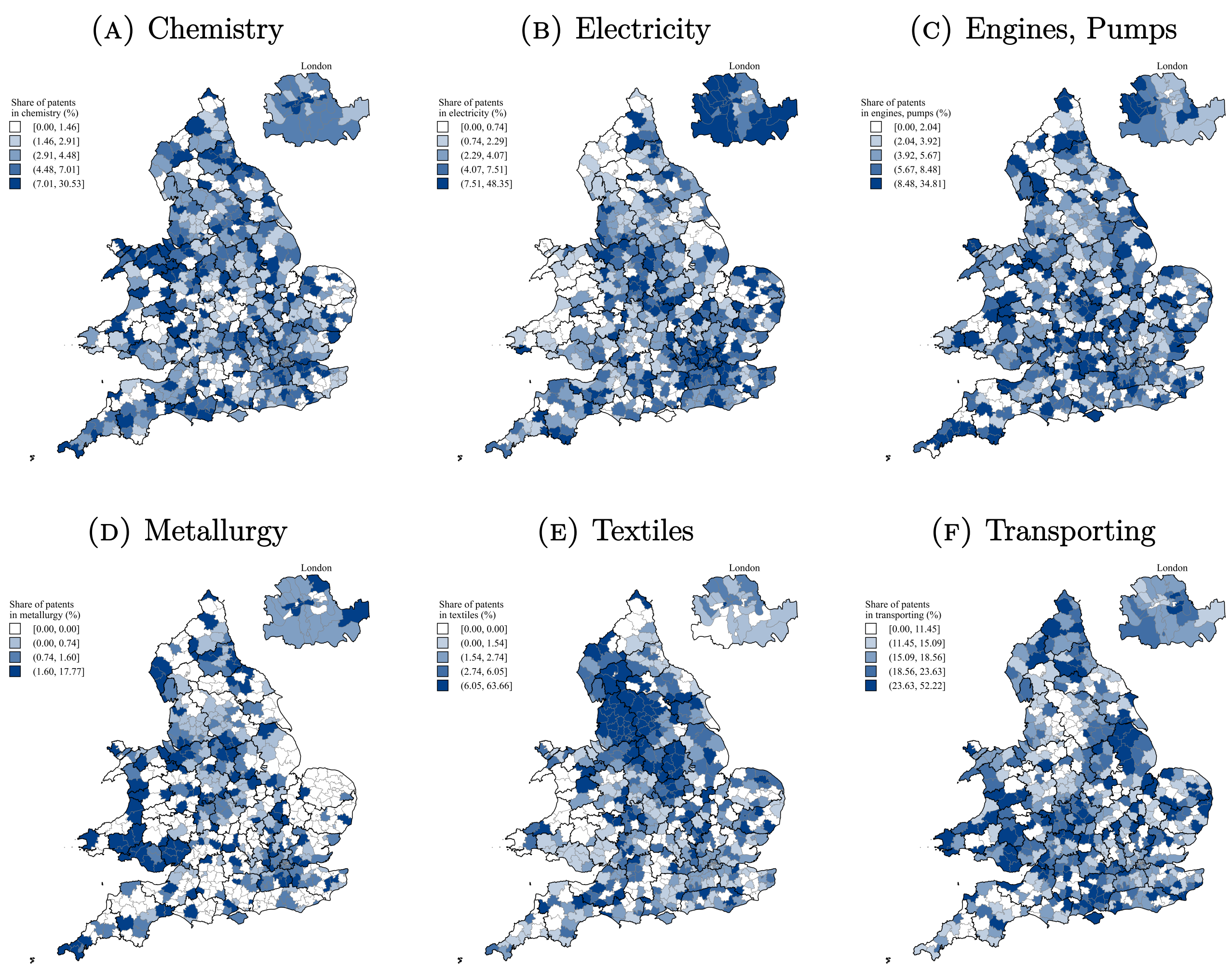


(B) Composition of Patents Across Technologies



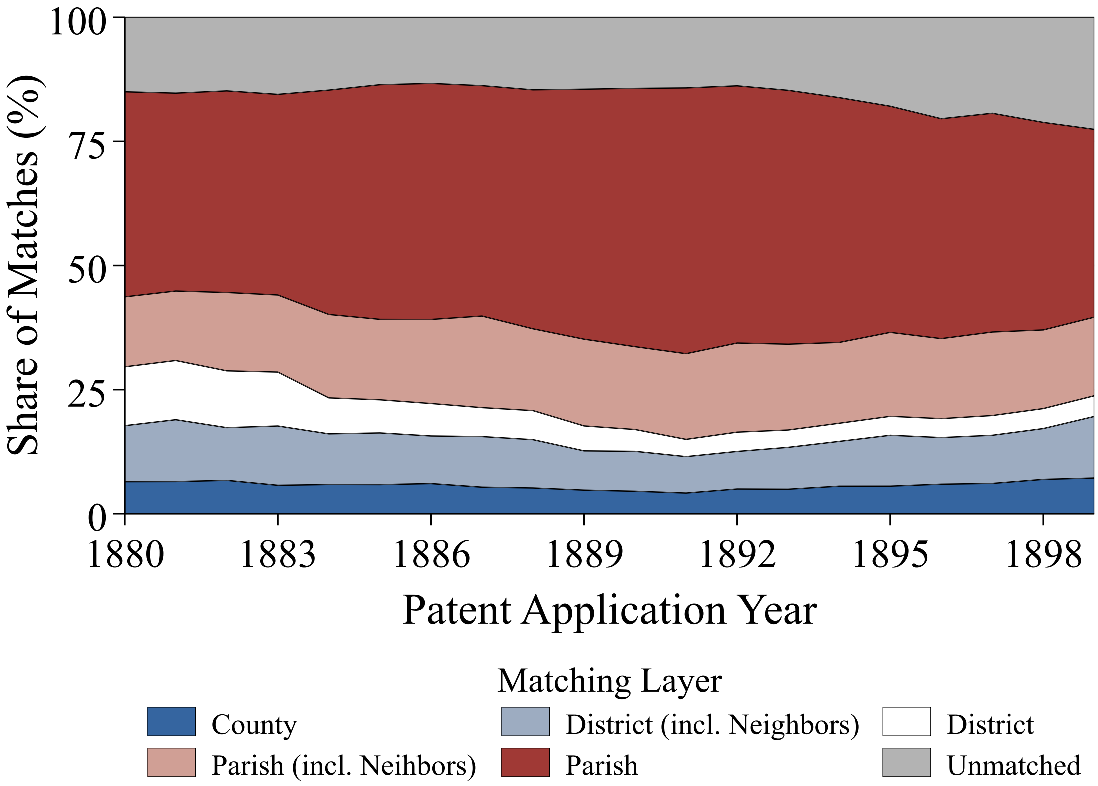
Notes. This Figure reports time-series information on the innovation activity in Britain between 1853 and 1939. In Panel A.4a, we report the total number of patents granted in the UK over the period. The blue dots report the newly digitized data that we assembled for this paper; the red dots report tabulations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

Figure A.5. Geographic Distribution of Patents Across Technologies



Notes. This Figure reports the intensity of patenting activity across districts over 1880–1939 for selected technology classes. Districts are displayed at 1891 borders. To assign patents to districts, we geo-reference the address of each author listed in the patent document and assign districts based on historical district borders. Black edges display county borders. The London area is displayed separately. Darker shades of blue indicate increasing quantiles of the patenting rate, defined as the percentage ratio between the number of patents in a given technology class and the overall number of patents produced. Referenced on page(s) A7.

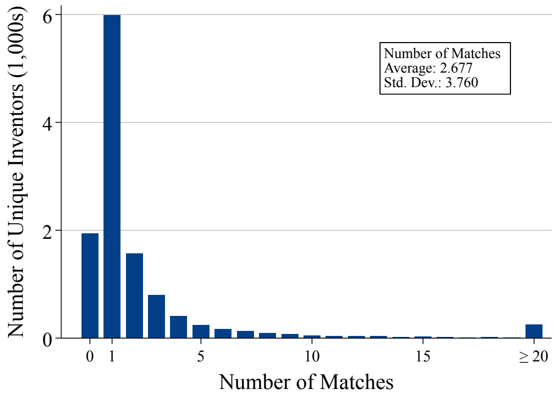
Figure A.6. Matching Rate of Linked Inventors Sample



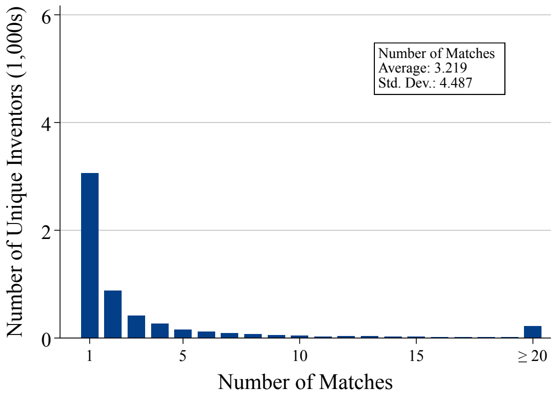
Notes. This Figure reports the matching rate of the linked sample of inventors. The records of the inventors who obtained a patent between 1880 and 1899 are linked to the 1891 population census, as detailed in the main text. The matching rate, i.e., the share of inventors successfully matched to the census, is reported on the y-axis in percentage points. The matching rate is broken down by the geographic layer of aggregation where the match is attained. Hence, we match over 75% inventors to the census throughout the period, and among those, the census record of slightly more than 50% of them is found in the same parish where the inventor is recorded living on the patent document. Referenced on page(s) A11.

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

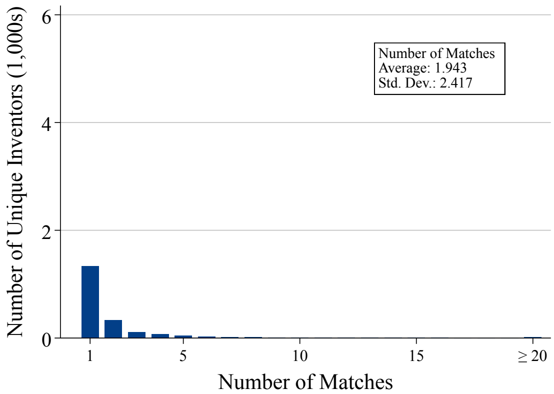
(a) Pooled Data



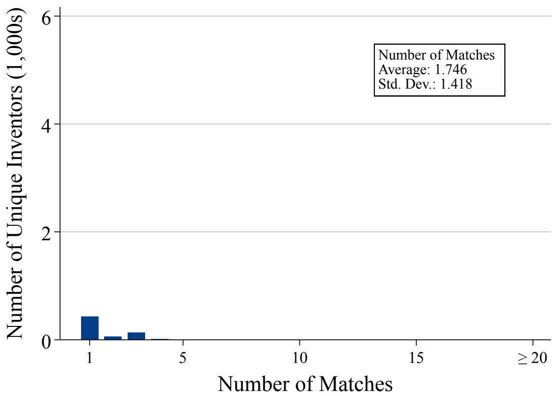
(b) Parish



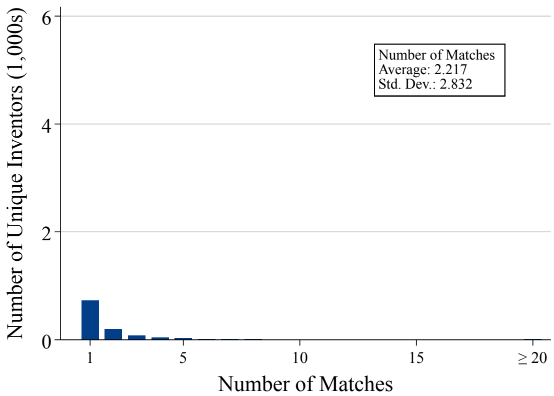
(c) Parish (incl. Neighbors)



(d) District

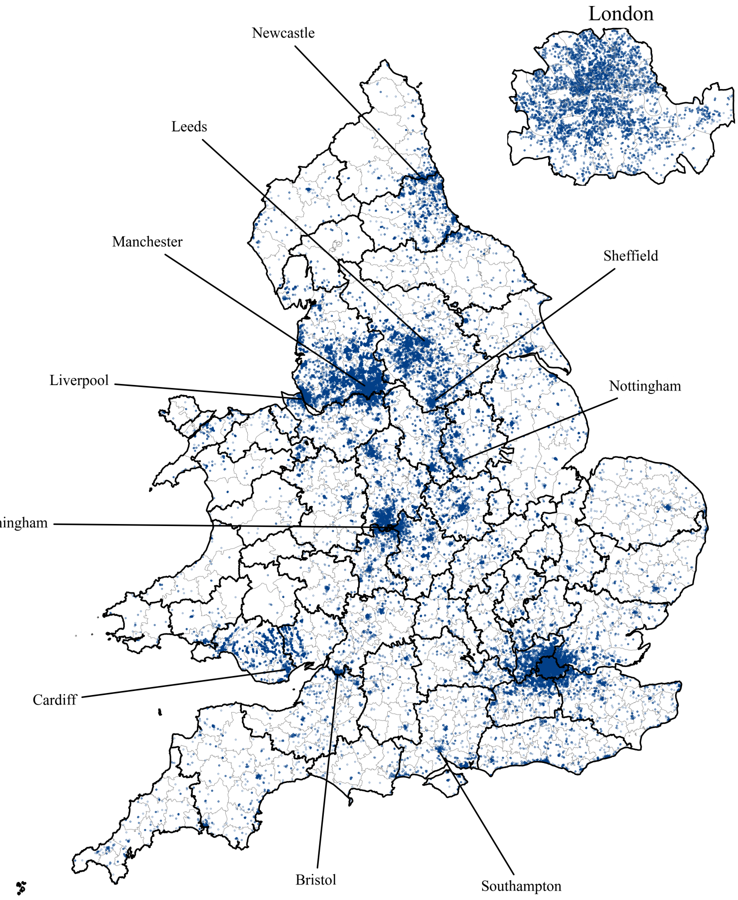


(e) District (incl. Neighbors)



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

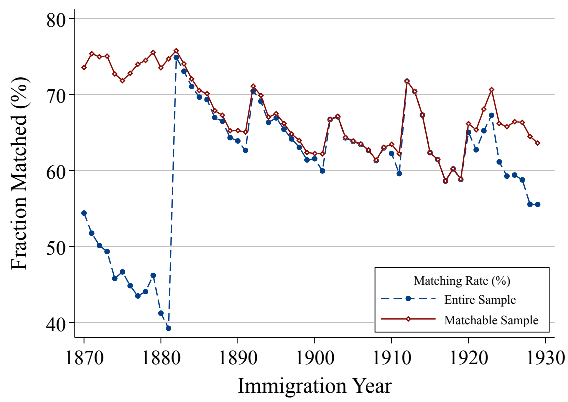
FIGURE A.8. Geographic Distribution of Linked Inventors



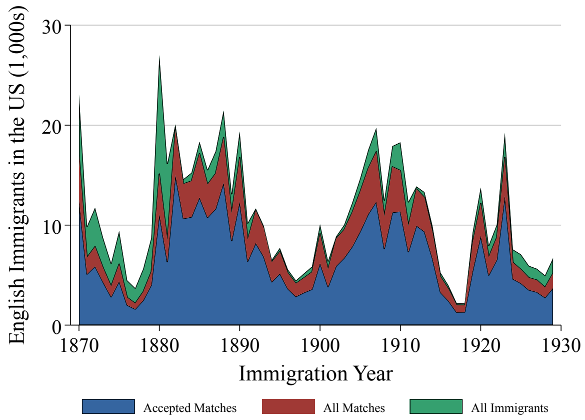
Notes. This Figure displays the spatial distribution of inventors across districts between 1880 and 1900. Each marker reports one inventor, defined as an individual who obtains at least one patent over the sample period. To retrieve the coordinates of the inventors, we first link population censuses, whose entries are, in turn, geo-referenced. The background map displays black counties and gray districts at historical borders in 1891. We highlight the ten largest urban centers at the time. Referenced on page(s) 13.

FIGURE A.9. Matching Rate and Number of Matches

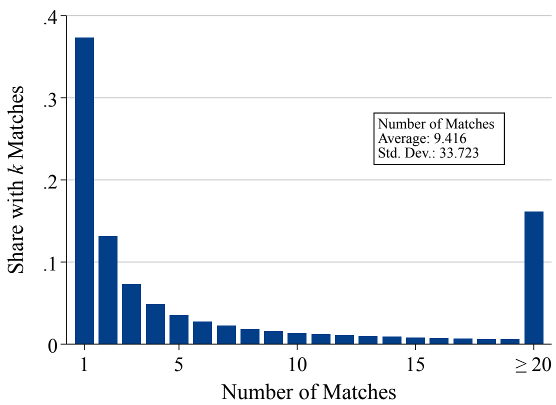
(A) Share of Matched Emigrants



(B) Number of Matched Emigrants



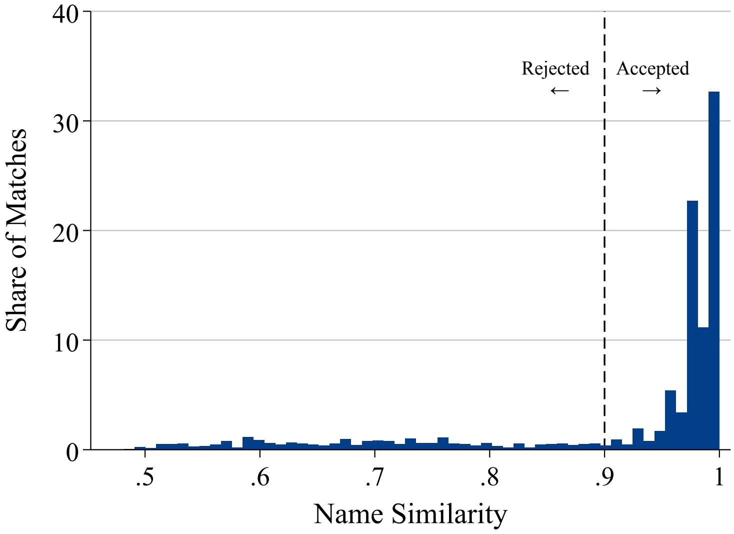
(C) Number of Matches



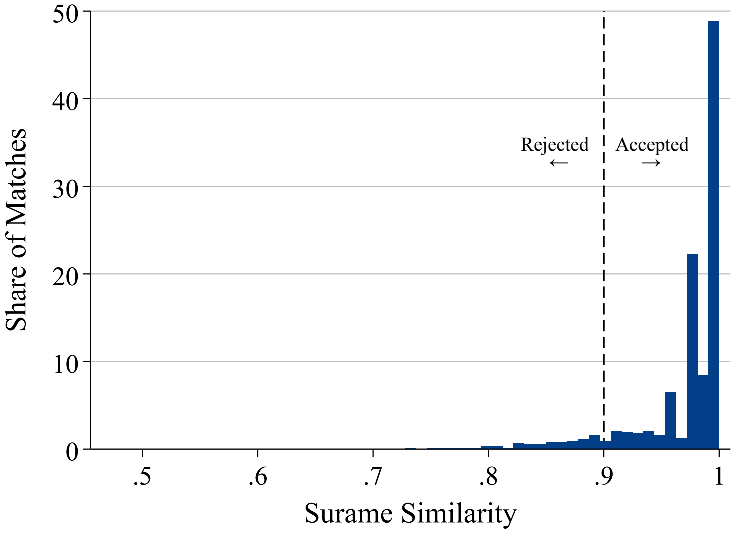
Notes. This Figure reports information on the matching performance of the emigrants' linked sample. In Panel A.9a, we display the matching rate, i.e., the share of emigrants recorded in the US census that are successfully linked to at least one entry in the UK census. The blue dots refer to the overall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

FIGURE A.10. Quality of Matches

(A) Name Similarity

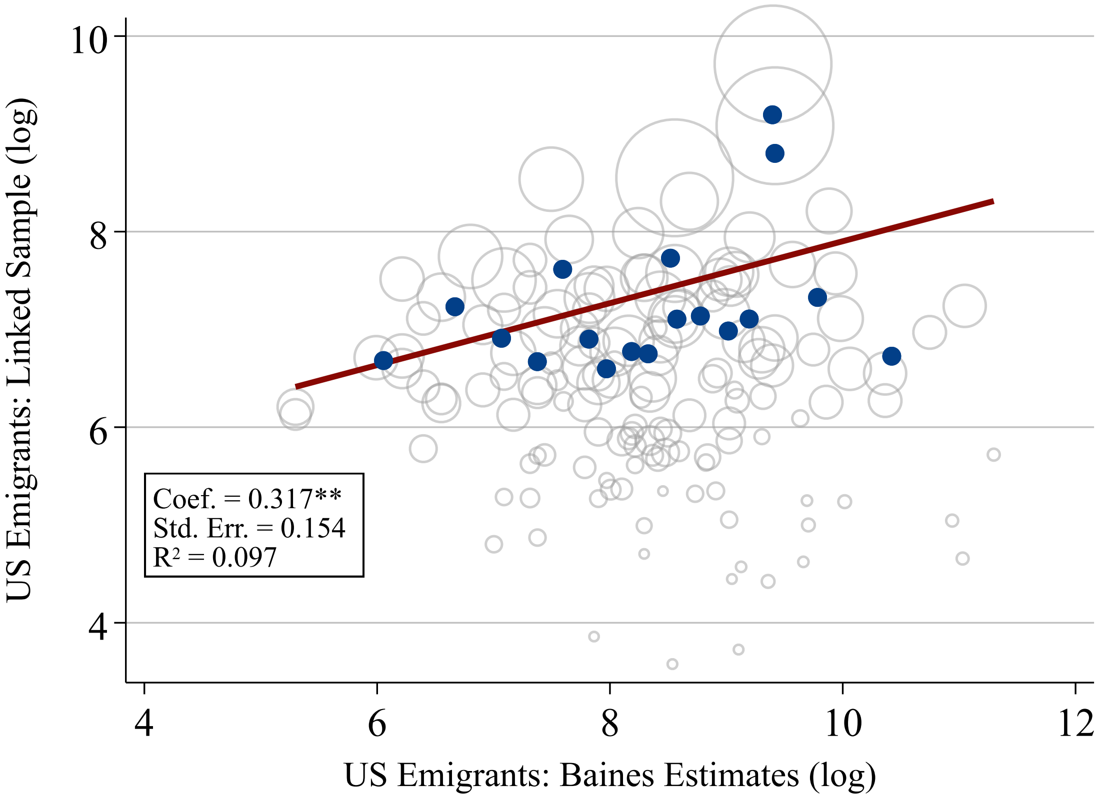


(B) Surname Similarity



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

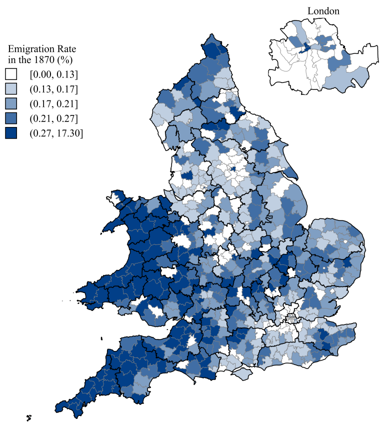
FIGURE A.11. External Validation with Out-Migration Estimates



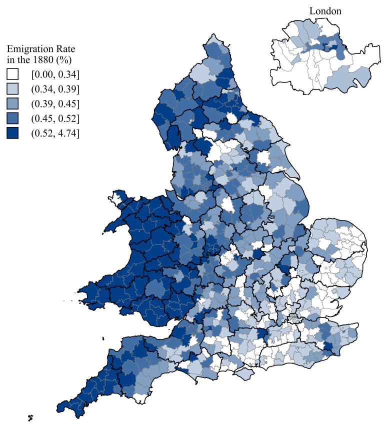
Notes. This Figure reports the correlation between out-migration in our linked inter-census sample and the estimates produced by Baines (2002). These estimates are at the county level and span 1880–1910. We thus aggregate our data by county. Furthermore, we exclude the London area because we cannot map our data to Baines' geographical divisions. Both variables are expressed in log terms. The gray dots display county-decade observations. Counties are weighted by their population. The blue dots report binned means. The red line overlays a linear fit between the variables. The Figure also reports the regression coefficient, its robust standard error, and the R^2 of the regression. Referenced on page(s) 11, A16.

Figure A.12. Geographic Distribution of Emigration Rates Over Time

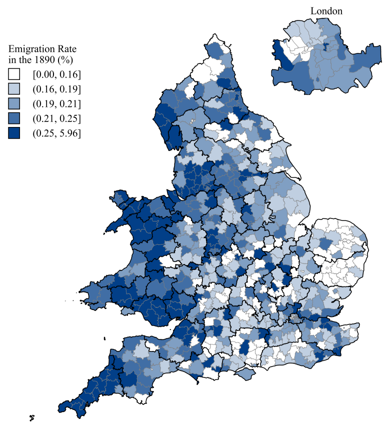
(A) 1870s



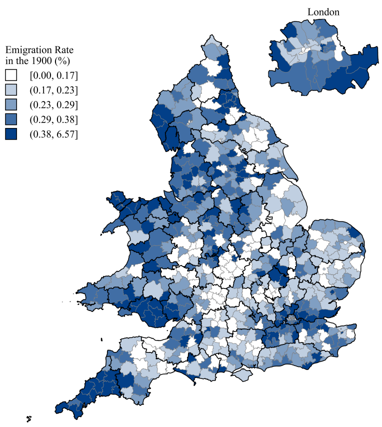
(B) 1880s



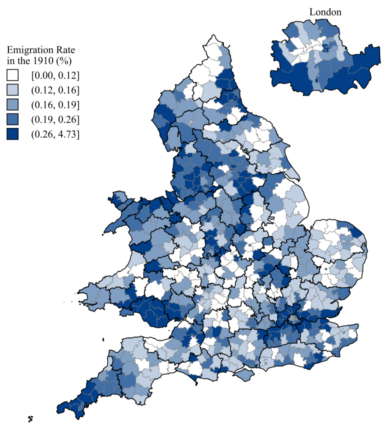
(C) 1890s



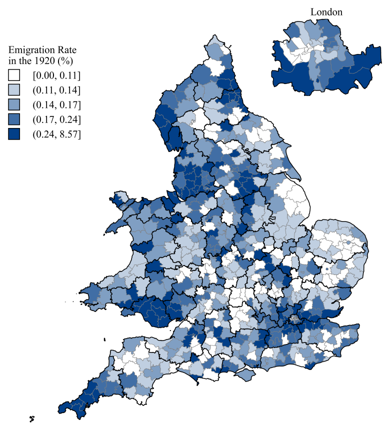
(D) 1900s



(E) 1910s



(F) 1920s



Notes. This Figure reports the distribution of US emigrants across districts in England and Wales over the period 1870–1940 by decade. Data are from the matched emigrants sample. The number of emigrants in each decade is normalized by population in 1891 and is expressed in percentage terms. Districts are displayed at their 1891 historical borders. Black edges also display historical county borders. Out-migration is cross-walked to consistent historical borders. Lighter to darker shades of blue indicate increasing quantiles of the emigration rate. The London area is displayed separately. Referenced on page(s) 11.

B Additional Results

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

B.I Selection of British Emigrants

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

Table B.1 presents. Column (1) refers to non-migrants, and columns (2) and (5) refer to emigrants and return migrants, respectively. In columns (3) and (6), we compute the difference between non-migrants and emigrants and non-migrants and return migrants, respectively. Emigrants are more likely than stayers to work as engineers and as manufacturing workers in metallurgy and textiles. Unsurprisingly, they are less likely to work in public administration and as liberal professionals, since those occupations could not be transferred overseas. Overall, these patterns confirm historical evidence by, among others, Baines (2002), who describes transatlantic emigrants as a positively selected group of entrepreneurial individuals well-versed in the manufacturing crafts, especially in the second half of the Nineteenth century. Return migrants appear somewhat different. They are more likely to work in construction, a lesser-skilled sector compared to emergent industrial jobs, and are more likely to work as professionals, public officers, in transports—which, in this period, would mainly comprise railway workers—and as utility workers. According to this sketched first inquiry, it is likely that the decision to return to the United Kingdom was relatively more common among those who were less successful in their American enterprise.

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

B.II Long-Run Association between Emigration and Innovation

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

We estimate the following regression:

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

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

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

B.III Assortative Matching of Emigrants in the United States

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

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

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

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

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

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

We test this prediction in Table B.3. We find limited evidence that the innovation similarity between origin and destination areas explains migration patterns. The regression coefficients are positive but they are seldom significant both contemporaneously (column 1) as well as lagging them by one (column 2) or two (column 3) decades. In columns (4–6), we estimate the same regressions on the smaller sample of county-district pairs with positive migration flows. The results remain largely similar. We do not wish to over-emphasize these results. Our measure of innovation similarity, while intuitive, may be subject to measurement error which may reduce the precision of the estimates. Taken together, however, we interpret these results as evidence that it is unlikely that assortative matching played a pivotal role in determining the location choice of immigrants. This insight is consistent with the fact that the OLS and the 2SLS estimates presented in the text are quantitatively similar.

B.IV Anecdotal Evidence of Return Innovation

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

In this section, we provide three instructive examples. All three are cases of return migration. Historical records typically focus on successful migrants who, upon returning, bring their technology to their origin areas and promote economic development there. The statistical analysis that we present later, however, suggests that this was only part of the story. In fact, we find that emigrants interacted with their origin communities even without returning.

B.IV.1 British Puddlers and the Kelly-Bessemer Process

An 1856 article published in Scientific American described a new patent granted in the UK to Henry Bessemer (Wagner, 2008). Bessemer had discovered a new process, the would-be eponymous Bessemer process, that, for the first time, allowed the production of inexpensive steel from molten pig iron.[11] American inventor William Kelly complained:

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

The veracity of Kelly's allegations remains unverified. They nonetheless indicate three important elements. First, American inventors knew that British immigrants posed a threat to the secrecy of their inventions. Second, technology transfer did not necessitate the very upper tail of the human capital distribution. Skilled workers, such as puddlers, could be the agents of technology diffusion. Finally, the precise mechanism that emerges is return migration. Kelly expects British puddlers to speak of "his" invention upon returning to England.

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

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

Henry Rowland Marsden was born in Leeds to poor parents in 1823 (Curtis, 1875). At age twenty-five, he emigrated to the United States, first to New York and then to Connecticut. There, he took on apprenticeships in engineering and metal-working firms. He obtained several engineering patents—chiefly related to steam engines and pumps, including a "stone-crusher" which is still in use today. In 1862, Marsden returned to Leeds, where he set up a flourishing business centered around his newly patented inventions. A wealthy man respected for his philanthropic endeavors, he was elected mayor of Leeds in 1873. He died in 1878 and is credited as one of the most prominent figures in the industrial development of Leeds.

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

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

Table B.1. Comparison between English Emigrants and Stayers

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

Notes. This Table displays the selection of emigrants and returning migrants to and from the United States relative to the rest of the British population. The unit of observation is an individual. In each row, the dependent variable is equal to 100 if the individual belongs to the given category (e.g., if they are employed in agriculture) and zero otherwise. In columns (1), (2), and (5) we report the averages for non-migrants, migrants, and return migrants. In columns (3) and (6), we display the difference between non-migrants and, respectively, migrants and return migrants. The associated robust standard errors are displayed in parentheses in columns (4) and (7). Referenced on page(s) B34.
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

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

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

Notes. This Table reports the long-run association between exposure to US technology and innovation in the UK. The unit of observation is a district-technology pair observed at a decade frequency between 1940 and 2010. The dependent variable is the (log) number of patents. The independent variable is an interaction term between exposure to US technology over the 1930s, decade dummies, and technology dummies. The 2010 decade dummy serves as the baseline category. The regression includes district-by-time, district-by-technology, and technology-by-time fixed effects. Standard errors are shown in parentheses and are clustered two-way by district and technology. Referenced on page(s) B35.

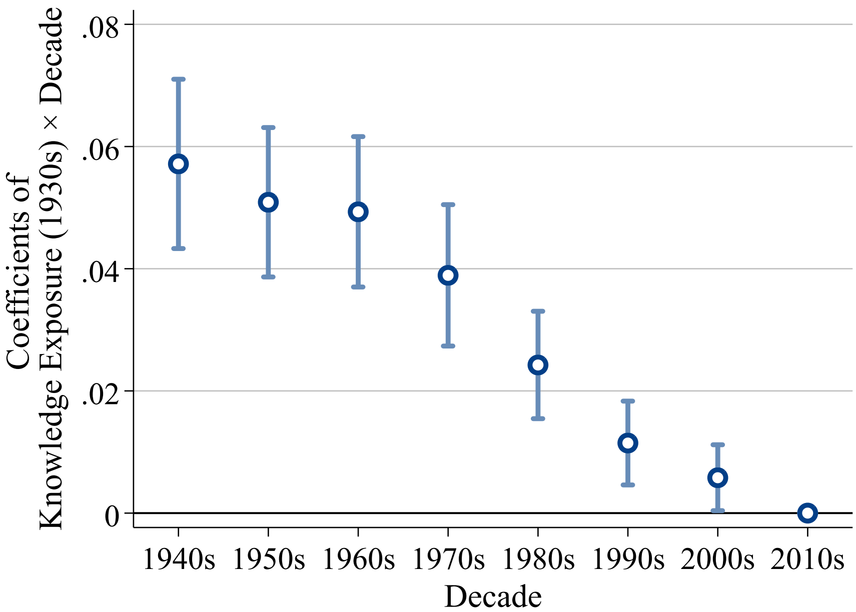
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

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

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

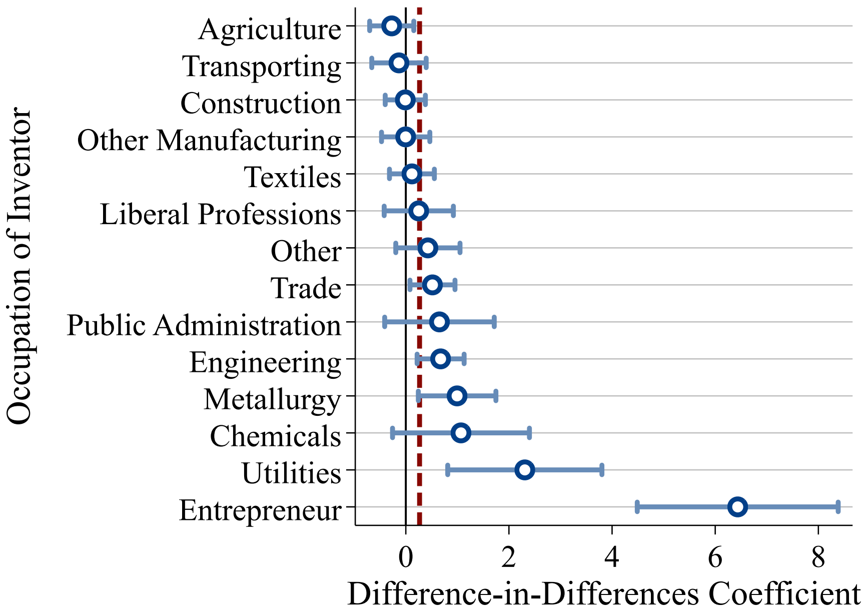
Notes. This Table tests the hypothesis that British immigrants settled in US counties that innovated in the same fields of their district of origin. The unit of observation is a district-county pair. Units are observed at a decade frequency between 1870 and 1930. The dependent variable is the number of migrants between the (UK) district and the (US) county. The independent variable is the similarity of the patent portfolios between the district and the county. The details of the similarity metric are explained in the main text. In columns (1–3), we include all district-county dyads; in columns (4–6), we include only the pairs with positive migration ties. All regressions include county-by-district, county-by-time, and district-by-time fixed effects. Standard errors are clustered two-way by district and county and are displayed in parentheses. Referenced on page(s) 17, B36. *: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

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



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

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



Notes. This Figure reports how emigrants to the United States impact the innovation activity fulfilled by their neighbors who remain in the UK. The unit of observation is an individual inventor observed at a yearly frequency between 1880 and 1900. The analysis sample is the universe of inventors linked to the 1891 population census, as detailed in the main text. The dependent variable is the (log) number of patents produced by the members of the family. Each coefficient refers to an interaction term between the baseline treatment—an indicator equal to one after the first neighbor of the inventor moves to the US, and zero otherwise—and a dummy variable that codes the occupation of the inventor. The dashed red line indicates the average treatment effect reported in the main text. All regressions include inventor and year-fixed effects. Standard errors are clustered at the county level. The bands report 95% confidence intervals. Referenced on page(s) 30.

C Robustness Analyses

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

C.I OLS and 2SLS Analysis

C.I.1 Alternative Dependent Variables

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

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

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

C.I.2 Alternative Definitions of Knowledge Exposure

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

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

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

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

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

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

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

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

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

C.I.3 Alternative Standard Errors

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

C.I.4 Instrumental Variable Strategy

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

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

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

$$\begin{aligned}\text{Immigrant Share}_{j,t} &= \alpha_j + \alpha_t + \beta \text{Immigrant Share}_{j,t-1} + \gamma I^{\text{Rail}}_{j,t-1} + \\ &+ \delta \left(I^{\text{Rail}}_{j,t-1} \times \text{Immigrant Flow}_{t-1} \right) + \zeta \left(\text{Industrialization}_{t-1} \times I^{\text{Rail}}_{j,t-1} \right) + \\ &+ \eta \left(\text{GDP Growth}_{t-1} \times I^{\text{Rail}}_{j,t-1} \right) + X^{\intercal}_{j,t-1} \Theta + \varepsilon_{j,t}\end{aligned} \tag{C.5}$$

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

Leave-out Instrument An alternative approach—which, for brevity, we refer to as "leave-out instrument"—to randomize immigration across counties is constructing an instrument along the lines of Card (2001) and Tabellini (2020). The instrument predicts the county-level immigrant share by interacting the number of immigrants by country of origin that had settled in the county before the sample period—i.e., in the 1860s—with the subsequent aggregate inflow of immigrants over time. We exclude British immigrants from the calculation.[12]

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

C.II Double and Triple Differences Analysis

C.II.1 Alternative Dependent Variables

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

C.II.2 Alternative Definitions of the Shocks

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

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

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

C.II.3 Alternative Standard Errors

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

C.II.4 Alternative Estimator

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

C.II.5 Innovation Shocks in the United States

The double and triple differences analyses implicitly rely on the fact that our methodology to flag innovation shocks in US counties can reliably isolate periods of intense patenting activity. We test this assumption in Table C.12. We consider the universe of US counties (in columns 1–2) and county-technology pairs (columns 3–4). The dependent variable is the (log) number of patents (in columns 1 and 3) and the (log) number of original patents (in columns 2 and 4), and the treatment returns a value of one for treated units, i.e. counties or county-technology pairs after the innovation shocks occur, and zero otherwise. The regressions are saturated with fixed effects. These double and triple differences regressions do not capture a causal effect but rather indicate the actual intensity of the shock to US innovation activity.

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

C.III Neighborhood Analysis

We consider two very simple departures from the baseline scenario described in the paper.

First, we exclude Wales from the estimation sample in Table C.13. We apply this sample cut because, as shown in Table A.2, the number of matches in the sample of inventors linked to the census is larger for inventors residing in Wales. By excluding them from the estimation sample, we thus ensure that this imbalance does not drive the results. The estimates presented in the Table suggest that this does not appear to be the case.

Second, in the paper, we consider an emigrant to be in the neighborhood of an inventor if, before migrating, he lived within five kilometers of the inventor. In Figure C.7, we consider five alternative threshold distances: one, two, three, ten, and twenty kilometers. We then estimate the baseline regression for these various thresholds and separately report the estimated average treatment effects. The coefficients remain positive for all thresholds, but they are largest in magnitude—and statistically significant at the 1% level—for three- and five-kilometer neighborhoods. The estimate for the ten-kilometer definition is similar, albeit noisier. From an economic perspective, it is plausible that as the intensity of the social connections vanishes with distance, including larger neighborhoods in the treatment introduces noise which reduces the precision of the estimates. In turn, a very low threshold discards a large number of emigrants, thus artificially dampening the treatment effect. This notwithstanding, it is reassuring that the coefficients remain positive throughout.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

TABLE C.7. Alternative Leave-Out Instrumental Variables

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

Notes. This Table reports the association between emigration, exposure to US technology, and innovation in the United Kingdom using alternative leave-out strategies to construct the instrumental variable. The unit of observation in Panel A (resp. B) is a district (resp. district-technology class pair) at a yearly frequency between 1870 and 1930. The dependent variable is the (log) number of patents. In Panel A (resp. B), the independent variable is the number of US emigrants (resp. exposure to US technology through emigration ties). The coefficients report the two-stage least squares estimates using the leave-out instrument constructed by including different groups of immigrants in the shift term: in column (1), we exclude UK immigrants (the baseline scenario); in column (2), we exclude immigrants from Sweden, Denmark, Norway, and Finland; column (3) excludes immigrants from Spain, Portugal, Italy, France, Greece; column (4) excludes immigrants from Poland, Romania, Russian Empire, Bulgaria; column (5) excludes immigrants from Germany, Austria-Hungary, Switzerland, Netherlands, and Belgium. In Panel A, all regressions include district and year-fixed effects; in Panel B, all regressions include district-by-time and technology-fixed effects. Standard errors are shown in parentheses and are clustered by district level. Each column and Panel reports the first-stage F-statistic. Referenced on page(s) C46.
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

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

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

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

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

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

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

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

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

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

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

Notes. This Table displays the effect of shocks to US innovation activity on innovation produced in the United Kingdom. In Panel A, the unit of observation is a district observed at a yearly frequency between 1870 and 1930; in Panel B, the unit of observation is a district-technology pair observed over the same period. The dependent variables are: in column (1), the number of patents; in column (2), the log(1+) number of patents; in column (3), the log(0.1+) number of patents; in column (4), the inverse hyperbolic sine of the number of patents; in columns (5–7), the number of patents in the top 20% of distribution of originality in the previous 1, 5, and 10 years; in columns (8–10), the dependent variable is the similarity of British patents with American patents issued in the previous 1, 5, and 10 years. The independent variable is equal to one for all years after the observation unit is exposed to a shock to US innovation activity. The definition of exposure is provided in the main text. All regressions in Panel A include district and year fixed effects; all regressions in Panel B include district-by-year, technology-by-year, and district-by-technology fixed effects. Standard errors are clustered by district and are displayed in parentheses. Referenced on page(s) 24, 24, A9, C49.
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

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

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

Notes. This Table displays the effect of shocks to US innovation activity on innovation produced in the United Kingdom. In Panel A, the unit of observation is a district observed at a yearly frequency between 1870 and 1930; in Panel B, the unit of observation is a district-technology pair observed over the same period. The dependent variable is in columns (1–3), the (log) number of patents; in columns (4–6), the number of patents in the top 20% of the originality distribution; in columns (7–9), the average text-based similarity between British patents and American patents produced in the previous five years. The independent variable is equal to one for all years after the observation unit is exposed to a shock to US innovation activity. A unit is exposed to a shock if the number of emigrants from that unit that are exposed to a US innovation shock is in the k-percentile of the distribution. We consider three alternative values for the k threshold, 10%, 5% (the baseline), and 1%. All regressions in Panel A include district and year fixed effects; all regressions in Panel B include district-by-year, technology-by-year, and district-by-technology fixed effects. Standard errors are clustered by district and are displayed in parentheses. Referenced on page(s) 26, C49.
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

TABLE C.12. Shocks to United States Innovation Activity

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

Notes. This Table reports the "first stage" of the shocks to US innovation, i.e., how much innovation in the US increases in the period following what we define as a shock. In columns (1–2) (resp. 3–4), the unit of observation is a county (resp. county-technology pair) observed at a yearly frequency between 1870 and 1930. In columns (1) and (3), the dependent variable is the (log) number of patents; in columns (2) and (4), it is the (log) number of patents in the top 20% of the novelty distribution. The independent variable is an indicator equal to one in the years after the observation units undergo an innovation shock. Regressions in columns (1–2) include county and year-fixed effects; regressions in columns (3–4) include county-by-year, technology-by-year, and county-by-technology fixed effects. Standard errors are shown in parentheses and are clustered at the county level. Referenced on page(s) 18, 18, 22, C50.
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

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

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

Notes. This Table reports how emigrants to the United States impact the innovation activity fulfilled by their neighbors who remain in the UK. The unit of observation is an individual inventor observed at a yearly frequency between 1880 and 1900. The analysis sample is the universe of inventors linked to the 1891 population census, as detailed in the main text, excluding inventors residing in Wales. The dependent variable is in columns (1–4), the (log) number of patents produced by the members of the family; in column (5), an indicator equal to one for the same variable; in column (6), the average text-based similarity between British patents and American patents produced in the previous five years; in column (7), the number of patents in the top 20% of the novelty distribution. In Panel A, the baseline treatment is an indicator equal to one after the first neighbor of the inventor moves to the US, and zero otherwise; in Panel B, we restrict the treatment to non-return emigrants. In column (4), we further include an interaction with the number of emigrants. All regressions include inventor and year fixed effects; in columns (2) and (3) we include, respectively, district-by-year and parish-by-year fixed effects. Standard errors are clustered at the county level and are displayed in parentheses. Referenced on page(s) C51.

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

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

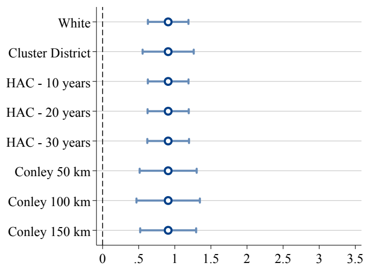
	Mentions of U.S.			Mentions of U.S. States			Mentions of U.S. Counties		
	(1) Railway	(2) Leave-out	(3) Overid.	(4) Railway	(5) Leave-out	(6) Overid.	(7) Railway	(8) Leave-out	(9) Overid.
Dependent Variable Mean	*1.491*	*1.491*	*1.491*	*1.019*	*1.019*	*1.019*	*0.001*	*0.001*	*0.001*
US Emigrants (1,000s)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

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

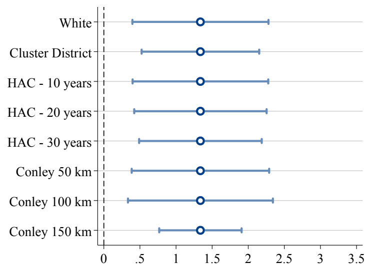
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

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

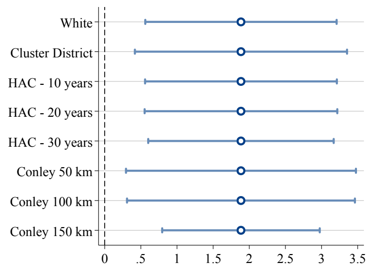
(A) Emigration: OLS



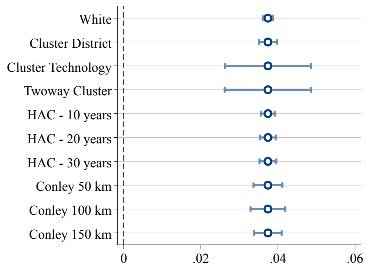
(B) Emigration: Railway IV



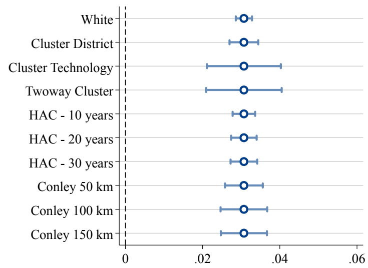
(C) Emigration: Leaveout IV



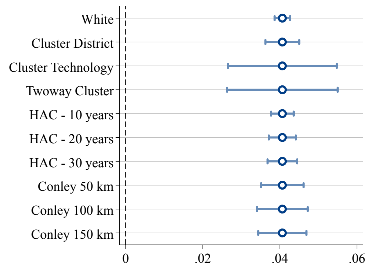
(D) Knowledge Exposure: OLS



(E) Knowledge Exposure: Railway IV



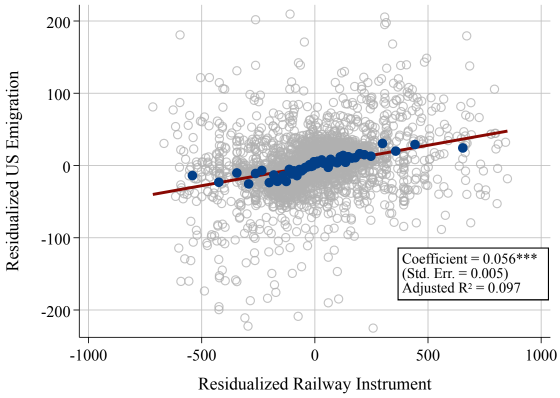
(F) Knowledge Exposure: Leaveout IV



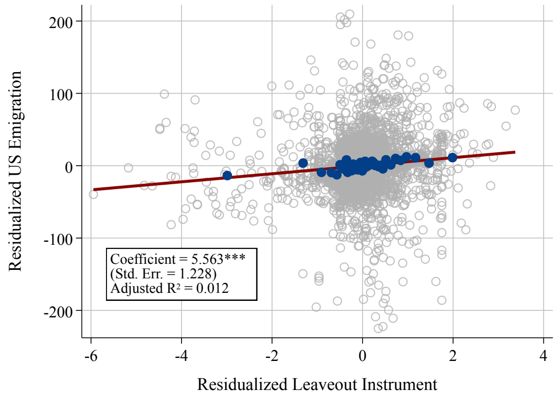
Notes. This Figure displays the association between emigration and innovation in the United Kingdom using various estimators for the standard errors. In Panels C.1a–C.1c (resp. C.1d–C.1f), the unit of observation is a district (resp. district-technology pair) observed at a decade frequency between 1870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

FIGURE C.2. First-Stage Correlations

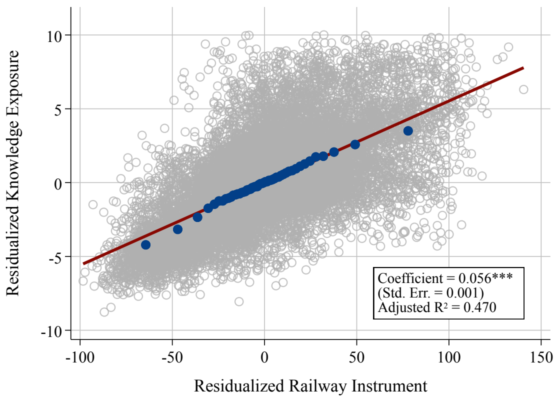
(A) Emigration: Railway IV



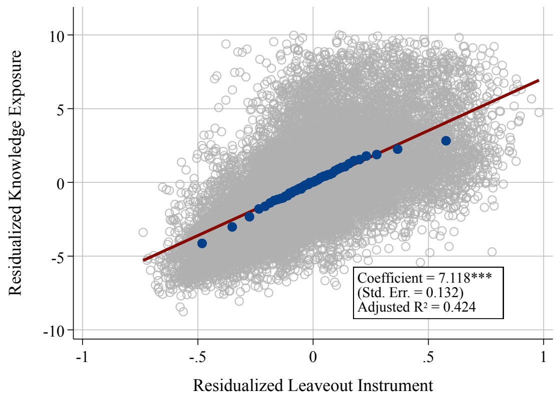
(B) Emigration: Leave-out IV



(C) Knowledge Exposure: Railway IV



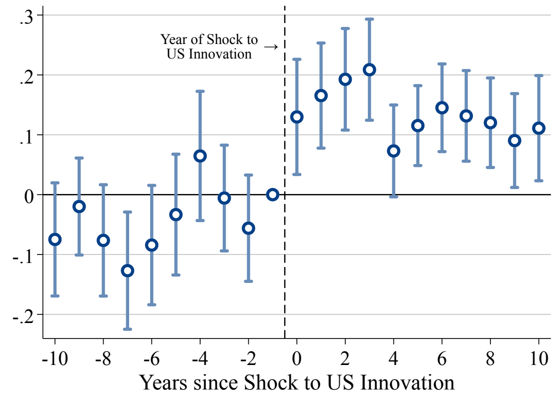
(D) Knowledge Exposure: Leave-out IV



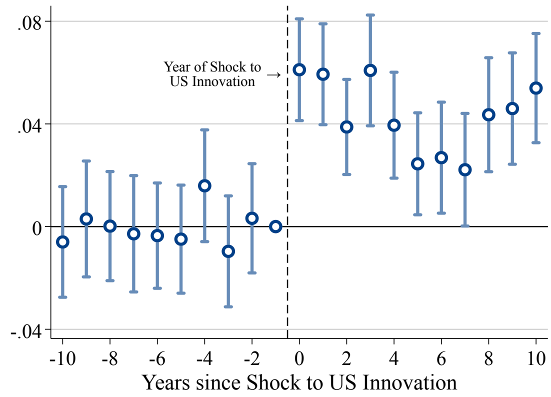
Notes. This Figure reports the first-stage association between out-migration (Panels C.2a–C.2b), exposure to US technology (Panels C.2c–C.2d), and the railway and leave-out instrumental variables described in the main text. In Panels C.2a–C.2b (resp. C.2c–C.2d), the unit of observation is a district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

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

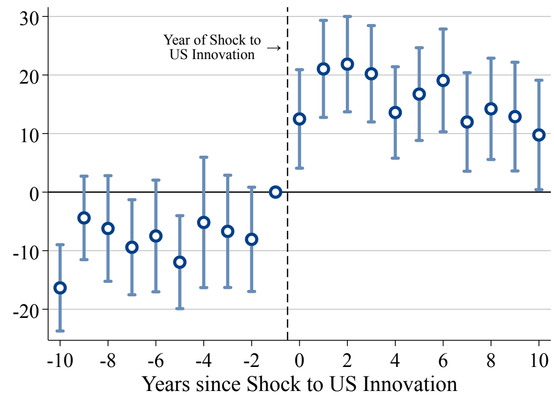
(a) Patents Pooled Across Technologies



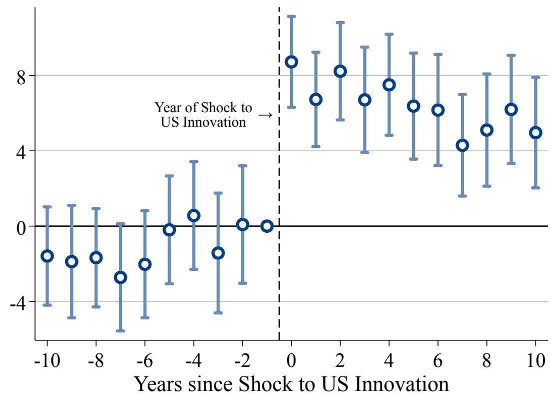
(b) Patents by Technology



(c) Similarity to US Patents: Pooled



(d) Similarity to US Patents: By Technology

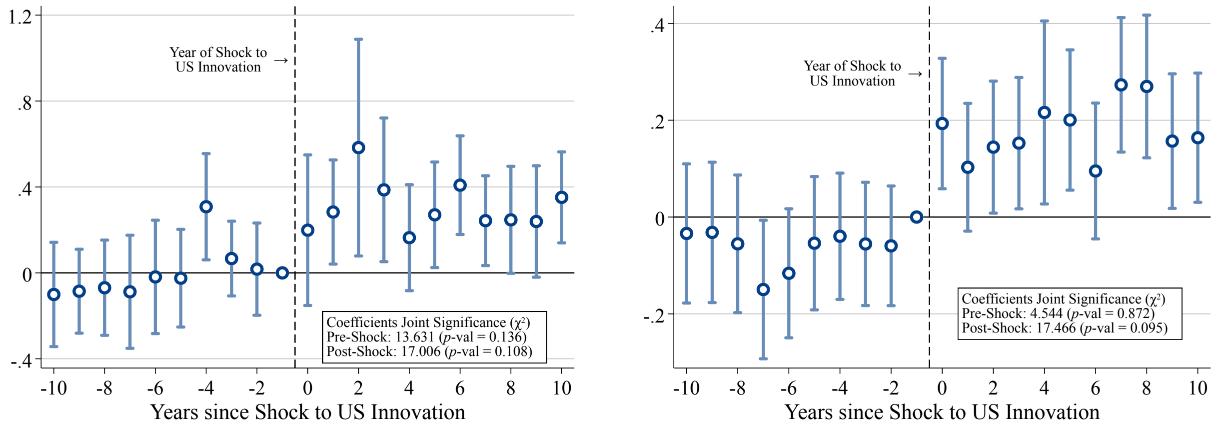


Notes. This Figure displays how shocks to US innovation activity impact innovation in the UK. In Panels C.3a and C.3c, the unit of observation is a district observed at a yearly frequency between 1870 and 1930; in Panels C.3b and C.3d, the unit of observation is a district-technology pair, observed over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

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

(a) Patents Pooled Across Technologies

(b) Patents by Technology

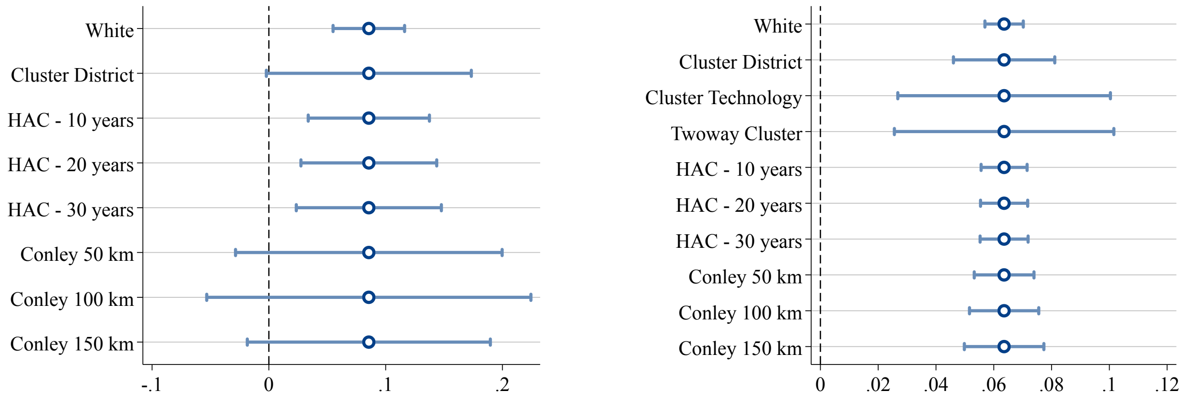


Notes. This Figure displays how shocks to US innovation activity impact innovation in the UK. In Panel C.4a, the unit of observation is a district observed at a yearly frequency between 1870 and 1930; in Panel C.4b, the unit of observation is a district-technology pair observed over the same time. The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

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

(A) Patents Pooled Across Technologies

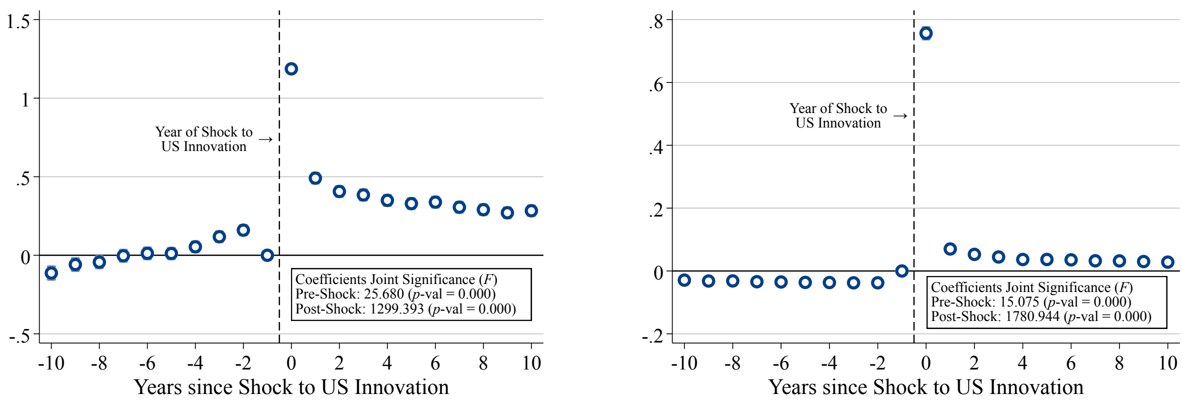
(B) Patents by Technology



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

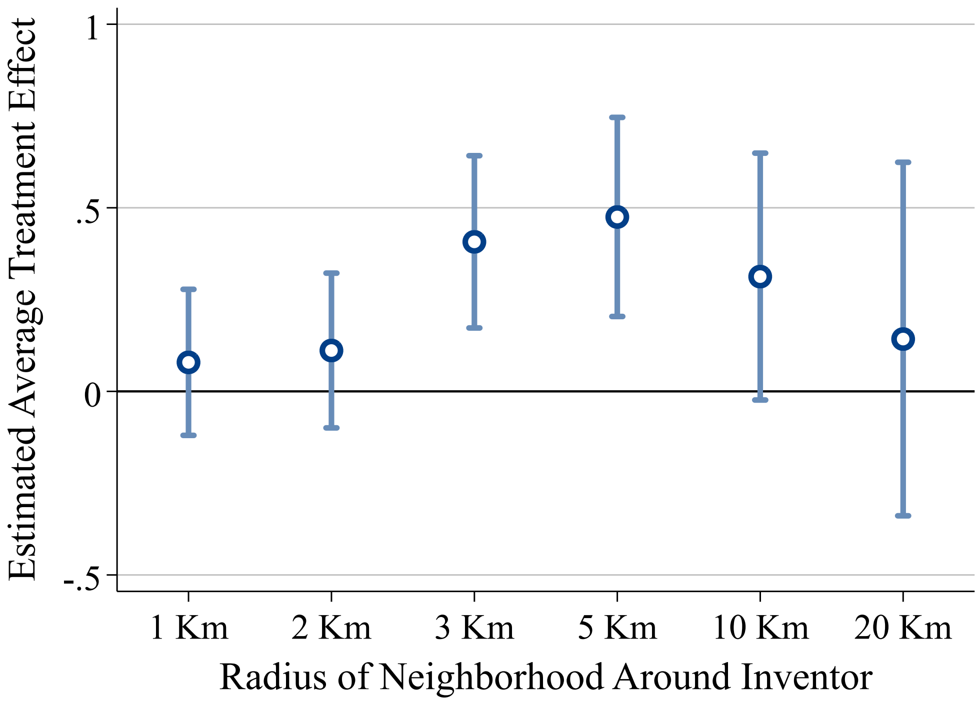
FIGURE C.6. Dynamic Shocks to US Innovation Activity

(A) Double Differences (B) Triple Differences



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

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



Notes. This Figure reports how emigrants to the United States impact the innovation activity fulfilled by their neighbors who remain in the UK. The unit of observation is an individual inventor observed at a yearly frequency between 1880 and 1900. The analysis sample is the universe of inventors linked to the 1891 population census, as detailed in the main text. The dependent variable is the (log) number of patents produced by the members of the family. Each dot reports the coefficient of a treatment variable equal to one after the first neighbor of the inventor moves to the US, and zero otherwise. A neighborhood is defined as a k-kilometer radius area around each inventor. Each dot refers to a different definition of neighborhood with k equal to 1, 2, 3, 5, 10, and 20 kilometers. The baseline case reported in the main text assumes $k = 5$. All regressions include inventor and year-fixed effects. Standard errors are clustered at the county level. Bands report 95% confidence intervals. Referenced on page(s) 29, C51.

Appendix References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

1617 to 1895." MFIS [data collection].

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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