European Immigrants and the United States' Rise to the Technological Frontier in the 19th Century

Costas Arkolakis (Yale, NBER), Sun Kyoung Lee (Columbia) and Michael Peters (Yale, NBER)

January 2018

Abstract

We study whether immigrants from Europe accelated the rise of the US to the technological frontier in the 19th century through the "import" of new knowledge and productive ideas. To do so, we exploit novel historical data on immigrants' pre-migration occupations. By linking immigration records with the US Federal Population Census, we can construct empirical measures of shocks to the supply of experienced workers across all US counties. We then relate these shocks to patent activity and measures of regional productivity growth obtained from the manufacturing census. Our preliminary results suggest that immigrants increased patent activity and that patent activity is positively correlated with future productivity growth.

1 Introduction

The transformation of the US economy in the last two hundred years has been remarkable. While being primarily rural in the beginning of the 19th century, the US had developed into an essentially industrial nation when the century came to an end. More strikingly, after lagging behind the technological frontier (represented by the UK) for most of the 19th century, the US entered the twentieth century as the global technological leader and the richest nation on the globe (Gordon, 2017). During this time period, which is also referred to as the "The Second Industrial Revolution", the US economy also experienced a massive in-flow of immigrants, mostly from the European continent (see Abramitzky and Boustan (2017) for a recent survey).

In this paper, we develop a massive comprehensive dataset that combines longitudinal information on immigrants and their occupations along with measures of economic outcomes such as, output, wages, and innovation at the disaggregated county-industry level. Using this information we investigate to what extent this influx of immigrants was an important contributor to the transformation of the American economic landscape between 1850 and 1929.

At heart of our empirical strategy are the original immigration records and historical passenger lists from the ships heading from Europe to the US.[1] These data sources are real treasure troves for empirical researchers, as they contain direct micro data on immigrants' *pre*-migration occupations.[2] In particular, both in the immigration records and in the passenger lists, all immigrants were

[1]The immigration database of 13 million immigrants and the passenger lists of around 5 million immigrants leaving for the US via the German port Hamburg, the so-called "Hamburg Passenger Lists" were provided to us for research purposes by the Battery Conservancy and the Archives of the city of Hamburg, respectively. See http://www.castlegarden.org and http://www.germanroots.com/hamburg.html for additional information. To the best of our knowledge, these data sources have yet to be used in empirical research.

[2]We constructed a crosswalk between these published occupational strings and the *Historical International Classification of Occupations (HISCO)*. For more details on the Historical International Classification of Occupations, see http://historyofwork.iisg.nl/index.php.

required to give a detailed account of their last occupation in Europe along with other important information such as the time of arrival in the US, the place of residence in Europe, their age and other characteristics. We further merge these immigration datasets with the restricted use Complete Count Federal Demographic Decennial Censuses of the US available from 1810-1940 using modern record-linking techniques. To link these datasets, we exploit the fact that both the immigration records (and passenger lists) and the Federal Census contain information about the names and other time-invariant individual characteristics. In that way we can construct a large-scale micro *panel* data set for immigrants with information on their *pre*-migration occupations and their *post*-migration labor market outcomes and spatial mobility patterns over their whole life-time.

Using this integrated dataset, we explore one specific mechanism by which immigration could have led to American prosperity: the transmission of *new* knowledge. While anecdotal evidence of this channel, whereby European immigrants brought novel ideas to the American shore, is abundant, there is no systematic evidence whether this mechanism was quantitatively important. To fill that gap, we first show that the information on immigrants' pre-immigration occupations allows us to construct empirical proxies for the flow of novel ideas, or in short, *flows of knowledge*. We then combine this immigrant micro data with novel measures of wages, productivity growth and patent activity. In particular, by digitizing the published results for the decennial Historical Manufacturing Census from 1860 to 1939, we construct data on wages and productivity measures at the county-industry level. Secondly, we also complement our data with new county-industry measures of patent activity and patent novelty.[3]

This combination of micro-information on immigrants and macro-measures of productivity and spatial idea creation allows us to relate knowledge flows (as proxied by inflowing immigrants with pre-migration expertise) to data on productivity growth and patent activity for the study period. By doing so, we hope to provide novel evidence on potential mechanisms by which past immigrant settlements could affect economic outcomes (see e.g. Nunn et al. (2017) and Akcigit et al. (2017)). Furthermore, we think that our study period is not only interesting in itself, but it also provides an ideal laboratory to empirically identify the importance of idea flows, which feature prominently in recent theories of economic growth (Kortum, 1997; Lucas Jr and Moll, 2014; Perla and Tonetti, 2014). As communication flows and technology were far less developed than those of today, the importance of embodied knowledge transmission was arguably much more important at that time.

2 A guiding example: The Story of Heinrich Engelhard Steinweg

To illustrate how our novel data can be merged with existing data sources to shed light on the process of technology transfer from old Europe, we start by a particular example: the case of Heinrich Engelhard Steinweg, later known as Henry Engelhard Steinway, the founder of renowned piano manufacturing company *Steinway & Sons*.

Heinrich Steinweg left Germany on May 28th 1850 via the port of Hamburg. This information is declared on the *Hamburg Passenger Lists (1850-1934),* which is available to us through a cooperation with the Archives of the city of Hamburg. His shipment record also indicates his pre-immigration occupation in Germany as *Instrumentenmacher* (instrument maker).[4] As we can see

[3] Our measure of spatial novelty, which we describe in more detail below, is inspired by Kelly et al. (2017).

[4] Henry Steinway started working on producing pianos early on with immediate success. But the unstable political climate following the revolutions of 1848 and the limited economic opportunities for a man working outside a guild let him to immigrate to the US. See Claudius Torp, "Heinrich Engelhard Steinway." In Immigrant Entrepreneurship: German-American Business Biographies, 1720 to the Present, vol. 2, edited by William J. Hausman. German Historical Institute.

from the same record, shown in Figure 1, his destination was New York and he was accompanied by four family members.



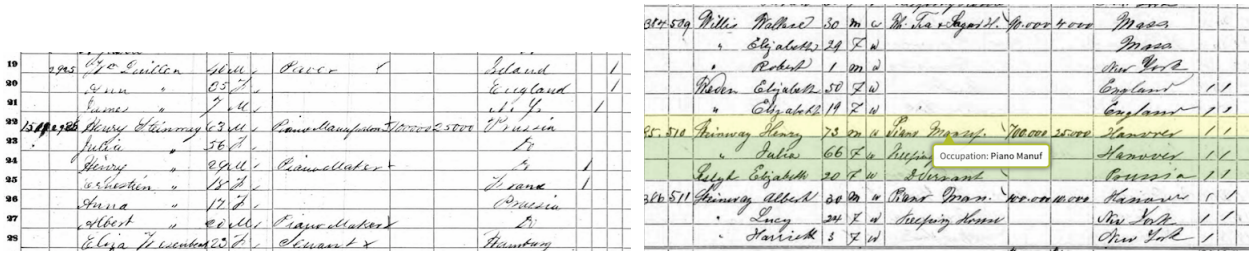
H Steinweg
in the **Hamburg Passenger Lists, 1850-1934**

Name:	H Steinweg
Occupation:	Instrumentenmacher
Birth Place:	Seesen, Hannover
Departure Date:	28 Mai 1850 (28 May 1850)
Port of Departure:	Hamburg
Destination:	New York (New York City (All Boroughs))
Port of Arrival:	New York (New York City (All Boroughs))
Ship Name:	Helena Sloman
Captain:	Paulsen, P N
Shipping Clerk:	Rob. M. Sloman
Shipping line:	Rob. M. Sloman
Ship Type:	Dampfschiff
Accommodation:	ohne Angabe
Volume:	373-7 I, VIII A 1 Band 001
Household Members:	

Name	Age
H Steinweg	
Steinweg	
Steinweg	7
Steinweg	
Steinweg	23

Figure 1: Heinrich Steinweg in the *Hamburg Passenger Lists, 1850-1934*

We can track Mr. Steinweg, now Mr. Steinway, in the subsequent US population censuses, shown in Figure 2. Both in in 1860 and 1870, Mr. Steinway and his family are recorded to reside in New York. Furthermore, his occupation *piano manufacturer* indicates Steinway's successful transition from a piano maker in Germany to the piano manufacturer of the US.



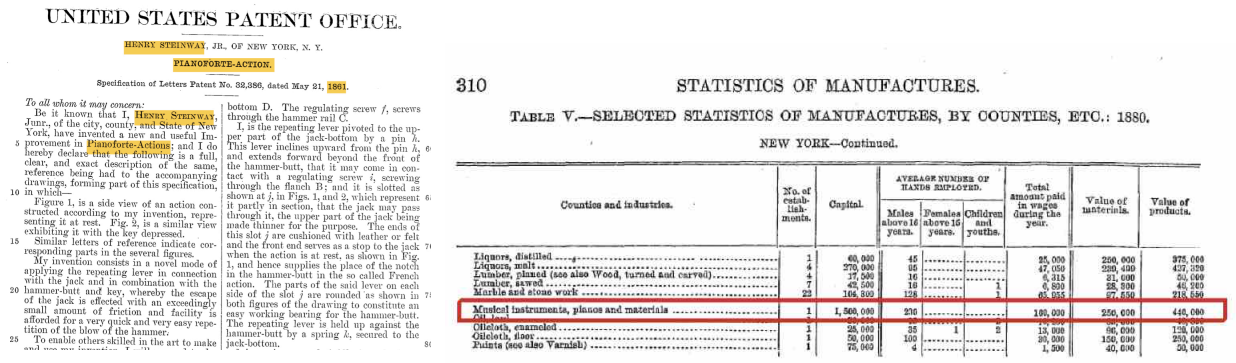
Notes: The figure shows Mr Steinway's US census records in 1860 (left panel) and 1870 (right panel).

Figure 2: Henry Steinway in the *US Census Schedules 1860 and 1870*

That this successful career trajectory might have been in part due to Mr Steinweg's prior knowledge is consistent with micro data on patenting. Using digitized historical patent data from the United States Patent and Trademark Office, we could extract a number of patents granted to him and his sons' names. For example, Steinway's famous piano-forte patent, dated 1862, is shown in the left panel of Figure 3.

Finally, we can use our newly digitized data from the US Census of Manufactures to learn about the economic magnitude of Mr. Steinway's success. While the US Census of Manufacturers data is not at the plant level (but reported at industry-by-county cells), the information is detailed enough to identify the main manufacturing plant of Steinway & Sons in Queens, NY. As the right panel of Figure 3 shows, the Steinway family had an enormous impact on manufacturing production in the New York area. The digitized Census of Manufacturers for the year 1880, for example, reveals that this single piano manufacturing plant was one of the most capital intensive sectors in New York

City with more than $1.5 millions of capital and sales close to a half a million dollars.



UNITED STATES PATENT OFFICE.

HENRY STEINWAY, JR., OF NEW YORK, N. Y.

PIANOFORTE-ACTION.

Specification of Letters Patent No. 32,386, dated May 21, 1861.

To all whom it may concern:

Be it known that I, HENRY STEINWAY, Junr., of the city, county, and State of New York, have invented a new and useful Improvement in Pianoforte-Actions; and I do hereby declare that the following is a full, clear, and exact description of the same, reference being had to the accompanying drawings, forming part of this specification, in which—

Figure 1, is a side view of an action constructed according to my invention, representing it at rest. Fig. 2, is a similar view exhibiting it with the key depressed.

Similar letters of reference indicate corresponding parts in the several figures.

My invention consists in a novel mode of applying the repeating lever in connection with the jack and in combination with the hammer-butt and key, whereby the escape of the jack is effected with an exceedingly small amount of friction and facility is afforded for a very quick and very easy repetition of the blow of the hammer.

To enable others skilled in the art to make

bottom D. The regulating screw *f*, screws through the hammer rail C.

I, is the repeating lever pivoted to the upper part of the jack-bottom by a pin *h*. This lever inclines upward from the pin *h*, and extends forward beyond the front of the hammer-butt, that it may come in contact with a regulating screw *i*, screwing through the flanch B; and it is slotted as shown at *j*, in Figs. 1, and 2, which represent it partly in section, that the jack may pass through it, the upper part of the jack being made thinner for the purpose. The ends of this slot *j* are cushioned with leather or felt and the front end serves as a stop to the jack when the action is at rest, as shown in Fig. 1, and hence supplies the place of the notch in the hammer-butt in the so called French action. The parts of the said lever on each side of the slot *j* are rounded as shown in both figures of the drawing to constitute an easy working bearing for the hammer-butt. The repeating lever is held up against the hammer-butt by a spring *k*, secured to the jack-bottom.

310 STATISTICS OF MANUFACTURES.

TABLE V.—SELECTED STATISTICS OF MANUFACTURES, BY COUNTIES, ETC.: 1880.

NEW YORK—Continued.

Counties and industries.	No. of establishments.	Capital.	Average number of hands employed. Males above 16 years.	Females above 15 years.	Children and youths.	Total amount paid in wages during the year.	Value of materials.	Value of products.
Liquors, distilled	1	60,000	45			25,000	250,000	375,000
Liquors, malt	4	270,000	95			47,050	220,400	457,370
Lumber, planed (see also Wood, turned and carved)	4	17,500	16			6,315	21,000	34,000
Lumber, sawed	7	42,500	18		1	6,830	28,300	46,265
Marble and stone work	23	166,800	128		1	63,955	97,850	218,840
Musical instruments, pianos and materials	1	1,500,000	236			126,000	250,000	448,000
Oilcloth	2	250,000	90			30,000	80,000	150,000
Oilcloth, enameled	1	25,000	35	1	2	13,000	90,000	128,000
Oilcloth, floor	1	50,000	100			30,000	150,000	250,000
Paints (see also Varnish)	1	75,000	4			1,200	40,000	50,000

Figure 3: Mr Steinway's Pianoforte patent and Steinway & Sons in the *US Census of Manufacturers 1880*

3 Primary Data Sources and Methodology

In this section we give more details on the data collection. In Section 3.1 we describe our main historical data sets, in Section 3.2 we discuss how we harmonize and link these data sources. An overview is contained in Figure 4. We rely on four distinct datasets. First of all, we digitize and harmonize two novel historical datasets: historical immigration records to construct a new database of millions of immigrants who arrived in the US during the Era of Great Migration and historical Census of Manufacturers data at the county/city-industry level to measure output and productivity.

We then combine these datasets with two other large datasets. Using record matching techniques, we link the data on immigrants with individual-level data from the US Population Census. This step yields an unprecedented panel dataset with pre- and post-immigration labor market outcomes for millions of immigrants. To construct direct measures of innovative activity, we also incorporate the population of US historical patents since 1790. The combination of these datasets allows us to systematically explore the relationship between immigrants' prior expertise, productivity growth and patent activity at the county-industry level.

3.1 Historical Data on Immigrants, Local Productivity and Innovation

In this section we describe the four datasets, which form the basis of our analysis in detail.

3.1.1 *The Immigration Database: Immigration Records and Passenger Lists (1820 - 1914)*

We construct our immigration database from two primary sources: the Castle Garden Immigration Database and the Hamburg Passenger Lists. The *Castle Garden Immigration Database (1820-1892)* is an educational project of the Battery Conservancy. The database contains the list of all immigrants entering the US via the port of New York between 1820 to 1892. In total, the database comprises approximately 11 million micro-records. Through a cooperation with the Battery Conservancy, we have access to the entire Castle Garden Immigration database. Importantly, these immigration records contain detailed pre-immigration occupation information of respective immigrants.

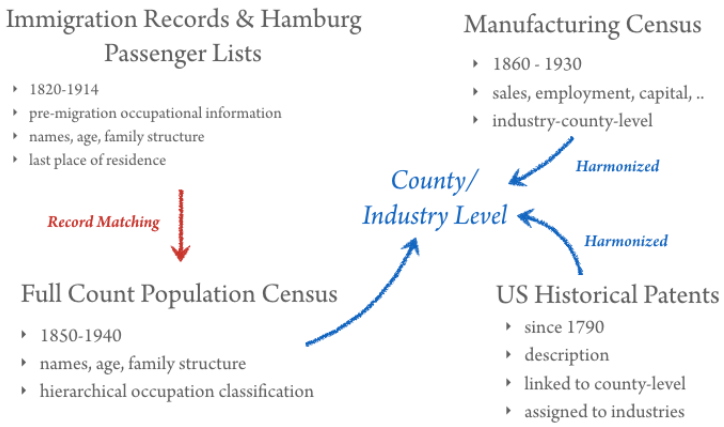


Figure 4: Data Construction: Combining the Data Sets

We complement this database with original passenger lists of all immigrants leaving from the port of Hamburg to US between 1850 and 1914 called the *Hamburg Passenger List Database (1850-1914)*. We have access to the complete records through a cooperation with the Hamburg State Archive. For our main analysis, we translate these passenger lists with pre-immigration occupation in German to English and construct a panel data of immigrants leaving from the port of Hamburg to US demographic census records by using time-invariant information such as name, gender, year of birth, arrival year and place of birth. Details regarding record linking are discussed in 3.2.

3.1.2 *The Complete Count Federal Demographic Decennial Census (1850-1940)*

To measure the impact of immigration, we require information on immigrants' characteristics after their arrival in the US. We do so by linking the individual immigration records from the Immigration Database with the Complete Count Federal Demographic Census. We take advantage the complete transcription of Federal Census Records between 1850 to 1940. Although the federal census year records exist for all years except for 1890 (which was lost due to fire), we currently focs on the records where information on the occupation of individuals are available. This is the case in the censuses of 1850, 1880, 1900, 1910, 1920, 1930, 1940. The transcription of the occupation variable of 1860, 1870 complete count census records are already underway, and this will be integrated once the datasets become available.

Our procedure to link the immigration data to the US census data, which we describe in detail in Section 3.2, allows us to construct a unique dataset, where we can observe both the pre-migration occupation and the entire employment lifecycle in the US for hundreds of thousands of immigrants in the second half of the 19th century. In terms of observables, the complete-count Demographic Census contains various socio-economic characteristics. In particular, occupation classification following the *Historical International Classification of Occupations (OCCHISCO)* is extremely detailed and well-suited to measure immigration induced knowledge flows.

3.1.3 *Measuring Productivity: The Historical Census of Manufacturers (1870 - 1929)*

Our empirical strategy heavily relies on spatial variation in productivity growth, innovation activity and the settlement of immigrants. Measuring productivity at a fine spatial resolution in the 19th Century is difficult. First of all, there are no measures of wages at the county-level. While

the available individual-level data in the decennial Population Census contains county-identifiers, earnings have only been reported starting in 1940. Secondly, information on labor earnings stemming from the National Accounts is available in the 19th Century, but the data does not have a spatial dimension. To overcome this problem, we digitized the published results from the Census of Manufacturers. These tables are published at either the county-industry level or city-industry level and report standard information from firms balance sheets. In particular, they report the number of manufacturing establishments, total number of workers, value of manufactured output, wage-bill and value of the capital stock. As an example, recall the information on the Steinway & Son Factory shown in Figure 3. As our main measures of spatial productivity, we consider total value added per manufacturing employee, manufacturing value added relative to the wage bill or manufacturing revenue TFP, i.e. $Y_r/\left(K_r^{\alpha}L_r^{1-\alpha}\right)$, where Y_r is total value added and K_r and L_r denote the capital stock and employment. We digitized this data at the *county*-industry level for the years 1860, 1870, 1880 and 1929 and at the *city*-industry level for the years 1880, 1900, 1909, 1919, 1929 and 1939. In Section below, we explain how we harmonized this information and combined with other sources.

3.1.4 *Measuring Innovative Activity: Data on Patenting*

To measure the extent of innovative activity at the county level, we exploit information on patenting. The United States Patent and Trademark Office (USPTO) granted millions of patents since 1790 and all patents contain information about the location of the patent and can therefore be geo-referenced. While this information is publicly available in the "HistPat Dataset", we need to harmonize the information on patents at the level of standard industrial classifications.

We use the information on patenting in two ways. We first use the extent of patenting in a given region-sector cell as a measure of idea creation. As we show in Section 4 below, this measure is consistent with a simple idea-based growth model and can easily be mapped to productivity growth in a model-consistent way.[5] More importantly, we use the patent data to devise a novel measure of *spatial idea novelty.* Our theoretical mechanism stresses the importance of European immigrants bringing *novel* insights to the US. To relate immigration inflows to the novelty of the type of ideas invented in a particular county, we use textual analysis to measure the extent to which patents originating in a particular region are similar to the patents that have been invented in that region in the past. Intuitively, as in Kelly et al. (2017), we measure the similarity between two patents as the correlation in the words, respective patents use. Given this measure of patent-to-patent similarity, we can then calculate the similarity of ideas invented in a region as the average patent-to-patent similarity between new patents originated in a particular region and the set of patents stemming from a particular region in the past. Our measure of *spatial idea novelty* is then simply the inverse of this average similarity and we can relate it systematically to the spatial inflow of immigrants.

3.2 Dataset Construction: Record Matching And Harmonization

To combine the four datasets introduced in Section 3.1 into a single database, we need to (i) link the micro-level Immigration Database with the Full Count US Demographic Census and (ii) provide consistent crosswalks to merge patents, measures of productivity and immigrant inflows at the spatial and sectoral level. In this section we provide more details for these two steps.

[5]Many contributions using more recent data on patenting stress the importance of weighing patents by their subsequent citations to adjust patents for their quality. Such quality adjustments are not possible in the 19th century patent data as the information on citations patterns is too scarce.

Year	Record Type	Germany	Italy	Ireland	UK	Total
1880	Immigration records	1,530,063	10,843	732,100	392,528	2,665,534
	Census record	1,081,249	31,292	887,609	514,476	2,514,626
	Matches	65,962	543	21,823	26,305	114,633
1900	Immigration records	2,543,625	616,566	761,317	438,249	4,359,757
	Census record	1,431,253	315,105	746,408	639,007	3,131,773
	Matches	309,198	48,527	62,119	47,790	467,634
1910	Immigration records	2,528,249	852,445	761,317	438,251	4,580,262
	Census record	1,343,333	887,044	615,329	676,429	3,522,135
	Matches	281,619	88,703	45,571	40,611	456,504
1920	Immigration records	2,384,298	859,328	713,603	400,094	4,357,323
	Census record	862,070	955,014	460,635	605,809	2,883,528
	Matches	194,219	75,086	25,707	27,590	322,602

Notes: The table contains the total number of immigrants (by county of origin) from the Castle Garden immigration records that migrated to the US *before* the respective year ("Immigration records"), the number of immigrants (by county of origin) in the US Population Census and the number of matches. The last column contains the sum of the four origin countries.

Table 1: Linking Immigration Records and the US Population Census

3.2.1 Record Matching: Linking Immigrants and the Complete Count Federal Demographic Decennial Census, 1850-1940

We have access to the complete individual-level complete-count US demographic federal census records from 1850-1940.[6] By linking our novel immigration records to the US census records, we can measure the entire life-cycle of immigrants since they entered the US. Our record linking procedure has the following characteristics: (1) we rely on all *complete-count* US Federal Census records with *occupation and industry information* (a newly transcribed variable from the original census records), (2) we link people at *more than two points in time* and (3) we use both *individual and household level information* to improve the matching of individuals.

The current results of our record matching procedure are contained in Table 1. The results are preliminary as we have currently only matched four major immigrants groups by sending country (Germany, Italy, Ireland, and UK). Although not included in the table, we also matched the records for the years 1850 and 1930. As in Table 1, our procedure is expected to yield millions of matched immigrant records which we can track their lives for multiple decades. We have already matched more than a half million immigrant records from four major immigrant sending countries between 1850 and 1930.

3.2.2 Data Harmonization and Integration at the Country-Industry Level

We need to combine the microdata from the matched Population Census with additional data sets at the macro, i.e. county-industry level. To do so, it is essential to harmonize the different datasources. As the occupation information in the *Hamburg Passenger Lists* and the *Castle Garden Database* is not classified er constructed a crosswalk to official HISCO occupational classification.

[6] We use the restricted complete-count US demographic census from 1850 to 1940 to link individual-level records by implementing random forest classification. This record linking methodology is similar in spirit to Minnesota Population Center (MPC) Record record linkage project of the 1850-1930 sample census records to the 1880 complete count census records. However, MPC implemented a support vector machines to automate the record linking. Feigenbaum (2016) discusses a machine learning approach to census record linking and compares different possible matching algorithms. He proposes a probit-based method as an ideal choice.

Similarly, we have to match the reported sectoral groups in the Census of Manufacturers to a standard classification of industries. As seen in Figure 3, our Manufacturing Census data is of considerable detail - the single establishment of "distilled liquor" is distinct from the four "malted liquor" plants in Queens, NY. Using these digitized sectoral strings, one can construct a cross-walk to the sectoral classification employed in the Population Census.

Finally, we have to be able to merge all these datasets at a unified spatial level. While the Population Census and the Patent Data is already geo-referenced, we have to perform this geo-referencing using the original geographical information in the Manufacturing Census. As counties and state boundaries have changed over time, we take the county shapefiles from National Historical Geographic Information System, and create the geographic location-based crosswalks of counties across time.

4 Analysis

4.1 Theory: A Simple Romer Model

To guide the analysis and to provide a simple link between the measures of output, innovation and immigration exposure, we consider a simple model in the spirit of Romer (1990). Labor is the only factor production and workers can be employed either as production workers or as "researchers" to generate new innovations ("ideas").

Output is produced using a neoclassical production function $Y_{rt} = A_{rt}L_{rt}^{P}$, where A_{rt} denotes productivity in region r and time t and L_{rt}^{P} is the number of production workers. Productivity A_{rt} is determined by an exogenous technological factor, Q_{rt}, and an endogenous component which we term *knowledge* , N_{rt}. In particular, $A_{rt} = Q_{rt}N_{rt}$. As standard in many idea-based models of growth, we model the growth of knowledge as

$$\dot{N}_{rt} \quad = \quad Z_{rt}N_{rt}^{\lambda}L_{rt}^{R}. \tag{1}$$

Here, the change in knowledge depends on the level existing knowledge, N_{rt}, the innovation productivity Z_{rt} and the amount of innovation resources L_{rt}^{R}. Moreover, $\lambda > 0$ determines the extent of knowledge spillovers. Furthermore, to operationalize the concept of knowledge growth empirically, we assume a proportional relationship between patenting and knowledge growth, i.e. $P_{rt} = \nu\dot{N}_{rt}$, where P_{rst} denotes the *flow* of patents and ν is a normalizing constant. Finally, suppose that the amount of innovation resources L_{rt}^{R} can be written as

$$L_{rt}^{R} = h\left(\pi_{rt}^{I}, E_{rt}\right) L_{rt}, \tag{2}$$

where π_{rt}^{I} is the share of immigrants , E_{rst} is a measure of immigrants' expertise and L_{rt} denotes the size of the population. Hence, the function $h\left(\pi_{srt}^{I}, E_{rst}\right)$ translates raw labor L_{rt} into efficiency units and we assume that both the share of immigrants and their expertise affect this transformation. In our empirical analysis we aim to study the productivity-knowledge relationship and the link between immigrants and knowledge creation.

4.2 Empirical Strategy and Preliminary Results

Patents and Productivity We start with the relationship between local productivity growth and the growth of the stock of knowledge. The model above implies that $g_{A,r} = g_{Q,r} + \frac{1}{\nu}\frac{P_r}{N_r}$, where

Dep. Var.: Prod. Growth 1870 - 1880			
$P_{rs}^{1880}/N_{rs}^{1870}$	0.026^{***}	0.023^{***}	0.022^{**}
	(0.008)	(0.008)	(0.010)
Industry FE	✓	✓	✓
Lagged productivity	✓	✓	✓
Spatial controls		✓	
County FE			✓
Observations	3363	3363	3363
R^2	0.552	0.556	0.667

Notes: Robust standard errors in parentheses. *, **, and *** denotes significance at 10%, 5% and 1% level. Productivity is measured as labor productivity (VA_{rs}/L_{rs}), where VA denotes value added and L denotes employment. Both VA and L are measured from the Historical Census of Manufacturers. P_{rs}^{1880} denotes the flow of patents between 1870 and 1880. N_{rs}^{1870} denotes the accumulated flow of patents with sector-region cells since 1790. The spatial controls (i.e. log population, the share of the urban population and the manufacturing employment share) are measured from the Federal Demographic Census Data in 1870.

Table 2: The Patent-Productivity Relationship in 1880

$g_{X,r} = \frac{\dot{X}_{rt}}{X_{rt}}$ .To implement this prediction empirically, we consider the specification

$$\Delta \ln A_{rst} = \beta \left(P_{rst}/N_{rst} \right) + \alpha_s + \gamma \ln \tilde{A}_{rst} + x_{rt}^{'} \zeta + u_{rst}, \tag{3}$$

where $\Delta \ln A_{rst} = \ln A_{rst+1} - \ln A_{rst}$ denotes productivity growth (measured as value added per worker), x_{rt} includes a number of controls to proxy for systematic differences in exogenous productivity growth $\left(\frac{\dot{Q}_{rst}}{Q_{rst}}\right)$ and u_{rst} is an error term. Note that we estimate (3) at the region-sector level and hence control for sector fixed effects (α_s) to isolate the spatial variation. We measure the *stock* of existing knowledge N_{rst}, by the cumulative patent flow since 1790.[7] . The results of estimating (3) for the period 1870-80 are contained in Table 2.

In column 1 we consider a parsimonious specification and find a positive relationship between patent growth and subsequent productivity growth. In column 2 we add additional spatial characteristics, namely the manufacturing share, the extent or urbanization and market size to control for exogenous productivity growth. This reduces the coefficient on patenting only slightly. In the last columns, we introduce county fixed effects.This again yields quite similar results. To the best of our knowledge, we are the first to provide systematic evidence of a positive relationship between patenting and productivity growth in the 19th century at this level of aggregation.

Immigrants and Innovative Activity After establishing this systematic relationship between innovative activity and regional productivity growth, we turn to the role of immigrants and their prior expertise in generating new insights. Using (1) and (2) we get that

$$\ln P_{rt} = \ln \dot{N}_{rt} = \ln Z_{rt} + \lambda \ln N_{rt} + \ln h \left(\pi_{rt}^{I}, E_{rt} \right) + \ln L_{rt}.$$

To implement this relationship empirically, we require a measure of foreign expertise E_{rt}. For this measure, we exploit our novel immigration data. In particular, we construct a crosswalk between immigrants' *pre*-migration occupations and the OCCHISCO occupational measure contained in

[7]From our measurement equation we have $\dot{N}_t = P_t$. Hence, $N_t = \int_{\tau=0}^{t} P_\tau d\tau + N_0$.If $N_{1790} = 0$, we have that $N_t = \int_{\tau=0}^{t} P_\tau d\tau$, i.e. the knowledge stock is equal to the cumulated number of patents.

	Panel A:			*Panel B:*		
	Dep. Var.: log patents 1870-80			*Dep. Var.: Patent Similarity 1870-80*		
	OLS		IV	OLS		IV
Share of immigrants	0.573***	0.559***	0.578*	-0.161***	-0.190***	-0.571**
in 1870	(0.125)	(0.121)	(0.333)	(0.054)	(0.056)	(0.257)
Share of immigrants	0.088***	0.050*	0.056*	-0.128**	-0.134***	-0.138**
with prior expertise in 1870	(0.025	(0.025)	(0.033)	(0.052)	(0.052)	(0.061)
Industry FE	✓	✓	✓			
Regional Controls	✓	✓	✓	✓	✓	✓
Stock of knowledge	✓	✓	✓			
Share of skilled individuals		✓	✓		✓	✓
Lagged similarity index				✓	✓	✓
log flow of new patents				✓	✓	✓
Observations	2827	2827	2729	883	883	754
R^2	0.555	0.566	0.564	0.376	0.384	0.347

Notes: Robust standard errors in parentheses. *, **, and *** denotes significance at 10%, 5% and 1% level. The spatial controls (i.e. log population, the share of the urban population and the manufacturing employment share) are measured from the Federal Demographic Census Data in 1870. The share of skilled individuals is the fraction of the native population that work in occupations, which are characterized as "conducting research", The lagged similarity index is measured between 1860 and 1870. The flow of patents is the number of patents originated between 1870 and 1880.

Table 3: The Empirical Innovation Immigration Relationship

the US Federal Census, and classify immigrants as having prior knowledge, if their *pre*-migration occupations are categorized as *"conducting research and applying scientific knowledge to the solution of a variety of technological, economic, social and industrial problems".* Hence, we consider a regression specification of

$$\ln P_{rst} = \beta \pi^I_{rt} + \vartheta e^I_{rt} + \lambda \ln N_{rst} + \gamma \ln L_{rt} + \alpha_s + x'_{rt}\zeta + u_{rst}, \tag{4}$$

where π^I_{rt} is the share of immigrants in country r and e^I_{rt} is the share of immigrants with pre-migration expertise.[8]

In the first two columns of Panel A in Table 3 we report the results of estimating (4) using OLS. The results show that there is a significant positive relationship between the share of immigrants within a location and subsequent patent activity. In particular, there is an independent effect of the share of migrants with prior expertise in fostering patent activity. These results are consistent with immigrants, in particular immigrants with prior expertise, acting as conduits of novel techniques.

To provide further evidence for this channel, we focus directly on the *novelty* of the set of newly invented patents. Using textual analysis, we construct a measure of *spatial patent novelty* from the detailed patent descriptions. Intuitively, we compare the similarity of the words used in a given patent from region r at time t, with the set of patents from region r *up to* time t. We then construct an index of average similarity, which indicates whether a region generates - on average - similar patents at time t as it used to. If immigrants are indeed a source of novel techniques, we expect regions with more immigrants to have lower values of the similarity index. This is exactly what we find. Consider the first two columns in Panel B in Table 3, where we show that both the share

[8]Again, N_{rst} and L_{rt} is the stock of knowledge and the size of the workforce in county r and x_{rt} are again the regional characteristics (the urbanization rate and the manufacturing employment share) used above to control for unobserved differences in research productivity $\ln Z_{rst}$. For brevity we do not report the respective coefficients.

of immigrants and the share of immigrants with prior expertise are negatively correlated with the similarity index.[9]

The specification in (4) is subject to endogeneity concerns as the share of immigrants, in particular the share of immigrants with foreign expertise, might be correlated with regional differences in innovation productivity, Z_{rst}. To address this endogeneity, we currently consider a instrumental variable strategy in the spirit of Card (2001), where we instrument the share of immigrants in county r at time t with the *predicted* share of immigrants using the cross-sectional variation of immigrants of origin o thirty years prior.[10] In the respective last columns of Table 3, we show the point estimates qualitatively similar.

References

Abramitzky, R., Boustan, L., 2017. Immigration in american economic history. Journal of Economic Literature 55(4), 1311–45.

Akcigit, U., Grigsby, J., Nicholas, T., 2017. Immigration and the rise of american ingenuity. Technical report, National Bureau of Economic Research.

Card, D., 2001. Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. Journal of Labor Economics 19(1), 22–64.

Feigenbaum, J. J., 2016. Automated census record linking: A machine learning approach. Technical report, http://jamesfeigenbaum.github.io/research/census-link-ml.html.

Gordon, R. J., 2017. The rise and fall of American growth: The US standard of living since the civil war. Princeton University Press.

Kelly, B., Papanikolaouz, D., Seru, A., Taddy, M., 2017. Measuring technological innovation over the long run. Working paper .

Kortum, S., 1997. Research, patenting, and technological change. Econometrica 65(6), 1389–1419.

Lucas Jr, R. E., Moll, B., 2014. Knowledge growth and the allocation of time. Journal of Political Economy 122(1), 1–51.

Nunn, N., Qian, N., Sequeira, S., 2017. Migrants and the making of america: The short and long run effects of immigration during the age of mass migration .

Perla, J., Tonetti, C., 2014. Equilibrium imitation and growth. Journal of Political Economy 122(1), 52–76.

Romer, P. M., 1990. Endogenous technological change. The Journal of Political Economy 98(5), S71–S102.

[9] Note that we explicitly control for lagged similarity (as some regions might just always generate more novel ideas) and the flow of new patents (as many new patents might mechanically reduce the similarity index).

[10] More specifically, let $\Delta I_o^{1850\to t}$ be the number of new immigrants entering the US from county o between 1850 and year t. Let I_{rt}^o be the number of immigrants from origin o in county r at time t. We can calculate the *predicted* number of immigrants as $\hat{I}_{rt}^o = I_{r1850}^o + \sum_{\tau>1850}^{t} \Delta I_o^{1850\to t} \times \frac{I_{r1850}^o}{\sum_j I_{j1850}^o}$ and use $\hat{I}_{rt}^o$ as an instrument for I_{rt}^o.