Online Appendix for

Migrants and the Making of America: The Short- and Long-Run Effects of Immigration during the Age of Mass Migration

Sandra Sequeira

London School of Economics, CEPR

Nathan Nunn

Harvard University, NBER, BREAD

Nancy Qian

Yale University, NBER, BREAD

(Not for Publication)

March 2017

A1. Data

A. *Railway Data*

Construction of the digitized railway network occurred using a "process of subtraction". We first obtained an accurate and geo-referenced shape file of the railway network in the modern period from the U.S. Department of Transportation. The organization has compiled the *National Transportation Atlas Railroads* (NTAR) dataset in GIS format at 1:100,000 scale, which was current as of 2009. The NTAR railroad dataset depicts railroads with greater detail than was needed for our purposes. Specifically, it contains very fine grained information on all of the railway tracks within railway yards of cities. Since these are not relevant for the measure that we construct, these additional tracks were deleted from the modern shapefile, and therefore are not included in the historical shapefiles.

We then laid the modern shapefile over a digitized version of a paper map of the most recent historical time period of interest, which is 1920. We proceeded to remove all railway lines that exist today but did not exist in 1920. We then repeated this for each earlier time period in sequence – i.e., 1910, then 1900, then 1890, etc. At each point, we removed railway lines that existed in the latter period but did not exist in the period in question. This procedure ensures the greatest precision in digitizing the exact location of the railway lines. Because of mapping imprecisions in the original historical maps and changes in mapping technology across time, simply tracing the lines from each paper map would have generated inaccurate maps of historical railway networks.

The set of historical maps that were used are summarized in Table A1. The top section of the table reports the list of all maps that cover the full railway network that existed at the time. In addition, we were able to access a number of maps that did not have full coverage, but only reported the railway network for a specific region. These are listed in the bottom half of the table. These were used to cross-check and validate the full-coverage maps. Digital images of all maps were obtain from the website of the

Library of Congress: http://memory.loc.gov/ammem/gmdhtml/rrhtml/rrmap.html. The digitization of the railway network was done manually by a professional company that specializes in GIS-based data construction, Shofield Brothers of New England, Inc.

Shapefiles were created for every year for which a full-coverage map was available. Therefore, as reported in table A1, shapefiles were created at frequencies higher than every decade, and for years prior to 1850 (the earliest year we use in our analysis). We match the higher frequency railway data to our decade intervals by choosing the year that was closest to the census years. In figures A1–A11, we report, for each year of railway data that is used in our analysis (i.e. decade), the full-coverage map that is the basis of the data for that period, as well as the digitized railway network for that year.

Table A1: List of Maps used to Construct Digitized Railway Data.

Map Title	Scale	Year
National Maps		
Map of the Railroads and Canals of the U.S.	not given	1830
Railroads and Canals	not given	1834
Map of the United States	not given	1839
United States of America Railroad Guide	not given	1847
New and Complete Map of the U.S.	not given	1850
A New and Complete Railroad Map of the United States	not given	1857
Lloyd's Railroad, Telegraph and Express Map of the United States and Canada	not given	1867
Railroad map of the United States	not given	1873
New Country and Railway Map of the United States	1:2,027,520	1876
New Country and Railway Map of the United States	not given	1883
Official Railroad Map of the U.S.	1:3,168,000	1890
Rand McNally &Co.'s Official Railroad Map of the Unites States	1:2,280,960	1893
Gray's New Trunk Railway Map of the United States	not given	1898
Canadian Pacific Railway and Connecting Lines	1:8,236,800	1912
Map of Principal Transportation Lines of the United States	1:7,286,400	1921
Regional Maps		
California		
General Map of a Survey of California in Connection with Examinations for Railroad Routes to the Pacific Ocean	1:1,100,000	1855
Railroad Map of the Central Part of California	1:253,440	1865
Map of California Southern Pacific R.R.	1:2,090,880	1876
Rand McNally & Co.'s Railroad and County map of California	1:1,200,000	1883
Illinois		
Railway Guide for Illinois	1:850,000	1855
G. Woolworh Colton's Railroad Map of Illinois	1:380,160	1861
Burlington, Cedar Rapids and Minnesota Railway and Connections	1:1,267,200	1868
Louisville, New Albany, and St. Louis Air Line Railroad and Connections	1:2,217,600	1872
Map of the Danville, Olney, & Ohio River Railroads	1:1,267,200	1881
Railroad Map of Illinois	1:900,000	1898
Indiana		
Railroad Map of Indiana	1:1,000,000	1850
Johnson's map of Indiana Railroads and Townships	1:675,000	1858
Map of the Iowa & Missouri State Line Railroad and its connections	1:1,267,200	1868
Map of the Danville, Olney, & Ohio River Railroads	1:1,267,200	1881
Indiana	1:850,000	1888
Calbraith's Railway Mail Service Map of Indiana	not given	1897
New England		
Map of Massachusetts, Connecticut and Rhode Island	not given	1831
Map of Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, and Connecticut Showing the Post Offices, Post Roads, Canals, Rail Roads, etc.	not given	1838
Railways in New England and Part of New York	not given	1847
Railroad Map Eastern	not given	1859
New Haven, Middleton, and Boston Railroads and connections	1:900,000	1867
Connecticut Western Railroad and It's connections	1:1,750,000	1871
Boston & Maine Railroad and Connections	not given	1898

Figure A1: Railway map and shapefile for 1850.

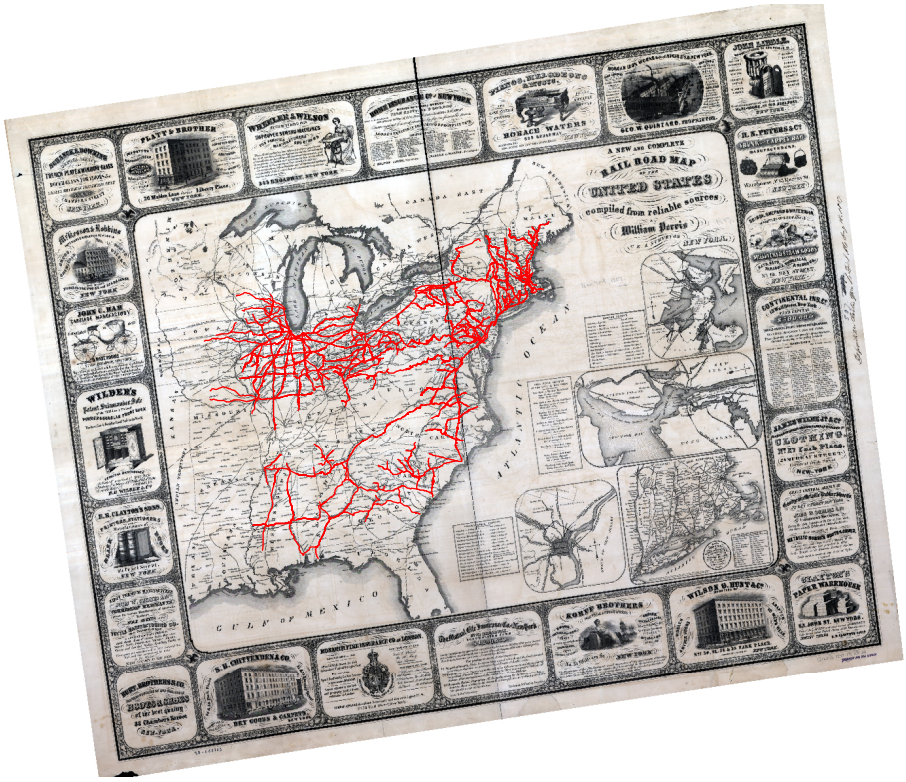


Figure A2: Railway map and shapefile for 1857.

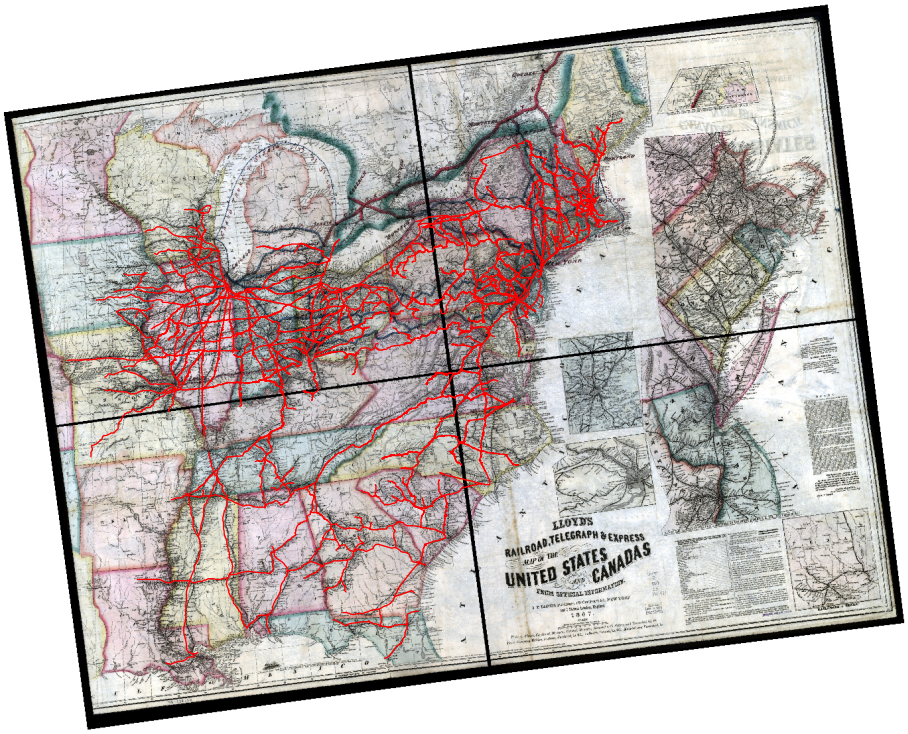


Figure A3: Railway map and shapefile for 1867.

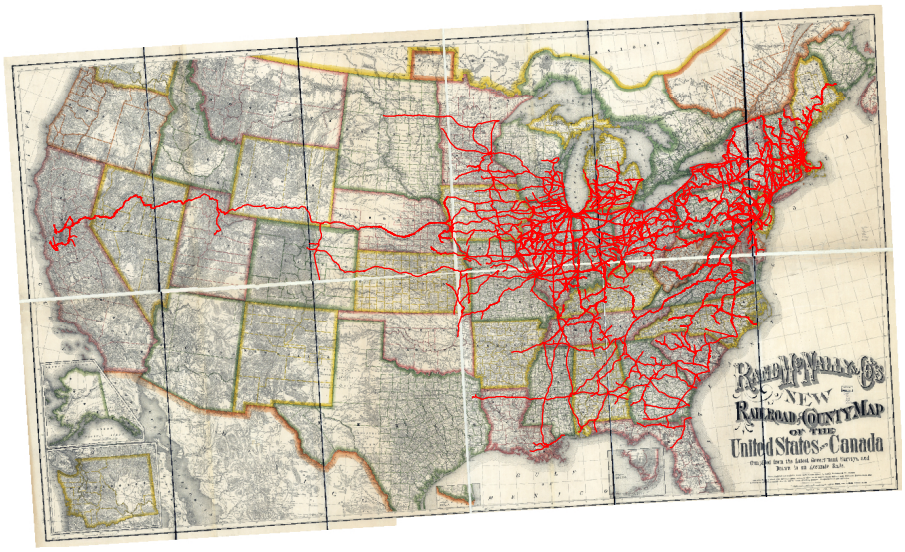


Figure A4: Railway map and shapefile for 1873.

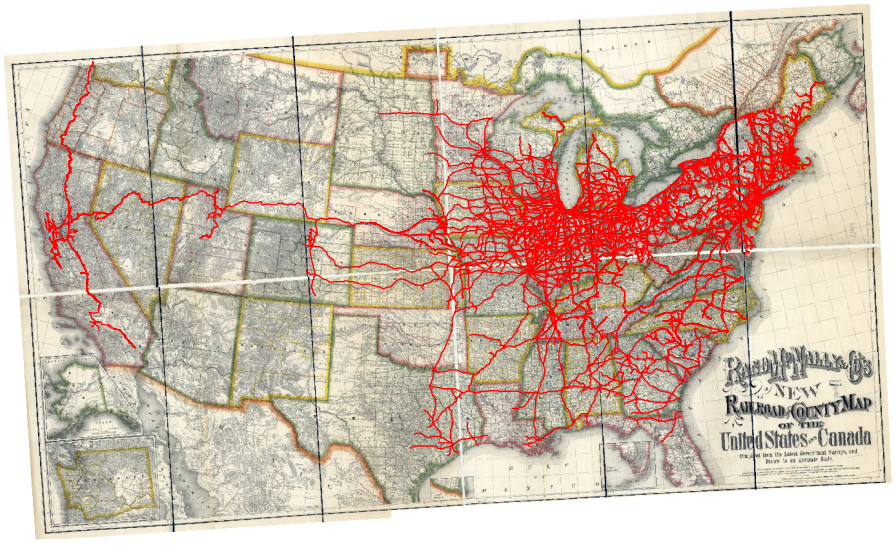


Figure A5: Railway map and shapefile for 1876.

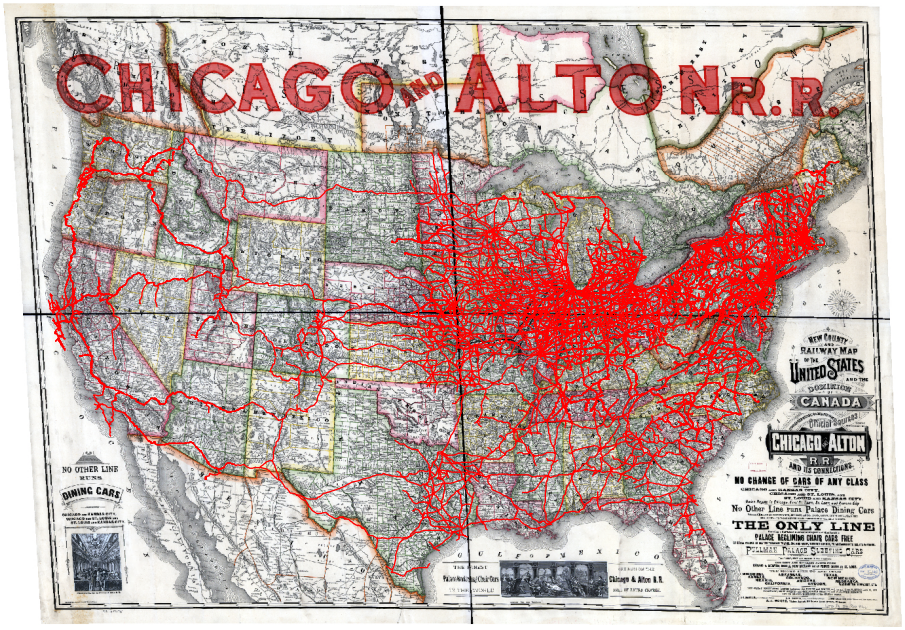


Figure A6: Railway map and shapefile for 1883.

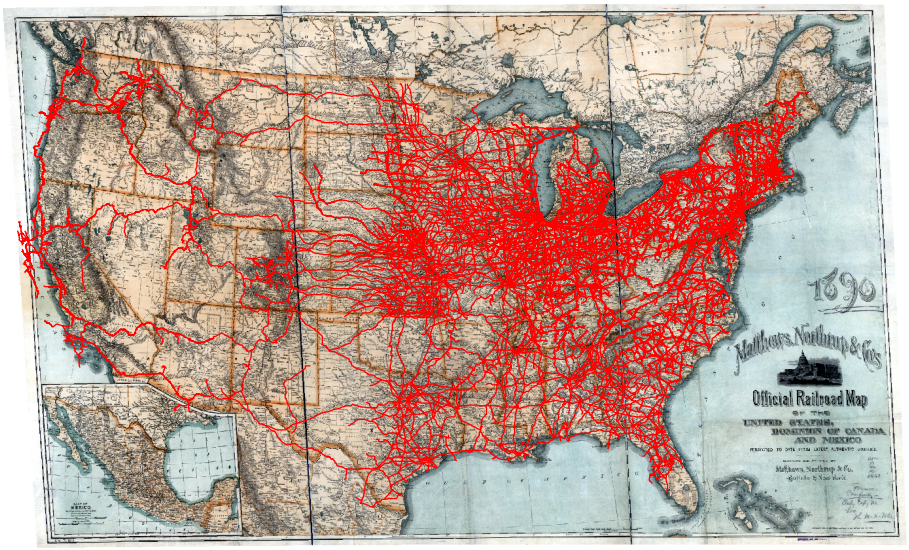


Figure A7: Railway map and shapefile for 1890.

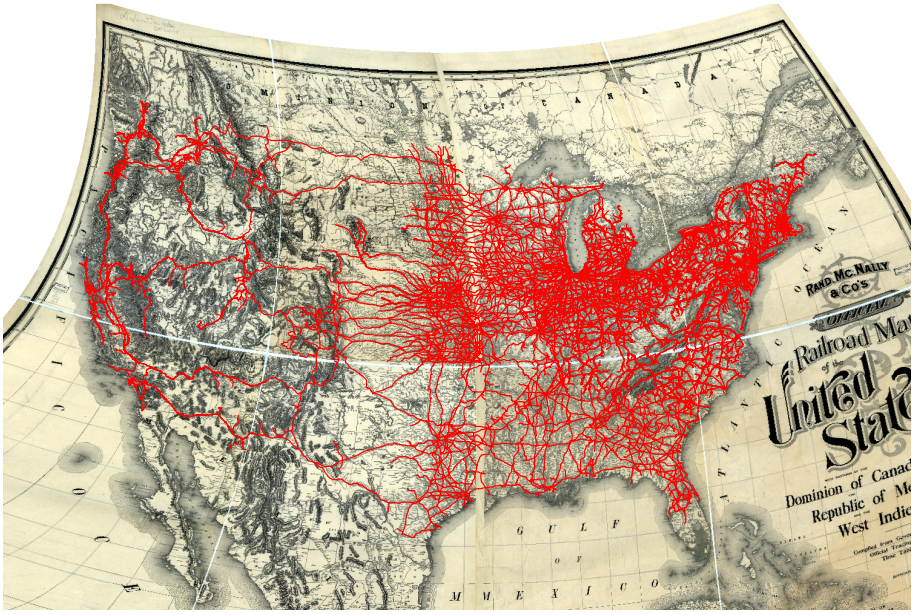


Figure A8: Railway map and shapefile for 1893.

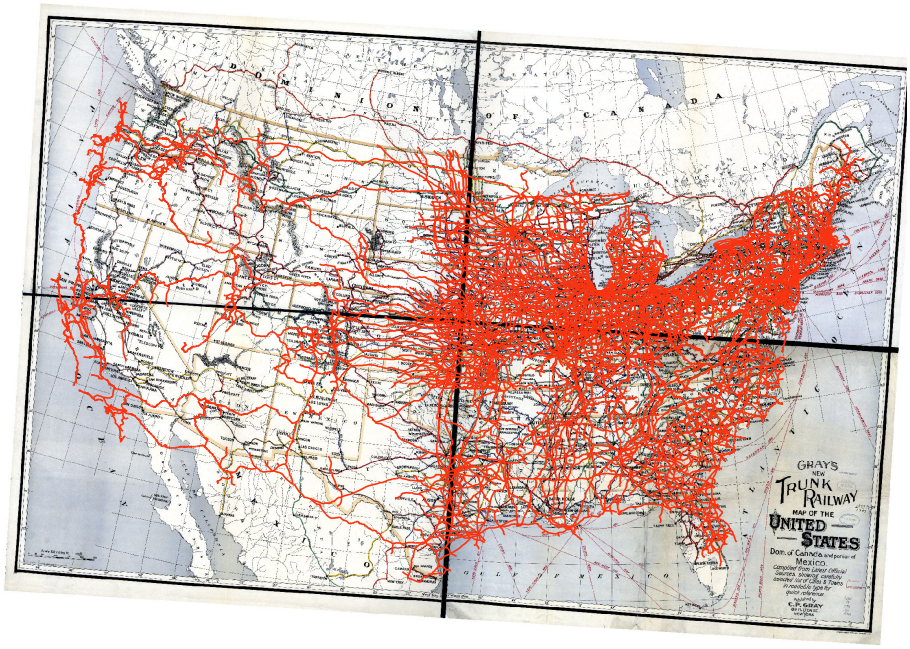


Figure A9: Railway map and shapefile for 1898.

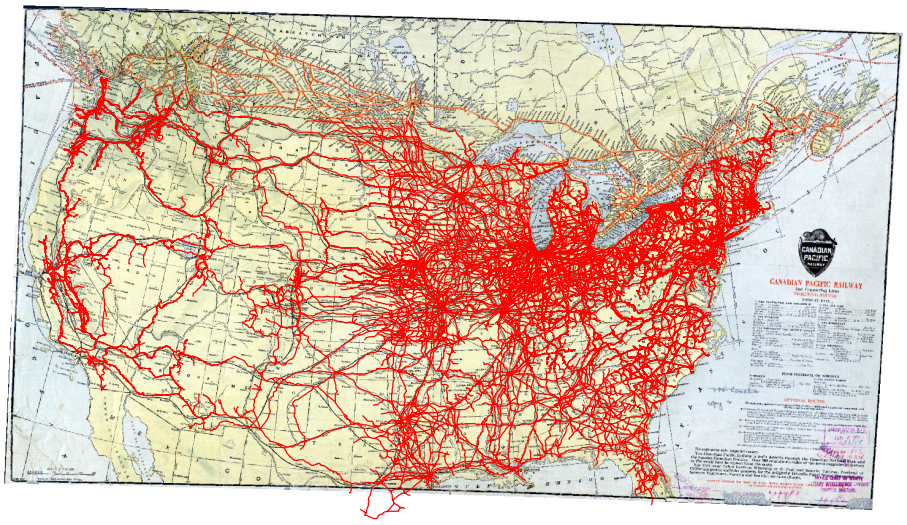


Figure A10: Railway map and shapefile for 1912.

Figure A11: Railway map and shapefile for 1921.

B. ***Other Data***

European immigration data were obtained from *International Migration Statistics* in Willcox (1929-1931). We use this source, rather than other potential sources, because it reports immigration into the United States by sending country. Although this information is not needed for our baseline analysis, it is necessary for our robustness check where we use origin-country weather shocks to predict out-migration to the United States.

Data on income, poverty, population density, education, manufacturing production, agricultural production, total population, foreign born populations, and urbanization rates are obtained from the United States Decennial Census, accessed through the NHGIS, Social Explorer, or ICSPR interfaces. Shapefiles of historical county boundaries were obtained from NHGIS: `https://nhgis.org/documentation/gis-data`. During our time-period of interest, 1860-1920, for a significant proportion of counties in 2000, their boundaries were established after the first period of our sample, 1860. Thus, the number of counties in our zero-stage panel estimates increases overtime. In other words, our zero-stage panel is unbalanced. In 1860, there are 1,600 counties in our sample; in 1870 there are 1,974 counties; in 1880 there are 2,216 counties; in 1890 there are 2,468; in 1900 there are 2,728; in 1910 there are 2,797; and in 1920 there are 2,946. We match counties across time using the nominally integrated series available from NHGIS. A detailed explanation of NHGIS' matching strategy is available at `https://nhgis.org/documentation/time-series#geographic-integration`.

Average years of schooling in 2000 are calculated based on the share of the population that has completed each level of schooling – primary, secondary and tertiary – in each county. Data on patents are obtained from Lexis Nexis and the United States Patent and Trademark Office. Data on voter turnout in 2000 were obtained from David Leips' *Atlas of U.S. Presidential Elections* accessed at `http://uselectionatlas.org`. The measure of social capital is taken from Rupasingha and Goetz (2008). Data on European weather shocks were obtained from Pauling, Luterbacher, Casty and Wanner (2006) for precipitation and Luterbacher, Dietrich, Xoplaki, Grosjean and Wanner (2004) for temperature.

Data on crop area in European countries were obtained from Ramankutty and Foley (1999), and accessed through `earthstat.org`. The industrialization index series between 1831 and 1915 was obtained from Davis (2004).

County-level crime rates were obtained from the Inter-University Consortium for Political and Social Research (ICPSR), with the original source being the "United States Department of Justice. Federal Bureau of Investigation. Uniform Crime Reporting Program Data [United States]: County-Level." We aggregated the number of all types of crime against the person and against property, divided by county population to produce a measure of total crimes per capita.

Data on patents was drawn from Lexis Nexis and the United States Patent and Trademark Office (https://www.uspto.gov/learning-and-resources/electronic-data-products/historical-patent-data-files). We obtained data on 1,297,086 utility patents from 1850 to 1919. Eighty-six per cente of this sample had both a county and state identifier. For the nationality analysis, we used only 50% of the total sample that included the nationality of the inventor directly reported in the patent document and the county and state of registration. We assumed US citizenship when the applicant reports being from an unspecified US county and state and no foreign country is mentioned. Citizenship was more consistently reported from 1890 onwards.

A2. Additional Estimates

Table A2: Zero-stage OLS panel estimates: Placebo test examining the Northeast only.

	(1)
Dependent Variable	**Migrant Share of Total County Population**
	Northeast Only
Lag Rail Access	-0.048
x Lag Migrant Inflow/Total US Population	[0.089]
Lag Rail Access	
x Lag Log Industrialization Index	Yes
Lag Rail Access	Yes
Lag Migrant Share	Yes
Lag Urban Indicator	Yes
Lag Urban Indicator	Yes
x Lag Predicted Migrant Inflow/Total US Population	Yes
Log County Population Density	Yes
County Fixed Effects	Yes
Decade Fixed Effects	Yes
Observations	1,023
R-squared	0.960

Notes : OLS estimates are reported. An observation is a county in a time period (1860, 1870, 1880, 1890, 1900, 1910 or 1920). The dependent variable "Migrant Share of Total County Population" is the proportion of a county's population that is foreign born in period t. "Lag Rail Access" is an indicator variable that equals one if a county has a railway in period t-1. Conley standard errors are reported in square brackets. ***, **, and * indicate significance at the 1, 5 and 10% levels.

Table A3: Zero-stage OLS panel estimates, robustness to omission of a lagged dependent variable.

Dependent Variable	(1)	(2)	(3)	(4)
	Migrant Share of Total County Population			
	All Counties	**Excluding Northeast**	**Excluding South**	**Midwest and West**
Lag Rail Access	0.134***	0.142***	0.204***	0.221***
x Lag Migrant Inflow/Total US Population	[0.031]	[0.033]	[0.053]	[0.059]
Control Variables:				
Lag Rail Access				
x Lag Log Industrial Index	Yes	Yes	Yes	Yes
Lag Migrant Share	No	No	No	No
Lag Rail Access	Yes	Yes	Yes	Yes
Lag Urban Indicator	Yes	Yes	Yes	Yes
x Lag Migrant Inflow/Total US Population	Yes	Yes	Yes	Yes
Lag Urban Indicator	Yes	Yes	Yes	Yes
Log County Population Density	Yes	Yes	Yes	Yes
County Fixed Effects	Yes	Yes	Yes	Yes
Decade Fixed Effects	Yes	Yes	Yes	Yes
Observations	16,729	15,706	11,591	10,568
R-squared	0.895	0.899	0.885	0.893
Mean of Dependent Variable	0.087	0.084	0.115	0.113

Notes: OLS estimates are reported. An observation is a county in a time period (1860, 1870, 1880, 1890, 1900, 1910 or 1920). The dependent variable "Migrant Share of Total County Population" is the proportion of a county's population that is foreign born in period *t*. "Lag Rail Access" is an indicator variable that equals one if a county has a railway in period *t-1*. Conley standard errors are reported in square brackets. ***, **, and * indicate significance at the 1, 5 and 10% levels.

Table A4: OLS and 2SLS estimates of the impacts of historical immigration on the health of the economy today, omitting the U.S. Northeast and South.

	(1)	(2)	(3)	(4)	(5)
Dependent Variable	**Log Average per Capita Income, 2000**	**Prop of Population Below Poverty Line, 2000**	**Unemployment Rate, 2000**	**Urbanization Share, 2000**	**Average Years of Schooling, 2000**
			A. OLS Estimates		
Average Migrant Share,	0.189**	-0.002	0.016*	0.730***	-0.087
1860-1920	(0.076)	(0.017)	(0.009)	(0.093)	(0.191)
			B. 2SLS Estimates		
Average Migrant Share,	3.544**	-0.530*	-0.428*	5.454**	10.298**
1860-1920	(1.603)	(0.314)	(0.234)	(2.446)	(4.598)
			C. First Stage Estimates		
			Dependent Variable: Average Migrant Share, 1860-1920		
Predicted Avg. Migrant Share,	4.895***	4.895***	4.895***	4.895***	4.895***
1860-1920	[1.890]	[1.890]	[1.890]	[1.890]	[1.890]
Kleibergen Paap *F*-statistic	6.71	6.71	6.71	6.71	6.71
Controls (in all Panels):					
Industrialization-Based Predicted Migrant Share	Yes	Yes	Yes	Yes	Yes
Date of RR Connection (Years as of 2000)	Yes	Yes	Yes	Yes	Yes
Latitude	Yes	Yes	Yes	Yes	Yes
Longitude	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	1,814	1,814	1,814	1,814	1,814
Mean of Dep. Var. (2nd-Stage and OLS)	10.03	0.213	0.045	0.400	11.52

Notes : An observation is a county. Panels A and B reports OLS estimates and 2SLS estimates, respectively. Panel C reports the first-stage estimates from the 2SLS. Coefficient estimates are reported, with Conley standard errors reported in square brackets. ***, **, and * indicate significance at the 1, 5 and 10% levels.

Table A5: OLS and 2SLS estimates of the impacts of historical immigration on the health of the economy today, omitting the U.S. West and Midwest.

	(1)	(2)	(3)	(4)	(5)
Dependent Variable	**Log Average per Capita Income, 2000**	**Prop of Population Below Poverty Line, 2000**	**Unemployment Rate, 2000**	**Urbanization Share, 2000**	**Average Years of Schooling, 2000**
			A. OLS Estimates		
Average Migrant Share,	0.214	0.063	0.090**	1.627***	-0.637
1860-1920	(0.241)	(0.044)	(0.043)	(0.184)	(0.626)
			B. 2SLS Estimates		
Average Migrant Share,	4.006**	-0.543	-0.637**	4.596**	9.543**
1860-1920	(1.664)	(0.407)	(0.283)	(2.227)	(4.759)
			C. First Stage Estimates		
			Dependent Variable: Average Migrant Share, 1860-1920		
Predicted Avg. Migrant Share,	5.140***	5.140***	5.140***	5.140***	5.140***
1860-1920	[1.570]	[1.570]	[1.570]	[1.570]	[1.570]
Kleibergen Paap *F*-statistic	10.72	10.72	10.72	10.72	10.72
Controls (in all Panels):					
Industrialization-Based Predicted Migrant Share	Yes	Yes	Yes	Yes	Yes
Date of RR Connection (Years as of 2000)	Yes	Yes	Yes	Yes	Yes
Latitude	Yes	Yes	Yes	Yes	Yes
Longitude	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	1,121	1,121	1,121	1,121	1,121
Mean of Dep. Var. (2nd-Stage and OLS)	10.01	0.150	0.05	0.403	11.33

Notes : An observation is a county. Panels A and B reports OLS estimates and 2SLS estimates, respectively. Panel C reports the first-stage estimates from the 2SLS. Coefficient estimates are reported, with Conley standard errors reported in square brackets. ***, **, and * indicate significance at the 1, 5 and 10% levels.

Table A6: Zero-stage OLS panel estimates using predicted migrant flows based on home-country weather shocks.

Dependent Variable	(1)	(2)	(3)	(4)
	Migrant Share of Total County Population			
	All Counties	Excluding Northeast	Excluding South	Midwest and West
Interaction of Interest:				
Lag Rail Access	0.263***	0.284***	0.318***	0.362***
x Lag Predicted Migrant Inflow/Total US Population	[0.046]	[0.047]	[0.080]	[0.086]
Other Variables:				
Lag Rail Access	-0.002	0.006	-0.019	-0.004
	[0.008]	[0.008]	[0.014]	[0.015]
Lag Rail Access	-0.005*	-0.008***	0.000	-0.005
x Lag Log Industrialization Index	[0.003]	[0.004]	[0.006]	[0.006]
Lag Rail Access	Yes	Yes	Yes	Yes
Lag Migrant Share	Yes	Yes	Yes	Yes
Lag Urban Indicator	Yes	Yes	Yes	Yes
Lag Urban Indicator	Yes	Yes	Yes	Yes
x Lag Predicted Migrant Inflow/Total US Population	Yes	Yes	Yes	Yes
Log County Population Density	Yes	Yes	Yes	Yes
County Fixed Effects	Yes	Yes	Yes	Yes
Decade Fixed Effects	Yes	Yes	Yes	Yes
Observations	16,729	15,706	11,591	10,568
R-squared	0.927	0.927	0.917	0.919
Mean of Dependent Variable	0.087	0.084	0.115	0.113

Notes : OLS estimates are reported. An observation is a county in a time period (1860, 1870, 1880, 1890, 1900, 1910 or 1920). The dependent variable "Migrant Share of Total County Population" is the proportion of a county's population that is forein born in period *t* . "Lag Rail Access" is an indicator variable that equals one if a county has a railway in period *t* -1. Conley standard errors are reported in parentheses. ***, **, and * indicate significance at the 1, 5 and 10% levels.

Table A7: OLS panel estimates of the impact of immigrants on railroad placement.

Dependent Variable	(1)	(2)	(3)	(4)
	Dependent Variable: Indicator for access to the railway			
	All Counties	Excluding Northeast	Excluding South	Midwest West
Lag Migrant Share in County	-0.025	0.051	-0.070	-0.012
	[0.068]	[0.071]	[0.070]	[0.075]
Lag Rail Access	Yes	Yes	Yes	Yes
Log Population Density	Yes	Yes	Yes	Yes
Lag Urban Indicator	Yes	Yes	Yes	Yes
Decade Fixed Effects	Yes	Yes	Yes	Yes
County Fixed Effects	Yes	Yes	Yes	Yes
Observations	16,728	15,705	11,591	10,568
R-squared	0.669	0.679	0.683	0.700
Mean of Dependent Variable	0.087	0.084	0.115	0.113

Notes : OLS estimates are reported. An observation is a county in a time period (1860, 1870, 1880, 1890, 1900, 1910 or 1920). The independent variable "Lag Migrant Share of Total County Population" is the proportion of a county's population that is forein born in period *t-1* . "Lag Rail Access" is an indicator variable that equals one if a county has a railway in period *t* -10. Conley standard errors are reported in square brackets. ***, **, and * indicate significance at the 1, 5 and 10% levels.

Table A8: OLS and 2SLS estimates of the impacts of historical immigration on the health of the economy today, with an instrument that accounts for decades of connection to rail.

	(1)	(2)	(3)	(4)	(5)
Dependent Variable	**Log Average per Capita Income, 2000**	**Prop of Population Below Poverty Line, 2000**	**Unemployment Rate, 2000**	**Urbanization Share, 2000**	**Average Years of Schooling, 2000**
			A. OLS Estimates		
Average Migrant Share,	0.197**	0.011	0.035***	0.970***	-0.165
1860-1920	[0.079]	[0.016]	[0.013]	[0.081]	[0.206]
			B. 2SLS Estimates		
Average Migrant Share,	6.120***	-1.130***	-0.661***	9.867***	18.743***
1860-1920	[1.927]	[0.388]	[0.241]	[3.096]	[5.976]
			C. First Stage Estimates		
			Dependent Variable: Average Migrant Share, 1860-1920		
Predicted Avg. Migrant Share,	4.889***	4.889***	4.889***	4.889***	4.889***
1860-1920	[1.453]	[1.453]	[1.453]	[1.453]	[1.453]
Kleibergen Paap *F*-statistic	11.33	11.33	11.33	11.33	11.33
Controls (in all Panels):					
Industrialization-Based Predicted Migrant Share	Yes	Yes	Yes	Yes	Yes
Never Connected to the Rail [0-1]	Yes	Yes	Yes	Yes	Yes
Latitude	Yes	Yes	Yes	Yes	Yes
Longitude	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	2,935	2,935	2,935	2,935	2,935
Mean of Dep. Var. (2nd-Stage and OLS)	10.02	0.136	0.047	0.401	11.45

Notes : An observation is a county. Panels A and B reports OLS estimates and 2SLS estimates, respectively. Panel C reports the first-stage estimates from the 2SLS. Coefficient estimates are reported, with Conley standard errors reported in square brackets. ***, **, and * indicate significance at the 1, 5 and 10% levels.

Table A9: OLS and 2SLS estimates of the impacts of historical immigration on social outcomes today, with an instrument that accounts for decades of connection to rail.

Dependent Variable	(1) **Social Capital, 2000**	(2) **Voting Turnout, 2000**	(3) **Total Crime Rate, 2000**	(4) **Crimes Against Persons, 2000**	(5) **Crimes Against Property, 2000**
			A. OLS Estimates		
Average Migrant Share,	-1.306***	-0.080***	0.006***	0.001***	0.003***
1860-1920	(0.348)	(0.027)	(0.001)	(0.000)	(0.001)
			B. 2SLS Estimates		
Average Migrant Share,	-0.153	-0.023	0.041**	0.004	0.036**
1860-1920	(4.672)	(0.365)	(0.021)	(0.004)	(0.015)
			C. First Stage Estimates		
			Dependent Variable: Average Migrant Share, 1860-1920		
Predicted Avg. Migrant Share,	4.889***	4.889***	4.889***	4.889***	4.889***
1860-1920	[1.453]	[1.453]	[1.453]	[1.453]	[1.453]
Kleibergen Paap *F* -statistic	11.326	11.326	11.326	11.326	11.326
Controls (in all Panels):					
Industrialization-Based Predicted Migrant Share	Yes	Yes	Yes	Yes	Yes
Never Connected to the Rail [0-1]	Yes	Yes	Yes	Yes	Yes
Latitude	Yes	Yes	Yes	Yes	Yes
Longitude	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	2,934	2,925	2,935	2,935	2,935
Mean of Dep. Var. (2nd-Stage and OLS)	-0.004	0.540	0.006	0.001	0.004

Notes : An observation is a county. Panels A and B reports OLS estimates and 2SLS estimates, respectively. Panel C reports the first-stage estimates from the 2SLS. Coefficient estimates are reported, with Conley standard errors reported in square brackets. ***, **, and * indicate significance at the 1, 5 and 10% levels.

Table A10: OLS and 2SLS estimates of the impacts of historical immigration on the health of the economy today, using only counties that existed in 1860 and whose borders have remained stable from 1860 to 2000.

Dependent Variable	(1) **Log Average per Capita Income, 2000**	(2) **Prop of Population Below Poverty Line, 2000**	(3) **Unemployment Rate, 2000**	(4) **Urbanization Share, 2000**	(5) **Average Years of Schooling, 2000**
			A. OLS Estimates		
Average Migrant Share,	0.629***	-0.084***	-0.000	1.034***	1.052***
1860-1920	[0.097]	[0.019]	[0.014]	[0.130]	[0.219]
			B. 2SLS Estimates		
Average Migrant Share,	8.907**	-1.498*	-1.163*	12.902*	28.045*
1860-1920	[4.321]	[0.807]	[0.707]	[6.243]	[15.647]
			C. First Stage Estimates		
			Dependent Variable: Average Migrant Share, 1860-1920		
Predicted Avg. Migrant Share,	3.712**	3.712**	3.712**	3.712**	3.712**
1860-1920	[1.815]	[1.815]	[1.815]	[1.815]	[1.815]
Kleibergen Paap *F*-statistic	4.180	4.180	4.180	4.180	4.180
Controls (in all Panels):					
Industrialization-Based Predicted Migrant Share	Yes	Yes	Yes	Yes	Yes
Date of RR Connection (years as of 2000)	Yes	Yes	Yes	Yes	Yes
Latitude	Yes	Yes	Yes	Yes	Yes
Longitude	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	1,596	1,596	1,596	1,596	1,596
Mean of Dep. Var. (2nd-Stage and OLS)	10.020	0.138	0.049	0.405	11.430

Notes : An observation is a county. Panels A and B reports OLS estimates and 2SLS estimates, respectively. Panel C reports the first-stage estimates from the 2SLS. Coefficient estimates are reported, with Conley standard errors reported in square brackets. ***, **, and * indicate significance at the 1, 5 and 10% levels.

Table A11: OLS and 2SLS estimates of the impacts of historical immigration on social outcomes today, using only counties that existed in 1860 and whose borders have remained stable from 1860 to 2000.

Dependent Variable	(1) **Social Capital, 2000**	(2) **Voting Turnout, 2000**	(3) **Total Crime Rate, 2000**	(4) **Crimes Against Persons, 2000**	(5) **Crimes Against Property, 2000**
			A. OLS Estimates		
Average Migrant Share,	0.448	0.008	0.009***	0.002***	0.006***
1860-1920	(0.464)	(0.034)	(0.001)	(0.000)	(0.001)
			B. 2SLS Estimates		
Average Migrant Share,	-2.340	0.564	0.058	0.014	0.034
1860-1920	(8.502)	(0.872)	(0.048)	(0.011)	(0.032)
			C. First Stage Estimates		
			Dependent Variable: Average Migrant Share, 1860-1920		
Predicted Avg. Migrant Share,	3.712**	3.712**	3.712**	3.712**	3.712**
1860-1920	[1.815]	[1.815]	[1.815]	[1.815]	[1.815]
Kleibergen Paap *F*-statistic	4.18	4.18	4.18	4.18	4.18
Controls (in all Panels):					
Industrialization-Based Predicted Migrant Share	Yes	Yes	Yes	Yes	Yes
Date of RR Connection (Years as of 2000)	Yes	Yes	Yes	Yes	Yes
Latitude	Yes	Yes	Yes	Yes	Yes
Longitude	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	1596	1596	1596	1596	1596
Mean of Dep. Var. (2nd-Stage and OLS)	-0.004	0.540	0.006	0.001	0.004

Notes : An observation is a county. Panels A and B reports OLS estimates and 2SLS estimates, respectively. Panel C reports the first-stage estimates from the 2SLS. Coefficient estimates are reported, with Conley standard errors reported in square brackets. ***, **, and * indicate significance at the 1, 5 and 10% levels.

Table A12: OLS and 2SLS estimates, accounting for spillovers to contiguous counties.

Dependent Variable	(1) Log Average per Capita Income, 2000	(2) Prop of Population Below Poverty Line, 2000	(3) Unemployment Rate, 2000	(4) Urbanization Share, 2000	(5) Average years of Schooling, 2000
			A. OLS Estimates		
Average Migrant Share, 1860-1920	0.109	0.044*	0.011	1.098***	-0.120
	[0.101]	[0.023]	[0.017]	[0.126]	[0.256]
Average Migrant Share in Contiguous Counties, 1860-1920	0.137	-0.052	0.045**	-0.284	-0.095
	[0.137]	[0.035]	[0.019]	[0.174]	[0.327]
			B. 2SLS Estimates		
Average Migrant Share, 1860-1920	4.425	-0.605	-0.692	7.015**	13.363*
	[3.229]	[0.573]	[0.424]	[3.365]	[7.660]
Average Migrant Share in Contiguous Counties, 1860-1920	5.982	-1.146*	-0.571	3.020	10.280
	[3.872]	[0.687]	[0.533]	[4.216]	[9.616]
			C. First Stage Estimates		
			Dep. Var.: Avg Migrant Share in County, 1860-1920		
Predicted Average Migrant Share, 1860-1920	3.879***	3.879***	3.879***	3.879***	3.879***
	[1.475]	[1.475]	[1.475]	[1.475]	[1.475]
Predicted Avg Migrant Share in Contiguous Counties, 1860-1920	-0.313	-0.313	-0.313	-0.313	-0.313
	[2.997]	[2.997]	[2.997]	[2.997]	[2.997]
Angrist-Pischke *F*-statistic	7.81	7.81	7.81	7.81	7.81
			Dep. Var.: Avgerage Migrant Share in Contiguous Counties, 1860-1920		
Predicted Average Migrant Share, 1860-1920	-1.114	-1.114	-1.114	-1.114	-1.114
	[1.193]	[1.193]	[1.193]	[1.193]	[1.193]
Predicted Avg Migrant Share in Contiguous Counties, 1860-1920	6.346***	6.346***	6.346***	6.346***	6.346***
	[2.014]	[2.014]	[2.014]	[2.014]	[2.014]
Angrist-Pischke *F*-statistic	9.89	9.89	9.89	9.89	9.89
Controls (in all Panels):					
Industrialization-Based Predicted Migrant Share	Yes	Yes	Yes	Yes	Yes
Date of RR Connection (years as of 2000)	Yes	Yes	Yes	Yes	Yes
Latitude	Yes	Yes	Yes	Yes	Yes
Longitude	Yes	Yes	Yes	Yes	Yes
Contiguous Counties: Avg. Date of RR Connection	Yes	Yes	Yes	Yes	Yes
Contiguous Counties: Average Latitude	Yes	Yes	Yes	Yes	Yes
Contiguous Counties: Average Longitude	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	2,931	2,931	2,931	2,931	2,931
Mean of Dep. Var. (2nd-Stage and OLS)	10.02	0.136	0.047	0.401	11.44

Notes: An observation is a county. Panels A and B reports OLS estimates and 2SLS estimates, respectively. Panels C reports the first-stage estimates from the 2SLS. Weighted Average Migrant Share in Contiguous Counties corresponds to the share of migrants in contiguous counties weighted by the length of the shared border with the county. Coefficient estimates are reported, with Conley standard errors in square brackets. ***, **, and * indicate significance at the 1, 5 and 10% levels.

Table A13: OLS and 2SLS estimates, accounting for state-level spatial spillovers.

Dependent Variable	(1) Log Average per Capita Income, 2000	(2) Prop of Population Below Poverty Line, 2000	(3) Unemployment Rate, 2000	(4) Urbanization Share, 2000	(5) Average years of Schooling, 2000
			A. OLS Estimates		
Average Migrant Share, 1860-1920	0.140	0.017	0.044***	0.897***	-0.273
	[0.096]	[0.019]	[0.017]	[0.093]	[0.260]
Average Migrant Share in other Counties	-3.448	0.397	0.635*	-2.850	-6.110
in the Same State, 1860-1920	[2.644]	[0.541]	[0.371]	[2.614]	[7.782]
			B. 2SLS Estimates		
Average Migrant Share, 1860-1920	4.375***	-0.589*	-0.639**	6.615***	12.954***
	[1.549]	[0.305]	[0.249]	[2.349]	[4.622]
Average Migrant Share in other Counties	19.293	0.113	-1.531	27.635	43.922
in the Same State, 1860-1920	[22.543]	[3.575]	[3.077]	[33.074]	[65.399]
			C. First Stage Estimates		
			Dep. Var.: Avg Migrant Share in County, 1860-1920		
Predicted Average Migrant Share, 1860-1920	2.759	2.759	2.759	2.759	2.759
	[2.068]	[2.068]	[2.068]	[2.068]	[2.068]
Predicted Average Migrant Share in other Counties	-95.214	-95.214	-95.214	-95.214	-95.214
in the Same State, 1860-1920	[106.9]	[106.9]	[106.9]	[106.9]	[106.9]
Angrist-Pischke *F* -statistic	9.72	9.72	9.72	9.72	9.72
			Dep. Var.: Average Migrant Share in All Other Counties, 1860-1920		
Predicted Average Migrant Share,	0.402**	0.402**	0.402**	0.402**	0.402**
1860-1920	[0.160]	[0.160]	[0.160]	[0.160]	[0.160]
Predicted Average Migrant Share in other Counties	27.982***	27.982***	27.982***	27.982***	27.982***
in the Same State, 1860-1920	[9.809]	[9.809]	[9.809]	[9.809]	[9.809]
Angrist-Pischke *F* -statistic	7.53	7.53	7.53	7.53	7.53
Controls (in all Panels):					
Industrialization-Based Predicted Migrant Share	Yes	Yes	Yes	Yes	Yes
Date of RR Connection (years as of 2000)	Yes	Yes	Yes	Yes	Yes
Latitude	Yes	Yes	Yes	Yes	Yes
Longitude	Yes	Yes	Yes	Yes	Yes
All Other State Counties: Avg. Date of RR Connect.	Yes	Yes	Yes	Yes	Yes
All Other State Counties: Average Latitude	Yes	Yes	Yes	Yes	Yes
All Other State Counties: Average Longitude	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	2,934	2,934	2,934	2,934	2,934
Mean of Dep. Var. (2nd-Stage and OLS)	10.02	0.136	0.047	0.401	11.44

Notes: An observation is a county. Panels A and B reports OLS estimates and 2SLS estimates, respectively. Panels C reports the first-stage estimates from the 2SLS. Weighted Average Migrant Share in Contiguous Counties corresponds to the share of migrants in contiguous counties weighted by the length of the shared border with the county. Coefficient estimates are reported, with Conley standard errors in square brackets. ***, **, and * indicate significance at the 1, 5 and 10% levels.

Table A14: OLS and 2SLS estimates, accounting for state-level spatial spillovers excluding contiguous counties.

	(1)	(2)	(3)	(4)	(5)
Dependent Variable	**Log Average per Capita Income, 2000**	**Prop of Population Below Poverty Line, 2000**	**Unemployment Rate, 2000**	**Urbanization Share, 2000**	**Average years of Schooling, 2000**
			A. OLS Estimates		
Average Migrant Share, 1860-1920	0.141	0.017	0.044***	0.899***	-0.270
	[0.096]	[0.019]	[0.017]	[0.093]	[0.260]
Average Migrant Share in Non-Contiguous	-3.404	0.388	0.634*	-2.712	-5.911
Counties in the same State, 1860-1920	[2.640]	[0.540]	[0.371]	[2.616]	[7.731]
			B. 2SLS Estimates		
Average Migrant Share, 1860-1920	4.265***	-0.572*	-0.623**	6.329***	12.634***
	[1.508]	[0.298]	[0.243]	[2.280]	[4.507]
Average Migrant Share in Non-Contiguous	11.804	1.370	-0.405	10.263	20.859
Counties in the same State, 1860-1920	[21.575]	[3.573]	[3.105]	[30.978]	[63.110]
			C. First Stage Estimates		
			Dep. Var.: Avg Migrant Share in County, 1860-1920		
Predicted Average Migrant Share, 1860-1920	3.433*	3.433*	3.433*	3.433*	3.433*
	[2.073]	[2.073]	[2.073]	[2.073]	[2.073]
Predicted Average Migrant Share in Non-Contiguous	-56.938	-56.938	-56.938	-56.938	-56.938
Counties in the same State, 1860-1920	[110.586]	[110.586]	[110.586]	[110.586]	[110.586]
Angrist-Pischke *F* -statistic	10.07	10.07	10.07	10.07	10.07
			Dep. Var.: Avgerage Migrant Share in Contiguous Counties, 1860-1920		
Predicted Average Migrant Share, 1860-1920	0.379***	0.379***	0.379***	0.379***	0.379***
	[0.161]	[0.161]	[0.161]	[0.161]	[0.161]
Predicted Average Migrant Share in Non-Contiguous	27.598***	27.598***	27.598***	27.598***	27.598***
Counties in the same State, 1860-1920	[10.280]	[10.280]	[10.280]	[10.280]	[10.280]
Angrist-Pischke *F* -statistic	10.07	10.07	10.07	10.07	10.07
Controls (in all Panels):					
Industrialization-Based Predicted Migrant Share	Yes	Yes	Yes	Yes	Yes
Date of RR Connection (years as of 2000)	Yes	Yes	Yes	Yes	Yes
Latitude	Yes	Yes	Yes	Yes	Yes
Longitude	Yes	Yes	Yes	Yes	Yes
Oth. State Counties (excl. contig.): Avg. Date of RR Connect.	Yes	Yes	Yes	Yes	Yes
Oth. State Counties (excl. contig.): Average Latitude	Yes	Yes	Yes	Yes	Yes
Oth. State Counties (excl. contig.): Average Longitude	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	2,934	2,934	2,934	2,934	2,934
Mean of Dep. Var. (2nd-Stage and OLS)	10.02	0.136	0.047	0.401	11.44

Notes: An observation is a county. Panels A and B reports OLS estimates and 2SLS estimates, respectively. Panels C reports the first-stage estimates from the 2SLS. Weighted Average Migrant Share in Contiguous Counties corresponds to the share of migrants in contiguous counties weighted by the length of the shared border with the county. Coefficient estimates are reported, with Conley standard errors in square brackets. ***, **, and * indicate significance at the 1, 5 and 10% levels.

Table A15: OLS and 2SLS estimates of the impact of historical immigration on subsequent immigration.

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Dependent Variable	**Share of Foreign Born in:**							
	1920	**1930**	**1940**	**1950**	**1970**	**1980**	**1990**	**2000**
	A. OLS Estimates							
Average Migrant Share,	0.645***	0.483***	0.311***	0.238***	0.129***	0.148***	0.154***	0.179***
1860-1920	[0.016]	[0.016]	[0.012]	[0.012]	[0.010]	[0.016]	[0.019]	[0.022]
	B. 2SLS Estimates							
Average Migrant Share,	0.353**	0.348**	0.174*	0.081	0.108*	-0.006	0.038	0.045
1860-1920	[0.158]	[0.139]	[0.096]	[0.098]	[0.065]	[0.123]	[0.137]	[0.190]
	C. First Stage Estimates							
	Dependent Variable: Average Migrant Share, 1860-1920							
Predicted Avg. Migrant Share,	4.423***	4.423***	4.423***	4.423***	4.423***	4.423***	4.423***	4.423***
1860-1920	[1.369]	[1.369]	[1.369]	[1.369]	[1.369]	[1.369]	[1.369]	[1.369]
Kleibergen Paap *F*-statistic	10.43	10.43	10.43	10.43	10.43	10.43	10.43	10.43
Controls (in all Panels):								
Industrialization-Based Predicted Migrant Share	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date of RR Connection (Years as of 2000)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Latitude	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Longitude	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,935	2,934	2,935	2,935	2,935	2,935	2,935	2,935
Mean of Dep. Var. (2nd-Stage and OLS)	0.067	0.054	0.037	0.028	0.015	0.020	0.021	0.033

Notes : An observation is a county. Panels A and B reports OLS estimates and 2SLS estimates, respectively. Panel C reports the first-stage estimates from the 2SLS. Coefficient estimates are reported, with Conley standard errors in square brackets. ***, **, and * indicate significance at the 1, 5 and 10% levels.

Table A16: OLS and 2SLS estimates of the impacts of historical immigration on the health of the economy today, controlling for contemporaneous immigration.

Dependent Variable	(1) **Log Average per Capita Income, 2000**	(2) **Prop of Population Below Poverty Line, 2000**	(3) **Unemployment Rate, 2000**	(4) **Urbanization Share, 2000**	(5) **Average Years of Schooling, 2000**
			A. OLS Estimates		
Average Migrant Share,	0.019	0.025	0.039***	0.445***	-0.311
1860-1920	[0.079]	[0.016]	[0.011]	[0.087]	[0.201]
			B. 2SLS Estimates		
Average Migrant Share,	4.133***	-0.611**	-0.620**	6.243***	12.572***
1860-1920	[1.519]	[0.299]	[0.246]	[2.280]	[4.458]
			C. First Stage Estimates		
			Dependent Variable: Average Migrant Share, 1860-1920		
Predicted Avg. Migrant Share,	4.322***	4.322***	4.322***	4.322***	4.322***
1860-1920	[1.388]	[1.388]	[1.388]	[1.388]	[1.388]
Kleibergen Paap *F*-statistic	9.69	9.69	9.69	9.69	9.69
Controls (in all Panels):					
Share of Foreign Born in 2000	Yes	Yes	Yes	Yes	Yes
Industrialization-Based Predicted Migrant Share	Yes	Yes	Yes	Yes	Yes
Date of RR Connection (Years as of 2000)	Yes	Yes	Yes	Yes	Yes
Latitude	Yes	Yes	Yes	Yes	Yes
Longitude	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	2,935	2,935	2,935	2,935	2,935
Mean of Dep. Var. (2nd-Stage and OLS)	10.02	0.136	0.047	0.401	11.45

Notes : An observation is a county. Panels A and B reports OLS estimates and 2SLS estimates, respectively. Panel C reports the first-stage estimates from the 2SLS. Coefficient estimates are reported, with Conley standard errors reported in square brackets. ***, **, and * indicate significance at the 1, 5 and 10% levels.

Table A17: OLS and 2SLS estimates of the impacts of historical immigration on human capital.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent Variable	**Median Years of Schooling**			**Average Years of Schooling**			
	1950	**1960**	**1970**	**1970**	**1980**	**1990**	**2000**
				A. OLS Estimates			
Average Migrant Share, 1860-1920	-1.751***	-1.365***	-1.158***	-0.741***	-0.402*	-0.335	-0.210
	[0.406]	[0.395]	[0.389]	[0.204]	[0.228]	[0.217]	[0.206]
Standardized 'beta' Coefficients	*-0.140*	*-0.104*	*-0.093*	*-0.086*	*-0.053*	*-0.055*	*-0.042*
				B. 2SLS Estimates			
Average Migrant Share, 1860-1920	7.39	9.641*	17.553**	11.489***	12.248***	13.380***	12.302***
	[5.340]	[5.808]	[7.212]	[4.950]	[4.750]	[4.756]	[4.345]
Standardized 'beta' Coefficients	*0.593*	*0.738*	*1.405*	*1.310*	*1.611*	*2.200*	*2.437*
				C. First Stage Estimates			
				Dependent Variable: Average Migrant Share, 1860-1920			
Predicted Avg. Migrant Share, 1860-1920	4.470***	4.463***	4.463***	4.423***	4.423***	4.423***	4.423***
	[1.387]	[1.370]	[1.370]	[1.357]	[1.357]	[1.357]	[1.357]
Kleibergen Paap *F*-statistic	10.38	10.67	10.67	10.431	10.431	10.431	10.431
Controls (in all Panels):							
Industrialization-Based Predicted Migrant Share	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date of RR Connection (years as of 2000)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Latitude	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Longitude	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,889	2,933	2,933	2,934	2,935	2,935	2,935
Mean of Dep. Var. (2nd-Stage and OLS)	5.29	9.27	10.89	9.19	10.18	10.98	11.45

Notes : An observation is a county. Panels A and B reports OLS estimates and 2SLS estimates, respectively. Panel C reports the first-stage estimates from the 2SLS. Coefficient estimates are reported, with Conley standard errors reported in square brackets. ***, **, and * indicate significance at the 1, 5 and 10% levels.

Table A18: OLS and 2SLS estimates of the impacts of historical immigration on income.

	(1)	(2)	(3)	(4)	(5)
Dependent Variable	**Log Average Income per Capita**				
	1960	**1970**	**1980**	**1990**	**2000**
	A. OLS Estimates				
Average Migrant Share,	0.268***	0.128*	0.136*	0.191**	0.183**
1860-1920	[0.082]	[0.068]	[0.069]	[0.077]	[0.080]
Standardized 'beta' Coefficients	*0.094*	*0.061*	*0.077*	*0.097*	*0.100*
	B. 2SLS Estimates				
Average Migrant Share,	2.507*	2.370**	2.593**	3.969***	4.080***
1860-1920	[1.350]	[1.080]	[1.050]	[1.440]	[1.463]
Standardized 'beta' Coefficients	*0.874*	*1.131*	*1.466*	*2.010*	*2.224*
	C. First Stage Estimates				
	Dependent Variable: Average Migrant Share, 1860-1920				
Predicted Avg. Migrant Share,	4.423***	4.423***	4.423***	4.423***	4.423***
1860-1920	[1.369]	[1.369]	[1.369]	[1.369]	[1.357]
Kleibergen Paap *F* -statistic	10.43	10.43	10.43	10.43	10.43
Controls (in all Panels):					
Industrialization-Based Predicted Migrant Share	Yes	Yes	Yes	Yes	Yes
Date of RR Connection (years as of 2000)	Yes	Yes	Yes	Yes	Yes
Latitude	Yes	Yes	Yes	Yes	Yes
Longitude	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	2,933	2,933	2,933	2,933	2,935
Mean of Dep. Var. (2nd-Stage and OLS)	8.53	8.91	9.21	9.29	10.02

Notes : An observation is a county. Panels A and B reports OLS estimates and 2SLS estimates, respectively. Panel C reports the first-stage estimates from the 2SLS. Coefficient estimates are reported, with Conley standard errors reported in square brackets. ***, **, and * indicate significance at the 1, 5 and 10% levels.

A3. Figures

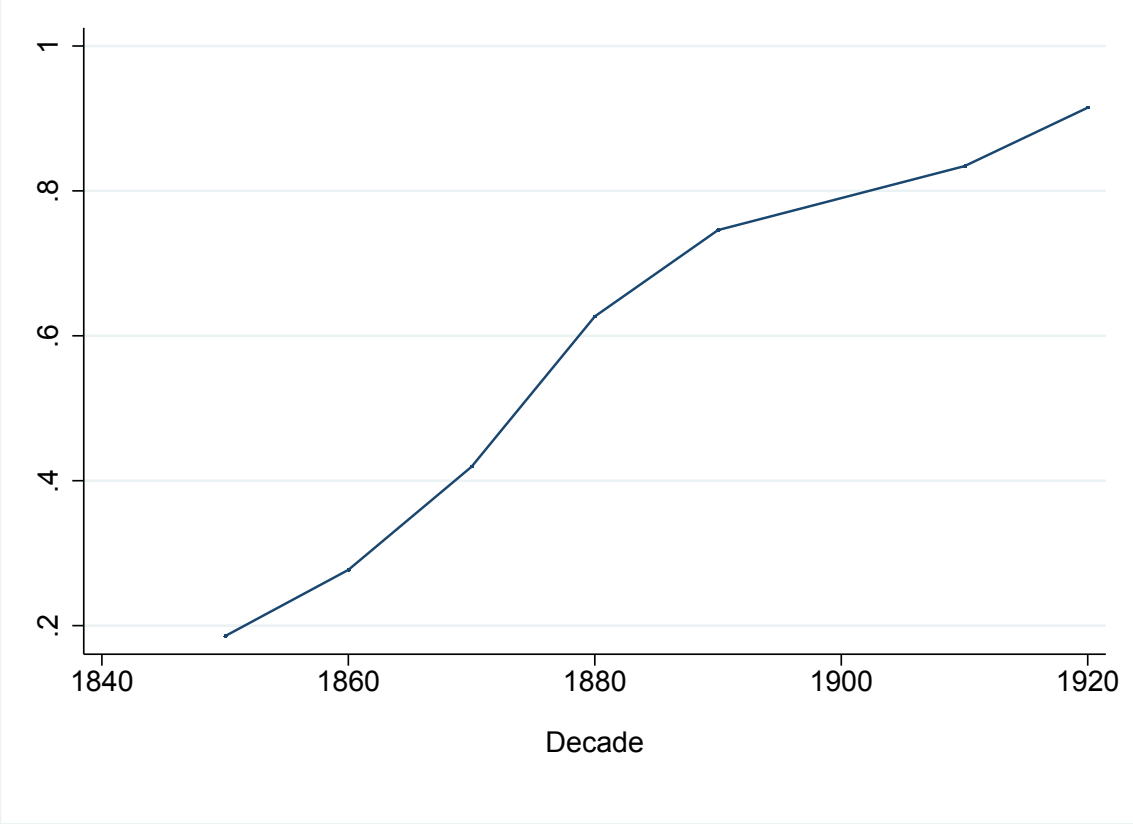
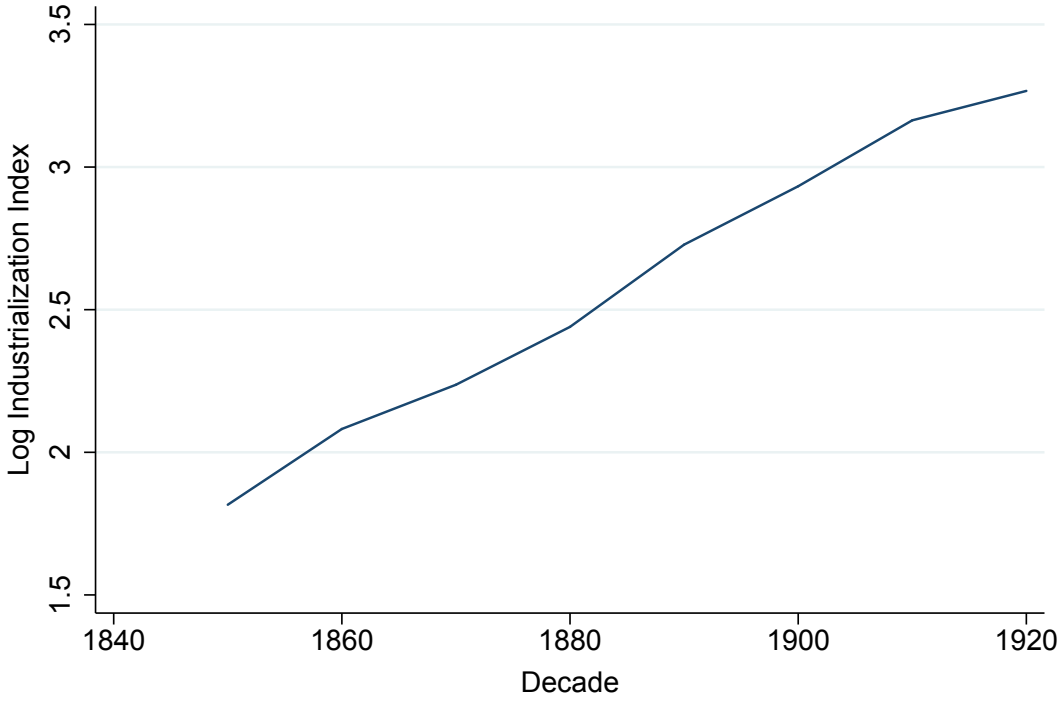


Figure A12: The expansion of the railway over time, 1850–1920.



Source: Davis (2004)

Figure A13: Index of industrial production, 1850–1920.

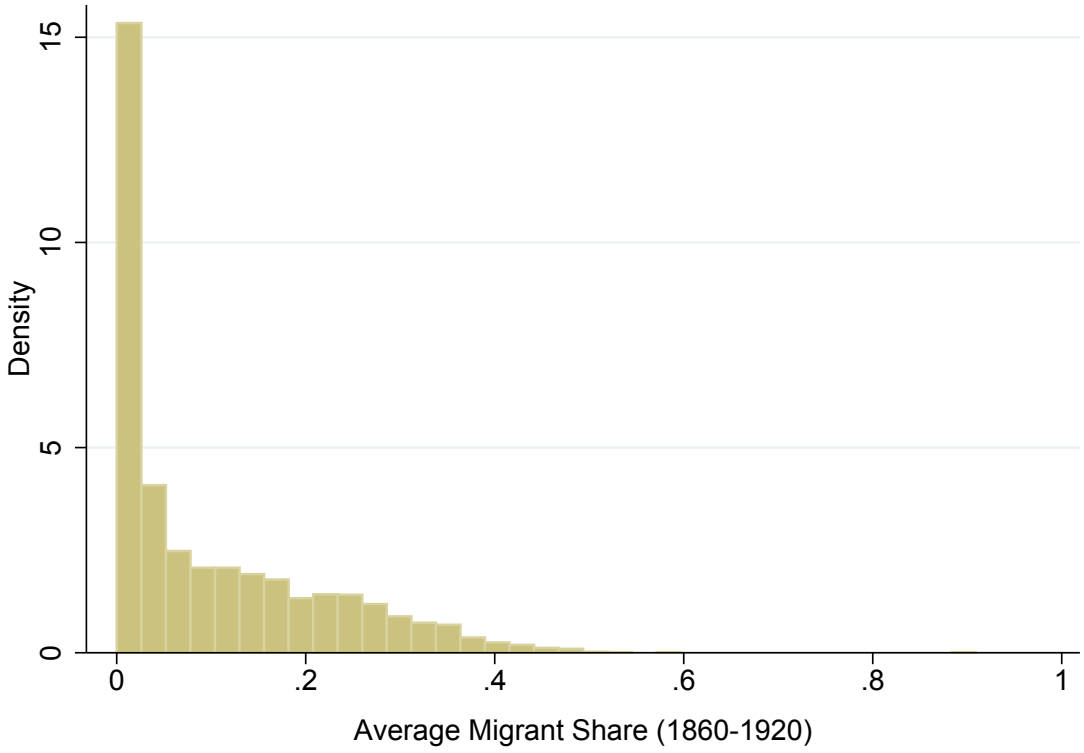


Figure A14: Histogram of the average migrant share, 1860–1920.

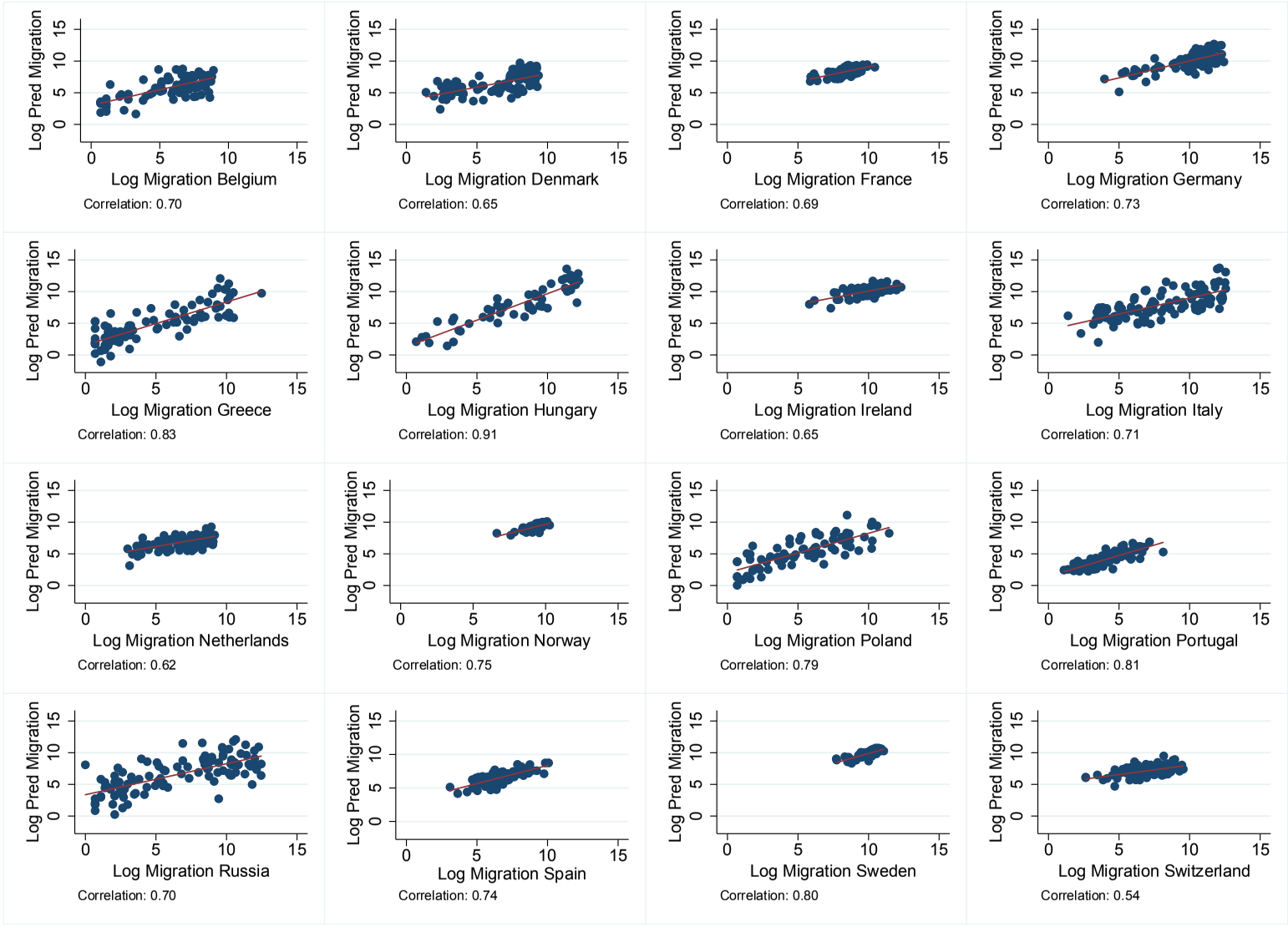


Figure A15: Actual versus predicted migration using weather shocks, by origin country

References

Davis, Joseph, "An Annual Index of U.S. Industrial Production, 1790–1915," *Quarterly Journal of Economics*, 1177–1215 2004, *119* (4), 642–658.

Luterbacher, Jurg, Daniel Dietrich, Elena Xoplaki, Martin Grosjean, and Heinz Wanner, "European Seasonal and Annual Temperature Variability, Trends, and Extremes Since 1500," *Science*, 2004, *303* (5663), 1499–1503.

Pauling, Andreas, Jurg Luterbacher, Carlo Casty, and Heinz Wanner, "Five Hundred Years of Gridded High-Resolution Precipitation Reconstructions over Europe and the Connection to Large-Scale Circulation," *Climate Dynamics*, 2006, *26*, 387–405.

Ramankutty, Navin and Jonathan A. Foley, "Estimating Historical Changes in Global Land Cover: Croplands from 1700 to 1992," *Global Biogeochemical Cycles*, December 1999, *13* (4), 997–1027.

Rupasingha, Anil and Stephan J. Goetz, "U.S. County-Level Social Capital Data, 1990–2005," 2008. The Northeast Regional Center for Rural Development, Penn State University, University Park, PA.

Willcox, Walter, *International Migrations*, Cambridge: National Bureau of Economic Research, 1929-1931.