Online Appendix for

Like an Ink Blot on Paper
Testing the Diffusion Hypothesis of Mass Migration, Italy 1876–1920

Yannay Spitzer
The Hebrew University of Jerusalem & CEPR

Ariell Zimran
Vanderbilt University & NBER

February 12, 2024

B Additional Tables and Figures

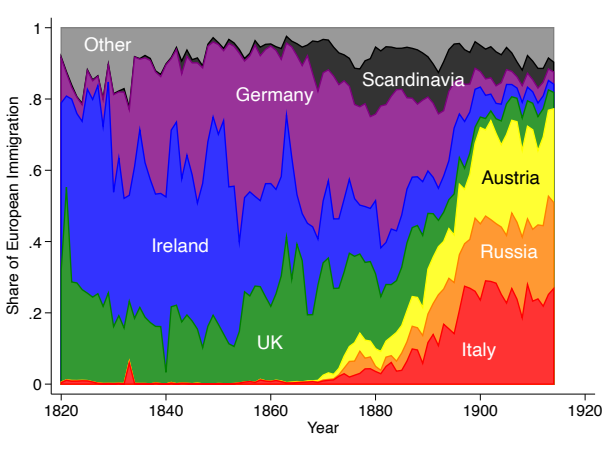


Figure B.1: Distribution of origin countries for US immigration from Europe

Source: Barde, Carter, and Sutch (2006)

Note: This graph shows the share of European immigration to the United States coming from each source country. The "Austria" data are from Barde, Carter, and Sutch's (2006) data for "Other Central Europe," which cover Central Europe other than Germany and Poland.

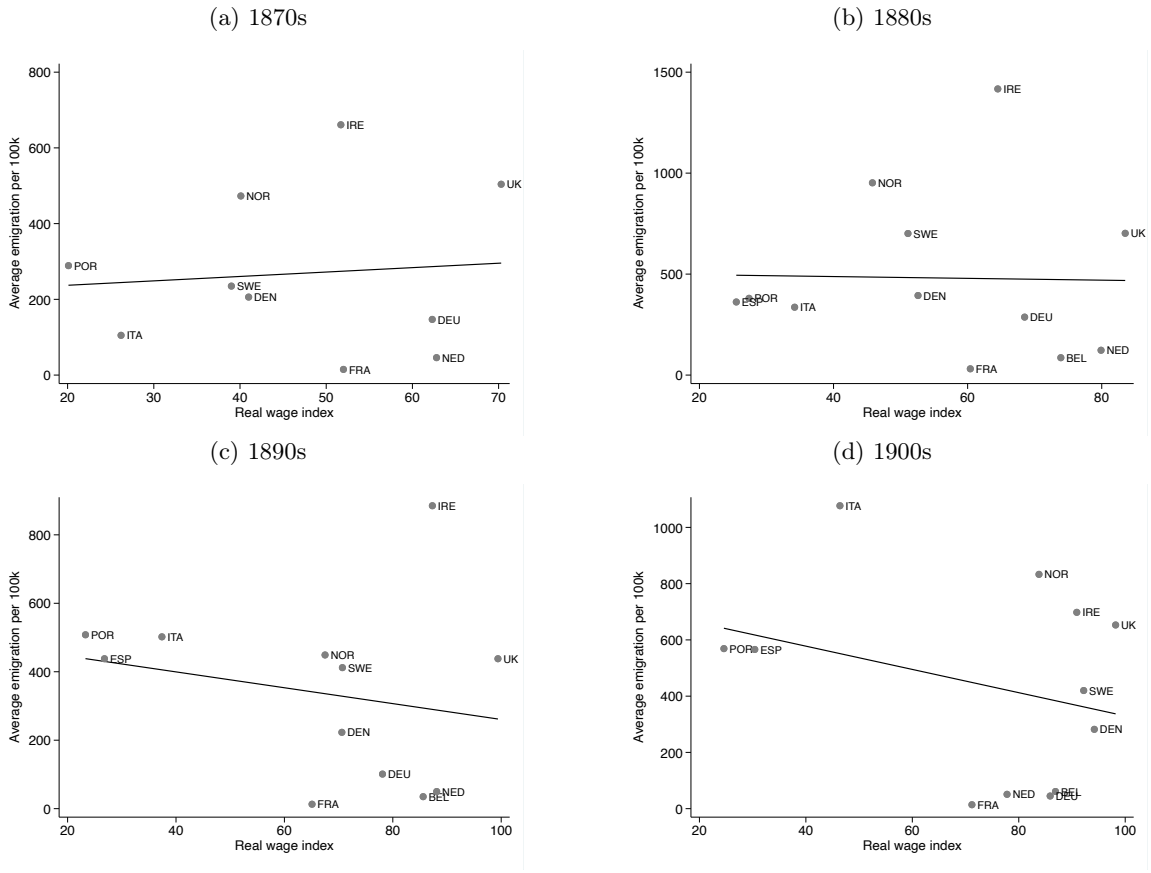


Figure B.2: Emigration and real wages

Source: Emigration data are from Ferenczi and Willcox (1929). Wage data are from Hatton and Williamson (1998).

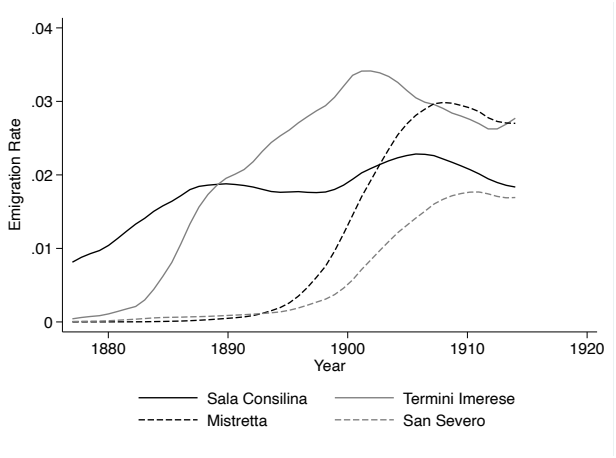


Figure B.3: Time series of emigration to North America from selected districts

Note: The four districts in this figure are Sala Consilina in Salerno, Termini Imerese in Palermo, Mistretta in Messina, and San Severo in Foggia. The epicenter district of Sala Consilina was selected because it had the highest emigration rate to North America in the period 1876–1883. The remaining three districts were selected because their estimated pre-1884 emigration rates as implied by their observables were among the most similar to that predicted for Sala Consilina. Some discretion was exercised in the choice of these example districts for purposes of exposition. The time series are smoothed using a local linear regression. The main takeaway in this figure is that, even though the four districts were observationally very similar, they experienced very different time series of emigration, surging into S-shapes in order of their distance from the nearest epicenter (not necessarily Sala Consilina).

(a) Regions (b) Provinces

Figure B.4: Maps of Italian provinces and regions

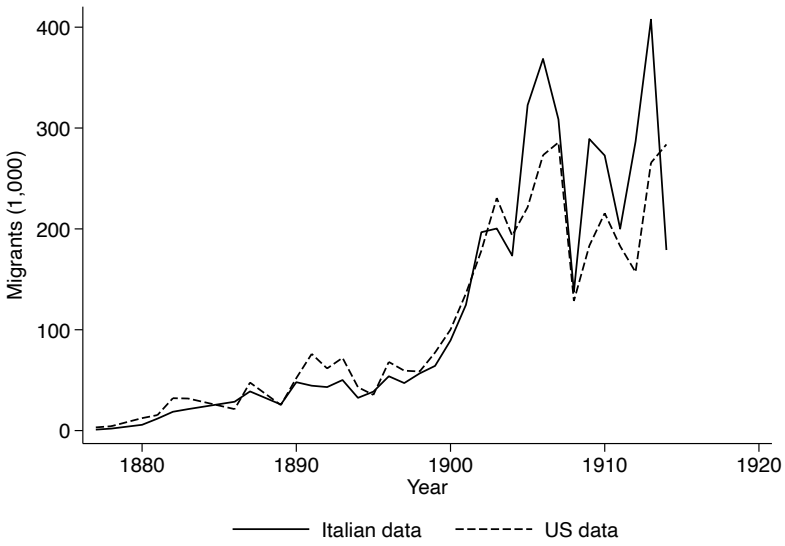


Figure B.5: Comparison of Italian emigration data and US immigration data

Note: The Italian data are for North America-bound emigrants from our transcriptions of the *Statistica della Emigrazione Italiana per l'Estero* and are based on calendar years. The US data are for immigrants arriving from Italy from Barde, Carter, and Sutch (2006) and are based on fiscal years.

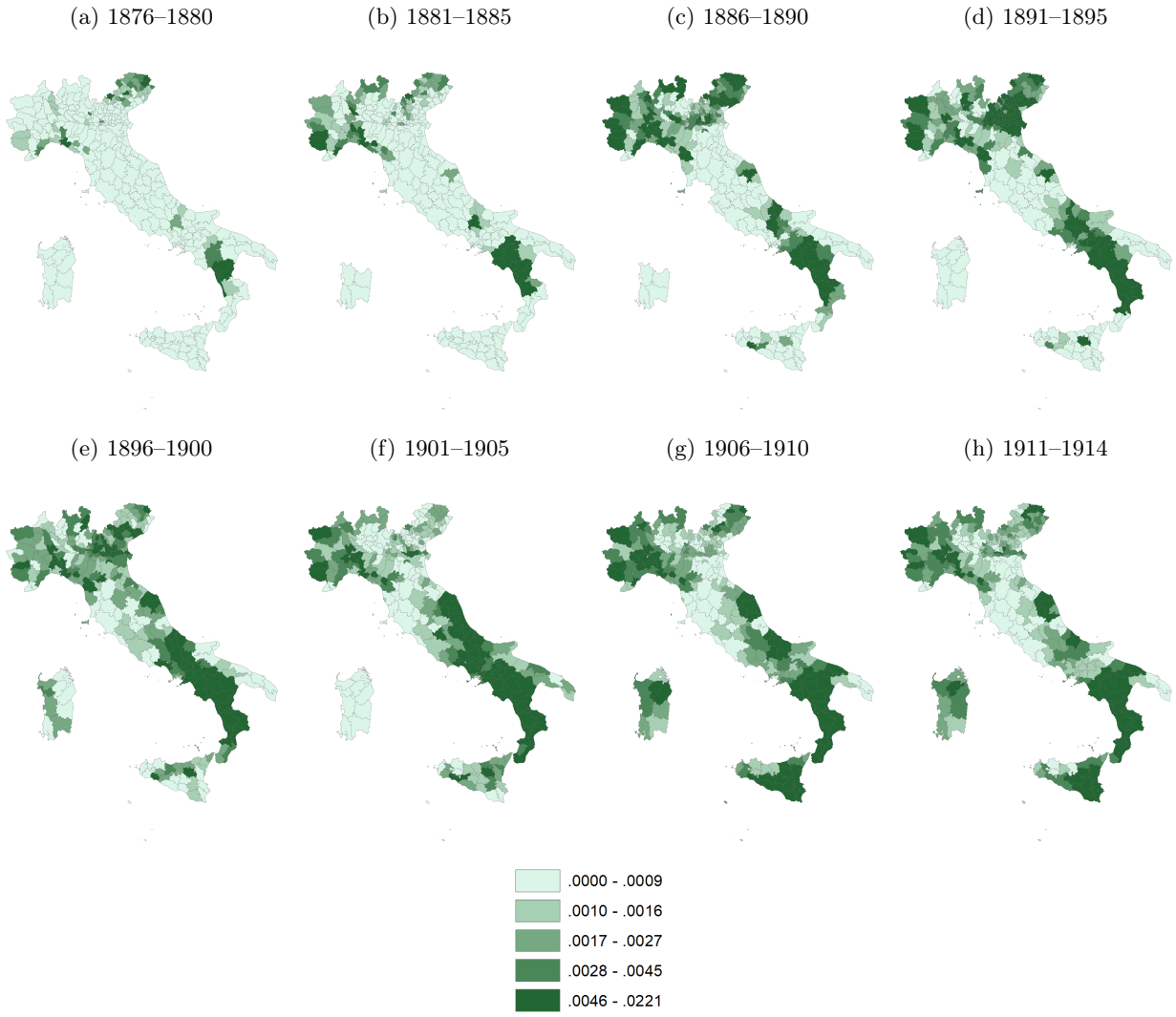


Figure B.6: District-level emigration rates to South America

Note: Each panel presents a district's average annual emigration rate to South America in the period in question. Scale is based on quintiles of emigration rates in 1911–1914.

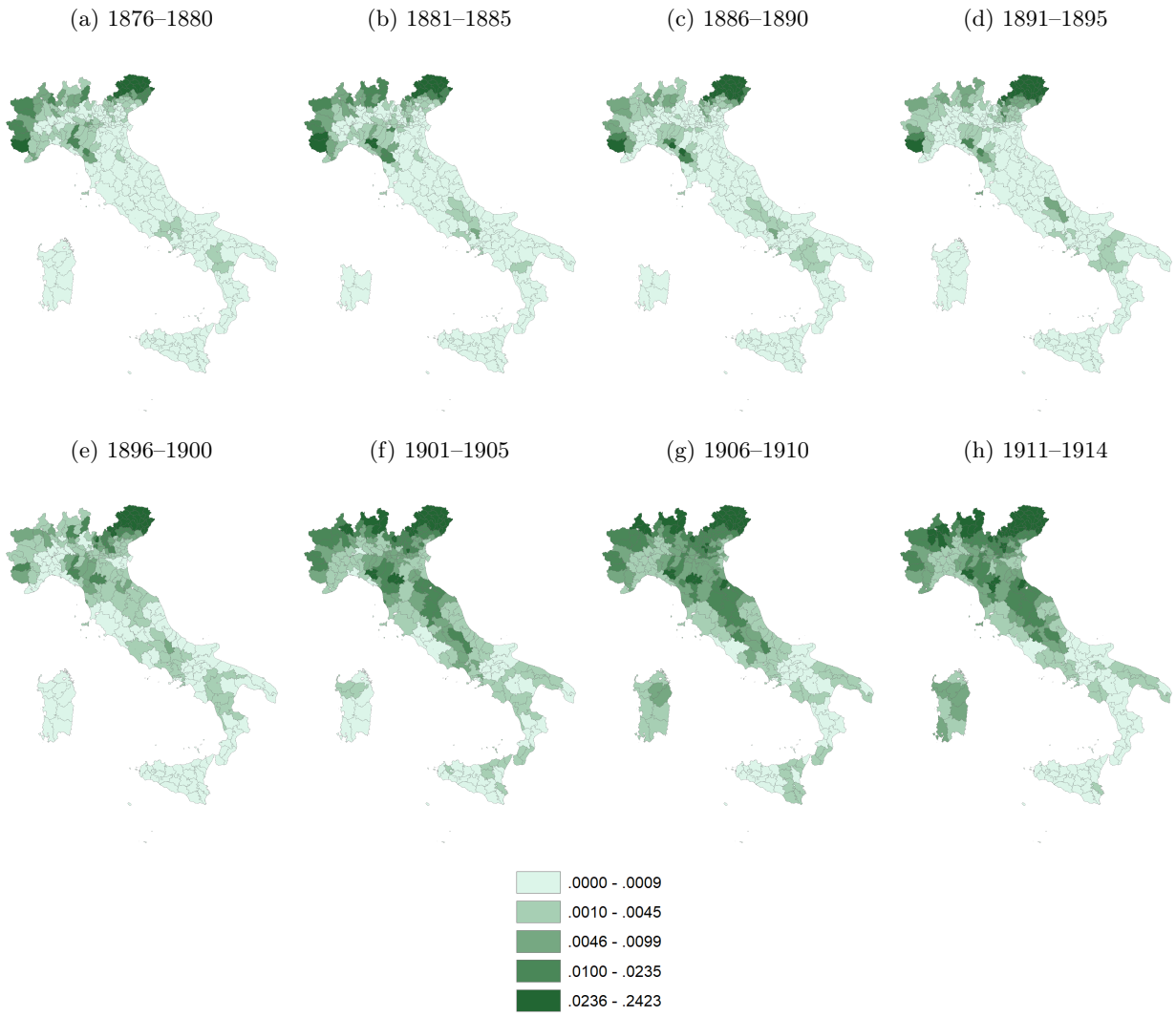


Figure B.7: District-level emigration rates to Europe

Note: Each panel presents a district's average annual emigration rate to Europe in the period in question. Scale is based on quintiles of emigration rates in 1911–1914.

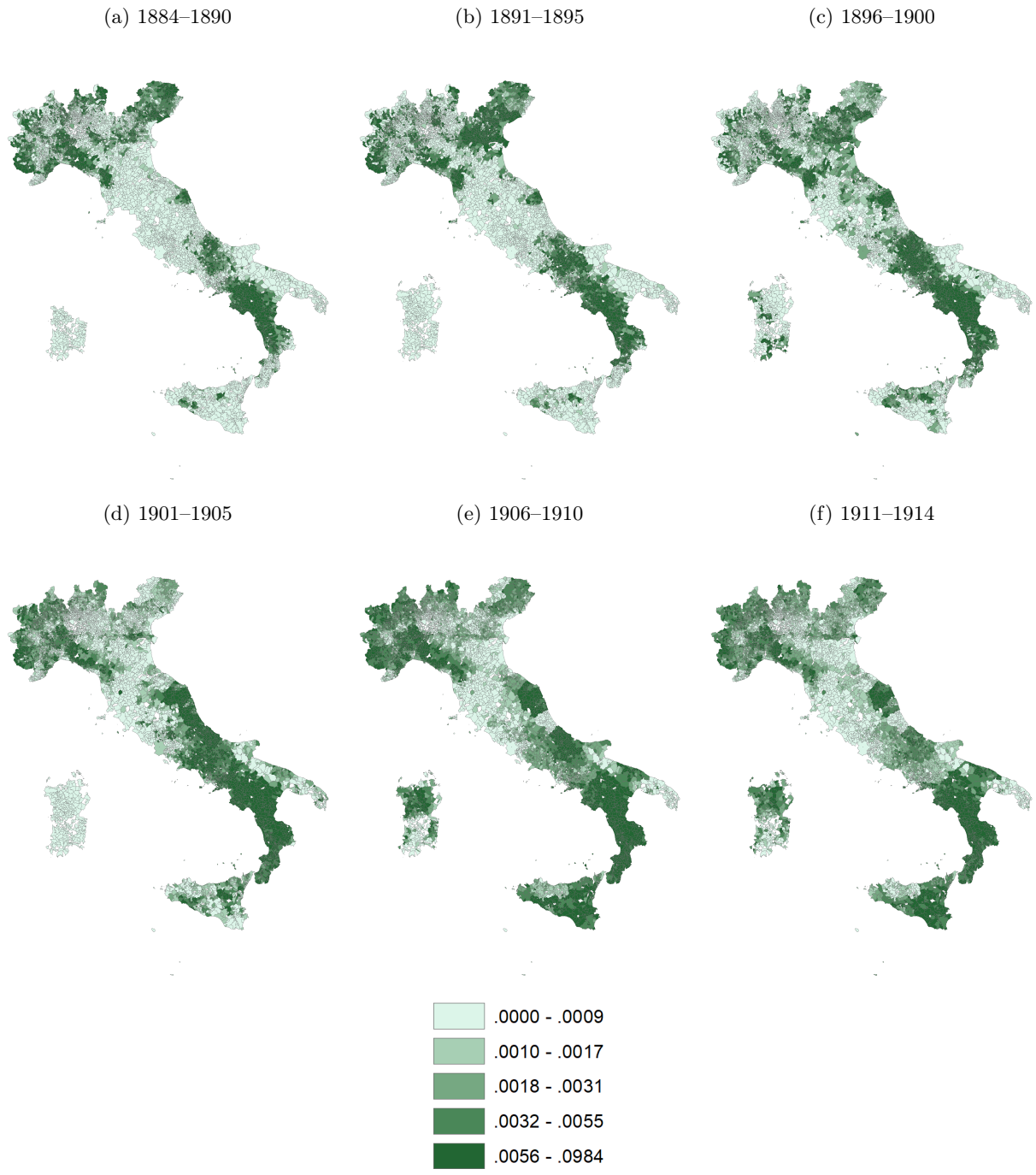


Figure B.8: Municipality-level emigration rates to South America

Note: Each panel presents a municipality's average annual emigration rate to South America in the period in question. Scale is based on quintiles of emigration rates in 1911–1914.

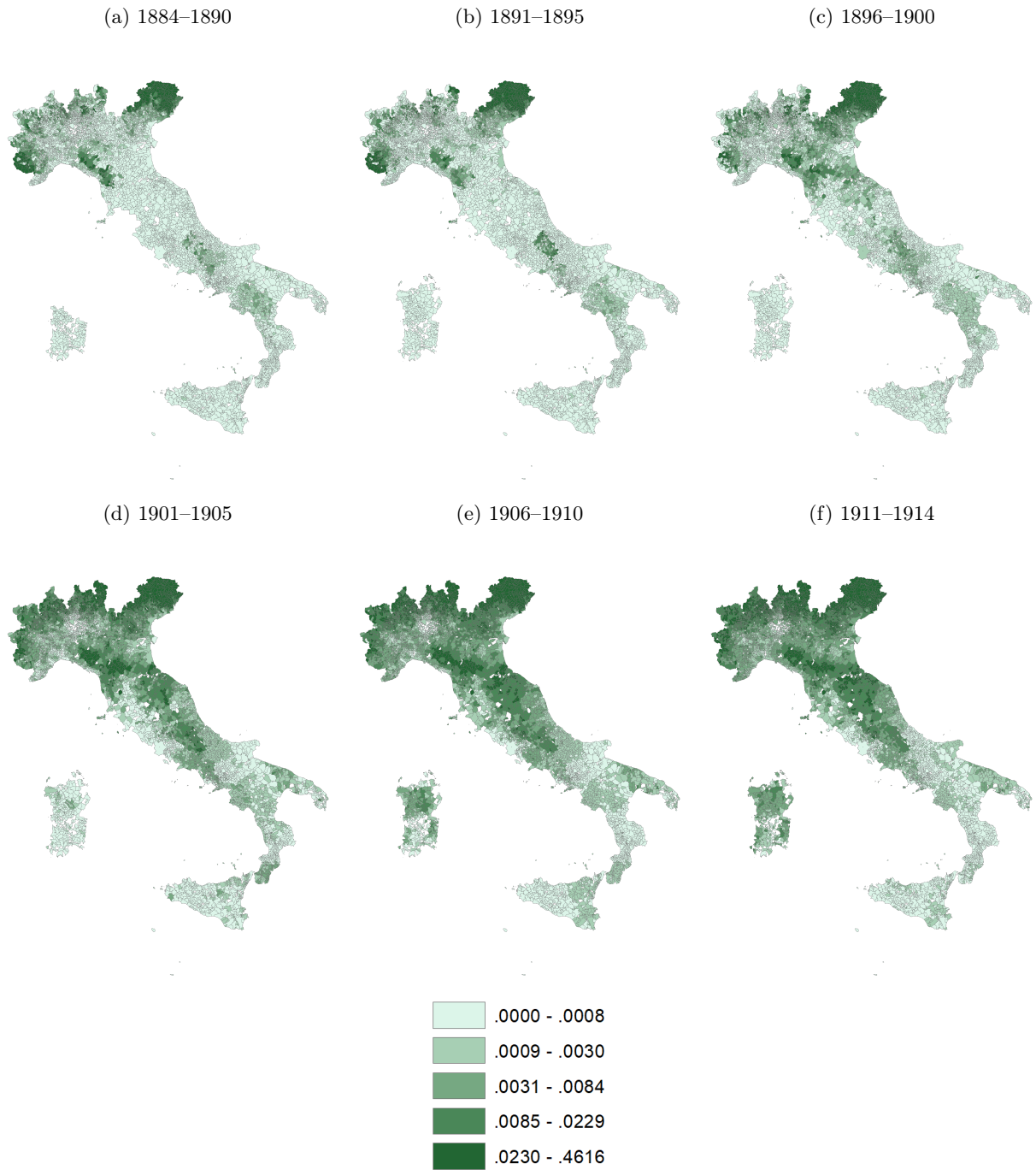


Figure B.9: Municipality-level emigration rates to Europe

Note: Each panel presents a municipality's average annual emigration rate to Europe in the period in question. Scale is based on quintiles of emigration rates in 1911–1914.

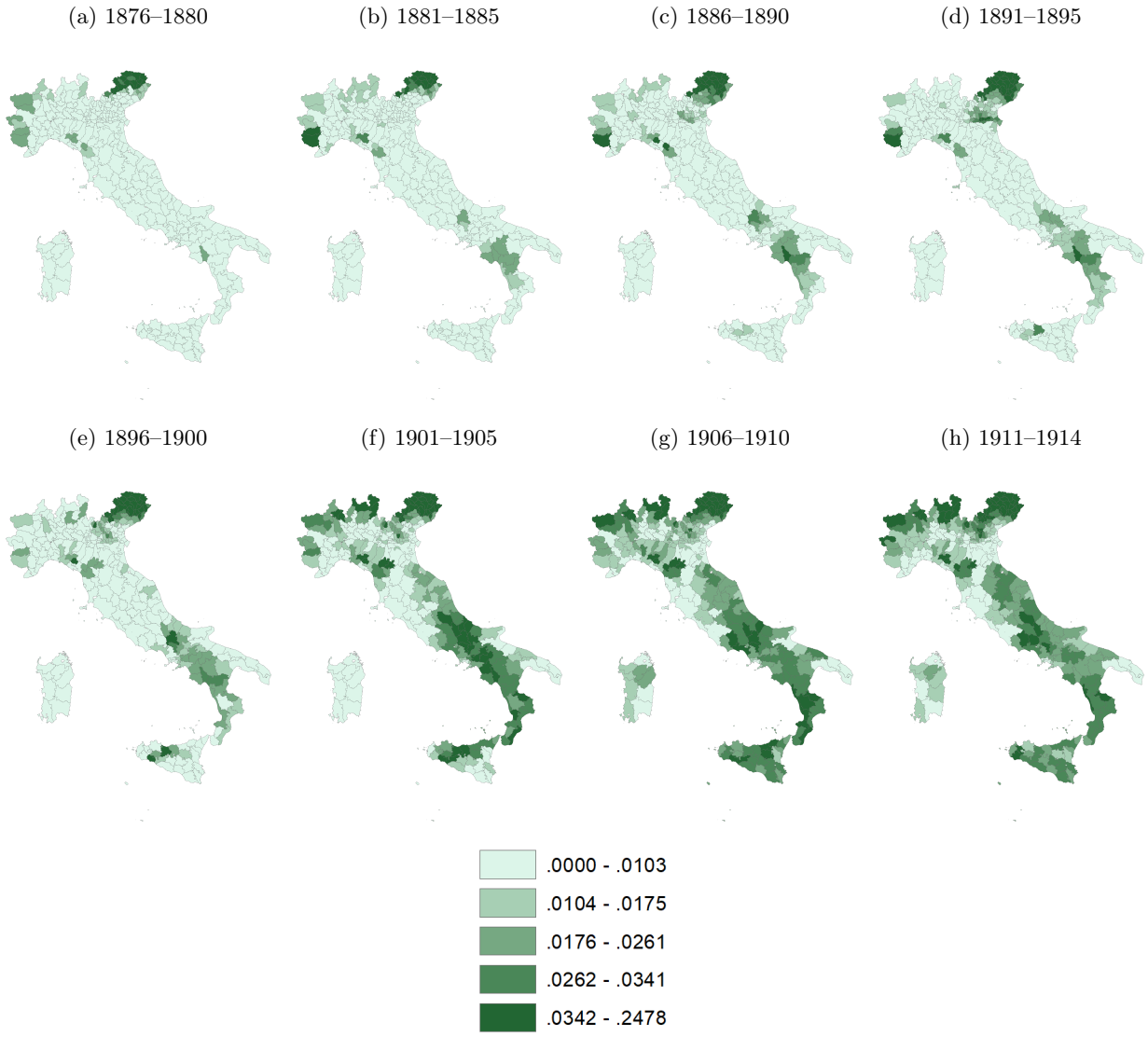


Figure B.10: District-level emigration rates to any destination

Note: Each panel presents a district's average annual emigration rate to any destination in the period in question. Scale is based on quintiles of emigration rates in 1911–1914.

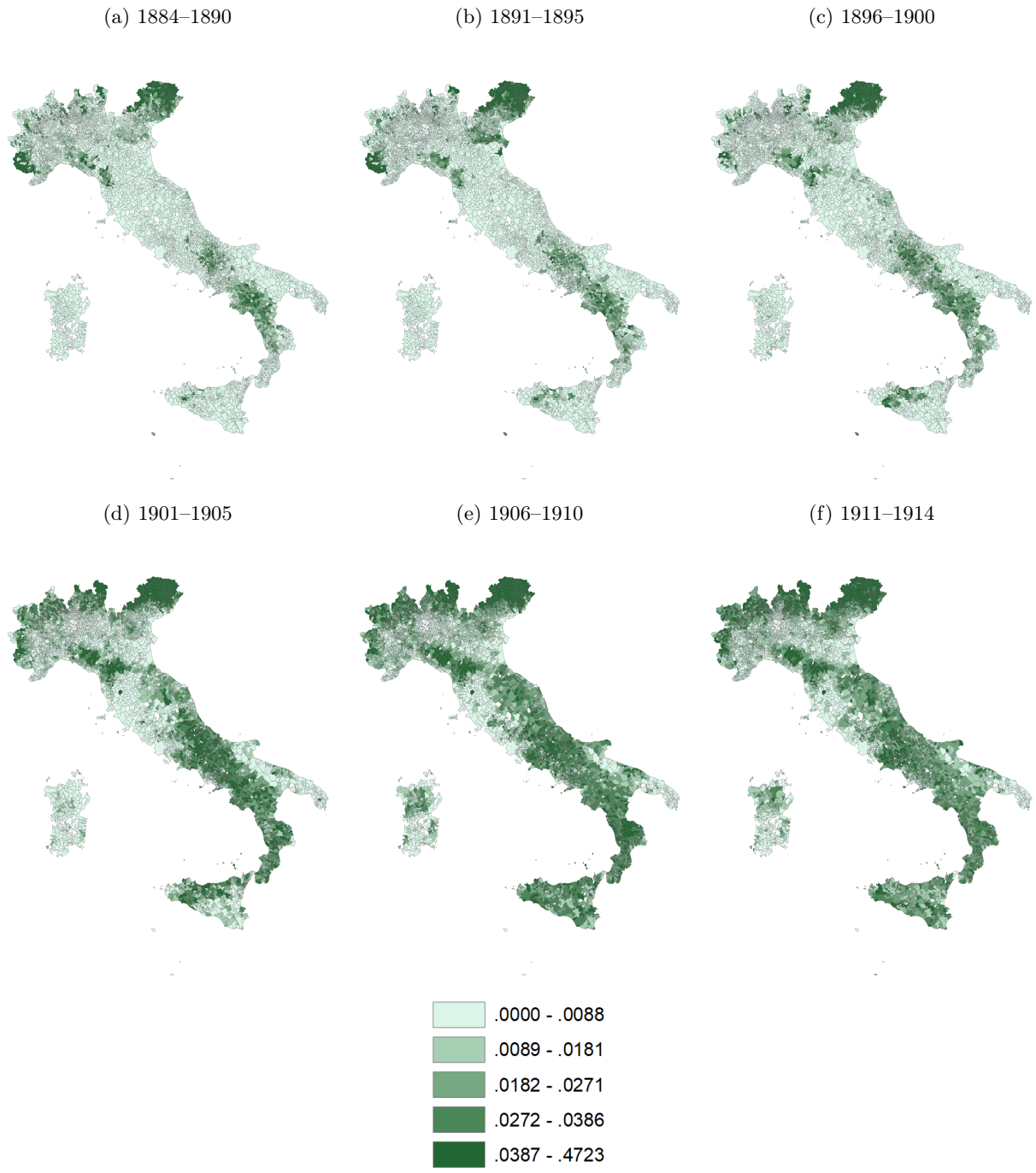


Figure B.11: Municipality-level emigration rates to any destination

Note: Each panel presents a municipality's average annual emigration rate to any destination in the period in question. Scale is based on quintiles of emigration rates in 1911–1914.

Figure B.12: Elevation

Source: Shuttle Radar Topography Mission (Jet Propulsion Laboratory 2014)

Note: Darker shading indicates greater elevation.

Table B.1: Summary statistics for time-varying variables

	District	Commune						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variable	All	All	1884–1890	1891–1895	1896–1900	1901–1905	1906–1910	1911–1914
Any Emigration	0.990	0.772	0.582	0.566	0.674	0.969	1.000	0.995
	(0.100)	(0.420)	(0.493)	(0.496)	(0.469)	(0.173)	(0.000)	(0.070)
Emigration Rates (per k)								
All Destinations	15.564	15.674	9.509	10.039	11.518	22.896	26.373	27.480
	(21.615)	(22.899)	(20.080)	(21.942)	(24.335)	(26.095)	(20.669)	(24.590)
North America	3.334	4.805	1.446	1.502	2.009	7.184	9.035	8.935
	(6.339)	(8.843)	(4.382)	(4.114)	(5.042)	(11.952)	(11.166)	(10.706)
South America	2.658	3.249	3.151	3.660	3.387	3.903	4.472	3.478
	(3.882)	(5.068)	(5.530)	(6.852)	(5.284)	(5.173)	(4.864)	(3.717)
Europe	9.173	8.448	4.313	4.669	5.919	11.273	12.450	14.622
	(20.045)	(19.038)	(15.955)	(18.314)	(21.940)	(21.908)	(17.203)	(21.770)
Mass Emigration								
All destinations (>10 per thousand)	0.464	0.446	0.272	0.276	0.310	0.613	0.772	0.770
	(0.499)	(0.497)	(0.445)	(0.447)	(0.463)	(0.487)	(0.420)	(0.421)
North America (> 5 per thousand)	0.184	0.246	0.084	0.093	0.125	0.312	0.426	0.452
	(0.388)	(0.431)	(0.278)	(0.290)	(0.331)	(0.463)	(0.494)	(0.498)
South America (> 5 per thousand)	0.164	0.217	0.216	0.237	0.241	0.277	0.305	0.228
	(0.370)	(0.412)	(0.412)	(0.425)	(0.428)	(0.448)	(0.460)	(0.420)
Europe (> 5 per thousand)	0.327	0.307	0.134	0.132	0.163	0.417	0.499	0.521
	(0.469)	(0.461)	(0.341)	(0.338)	(0.369)	(0.493)	(0.500)	(0.500)
Within North American Frontier	0.173	0.194	0.031	0.069	0.095	0.132	0.281	0.366
	(0.379)	(0.395)	(0.173)	(0.253)	(0.293)	(0.339)	(0.450)	(0.482)
Distance to North American Frontier (km)	288.062	246.347	410.926	392.245	391.055	349.372	74.415	53.385
	(241.148)	(243.451)	(221.203)	(238.929)	(240.525)	(226.968)	(73.068)	(70.924)
Observations	2,545	56,083	7,909	8,029	8,029	8,029	8,029	8,029
Units	284	8,029	7,909	8,029	8,029	8,029	8,029	8,029

Notes: Observations are at the district-half decade level in column (1) and at the municipality-half decade level in columns (2)–(8). Columns (1) and (2) have multiple observations for each municipality or district (one for each half decade). Columns (3)–(8) have one observation for each municipality. Standard deviations in parentheses.

Table B.2: Summary statistics for time-invariant variables

Variable	(1) All	(2) North	(3) Center	(4) South
Panel A: District-level Data				
District Share of Male Labor in Agriculture	0.547	0.555	0.547	0.534
	(0.109)	(0.101)	(0.135)	(0.109)
District Share of Male Labor in Industry	0.216	0.237	0.191	0.192
	(0.070)	(0.074)	(0.058)	(0.056)
District Adult Male Literacy Rate	0.468	0.582	0.453	0.279
	(0.183)	(0.149)	(0.127)	(0.055)
District Population Fraction Under Age 15	0.328	0.334	0.309	0.327
	(0.024)	(0.024)	(0.026)	(0.018)
Observations	284	154	41	89
Panel B: Municipality-level Data				
Distance to Railroad (1881, km)	9.853	8.871	7.992	12.486
	(12.437)	(12.178)	(9.209)	(13.758)
Mean Elevation (m)	451.418	471.039	390.918	445.771
	(425.914)	(503.853)	(260.907)	(324.431)
Distance to North America Epicenter (km)	149.796	162.926	138.351	132.067
	(90.280)	(68.928)	(78.142)	(120.294)
Distance to South America Epicenter (km)	168.408	124.097	219.139	222.421
	(101.971)	(49.260)	(107.815)	(127.637)
Distance to European Border (km)	240.432	45.426	238.222	586.443
	(264.104)	(37.066)	(122.753)	(173.120)
Distance to Coast (km)	68.738	106.680	30.388	20.045
	(57.649)	(50.310)	(26.493)	(18.982)
Birth Rate (1881)	0.037	0.036	0.036	0.040
	(0.009)	(0.008)	(0.007)	(0.010)
Death Rate (1881)	0.026	0.024	0.027	0.030
	(0.010)	(0.008)	(0.010)	(0.012)
Mutual Aid Members per capita (1878)	0.007	0.009	0.007	0.002
	(0.026)	(0.033)	(0.018)	(0.011)
Postal Savings Deposits per capita (1886, 100s)	5.266	5.970	8.182	2.618
	(21.929)	(24.037)	(32.064)	(5.921)
Observations	8,028	4,371	1,187	2,470

Notes: Observations are at the district level in Panel A and at the municipality level in Panel B. Standard deviations in parentheses. Observation numbers are the minimum with observations for all variables, excluding places whose emigration rates cannot be calculated due to a lack of population data. Distance from railroad is 0 for any municipality with a rail line passing through it in 1881, and the distance from the nearest municipality border to the rail line for all other municipalities. Distance to the epicenters of North America- and South America-bound emigration are from the municipality centroid to the centroid of the epicenter district's capital city. Distance to the European border is from the municipality centroid.

Table B.3: Destination dissimilarity and distance between provinces

Variables	(1) All	(2) All	(3) Major	(4) Minor
log(Distance)	0.134[a]	0.120[a]	0.113[a]	0.073[a]
	(0.003)	(0.004)	(0.004)	(0.004)
Observations	21,114	21,114	21,114	21,114
R-squared	0.303	0.313	0.220	0.127
Controls	No	Yes	Yes	Yes

Significance levels: [a] p<0.01, [b] p<0.05, [c] p<0.1

Notes: Dependent variable is the dissimilarity index in the emigration destination distribution of the two provinces making up a province pair in a given half decade. Unit of observation is a province pair-half decade. Standard errors clustered by province pair. All regressions include half-decade fixed effects. Major destinations are US, Canada, France, Argentina, Uruguay, Switzerland, Austria-Hungary, Germany, and Brazil. Controls are absolute differences in agricultural and industrial employment shares, literacy rates, fraction under age 15, birth rate, death rate, mutual aid society members per capita, and log postal savings deposits per capita.

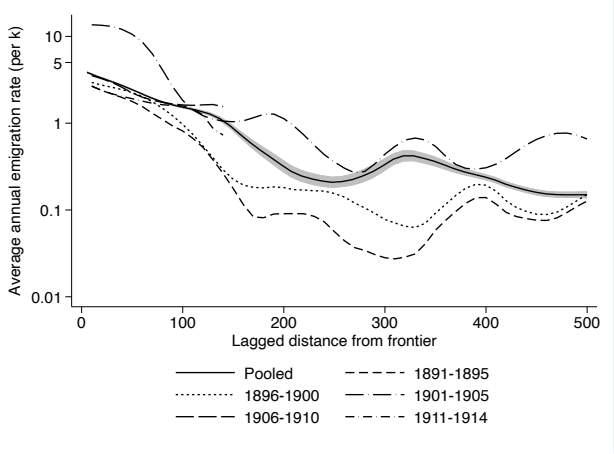


Figure B.13: Emigration rates to North America by distance to the mass migration frontier (km)

Note: This figure is analogous to panel (a) of Figure 11, but focuses on the municipality as the unit of analysis.

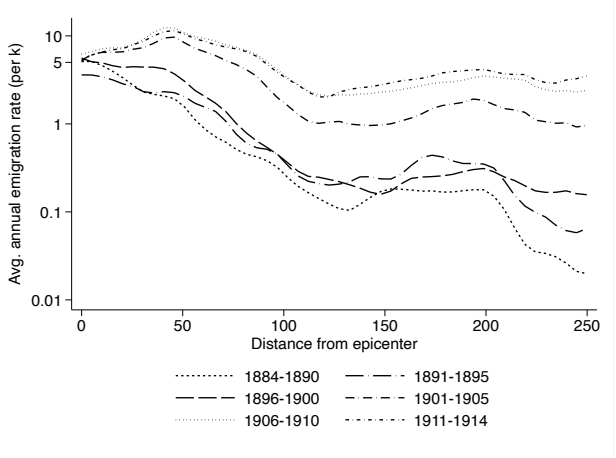


Figure B.14: Emigration rates to North America by distance to epicenter (km)

Note: This figure is analogous to panel (a) of Figure 12, but focuses on the municipality as the unit of analysis.

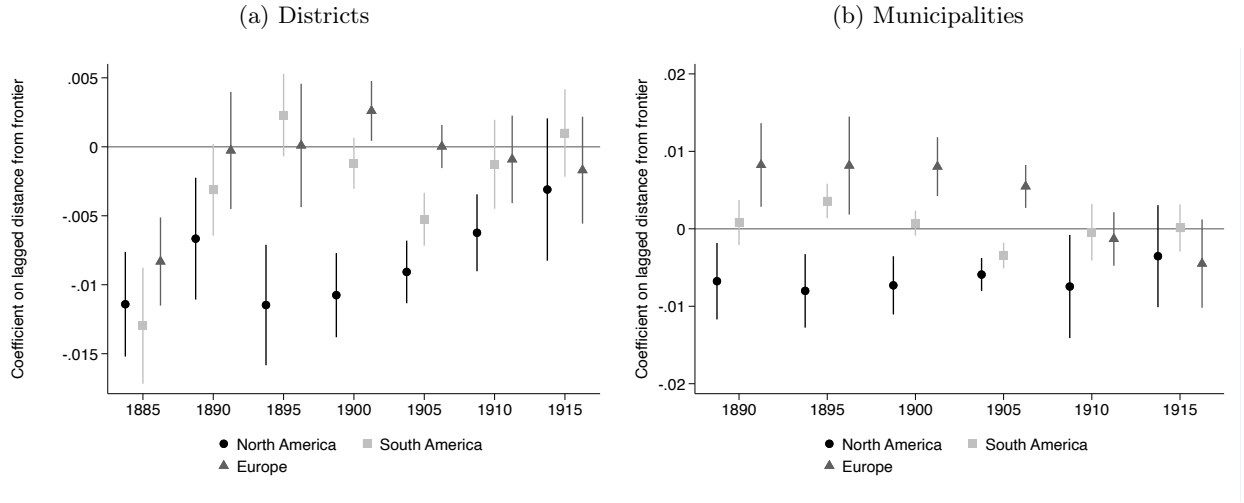


Figure B.15: Emigration to various destinations by distance to the mass migration frontier for North America (km)

Note: This figure repeats the binomial maximum likelihood regressions of panel (c) of Figure 11, but includes results for migration to South America and Europe in addition to those for migration to North America.

Table B.4: Spatial contagion results, OLS

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: 50–250km from epicenters								
Lagged Emigration Exposure	0.957[a]	0.942[a]	0.804[a]	0.836[a]	0.657[a]	0.719[a]	0.528[a]	0.597[a]
	(0.031)	(0.030)	(0.032)	(0.032)	(0.031)	(0.030)	(0.031)	(0.035)
Observations	31,463	31,463	31,463	31,463	31,463	31,463	31,463	31,463
R-squared	0.717	0.751	0.763	0.773	0.782	0.813	0.793	0.840
Panel B: All municipalities								
Lagged Emigration Exposure	0.979[a]	0.932[a]	0.818[a]	0.841[a]	0.666[a]	0.714[a]	0.554[a]	0.590[a]
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.024)	(0.027)	(0.031)
Observations	41,169	41,165	41,165	41,165	41,165	41,165	41,165	41,165
R-squared	0.732	0.763	0.774	0.785	0.792	0.823	0.804	0.850
Additional FE	None	None	C	CT	P	PT	D	DT
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Significance levels: [a] p<0.01, [b] p<0.05, [c] p<0.1

Notes: Panel A limits the sample to municipalities between 50 and 250km of the epicenters of mass migration to North America, whereas Panel B does not. Standard errors clustered at the district level. All specifications include at least half-decade fixed effects and control for half decade-specific functions of own predicted lagged emigration based on distance from the emigration source, local population, distance to coast, and distance to the European frontier. Dependent variable is the log of the emigration rate to North America. Unit of observation is a municipality-half decade. Controls include half decade-specific functions of latitude, longitude, elevation, agricultural employment share, industrial employment share, literacy rate, fraction under age 15, distance to railroad, birth rate, death rate, mutual aid members per capita, and log postal savings deposits per capita. C denotes region (compartimento)-level fixed effects. P denotes province-level fixed effects. D denotes district-level fixed effects. CT, PT, and DT denote region-time, province-time, and district-time fixed effects.

Table B.5: Spatial contagion results, standard instrumentation approach

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Lagged Emigration Exposure	0.619[a]	0.826[a]	0.803[a]	0.794[a]	0.585[a]	0.619[a]	0.503[a]	0.513[a]
	(0.038)	(0.149)	(0.181)	(0.170)	(0.128)	(0.114)	(0.096)	(0.067)
Lagged Own Emigration	0.348[a]	0.258[a]	0.245[a]	0.267[a]	0.284[a]	0.339[a]	0.263[a]	0.356[a]
	(0.030)	(0.066)	(0.067)	(0.061)	(0.038)	(0.027)	(0.028)	(0.018)
Observations	35,332	35,329	35,329	35,329	35,329	35,327	35,329	35,284
Additional FE	None	None	C	CT	P	PT	D	DT
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic	392.6	43.93	34.30	37.09	86.73	69.22	117	119

Significance levels: [a] p<0.01, [b] p<0.05, [c] p<0.1

Notes: Standard errors clustered at the district level. All specifications include at least half-decade fixed effects. Dependent variable is the log of the emigration rate to North America. Unit of observation is a municipality-half decade. Controls include half decade-specific functions of latitude, longitude, elevation, agricultural employment share, industrial employment share, literacy rate, fraction under age 15, distance to coast, distance to the European land border, population, distance to railroad, birth rate, death rate, mutual aid members per capita, and log postal savings deposits per capita. C denotes region (compartimento)-level fixed effects. P denotes province-level fixed effects. D denotes district-level fixed effects. CT, PT, and DT denote region-time, province-time, and district-time fixed effects.

Table B.6: Spatial contagion results, epicenter-based IV, other catchments

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Epicenter-based instrument								
Lagged Emigration Exposure	0.745^a	0.740^a	0.534^b	0.452	0.665^a	0.489	1.025^b	−0.636
	(0.199)	(0.147)	(0.267)	(0.458)	(0.157)	(0.366)	(0.481)	(1.523)
Observations	31,463	31,463	31,463	31,463	31,463	31,462	31,463	31,427
F-statistic	11.75	16.43	6.343	2.382	17.22	3.698	9.771	1.004
Panel B: Frontier-based instrument								
Lagged Emigration Exposure	0.362	0.908^a	0.784^a	0.662^b	1.518^a	0.612	0.104	0.667
	(0.237)	(0.225)	(0.267)	(0.306)	(0.529)	(0.378)	(0.828)	(0.460)
Observations	11,195	11,194	11,194	11,193	11,194	11,184	11,194	11,158
F-statistic	38.90	28.73	25.01	19.83	25.40	13.41	13.71	10.78
Additional FE	None	None	C	CT	P	PT	D	DT
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Significance levels: ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1

Notes: Panel A uses instruments constructed on the basis of a municipality's distance to the nearest epicenter of mass migration, but basing the predicted estimation only on municipalities in other epicenter catchments. Panel B uses instruments constructed on the basis of distance to the frontier of mass migration, but basing the prediction only on municipalities in other provinces. Sample limited to municipality-half decades between 50 and 250km of the migration source (i.e., the epicenter or frontier). Standard errors clustered at the district level. All specifications include at least half-decade fixed effects and control for half decade-specific functions of own predicted lagged emigration based on distance from the emigration source, local population, distance to coast, and distance to the European frontier. Dependent variable is the log of the emigration rate to North America. Unit of observation is a municipality-half decade. Controls include half decade-specific functions of latitude, longitude, elevation, agricultural employment share, industrial employment share, literacy rate, fraction under age 15, birth rate, death rate, mutual aid society membership per capita, and log postal savings deposits per capita. C denotes region (compartimento)-level fixed effects. P denotes province-level fixed effects. D denotes district-level fixed effects. CT, PT, and DT denote region-time, province-time, and district-time fixed effects.

C The Migration of Antonio Squadrito's Group in October 1903

Antonio Squadrito was born around 1877 in the small Sicilian town of Gualtieri-Sicamino, near Messina. In 1898 he decided to migrate to the United States, among the first in his municipality to do so. In New York he had a "distant relative from a northern province," and his passenger manifest listed his American contact as an uncle living on 21st Street, but he paid for his travel by borrowing money from his father, Giovanni. He arrived at the Battery on July 7, 1898, and his first job was in a quarry in Rhode Island. Soon thereafter, several opportunities arose. He befriended another Italian who owned a barbershop in Stonington, CT, and joined him as an employee, gradually paying off his loan. Shortly after, his boss had to leave the business and forced Antonio to take a loan to purchase the shop. The shop prospered, and Antonio had his older married brother Giuseppe come in to help him. Giuseppe was followed by their father and two younger brothers. In June 1903 Antonio married Harriet H. Burtch-Gardiner, who, at 66, was 41 years his senior.[93] That the same summer, he travelled back to his hometown with the purpose of helping the migration of a large number of friends and relatives. By that time, five years after Antonio had first left Sicily, emigration was already widespread in Gualtieri-Sicamino, as can be seen in Figure C.1, which plots migration rates for the municipality and the broader district of Messina.

While in Sicily, Antonio collected a large group of individuals whose migration he facilitated (listed in Table C.1), mainly close and more distant relatives from Gualitieri-Sicamino and from other neighboring places. Among them were his sister-in-law, her four-year-old niece, her brother, and her nephew, all from Gualtieri-Sicamino and destined for Stonington. The others had other destinations in the United States, where they reported having relatives. A sixteen-year-old girl—a cousin from the neighboring municipality of San Filipo—and five young men—all neighbors and family friends from Gualtieri-Sicamino—were traveling to Boston and to New York. Four farmer boys from Soccorso,[94] a detachment (*frazioni*) of Gualtieri-Sicamino, were on their way to the mines in Pennsylvania. They reported relatives in Philadelphia, but in reality they were illegally contracted laborers, and the uncle of one of them was the middleman (perhaps a *padrone* of sorts) who helped to recruit them. The entire group left for Messina en route to Napoli. From Napoli they embarked on the steamship *Prinzess Irene* on October 2, 1903, and arrived at Ellis Island on October 14. Broughton Brandenburg, the journalist and self-proclaimed immigration specialist who followed Squadrito's entourage and documented their migration, noted that this sort of group migration organized by a friend or relative was so common, that "The most notable feature was the ease with which one could detect that

[93] Her wealth, estimated at $60,000, was inherited from her deceased husband, a whaling ship captain (Brandenburg 1904, p. 44).

[94] Brandenburg (1904) mistakenly referred to it as "Socosa."

every seventh or eighth person had been to America before, and now had gathered around him a group of from two to thirty friends, relatives, and neighbors, going over in his care, just as our party was going in the care of Antonio Squadrito and myself" (Brandenburg 1904, p. 172).

In fact, the group was planned to be larger, as they had expected passengers from other municipalities to join them in Messina. These were Giuseppe Cardillo, accompanied by a few other people, and the Papalia family from Monforte San Giorgio, a small town situated about ten kilometers west of Gualtieri-Sicamino. Cardillo's hometown is unknown and so is the specific relation between the two families and the Squadritos. Eventually, according to Brandenburg (1904, p. 133), Cardillo's group decided to postpone their travel and the Papalias ended up taking the next steamer. Indeed, two weeks later, on October 28, Michele and Maria Papalia, originally from Monforte San Giorgio, and their five-year-old daughter Rosina were recorded arriving at Ellis Island on board the steamship *Lahn*, where they were listed as American citizens returning home to New York. All in all, the extended group that planned their joint voyage comprised of neighbors, friends, relatives, and other acquaintance from five different localities, at least four of which were within a short distance from one another.

How does this case fit the theoretical framework proposed in section 3? Clearly, it shows that the reality was more complex than the stylized story about a linear chain in which one individual links others in his geographic environment who depend on him, and leads them to the same destination. It is not clear, for example, how crucial the role played by Antonio Squadrito's relative was in enabling his own migration in 1898, and therefore it is impossible to tell whether or not he was a real pioneer. Even if he were linked by his relative, it is hard to tell whether this linkage conformed to our assumption that social contacts were largely local, because although he was a relative, according to Brandenburg he was from a "northern province" (Brandenburg 1904, p. 43). Furthermore, many in the group relied on additional contacts in the United States. They were supported by Antonio, but he was not their sole sponsor, and it is probable that they would have migrated even without his help. Indeed, only a few were destined to join him in Stonington. Networks merged and diverged to different destinations, and it is unknown whether the emigration from Gualtieri-Sicamino could be traced back to a single local founding father or to several ancestors separately linked from other municipalities, and whether any of them were virtual pioneers. Nevertheless, those going to other destinations were still relying on other personal links, usually family members. Even those who were in reality contracted laborers were recruited through a relative. If the case of Squadrito's group is indicative, then in a broad sense, the Italian transatlantic movement occurred within local networks based both on intra-place and on short-distance inter-place links. This is precisely the core insight that the theoretical

framework that we propose is meant to capture.

Table C.1: Antonio Squadrito's group, on board *Prinzess Irene*, arriving October 14, 1903

First Name	Last Name	Sex	Age	Relation to Antonio Squadrito	Place of Origin	Joining	Destination
Antonio	Squadrito	M	26		Gualtieri-Sicamino	Brothers, Giuseppe, Carmelo, and Gaetano	Stonington, CT
Carmela	Squadrito	F	32	Sister in law	Gualtieri-Sicamino	Husband, Giuseppe Squadrito (Antonio's brother)	Stonington, CT
Caterina	Squadrito	F	4	Niece	Gualtieri-Sicamino	Father, Giuseppe Squadrito	Stonington, CT
Giovanni	Pulejo	M	49	Brother in law, probably also a cousin	Gualtieri-Sicamino	Brother, Nicola	Boston, MA
Felice	Pulejo	M	16	Nephew	Gualtieri-Sicamino	Uncle, Nicola	Boston, MA
Concetta	Fomica	F	15	Cousin	San Filipo	Uncle, Stefano Senedile, Boston	Boston, MA
Antonio	Nastasia	M	16	Neighbor	Gualtieri-Sicamino	Uncle, Tommaso Trovato, Boston	Boston, MA
Gaetano	Mullura	M	16	Neighbor	Gualtieri-Sicamino	Uncle, Nicolo Puleo, Boston	Boston, MA
Nicola	Curro	M	27	Family friend	Gualtieri-Sicamino	Cousin, Angelo Ragusa, New York	New York, NY
Nunzio	Giunta	M	23	Fellow townsman	Gualtieri-Sicamino	Cousin, New York	New York, NY
Antonio	Genino	M	21	Fellow townsman	Gualtieri-Sicamino	Uncle, Giuseppe Maucino, Philadelphia	Philadelphia, PA
Salvatore	Niceta	M	20	Farm boy from detached village	Soccorso	Brother, Giuseppe Niceta, Philadelphia	Philadelphia, PA
Benedetto	Runzio	M	21	Farm boy from detached village	Soccorso	Cousin, Giuseppe Niceta, Philadelphia	Philadelphia, PA
Luciano	Sofia	M	17	Farm boy from detached village	Soccorso	Cousin, Giuseppe Niceta, Philadelphia	Philadelphia, PA
Salvatore	Damico	M	23	Farm boy from detached village	Soccorso	Brother in law, Antonio Salvatore, Philadelphia	Philadelphia, PA

Sources: Brandenburg (1904) and the Statue of Liberty-Ellis Island Foundation

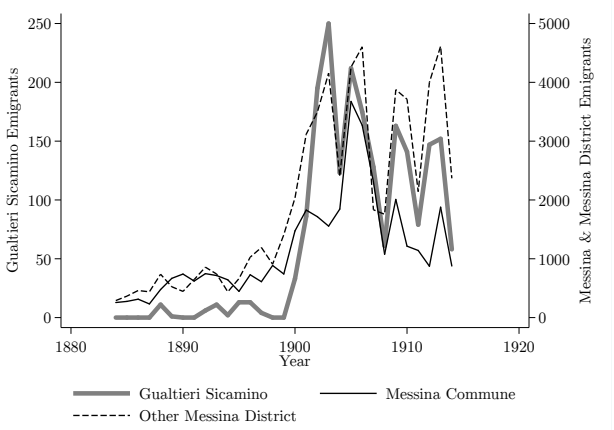


Figure C.1: Emigration from Gualtieri-Sicamino and Messina District

Note: This figure presents migrant counts for Gualtieri-Sicamino from 1884–1914. Gualtieri-Sicamino is part of the district of Messina. The other lines in the figure present the migrant counts for the municipality of Messina and for all of the district of Messina except Gualtieri-Sicamino and the municipality of Messina.

D Model Simulations

This appendix presents the results of a simulation that generates the main predicted patterns discussed in Section 3.3.

D.1 Demand for Migration

We simulate a diffusion process over a 40×40 grid. Each square represents a place whose population is drawn from a distribution similar to that of 1881 Italian municipalities according to Zipf's Law.[95] The distance between each pair of places is the Euclidean distance between the centers of the squares. There are two destinations, a and b, and we arbitrarily assign two epicenters, one for each destination, in opposite corners of the grid. Each epicenter is assigned an initial share of the population $L_1 = 0.1$ that is already linked to the respective destination at the beginning of period 1; other than that, there has been no prior migration and there are no linked individuals in any other place.

The diffusion parameters are as follows:

$$\Theta = \{\alpha, \lambda, \delta, \pi\} = \{10^{-4}, 25, 20, 4\}$$

For simplicity, we assume symmetry between the two destinations, homogeneity across origins, and immutability over time. The demand for migration has a simple logit form. At any period, the probability of migration to destination $d \in \{a, b\}$ for an individual from place i who is linked to destination d alone is:

$$m_i^d = \frac{exp^{\eta_i}}{1 + \exp^{\eta_i}}$$

where η_i stands for the strength of the local push factors in place i, and in the benchmark simulation is simply $\eta_i = \eta = -4.5$. The demand for migration to destination $d \in \{a, b\}$ for an individual linked to both destinations is:

$$m_i^d = \frac{exp^{\eta_i}}{1 + 2\exp^{\eta_i}}.$$

D.2 Predictions

We present here a series of figures generated from a single simulation that exemplifies each of the main predicted patterns. The simulation was run over 60 periods, during which the country moved from only a

[95] Specifically, the largest place has a population of 1,000,000, the n^{th} largest place has a population of $\frac{1,000,000}{n}$, and places are assigned a random location across the grid.

small fraction of the population linked, in the two corner epicenters, to virtual saturation.

D.2.1 Convergence

Figure D.1(a) demonstrates σ-convergence, corresponding to real-data Figure 7. The vertical axis presents the coefficient of variation of emigration rates to destination a across all places. Over time, the variation decreases monotonically until all places are saturated.[96] Figure D.1(b) demonstrates β-convergence, corresponding to real-data Figure 8. Each dot represents one place. The horizontal axis has the rate of emigration to destination a in the first 30 periods, and the vertical axis has the log of the ratio between the rate of emigration in the remaining 30 periods and the first 30 periods. The pattern of β-convergence is apparent in that the increase in the rate of emigration is negatively correlated with the rate of emigration in the early period.

D.2.2 S-Shaped Local Trends

Figure D.2 demonstrates how the typical emigration path from all places follows an S-shaped trend, regardless of the timing of the onset of mass emigration. As in real-data Figure 9, the time scale is shifted for each place such that period zero is the first period in which the emigration to destination a had reached one per thousand. The places are binned into four quartiles, by the period in which mass emigration was reached. Each bin is represented by three curves, for the mean and for the 25^{th} and the 75^{th} percentiles in each period since mass emigration.

D.2.3 Correlated Destinations

In Figure D.3, corresponding to Figure 10, each curve represents the relationship between the distance between each pair of places and the dissimilarity of their destination choices, where the dissimilarity index is based on total emigration over segments of 10 periods. The upward slope of all curves over the bulk of the range of distance means that places farther from one another have less similar distributions of destinations.[97]

[96] Notice that the coefficient of variation converges to zero because the places are assumed to be homogeneous; if they were heterogeneous in terms of push factors (η), then it would converge to a positive value.

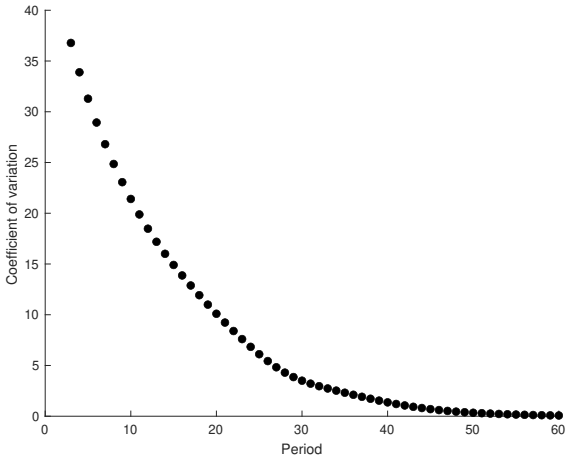
[97] That the curves for the early periods begin to slope downwards at the longest distances is the product of the specific structure of the simulation. Suppose that the epicenters are placed in the northwest and southeast corners of the grid. The northeast and southwest corners are at the maximum possible distance from one another, but are reached by the spreading waves of migration to each destination roughly simultaneously, and thus develop similar destination choice profiles. This will arise in any case in which there are two destinations with one epicenter each, and will occur along the axis orthogonal to that between the epicenters. For two reasons we do not consider this part of the prediction dispositive in our case. First, in the case of a long narrow strip (such as Italy) between the epicenters that excludes the northeast and southwest corners, the same pattern cannot arise. Second, with more than one destination and more than one epicenter per destination, the pattern is less likely to arise.

That the slopes are shifted downward over time means that all pairs become increasingly similar in their destination choices.

D.2.4 Spatial expansion and the frontier effect

Figure D.4(a), corresponding to real-data Figure 11, plots for each 10-period segment the relationship between the distance to the frontier of mass emigration to destination a in the previous segment (using the threshold of 1 per thousand) and current segment's log emigration rate.[98] The prediction is that within each segment of time, all curves are strongly downward sloping, indicating the close dependence of emigration on proximity to the frontier. Figure D.4(b), corresponding to real-data Figure 12, does the same with respect to distance to the nearest epicenter. The downward sloping curves become increasingly flat over time, indicating an initially strong association between proximity to the epicenters and emigration that gradually diminishes as the country becomes saturated.

(a) σ-convergence



(b) β-convergence

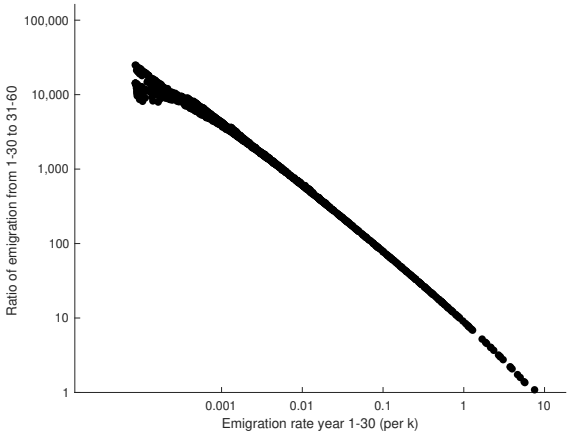
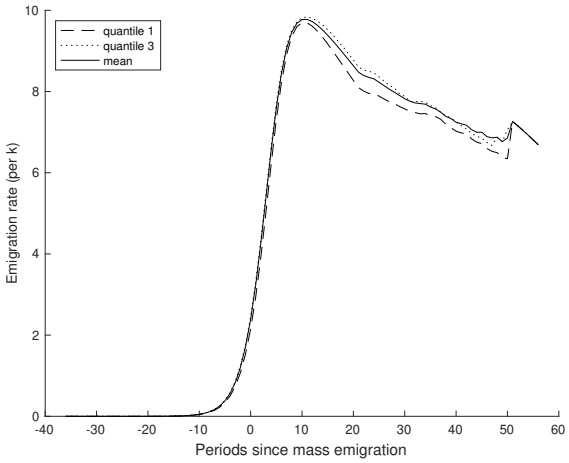


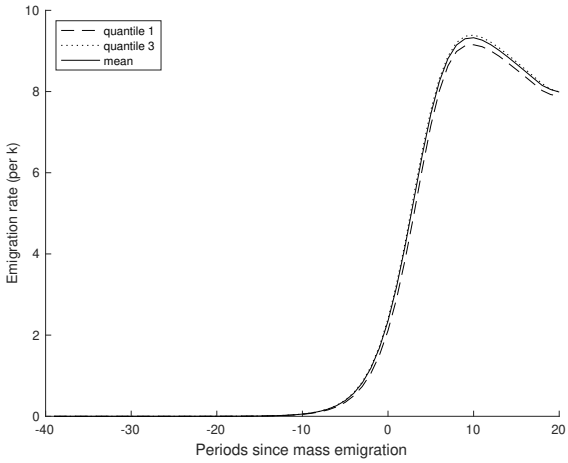
Figure D.1: Convergence

[98] The distance to the frontier in periods 1–10 is arbitrarily defined to be the distance to the epicenters, since there is no prior emigration.

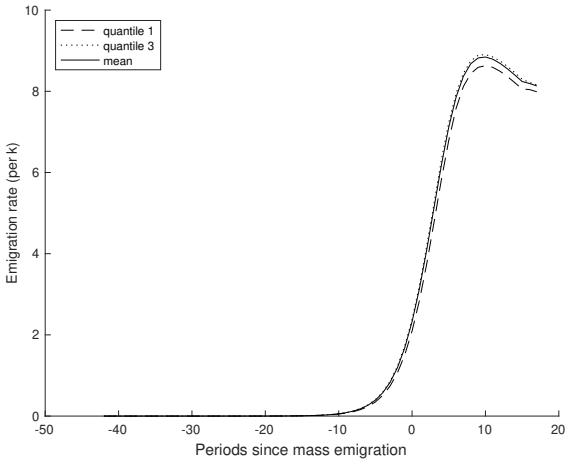
(a) Quartile 1



(b) Quartile 2



(c) Quartile 3



(d) Quartile 4

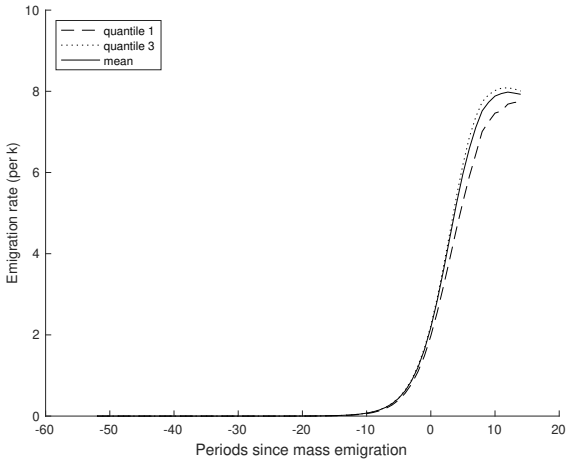


Figure D.2: S-shaped local trends

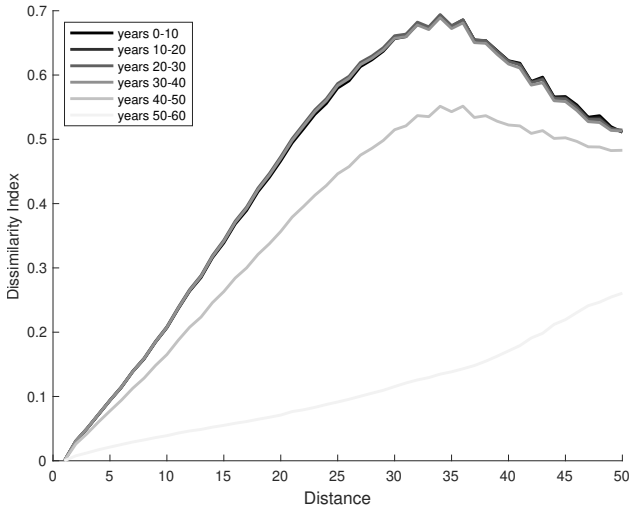


Figure D.3: Correlated destinations

(a) Emigration and distance from the frontier

(b) Emigration and distance from the epicenters

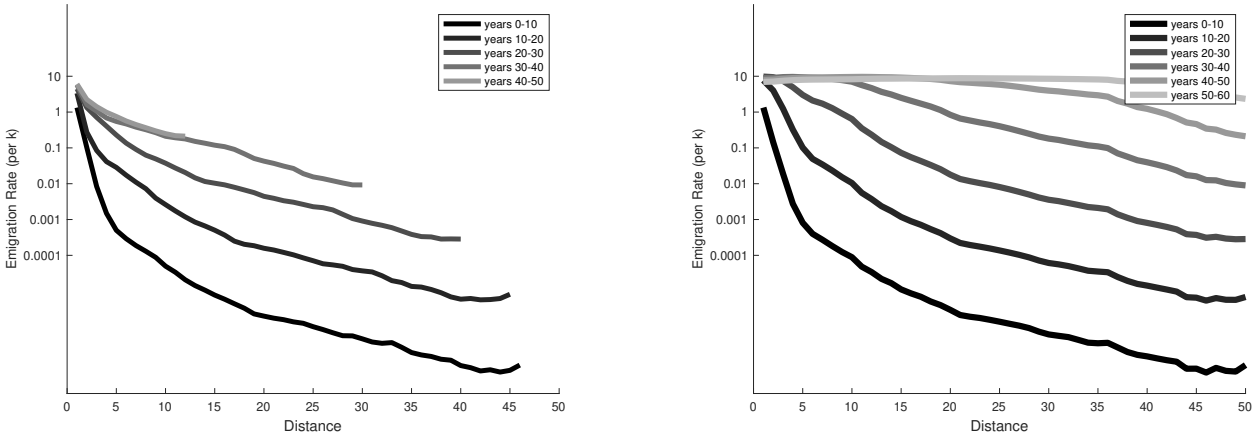


Figure D.4: Spatial expansion and the frontier effect

E List of Historical Statistical Publications

This appendix lists the historical statistical publications that provided our raw data. All of our emigration data (and our population data for 1901) were from *Statistica della Emigrazione Italiana per l'Estero* (some with an additional subtitle), published by the Ministero di Agricoltura, Industria, e Commercio, Direzione Generale della Statistica. Table E.1 lists the publication year and press of each volume that we used.

Our 1881 census data are from the *Censimento della Popolazione del Regno d'Italia al 31 Dicembre 1881* published by the Ministero di Agricoltura, Industria, e Commercio, Direzione Generale della Statistica and Tipografia Bodoniana. The literacy and age data are from Table II of Volume II, "Popolazione classificata per età, sesso, stato civile, e istruzione elementare," published in 1883. The employment-by-industry data are from Table I of Volume III, "Popolazione classificata per professioni o condizione," published in 1884.

Our data on municipality population for 1881 are from the *Comuni e Loro Popolazione ai Censimenti dal 1861 al 1951*, published by the Istituto Centrale di Statistica and Azienda Beneventana Tipografica Editoriale in 1960.

Our data on births and deaths in 1881 are from Table I of the *Movimento dello Stato Civile Anno XX.—1881*, published by the Ministero di Agricoltura, Industria, e Commercio, Direzione Generale della Statistica and Tipografia Bodoniana in Rome in 1882.

Our data on postal savings credit in 1886 are from the *Relazione intorno al servizio delle Casse Postale de Risparmio durante l'Anno 1886*, published by the Stamperia Reale in Rome in 1888.

Our data on membership in mutual aid societies are from the *Statistica delle Società di Mutuo Soccorso, Anno 1878*, published by the Ministero di Agricoltura, Industria, e Commercio, Direzione Generale della Statistica and the Stamperia Reale in Rome in 1880.

Table E.1: Sources of emigration data

Years of Coverage	Publication Year	Press
1876	1877	Elzeviriana
1877–1878	1880	E. Sinimberghi
1879	1880	Cenniniana
1880–1881	1882	Bodoniana
1882	1883	Fratelli Centenari
1883	1884	Camera dei Deputati
1884–1885	1886	Camera dei Deputati
1886	1887	Aldina
1887	1888	Aldina
1888	1889	Aldina
1889	1890	dell'Opinione
1890	1891	dell'Opinione
1891	1892	dell'Opinione
1892	1893	Cooperativa Romana
1893	1894	Cooperativa Romana
1894–1895	1896	Bontempelli
1896–1897	1899	Nazionale G. Bertero
1898–1899	1900	Nazionale G. Bertero
1900–1901	1903	Nazionale G. Bertero
1902–1903	1904	Nazionale G. Bertero
1904–1905	1906	Nazionale G. Bertero
1906–1907	1908	G. Civelli
1908–1909	1910	Nazionale G. Bertero
1910–1911	1913	Nazionale G. Bertero
1912–1913	1915	Ludovico Cecchini
1914–1915	1918	Ludovico Cecchini
1918–1920	1925	Provveditorato Gener. dello Stato

F Preparing Official Statistics for Analysis

The data that we collected from the *Statistica della Emigrazione Italiana per l'Estero* volumes and from the 1881 Italian census required considerable preparation before they could be used for analysis. At the municipality level, the main difficulties are the changing of municipality names over time, and the combination or division of municipalities to form other municipalities. A key source for this effort was the *Comuni e Loro Popolazione ai Censimenti dal 1861 al 1951*, published by ISTAT (the Italian statistical bureau) in 1960. This publication describes the changing borders of municipalities, allowing us to create consistently defined municipalities over the entire sample period, based on borders in 1904. Another difficulty arose from the existence of two sometimes conflicting records for the same municipality-year in cases when two different volumes presented data for the same year. In this case, we used data from the later-published volume.

Our analysis also requires knowing the geographic location of each municipality. For municipalities that still exist (the vast majority), we were able to simply match the list of municipality names to a GIS file of modern municipalities (ISTAT 2018) whose historical provinces could be determined using a shapefile of historic province boundaries (ISTAT 2019). This was more difficult in the case of historic municipalities that were consistently defined throughout our study period but have since ceased to exist. For instance, the municipality of Santo Stefano di Briga existed throughout our study period, but has since been incorporated into the municipality of Messina. The best guess of geographic location that we are able to derive is thus to place Santo Stefano di Briga in the same place as Messina. This simplification is a possible source of error, but because most municipalities are quite small, the resulting error is likely to be small.

Another issue was the mapping of districts. To our knowledge, no shapefile of Italian districts existed at the time that we cleaned our data, though one has since become available (ISTAT 2019). We constructed the shapefile that we use by merging the polygons of all municipalities assigned to a particular district. For municipalities that were created after our study period, we determined the municipality of which they were once a part, and assign the modern municipality to the district of the historic municipality from which it was split. Comparison of our resultant shapefile to a map that we were able to locate of historical districts, as well as that provided by ISTAT, shows that our generated shapefile is extremely accurate.

Another issue arose from the fact that northern provinces that were previously part of the Austro-Hungarian Empire had *distretti* instead of *circondari*. We treat both of these as districts, but the *distretti* were smaller and were eventually eliminated, creating provinces with a single *circondario*. For the emigration data, we can reconstruct the *distretti* totals from the municipality-level data. For the census data, we must use province-level data on literacy and employment for these northern provinces.

G Results Including Data on "Other Municipalities"

This appendix addresses the fact that for years 1903 and earlier, the emigration of some municipalities was not listed in the *Statistica della Emigrazione Italiana per l'Estero*, but was instead included in an aggregate report for each district under the header of "Other Municipalities in this District." To ensure that this is not responsible for driving results, we allocate this unassigned emigration equally to the excluded municipalities and repeat the main results. Since this does not affect the district-level data, those results are not repeated here.

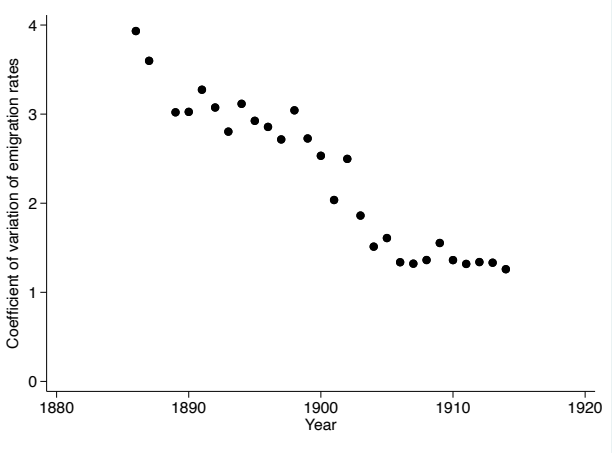


Figure G.1: σ-convergence in emigration rates to North America

Note: Each point represents the coefficient of variation in emigration rates to North America in a particular year.

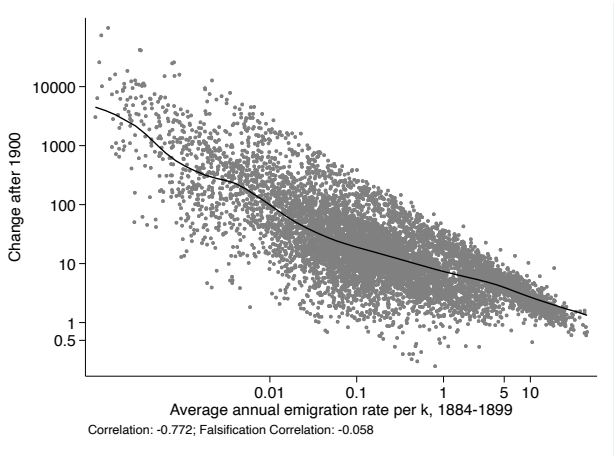


Figure G.2: β-convergence in emigration rates to North America

Note: Each point represents a municipality. The x-axis is the average annual emigration rate for 1884–1899 on a log scale. The y-axis is the ratio of the average emigration rate before and after 1900, also on a log scale. The falsification correlation is the correlation of the change in emigration and emigration after 1900; that it is not positive indicates that the negative relationship shown in the graphs is unlikely to be spurious, as explained in section 5.1.

(a) Municipalities

(b) Municipalities, grouped

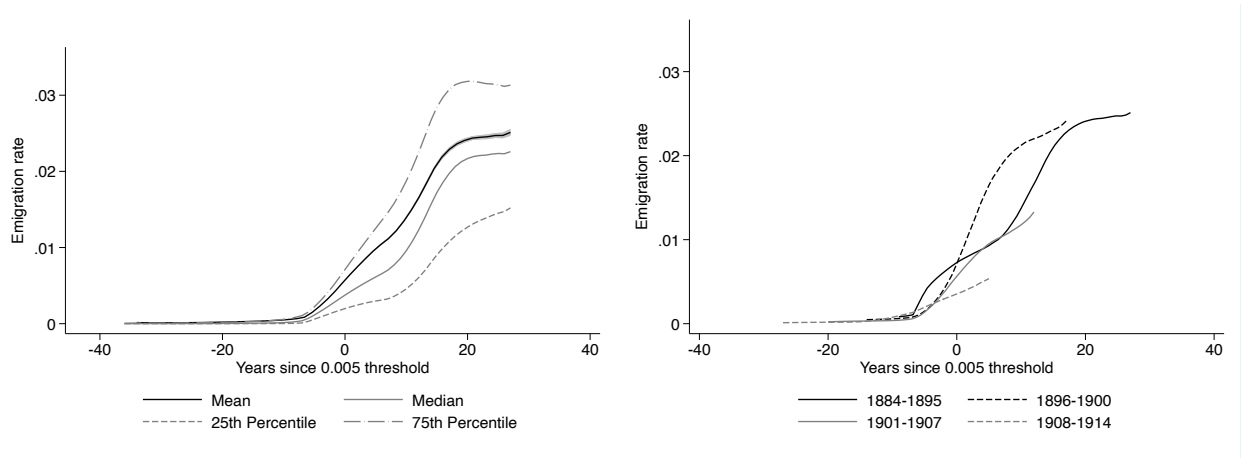


Figure G.3: S-shaped time series of migration to North America

Note: Panels (a) plots a non-parametric regression of emigration rates to North America against time, normalized so that year 0 is the first year in which a place had an emigration rate of at least 5 per thousand. The shaded area is a 95-percent confidence interval. Panel (b) divides municipalities according to the half decade in which they crossed the threshold.

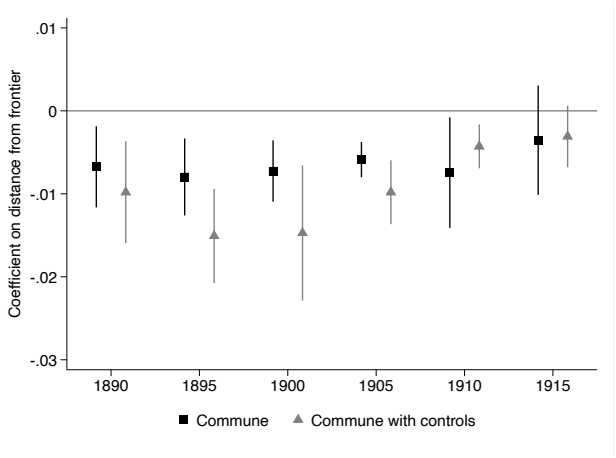


Figure G.4: Emigration rates to North America by distance to the mass migration frontier (km)

Note: This figure estimates a binomial maximum likelihood regression of emigration rates on half decade-specific functions of lagged distance from the frontier of mass migration to North America and plots the coefficients on lagged distance from the frontier. Panel (b) also includes a regression controlling for half decade-specific functions of various controls.

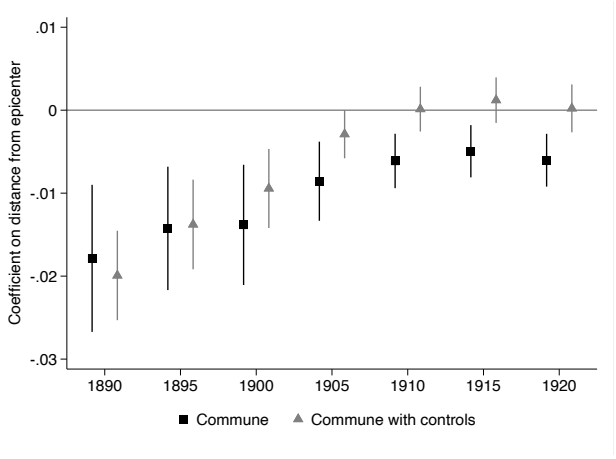


Figure G.5: Emigration rates to North America by distance to epicenter (km)

Note: This figure estimates a binomial maximum likelihood regression of emigration rates on half decade-specific functions of distance from the nearest epicenter of emigration to North America and plots the coefficients on distance from epicenter. It also includes a regression controlling for half decade-specific functions of various controls.

Table G.1: Spatial contagion results

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Epicenter-based instrument								
Lagged Emigration Exposure	0.936^{a}	1.013^{a}	0.763^{a}	0.670^{a}	0.756^{a}	0.493^{b}	1.055^{a}	0.244
	(0.110)	(0.109)	(0.181)	(0.222)	(0.161)	(0.228)	(0.348)	(0.281)
Observations	36,697	36,697	36,697	36,697	36,697	36,697	36,697	36,668
F-statistic	37.86	55.91	20.68	15.46	26.69	12.75	34.15	12.66
Panel B: Frontier-based instrument								
Lagged Emigration Exposure	0.834^{a}	0.875^{a}	0.534^{b}	0.622^{a}	0.611^{a}	0.484^{a}	0.315	0.358^{c}
	(0.192)	(0.184)	(0.216)	(0.207)	(0.179)	(0.170)	(0.384)	(0.187)
Observations	12,670	12,668	12,668	12,667	12,668	12,658	12,668	12,639
F-statistic	77.19	83.45	78.03	70.23	81.45	79	61.16	53.69
Additional FE	None	None	C	CT	P	PT	D	DT
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Significance levels: ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1

Notes: Panel A uses instruments constructed on the basis of a municipality's distance to the nearest epicenter of mass migration. Panel B uses instruments constructed on the basis of distance to the frontier of mass migration. Sample limited to municipality-half decades between 50 and 250km of the migration source (i.e., the epicenter or frontier). Standard errors clustered at the district level. All specifications include at least half-decade fixed effects and control for half decade-specific functions of own predicted lagged emigration based on distance from the emigration source, local population, distance to coast, and distance to the European frontier. Dependent variable is the log of the emigration rate to North America. Unit of observation is a municipality-half decade. Controls include half decade-specific functions of latitude, longitude, elevation, agricultural employment share, industrial employment share, literacy rate, fraction under age 15, birth rate, death rate, mutual aid society membership per capita, and log postal savings deposits per capita. C denotes region (compartimento)-level fixed effects. P denotes province-level fixed effects. D denotes district-level fixed effects. CT, PT, and DT denote region-time, province-time, and district-time fixed effects.

H Results with 1881 Population

The main results use 1901 population as the denominator in calculating emigration rates. This appendix repeats the main results using 1881 population as the denominator.

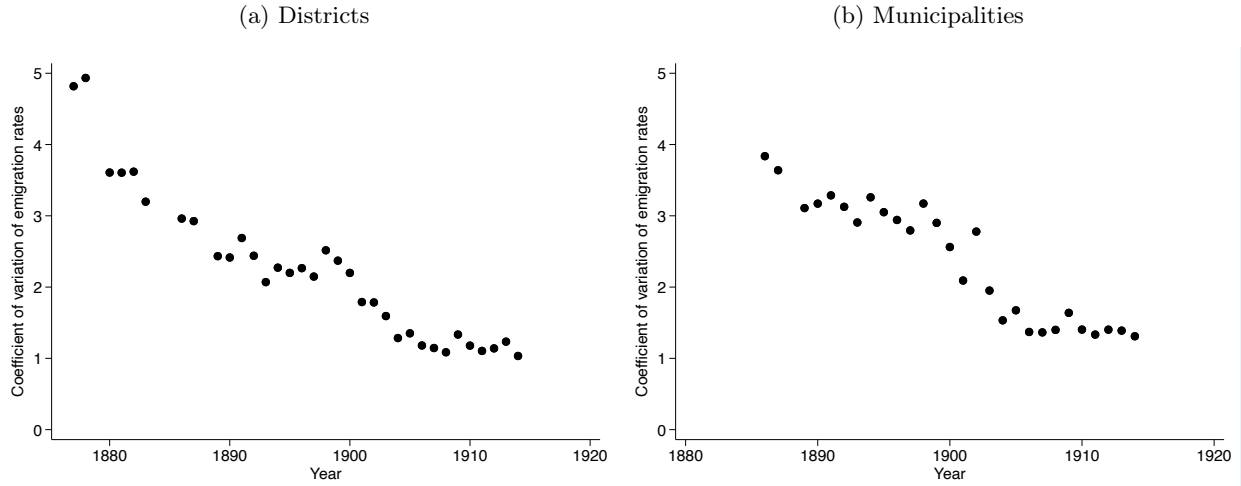


Figure H.1: σ-convergence in emigration rates to North America

Note: Each point represents the coefficient of variation in emigration rates to North America in a particular year.

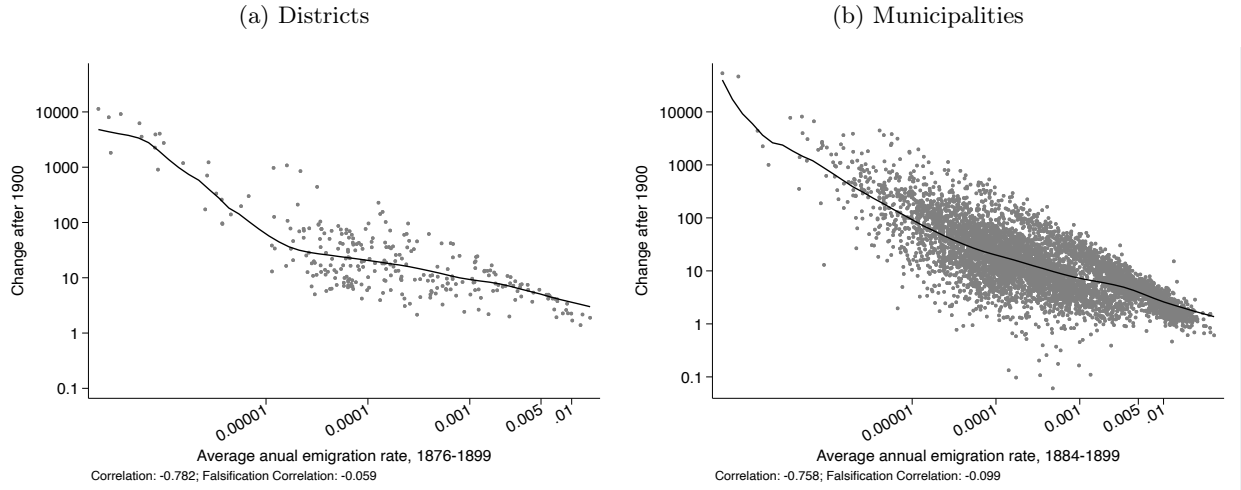


Figure H.2: β-convergence in emigration rates to North America

Note: Each point represents a municipality or district. The x-axis is the average annual emigration rate for a district for 1876–1899 or a municipality for 1884–1899 on a log scale. The y-axis is the ratio of the average emigration rate before and after 1900, also on a log scale. The falsification correlation is the correlation of the change in emigration and emigration after 1900; that it is not positive indicates that the negative relationship shown in the graphs is unlikely to be spurious, as explained in section 5.1.

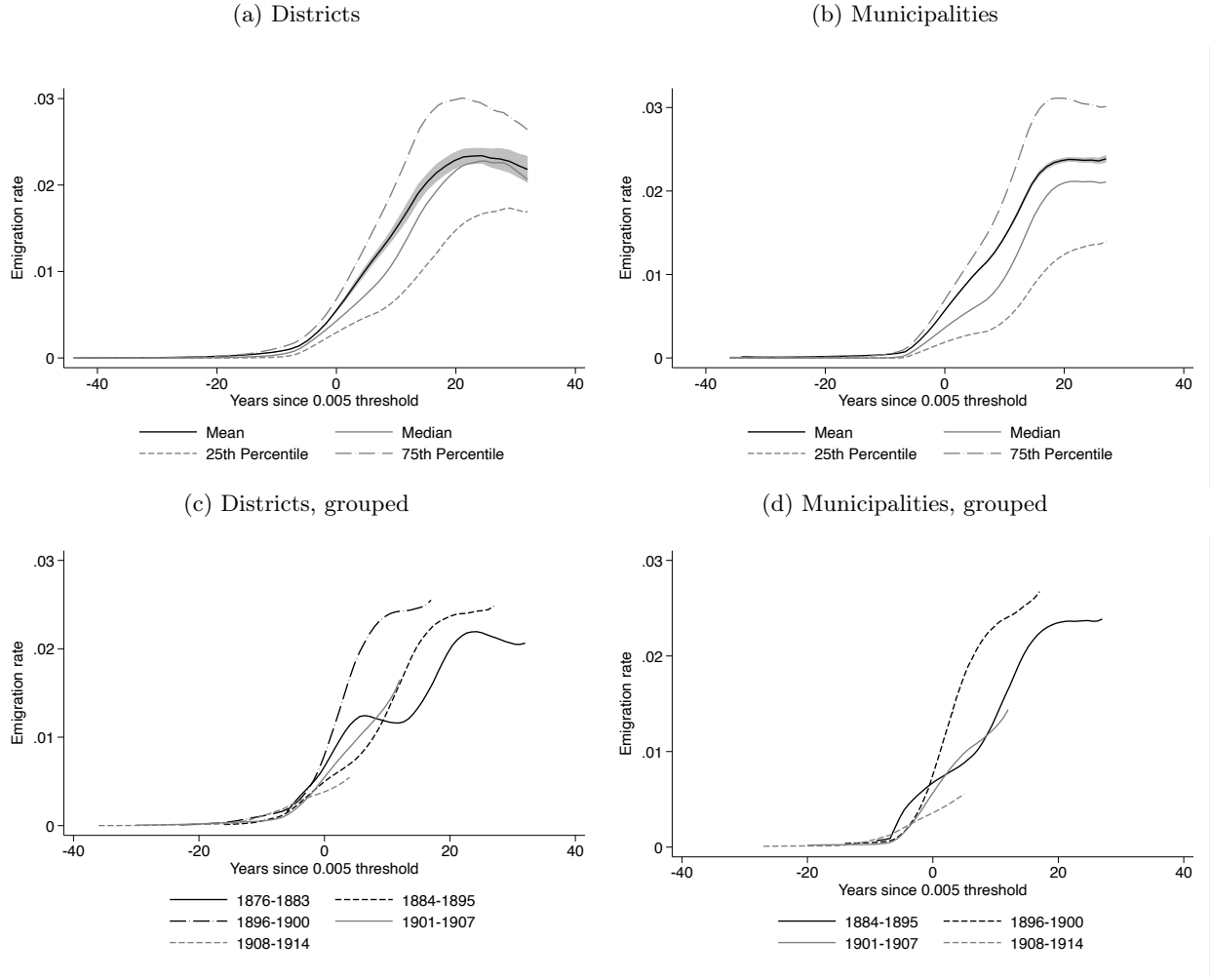


Figure H.3: S-shaped time series of migration to North America

Note: Panels (a) and (b) plot a non-parametric regression (the mean), as well as quartiles of emigration rates to North America against time, normalized so that year 0 is the first year in which a place had an emigration rate of at least 5 per thousand. Shaded areas are 95-percent confidence intervals for the mean. Panels (c) and (d) are the same as (a) and (b) but divide areas according to the half decade in which they crossed the threshold.

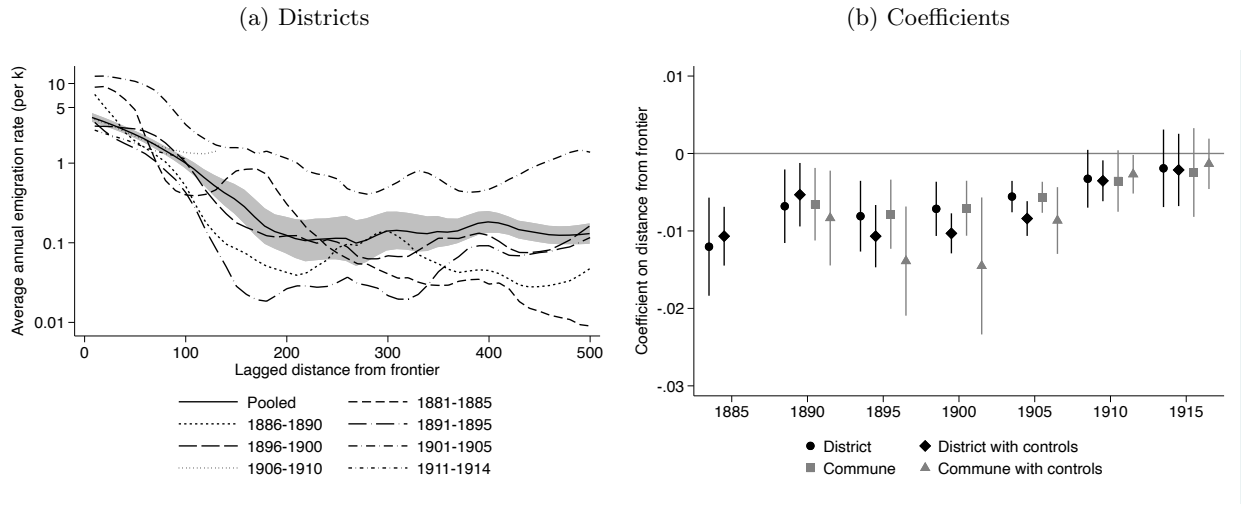


Figure H.4: Emigration rates to North America by distance to the mass migration frontier (km)

Note: Panel (a) presents non-parametric regressions of the log of average annual migration rates for the whole sample and for each half decade on the distance from a district that had ever achieved an average annual migration rate of at least 5 per thousand by the previous half decade, limiting the sample to districts that had not yet achieved this threshold. Shaded areas are 95-percent confidence intervals. Panel (b) estimates a binomial maximum likelihood regression of emigration rates on half decade-specific functions of lagged distance from the frontier of mass migration to North America and plots the coefficients on lagged distance from the frontier. Panel (b) also include regressions controlling for half decade-specific functions of various controls.

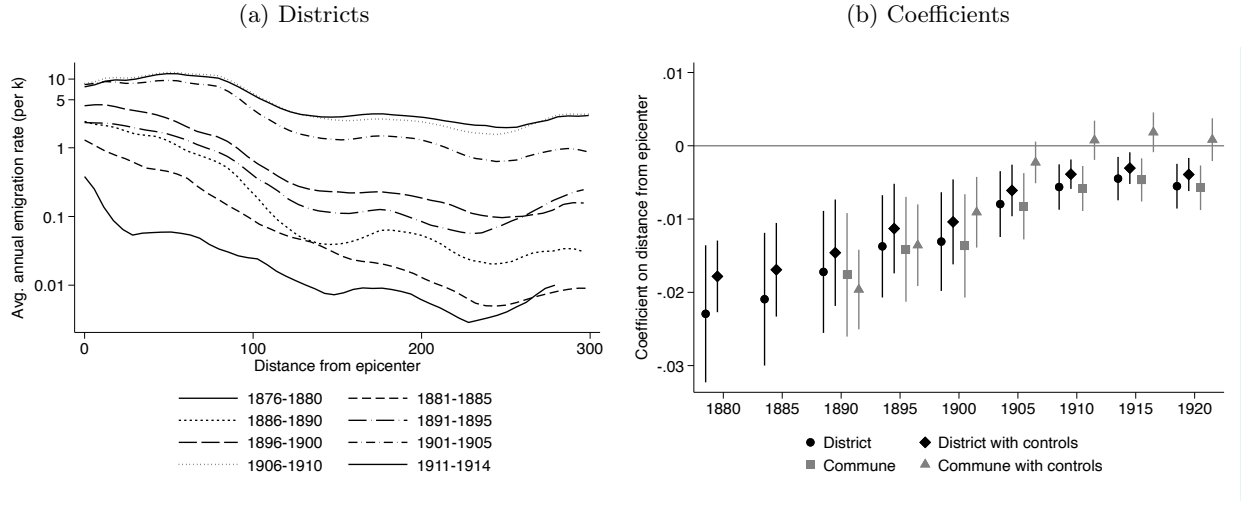


Figure H.5: Emigration rates to North America by distance to epicenter (km)

Note: Panel (a) plots non-parametric regressions of the log of the average annual emigration rate for each half decade against distance to the nearest epicenter of emigration to North America. Panel (b) estimates a binomial maximum likelihood regression of emigration rates on half decade-specific functions of distance from the nearest epicenter of emigration to North America and plots the coefficients on distance from epicenter. Panel (b) also includes regressions controlling for half decade-specific functions of various controls.

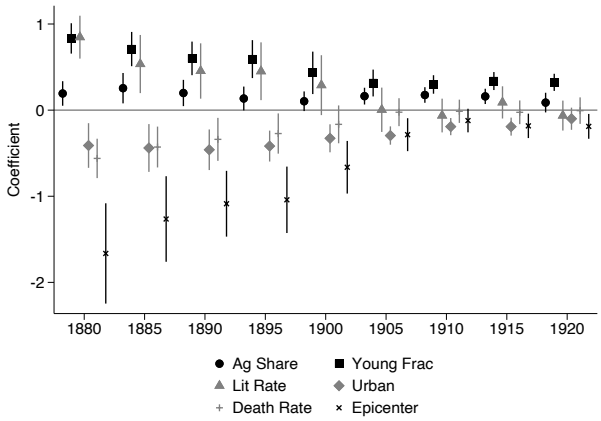


Figure H.6: Relationship of migration to various local characteristics

Note: This figure presents the results of a regression of emigration to any destination on year-specific functions of various district characteristics. All explanatory variables are standardized to have mean zero and standard deviation one.

Table H.1: Spatial contagion results

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Epicenter-based instrument								
Lagged Emigration Exposure	0.957^a	0.882^a	0.629^a	0.632^b	0.645^a	0.621^a	0.946^b	0.389
	(0.095)	(0.125)	(0.199)	(0.245)	(0.140)	(0.213)	(0.368)	(0.394)
Observations	31,573	31,573	31,573	31,573	31,573	31,572	31,573	31,537
F-statistic	35.88	41.18	16.07	9.913	24.60	10.41	25.09	6.645
Panel B: Frontier-based instrument								
Lagged Emigration Exposure	0.925^a	0.909^a	0.701^a	0.834^a	0.714^a	0.686^a	0.312	0.545^b
	(0.126)	(0.148)	(0.151)	(0.159)	(0.171)	(0.157)	(0.553)	(0.219)
Observations	11,372	11,371	11,371	11,370	11,371	11,361	11,371	11,335
F-statistic	78.87	76.82	73.46	53.58	56.10	52.12	19.99	30.95
Additional FE	None	None	C	CT	P	PT	D	DT
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Significance levels: ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1

Notes: Panel A uses instruments constructed on the basis of a municipality's distance to the nearest epicenter of mass migration. Panel B uses instruments constructed on the basis of distance to the frontier of mass migration. Sample limited to municipality-half decades between 50 and 250km of the migration source (i.e., the epicenter or frontier). Standard errors clustered at the district level. All specifications include at least half-decade fixed effects and control for half decade-specific functions of own predicted lagged emigration based on distance from the emigration source, local population, distance to coast, and distance to the European frontier. Dependent variable is the log of the emigration rate to North America. Unit of observation is a municipality-half decade. Controls include half decade-specific functions of latitude, longitude, elevation, agricultural employment share, industrial employment share, literacy rate, fraction under age 15, birth rate, death rate, mutual aid society membership per capita, and log postal savings deposits per capita. C denotes region (compartimento)-level fixed effects. P denotes province-level fixed effects. D denotes district-level fixed effects. CT, PT, and DT denote region-time, province-time, and district-time fixed effects.

I Results for Migration to All Destinations

This appendix repeats the main results of the paper, but focuses on migration to all destinations rather than on migration to North America alone. This addresses the concern that some of the local correlation in emigration rates could be the product of the fact that the emigration-by-destination data are available only at the province level. It focuses on distance to all epicenters rather than only distance to epicenters of emigration to North America and on the frontier of mass emigration to any destination rather than only on the frontier of mass migration to North America. The results are, for the most part, qualitatively unchanged, with two exceptions. The first concerns the S-shaped time series (Figure I.3). For the case of all destinations, these are not S-shaped, but continuously increasing after places cross the mass migration threshold. This may be the consequence of surges to several destinations combining to create a continuous increase, and shows that even in this case, the surge is rapid and virtually irreversible once mass emigration begins in a place. The second concerns the results of the spatial contagion model using the instrument based on distance to the frontier of mass migration (Panel B of Table I.1). In this case, all of the sample limitations imposed in the analysis, together with the loss of observations of zero emigration, yield a sample so small that statistical power is substantially weakened.

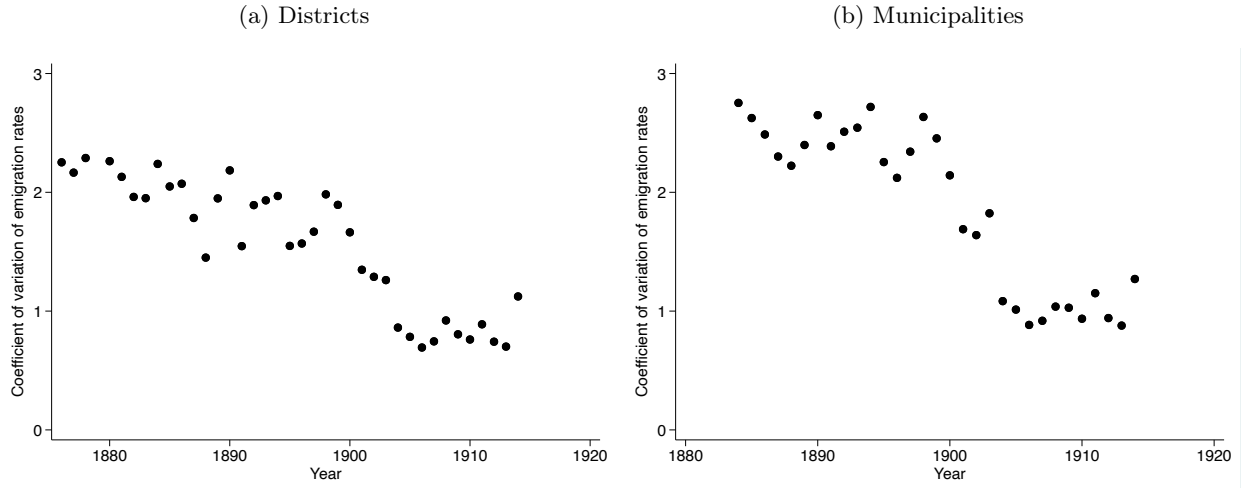


Figure I.1: σ-convergence in emigration rates to all destinations

Note: Each point represents the coefficient of variation in emigration rates to all destinations in a particular year.

(a) Districts

(b) Municipalities

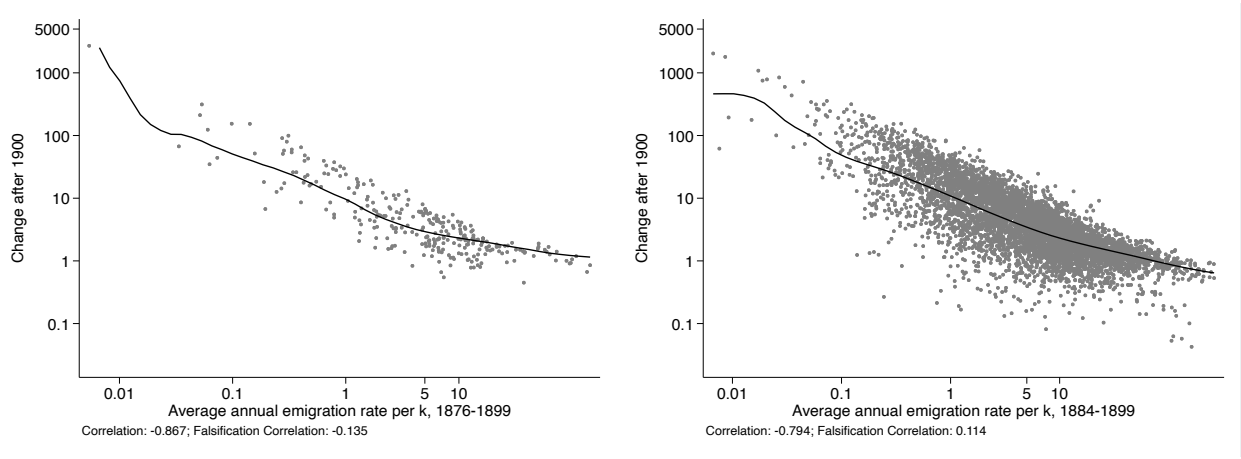


Figure I.2: β-convergence in emigration rates to all destinations

Note: Each point represents a municipality or district. The x-axis is the average annual emigration rate for a district for 1876–1899 or a municipality for 1884–1899 on a log scale. The y-axis is the ratio of the average emigration rate before and after 1900, also on a log scale. The falsification correlation is the correlation of the change in emigration and emigration after 1900; that it is not positive (or if it is, that its magnitude is considerably less than the plotted negative correlation) indicates that the negative relationship shown in the graphs is unlikely to be spurious, as explained in section 5.1.

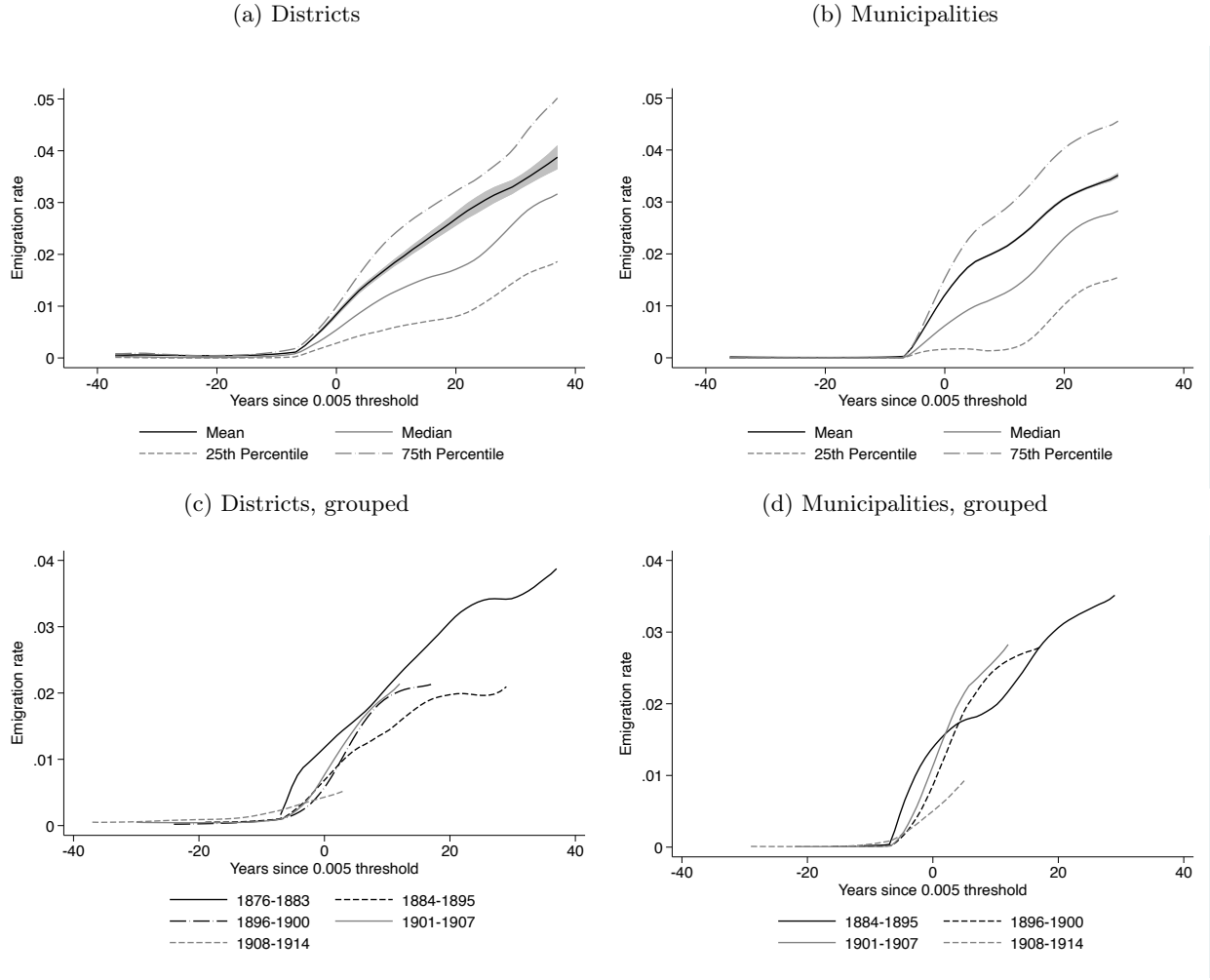


Figure I.3: (Non-)S-shaped time series of migration to all destinations

Note: Panels (a) and (b) plot a non-parametric regression (the mean), as well as quartiles of emigration rates to any destination against time, normalized so that year 0 is the first year in which a place had an emigration rate of at least 5 per thousand. Shaded areas are 95-percent confidence intervals for the mean. Panels (c) and (d) are the same as (a) and (b) but divide areas according to the half decade in which they crossed the threshold.

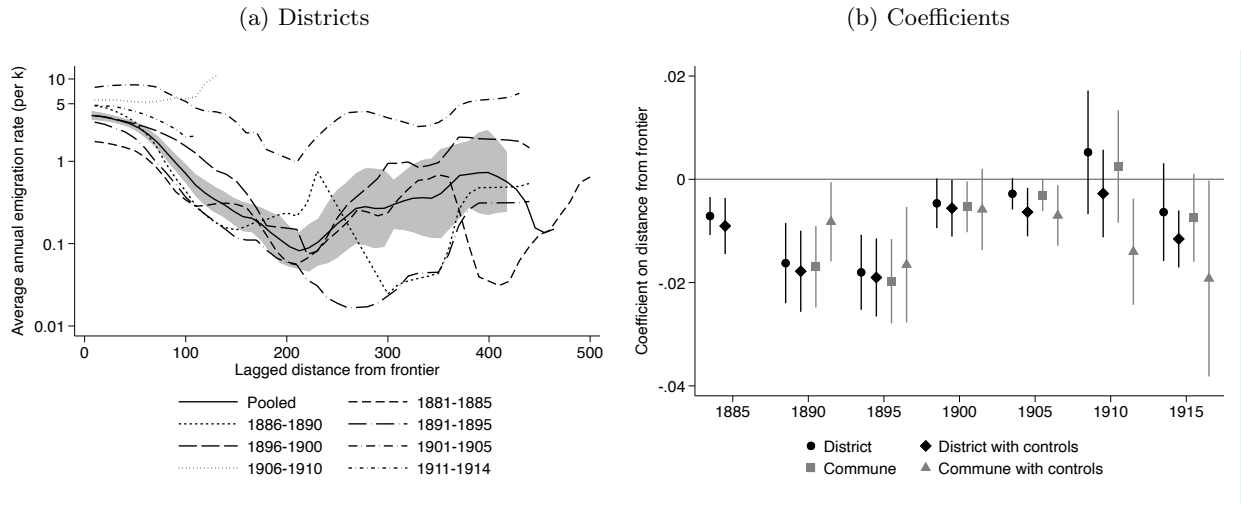


Figure I.4: Emigration rates to all destinations by distance to the mass migration frontier (km)

Note: Panel (a) presents non-parametric regressions of the log of average annual migration rates for the whole sample and for each half decade on the distance from a district that had ever achieved an average annual migration rate of at least 5 per thousand by the previous half decade, limiting the sample to districts that had not yet achieved this threshold. Shaded areas are 95-percent confidence intervals. Panel (b) estimates a binomial maximum likelihood regression of emigration rates on half decade-specific functions of lagged distance from the frontier of mass migration and plots the coefficients on lagged distance from the frontier. Panel (b) also includes regressions controlling for half decade-specific functions of various controls.

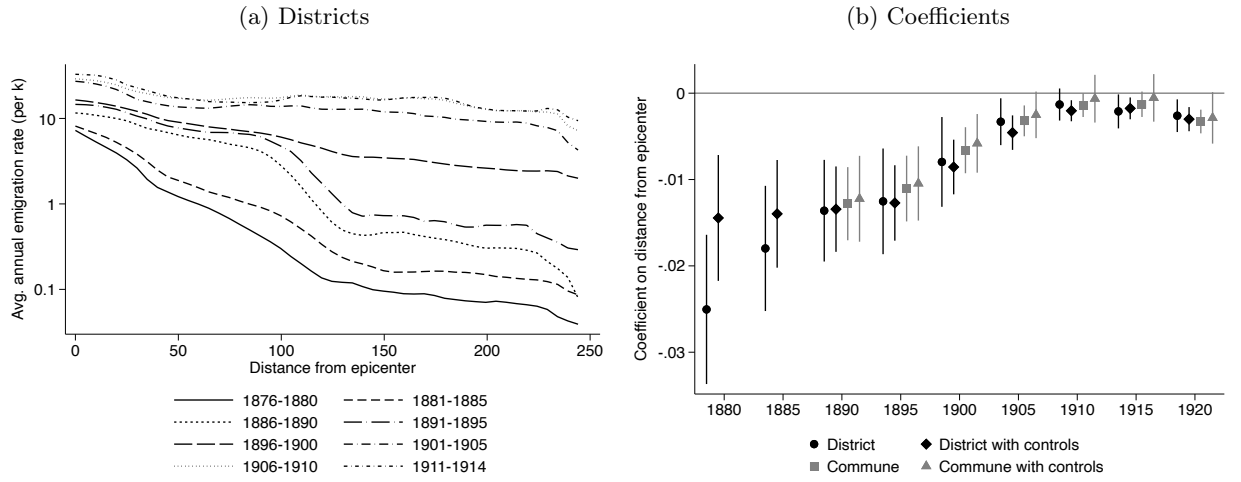


Figure I.5: Emigration rates to all destinations by distance to epicenter (km)

Note: Panel (a) plots non-parametric regressions of the log of the average annual emigration rate for each half decade against distance to the nearest epicenter of emigration. Panel (b) estimates a binomial maximum likelihood regression of emigration rates on half decade-specific functions of distance from the nearest epicenter of emigration and plots the coefficients on distance from epicenter. Panel (b) also includes regressions controlling for half decade-specific functions of various controls.

Table I.1: Spatial contagion results

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Epicenter-based instrument								
Lagged Emigration Exposure	−0.222	0.775[a]	0.647[a]	0.588[a]	0.715[a]	0.662[b]	0.460	0.630
	(1.268)	(0.142)	(0.225)	(0.223)	(0.229)	(0.328)	(0.326)	(0.451)
Observations	20,238	20,238	20,238	20,238	20,238	20,238	20,238	20,186
F-statistic	1.259	23.59	17.92	14.55	19.46	11.80	7.421	6.900
Panel B: Frontier-based instrument								
Lagged Emigration Exposure	0.315	0.303	0.267	0.483[c]	0.358	0.636[b]	0.210	0.024
	(0.283)	(0.276)	(0.335)	(0.265)	(0.322)	(0.289)	(0.461)	(0.381)
Observations	1,689	1,688	1,688	1,684	1,688	1,682	1,685	1,668
F-statistic	27.43	36.24	27.67	30.46	47.89	53.13	33.38	38.76
Additional FE	None	None	C	CT	P	PT	D	DT
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Significance levels: [a] p<0.01, [b] p<0.05, [c] p<0.1

Notes: Panel A uses instruments constructed on the basis of a municipality's distance to the nearest epicenter of mass migration. Panel B uses instruments constructed on the basis of distance to the frontier of mass migration. Sample limited to municipality-half decades between 50 and 250km of the migration source (i.e., the epicenter or frontier). Standard errors clustered at the district level. All specifications include at least half-decade fixed effects and control for half decade-specific functions of own predicted lagged emigration based on distance from the emigration source, local population, distance to coast, and distance to the European frontier. Dependent variable is the log of the emigration rate to any destination. Unit of observation is a municipality-half decade. Controls include half decade-specific functions of latitude, longitude, elevation, agricultural employment share, industrial employment share, literacy rate, fraction under age 15, birth rate, death rate, mutual aid society membership per capita, and log postal savings deposits per capita. C denotes region (compartimento)-level fixed effects. P denotes province-level fixed effects. D denotes district-level fixed effects. CT, PT, and DT denote region-time, province-time, and district-time fixed effects.

J Robustness Checks for β-Convergence

Table J.1 provides a more detailed test of the β-convergence prediction than the one presented in Figure 8, including exercises that address concerns over spurious correlations. We estimate analogous regressions of the form

$$\log\left(\frac{e_{i1}}{e_{i0}}\right) = \alpha + \beta \log(e_{i0}) + \mathbf{x}_i'\gamma + \varepsilon_i, \tag{J.1}$$

where e_{i0} and e_{i1} are the rates of emigration from municipality i before and after 1900, respectively, and $\mathbf{x}_i$ is a vector of controls. Column (1) performs this estimation with no added controls, essentially replicating the findings of Figure 8. Columns (2)–(4) of Table 3 repeat the same estimation with the addition of a variety of control variables and then of province and district fixed effects. These four columns include, under "Falsification," the coefficient from repeating the estimation of equation (J.1) with emigration in the second period, $\log(e_{i1})$, as the regressor. The fact that this coefficient is positive in columns (3) and (4) indicates that there is some merit to concerns that the relationship may in part be the product of measurement error or idiosyncratic shocks; but since the absolute value of the falsification coefficient is more than 4 times smaller, we can still conclude that, even if the coefficient is biased in this way, the convergence is not fully spurious. Figure 8 shows the results of a similar robustness check. The negative coefficient in this case for a regression of the change in the emigration rate and the emigration rate in the *later* period (the "Falsification Correlation" in the notes to the figures) implies that the relationship is indeed driven by convergence, even if the slope is downward biased. If the true relationship were zero, then this falsification coefficient would be positive and of the same absolute magnitude.[99]

In columns (5)–(8) of Table J.1, we take a direct approach to addressing concerns of a spuriously negative relationship. Specifically, we use distance from the nearest epicenter of emigration to North America as an instrument for $\log(e_{i0})$. This strategy exploits the fact that emigration expanded spatially from these initial epicenters. To be clear, the object of this analysis is not to identify a causal effect, but to clear the possible source of spurious correlation discussed above. Although the first-stage F-statistics become weak in the most restrictive specifications, the coefficients are statistically significant and closely match those of the OLS regressions. This strongly suggests that the strong β-convergence was not a result of the suspected mechanical bias.

[99] See Spitzer (2021) and Spitzer, Tortorici, and Zimran (2022) for similar analyses.

Table J.1: β-convergence

Variables	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) IV	(6) IV	(7) IV	(8) IV
Lagged Own Emigration	-0.521^a	-0.739^a	-0.831^a	-0.855^a	-0.514^a	-0.793^a	-0.823^a	-0.913^a
	(0.021)	(0.017)	(0.012)	(0.012)	(0.043)	(0.062)	(0.157)	(0.208)
Observations	5,856	5,855	5,855	5,855	5,856	5,855	5,855	5,849
R-squared	0.592	0.770	0.882	0.915	0.592	0.766	0.801	0.823
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
1st Stage F					55.107	94.214	11.507	7.449
FE	None	None	P	D	None	None	P	D
Falsification	−0.150	−0.010	0.183	0.176				
	(0.054)	(0.073)	(0.065)	(0.057)				

Significance levels: ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1

Notes: Standard errors clustered at the district level. Unit of observation is a municipality. Dependent variable is the change in the log of the emigration rate to North America from the pre-1900 period to the period 1900 and later. Controls include latitude, longitude, elevation, agricultural employment share, industrial employment share, literacy rate, fraction under age 15, distance to railroad, birth rate, death rate, mutual aid society members per capita, and log postal savings deposits per capita. Instrument is the distance to the nearest epicenter of emigration to North America. P denotes province-level fixed effects included. D denotes district-level fixed effects included. The falsification coefficient is the coefficient from regressing the change in emigration on emigration in the post-1900 period; if it is either negative or positive but of a smaller magnitude than the main coefficient of interest, this is evidence that the relationship is not spurious.

K Results Including Observations with No Migration

Whenever the object of interest is the logarithm of emigration, municipalities with no emigration in a particular half decade must be excluded. In this appendix, we repeat the main results using $\log(e_{it} + \varepsilon)$, where $\varepsilon = 10^{-5}$ instead of $\log(e_{it})$ in order to incorporate these municipality-half decades into the analysis. This is not necessary when the binomial maximum likelihood regression is used (since that is designed to account for cases of zero migration), and so this appendix only repeats the results where the change is necessary. The results are, for the most part, qualitatively unchanged. The results using the epicenter distance-based instrument (Panel A of Table K.1) are somewhat weaker than those of the main text, however, and in some cases the combination of a decline in the magnitude of the coefficient and an increase in the standard error has caused some estimates to cross past conventional levels of statistical significance. On the whole, however, these results, in particular when viewed in combination with those of Panel B of Table K.1, are consistent with those of the main text.

Table K.1: Spatial contagion results

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Epicenter-based instrument								
Lagged Emigration Exposure	0.725[a]	0.763[a]	0.429[c]	0.432	0.410[c]	0.424	0.214	0.125
	(0.134)	(0.131)	(0.250)	(0.272)	(0.225)	(0.296)	(0.474)	(0.441)
Observations	37,128	37,128	37,128	37,128	37,128	37,128	37,128	37,104
F-statistic	29.04	47	18.60	13.12	24.75	10.38	27	8.256
Panel B: Frontier-based instrument								
Lagged Emigration Exposure	0.723[a]	0.950[a]	0.675[a]	0.775[a]	0.703[a]	0.611[a]	0.268	0.479[c]
	(0.185)	(0.213)	(0.207)	(0.198)	(0.223)	(0.208)	(0.537)	(0.271)
Observations	12,877	12,875	12,875	12,873	12,875	12,864	12,875	12,850
F-statistic	60.61	48.56	45.66	42.29	41.83	40.97	30.60	23.06
Additional FE	None	None	C	CT	P	PT	D	DT
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Significance levels: [a] p<0.01, [b] p<0.05, [c] p<0.1

Notes: Panel A uses instruments constructed on the basis of a municipality's distance to the nearest epicenter of mass migration. Panel B uses instruments constructed on the basis of distance to the frontier of mass migration. Sample limited to municipality-half decades between 50 and 250km of the migration source (i.e., the epicenter or frontier). Standard errors clustered at the district level. All specifications include at least half-decade fixed effects and control for half decade-specific functions of own predicted lagged emigration based on distance from the emigration source, local population, distance to coast, and distance to the European frontier. Dependent variable is the log of the emigration rate to North America. Unit of observation is a municipality-half decade. Controls include half decade-specific functions of latitude, longitude, elevation, agricultural employment share, industrial employment share, literacy rate, fraction under age 15, birth rate, death rate, mutual aid society membership per capita, and log postal savings deposits per capita. C denotes region (compartimento)-level fixed effects. P denotes province-level fixed effects. D denotes district-level fixed effects. CT, PT, and DT denote region-time, province-time, and district-time fixed effects.

L Robustness to Choice of θ

This appendix verifies the robustness of the results in section 6 to alternative choices of the parameter θ, which governs the rate at which the influence of other municipalities on the emigration exposure of a municipality declines with the distance between them. In particular, two alternate values of θ are considered. Whereas that in the main text was chosen on the basis of estimating equation (2) by non-linear least squares without controls, the alternate values in this appendix were chosen after performing this estimation with controls and with province-half decade fixed effects. The results are qualitatively unaffected.

L.1 θ=-2.86

Table L.1: Spatial contagion results

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Epicenter-based instrument								
Lagged Emigration Exposure	0.718^a	0.827^a	0.596^a	0.582^b	0.567^a	0.459^c	0.955^b	−0.095
	(0.182)	(0.127)	(0.185)	(0.234)	(0.149)	(0.247)	(0.452)	(0.577)
Observations	31,463	31,463	31,463	31,463	31,463	31,462	31,463	31,427
F-statistic	11.92	33.05	15.19	8.903	19.93	7.495	18.99	4.241
Panel B: Frontier-based instrument								
Lagged Emigration Exposure	0.769^a	0.835^a	0.620^a	0.731^a	0.660^a	0.600^a	0.168	0.343
	(0.179)	(0.163)	(0.164)	(0.170)	(0.183)	(0.163)	(0.652)	(0.242)
Observations	11,207	11,206	11,206	11,205	11,206	11,196	11,206	11,170
F-statistic	28.90	42.32	47.11	36.80	35.56	32.87	14.22	20.05
Additional FE	None	None	C	CT	P	PT	D	DT
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Significance levels: ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1

Notes: Panel A uses instruments constructed on the basis of a municipality's distance to the nearest epicenter of mass migration. Panel B uses instruments constructed on the basis of distance to the frontier of mass migration. Sample limited to municipality-half decades between 50 and 250km of the migration source (i.e., the epicenter or frontier). Standard errors clustered at the district level. All specifications include at least half-decade fixed effects and control for half decade-specific functions of own predicted lagged emigration based on distance from the emigration source, local population, distance to coast, and distance to the European frontier. Dependent variable is the log of the emigration rate to North America. Unit of observation is a municipality-half decade. Controls include half decade-specific functions of latitude, longitude, elevation, agricultural employment share, industrial employment share, literacy rate, fraction under age 15, birth rate, death rate, mutual aid society membership per capita, and log postal savings deposits per capita. C denotes region (compartimento)-level fixed effects. P denotes province-level fixed effects. D denotes district-level fixed effects. CT, PT, and DT denote region-time, province-time, and district-time fixed effects.

L.2 θ=-2.98

Table L.2: Spatial contagion results

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Epicenter-based instrument								
Lagged Emigration Exposure	1.109^a	0.928^a	0.710^a	0.691^a	0.626^a	0.477^c	1.011^b	0.380
	(0.081)	(0.119)	(0.177)	(0.210)	(0.163)	(0.257)	(0.394)	(0.393)
Observations	31,463	31,463	31,463	31,463	31,463	31,462	31,463	31,427
F-statistic	52.15	51.83	23.03	15.40	22.24	9.397	26.21	9.081
Panel B: Frontier-based instrument								
Lagged Emigration Exposure	1.079^a	0.930^a	0.650^a	0.769^a	0.674^a	0.664^a	0.304	0.485^b
	(0.119)	(0.155)	(0.160)	(0.166)	(0.174)	(0.166)	(0.539)	(0.231)
Observations	11,207	11,206	11,206	11,205	11,206	11,196	11,206	11,170
F-statistic	96.83	65.04	69.36	53.38	60.59	54.27	25.23	35.81
Additional FE	None	None	C	CT	P	PT	D	DT
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Significance levels: ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1

Notes: Panel A uses instruments constructed on the basis of a municipality's distance to the nearest epicenter of mass migration. Panel B uses instruments constructed on the basis of distance to the frontier of mass migration. Sample limited to municipality-half decades between 50 and 250km of the migration source (i.e., the epicenter or frontier). Standard errors clustered at the district level. All specifications include at least half-decade fixed effects and control for half decade-specific functions of own predicted lagged emigration based on distance from the emigration source, local population, distance to coast, and distance to the European frontier. Dependent variable is the log of the emigration rate to North America. Unit of observation is a municipality-half decade. Controls include half decade-specific functions of latitude, longitude, elevation, agricultural employment share, industrial employment share, literacy rate, fraction under age 15, birth rate, death rate, mutual aid society membership per capita, and log postal savings deposits per capita. C denotes region (compartimento)-level fixed effects. P denotes province-level fixed effects. D denotes district-level fixed effects. CT, PT, and DT denote region-time, province-time, and district-time fixed effects.

M Results With Grid Fixed Effects

This appendix repeats the estimation of section 6, but instead of using fixed effects based on actual geographic divisions (i.e., region, province, and district), we use fixed effects for grids of various sizes, ranging from grid cells of 90-by-90 kilometers to 15-by-15 kilometers. This method, based on that used by Barsbai et al. (2017), is intended to show that the estimates are not the product of bias caused by unobservables by making a coefficient stability argument—if the coefficients are largely unchanged in the face of fixed effects for finer and finer grid cells, it is unlikely that local characteristics are responsible for the relationship. Although the results for the epicenter-based instrument are in many cases rendered statistically insignificant by these very fine controls, the results for the frontier-based instrument are robust, and moreover are largely stable across specifications.

Table M.1: Spatial contagion results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables								
Panel A: Epicenter-based instrument								
Lagged Emigration Exposure	0.524^{c}	0.385	0.584^{a}	0.429	0.622^{a}	0.682	0.708^{a}	1.766^{c}
	(0.297)	(0.768)	(0.218)	(0.624)	(0.176)	(0.636)	(0.157)	(0.926)
Observations	31,463	31,435	31,463	31,409	31,463	31,280	31,463	30,341
F-statistic	12.32	2.087	24.07	3.466	37.96	3.028	49.69	3.968
Panel B: Frontier-based instrument								
Lagged Emigration Exposure	0.664^{a}	0.777^{a}	0.820^{a}	0.829^{a}	0.625^{c}	0.684^{b}	0.987^{c}	0.579
	(0.197)	(0.186)	(0.228)	(0.191)	(0.328)	(0.298)	(0.572)	(0.470)
Observations	11,205	11,182	11,202	11,166	11,199	11,093	11,159	10,674
F-statistic	49.37	53.63	34.57	59.16	25.40	36.49	15.26	26.68
Additional FE	G	GT	G	GT	G	GT	G	GT
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grid Size	90	90	60	60	30	30	15	15

Significance levels: ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1

Notes: Panel A uses instruments constructed on the basis of a municipality's distance to the nearest epicenter of mass migration. Panel B uses instruments constructed on the basis of distance to the frontier of mass migration. Sample limited to municipality-half decades between 50 and 250km of the migration source (i.e., the epicenter or frontier). Standard errors clustered at the district level. All specifications include at least half-decade fixed effects and control for half decade-specific functions of own predicted lagged emigration based on distance from the emigration source, local population, distance to coast, and distance to the European frontier. Dependent variable is the log of the emigration rate to North America. Unit of observation is a municipality-half decade. Controls include half decade-specific functions of latitude, longitude, elevation, agricultural employment share, industrial employment share, literacy rate, fraction under age 15, birth rate, death rate, mutual aid society membership per capita, and log postal savings deposits per capita. G denotes the inclusion of grid fixed effects for grid cells of the specified size4 (e.g., 90-by-90km). GT denotes the inclusion of grid-half decade fixed effects.

References

Barde, Robert, Susan B. Carter, and Richard Sutch (2006). "Table Ad106–120: Immigrants, by country of last residence—Europe, 1820–1997." In *Historical Statistics of the United States*. Susan B. Carter, Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright (ed.). Cambridge: Cambridge University Press, pp. 1.560–1.563.

Barsbai, Toman, Hillel Rapoport, Andreas Steinmayr, and Christoph Trebesch (2017). "The Effect of Labor Migration on the Diffusion of Democracy: Evidence from a Former Soviet Republic." *American Economic Journal: Applied Economics* 9:3, pp. 36–69.

Brandenburg, Broughton (1904). *Imported Americans: The Story of the Experiences of a Disguised American and His Wife Studying the Immigration Question*. New York: Frederick A. Stokes Company.

Ferenczi, Imre and Walter F. Willcox (1929). *International Migrations*. New York: National Bureau of Economic Research.

Hatton, Timothy J. and Jeffrey G. Williamson (1998). *The Age of Mass Migration: Causes and Economic Impact*. New York: Oxford University Press.

ISTAT (2018). *Confini della unità amministrative a fini statistici al 1 gennaio 2018* [machine-readable database]. Rome: Istituto Nazionale di Statistica.

——— (2019). *Confini statistico-amministrativi: Analisi storica*. Rome: Istituto Nazionale di Statistica.

Jet Propulsion Laboratory (2014). *Shuttle Radar Topography Mission* [machine-readable database]. Pasadena: California Institute of Technology.

Spitzer, Yannay (2021). "Pogroms, Networks, and Migration: The Jewish Migration from the Russian Empire to the United States 1881–1914." Mimeo., Hebrew University of Jerusalem.

Spitzer, Yannay, Gaspare Tortorici, and Ariell Zimran (2022). "International Migration Responses Modern Europe's Most Destructive Earthquake: Messina and Reggio Calabria, 1908." NBER Working Paper 27506.