ONLINE APPENDIX

Emigration Restrictions and Economics Development

Evidence from the Italian Mass Migration to the U.S.

Davide M. COLUCCIA & Lorenzo SPADAVECCHIA

September, 2024

A DATA APPENDIX

A.I DATA SOURCES

This section lists the sources, methodology, and coverage of the data that we assemble. We defer a more detailed description of the emigration data to section A.III. Table A.1 summarizes the content of this section.

Population Censuses. We extract information from population censuses in 1881, 1901, 1911, 1921, 1931, and 1936. No census was taken in 1891. We digitize district-level information on the number of people employed in all major sectors in 1881–1921 censuses. The same information is available at the municipality level for 1931 and 1936. Hence we aggregate it at the district level. To assign municipalities to districts, we geo-code each municipality and overlay the coordinates to the 1921 district shape file. From population censuses, we also extract information on employment by industry: this is available by districts until 1921 but only by provinces in 1931 and 1936. We thus aggregate industry employment to provinces. Population data have been tabulated by ISTAT for each municipality. We aggregate them at the district and province levels. We also code two types of urbanization variables: the share of the population living in districts with at least *k*-thousand municipalities and the share of municipalities with at least *k*-thousand inhabitants. The ISTAT population tables report information on the area, altitude, and access to the sea for each municipality.

Manufacture Censuses. Manufacture censuses were taken in 1911, 1927, and 1937. They report province-level information on the universe of firms. We digitize data on a set of proxies for capital investment: the number of firms, the number of firms with at least one installed engine, the number of installed mechanical and electrical engines, and the number of installed mechanical and electrical horsepower.

World War One Deaths. WW1 deaths are from `cadutigrandeguerra.it`, a dataset maintained by ISTORECO, a branch of the *Associatione Nazionale Partigiani d'Italia* (ANPI). The underlying data were collected by the Fascist regime for propaganda purposes. The honor roll call contains information on 570,355 Italian soldiers who lost their lives during the war. The data appears to be comprehensive since most estimates place the total death toll at 650,000. For each individual, we know the name and surname, the birth year, the municipality of origin—which we map to the municipalities listed in the ISTAT data—, the military rank at death, the regiment, when, where, and why they died, and the decoration, if any. Deaths span the war years (1915–1918). We aggregate them at the district and province levels to have a cross-sectional indicator of mortality.

Railway Data. We reconstruct the network of Italian railways between 1839 and 1926 from a volume curated by the Italian Statistical Office and edited in 1927. This is the first paper using these data. The unit of observation in the data is truck lines connecting two stations. We have information on the precise date when each line opened and the distance it covered, and the name of the station it connected. We geo-code the location of each station and generate a dummy variable that indicates whether a given municipality had at least one station, as well as a count variable returning the number of stations. We also define a variable that returns the number of kilometers that separate every given municipality from the closest transatlantic migration port. Calabrese (2017) suggests that railway access to Genoa, Palermo, or Naples—the only ports with ships sailing toward the United States—was a crucial condition to ignite migration movements. We thus compute the shortest path connecting each municipality with each transatlantic port given the state of the railway network in every given year and take the minimum among the three. We aggregate this variable by district and province, taking a population-weighted average of the shortest railway distances to emigration ports.

A.II Construction of the Sample

All variables that are not computed from geo-coded data are cross-walked to consistent 1921 district and border geographies using the method described by Eckert *et al.* (2020). To do so, we use GIS boundary files publicly provided by ISTAT for each census year. Even though this yields quantitatively minor corrections, it is important to ensure that geographies remain consistent because the Fascist regime undertook an extensive revision of local government divisions, which ultimately abolished districts in 1927.

We forcibly exclude areas that were annexed as a consequence of World War One in 1918—Trento and Trieste, Südtirol, Istria, and Zara—because we do not observe pre-Quota outcomes in those regions.

Unlike for US emigrants, we do not have district-level data on overall emigration. To assign province-level emigration to districts, we assume that emigration rates were constant within each province. We then impute district emigration by multiplying province emigration by the share of inhabitants in each district compared to the province population before the mass migration (in 1881).

We assemble three distinct samples (Samples 1, 2, and 3). Sample 1, which comprises population and employment-by-sector data, is at the district level and covers the years 1881, 1901, 1911, 1921, 1931, and 1936. Agriculture employment is not available in 1931. All data in Sample 1 are either digitized from population censuses or are aggregated from data tabulated by ISTAT. Sample 2 runs at the province level, covers capital variables, and is available in 1911, 1927, and 1937. Capital variables are retrieved from manufacturing censuses. Sample 3 runs at the province level, covers employment-by-industry variables, and spans the years 1901, 1911, 1921, 1927, and 1936. For the first three decades, the data are from population censuses; for the latter two, we digitize them from manufacturing censuses. Since capital variables exhibit substantial skewness, we winsorize them at the 5% level.

A.III Details on the Emigration Data

In this section, we document in detail the emigration data that we collect. The raw data can be found at `https://heritage.statueofliberty.org/`. First, we describe the methodology that we adopt to assemble the data. Second, we show how to validate this dataset with external sources. Last, we present some stylized facts that the new data allow us to document.

This dataset responds to a key limitation of commonly used US census data (Ruggles *et al.*, 2021). These list the country of origin of immigrants residing in the US, but they do not report where immigrants originated from within their home country. This issue applies to all countries. Hence separate papers developed strategies to reconstruct such information for, among others, Norway (Abramitzky *et al.*, 2014) and England (Coluccia and Dossi, 2023). This paper looks at emigrants from Italy, a major emigration country among the so-called "second-wave" nations.

A.III.1 Methodology

We run queries over a comprehensive set of the most common 20,000 Italian surnames between 1890 and 1930. We collect individual-level information on the name and surname of immigrants, their municipality of origin, their immigration year, and whether they can read or write.

The municipality of origin is recorded consistently only between 1892 and 1924. Names, surnames, and municipalities are frequently coded with spelling errors, possibly because they were recorded by American enumerators. In this paper, we are interested in the municipality of origin of immigrants. We tackle this data quality issue in two steps. First, we pick the 1,000 most common origin municipalities in the data, and we correct eventual coding errors in those manually. We also discard entries that are too coarse, such as "Italy" or "Sicily." Then, we geo-code the remaining entries using Google Maps' auto-correction algorithm. Then, we manually checked that the return geo-coded locations are reliable for a subset of 200 municipalities. The algorithm successfully matches 189 out of 200 municipalities. The remaining 11 are impossible to match even by hand. We assess the plausibility of this matching exercise in section A.III.2.

The municipality of origin is missing for a non-negligible sub-sample of immigrants. In Figure A.1, we report graphical evidence. The left panel reports the absolute number of recorded immigrants (in blue) and those that we match to a municipality. The right panel reports the share of immigrants with at least one listed origin (in blue) and the share that we match to a municipality. Municipalities after the 1924 Immigration Act were seldom recorded, but we never use this sample period in this paper.

In the analysis, the dataset is aggregated by district or province depending on the part of the analysis using boundaries in 1921 from historical shape files provided by ISTAT.

A.III.2 Validation

The granular nature of the dataset implies that we cannot validate it with existing data at similar levels of aggregation. Our strategy, instead, is to aggregate it at the regional level and compare it with data from official statistics on Italian emigration to the United States collected by the Commissioner General for Emigration. These span the period 1877–1925 and are available by region.

In Table A.2, we report the correlation between the Ellis Island dataset and US emigration as recorded in official statistics. In columns (1–5), we report the correlation between the raw series, while in columns (6–7), we take logs. We find a positive and large unconditional correlation between the two (columns 1 and 6). In particular, Ellis Island migration explains more than 80% of the region-level variation in US emigration as measured in official statistics. This correlation remains conditioning on year (2 and 7), region (3 and 8), and year and region (4 and 9) fixed effects. Importantly it holds within the sub-sample period we use to compute the treatment (5 and 10). Figure A.2 displays the unconditional correlation between the two series (A.2a) and the one absorbing region and year fixed effects (A.2b). These exercises document a positive, large, and statistically significant correlation between the Ellis Island US emigration and data from official statistics. Finally, in Figure A.3, we check that the correlation remains high in each year of the observation sample. Each dot in the figure reports the correlation between the two series in one year between 1892 and 1924. The correlation remains stable, positive, and statistically significant throughout the sample period.

A.III.3 Stylized Facts

Dissecting the specific features of mass emigration to the United States is beyond the scope of this paper. Instead, we present two suggestive facts.

First, in Table A.3, we list the districts that were relatively more exposed to the US migration. In columns (1–3), districts are ranked by the absolute number of emigrants. In columns (4–6), we rank them by the emigration rate, expressed as the ratio between overall US emigrants and the 1921 population. The vast majority of top-ranked districts are located in Southern regions. Emigration rates are higher in Sicily and Campania. The district of Palermo alone accounts for almost 90,000 emigrants out of a population of 850,000.

We then provide evidence supporting the S-hypothesis advanced by Gould (1980) and recently analyzed by Spitzer and Zimran (2023). This maintains that local out-migration patterns followed a logistic-type dynamic, with initially low uptake, large increases in a relatively short time period, and subsequent plateau. Gould (1980) connects these dynamics to information diffusion within the population. To test this hypothesis, we mark the beginning of the mass migration in each district when the US emigration rate exceeded 0.1‰. This generates a setting akin to a staggered treatment roll-out. We then use the method of Borusyak *et al.* (2022) to estimate the dynamics of US out-migration. Importantly, this approach ensures that we compare emigration districts with areas where the migration had not already begun. We find that emigration followed Gould's S-shaped pattern as documented by Spitzer and Zimran (2023). The event-study figures associated with the resulting model are listed in A.4 and show that out-migration follows an S-shaped pattern as argued by Gould (1980). We interpret this finding as additional evidence supporting the quality of the underlying data.

A.III.4 Linked Sample

To produce the instrumental variable for Quota Exposure, we require information on the origin district and province of Italian immigrants by US county. This information is not available in the US census or reported in the Ellis Island data. To circumvent this limitation, we link the full-count non-anonymized US population census (Ruggles *et al.*, 2021) and the Ellis Island records. To the best of our knowledge, ours is the first attempt to produce a linked sample between these two uniquely rich sources.

The algorithm builds on similar automated linking procedures (e.g., Abramitzky *et al.*, 2021). First, we translate the names of Italian-born individuals recorded in the US census to their Italian version.[29] For each record in the Ellis Island dataset i who immigrates in year t_i, we pick the set of Italian-born individuals J in the 1900 US census whose initial name and surname Soundex-adjusted letters are the same as i's and whose immigration year recorded in the US census is in the window $[t_i - 1, t_i + 1]$.[30] We then compute the Monge-Elkan similarity with Jaro-Winkler inner word distance between the name and surname of i and those of the individuals in J. Among those, we pick the j's with the highest name and surname similarity as potential matches. If both the maximal name and surname similarities are above a given quality threshold, which in the baseline exercise is set at .9, the match(es) is (are) accepted; otherwise, they are discarded.

In Figure A.5, we report the distribution of name and surname similarities for the sample of individuals with at least one potential match. There is substantial mass at 1, where matches are literal. In Figure A.6, we report the share of Ellis Island records with at least one match, in blue, and one *accepted* match, in red. The gross matching rate remains constant throughout the sample at approximately 50%. There are several reasons why someone recorded at Ellis Island may not appear in the 1900 census. First, that person may have left before 1900. Second, women could change their surname. Alternatively, Italians could choose to change their surname as an assimilation effort. If the name change did not simply consist of a translation, we would fail to detect this practice. Finally, the immigration year in the census may be coded with errors. Of the 50% fraction with at least one match, between one-third and one-half presents at least one accepted match with sufficiently high

[29]For instance, we convert "Peter" to "Pietro." This procedure ensures that Anglicizations of Italian names that occur in the US census but not in the Ellis Island records do not artificially deflate our matching rate.

[30]We use the Soundex-adjusted algorithm to ensure that different spellings with similar phonetics are treated in the same way. For instance, the Soundex-adjusted initial of "Katherine" and "Catherine" is encoded as the same hard "c."

quality. The resulting 22% matching rate is not very distant from benchmark rates for intergenerational linked samples using US census data (Abramitzky *et al.*, 2021).

A key concern for the empirical strategy is that the probability of matching individuals from the Ellis Island data is correlated with their area of origin. This possibility would induce selection in the resulting linked sample, thus ultimately invalidating the relevance of the shift-share instrument. While we provide quantitative evidence to support the relevance of the first stage, we can check whether any systematic selection pattern emerges directly from the linked data. In Figure A.7, we report the correlation between the probability of matching and the origin region (panel A.2a) and province (panel A.2b) of immigrants. There appears to be no systematic selection pattern. Individuals from Calabria and Sicily are marginally more likely to be matched, even though this difference likely reflects a larger sample size of emigrants from those regions. Overall, the resulting instrumental variable retains a considerable correlation with the observed emigration outflows.

A.IV TABLES

TABLE A.1. SUMMARY OF THE DATA SOURCES AND COVERAGE

Variable (1)	Observation Unit (2)	Source (3)	Observed Years (4)
Panel A. Demographics			
Population	District (1881-1921), Municipality (1931-1936)	Population Censuses	1881-1936, excl.1891
Area	District (1881-1921), Municipality (1931-1936)	Population Censuses	1881-1936, excl.1891
Urbanization	District (1881-1921), Municipality (1931-1936)	Population Censuses	1881-1936, excl.1891
Literacy	Municipality	Population Censuses	1881-1936, excl.1891
Panel B. Employment, by Sector			
Manufacture	District (1881-1921), Municipality (1931-1936)	Population Censuses	1881-1936, excl.1891
Agriculture	District (1881-1921), Municipality (1931-1936)	Population Censuses	1881-1936, excl.1891
Trade	District (1881-1921), Municipality (1931-1936)	Population Censuses	1881-1936, excl.1891
Liberal Professions	District (1881-1921), Municipality (1931-1936)	Population Censuses	1881-1936, excl.1891
Public Administration	District (1881-1921), Municipality (1931-1936)	Population Censuses	1881-1936, excl.1891
Panel C. Capital			
Firms	Province	Manufacture Censuses	1911, 1927, 1937
Firms with Engine	Province	Manufacture Censuses	1911, 1927, 1937
Mechanical Engines	Province	Manufacture Censuses	1911, 1927, 1937
Electrical Engines	Province	Manufacture Censuses	1911, 1927, 1937
Mechanical Horsepower	Province	Manufacture Censuses	1911, 1927, 1937
Electrical Horsepower	Province	Manufacture Censuses	1911, 1927, 1937
Panel D. Manufacturing Employment, by Industry			
Agriculture	District (1901–1921), Province (1927–1936)	Population Censuses, Manufacture Census	1901–1936
Chemicals	District (1901–1921), Province (1927–1936)	Population Censuses, Manufacture Census	1901–1936
Construction	District (1901–1921), Province (1927–1936)	Population Censuses, Manufacture Census	1901–1936
Metalworking	District (1901–1921), Province (1927–1936)	Population Censuses, Manufacture Census	1901–1936
Mining	District (1901–1921), Province (1927–1936)	Population Censuses, Manufacture Census	1901–1936
Textiles	District (1901–1921), Province (1927–1936)	Population Censuses, Manufacture Census	1901–1936
Panel E. Emigration			
US Emigration	Municipality	Ellis Island Data	1892-1924
Overall Emigration	Province, imputed to Districts	Official Statistics of the Commissioner General	1877-1925
Panel F. Other			
WW1 deaths	Municipality	ISTORECO, ANPI	1915-1918
Railways	Municipality	ISTAT	1839-1926
US GDP	National	Maddison (2007)	
Panel G. GIS Files			
Shapefiles	District, Provinces	ISTAT	1881–1936, excl. 1891

Notes. This table reports all variables used in the paper. Column (2) returns the level of aggregation at which the variable is measured. Column (3) displays the type of source the raw data are extracted from. Further references to original sources can be found in the text's main body. Column (4) reports the years when the raw data is available; "excl." indicates the years when the data do not exist. ISTAT: Italian Statistical Office or previous denominations. ISTORECO: Istituto per la storia della Resistenza e della società contemporanea, part of the Associazione Nazionale Partigiani Italiani (ANPI). Referenced on pages 9, A1.

TABLE A.2. CORRELATION BETWEEN ELLIS ISLAND AND OFFICIAL STATISTICS US EMIGRATION

	Official Statistics Emigrants					ln(1+Official Statistics Emigrants)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ellis Island Migrants	0.906***	0.906***	0.591***	0.591***	0.578***					
	(0.108)	(0.112)	(0.067)	(0.070)	(0.110)					
ln(1+Ellis Island Migrants)						0.957***	1.034***	0.835***	0.618***	0.722***
						(0.048)	(0.087)	(0.045)	(0.059)	(0.143)
Region FE	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Year FE	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes
Sample Years	All	All	All	All	1900–1914	All	All	All	All	1900–1914
N. of Regions	16	16	16	16	16	16	16	16	16	16
N. of Observations	527	527	527	527	240	527	527	527	527	240
R^2	0.820	0.820	0.971	0.971	0.982	0.820	0.882	0.894	0.955	0.972

Notes. This Table compares the number of US emigrants recorded in Italian official statistics with data from the Ellis Island Foundation dataset. The unit of observation is a region at a yearly frequency. The sample period spans 1892 to 1925. In columns (1–5), the dependent and independent variables are the number of emigrants recorded in official statistics and at Ellis Island, respectively. In columns (6–10), both variables are taken as log(1+). Columns (1) and (6) display the unconditional correlation; in columns (2) and (7), (3) and (8), and (4) and (9), we include year, region, and year and region fixed effects. In columns (5) and (10), we restrict the observation sample to the years 1900–1914, which is the period with a lower share of missing municipalities. Standard errors are always clustered at the region level and are displayed in parentheses. Referenced on page A4.

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

TABLE A.3. MOST COMMON EMIGRATION DISTRICTS

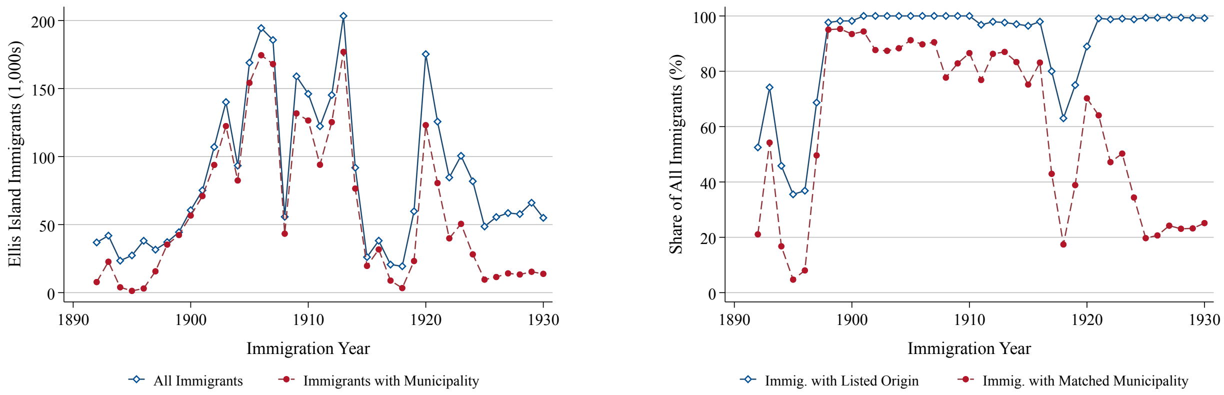
	Most Common Origin Districts					
	Absolute Number of Emigrants			Emigration Rate		
	(1) District	(2) Emigrants	(3) Emigration Rate (‰)	(4) District	(5) Emigrants	(6) Emigration Rate (‰)
1	Palermo	89546	14.097	Termini Imerese	26069	26.358
2	Caserta	40586	11.970	Piedimonte d'Alife	11256	22.794
3	Cosenza	38821	16.673	Campobasso	26965	21.390
4	Bari delle Puglie	37918	8.485	Avellino	37422	18.932
5	Avellino	37422	18.932	Sulmona	18928	18.294
6	Girgenti	34467	12.135	Cefalù	19185	17.604
7	Salerno	33096	10.414	Cosenza	38821	16.673
8	Frosinone	29422	13.111	Isernia	22900	16.466
9	Campobasso	26965	21.390	Asiago	5214	16.429
10	Termini Imerese	26069	26.358	Sant'Angelo de' Lombardi	21404	16.020
11	Messina	25765	8.735	Cerreto Sannita	13676	15.582
12	Napoli	24625	2.479	Nola	17583	15.566
13	Isernia	22900	16.466	Corleone	9192	15.482
14	Sant'Angelo de' Lombardi	21404	16.020	Nicastro	19631	15.360
15	Gerace Marina	21094	13.711	Benevento	18385	14.410
16	Catanzaro	20535	11.606	Palermo	89546	14.096
17	Gaeta	20329	11.228	Gerace Marina	21094	13.711
18	Potenza	19792	12.651	Mistretta	8763	13.612
19	Nicastro	19631	15.360	Cotrone	11811	13.593
20	Cefalù	19185	17.604	Campagna	14273	13.572
21	Sulmona	18928	18.294	Bivona	10908	13.390
22	Aquila degli Abruzzi	18562	12.411	Melfi	13898	13.231
23	Benevento	18385	14.410	Sala Consilina	10458	13.143
24	Caltanissetta	17715	10.747	Frosinone	29422	13.111
25	Lucca	17626	4.748	Castroreale	15883	13.047
26	Nola	17583	15.566	Potenza	19792	12.650
27	Oristano	17183	12.268	Aquila degli Abruzzi	18562	12.411
28	Castellammare di Stabia	16420	7.290	Oristano	17183	12.268
29	Castroreale	15883	13.047	Bovino	6784	12.224
30	Roma	15831	1.616205	Patti	15739	12.18462

Notes. This Table reports the districts with the largest number of US emigrants (columns 1–3) and US emigration rates relative to the 1921 population (columns 4–6). We list the top 30 origin districts in each category. Emigration rates are expressed in per-thousand units. Referenced on page A4.

A.V Figures

Figure A.1. Missing Origin in the Ellis Island Dataset

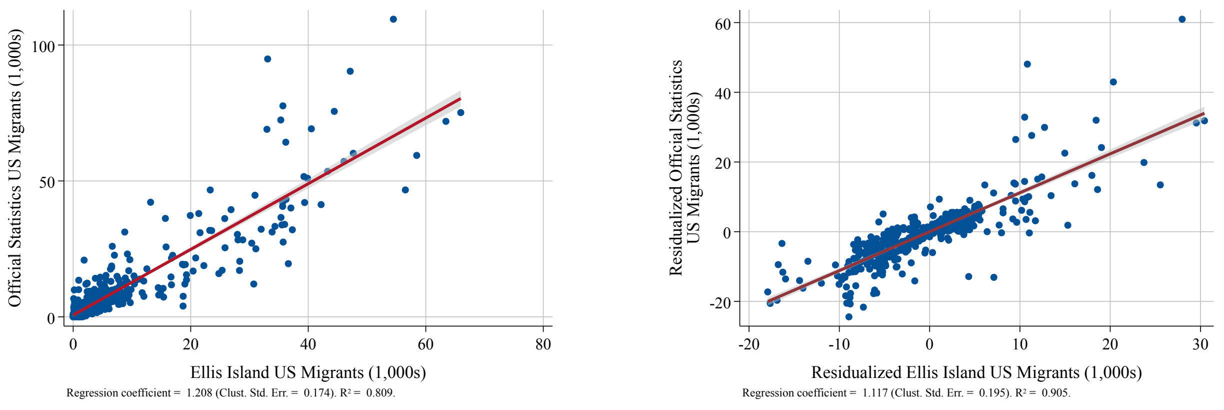
(a) Number of US Immigrants | (b) Share of US Immigrants



Notes. This Figure reports data on the missing municipality of origin in the Ellis Island dataset. In Panel A.1a, we display the absolute number of immigrants (in blue) and immigrants for whom we can assign a municipality of origin (in red). In Panel A.1b, we display the share of immigrants with at least one listed place of origin (in blue) and those for whom we can assign the listed origin to a municipality. Referenced on page 11, A3.

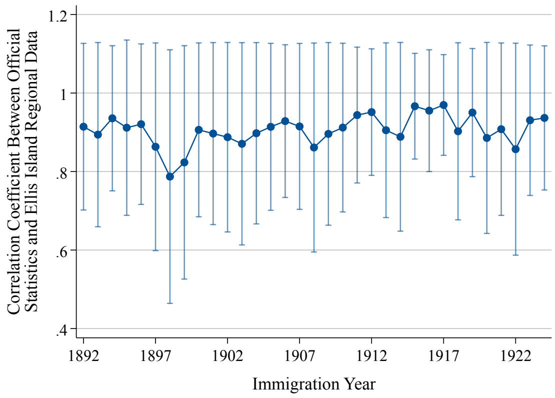
Figure A.2. Correlation between Official Statistics and Ellis Island US Emigration

(a) Unconditional Correlation | (b) Conditional Correlation



Notes. This Figure reports the correlation between the number of US emigrants recorded in the Italian official statistics and the number of Italian immigrants recorded in the Ellis Island dataset. The unit of observation is a region at a yearly frequency between 1892 and 1925. In Panel A.2a, we report the unconditional correlation. Panel A.2b displays the correlation between the variables after residualizing region and year-fixed effects. Both graphs report the fitted values of a linear regression between the variables. Referenced on page 11, A4.

FIGURE A.3. YEAR-BY-YEAR CORRELATION BETWEEN ELLIS ISLAND AND OFFICIAL STATISTICS US EMIGRANTS

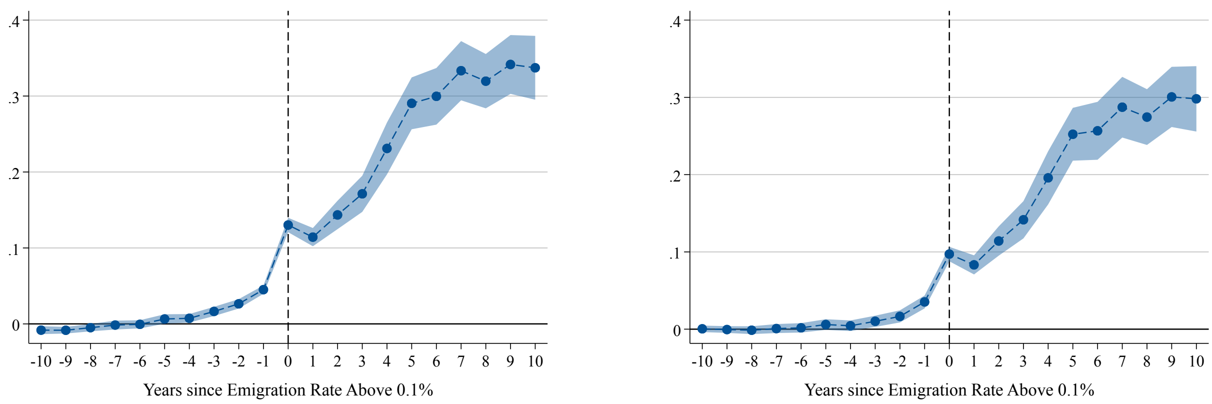


Notes. This Figure reports the correlation between the number of US emigrants recorded in the Italian official statistics and the number of Italian immigrants recorded in the Ellis Island dataset. Each dot reports the correlation between the two variables in one specific year between 1892 and 1925. In each dot, the unit of observation is a region. Both variables are standardized to have zero mean and unitary standard deviation in each year for readability. The bars report 95% confidence intervals from standard errors clustered at the regional level. Referenced on page 11, A4.

FIGURE A.4. TESTING THE S-SHAPED HYPOTHESIS

(A) UNCONDITIONAL EMIGRATION

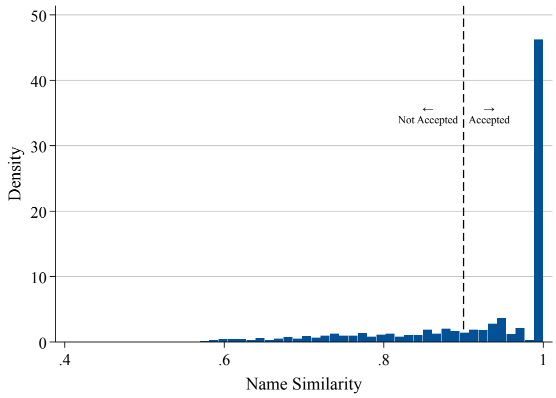
(B) EMIGRATION NET OF FIXED EFFECTS



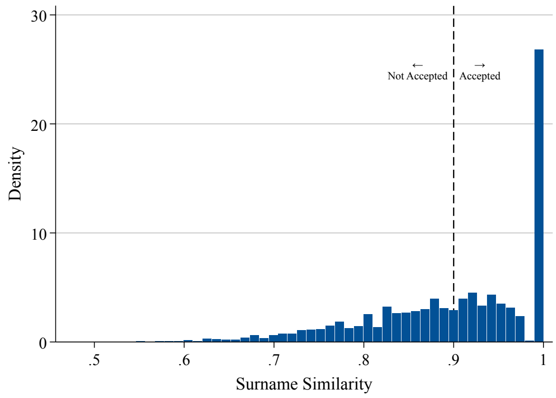
Notes. These figure test the S-shape emigration hypothesis of Gould (1980). The dependent variable is US emigration, observed at the district level every year between 1892 and 1924. For each district, we define an indicator variable that returns the number of years since the ratio between US emigrants and population in 1921 exceeds 0.1%. We report the estimated difference-in-differences coefficients obtained using the method of Borusyak *et al.* (2022) around this threshold. Panel A.4a does not include any fixed effect in the regression; in Panel A.4b, we include district and year fixed effects. Standard errors are clustered at the district level; bands report 99% confidence intervals. Referenced on page 18, A4.

FIGURE A.5. NAME AND SURNAME SIMILARITY IN THE LINKED ELLIS ISLAND–US CENSUS SAMPLE

(A) NAME SIMILARITY

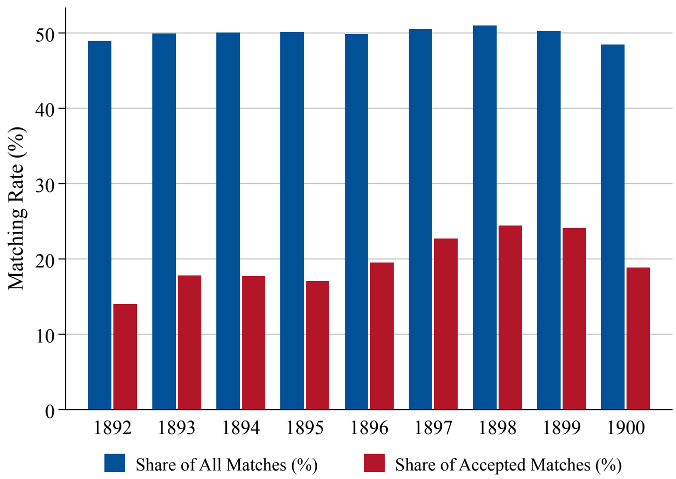


(B) SURNAME SIMILARITY



Notes. These figures reports the name (panel A.5a) and surname (panel A.5b) similarities between the records of Ellis Island and those that appear in the US census. To compare name and surname matches, we adopt the Monge-Elkan method with the embedded Jaro-Winkler word measure. The resulting values are normalized between zero and one, with values closer to one indicating closer comparisons. For each name and surname recorded in the Ellis Island data, we select the individual in the US census with the highest name and surname similarity among those whose initial Soundex-adjusted letter is the same as the Ellis Island record to be matched. The dashed lines mark the quality thresholds below which we reject the matches as uncertain. Referenced on page A5.

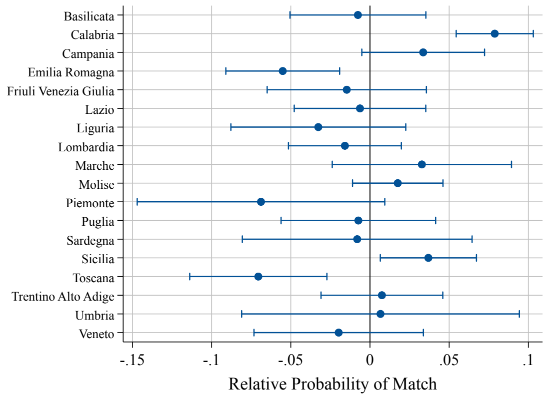
FIGURE A.6. MATCHING RATE IN THE LINKED ELLIS ISLAND–US CENSUS SAMPLE



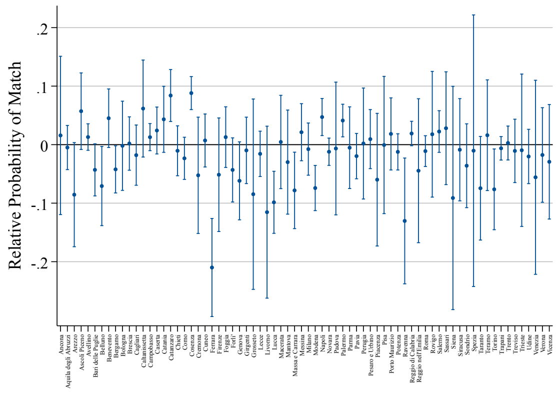
Notes. This Figure reports the matching rate in the linked sample between the Ellis Island data and the records of Italians living in the US as recorded in the population census. The blue bars report the share of individuals with at least one match in the US census. The red bars report the share of individuals with at least one acceptable match—i.e., with name and surname similarity above 0.9—in the US census. Matching rates are expressed in percentage terms and by immigration year. Referenced on page A5.

FIGURE A.7. PROBABILITY OF MATCH: BREAKDOWN BY ORIGIN OF ELLIS ISLAND IMMIGRANTS

(A) CORRELATION BY REGION OF ORIGIN



(B) CORRELATION BY PROVINCE OF ORIGIN



Notes. These figures report the correlation between the matching status and the region (panel A.7a) and the province (panel A.7b) of origin in the Ellis Island–US Census linked sample. The sample comprises all individuals who appear in the Ellis Island dataset who immigrated between 1892 and 1900. The dependent variable is an indicator equal to one if the individual has at least one accepted match in the linked sample and zero otherwise. The right-hand side consists of a series of indicator variables which tag regions, in panel A.7a, or provinces, in panel A.7b, of origin. Standard errors are clustered at the year level, and the bands report the associated 95% confidence intervals. Referenced on page A6.

B Additional Facts and Results

B.I Tables

Table B.1. First-Stage Regressions

	Dependent Variable: Measured US Emigrants							
	Weighed				Unweighed			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Predicted US Emigrants	0.511***	0.313***	0.341***	0.425***	0.580***	0.402***	0.400***	0.436***
	(0.048)	(0.043)	(0.047)	(0.058)	(0.039)	(0.043)	(0.049)	(0.059)
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Region FE	No	No	Yes	–	No	No	Yes	–
Province FE	No	No	No	Yes	No	No	No	Yes
N. of Districts	192	192	191	177	201	201	201	190
R^2	0.503	0.672	0.755	0.825	0.557	0.673	0.748	0.827
Mean Dep. Var.	8.600	8.600	8.598	8.626	8.544	8.544	8.544	8.564
Std. Beta Coef.	0.709	0.434	0.473	0.600	0.805	0.558	0.555	0.613

Notes. This Table reports the correlation between the number of US emigrants and the shift-share instrument constructed from equation (2). The units of observation are districts. Each district is observed once, and the outcome and the dependent variables are aggregated over time. The outcome variable is measured exposure to the US Quota Acts. In columns (1–4), districts are weighed by 1881 population; in columns (5–8), districts are not weighed. Columns (1) and (5) report the unconditional correlation; in columns (2) and (6), we include the number of emigrants and an indicator for southern regions as controls; in columns (3) and (7), we include region fixed effects; columns (4) and (8) control for province fixed effects. Standard errors clustered at the district level are reported in parentheses. Referenced on page 16.

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

TABLE B.2. REGIONAL US AND OVERALL EMIGRATION, 1876–1925

	Emigrants to US					Emigrants to all destinations					Share
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	76-87	88-99	00-12	13-25	Total	76-87	88-99	00-12	13-25	Total	
Abruzzi e Molise	26.9	68.0	371.0	161.6	627.4	58.3	164.1	585.7	241.6	1049.7	59.8
Basilicata	28.4	53.3	108.1	38.5	228.3	74.1	106.5	179.8	70.5	431.0	53.0
Calabria	15.0	58.5	457.7	125.1	656.3	74.1	178.5	539.8	253.6	1046.1	62.7
Campania	44.3	157.5	637.8	241.5	1081.2	131.3	339.6	871.0	360.7	1702485	63.5
Emilia Romagna	1.3	8.4	62.0	24.0	95.8	60.5	137.7	422.4	178.7	799.2	12.0
Lazio	0.02	2.3	109.4	50.1	161.9	0.4	14.0	151.4	72.9	238.6	67.8
Liguria	8.2	10.8	27.2	10.6	56.8	63.0	51.1	89.0	92.9	296.1	19.2
Lombardia	4.4	11.0	56.7	28.6	100.8	237.9	259.7	675.8	441.6	1615.2	6.2
Marche	0.2	2.0	62.0	30.6	94.8	12.7	48.0	280.6	131.1	472.3	20.1
Piemonte	5.2	12.3	109.8	43.4	170.8	353.3	332.5	697.2	527.9	1910.8	8.9
Puglie	1.3	12.9	164.7	107.9	286.9	8.1	37.2	283.4	172.4	501.2	57.2
Sardegna	0.01	0.03	8.5	5.7	14.2	1.3	6.2	72.8	43.9	124.1	11.5
Sicilia	12.6	117.2	687.7	356.1	1173.6	26.8	170.9	946.5	516.4	1660.6	70.7
Toscana	3.3	12.9	89.6	42.0	147.8	110.7	157.5	412.4	230.6	911.2	16.2
Umbria	0.1	0.5	24.1	11.8	36.6	0.5	6.0	129.9	59.4	195.7	18.7
Veneto	1.0	6.0	52.7	48.4	108.1	486.3	1197.6	1298.2	651.0	3633.1	3.0
Total	152.1	533.9	3029.1	1326.0	5041.3	1699.3	3206.9	7635.8	4045.4	16587.4	30.4

Notes. This Table reports regional emigration towards the US and total emigration from 1876 to 1925. Figures are in thousands. Column (11) indicates the percentage of total emigrants towards US relative to all emigrants from the given region in the whole period 1876-1925. Referenced on page 6.

Source: our elaboration on data from the *Annuario statistico dell'emigrazione italiana dal 1876 al 1925: con notizie sull'emigrazione negli anni 1869-1875*, Italian Statistical Office (ISTAT), Roma, 1926.

TABLE B.3. INTERNAL AND INTERNATIONAL MIGRATIONS, 1921–1931

	Absolute numbers			Share over Population	
	(1) Population	(2) Internal Migrants	(3) Emigrants	(4) Internal Migrants	(5) Emigrants
Abruzzo	1317.2	19.3	170.3	1.5	12.9
Basilicata	524.5	5.6	52.4	1.1	10.0
Calabria	1257.9	8.2	219.4	0.7	17.4
Campania	2896.6	1.2	248.4	0.0	8.6
Emilia Romagna	2183.4	78.7	165.3	3.6	7.6
Lazio	903.5	-133.8	88.2	-14.8	9.8
Liguria	892.4	-60.5	112.7	-6.8	12.6
Lombardia	3680.6	-198.0	460.6	-5.4	12.5
Marche	939.3	25.2	99.2	2.7	10.6
Piemonte	3070.3	-111.9	469.3	-3.6	15.3
Puglia	1589.1	52.9	117.8	3.3	7.4
Sardegna	682.0	2.8	27.7	0.4	4.1
Sicilia	2927.9	31.7	333.4	1.1	11.4
Toscana	2208.9	27.2	198.0	1.2	9.0
Umbria	572.1	-1.0	37.1	-0.2	6.5
Veneto	2814.2	139.8	639.8	5.0	22.7

Notes. This Table reports internal migration and out-migration flows over the period 1921-1931. Column (1) reports the population in 1881. Column (2) is the net internal migrant flow. To compute net internal migration flows, we take the difference in the outflow of people leaving a given region and the inflow of people arriving in that region during the decade 1921-1931. Since Census data only report the stock of people born in a given region living in another region in 1921 and 1931, to compute the outflow of people leaving a region during that decade, we take the difference across years of the total number of people born in that region and living in any other Italian region. Similarly, to compute the inflow of people arriving in a region during that decade, we take the difference across years of the total number living in that region who were born in any other Italian region. Positive (negative) figures imply a net population loss (gain) due to internal migrations. Column (3) reports the number of international emigrants. Figures are in thousands. Columns (4–5) report net internal and international migration figures relative to the 1881 population. Figures are in percentage terms. Referenced on page 6.

Source. Our elaboration on data from the *Annuario statistico dell'emigrazione italiana dal 1876 al 1925: con notizie sull'emigrazione negli anni 1869-1875*, Italian Statistical Office (ISTAT), Roma, 1926, and from *Censimento della Popolazione Italiana*, Italian Statistical Office (ISTAT), Roma, 1921 and 1931.

C ROBUSTNESS CHECKS

C.I TABLES

TABLE C.1. EXPOSURE TO THE US QUOTA ACTS AND POPULATION

	Difference-in-Differences					Instrumented DiD				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
US Emigrants × Post (Increasing weight)	0.077**					0.035**				
	(0.034)					(0.015)				
US Emigrants × Post (Decreasing weight)		0.080*					0.035**			
		(0.041)					(0.015)			
US Emigrants × Post (Between 1892 and 1924)			0.081**					0.035**		
			(0.040)					(0.015)		
US Emigrants × Post (Between 1892 and 1920)				0.080**					0.035**	
				(0.040)					(0.015)	
US Emigrants × Post (Between 1892 and 1914)					0.077*					0.035**
					(0.040)					(0.015)
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. of Districts	202	202	202	202	202	192	192	192	192	192
N. of Observations	1,198	1,198	1,198	1,198	1,198	1,140	1,140	1,140	1,140	1,140
R^2	0.372	0.372	0.372	0.372	0.372	0.369	0.369	0.369	0.369	0.369
Mean Dep. Var.	1.734	1.734	1.734	1.734	1.734	1.788	1.788	1.788	1.788	1.788
Std. Beta Coef.	0.121	0.132	0.151	0.150	0.143	0.020	0.020	0.016	0.016	0.017

Notes. This Table reports the estimated effect of district-level exposure to the US Quota Acts on population. The unit of observation is a district observed at a census-decade frequency between 1881 and 1936. Each column represents the estimates of a regression where the treatment is an interaction between a post-Quota (1921) indicator variable and the district-level number of US emigrants. We use five different approaches to compute US emigration, constructed using different weighting of the number of emigrants to the US or computing the number of US emigrants over different periods. In columns 1 and 2, we assign a higher weight to emigrants who departed closer to the Quota Acts date and much earlier: for both measures, we adopt an exponential weighting of factor 0.9. In columns 3, 4, and 5, we construct the Quota Exposure using the number of emigrants to the US who departed between 1892 and 1924, between 1892 and 1920, and between 1892 and 1914. Each regression controls for the number of emigrants and an interaction term between an indicator variable for Southern regions and a post-1921 indicator. We replicate the analysis in columns 6-10, substituting the measured US emigration with the shift-share instrument. Units are weighed by their population in 1881. Standard errors clustered at the district level are reported in parentheses. Referenced on page 22.

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

TABLE C.2. EXPOSURE TO THE US QUOTA ACTS AND CAPITAL INVESTMENT

	Observed Quota Exposure						Instrumented Quota Exposure					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Firms	Firms with Engines	Mechanical Engines	Electrical Engines	Mechanical Horsepower	Electrical Horsepower	Firms	Firms with Engines	Mechanical Engines	Electrical Engines	Mechanical Horsepower	Electrical Horsepower
Panel A. Increasing weight												
US Emigrants × Post	-0.165**	-0.089	-0.189***	-0.393	-0.247***	-0.212						
	(0.076)	(0.152)	(0.053)	(0.369)	(0.095)	(0.288)						
IV US Emigrants × Post							-0.053**	-0.092***	-0.012	-0.109	-0.069*	-0.101**
							(0.027)	(0.020)	(0.027)	(0.070)	(0.041)	(0.050)
Panel B. Decreasing weight												
US Emigrants × Post	-0.185**	-0.052	-0.198***	-0.356	-0.170*	-0.148						
	(0.079)	(0.167)	(0.054)	(0.397)	(0.093)	(0.304)						
IV US Emigrants × Post							-0.053**	-0.092***	-0.011	-0.113	-0.067	-0.103**
							(0.027)	(0.020)	(0.027)	(0.071)	(0.042)	(0.051)
Panel C. US emigrants between 1892 and 1924												
US Emigrants × Post	-0.168**	-0.058	-0.198***	-0.335	-0.205**	-0.148						
	(0.074)	(0.162)	(0.052)	(0.382)	(0.094)	(0.296)						
IV US Emigrants × Post							-0.053**	-0.092***	-0.011	-0.111	-0.068	-0.102**
							(0.027)	(0.020)	(0.027)	(0.071)	(0.042)	(0.051)
Panel D. US emigrants between 1892 and 1920												
US Emigrants × Post	-0.166**	-0.058	-0.199***	-0.325	-0.208**	-0.143						
	(0.073)	(0.161)	(0.052)	(0.379)	(0.094)	(0.294)						
IV US Emigrants × Post							-0.053**	-0.092***	-0.012	-0.110	-0.068*	-0.102**
							(0.027)	(0.020)	(0.027)	(0.071)	(0.041)	(0.050)
Panel E. US emigrants between 1892 and 1914												
US Emigrants × Post	-0.154**	-0.055	-0.198***	-0.307	-0.213**	-0.132						
	(0.070)	(0.159)	(0.051)	(0.373)	(0.096)	(0.290)						
IV US Emigrants × Post							-0.053**	-0.092***	-0.011	-0.109	-0.069*	-0.102**
							(0.027)	(0.020)	(0.027)	(0.070)	(0.041)	(0.050)
Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. of Provinces	68	68	68	68	68	68	68	68	68	68	68	68
N. of Observations	204	204	204	204	204	204	204	204	204	204	204	204
Mean Dep. Var.	7.277	1.253	0.545	3.437	14.095	24.245	7.277	1.253	0.545	3.437	14.095	24.245

Notes. This Table reports the estimated effect of exposure to the US Quota Acts on capital investment. The unit of observation is a province observed at a census-decade frequency between 1911 and 1936. The dependent variables are the number of firms (column 1), the number of firms with at least one engine (column 2), the number of mechanical engines (column 3), the number of electrical engines (column 4), the horsepower generated by mechanical (column 5) and electrical (column 6) engines. The treatment is an interaction between a post-Quota (1921) indicator variable and US emigration. In the five panels, we construct US emigration using different approaches. In Panel A and B we respectively assign a higher weight to emigrants departed closer to the Quota Acts date and much earlier in time: for both measures we adopt an exponential weighting of factor 0.9. In Panel C, D, and E, we respectively construct the Quota Exposure using the number of emigrants to the US who departed, respectively, between 1892 and 1924, between 1892 and 1920, and between 1892 and 1914. Each regression controls for the number of emigrants and an interaction term between an indicator variable for Southern regions and a post-1921 indicator. In columns (7–12), we replicate the same analysis but substitute the observed values of US emigration with the shift-share instrument. Units are weighed by their population in 1881. Standard errors clustered at the province level are reported in parentheses. Referenced on page 22.

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

TABLE C.3. EXPOSURE TO THE US QUOTA ACTS AND MANUFACTURE EMPLOYMENT

	Difference-in-Differences					Instrumented DiD				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
US Emigrants × Post (Increasing weight)	0.113**					0.068**				
	(0.053)					(0.031)				
US Emigrants × Post (Decreasing weight)		0.077					0.068**			
		(0.054)					(0.031)			
US Emigrants × Post (Between 1892 and 1924)			0.094*					0.068**		
			(0.056)					(0.031)		
US Emigrants × Post (Between 1892 and 1920)				0.093*					0.068**	
				(0.056)					(0.031)	
US Emigrants × Post (Between 1892 and 1914)					0.095*					0.068**
					(0.055)					(0.031)
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. of Districts	202	202	202	202	202	192	192	192	192	192
N. of Observations	1,198	1,198	1,198	1,198	1,198	1,140	1,140	1,140	1,140	1,140
R^2	0.598	0.597	0.597	0.597	0.597	0.596	0.596	0.596	0.596	0.596
Mean Dep. Var.	2.114	2.114	2.114	2.114	2.114	2.185	2.185	2.185	2.185	2.185
Std. Beta Coef.	0.082	0.059	0.081	0.081	0.081	0.018	0.018	0.014	0.014	0.015

Notes. This Table reports the estimated effect of exposure to the US Quota Acts on manufacturing employment. The unit of observation is a district observed at a census-decade frequency between 1881 and 1936. Each column represents the estimates of a regression where the treatment is an interaction between a post-Quota (1921) indicator variable and the number of US emigrants. We use five different approaches to compute the number of US emigrants. In columns 1 and 2, we assign a higher weight to emigrants who departed closer to the Quota Acts date and much earlier: for both measures, we adopt an exponential weighting of factor 0.9. In columns 3, 4, and 5, we construct the Quota Exposure using the number of emigrants to the US who departed between 1892 and 1924, between 1892 and 1920, and between 1892 and 1914. Each regression controls for the number of emigrants and an interaction term between an indicator variable for Southern regions and a post-1921 indicator. We replicate the analysis in columns 6-10, substituting the measured US emigration with the shift-share instrument. Units are weighed by their population in 1881. Standard errors clustered at the district level are reported in parentheses. Referenced on page 22.

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

Table C.4. Exposure to the Quota Acts and Agriculture Employment

	Difference-in-Differences					Instrumented DiD				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
US Emigrants × Post (Increasing weight)	-0.040					-0.016				
	(0.045)					(0.021)				
US Emigrants × Post (Decreasing weight)		-0.050					-0.017			
		(0.049)					(0.021)			
US Emigrants × Post (Between 1892 and 1924)			-0.043					-0.017		
			(0.048)					(0.021)		
US Emigrants × Post (Between 1892 and 1920)				-0.042					-0.017	
				(0.048)					(0.021)	
US Emigrants × Post (Between 1892 and 1914)					-0.036					-0.016
					(0.047)					(0.021)
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. of Districts	202	202	202	202	202	192	192	192	192	192
N. of Observations	996	996	996	996	996	948	948	948	948	948
R^2	0.313	0.313	0.313	0.313	0.313	0.304	0.304	0.304	0.304	0.304
Mean Dep. Var.	4.089	4.089	4.089	4.089	4.089	4.215	4.215	4.215	4.215	4.215
Std. Beta Coef.	-0.044	-0.060	-0.058	-0.056	-0.048	-0.007	-0.007	-0.005	-0.005	-0.005

Notes. This Table reports the estimated effect of exposure to the US Quota Acts on agriculture employment. The unit of observation is a district observed at a census-decade frequency between 1881 and 1936. Each column represents the estimates of a regression where the treatment is an interaction between a post-Quota (1921) indicator variable and the number of US emigrants. We use five approaches to compute the number of US emigrants. In columns 1 and 2, we assign a higher weight to emigrants who departed closer to the Quota Acts date and much earlier: for both measures, we adopt an exponential weighting of factor 0.9. In columns 3, 4, and 5, we respectively construct the Quota Exposure using the number of emigrants to the US who departed, respectively, between 1900 and 1924, between 1900 and 1920, and between 1890 and 1910. Each regression controls for the number of emigrants and an interaction term between an indicator variable for Southern regions and a post-1921 indicator. We replicate the analysis in columns 6-10, substituting the measured US emigration with the shift-share instrument. Units are weighed by their population in 1881. Standard errors clustered at the district level are reported in parentheses. Referenced on page 22.

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

TABLE C.5. EXPOSURE TO THE US QUOTA ACTS AND POPULATION

	Difference-in-Differences						Instrumented DiD					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
US Emigrants × Post	0.101**	0.073*	0.109**	0.101**	0.064	0.103**						
	(0.046)	(0.044)	(0.048)	(0.048)	(0.050)	(0.048)						
US $\widehat{\text{Emigrants}}$ × Post							0.046***	0.045***	0.047**	0.046**	0.041**	0.047**
							(0.018)	(0.017)	(0.019)	(0.019)	(0.020)	(0.019)
Literacy × Post	0.000						0.000					
	(0.000)						(0.000)					
Urbanization × Post		0.000***						0.000***				
		(0.000)						(0.000)				
Altitude × Post			-0.030*						-0.022			
			(0.018)						(0.022)			
Railway × Post				0.065*						0.096***		
				(0.038)						(0.037)		
WW1 deaths × Post					0.000**						0.000***	
					(0.000)						(0.000)	
Emigrants × US GDP growth						-0.000						-0.001
						(0.001)						(0.001)
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. of Districts	202	202	202	202	202	202	192	192	192	192	192	192
N. of Observations	1,198	1,198	1,198	1,198	1,198	1,198	1,140	1,140	1,140	1,140	1,140	1,140
R^2	0.372	0.373	0.372	0.372	0.373	0.372	0.369	0.370	0.369	0.369	0.370	0.369
Mean Dep. Var.	1.734	1.734	1.734	1.734	1.734	1.734	1.788	1.788	1.788	1.788	1.788	1.788
Std. Beta Coef.	0.188	0.136	0.203	0.189	0.120	0.193	0.018	0.018	0.019	0.018	0.016	0.019

Notes. This Table reports the estimated effect of district-level exposure to the US Quota Acts on population. Each column includes the interaction between the post-quota indicator variable and a district- or country-specific variable as one further control. In column 1, we use the district's literacy rate in 1901; in column 2, the urbanization rate in 1901, measured as the share of people living in towns with more than 5,000 inhabitants in the district; in column 3, the altitude at which the district is located; in column 4 an indicator variable specifying whether the district was connected to the railway network before 1901; in column 5 district's number of deaths caused by WW1. In column 6, we use the US GDP growth interacted with the volume of international emigrants. Each regression controls for the number of emigrants and an interaction term between an indicator variable for Southern regions and a post-1921 indicator. We replicate the analysis in columns 7-12, substituting the measured US emigration with the shift-share instrument. Units are weighed by their population in 1881. Standard errors clustered at the district level are reported in parentheses. Referenced on page 22, 23.

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

TABLE C.6. EXPOSURE TO THE US QUOTA ACTS AND CAPITAL INVESTMENT

	Observed Quota Exposure						Instrumented Quota Exposure					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Firms	Firms with Engines	Mechanical Engines	Electrical Engines	Mechanical Horsepower	Electrical Horsepower	Firms	Firms with Engines	Mechanical Engines	Electrical Engines	Mechanical Horsepower	Electrical Horsepower
Panel A. Control: Literacy × Post												
US Emigrants × Post	-0.181***	-0.042	-0.216***	-0.323	-0.247**	-0.140						
	(0.067)	(0.147)	(0.055)	(0.340)	(0.108)	(0.280)						
IV US Emigrants × Post							-0.041	-0.124**	-0.043	-0.149	-0.216**	-0.245**
							(0.051)	(0.055)	(0.046)	(0.132)	(0.085)	(0.109)
Panel B. Control: Urbanization × Post												
US Emigrants × Post	-0.216***	-0.068	-0.203***	-0.395	-0.253**	-0.176						
	(0.083)	(0.174)	(0.055)	(0.454)	(0.103)	(0.332)						
IV US Emigrants × Post							-0.124***	-0.149***	-0.035	-0.369**	-0.156***	-0.274***
							(0.048)	(0.047)	(0.035)	(0.161)	(0.057)	(0.102)
Panel C. Control: Altitude × Post												
US Emigrants × Post	-0.233**	-0.065	-0.200***	-0.646	-0.206*	-0.265						
	(0.102)	(0.196)	(0.054)	(0.491)	(0.124)	(0.373)						
IV US Emigrants × Post							-0.096*	-0.151***	-0.034	-0.445***	-0.138**	-0.323***
							(0.056)	(0.055)	(0.035)	(0.150)	(0.070)	(0.107)
Panel D. Control: Railway access × Post												
US Emigrants × Post	-0.185***	-0.069	-0.219***	-0.320	-0.222**	-0.155						
	(0.072)	(0.168)	(0.050)	(0.372)	(0.090)	(0.299)						
IV US Emigrants × Post							-0.086*	-0.150***	-0.049	-0.295**	-0.136**	-0.271***
							(0.050)	(0.039)	(0.035)	(0.115)	(0.059)	(0.073)
Panel E. Control: WW1 death rate × Post												
US Emigrants × Post	-0.021	0.061	-0.215***	0.338	-0.184*	0.245						
	(0.071)	(0.173)	(0.052)	(0.344)	(0.109)	(0.298)						
IV US Emigrants × Post							-0.028	-0.110**	-0.043	-0.186**	-0.130**	-0.179**
							(0.045)	(0.044)	(0.037)	(0.086)	(0.061)	(0.084)
Panel F. Control: US GDP growth × I(Quota Exposure) × Post												
US Emigrants × Post	-0.153*	-0.103	-0.327***	-0.272	-0.244***	-0.120						
	(0.086)	(0.170)	(0.073)	(0.362)	(0.093)	(0.274)						
IV US Emigrants × Post							-0.087*	-0.147***	-0.050	-0.319***	-0.150**	-0.262***
							(0.051)	(0.042)	(0.035)	(0.114)	(0.061)	(0.082)
Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. of Provinces	68	68	68	68	68	68	68	68	68	68	68	68
N. of Observations	204	204	204	204	204	204	204	204	204	204	204	204
Mean Dep. Var.	7.277	1.253	0.545	3.437	14.095	24.245	7.277	1.253	0.545	3.437	14.095	24.245

Notes. This Table reports the effect of the Quotas on capital investment. In the five panels, we add a different control, given by the interaction between the post-quota indicator variable and a province-specific variable. We use, respectively, the literacy rate in 1901, the urbanization rate in 1901 (share of people in towns with more than 5'000 inhabitants), the altitude, the share of municipalities connected to the railway network before 1901, the number of deaths caused by WW1; the US GDP growth interacted with the province's number of emigrants. Each regression controls for the number of emigrants and an interaction term between an indicator variable for Southern regions and a post-1921 indicator. In columns (7–12), we replicate the exercises but substitute measured US emigration with the shift-share instrument. Standard errors clustered at the province level are reported in parentheses. Referenced on page 22, 23.

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

TABLE C.7. EXPOSURE TO THE US QUOTA ACTS AND MANUFACTURE EMPLOYMENT

	Difference-in-Differences						Instrumented DiD					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
US Emigrants × Post	0.090*	0.105*	0.131**	0.124**	0.107	0.128**						
	(0.050)	(0.059)	(0.065)	(0.063)	(0.067)	(0.063)						
US $\widehat{\text{Emigrants}}$ × Post							0.050	0.080**	0.084**	0.083**	0.081**	0.084**
							(0.036)	(0.038)	(0.036)	(0.036)	(0.039)	(0.036)
Literacy × Post	0.000***						0.000***					
	(0.000)						(0.000)					
Urbanization × Post		0.000						0.000*				
		(0.000)						(0.000)				
Altitude × Post			-0.016						-0.004			
			(0.033)						(0.031)			
Railway × Post				0.160						0.187*		
				(0.098)						(0.111)		
WW1 deaths × Post					0.000						0.000	
					(0.000)						(0.000)	
Emigrants × US GDP growth						-0.000						-0.001
						(0.004)						(0.004)
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. of Districts	202	202	202	202	202	202	192	192	192	192	192	192
N. of Observations	1,198	1,198	1,198	1,198	1,198	1,198	1,140	1,140	1,140	1,140	1,140	1,140
R^2	0.599	0.598	0.598	0.598	0.598	0.597	0.598	0.597	0.597	0.597	0.597	0.597
Mean Dep. Var.	2.114	2.114	2.114	2.114	2.114	2.114	2.185	2.185	2.185	2.185	2.185	2.185
Std. Beta Coef.	0.078	0.091	0.113	0.107	0.093	0.110	0.009	0.015	0.015	0.015	0.015	0.015

Notes. This Table reports the estimated effect of district-level exposure to the US Quota Acts on population. In each column, we control for the interaction between the post-quota indicator variable and a district- or country-specific variable. In column 1, we use the district's literacy rate in 1901; in column 2, the urbanization rate in 1901, measured as the share of people living in towns with more than 5'000 inhabitants in the district; in column 3, the altitude at which the district is located; in column 4 an indicator variable specifying whether the district was connected to the railway network before 1901; in column 5 district's number of deaths caused by WW1. In column 6, we use the US GDP growth interacted with the volume of international emigrants. Each regression controls for the number of emigrants and an interaction term between an indicator variable for Southern regions and a post-1921 indicator. In columns (7–12), we replicate the analysis by substituting the measured US emigration with the shift-share instrument. Units are weighed by their population in 1881. Standard errors clustered at the district level are reported in parentheses. Referenced on page 22, 23.

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

TABLE C.8. EXPOSURE TO THE QUOTA ACTS AND AGRICULTURE EMPLOYMENT

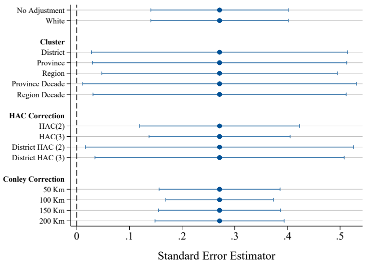
	Difference-in-Differences						Instrumented DiD					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
US Emigrants × Post	-0.004	-0.006	-0.001	-0.008	-0.010	-0.009						
	(0.039)	(0.035)	(0.036)	(0.036)	(0.035)	(0.036)						
US $\widehat{\text{Emigrants}}$ × Post							0.003	-0.019	-0.014	-0.016	-0.016	-0.016
							(0.019)	(0.024)	(0.021)	(0.021)	(0.022)	(0.021)
Literacy × Post	-0.000***						-0.000***					
	(0.000)						(0.000)					
Urbanization × Post		-0.000						-0.000				
		(0.000)						(0.000)				
Altitude × Post			-0.032*						-0.030*			
			(0.017)						(0.018)			
Railway × Post				-0.006						-0.002		
				(0.038)						(0.041)		
WW1 deaths × Post					0.000						0.000	
					(0.000)						(0.000)	
Emigrants × US GDP growth						-0.007***						-0.007***
						(0.002)						(0.003)
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. of Districts	202	202	202	202	202	202	192	192	192	192	192	192
N. of Observations	996	996	996	996	996	996	948	948	948	948	948	948
R^2	0.314	0.312	0.313	0.312	0.312	0.313	0.306	0.304	0.305	0.304	0.304	0.305
Mean Dep. Var.	4.089	4.089	4.089	4.089	4.089	4.089	4.215	4.215	4.215	4.215	4.215	4.215
Std. Beta Coef.	-0.005	-0.008	-0.002	-0.011	-0.013	-0.012	0.001	-0.005	-0.004	-0.004	-0.004	-0.004

Notes. This Table reports the estimated effect of district-level exposure to the US Quota Acts on population. In each column, we control for the interaction between the post-quota indicator variable and a district- or country-specific variable. In column 1, we use the district's literacy rate in 1901; in column 2, the urbanization rate in 1901, measured as the share of people living in towns with more than 5'000 inhabitants in the district; in column 3, the altitude at which the district is located; in column 4 an indicator variable specifying whether the district was connected to the railway network before 1901; in column 5 district's number of deaths caused by WW1. In column 6 we use the US GDP growth interacted with the volume of international emigrants. Each regression controls for the number of emigrants and an interaction term between an indicator variable for Southern regions and a post-1921 indicator. In columns (7–12), we replicate the analysis, substituting the measured US emigration with the shift-share instrument. Units are weighed by their population in 1881. Standard errors clustered at the district level are reported in parentheses. Referenced on page 22, 23.
***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.10$

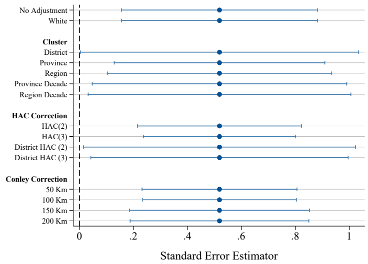
C.II FIGURES

FIGURE C.1. ALTERNATIVE STANDARD ERRORS: DISTRICT-LEVEL REGRESSIONS

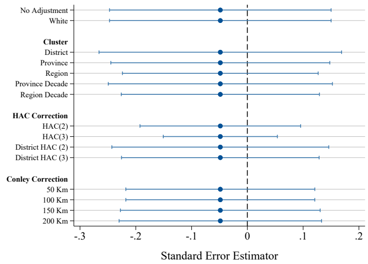
(A) POPULATION



(B) MANUFACTURING EMPLOYMENT



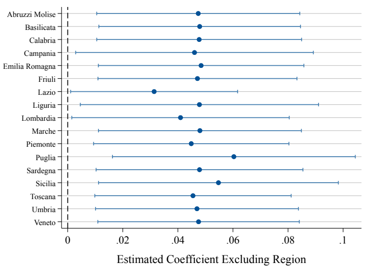
(C) AGRICULTURE EMPLOYMENT



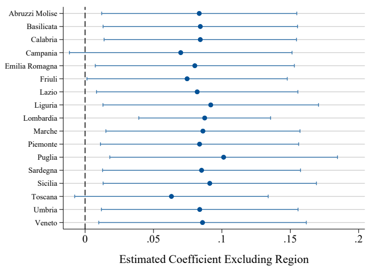
Notes. This Figure reports alternative estimators of the standard errors in the baseline district-level regressions. The unit of observation is a district, at a decade frequency between 1881 and 1936. The dependent variable is population (Panel C.1a), manufacturing employment (Panel C.1b), and agriculture employment (Panel C.1c). Regressions are estimated through OLS and include district and decade fixed effects. The treatment is an interaction term between the instrumented US emigration and a post-1921 term. Each regression controls for the number of emigrants and an interaction term between an indicator variable for Southern regions and a post-1921 indicator. Units are weighed by their population in 1881. Bands report 95% confidence intervals. Referenced on page 23.

FIGURE C.2. EXCLUDING ONE REGION AT THE TIME: DISTRICT-LEVEL REGRESSIONS

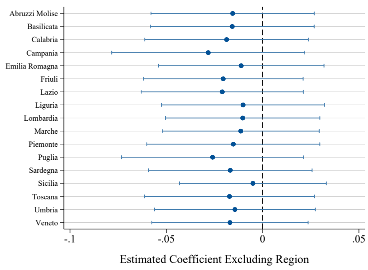
(A) POPULATION



(B) MANUFACTURING EMPLOYMENT



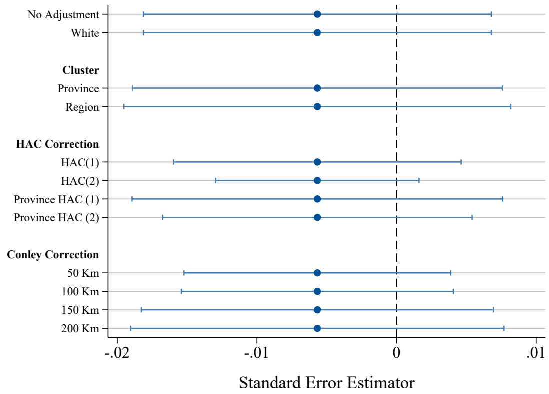
(C) AGRICULTURE EMPLOYMENT



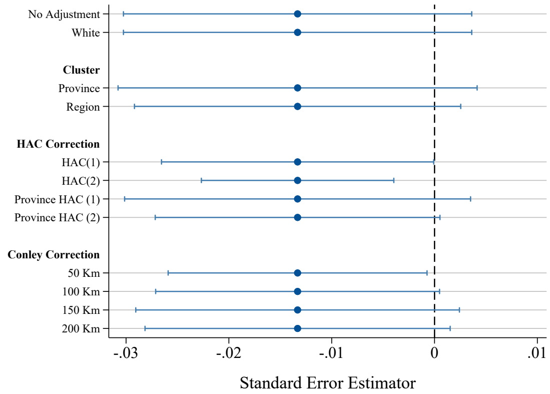
Notes. This Figure reports the estimates of the baseline district-level regressions when excluding one region. The unit of observation is a district, at a decade frequency between 1881 and 1936. The dependent variable is population (Panel C.2a), manufacturing employment (Panel C.2b), and agriculture employment (Panel C.2c). Regressions include district and decade-fixed effects. The treatment is an interaction term between the instrumented US emigration and a post-1921 term. Each regression controls for the number of emigrants and an interaction term between an indicator variable for Southern regions and a post-1921 indicator. The left axis indicates the excluded region. Units are weighed by their population in 1881. Bands report 95% confidence intervals. Standard errors clustered at the district level. Referenced on page 22.

FIGURE C.3. ALTERNATIVE STANDARD ERRORS: PROVINCE-LEVEL REGRESSIONS

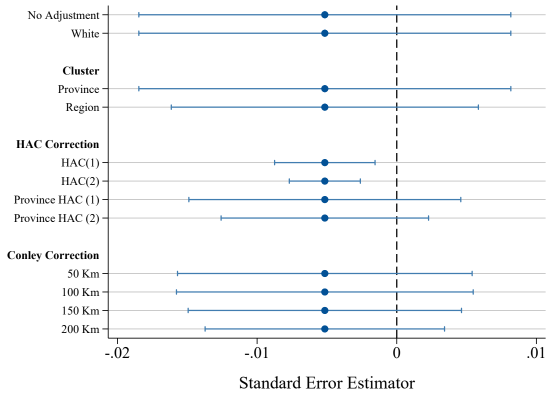
(A) FIRMS



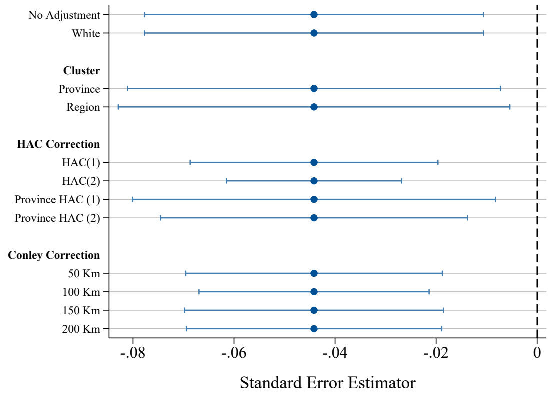
(B) FIRMS WITH ENGINE



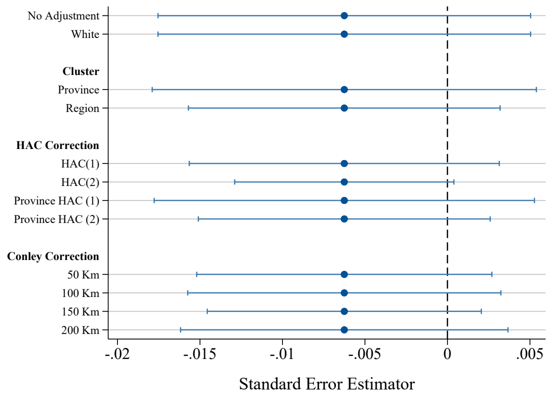
(C) MECHANICAL ENGINES



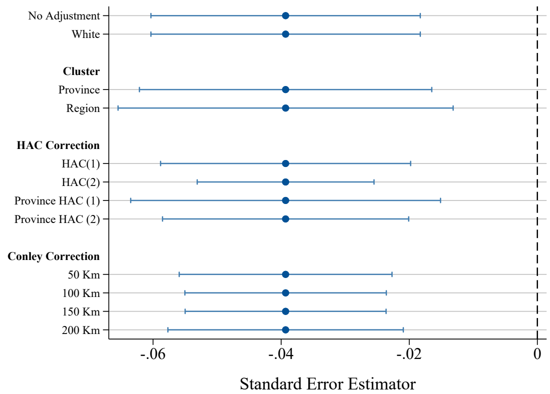
(D) ELECTRICAL ENGINES



(E) MECHANICAL HORSEPOWER



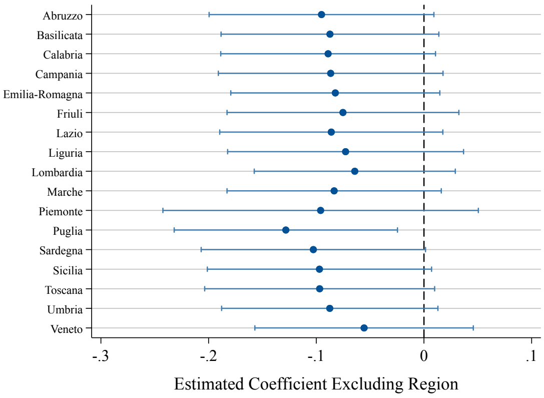
(F) ELECTRICAL HORSEPOWER



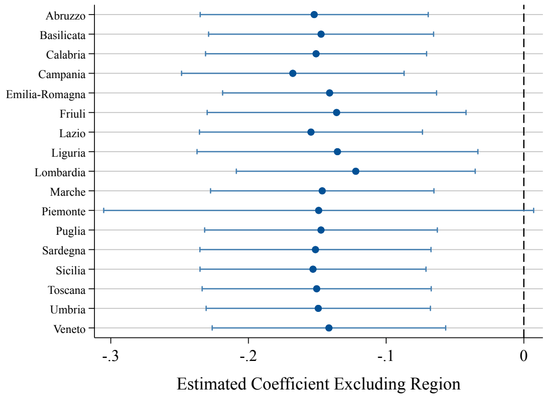
Notes. This Figure reports alternative estimators of the standard errors in the baseline province-level regressions. The unit of observation is a province, at a decade frequency between 1911 and 1937. Regressions are estimated through OLS on the logged outcomes and include province and decade fixed effects. The treatment is an interaction term between the instrumented US emigration and a post-1921 term. Each regression controls for the number of emigrants and an interaction term between an indicator variable for Southern regions and a post-1921 indicator. Units are weighed by their population in 1881. Bands report 95% confidence intervals. Referenced on page 23.

Figure C.4. Excluding one region at the time: Province-Level Regressions

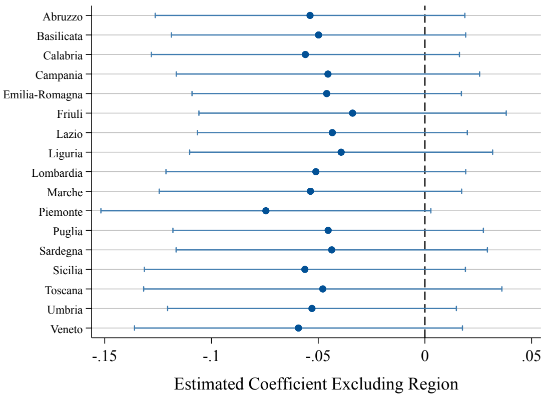
(a) Firms



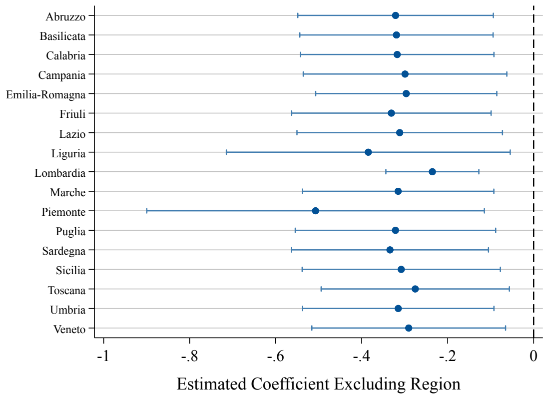
(b) Firms with Engine



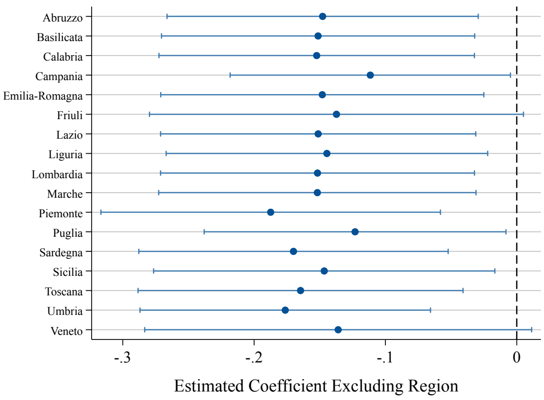
(c) Mechanical Engines



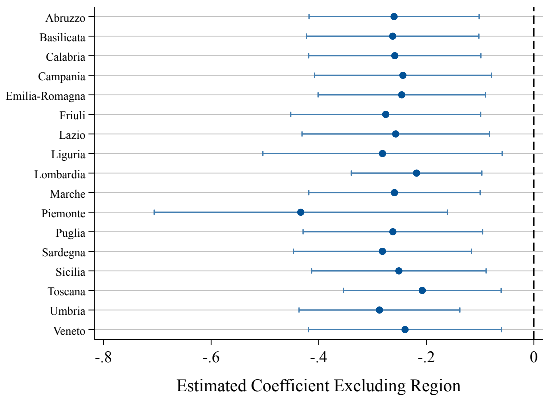
(d) Electrical Engines



(e) Mechanical Horsepower



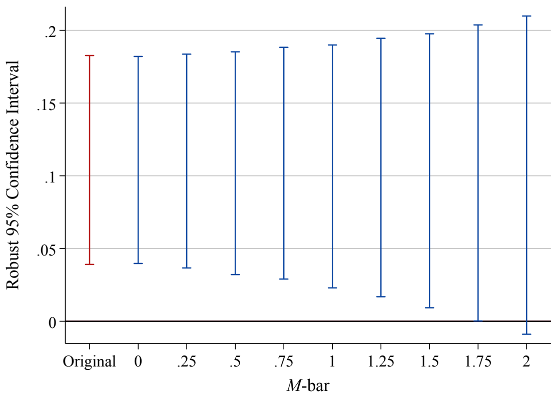
(f) Electrical Horsepower



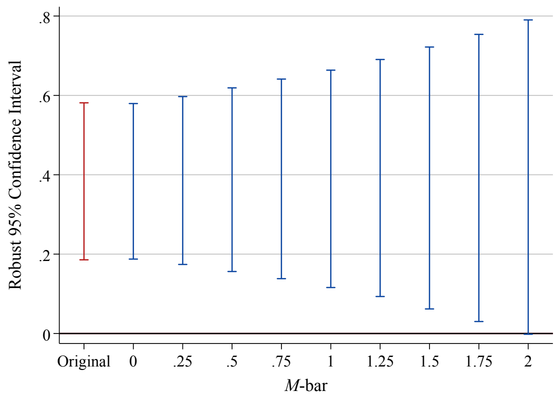
Notes. This Figure reports the estimates of the baseline province-level regressions when excluding one region. The unit of observation is a province, at a decade frequency between 1911 and 1937. Regressions include province and decade fixed effects. The left axis indicates the excluded region. The treatment is an interaction term between the instrumented US emigration and a post-1921 term. Each regression controls for the number of emigrants and an interaction term between an indicator variable for Southern regions and a post-1921 indicator. Units are weighed by their population in 1881. Bands report 95% confidence intervals. Standard errors clustered at the province level. Referenced on page 22.

FIGURE C.5. RELATIVE MAGNITUDE OF PARALLEL TRENDS VIOLATION

(A) POPULATION



(B) MANUFACTURING EMPLOYMENT



Notes. This Figure presents alternative confidence intervals for the post-Quota coefficient displayed in the event-study graphs for population (Panel C.5a) and manufacturing employment (Panel C.5b). The bands report 95% robust confidence intervals constructed following the method developed by Rambachan and Roth (2023). In both cases, the unit of observation is a district observed at a census-decade frequency between 1881 and 1936. Each dot reports the coefficient of an interaction term between period dummies and the cross-sectional instrument of US emigration. The regression includes district and decade-fixed effects and controls for the volume of international migrants and an indicator variable for Southern regions, both interacted with a post-1921 indicator. Units are weighed by their population in 1881. Referenced on page 23.

D A Model of Directed Technological Adoption

In this section, we develop a simple framework to rationalize our main findings in the context of labor-saving technical change theory. Proofs and further analytical insights on the baseline environment can be found in section D.III.

D.I Theoretical Framework

In this section, we develop a simple analytical framework inspired to Zeira (1998) and San (2023) to clarify the empirical implications of directed technical change and adoption theory. The core assumption we make is that capital goods—hereafter, machines—substitute labor as a production input. We thus implicitly restrict technological progress to be labor-saving, as in Acemoglu (2002, 2007). The decision of the firm to adopt productivity-enhancing machines will depend on their price relative to the cost of labor. In the equilibrium, a labor supply shock—such as the one induced by IRPs—dampens the incentive to adopt machines because it pushes down the wage, hence prompting firms to substitute capital with labor.

Consider a closed economy with one consumption good and a representative household supplying labor. The consumption good is produced by a continuum of tasks $j \in [0,1]$. Each task can be performed with either labor or machines. The amount of machines in task j is denoted by $x(j)$, whereas the amount of labor employed is $e(j)$. Note that each task can be fulfilled with either machines or labor, but not both. This is intended to model in a stylized manner labor-saving machines. To simplify the analysis and following **?** we assume that machines fully depreciate at the end of the period, hence the model is essentially static.

The final consumption good is produced by identical, perfectly competitive firms with the following production function:

$$Y = A\left[\int_0^{\iota} mx(j)^{\alpha}\,dj + \int_{\iota}^{1} e(j)^{\alpha}\,dj\right], \tag{D.1}$$

where A is a technology parameter, m is the relative productivity of machines and $\alpha \in (0,1)$ is a production parameter. We assume $m \in (0,1)$ following San (2023), and restrict machines to be equally productive across tasks j. The choice variable $\iota \in [0,1]$ denotes *industrialization* defined as the share of automatized tasks, which are those fulfilled by machines. We assume that tasks are ordered by degree of complexity. Because the marginal cost of producing machines—which we define below—is increasing in complexity, the price of machines is non-decreasing in j. It is therefore without loss of generality to assume that the first ι tasks are automatized. This is because the final good producer will first automatize tasks whose machine costs the least since the relative productivity of machines is constant across tasks. We assume that there is a fixed stock of labor $L > 0$ which is supplied inelastically by the household.

The problem of the representative final good producer is, therefore, to choose the industrialization level ι, and input quantities $x(j)$ and $e(j)$ for each task to maximize profits

$$\max_{\iota,\{x(j),e(j)\}_{j\in[0,1]}} Y - \int_0^{\iota} p(j)x(j)\,dj - w\int_{\iota}^{1} e(j)\,dj, \tag{D.2}$$

where $p(j)$ is the price of a machine for task j, w is the nominal wage, subject to the technology constraint (D.1). Note that the price of the consumption good is implicitly normalized to one. In section D.III, we formally show that the demand schedules for machines and labor are given by the following demand schedules:

$$x(j) = p(j)^{-\frac{1}{1-\alpha}} (\alpha A m)^{\frac{1}{1-\alpha}} \quad \forall j \in [0,\iota], \tag{D.3a}$$

$$e(j) = w^{-\frac{1}{1-\alpha}} (\alpha A)^{\frac{1}{1-\alpha}} \qquad \forall j \in [\iota,1]. \tag{D.3b}$$

Combining (D.3a)-(D.3b) with the first order condition for the industrialization rate, it follows that in the equilibrium ι^* is pinned down by the following:

$$m = \left[\frac{p(\iota^*)}{w}\right]^{\alpha}. \tag{D.4}$$

The economic intuition behind condition (D.4) is that at the marginal task, *i.e.* the last automatized task, the price of the machine fulfilling the task must be equal to the cost of labor, adjusted by the technology parameter and the relative productivity of machines.

Each machine is produced by a monopolist, following Zeira (1998). The machine producer will seek to set the monopoly price, which maximizes its profits subject to a demand for machines (D.3a). We assume that the marginal cost of machines $\psi(\cdot)$ is increasing in the complexity of tasks, *i.e.* $\psi'(\cdot) > 0$. Moreover, we assume that the marginal cost function satisfies basic Inada conditions.[31] This is intended to capture the idea that machines substituting low-skill tasks are not as expensive as those replacing tasks on the right side of the skill distribution of workers. The problem of the machine producer is therefore

$$\max_{p(j)} [p(j) - \psi(j)]\, x(j), \tag{D.5}$$

subject to (D.3a). In section D.III, we show that the first-order conditions imply

$$p(j) = \min\left\{mw, \frac{\psi(j)}{\alpha}\right\}, \tag{D.6}$$

where the minimum descends from the observation that because each task can be performed by labor as well as by machines, setting a price greater than the productivity-adjusted wage simply pushes the final goods producer not to automatize the task. We now obtain two technical results to ensure the existence and uniqueness of the equilibrium. The formal definition of the competitive equilibrium in this economy, as well as the proofs of all lemmas and propositions, can be found in section D.III.

[31] In this setting, this simply boils down to $\lim_{j\uparrow 1} \psi(j) = +\infty$ and $\lim_{j\downarrow 0} \psi(j) = 0$. The economic intuition behind these is that it is never profitable for the representative firm to automatize all tasks. Similarly, there is always at least one task that is automatized. Note that while these assumptions are sufficient for the existence of an equilibrium, they are not necessary.

Lemma D.1. *In the equilibrium, the marginal task ι^* is such that $p(\iota^*) = \psi(\iota^*)/\alpha = wm^{1/\alpha}$.*

Combining this result with the equilibrium conditions of the final goods producer, we derive the following strong existence result.

Proposition D.1. *There exists one and only one $\iota^* \in [0,1]$ which solves the problem of the final good producer* (D.3a)-(D.3b)-(D.4) *as well as the problem of the machine producers* (D.6) *and verifies labor market clearing. In particular, the equilibrium industrialization ι^* is the solution to the following:*

$$\psi(\iota^*) = L^{\alpha-1}(1-\iota^*)^{1-\alpha}\alpha^2 A m^{1/\alpha}.$$

This concludes our analytical characterization of the environment. We now exploit the model to deliver a number of testable predictions which will guide our empirical analysis.

D.II EMPIRICALLY TESTABLE IMPLICATIONS

Having established the existence of the equilibrium, we can now derive two key empirical implications of this directed technical adoption setting. First, note that Lemma D.1 conveys the basic intuition of the model. In particular, we have $\psi(\iota^*) = \alpha m^{1/\alpha} w$, hence an increase in the nominal wage induces industrialization to rise because $\psi'(\cdot) > 0$ by assumption. The economic intuition behind this result is that if the cost of labor increases, then the final good producer will seek to automatize more tasks in order to avoid paying the increase in the wage. This is summarized in the following implication statement.

Implication D.1. Following an exogenous increase (resp. decrease) in the nominal wage w, the share of tasks performed by machines ι^* increases (resp. decreases).

A similar comparative static result follows considering an increase in the labor stock. To see it, notice that because the nominal wage is invariant across tasks, from (D.3b) and labor market clearing the total labor stock L is evenly allocated across the $(1-\iota^*)$ non-automated tasks. Using this insight, we obtain the following empirical prediction.

Implication D.2. Following an exogenous increase (resp. decrease) in the labor supply stock L, the share of tasks performed by machines ι^* decreases (resp. increases).

This is the key implication of the model that we test in the paper. In our setting, we provide evidence that immigration restriction policies induce positive labor supply shocks, hence increasing the labor stock. We show that firms operating in districts that were more exposed to the Quota Acts decreased investment in machinery—section V.2—and increased employment—-section V.3. These findings are fully in line with the empirical predictions D.2 of the model and hence provide evidence in favor of labor-saving directed technical adoption.

D.III Proofs of Analytical Results

Solution of the problem of the final good producer. Plugging the technology constraint into problem (D.2), the problem of the final good producer reads out as follows:

$$\max_{\iota,\{x(j),e(j)\}_{j\in[0,1]}} A\left[\int_0^\iota mx(j)^\alpha\,dj+\int_\iota^1 e(j)^\alpha\,dj\right]-\int_0^\iota p(j)x(j)\,dj-w\int_\iota^1 e(j)\,dj.$$

The—necessary and sufficient—first-order conditions with respect to labor and capital in the generic task j are

$$x(j)=p(j)^{-\frac{1}{1-\alpha}}(\alpha Am)^{\frac{1}{\alpha}}\;\forall j\in[0,\iota],$$
$$e(j)=w^{-\frac{1}{1-\alpha}}(\alpha A)^{\frac{1}{\alpha}}\quad\forall j\in[\iota,1].$$

To obtain the first-order condition for the optimal industrialization rate, apply the Leibniz integral rule with respect to ι to get:

$$x(\iota^*)\left[mx(\iota^*)^{\alpha-1}-p(\iota^*)\right]=e(\iota^*)\left[e(\iota^*)^{\alpha-1}-w\right].$$

Plugging (D.3a)-(D.3b) into the expression above we get $m=(p(\iota^*)/w)^\alpha$. □

Solution of the problem of the monopolist. The solution is trivial upon plugging (D.3a) into the objective function (D.5). □

Proof of Lemma D.1. From (D.6) and (D.4), it is

$$p(\iota^*)=\min\left\{\frac{\psi(\iota^*)}{\alpha},mw\right\},$$
$$p(\iota^*)=m^{1/\alpha}w$$

Hence, we have

$$m=\left[\frac{\min\left\{\frac{\psi(\iota^*)}{\alpha},mw\right\}}{w}\right]^\alpha.$$

We can distinguish two cases. Assume $mw\le\psi(\iota^*)/\alpha$. This implies that $m=m^\alpha$, which is only verified if $m=1$ or $m=0$. Since by assumption $m\in(0,1)$, this can never hold. We are left with the case $mw>\psi(\iota^*)/\alpha$. We show that this is consistent with all the parameter restrictions. Note first that since $m\in(0,1)$, it must be $\psi(\iota^*)/\alpha<w$, since otherwise it would be $m\ge1$. We therefore have $\psi(\iota^*)/\alpha<w$ and $\psi(\iota^*)/\alpha<mw$. Because $m<1$, the only binding constraint is $\psi(\iota^*)/\alpha<mw$. It is

$$m=\left[\frac{\psi(\iota^*)}{\alpha}\cdot\frac{1}{w}\right]^\alpha,$$

which implies $\psi(\iota^*)/\alpha = wm^{1/\alpha}$. Because $m \in (0,1)$, $m^{1/\alpha} < m$ since $\alpha \in (0,1)$, and therefore $\psi(\iota^*)/\alpha = wm^{1/\alpha} < wm$. This implies that the solution is acceptable. Hence, $p(\iota^*) = \psi(\iota^*)/\alpha$ and this concludes the proof. □

Proof of Proposition D.1. Because $w(j) = w$ for all $j \in [0,1]$, from (D.3b) we get that $e(j)$ does not depend on j and:

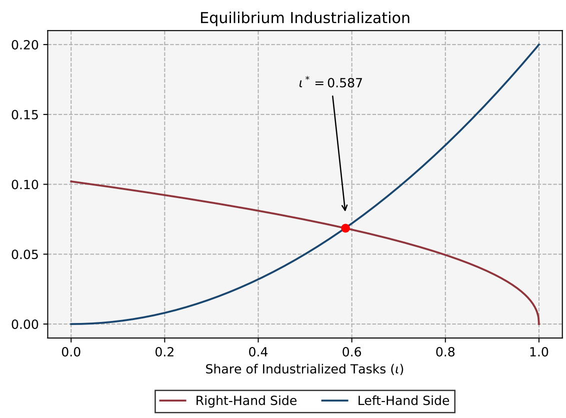
$$e(j) = e = w^{-\frac{1}{1-\alpha}}(\alpha A)^{\frac{1}{1-\alpha}} = \frac{L}{1-\iota^*},$$

where the last equality holds by labor market clearing, which requires $(1-\iota^*)e = L$. From Lemma D.1, it is $w = \psi(\iota^*)/(\alpha m^{1/\alpha})$. Plugging this into the previous equation, we get

$$\left(\frac{\psi(\iota^*)}{\alpha m^{1/\alpha}}\right)^{-\frac{1}{1-\alpha}}(\alpha\beta)^{\frac{1}{1-\alpha}} = \frac{L}{1-\iota^*}$$
$$\frac{\psi(\iota^*)}{\alpha m^{1/\alpha}}(\alpha\beta)^{-1} = \left(\frac{L}{1-\iota^*}\right)^{-1+\alpha}$$
$$\psi(\iota^*)L^{1-\beta} = (1-\iota^*)^{1-\alpha}\alpha^2 A m^{1/\alpha}.$$

Because $\psi'(\cdot) > 0$, the left-hand side is strictly increasing in ι^*. Moreover, because $\alpha \in (0,1)$, the right-hand side is strictly decreasing in ι^*. By the Inada conditions, $\lim_{z\uparrow 1}\psi(z) = +\infty$ and $\lim_{z\downarrow 0}\psi(z) = 0$. If $\iota^* = 0$, the right-hand side is strictly positive, whereas it is zero if $\iota^* = 1$. Hence, because both are trivially continuous, by the intermediate value theorem, there exists at least one ι^* which verifies the equation. Since both are strictly monotone, ι^* is unique. □

FIGURE D.1. EQUILIBRIUM OF THE MODEL



Notes. This figure plots the equilibrium of the model. The blue and red lines, respectively, display the left and right-hand sides of the final equation of the proof of Proposition D.1. We assume $\psi(j) = \gamma j^2$ even though quadratic costs do not verify the Inada conditions.

Parametrization: $\alpha = .55$, $\beta = .45$, $\gamma = .2$, $A = .5$, $L = 1$, $m = .5$.

Proof of Implication D.1. From Lemma D.1, it is $m^{1/\alpha} = \psi(\iota^*)/(\alpha w)$, or

$$\alpha w m^{1/\alpha} = \psi(\iota^*).$$

Because $\psi'(\cdot) > 0$, an increase in w in the equilibrium implies an increase in $\psi(\iota^*)$, hence in ι^*. $\square$

Proof of Implication D.2. First note that because w is invariant across tasks, then by (D.3b) $e(j) = e$ for all j. Moreover, since the productivity of labor is constant across tasks, it is optimal to divide evenly L across the $(1-\iota^*)$ non-automatized tasks. Therefore, by labor market clearing $e = L/(1-\iota^*)$. Plug this in the left-hand side of (D.3b), yielding

$$w^{-\frac{1}{1-\alpha}}(\alpha A)^{\frac{1}{1-\alpha}} = \frac{L}{1-\iota^*}.$$

Using Lemma D.1 into the previous equation we get

$$\begin{aligned}\frac{\psi(\iota^*)}{\alpha m^{1/\alpha}} &= \left(\frac{L}{1-\iota^*}\right)^{\alpha-1}\alpha A \\ L^{1-\alpha} &= \frac{(1-\iota^*)^{1-\alpha}}{\psi(\iota^*)}\alpha^2 A m^{1/\alpha}.\end{aligned}$$

Because $\alpha \in (0,1)$ and $\psi'(\cdot) > 0$, the right-hand side is decreasing in ι^*. Therefore, an exogenous increase in L leads to an increase in the right-hand side, hence a decrease in ι^*. Following an increase in the labor supply, the share of automatized tasks decreases. $\square$

Appendix References

ABRAMITZKY, R., L. BOUSTAN, K. ERIKSSON, J. FEIGENBAUM and S. PÉREZ (2021). "Automated Linking of Historical Data." *Journal of Economic Literature*, 59(3): 865–918.

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

ACEMOGLU, D. (2002). "Directed Technical Change." *The Review of Economic Studies*, 69(4): 781–809.

——— (2007). "Equilibrium Bias of Technology." *Econometrica*, 75(5): 1371–1409.

BORUSYAK, K., X. JARAVEL and J. SPIESS (2022). "Revisiting Event Study Designs: Robust and Efficient Estimation."

CALABRESE, V. (2017). "Land of Women: Basilicata, Emigration, and the Women Who Remained Behind, 1880-1914." *Dissertations, Theses, and Capstone Projects*.

COLUCCIA, D. M. and G. DOSSI (2023). "Return Innovation: The Knowledge Spillovers of the British Migration to the United States, 1870–1940."

ECKERT, F., A. GVIRTZ, J. LIANG and M. PETERS (2020). "A Method to Construct Geographical Crosswalks with an Application to US Counties since 1790."

GOULD, J. D. (1980). "European inter-continental emigration. The road home: return migration from the

U.S.A." *The Journal of European Economic History*.

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

RAMBACHAN, A. and J. ROTH (2023). "A More Credible Approach to Parallel Trends." *Review of Economic Studies*, 90(5): 2555–2591.

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

SAN, S. (2023). "Labor Supply and Directed Technical Change: Evidence from the Termination of the Bracero Program in 1964." *American Economic Journal: Applied Economics*, 15(1): 136–163.

SPITZER, Y. and A. ZIMRAN (2023). "Like an Ink Blot on Paper: Testing the Diffusion Hypothesis of Mass Migration, Italy 1876-1920."

ZEIRA, J. (1998). "Workers, Machines, and Economic Growth." *The Quarterly Journal of Economics*, 113(4): 1091–1117.