Southern (American) Hospitality:

Italians in Argentina and the US during the Age of Mass Migration

Santiago Pérez*

October 3, 2019

Abstract

Argentina and the US were the two largest destinations for Italians during the age of mass migration. Prior work finds that Italians assimilated faster in Argentina, but is inconclusive on whether this was due to differences in selection or host-country conditions. I assemble data following Italians from passenger lists to censuses of Argentina and the US, enabling me to compare migrants with similar pre-migration characteristics but who moved to different countries. Italians had faster assimilation in Argentina, and this advantage was unlikely to be due to selection. Migration path dependence can rationalize these differences in an era of open borders.

*Department of Economics, University of California, Davis and NBER. Contact e-mail: seperez@ucdavis.edu. I thank Enrique Pérez and María Fabiana Vaccaro for their help collecting the data. I have benefitted from feedback from Ran Abramitzky, Leah Boustan, Federico Curci, Walker Hanlon, Alejandra Irigoin, Herbert S. Klein, Giovanni Peri, Rowena Gray, Mateo Uribe-Castro, and from seminar participants at the NBER Summer Institute; the 12th Migration and Development Conference in Madrid; Statistics Norway Workshop on Intergenerational Mobility, Gender and Family Formation in the Long Run; UC Davis Migration Research Cluster; Economic History Association Meetings; Universidad de San Andrés and Universidad de los Andes.

1 Introduction

At the turn of the 20th century, the shift in migrants' regional origin toward Southern and Eastern Europe fueled the rise of anti-immigrant sentiments in the US. In 1907, the US Congress convened a special commission to analyze the social and economic life of immigrants. The *Immigration Commission* painted a dismal picture of Italians, the largest contributors to the surge of southern European immigration: Italians were consistently at the bottom in terms of family income, rates of home ownership and job skills (Dillingham, 1911). The conclusions of the Commission served as the basis for the imposition of country-of-origin quotas, which in 1924 limited the number of new Italian immigrants to just 4,000 per year.[1]

The situation of Italians in the US contrasts with their situation in Argentina, the second largest destination for Italians during the age of mass migration. For instance, in 1909 Italians owned 38 percent of the 28,632 commercial establishments in Buenos Aires, despite them being just 22 percent of the city's population (Martínez, 1910). According to Klein (1983, p. 306), "the sharp differences in the Italian immigrant experience within Argentina and the United States were fully perceived by both the immigrants themselves and virtually all contemporary observers."

These contrasting experiences raise two related questions: First, what was the source of these differences? Italians in Argentina might have had different outcomes than Italians in the US either because the Italians who went to Argentina were different from those who went to the US, or because of differences in how Italians fit socially and economically into each of the host countries. Second, why did some Italians choose one destination country over the other? By 1900, the end of my period of analysis, nearly one million Italians had moved to each of the destinations (both of which had nearly open borders for European immigration).

I study the selection and economic outcomes of Italians in Argentina and the US during the age of mass migration. To do so, I assembled data following Italian immigrants from passenger lists to population censuses. In these data, I observe the year of entry, port of origin and pre-migration occupation of a sample of Italians who resided in Argentina or the US by the late 19th century. These data enable me to assess, to a far greater extent than with the existing cross-sectional data, if the differences in economic outcomes at the destination can be explained by selection. In addition,

[1] For perspective, an average of 200,000 Italians per year entered the US in the decade following the country-of-origin quotas (Ferenczi, 1929).

by including individual-level data on pre-migration characteristics, it also enables me to shed light on the drivers of migrants' destination choices.

Beyond its historical significance, this episode can inform broader issues in the economics of immigration. First, both Argentina and the US maintained nearly open borders for European immigration, enabling me to study migrants' destination choices without the interference of modern immigration policy. Hence, comparing my results to modern studies can shed light on how these choices are influenced by immigration policy. For instance, by investigating the role of immigrant networks in a context with no family reunification visas, I can inform the discussions on the importance of such visas in explaining modern "diaspora" effects (Beine et al., 2011). Second, by observing immigrants' outcomes and characteristics both before and after they move, I am able to quantify the relative importance of pre-migration characteristics and host-country conditions in explaining immigrant assimilation. This is an important margin, as several countries rely on pre-migration characteristics to screen potential migrants.

In the first part of the paper, I compare Italians in the census cross-sections of 1895 Argentina and 1900 US. I focus on two main economic outcomes throughout the analysis: A person's occupation and whether he owned his home in the destination country. Consistent with the historical literature (Baily, 1983, 2004; Klein, 1983), I document that Italians in Argentina had higher rates of home ownership and were more likely to hold skilled occupations than Italians in the US.

The advantage of Italians in Argentina might have been driven by differences in the characteristics of those moving to each of the destinations. I use individual-level passenger-list data to compare the pre-migration characteristics of Italians moving to Argentina or the US before 1900. The main difference between both groups was the higher fraction of Italians departing from northern ports among those going to Argentina. However, I find small or no differences in pre-migration occupations or other demographic characteristics: Italians who moved to Argentina or the US were similar with respect to their age and gender, and were employed in similar (predominantly unskilled) occupations prior to migrating.

I next compare the outcomes of Italians in Argentina and the US using the passenger lists linked to census data. Here, I am able to narrow the comparison to immigrants who left Italy in the same year, from the same port, and who had the same pre-migration occupation and literacy level. The advantage of Italian migrants in Argentina is, in most cases, similar to that in the

cross-section. This similarity suggests a limited role for observable pre-migration characteristics (including regional origins within Italy) in explaining the advantage of Italians in Argentina.

As a last exercise, I further narrow the comparison to Italian immigrants who shared a surname but moved to different destinations. This comparison serves two purposes. First, Italian surnames are informative of regional origins (Guglielmino and De Silvestri, 1995; Spitzer and Zimran, 2018). Hence, surnames enable me to absorb a finer regional variation than the one captured by ports of origin. Second, Clark, Cummins, Hao, and Vidal (2015) and Güell, Rodríguez Mora, and Telmer (2014) show that there is substantial persistence in economic outcomes across family lines, an effect that a within-surname comparison would absorb. These results show a similar pattern, again suggesting a limited role for pre-migration characteristics in explaining the different outcomes of Italians across the Americas.

Which host-country conditions explain the differences in economic outcomes? First, I show that although Italian immigrants in Argentina and the US had similar levels of human capital (as proxied by literacy and numeracy rates), Italians in Argentina had higher levels of human capital *relative* to the native-born population. Second, I find evidence consistent with the closer linguistic distance between Italian and Spanish enabling Italians to enter a broader range of occupations in Argentina relative to the US. Finally, I provide qualitative and historical survey evidence showing the widespread prejudice against Italians in the US during this period.

While my data enable me to account for selection into destinations to a far greater extent than with the existing cross-sectional data, I cannot fully rule out that the relative success of Italians in Argentina was driven by unobserved differences between those who went to Argentina and those who went to the US. To assess the likely role of unobservable characteristics in explaining the differences in outcomes, I compare the stability of the coefficient estimates when expanding the set of control variables, as suggested in Altonji et al. (2005). The results of this exercise suggest that the degree of selection on unobservables would have needed to be as much as 11 times larger to rationalize the differences in home ownership rates, and 37 times larger to rationalize the differences in the likelihood of being employed in an unskilled occupation.

The analysis is based on a sample of individuals who chose to stay in the Americas (at least until the time of the censuses). However, a substantial fraction of Italians (both in Argentina and the US) eventually returned to Italy during this era (Bandiera, Rasul, and Viarengo, 2013).

One hypothesis for the better performance of Italians in Argentina is that Italian migrations to the US were less likely to be permanent, thus reducing the incentives to invest in host-country specific human capital. Indeed, the US Immigration Commission pointed to the high rates of return migration among southern Europeans as one of the main reasons for their lack of assimilation (and even recommended restricting temporary migrations).[2] My results provide evidence against this being the main explanation, as the advantage of Italians in Argentina was present even by the second generation. Moreover, rates of return migration were actually not very different for these cohorts of Italians in Argentina and the US.

The large differences in economic outcomes and the likely limited role of pre-migration characteristics pose a puzzle: Why did (in an era of open borders) some Italians choose a country that offered them limited prospects for upward mobility? One simple explanation is that, although upward mobility was lower, wages for unskilled workers were higher in the US than in Argentina. Indeed, while Argentina was among the top five countries in terms of per-capita income by 1900 (having been the fastest growing economy in the world in the previous three decades (Taylor, 2018)), wages for unskilled workers were about 25% lower than in the US (Williamson, 1995). Hence, Italians deciding between Argentina and the US might have faced a trade-off between higher wages in the short-term and better prospects for upward mobility in the long-term. While this hypothesis is plausible, it fails to explain *who* chose to move to each of the countries. For instance, one implication of such hypothesis is that younger Italians should have been more likely to move to Argentina (where there were lower initial wages but higher rates of upward mobility). Yet, arrival ages of Italians in Argentina and the US were remarkably similar. Similarly, it fails to explain why immigrants from certain Italian provinces were disproportionately more likely to choose one destination over the other.

An alternative explanation is that immigrant networks generated path dependence in destination choices: For Italians choosing where to migrate, having relatives or friends in one of the destinations might have been the decisive factor. In the last part of the paper, I use the passenger lists to test whether migrants were more likely to move to the destination where their "family and friends" had migrated in the past. Because I do not directly observe family or friend relationships among

[2] "As far as possible, the aliens excluded should be those who come to this country with no intention to become American citizens or even to maintain a permanent residence here, but merely to save enough, by the adoption, if necessary, of low standards of living, to return permanently to their home country." (Dillingham, 1911)

immigrants, I use the surnames of previous migrants to Argentina and the US to construct a proxy for the size of a migrant's network at each potential destination. I find that this measure is a strong predictor of where Italians moved, suggesting a role for path dependence in explaining destination choices. Consistent with network effects, the measure has stronger predictive power for women, children, and relatively unskilled migrants.

This paper is related to the literature on immigrant assimilation during the age of mass migration. Several papers have studied the economic assimilation of immigrants in specific receiving countries. For instance, Abramitzky, Boustan, and Eriksson (2014), Catron (2016) and Ferrie (1994, 1997) study the assimilation of immigrants in the US, whereas Inwood et al. (2016) and Green and MacKinnon (2001) study the case of Canada.[3] In previous work (Pérez, 2017), I studied the economic assimilation of European immigrants in 19th-century Argentina. However, no quantitative studies have looked at the *comparative* performance of immigrant groups across different receiving countries. The case of Italian migration to Argentina and the US is especially relevant, as it deals with the main sending country and the two largest destinations during this period. Italians are also an ideal case study because they migrated in large numbers to both North and South America and because of the availability of individual-level data with information on pre- and post-migration outcomes.[4]

2 Italian Mass Emigration

From 1876 to 1915, more than 14 million Italians emigrated to other countries in Europe and to the Americas: Italians represented the largest flow in absolute numbers during the age of mass migration. In per capita terms, Italian emigration rates were second only to the Irish (Taylor and Williamson, 1997).

The Italian case was distinct from that of other European countries in that Italians emigrated in large numbers to multiple destinations: About 60% moved to South and North America, and the rest moved to other countries within Europe. Argentina and the US were the two destinations

[3] Other examples include Hatton (1997) and Minns (2000) for the US, and Moya (1998) for Argentina.

[4] Italians are the only immigrant group for which there was significant overlap between Argentina and the US. The second largest sending country for Argentina was Spain: Combined, Italy and Spain account for more than 80% of all the immigrants who went to Argentina. However, there were only about 7,000 Spanish immigrants in 1900 US (Gibson and Jung, 1999).

with the largest transcontinental flow, receiving 2.5 and 4.5 million Italians, respectively, from 1857 to 1924 (Ferenczi, 1929). Figure 1 shows the yearly number of Italian arrivals in Argentina and the US in this period. While Italians were more likely to emigrate to Argentina than to the US from 1860 to 1880, both countries attracted similar numbers during the last two decades of the 19th century. After 1900, however, the majority of Italian migration was to the US.

The increase in migration to the US relative to Argentina coincided with a change in the regional origins of Italian migrants. During the second half of the 19th century, Italian emigration was more predominant in the relatively more developed northern Italy. By the turn of the 20th century, southern Italian migration took off (Gomellini and O'Grada, 2011).[5]

3 Data

I use two sources of individual-level data for Argentina and the US: passenger lists of immigrant arrivals and population censuses. The Argentine passenger lists were originally collected by Argentina's National Direction of Immigration and have been digitized by *Centro de Estudios Migratorios Latinoamericanos* and *Fondazione Rodolfo Agnelli*. The data include about 1,020,000 records of Italians who arrived in Argentina through the port of Buenos Aires between 1882 and 1920.[6] Each record contains the name, age, sex, civil status, literacy, occupation, date of arrival and port of origin of each passenger. Unfortunately, other than the port of origin, the data do not include any systematic information on last place of residence within Italy.

The US passenger lists come from the National Archives ("Italians to America" passenger data file) and are based on information collected by the US Customs Service. The data include about 845,000 passengers who arrived in the US between 1855 and 1900 and who identified their country of origin as Italy or one of the following regions: Lombardy, Piedmont, Sardinia, Sicily, or Tuscany. Most of the records are of passengers arriving to New York, although other US ports are also included. Each record contains information on the name, age, sex, literacy, occupation, town of last residence (although with incomplete coverage), destination within the US, date of arrival, port of origin and entry and class of travel of each passenger.

[5] Northern Italians were also overrepresented among those going to Brazil. Klein (1989) shows that the economic outcomes of Italians in Brazil were similar to those of Italians in Argentina.

[6] About 75% of immigrants entered Argentina through the port of Buenos Aires in this period (de Inmigración, 1925). I discuss the coverage and representativeness of the passenger list data in Online Appendix section 7.

I linked males in these passenger lists to national population censuses of Argentina and the US. In the case of Argentina, the 1895 census is the only census for which such linking is possible, since the previous national census (which took place in 1869) was conducted before the passenger list data started being systematically collected, and there are no surviving individual-level records for the next census (which took place in 1914). For the US, I linked the passenger lists to the 1900 census, which is the closest in time to the 1895 Argentine census.[7] To improve the comparability between the Argentine and US data, I restricted the US sample to arrivals on or after 1882. As a result of these data limitations, my analysis excludes Italians who arrived during the peak of Italian immigration to the US (after 1900).

The linking is based on country of birth, first and last name, and reported age. A challenge in linking these data is that some Italians declared their original name (in Italian) upon arrival but later adopted a Spanish/English version of it (see Biavaschi et al. (2017) and Carneiro et al. (2017) for these names changes in the context of the US). For instance, the *Giuseppes* were likely to become *Josés* in Argentina and *Josephs* in the US. To deal with this challenge, I first used a dictionary of first names to translate Italian names into their Spanish or English counterparts. Then, I used these translated names as an additional input in the linking procedure, following a similar procedure in Alexander et al. (2018) and Pérez (2017).

To link individuals from the passenger lists to the censuses, I implemented the following procedure (described in detail in Abramitzky et al. (2019)). In the first step, I identified a group of individuals in the passenger lists that I would attempt to match to the census. Then, I searched the full count census for a set of potential matches for each individual. I identified potential matches as individuals who: (1) reported Italy as their place of birth, (2) had a predicted age difference of no more than five years in absolute value, and (3) had first and last names starting with the same letter. Based on the similarity of their reported names and predicted years of birth, I calculated a linking score ranging from 0 to 1 for each pair of potential matches, with higher scores corresponding to pairs of records that were more similar to each other.[8]

To be considered a unique match for an individual in the passenger lists, a record in the census had to satisfy three conditions: (1) be the record with the highest linking score p_1 among all the

[7] Sadly, there are no surviving individual-level records of the 1890 US Census.

[8] To measure similarity in first and last names, I used the Jaro-Winkler string distance function (Winkler, 1990), whereas to measure similarity in reported ages I used the absolute value of the predicted years of birth.

potential matches for that individual, (2) have a linking score above a threshold ($p_1 > \underline{p}$, with $\underline{p} \in (0,1)$), and (3) have a linking score sufficiently higher than the second highest linking score ($p_2 < l$, with $l \in [0, \underline{p})$). In the baseline analysis, I only kept observations with a linking score of at least 0.7 and a second highest linking score of at most 0.5. In section 4.4, I show the robustness of the results to using more conservative choices of the linking parameters.

An important concern with using such data is that some of the links might be incorrect (Bailey et al., 2017). To address this concern, I chose a relatively conservative set of linking parameters. While this choice implies that I am able to uniquely match a relatively small fraction of records (due to a standard trade-off between type I and type II errors), it also implies that the quality of matches is likely higher. Indeed, Abramitzky et al. (2018) show that this method achieves low rates of false positives (below 5%), although at the expense of matching relatively few observations. Using my baseline parameters, I uniquely link around 6% of the Argentine observations and 4% of the US observations. Lower matching rates for the US are expected given slightly higher return migration (37 versus 30% for these cohorts, as shown below), combined with the fact that Italian names in the US were probably more likely to be severely misspelled than in Argentina (given the similarity between Italian and Spanish).[9] In the robustness section of the paper, I show that the results are robust to imposing an even higher threshold for considering an observation as a match.

An additional concern is whether this linking procedure generates representative samples of the populations of interest. Tables A1 and A2 compare immigrants in the passenger lists who were uniquely linked to the census to those who were not (for Argentina and the US, respectively). Column 1 in each table reports the average value of each of the included characteristics in the passenger lists, whereas column 2 reports the corresponding average in the linked data. In column 3, I report the average difference between both groups of observations.

There are some statistically significant differences between Italians in the passenger lists and those in the linked data, although the differences are, in all cases, fairly small. In both the Argentina and US samples, I am less likely to match individuals who report an unskilled occupation upon arrival and more likely to match individuals with white-collar occupations. There is also a correlation between age upon arrival (positive for Argentina, negative for the US) and the likeli-

[9] When comparing two independent transcriptions of the 1940 US census, Abramitzky et al. (2018) show that Italian surnames have very high rates of discrepancies: 32% of the surnames have at least a one character difference.

hood of matching. It is also worth noting that immigrants in the linked sample might differ from immigrants in the passenger lists data for reasons unrelated to the linking procedure (for instance, selective mortality or return migration).

As an alternative approach to assess the representativeness of the linked samples, in Tables A3 and A4 I compare Italians in the cross sections of 1895 Argentina and 1900 US to those in the linked data. One limitation of this comparison is that the 1895 Argentine census did not include a question on year of arrival to the country, which prevents me from restricting the sample to the relevant arrival cohorts in the cross section (that is, those arriving from 1882 to 1895). The main advantage, however, is that it enables me to compare Italians in the panel data to those in the cross section with respect to the main two outcomes I investigate in the paper: home ownership rates and the likelihood of holding an unskilled occupation. Both tables show small differences between the linked samples and the cross section with respect to the main outcomes of interest of the paper. In the Argentine panel data, Italians are 1.6 percentage points more likely to own their homes, compared to an average of 20% in the cross section. In the US, Italians in the panel data are similarly likely to be home owners than in the cross section. Italians in the Argentine panel are statistically indistinguishable from Italians in the cross section with respect to the likelihood of holding an unskilled occupation, whereas Italians in the US panel are 2 percentage points less likely to hold an unskilled job.

Overall, while the differences between the linked sample and the cross-sectional data are small, both comparisons suggest some selection into the linked samples. To address this concern, in section 4.4 I show that the results are similar when I reweight the data to account for selection with respect to these observable characteristics. Also, whenever possible, I show results based on the cross-sectional data (which do not rely upon linking). Indeed, the finding that Italians had higher rates of home ownership and were less likely to be employed in unskilled occupations is preserved when using these data.

Both the US passenger lists and the US 1900 census are fully digitized, including the information on occupations and other economic outcomes. The Argentine passenger lists are also fully digitized, but only the *indexes* of the 1895 census are. Hence, after linking the data, I manually digitized the economic information in the 1895 Argentine census (using the original manuscripts available in the genealogy website familysearch.org). The baseline linked samples include about 15,000 observations

for each destination country.

4 Results

4.1 Differences in the cross section

I start by comparing Italian immigrants in the census cross-sections of 1895 Argentina (Somoza, 1967) and 1900 US (Ruggles et al., 1997). To do so, I restrict both samples to Italians and estimate:

$$y_{ic} = \alpha + \beta Argentina_{ic} + \gamma X_{ic} + \epsilon_{ic} \tag{1}$$

where y_{ic} is an outcome of individual i in destination country c. Throughout the analysis, I focus on two outcomes that can be consistently measured in the Argentine and US censuses: The likelihood of home ownership and the likelihood of holding an unskilled occupation.[10] The coefficient of interest is β, which measures the economic advantage/disadvantage of Italians in Argentina relative to Italians in the US. The sample is restricted to males aged 18 to 60 years old at the time of the census. In all specifications, I control for an individual's age using fixed effects (captured by the vector X_{ic}).

One issue with this model is that the Italian advantage/disadvantage in Argentina might reflect aggregate differences between the Argentine and US economies rather than Italian-specific differences. Indeed, Klein (1983) argues that the preponderance of small artisan shops in Argentine manufacturing offered more opportunities for skilled blue-collar and white-collar jobs than the more industrialized US economy. Hence, I also estimate a model comparing Italians in Argentina and the US relative to natives or other immigrants in their respective destination countries:

$$y_{ic} = \alpha_0 + \beta_1 Italian_{ic} + \beta_2 Argentina + \beta_3 Italian_{ic} \times Argentina_{ic} + \gamma X_{ic} + \epsilon_{ic} \tag{2}$$

[10] To maximize the consistency of the occupational classifications, I first assigned each occupation a code from the Historical International Standard Classification of Occupations (HISCO), which I then mapped into occupational categories using the Historical International Social Class Scheme (HISCLASS) (Leeuwen et al., 2002). This classification is based on the International Standard Classification of Occupations (ISCO) and has been adapted to work with historical data. Unskilled jobs are those in HISCLASS categories 10 to 12. The 1895 Argentine census asked *Posee propiedad raíz?* ("Do you own real estate property?"). The 1900 US census asked "Is the person's home owned or rented?". Neither the 1895 Argentine nor the 1900 US census contains information on individual-level earnings, which prevents me from looking at earnings as an outcome variable.

Here, the coefficient of interest is β_3 which measures the advantage/disadvantage of Italians in Argentina relative to Italians in the US, net of aggregate differences between both destination countries.

Table 1 shows that first-generation Italians in Argentina were 4.3 percentage points more likely to own their home, relative to a baseline of 14.4 percent among Italians in the US. The relative advantage of Italians in terms of home ownership is much larger (above 25 percentage points) when including the native-born of both countries in the sample (column 2), or when comparing Italians to other immigrant groups (column 3), or to everyone in the male working-age population (column 4).

Italians in Argentina were also 28 percentage points less likely to be employed in an unskilled occupation, relative to a baseline of about 50 percent among Italians in the US. This gap is very similar when including the native born in the sample (column 6), and smaller but still large when comparing Italians to other immigrant groups (column 7), or to everyone in the male working-age population (column 8).

Table B1 in the Online Appendix shows that these results are not driven by Italians in the US being disproportionately more likely to locate in urban areas: There is a similar pattern when restricting the sample to urban areas (panel (a)). The relatively few Italians who lived in rural areas in the US were more likely to own property than the Italians in rural areas of Argentina. However, panel (b) show that this pattern reverses when I compare Italians to the other groups in the respective countries (reflecting that, in the US, home ownership rates were on average higher than in Argentina). Finally, Table B2 shows similar results when restricting the sample to just the cities of Buenos Aires and New York, the two largest Italian destinations in Argentina and the US and the original focus of Baily (1983, 2004).

4.2 Differences in selection

The above results confirm the relative economic success of Italians in Argentina documented in Klein (1983) and Baily (1983, 2004). However, this advantage might have just reflected differences in the pre-migration characteristics of those who went to each of the countries. For instance, Argentina received relatively more migration from northern Italy (which exhibited higher levels of

human capital than the South at the time (Federico et al., 2019)) than did the US.[11] Figure 2 shows the ten largest ports of origin of Italian arrivals in Argentina and the US in the 1882-1900 period. Genoa and Naples were the two largest ports of departure of Italian migrants, both for Argentina and the US. However, while close to 80% of the Italians entering Argentina departed from the port of Genoa (a northern port), less than 20% of those moving to the US did so. In contrast, Naples (a southern port) accounted for more than half of the arrivals in the US but for only 10% in Argentina.

Whether Argentina received relatively more skilled migrants than the US is controversial in the historical literature. On the one hand, Baily (1983, p. 295) argues that: "Those who migrated to Buenos Aires included more workers with higher levels of skill and of literacy, more individuals with experience in organization, and more people who intended to stay." On the other hand, Klein (1983, p. 329) argues that: "No significant factors in the Italian origin of the immigrants, or in their cultural make-up, can as fully explain the social and economic history of the Italians in the Americas."

I next use the passenger lists data to compare the pre-migration characteristics of Italians arriving in Argentina and the US. For this analysis, I use the data on the 1882 to 1900 arrivals, as these are the years for which the Argentine and US passenger lists overlap. Specifically, I estimate:

$$x_{it} = \alpha + \beta Argentina_{it} + \gamma Z_{it} + \epsilon_{it} \tag{3}$$

where x_{it} represents a pre-migration characteristic of immigrant i arriving in year t.

Table 2 shows the results of these regressions, where each row represents a different individual-level characteristic. In the first column, I report the average value of each of these variables in the US data. In the remaining columns, I report the value of β (the Argentina-US difference) as I progressively control for year of arrival and port of origin fixed effects (captured by the vector Z_{it}).

The top panel compares Italian migrants in Argentina and the US with respect to demographic characteristics: the fraction of males, average age and the fraction of children (defined as those aged 16 or less). These variables are important because a higher fraction of women and children is indicative of the prevalence of family migration and hence, of the intended permanence of migra-

[11] Several studies look at contemporary differences between Southern and Northern Italy. See for instance Ichino and Maggi (2000).

tions. In the raw data, Italians moving to Argentina were on average younger, more likely to be aged less than 16 years, and less likely to be male. However, this pattern reverses once I include year of entry and port of origin fixed effects, thus comparing Italians moving to the Americas on the same year and from the same port. Overall, even in the raw data, there were no large differences in the age structure and gender ratios of Italians moving to Argentina or the US: Both groups were largely comprised of working-age males, with males aged 18 to 60 years old accounting for close to 60% of the flow in both cases.[12]

I next look at differences in pre-migration occupations. Here, I focus on the sample of males aged 18 to 60 years old upon arrival. Italians who went to Argentina were overrepresented among those holding white-collar jobs (although the proportion of workers in this category was fairly small in both flows at around 3%), and underrepresented among those holding skilled/semi-skilled jobs. The differences in this regard are of similar magnitude when including the various fixed effects.

The most salient difference in terms of pre-migration occupations is that Italians who migrated to Argentina were more likely to report farming and less likely to report unskilled occupations. Part of this difference captures differences across regions of origin: Including port of origin fixed effects reduces the Argentina-US difference in the likelihood of holding a farming occupation from 20 to 13 percentage points.

However, the distinction between farm and general laborers is unlikely to have been very informative in this context: As late as 1911, about 60% of the Italian workforce was still employed in agriculture.[13] Indeed, the linked data enable me to explicitly test the informativeness of the distinction between general and farm laborers. Specifically, if this distinction captured some relevant information, we should observe differences in the outcomes at the destination of both types of workers. However, when using the linked data in section 4.3, I find that whether an individual declared a farming or an unskilled occupation upon arrival has little predictive power on his outcomes at the destination.[14]

[12] Figure B1 in the Online Appendix shows that the overall age structure of both groups (and not just the average age) was also very similar. This figure uses the pooled 1882-1900 data to plot a histogram of the ages of Italians arriving in Argentina or the US.

[13] For instance, Klein (1983, p. 313) writes that "the entire distinction between non-farm unskilled laborers and farm workers may have been rather artificial." Coletti (1912) declared that he and all other analysts of Italian emigration have found that "laborers, day laborers, and the like come in large part from the rural classes and for that reason should be added to the category of agricultural laborers in order to account fully for the rural contingent in the emigrant stream."

[14] Ferrie (1997) makes a similar point regarding the distinction between farmers, farm tenants, and unskilled workers

Literacy is another measure of skills that was collected in the passenger lists. Unfortunately, this variable is missing for about 60% of the individuals in the US data. One way to deal with this limitation is to measure literacy in the census cross-sections of 1895 Argentina and 1900 US. A concern with such approach is that differences in the cross-sectional data might exaggerate differences upon arrival if Italians were more likely to accumulate skills in one of the countries. Yet, the data show a relatively small difference between Italians in Argentina and the US with respect to literacy: 64% of the Italians aged 18 to 60 in 1895 Argentina were literate, compared to 59% in 1900 US% (own calculation based on Somoza (1967) and Ruggles et al. (1997)). This difference is much smaller than the difference between southern and northern Italians who remained in Italy: By 1901 only 30% of southerners were literate, compared to 65% of northerners (Klein, 1983).

In settings in which other measures are unavailable, age heaping –the tendency of individuals to report "attractive" numbers as their age, typically multiples of five– has been used as a proxy for quantitative skills. A'Hearn et al. (2009) show that, when age heaping and literacy are both observed, there is a high correlation between the two.[15] To test whether Italians moving to Argentina had higher numeracy, I define an indicator that takes a value of one if an individual reported a multiple of five as his or her age. The table shows that Italians in Argentina were less likely to report a multiple of five as their age but that this difference is very small in magnitude.[16]

Overall, the above analysis indicates that the main difference between Italians who moved to Argentina and those who moved to the US was their region of origin, with smaller or no differences in other pre-migration characteristics. However, one limitation of the analysis is that the available measures of skills are all relatively coarse. For instance, I do not observe an individual's wealth upon arrival or finer measures of human capital such as years of schooling. Spitzer and Zimran

in the US passenger lists of the mid 19th century. He writes: "Farmers are more problematic. This is probably less the result of carelessness on the part of those compiling the lists than it is the result of the presence of large numbers of farm tenants in Europe and the absence of the term 'tenant' in the ship lists. An individual who was a farm tenant in Europe would thus have been described as either a farmer or a laborer in the ship lists. Since there were no clear guidelines by which farm tenants were assigned to either group in the ship lists, we cannot separate those who were farm tenants in Europe (who were probably more akin to laborers in the amount of capital they possessed and the amount of supervision they received) from independent farmers. The same problem applies to a lesser degree for those reported as unskilled in the ship lists."

[15] In the census data described above, there is also a negative correlation between being literate and the likelihood of reporting a multiple of five as age.

[16] A number of studies have used this measure in the context of immigration. Mokyr and Grada (1982) and Collins and Zimran (2018) use this measure to analyze the selection of Irish famine migrants, and Stolz and Baten (2012) use this measure to test whether immigrant selection responded to relative inequality.

(2018) use the heights of Italians entering the US to study immigrant selectivity.[17] Unfortunately, there is no systematic data on heights for Argentina that would allow me to compare Italians in Argentina and the US with respect to this characteristic.[18]

4.3 Differences in the linked data

Italians who moved to Argentina were more likely to be northerners, but there were no large differences in other observable characteristics, including pre-migration occupations, literacy, and proxies for numeracy. How much of the relative advantage of Italians in Argentina can be explained by these differences in pre-migration characteristics?

4.3.1 Occupational mobility from arrival to 1895/1900

I start by constructing occupational transition matrices in which rows represent an immigrant's occupation in Italy, and columns represent his occupation in Argentina or the US. Panel (a) in Table 3 shows this transition matrix for Argentina, whereas panel (b) shows the corresponding matrix for the US. Despite Italian migrants were predominantly from rural backgrounds, they concentrated in urban areas in both Argentina and the US, particularly in the port of entry cities of Buenos Aires and New York. Indeed, a relatively small fraction (about 20%) of Italians in Argentina and the US worked as farmers (last row of each transition matrix). The majority of Italians in the US were employed as unskilled workers, whereas the largest category for Italians in Argentina was that of skilled/semi-skilled workers. Relative to Italians in the US, Italians in Argentina were also more likely to be employed in white-collar jobs.

Turning to the occupational transitions, I find that Italians in Argentina were less likely to experience occupational *downgrading* than Italians in the US. For instance, among Italians with a white-collar job in Europe, only 12% held an unskilled occupation in Argentina. In contrast, among Italians in the US, the chances of landing an unskilled job were substantial (36%) even for those previously employed as white-collar workers. Moreover, the likelihood of moving out of unskilled

[17] They find that migrants moving to the US were positively selected *within* their provinces of origin, but negatively selected overall.This pattern is consistent with the fact that, despite there were large differences in literacy rates between the South and the North and that Argentina attracted more Northerners, the literacy rates of Italians in Argentina and the US were close to each other.

[18] Kosack and Ward (2014) use heights to measure the selectivity of Mexican migration to the US in the early 20th century.

occupations was lower in the US (42%) than in Argentina (77%).

In the last row of the table, I simulate the occupational distribution of Italians in the US had they been exposed to the transition matrix of Italians in Argentina (following the approach in Collins and Wanamaker (2017)). This counterfactual distribution is quite different from the observed one for Italians in the US. For instance, while more than half of the Italians in the US worked in unskilled jobs, the counterfactual fraction of unskilled workers is just 20%. In addition, the counterfactual distribution for Italians in the US is very close to the observed one in Argentina, which is consistent with the fact that the occupations upon arrival were similar for both groups (as documented above).

4.3.2 Differences in home ownership rates and the likelihood of holding an unskilled occupation

I next estimate versions of equation 1 in which I include pre-migration characteristics as control variables. Table 4 shows the results of this exercise. In panel (a), I focus on the likelihood of home ownership, whereas in panel (b) I focus on the likelihood of holding an unskilled job. The first column of each panel (which does not include any controls other than age fixed effects) shows a similar pattern to the cross-sectional results in Table 1: Italians in Argentina were 5.6 percentage points more likely to own their home and 25 percentage points less likely to hold an unskilled occupation.

In the second column, I add indicator variables based on the number of years spent in each of the countries. Adding this variable slightly reduces both coefficients in magnitude (reflecting that Italians in Argentina had on average spent more years at the destination by the time of the censuses). Figure 3 shows the relationship between time spent in each of the countries and outcomes at the destination, net of age fixed effects. With just one cross-section of data, I cannot disentangle cohort effects from years since migration.[19] However, the data suggest that, in both countries, a longer stay was associated with a higher likelihood of home ownership and a lower likelihood of holding an unskilled job. Also, the figure suggests little convergence between Italians in Argentina and those in the US.

[19] An added difficulty is that the census data are from two different points in time in the two countries (1895 and 1900).

In the third column, I further absorb the port of origin of each migrant. This specification enables me to test if the different mix between northern and southern Italians in Argentina and the US could explain their different average outcomes at each of the destinations. This variable makes a difference in the home ownership results (where the coefficient declines by about a third), but makes little difference with respect to the likelihood of holding an unskilled job. The results are nearly identical if I restrict the sample to include only immigrants who departed from Italian ports (Table B3 in the Online Appendix). In Table B4 in the Online Appendix, I split Italians into four mutually excluding groups based on whether they departed from a southern or a northern port, and whether they migrated to Argentina or the US. The table shows that, both in Argentina and the US, northern Italians did better than southern Italians with respect to home ownership rates and the likelihood of holding an unskilled occupation.

In column 4 of Table 4, I include indicators for the occupational category declared upon arrival and for literacy (as reported in the census). Adding these variables increases the predictive power of the regressions (as reflected by the higher R-squared) but has relatively little impact on the estimated coefficients on both outcome variables. This pattern is not surprising given the balancing in these characteristics documented above.[20]

In column 5, I include surname fixed effects, thus comparing immigrants with the same surname but who moved to different destinations.[21] Because of errors in transcribed surnames, I use a phonetically equivalent version of surnames based on NYSIIS, although preserving the last letter of the original surname (as the last letter of a surname is a strong predictor of regional origins in the Italian case).[22] There are two reasons why surnames might provide information beyond the one contained in the other observable characteristics. First, Italian surnames are informative of regional origins within Italy (Spitzer and Zimran, 2018). Hence, exploiting within-surname variation enables me to net out differences in the region of origin of migrants beyond those captured by ports of origin. A further advantage of surnames is that they provide a measure of the region of origin that does not depend on accurately linking the passenger lists to the census. Second, Clark et al. (2015)

[20] One concern with this exercise is that, as discussed above, the passenger lists data is likely not entirely accurate in distinguishing farmers from unskilled workers. To address this issue, in Table B5 I re-estimate this specification while excluding from the sample everyone who reported either farming or an unskilled occupation.

[21] This strategy is used in Bleakley and Ferrie (2016).

[22] Not doing so would result, for instance, in assigning the same surname to migrants whose original surnames were "Russo" or "Rossi".

and Güell et al. (2014) highlight the persistence in a variety of outcomes across family lines. The findings of these studies suggest that surnames might also capture differences in broadly defined social status beyond those captured by occupations.[23]

The results using surname fixed effects (column 5) also indicate a higher likelihood of home ownership and a lower likelihood of being employed in an unskilled occupation for Italians in Argentina. This pattern is confirmed in column 6, where I estimate a regression including surname fixed effects but not including port of origin effects.[24] Note that, conditional on surname fixed effects, ports of origin do not add much predictive power to the regression, as indicated by the very similar R-squared in columns 5 and 6 of both panels. Table B6 in the Online Appendix shows that the results are similar (but less precisely estimated) if I instead include surname times port-of-origin fixed effects, thus enabling the informational content of surnames to depend on the port of origin.[25]

As discussed in Güell et al. (2014), the empirical distribution of Western surnames is highly skewed, with a large percentage of the population having relatively infrequent surnames. As a result, among individuals with rare surnames, surnames can be very informative of family linkages. I exploit this feature of surnames in Figure 4, where I re-estimate equation 1 while progressively excluding individuals with common surnames from the sample. To do so, I focus on surnames that show up at least once in both the Argentine and US datasets, and then rank them in terms of frequency. The idea behind this exercise is that, by progressively excluding common surnames, I am increasingly likely to restrict the comparisons to individuals who are actually related to each other (as, for instance, in Abramitzky et al. (2012) and Collins and Wanamaker (2014) who compare the outcomes of migrants to their siblings). In the last row of the figure, I just include surnames that are in the bottom 10% in terms of frequency (which corresponds to surnames that show up at most three times in the combined Argentina-US linked samples). The results show a similar pattern than in the baseline exercise (if anything, the gaps become larger as I restrict the sample to only include

[23] This specification requires overlap between the surnames of immigrants who moved to Argentina and those who moved to the US. As I will show in section 6, there was strong regional and family path dependence in destination choices, implying that migrants with the same surname tended to go to the same destination country. However, Figure B2 in the Online Appendix shows that there is still some overlap between the surnames of Italians going to Argentina and of those going to the US. In this figure, each dot represents a different surname. The y-axis represents the frequency of a given surname in the Argentine data and the x-axis represents such frequency in the US data.

[24] I estimate this model because the regional clustering of Italian surnames implies that conditional on a surname, there is little variation in ports of origin.

[25] Olivetti and Paserman (2015) instead use the informational content of first names to measure social status. Table B7 in the Online Appendix shows that the results are similar when I simultaneously include both first and last name fixed effects.

more infrequent surnames).

While these results enable me to account for selection to a much larger degree than with the cross-sectional data, I cannot rule out that the difference between Italians in Argentina and those in the US was driven by differences in unobserved characteristics of those moving to each of the countries. To assess the likely role of unobserved characteristics in explaining the differences in outcomes, I compare the stability of the coefficient estimates when expanding the set of control variables as suggested in Altonji et al. (2005). Specifically, I estimate the models that include port of origin fixed effects and assess the sensitivity of the results to adding controls for the skills of immigrants: occupational category fixed effects and literacy. Table B8 shows the results of this exercise: The amount of selection on unobservables would have needed to be as much as 11 times larger than the selection on observables to rationalize the differences in home ownership rates and 37 times larger to rationalize the differences in the likelihood of being employed in an unskilled occupation.[26]

4.4 Robustness to linking

One concern with the results is that incorrect links will result in pre-migration characteristics being measured with error. To address this possibility, Figure 5 progressively excludes relatively lower quality matches from the Argentina and US samples. In the second to last row of the figure, I only include observations with a linking score above the 75th percentile of the distribution of linking scores within the Argentine and US samples. The figure shows a similar pattern regardless of the sample that is used.

Another concern is that the results might be driven by selection into the linked samples. First, note that the results in the cross section (which do not rely upon linking) are consistent with the results that use the linked data. Yet, to further alleviate this concern, in the last row of Figure 5 I reweight the data to account for selection into the linked sample based on observable characteristics upon arrival.[27] The results are similar to those in the baseline sample, suggesting that selection into the linked samples (at least with respect to observable characteristics) is not driving the results.

[26] Collins and Wanamaker (2014) use a similar approach when estimating the gains from internal migration in the context of the Great Migration out of the American South.

[27] To estimate the weights, I estimate a model of the probability of being in the linked sample as a function of observable characteristics upon arrival (year of arrival fixed effects, age fixed effects, literacy, and occupational category fixed effects). I then use the inverse of that estimated probability as the weight.

4.5 Second-generation Italians in Argentina and the US

First-generation Italians in Argentina outperformed those who moved to the US. How persistent was this advantage? To answer this question, I estimate a version of equations 1 and 2, focusing on second-generation Italian immigrants.

A challenge in estimating this equation is that the 1895 Argentine census did not include information on parental place of birth. To obtain this information, I link males from the 1895 census to their childhood household in 1869, where they can be observed living with their parents. As a result, the sample is restricted to native-born males who were 26 to 44 years old in the 1895 census. That is, those who were old enough to be born by 1869 but young enough to still be in their parent's household by that year. Using these data, I am able to distinguish those with Italian parents from those with native-born parents. Further details on the construction of this sample are provided in Pérez (2017).

For the US, it is possible to identify the children of Italian immigrants without linking individuals across censuses, as the 1900 census included a question on parental place of birth. To improve the comparability with the Argentine data, I focus on US-born individuals who were at least 26 and at most 44 years old in 1900. Both in the Argentine and US cases, I define the sample based on the place of birth of the father.

Table 5 shows that the advantage of Italians in Argentina persisted into the second generation. Similar to the results in Table 1, this table shows the differences between children of Italian immigrants in Argentina and the US relative to different groups: the children of the native-born and the children of non-Italian immigrants. The advantage of second-generation Italians in Argentina with respect to home ownership is close to that of the first generation. In contrast, there is a smaller gap in the likelihood of holding unskilled occupations, suggesting some convergence at least in this dimension.

There are two important limitations of this exercise. First, second-generation Italians in this sample are *not* the children of Italian migrants in the linked passenger lists to census data: To enter the passenger lists sample, a migrant had to arrive to the Americas by 1882 or later, but to enter the second-generation sample, a person had to be in Argentina by 1869 or in the US by 1874. Second, because there was little Italian migration to the US before 1880, there are not many

children of Italians in the US that satisfy this condition. Moreover, the Italians who had migrated to the US by 1874 might have represented a relatively selected group of "pioneer" migrants.

5 Mechanisms

The analysis above suggests a limited role for pre-migration characteristics in explaining the differences in economic outcomes between Italians in Argentina and the US. Why were Italians in Argentina faster in assimilating?

5.1 Return migration

Research in history argues that differences in the "expected length of stay" of Italians in Argentina and the US might partly explain the different pattern of assimilation between the two countries.[28] Specifically, if Italians who migrated to Argentina (perhaps because of its closer cultural proximity with Italy) did so with a higher expectation of staying permanently, they might have had stronger incentives to invest in host-country specific capital.

There are three reasons why this mechanism is unlikely to fully explain the results. First, rates of return migration were actually similar in Argentina and the US for migrants in these arrival cohorts. Combining data on population stocks, arrivals, and mortality (following the approach in Bandiera et al. (2013)), I estimate the rate of return migration for the cohorts who entered the Americas in the last three decades of the 19th-century to be about 30% for Argentina and 37% for the US.[29] Of course, rates of return migration are themselves endogenous to outcomes at the destination, so the fact that more Italians returned from the US is not in itself conclusive evidence of weaker intentions to stay in that country. Second, as shown in section 4.2, gender ratios and the fraction of children upon arrival were remarkably close for both groups, indicating a similar predominance of family migration in both flows (at least for these cohorts). Rates of family migration are usually used as a proxy for the expected length of stay of a group, and are empirically

[28] For instance, (Baily, 1983, p. 295) states that "Those who migrated to Buenos Aires included [..] more people who intended to stay."

[29] To perform this calculation, I use the fact that $Emigrants_t = Immigrant\,Stock_{t-1} - Immigrant\,Stock_t + Arrivals_t - Mortality_t$. The information on population stocks is from the censuses of 1870 and 1890 in the US, and 1869 and 1895 in Argentina. The information on arrivals to both countries is from Ferenczi (1929) and the information on mortality rates is from Somoza (1973). The rate of return migration that I estimate for Italians is lower than the one in Bandiera et al. (2013). However, note that Bandiera et al. (2013) focus their analysis in later cohorts of Italian immigrants and that there was an upward trend in rates of return migration in this period.

correlated with rates of return migration (Dustmann and Görlach, 2016). Third, the results are similar when restricting the sample to Italians who had spent at least five years in the Americas and hence were likely to settle permanently (Table B9 in the Online Appendix).[30]

5.2 Location choices

Italians in the US tended to settle in the older regions of the country and predominantly in cities. By 1900, 72% of Italians lived in the Northeast and 75% lived in urban areas. Klein (1983) argues that the concentration of Italians in Northeastern cities hampered their prospects for long-term social mobility, as upward mobility in the US tended to be higher in younger and smaller places. This argument is consistent with recent quantitative evidence on historical differences in mobility across US regions.[31]

Italians in the US were indeed geographically concentrated compared to other immigrant groups, and also relative to Italians in Argentina. Figure B3 shows a residential dissimilarity index computed for Italians (as well as for the four largest sending countries in each destination country) relative to the native born in Argentina and the US. Specifically, I compute (based on an individual's department or county of residence):

$$Dissimilarity\ Index_c = \frac{1}{2}\sum_{i}^{D} \left| \frac{\#Immigrants_{ci}}{\#Immigrants_c} - \frac{\#Natives_i}{\#Natives} \right| \tag{4}$$

where D corresponds to the number of Argentine departments or US counties, i indexes the departments/counties, and c indexes different countries of origin.[32] This index ranges from zero to one, where zero corresponds to full residential segregation and one corresponds to full integration.[33] The figure shows that Italians in the US were more residentially segregated than the other major

[30] Dustmann and Görlach (2016) shows that contemporary rates of outmigration (both out of Europe and out of the US) decline sharply as an individual spends time at the destination.

[31] For instance, Feigenbaum (2017) shows that there was higher mobility in rural areas than in cities in the early 20th century. Similarly, Pérez (2019) finds higher rates of mobility in the Midwest than in the Northeast in the mid 19th century.

[32] Departments/counties are the smallest unit that allow for a meaningful cross-country comparison. There are 312 distinct departments in the 1895 Argentine data and 2826 counties in the US data (US population in 1900 is roughly 19 times the population of 1895 Argentina). My data do not enable to implement the approach in Logan and Parman (2017) in a comparable way between Argentina and the US, as identifying households in the Argentine data is not straightforward due to the absence of a question on relationship to household head.

[33] The index can be interpreted as the fraction of immigrants that would need to switch counties so as to mimic the geographic distribution of natives.

sending countries, as well as more segregated relative to Italians in Argentina.[34]

To assess if the concentration of Italians in cities of the US Northeast can explain the differences with respect to Argentina, I estimate versions of equations 1 and 2 in which I split the US into four different regions (Northeast, Midwest, South and West). Table B10 shows that there was a similar gap across US regions relative to Argentina in the likelihood that Italians would be employed in an unskilled occupation, ranging from 32 percentage points in the Northeast to 20 percentage points in the South. The results are similar regardless of whether I include natives or not as a comparison group.

Italians in the US Midwest and West had a higher likelihood of home ownership than Italians in Argentina. However, this advantage disappears when comparing Italians in Argentina and the US to the respective native-born population. Relative to the native born, Italians in all US regions were less likely to own property than Italians in Argentina. The smallest gap corresponds to the comparison between Argentina and the US West, while all the remaining gaps are of similar size. Overall, while Italians in the US were indeed concentrated geographically in the Northeast, their disadvantage relative to Italians in Argentina was not confined to Italians in this region.

5.3 Competition from natives and other immigrant groups

Italians in the US likely faced more competition from other previously established immigrant groups than Italians in Argentina. By 1870, there were only 17,000 Italians in the US compared to more than 2.5 million migrants from the British Isles and 1.7 million migrants from Germany (Gibson and Jung, 1999). In 1869 Argentina, the 70,000 Italians were the largest immigrant group and constituted 40% of all the European migrants in the country. As a result, Italians who arrived in Argentina likely benefitted from a denser and more established network. In contrast, newly arrived Italians to the US might have only been able to obtain "those jobs scorned by the native born and the second-generation children of immigrants" (Klein, 1983, p. 318).

There are, however, two caveats to the plausibility of this argument. First, standard models of the labor market impacts of immigration predict that the labor market outcomes of Italian migrants should have actually been *hurt* by a larger Italian population at the destination. Moreover, the size

[34] Using the measure of segregation in Logan and Parman (2017), Eriksson and Ward (2018) show that Italians at the turn of the 20th century were indeed among the most residentially segregated immigrant group in the US.

of the Italian flow to Argentina was much larger than the flow to the US when considered relative to the size of the native-born population (US population in 1900 was roughly 19 times larger than in 1895 Argentina). Second, several studies (for instance, Beine et al. (2011) and Spitzer and Zimran (2018)) show that a larger immigrant stock at the destination is usually associated with a more negatively selected migration flow (which would tend to worsen the outcomes of Italians in Argentina relative to the US).

In addition to competition from other immigrant groups, Italians in Argentina and the US also faced labor market competition from natives. While Italians in Argentina and the US had similar levels of human capital upon arrival (at least as captured by their occupational structure upon arrival, literacy, and numeracy levels), Italians in Argentina had higher levels of human capital *relative* to the native born. For instance, among native males aged 18 to 60 years old, 89% were literate in 1900 US but only 54% were literate in 1895 Argentina.

To test whether the advantage of Italians in Argentina was driven by their higher human capital relative to natives, I estimate a version of equation 2 (comparing Italians to natives in Argentina and the US) in which I add literacy as a control variable.[35] Table B11 in the Online Appendix shows that adding this variable reduces the gap in the likelihood of home ownership by 8 percentage points and the gap in the likelihood of holding an unskilled occupation by 10 percentage points. Hence, these results suggest a quantitatively important role for differences in relative human capital in accounting for the advantage of Italians in Argentina. However, the remaining differences between the outcomes of Italians in Argentina and the US are still large in size.

5.4 Linguistic distance

The next hypothesis is that the closer linguistic distance between Italian and Spanish enabled Italians in Argentina to sort into a broader range of occupations. The main implication of this hypothesis is that Italians in Argentina must have had an advantage in accessing communication-intensive occupations compared to Italians in the US (Peri and Sparber, 2009). To test this hypothesis, I classify occupations into "manual" and "non-manual" using the information from HISCLASS. I then estimate versions of equations 1 and 2 with an indicator that takes a value of one if the

[35] As noted above, using literacy as measured in the destination country is an imperfect measure of human capital upon arrival since immigrants might have become literate in the destination country.

individual held a "manual" job as the dependent variable.

Panel (a) of Table 6 shows evidence consistent with this hypothesis. Italians in the US were overrepresented among those employed in manual occupations, both when compared to Italians in Argentina and when including the native-born in the sample. However, compared to Italians in Argentina, Italians in the US were not overrepresented in manual occupations relative to other immigrant groups (column 3), probably reflecting the fact that about 25% of the immigrants in Argentina (or about half of the non-Italian migrants) were from another Spanish-speaking country (Spain).

The US census contains a question on English proficiency, which enables me to further investigate the relationship between language proficiency and occupational attainment. In the US, Italians were less likely to speak English than the average immigrant: About 60% of Italian males aged 18 to 60 spoke English in 1900, compared to 86% among the remaining immigrant groups. How important were these language barriers in preventing Italians from accessing non-manual occupations in the US? To investigate this issue, I use the 1900 US census to estimate:

$$Manual_{ic} = \alpha_c + \beta_1 Speaks\,English_{ic} + \epsilon_{ic} \tag{5}$$

where $Manual_{ic}$ takes a value of one if migrant i from sending country c was employed in a manual occupation in 1900 US, α_c are country of origin fixed effects, and $Speaks\,English_{ic}$ takes a value of one if the migrant reported being able to speak English.[36]

Panel (b) of Table 6 shows that there is an association between English proficiency and the likelihood of being employed in a manual occupation. In column 1, I only include Italians in the sample, whereas in column 2 I include all immigrant groups and control for country of origin fixed effects. Speaking English reduces the likelihood of being employed in a manual occupation by about 9 percentage points, relative to a baseline of 90%. This coefficient probably overestimates the true causal effect of English ability on occupational attainment since this ability is likely positively correlated with other unobserved dimensions of human capital.[37]

However, a back of the envelope calculation suggests that language ability cannot fully explain

[36] Specifically, the 1900 Census asked "Can the person speak English?".

[37] The coefficients are similar in size to those in Ward (2018), who uses within-worker variation to estimate the effects of English ability on labor market outcomes during the age of mass migration.

the overrepresentation of Italians in manual occupations in the US relative to Argentina. Taking the above estimates at face value, if all Italians had been proficient in English, their proportion in manual occupations in the US would have decreased from 90 to 88%. In Argentina, however, only 80% of Italians were employed in a manual occupation. Hence, these calculations suggest that up to 20% ($\frac{90-88}{10}$) of the overrepresentation of Italians in manual occupations in the US can be accounted for by their lower language ability.

The results on the second generation also suggest that linguistic distance cannot fully explain the disadvantage of Italians in the US: In 1900 US, more than 90% of US-born males aged 26 to 44 with an Italian father spoke English. However, the results in section 4.5 show that this group was also at a disadvantage relative to second-generation Italians in Argentina. Overall, while there is evidence in favor of this hypothesis, it seems unlikely to be the only explanation for the observed outcomes of Italians in the Americas.

5.5 Prejudice against Italians

A final hypothesis is that Italians in the US suffered more nativist prejudice and discrimination than Italians in Argentina. While it is hard to prove such a hypothesis empirically (for instance, due to the fact that discrimination might itself be endogenous to economic outcomes at the destination), it is certainly consistent with the existing qualitative evidence. First, it is clear that Italians were culturally closer to the native born in Argentina than in the US: Argentina and Italy are both predominantly Catholic countries, that speak a Romance language (which, as discussed above, might have facilitated both the social and economic integration of Italians in Argentina) and share a common Latin culture. In addition, the elites that governed Argentina during this period had a very positive view of European immigration, which they considered a source of "civilization" for the country.[38]

In the US, in contrast, Italians and other "new immigrants" were the focus of renewed anti-immigrant sentiments by the turn of the 20th century. Indeed, the 1921 and 1924 quota acts were written with the explicit goal of reducing the number of these migrants, among which Italians were

[38] For instance, Article 25 of the 1853 Constitution of Argentina stated that "The Federal Government shall encourage European immigration, and it may not restrict, limit, or burden with any tax whatsoever the entry into Argentine territory of foreigners whose purpose is tilling the soil, improving industries, and introducing and teaching the sciences and the arts."

the largest national group (Goldin, 1994). Nativist prejudice against Italians was also associated with anti-Catholicism (Fouka et al., 2018; Higham, 2002).

The reports of the Immigration Commission, a bipartisan Congressional Commission devoted to the study of European mass migration to the US, contain several instances of discriminatory remarks against Italians, particularly southerners.[39] For instance, the reports argue that northern Italians would be more desirable immigrants because they "..are more easily assimilated than their southern countrymen, who, because of their ignorance, low standards of living, and the supposedly great criminal tendencies among them are regarded by many as racially undesirable" (Dillingham, 1911, p. 177). As this quote makes clear, Italians were also portrayed as having high rates of criminality: The reports even warn about the mass emigration of Italian criminals to the US.[40] Overall, the Commission describes Italians as less desirable for the US than other immigrant groups.[41]

Italians in the US were even the subjects of nativist violence, particularly in the US South. In the 1890s, more than 20 Italians were lynched in the US (LaGumina, 1999). The worst episode of violence took place in New Orleans in 1891, where a mob killed 11 Italians after a jury found them to be innocent of a killing. Another famous example is the 1927 execution of two Italian immigrants, Sacco and Vanzetti, who were accused of robbery and murder. This execution resulted in international protests due to the perceived bias of the judge and the prosecution.

This evidence can be complemented with historical survey data collected with the aim of measuring the prejudice of Americans against various nationalities and races. Bogardus (1928) conducted a survey in which 1725 (predominantly white) American college students were asked to express their preferences over 40 different races and nationalities.[42] The survey asked students several questions, including whether they would be willing to accept someone from a given group

[39] Southern Italians were considered to be of a different "race" than northern Italians in the writings of the Commission. "The Bureau of Immigration, following the general practice of ethnologists, divides the people of Italy into two races North Italians and South Italians" (Dillingham, 1911, p. 141). Southern Italians were not considered completely white by Anglo-Saxon standards. "Sardinians have a considerable infusion of Spanish blood, while the Neopolitans are said to incline slightly toward the African or negro type." (Dillingham, 1911, p. 177)

[40] "It is certain that many Italian criminals, both those who had served sentences and others who had escaped punishment, have come to the United States during the past 30 years. It was frequently stated to members of the Commission in southern Italy and Sicily that crime had greatly diminished in many communities because most of the criminals had gone to America." (Dillingham, 1911, p. 209)

[41] "Treating an Italian from that standpoint, you will find the Italian as good as any of the other races from farther north, with the possible exception of those from Scandinavia and the British Isles." (Dillingham, 1911, p. 181). "As to the comparative value of the German and Italian emigrants to the south, I would say that the German has desirable qualities which the Italian has not. That is, the German is very solid; very solid; he does not get angry quickly." (Dillingham, 1911, p. 182)

[42] These data were recently used by Hilger (2016).

as a marriage partner, a friend, or a neighbor. Figure B4 shows that Italians fared similarly to other Southern Europeans in terms of the students' preferences but consistently below Northern Europeans. For example, only 15.4% of the students declared to be willing to marry an Italian. Similarly, only 25% to be willing to accept an Italian as member of one's social club, and 35% to be willing to accept an Italian as a neighbor. For reference, the comparable figures for Germans are 54, 67 and 78.7%, and for Scottish are 78.1, 89.1 and 91.3%, respectively. It is worth noting that while Southern Europeans were ranked consistently below Northern Europeans, they were still ranked above non-white groups such as African Americans, Asians and Mexicans.

Finally, rates of out-marriage (considered the "last" stage of assimilation (Gordon, 1964)) were extremely low for Italian immigrants in the US: only 5% of married Italian men in the US were married to a non-Italian in 1900, and just 4% were married to a US-born person (own calculation based on Ruggles et al. (1997)).[43] Figure B5 in the Online Appendix shows that Italians ranked consistently below other immigrant groups in their rates of out-group marriage.[44] For comparison, Baily (1980) shows that a much larger proportion (approximately 35%) of Italian men in Argentina married outside of their group in late 19th century Buenos Aires.[45] This is despite similar gender ratios upon arrival in both destination countries, and despite Italians were a much larger fraction of the population in Argentina than in the US (which would mechanically tend to increase in-group marriage rates).

6 Understanding destination choices

Italian migrants and their children had higher rates of home ownership and were less likely to be employed in unskilled occupations when they migrated to Argentina. Why did (in an era of open borders) some Italians choose a destination that offered them limited prospects for upward mobility?

[43] These rates are similarly low when restricting the analysis to those who likely got married in the US (based on their reported duration of marriage and year of arrival to the country).

[44] Because Italian migration was more recent than that of the other major sending countries, there were fewer second-generation Italians (which would tend to exaggerate the degree of Italian homogamy relative to other countries of origin). Hence, in this figure, marriages with a second-generation immigrant from the same origin country are also counted as "in-group" marriage.

[45] The 1895 census of Argentina is not well suited for computing rates of intermarriage, as it does not include a question on relationship to household head. In addition, because women in Argentina preserve their surname upon marriage, it is not straightforward to identify married couples in the data in the absence of information on relationship to household head.

One potential explanation is that migrants deciding between Argentina and the US faced a trade-off between higher wages in the short term (when they moved to the US) and higher chances of upward mobility in the long run (when they moved to Argentina). Indeed, while by 1900 Argentina was in the world's top five with respect to per-capita income, wages for unskilled workers were still about 25% lower than in the US (Williamson, 1995).[46]

While this hypothesis is plausible, a limitation with it is that it does not successfully explain *who* moved to each of the destinations. First, an implication of this hypothesis is that Italians who moved to Argentina should have been on average younger than those who went to the US (as Argentina offered lower initial wages but higher upward mobility). Yet, the age structure upon arrival was very similar for both groups. Second, this hypothesis also implies that the US should have been a relatively more attractive options for migrants who only wished to stay temporarily. However, as discussed above, rates of return migration for these cohorts of immigrants were actually similar in both destination countries. Third, this hypothesis is unlikely to explain why Italians from certain provinces were disproportionately more likely to move to one destination over the other.[47]

Another potential explanation is that immigrant networks (by reducing costs and mitigating the risks associated with migration) generated path dependence in destination choices within the Americas. In other words, for Italians deciding where to migrate, having relatives or friends in one of the destinations might have been the primary consideration.[48] Indeed, a number of scholars point to path dependence as the key driver of Italian migrants' destination choices within the Americas. For instance, Moretti (1999) writes that "In other words, neither wage differentials nor the probability of employment nor transportation costs can explain the choice of the destination country."[49] Moreover, the US Immigration Commission reported that more than 90% of Italians

[46] Note that this difference does not take into account that the gap between average wages and wages received by Italian workers might have been different in Argentina and the US. Both Argentina and the US had much higher wages than Italy: The ratio between US and Italian wages was 3.8 by 1900.

[47] Another potential explanation is that the cost of the passage to the US was lower than the cost to Argentina, which resulted in Argentina attracting more positively selected migrants and more migrants with the intention of staying. However, while the trip from Genoa or Naples to Buenos Aires was about twice as long as the trip to New York, there were no large differences in the monetary costs of going to each of these destinations. Baily (2004) reports that, in 1902, the average cost of the trip was 173.8 lire from Genoa or Naples to New York and 174.6 lire to Buenos Aires. According to Baily (2004), "the cost of passage was of marginal significance in the choice between the two destinations."

[48] Arroyo Abad and Sánschez-Alonso (2015) show that ethnic networks were an important consideration in the occupational choices of Italians in late 19th century Buenos Aires.

[49] Hatton and Williamson (1998) shows that the stock of emigrants from a given Italian province across different destination countries had a strong association with subsequent province-level migration flows. A similar argument is made by Gould (1980), who describes the process of spatial diffusion through which emigration spread across Italian

who entered the US were joining either friends or relatives (Baily, 2004).

I next use the passenger lists data to test whether Italian migrants were more likely to migrate to the destination to which their "friends and relatives" had migrated in the past. Because I do not directly observe family or friend relationships among immigrants, I use the surnames of previous migrants to Argentina and the US to compute a proxy measure of the strength of a migrant's network at each potential destination. Specifically, for each immigrant i with surname s arriving at time t, I compute the "Argentina Surname Index" (ASI) as:

$$ASI_{ist} = \frac{\frac{\#Italians\,with\,surname\;s\,in\,Argentina}{\#Italians\,in\,Argentina}}{\frac{\#Italians\,with\,surname\;s\,in\,Argentina}{\#Italians\,in\,Argentina} + \frac{\#Italians\,with\,surname\;s\,in\,US}{\#Italians\,in\,US}} \tag{6}$$

where the number of Italians in each of the sending countries is computed based on the number of arrivals up to year $t-1$.

This measure is based on Fryer Jr and Levitt (2004) and has been more recently used in economic history by Abramitzky et al. (2016).[50] It takes a value of one if the immigrant has a surname that, up to year $t-1$, can only be found among arrivals to Argentina and takes a value of zero if the migrant has a surname that can only be found among arrivals to the US.[51] Because ASI_{ist} is undefined for immigrants whose surnames show up for the first time at time t, in the main analysis I assign a value of 0.5 to these surnames (reflecting that they are "neutral" surnames). However, the results are similar if I instead drop these observations from the sample.

This index is an imperfect proxy of the relative network strength of a migrant. For instance, it does not capture connections to friends or relatives with a different surname. In addition, it introduces noise because some individuals that share a surname might not know each other (although due to the regional clustering of Italian surnames, ASI will also capture past migrations of individuals with similar regional origins).

Figure 6 shows a binned scatterplot of the likelihood of moving to Argentina (y-axis) on the ASI (x-axis). The figure clearly shows that Italians arriving in Argentina had surnames that were

regions. Spitzer (2013) studies spatial diffusion in the context of Jewish emigration from Russia.

[50] Connor (2018) and Wegge (1998) also use surnames to measure immigrant's networks in historical data. My measure is close in spirit to that in Connor (2018).

[51] For Argentina, the stock of Italian surnames in 1882 is constructed using data on Italian surnames in the 1869 census of population. For the US, the stock of Italian surnames in 1882 is measured using the surnames of immigrants arriving from 1855 to 1881 in the passenger lists data.

distinctively associated with those of previous Italian migrants to Argentina. To more formally test for the role of relative network strength in explaining migration choices, I estimate the following model:

$$Argentina_{ist} = \alpha_t + \beta\, ASI_{ist} + \gamma X_{ist} + \epsilon_{ist} \tag{7}$$

where $Argentina_{ist}$ is an indicator that takes a value of one if the observation belongs to the Argentine passenger lists, α_t represent arrival year fixed effects and X_{ist} is a vector of individual-level characteristics. The coefficient of interest is β, which measures the association between migration decisions and the relative network strength of a migrant at each of the destinations (as captured by ASI_{ist}). I emphasize that β does not necessarily capture a causal relationship, as there might be persistent characteristics that attracted immigrants to certain destinations within the Americas and which were also correlated with surnames, a common issue when measuring network effects using observational data (Manski, 1993).

Table 7 shows that ASI_{ist} is a strong predictor of whether an individual moved to Argentina or the US. Going from a distinctively "Argentine" to a distinctively "American" surname is associated with an increase in the probability of moving to Argentina of almost 50 percentage points (more than doubling the unconditional probability). Column 2 shows that the results are very similar when I control for a vector of individual-level characteristics: gender, age, and the occupational category declared upon arrival. Indeed, there is little increase in the predictive power of the regression from adding these variables, which is unsurprising given the limited amount of selection on observables documented above.[52]

Table 7 also shows some heterogeneous effects consistent with a path dependence interpretation of the results. Columns 3 and 4 show that the association is stronger for women and for migrants who were less than 16 years old at the time of arrival. This pattern is consistent with adult males migrating first and then being followed by their families. The last column (where I focus on working-age males) shows that the association is stronger for those employed as either farmers or unskilled workers prior to migrating. This pattern is consistent with the migration decisions of

[52]This exercise raises a concern about the within-surname comparisons in section 4.3. Specifically, that individuals who moved to one destination despite having a surname strongly associated with the other destination (that is, a value of ASI close to either zero or one) might have been selected in an idiosyncratic way. To address this concern, in Figure B6 in the Online Appendix I show that the results are similar when restricting the within-surname comparisons to individuals with relatively "neutral" surnames (that is, surnames which are not indicative of a preference for either Argentina or the US).

the unskilled being more sensitive to the presence of networks, perhaps due to liquidity constraints (McKenzie and Rapoport, 2007, 2010; Moraga, 2013; Wegge, 1998). Note, however, that in all cases the baseline effects are still large in size, suggesting that networks were important for migrants in general and not only for certain subgroups.[53]

7 Conclusions

Seven million Italians moved to either Argentina or the US during the age of mass migration. Prior work shows that Italians had faster assimilation in Argentina than in the US, but is inconclusive on whether this was due to differences in selection or to differences in host-country conditions. Using data linking Italian immigrants from passenger lists to population censuses, I showed that differences in observable pre-migration characteristics cannot explain the relative economic success of Italians in Argentina.

What are the broader implications of these findings? First, despite being a large group relative to the native population, Italians were able to successfully integrate into the economy of Argentina. This success suggests that the size of an immigrant group might not be *per se* an important driver of assimilation. Second, although Argentina and the US followed a similar policy of open borders for European immigration, Italians fared very differently in each of these countries. This finding raises doubts about the importance of immigration policy as a determinant of assimilation, and highlights instead the importance of host-country conditions. Finally, my findings provide further evidence that, even in the absence of policies such as family reunification visas, immigrant networks can play a crucial role in shaping migrants' destination choices.

References

Abramitzky, R., L. P. Boustan, and K. Eriksson (2012): "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration," *American Economic Review*, 102, 1832–1856.

[53] One concern with these results is that they might be driven by common misspellings of Italian surnames in the Argentine and US datasets. Table B12 in the online appendix shows that the results are nearly identical when computing ASI_{ist} based on a NYSIIS standardized version of the surname.

——— (2014): "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration," *Journal of Political Economy*, 122.

——— (2016): "Cultural assimilation during the age of mass migration," Tech. rep., National Bureau of Economic Research.

Abramitzky, R., L. P. Boustan, K. Eriksson, and S. Feigenbaum, James J.and Pérez (2018): "Automated Linking of Historical Data," Tech. rep., NBER WP 25825.

Abramitzky, R., R. Mill, and S. Pérez (2019): "Linking individuals across historical sources: A fully automated approach*," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 0, 1–18.

A'Hearn, B., J. Baten, and D. Crayen (2009): "Quantifying quantitative literacy: Age heaping and the history of human capital," *The Journal of Economic History*, 69, 783–808.

Alexander, R., Z. Ward, et al. (2018): "Age at Arrival and Assimilation during the Age of Mass Migration," *Journal of Economic History*.

Altonji, J. G., T. E. Elder, and C. R. Taber (2005): "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools," *Journal of political economy*, 113, 151–184.

Arroyo Abad, L. and B. Sánschez-Alonso (2015): "A city of trades: Spanish and Italian Immigrants in Late Nineteenth Century Buenos Aires. Argentina," Tech. rep., European Historical Economics Society (EHES).

Bailey, M., C. Cole, M. Henderson, and C. Massey (2017): "How Well Do Automated Linking Methods Perform in Historical Samples? Evidence from New Ground Truth," Tech. rep., Working Paper.

Baily, S. L. (1980): "Marriage patterns and immigrant assimilation in Buenos Aires, 1882-1923," *The Hispanic American Historical Review*, 60, 32–48.

——— (1983): "The Adjustment of Italian Immigrants in Buenos Aires and New York, 1870-1914," *The American Historical Review*, 281–305.

——— (2004): *Immigrants in the lands of promise: Italians in Buenos Aires and New York City, 1870-1914*, Cornell University Press.

Bandiera, O., I. Rasul, and M. Viarengo (2013): "The making of modern America: Migratory flows in the age of mass migration," *Journal of Development Economics*, 102, 23–47.

Beine, M., F. Docquier, and c. Ozden (2011): "Diasporas," *Journal of Development Economics*, 95, 30–41.

Biavaschi, C., C. Giulietti, and Z. Siddique (2017): "The economic payoff of name Americanization," *Journal of Labor Economics*, 35, 1089–1116.

Bleakley, H. and J. Ferrie (2016): "Shocking behavior: Random wealth in antebellum Georgia and human capital across generations," *The quarterly journal of economics*, 131, 1455–1495.

Bogardus, E. S. (1928): "Immigration and race attitudes." .

Carneiro, P., S. Lee, and H. Reis (2017): "Please call me John: name choice and the assimilation of immigrants in the United States, 1900-1930," .

Catron, P. (2016): "Made in America? Immigrant Occupational Mobility in the First Half of the Twentieth Century," *American Journal of Sociology*, 122, 325–378.

Clark, G., N. Cummins, Y. Hao, and D. D. Vidal (2015): "Surnames: A new source for the history of social mobility," *Explorations in Economic History*, 55, 3–24.

Coletti, F. (1912): *Dell'emigrazione italiana*, U. Hoepli.

Collins, W. J. and M. H. Wanamaker (2014): "Selection and Economic Gains in the Great Migration of African Americans: New Evidence from Linked Census Data," *American Economic Journal: Applied Economics*, 6, 220–252.

——— (2017): "Up from Slavery? African American Intergenerational Economic Mobility Since 1880," Tech. rep., National Bureau of Economic Research.

Collins, W. J. and A. Zimran (2018): "The Economic Assimilation of Irish Famine Migrants to the United States," Tech. rep., National Bureau of Economic Research.

Connor, D. S. (2018): "The cream of the crop? Inequality and migrant selectivity in Ireland during the age of mass migration," .

de Inmigración, D. G. (1925): "Resumen Estadístico del Movimiento Migratorio en la República Argentina. Años 1857–1924," .

Dillingham, W. P. (1911): *Emigration conditions in Europe*, vol. 4, US Government Printing Office.

Dustmann, C. and J.-S. Görlach (2016): "The economics of temporary migrations," *Journal of Economic Literature*, 54, 98–136.

Eriksson, K. and Z. A. Ward (2018): "The Ethnic Segregation of Immigrants in the United States from 1850 to 1940," Tech. rep., National Bureau of Economic Research.

Federico, G., A. Nuvolari, and M. Vasta (2019): "The origins of the Italian regional divide: Evidence from real wages, 1861–1913," *The Journal of Economic History*, 79, 63–98.

Feigenbaum, J. J. (2017): "Multiple Measures of Historical Intergenerational Mobility: Iowa 1915 to 1940," *mimeo*.

Ferenczi, I. (1929): "International Migration Statistics," in *International Migrations, Volume I: Statistics*, NBER, 47–76.

Ferrie, J. P. (1994): "The wealth accumulation of antebellum European immigrants to the US, 1840–60," *The Journal of Economic History*, 54, 1–33.

——— (1997): "The entry into the US labor market of antebellum European immigrants, 1840–1860," *Explorations in Economic History*, 34, 295–330.

Fouka, V., S. Mazumder, and M. Tabellini (2018): "From Immigrants to Americans: Race and Assimilation during the Great Migration," *Harvard Business School BGIE Unit Working Paper*.

Fryer Jr, R. G. and S. D. Levitt (2004): "The causes and consequences of distinctively black names," *The Quarterly Journal of Economics*, 767–805.

Gibson, C. and K. Jung (1999): "Historical Census Statistics on the Foreign-born Population of the United States: 1850-1990." Population Division Working Paper," .

Goldin, C. (1994): "The political economy of immigration restriction in the United States, 1890 to 1921," in *The regulated economy: A historical approach to political economy*, University of Chicago Press, 223–258.

Gomellini, M. and C. O'Grada (2011): "Outward and inward migrations in Italy: a historical perspective," .

Gordon, M. M. (1964): *Assimilation in American life: The role of race, religion, and national origins*, Oxford University Press on Demand.

Gould, J. D. (1980): "European International Emigration: The Role of Diffusion and Feedback," *Journal of European Economic History*, 9, 267.

Green, A. and M. MacKinnon (2001): "The slow assimilation of British immigrants in Canada: evidence from Montreal and Toronto, 1901," *Explorations in Economic History*, 38, 315–338.

Güell, M., J. V. Rodríguez Mora, and C. I. Telmer (2014): "The informational content of surnames, the evolution of intergenerational mobility, and assortative mating," *The Review of Economic Studies*, 82, 693–735.

Guglielmino, C. and A. De Silvestri (1995): "Surname sampling for the study of the genetic structure of an Italian province," *Human biology*, 613–628.

Hatton, T. J. (1997): "The Immigrant Assimilation Puzzle in Late Nineteenth-Centuty America," *The journal of economic history*, 57, 34–62.

Hatton, T. J. and J. G. Williamson (1998): *The age of mass migration: Causes and economic impact*, Oxford University Press on Demand.

Higham, J. (2002): *Strangers in the land: Patterns of American nativism, 1860-1925*, Rutgers University Press.

Hilger, N. (2016): "Upward mobility and discrimination: The case of Asian Americans," Tech. rep., National Bureau of Economic Research.

Ichino, A. and G. Maggi (2000): "Work environment and individual background: Explaining regional shirking differentials in a large Italian firm," *The Quarterly Journal of Economics*, 115, 1057–1090.

Inwood, K., C. Minns, and F. Summerfield (2016): "Reverse assimilation? Immigrants in the Canadian labour market during the Great Depression," *European Review of Economic History*, 20, 299–321.

Klein, H. S. (1983): "The Integration of Italian Immigrants into the United States and Argentina: A Comparative Analysis," *The American Historical Review*, 306–329.

——— (1989): "A integração dos imigrantes italianos no Brasil, na Argentina e Estados Unidos," *Novos Estudos CEBRAP*, 25, 95–117.

Kosack, E. and Z. Ward (2014): "Who Crossed the Border? Self-Selection of Mexican Migrants in the Early Twentieth Century," *The Journal of Economic History*, 74, 1015–1044.

LaGumina, S. J. (1999): *Wop!: A Documentary History of Anti-Italian Discrimination in the United States*, 32, Guernica Editions.

Leeuwen, M. v., I. Maas, and A. Miles (2002): *HISCO: Historical international standard classification of occupations*, Leuven: Leuven University Press.

Logan, T. D. and J. M. Parman (2017): "The national rise in residential segregation," *The Journal of Economic History*, 77, 127–170.

Manski, C. F. (1993): "Identification of endogenous social effects: The reflection problem," *The review of economic studies*, 60, 531–542.

Martínez, A. B. (1910): *Censo general de población, edificación, comercio é industrial de la ciudad de Buenos Aires, conmemorativo del primer centenario de la revolución de mayo 1810-1910, levantado en los días 16 al 24 de octubre de 1909, bajo la administración del Señor intendente don Manuel J. Güiraldes*, vol. 2, Compañía sud-americana de billetes de banco.

McKenzie, D. and H. Rapoport (2007): "Network effects and the dynamics of migration and inequality: theory and evidence from Mexico," *Journal of development Economics*, 84, 1–24.

——— (2010): "Self-selection patterns in Mexico-US migration: the role of migration networks," *The Review of Economics and Statistics*, 92, 811–821.

Minns, C. (2000): "Income, cohort effects, and occupational mobility: a new look at immigration to the United States at the turn of the 20th century," *Explorations in economic history*, 37, 326–350.

Mokyr, J. and C. Ó. Grada (1982): "Emigration and Poverty in Prefamine Ireland," *Explorations in Economic History*, 19, 360.

Moraga, J. F.-H. (2013): "Understanding different migrant selection patterns in rural and urban Mexico," *Journal of Development Economics*, 103, 182–201.

Moretti, E. (1999): "Social networks and migrations: Italy 1876–1913," *International Migration Review*, 33, 640–657.

Moya, J. C. (1998): *Cousins and strangers: Spanish immigrants in Buenos Aires, 1850-1930*, Univ of California Press.

Olivetti, C. and M. D. Paserman (2015): "In the name of the son (and the daughter): Intergenerational mobility in the United States, 1850–1940," *The American Economic Review*, 105, 2695–2724.

Pérez, S. (2017): "The (South) American Dream: Mobility and Economic Outcomes of First-and Second-Generation Immigrants in Nineteenth-Century Argentina," *The Journal of Economic History*, 77, 971–1006.

Pérez, S. (2019): "Intergenerational Occupational Mobility Across Three Continents," *The Journal of Economic History*, 79, 383–416.

Peri, G. and C. Sparber (2009): "Task specialization, immigration, and wages," *American Economic Journal: Applied Economics*, 1, 135–69.

Ruggles, S., M. Sobek, C. A. Fitch, P. K. Hall, and C. Ronnander (1997): *Integrated public use microdata series: Version 2.0*, Historical Census Projects, Department of History, University of Minnesota.

Somoza, J. L. (1967): *Muestras de los dos primeros censos nacionales de población, 1869 y 1895*, 46, Instituto Torcuato Di Tella, Centro de Investigaciones Sociales.

——— (1973): "La mortalidad en la Argentina entre 1869 y 1960," *Desarrollo Económico*, 12, 807–826.

Spitzer, Y. (2013): "The Dynamics of Mass Migration: Estimating the Effect of Income Differences on Migration in a Dynamic Model of Discrete Choice with Diffusion," .

Spitzer, Y. and A. Zimran (2018): "Migrant self-selection: Anthropometric evidence from the mass migration of Italians to the United States, 1907–1925," *Journal of Development Economics*, 134, 226–247.

Stolz, Y. and J. Baten (2012): "Brain drain in the age of mass migration: Does relative inequality explain migrant selectivity?" *Explorations in Economic History*, 49, 205–220.

Taylor, A. M. (2018): "The Argentina Paradox: microexplanations and macropuzzles," *Latin American Economic Review*, 27, 3.

Taylor, A. M. and J. G. Williamson (1997): "Convergence in the age of mass migration," *European review of economic history*, 1, 27–63.

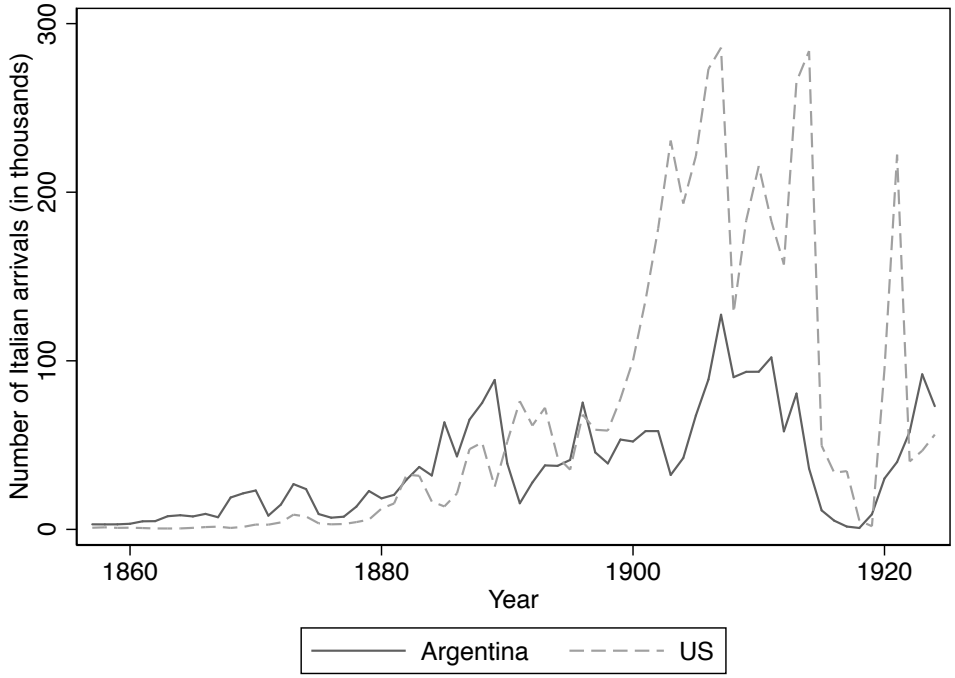
Ward, Z. (2018): "Have language skills always been so valuable? The low return to English fluency during the Age of Mass Migration," *Unpublished paper, Australian National University.*

Wegge, S. A. (1998): "Chain migration and information networks: Evidence from nineteenth-century Hesse-Cassel," *The Journal of Economic History*, 58, 957–986.

Williamson, J. G. (1995): "The evolution of global labor markets since 1830: background evidence and hypotheses," *Explorations in Economic History*, 32, 141–196.

Winkler, W. E. (1990): "String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage." .

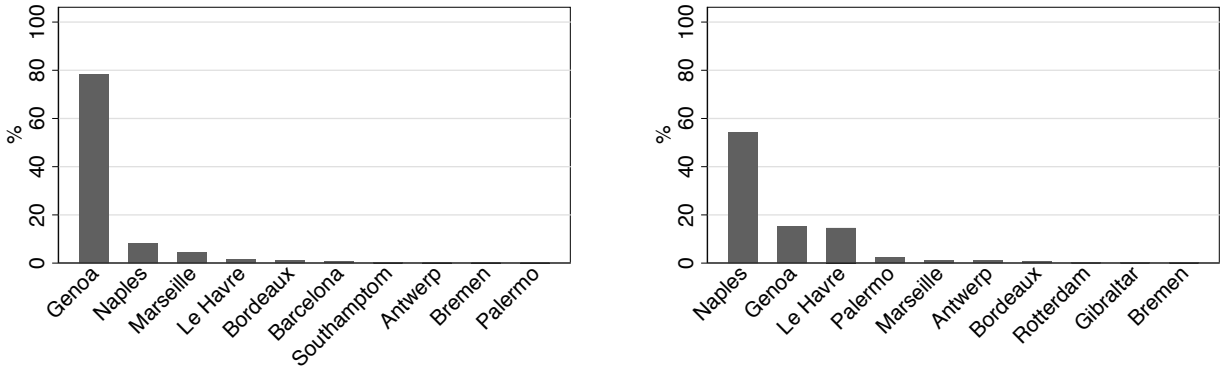
Figure 1: Number of yearly Italian arrivals in Argentina and the US



Notes: This figure shows the number of yearly Italian arrivals to Argentina and the US from 1857 to 1924.
Source: Ferenczi (1929).

Figure 2: Ports of departure of Italian migrants (1882-1900)

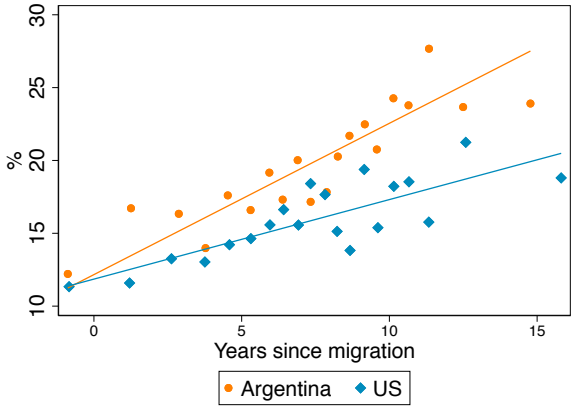
(a) Argentina **(b)** US



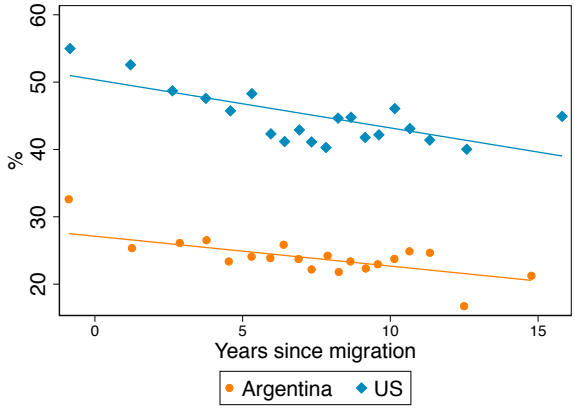
Notes: This figure shows the ten largest ports of departure of Italian immigrants to Argentina and the US.
Source: Passenger lists data as described in the main text.

Figure 3: Main economic outcomes, by years since migration

(a) Home ownership



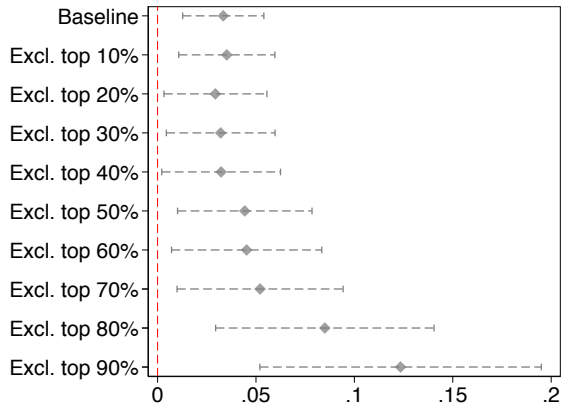
(b) Unskilled occupation



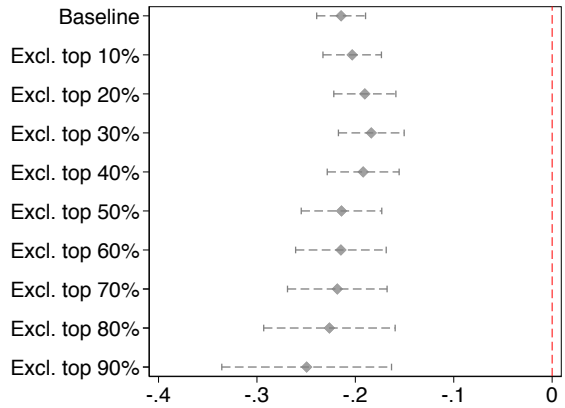
Notes: This figure shows a binned scatterplot of the main economic outcomes (y-axis) on years since migration (x-axis), net of age fixed effects, by country of destination.
Source: Data are from the samples linking passenger lists to the census as described in the main text.

Figure 4: Surname fixed effects, excluding common surnames

(a) Home ownership



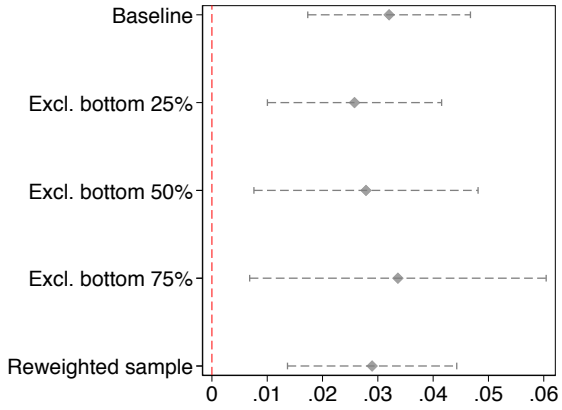
(b) Unskilled occupation



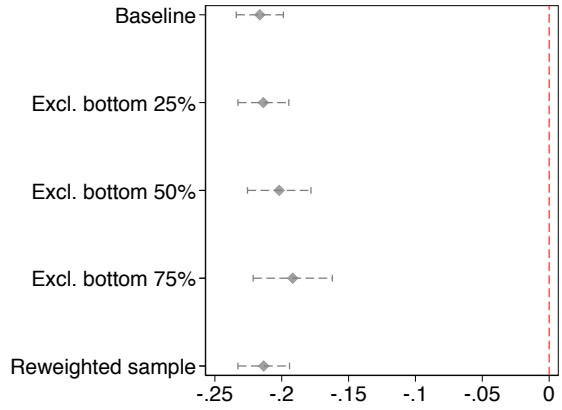
Notes: This figure shows the results of the specification using surname fixed effects after progressively excluding individuals with common surnames from the sample. In the last row, I only include those with a surname in the bottom 10% of frequency among those surnames that show up at least once in both the Argentine and US datasets. The specification corresponds to the one with the largest set of controls in Table 4 (including surname fixed effects).
Source: Data are from the samples linking passenger lists to the census as described in the main text.

Figure 5: Robustness to linking procedure

(a) Home ownership

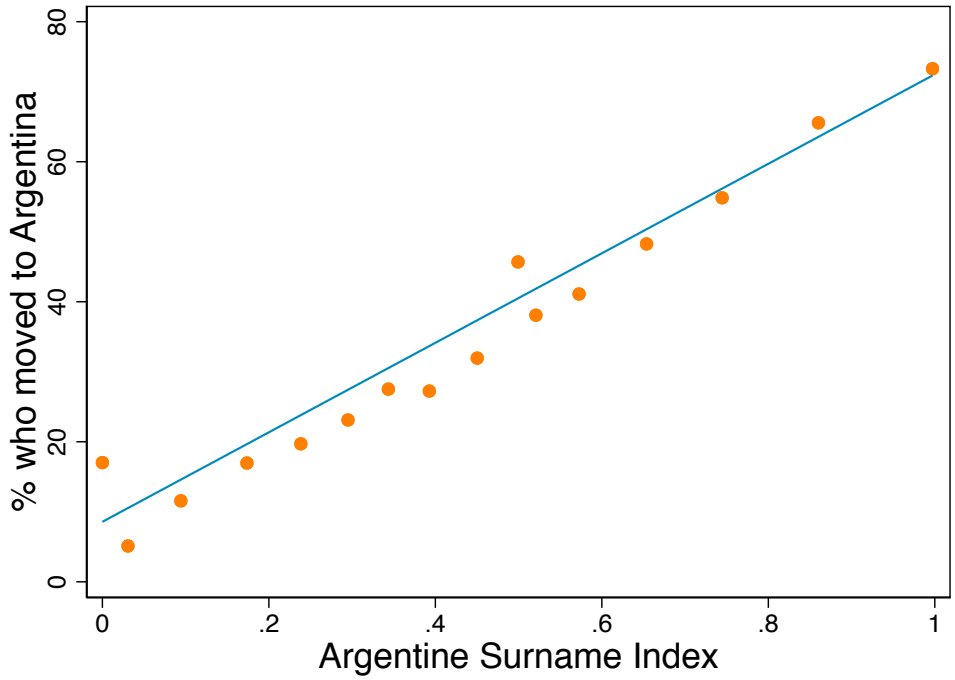


(b) Unskilled occupation



Notes: This figure shows the robustness of the results to progressively increasing the quality of matches, and to selection into the linked samples. In rows 2 to 4, I progressively exclude observations with a linking score in the bottom 25, 50 and 75% within the Argentina and US samples. In the last row, I reweight the sample to account for selection into the linked sample based on observable characteristics.
Sources: Data are from the samples linking passenger lists to the census as described in the main text.

Figure 6: Argentine Surname Index and the likelihood of moving to Argentina



Notes: This figure shows a binned scatterplot of the probability of moving to Argentina (y-axis) on the *Argentine Surname Index (ASI)* (x-axis). This index is computed based on the relative frequency of the surnames of all previous arrivals to Argentina and the US. A value of one indicates a surname that was only held by Italians who previously moved to Argentina, whereas a value of zero indicates a surname that was only held by Italians who moved to the US.
Source: Passenger lists data as described in the main text.

Table 1: First-generation immigrants in Argentina and the US, cross-sectional data

	Home ownership				Unskilled occupation			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Italian in Argentina	0.0430***	0.259***	0.227***	0.262***	-0.279***	-0.274***	-0.176***	-0.249***
	(0.00552)	(0.00841)	(0.00902)	(0.00808)	(0.00714)	(0.00740)	(0.00877)	(0.00722)
Including natives	No	Yes	No	Yes	No	Yes	No	Yes
Including other immigrants	No	No	Yes	Yes	No	No	Yes	Yes
Observations	19699	857935	244109	1082345	19699	857935	244109	1082345
Mean of dep. var.	0.144	0.144	0.144	0.144	0.490	0.490	0.490	0.490

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. This table reports differences between Italians in Argentina and the US with respect to rates of home ownership and the likelihood of holding an unskilled occupation. The mean of the dependent variable is computed among Italians in the United States. In columns 1 and 5, the sample is restricted to first-generation Italian migrants. In columns 2 and 6, I also include native-born individuals in Argentina and the US. I columns 3 and 7, I compare Italians to other immigrant groups in Argentina and the US. In columns 5 and 8, I compare Italians to the rest of the working-age population.
Sources: Argentine data are from Somoza (1967) and US data are from Ruggles et al. (1997).

Table 2: Selection of Italian immigrants in Argentina and the US

Variable	Mean (US) (1)	Argentina-US difference (2)	(3)	(4)
i. Demographic (N=1,350,461)				
Age	27.001	-0.286a	-0.178a	0.319a
Age less than 16	0.198	0.009a	0.015a	-0.009a
Male	0.743	-0.020a	-0.032a	0.012a
ii. Skills (N=1,350,461)				
Age heaping	0.220	-0.001	-0.006a	-0.010a
iii. Occupation (N= 755,764)				
White-collar	0.028	0.007a	0.015a	0.011a
Farmer	0.221	0.208a	0.206a	0.133a
Skilled/Semi-skilled	0.148	-0.039a	-0.023a	-0.023a
Unskilled	0.603	-0.176a	-0.199a	-0.121a
Year of entry FE	.	No	Yes	Yes
Port of origin FE	.	No	No	Yes

Notes: $a : p < 0.01$, $b : p < 0.05$, $c : p < 0.1$. This table compares Italian migrants who moved to Argentina and the US with respect to observable characteristics upon arrival. In column 1, I report the average value of each of these characteristics in the US data. Columns 2 to 4 report the coefficient of a regression of each of these variables on an indicator that takes a value of one if the observation belongs to the Argentine data. Sample is restricted to 1882-1900 arrivals.
Source: Passenger lists data as described in the main text.

Table 3: Occupational mobility from arrival and until the next census

Occupation in Italy	Occupation in destination				
	White collar	Farmer	Skilled/semi-skilled	Un-skilled	Row total
Argentina					
White Collar	**0.50**	0.10	0.28	0.12	1
	(**188**)	(39)	(105)	(46)	(378)
Farmer	0.18	**0.25**	0.33	0.25	1
	(939)	(**1307**)	(1742)	(1307)	(5295)
Skilled/semi-skilled	0.16	0.11	**0.57**	0.15	1
	(223)	(148)	(**783**)	(209)	(1363)
Unskilled	0.18	0.22	0.37	**0.23**	1
	(1013)	(1209)	(2028)	(**1301**)	(5551)
Column total	0.19	0.21	0.37	0.23	1
	(2363)	(2703)	(4658)	(2863)	(12587)
US					
White Collar	**0.30**	0.13	0.22	0.36	1
	(**136**)	(58)	(99)	(167)	(460)
Farmer	0.11	**0.12**	0.19	0.57	1
	(379)	(**413**)	(645)	(1930)	(3367)
Skilled/semi-skilled	0.11	0.26	**0.21**	0.42	1
	(298)	(717)	(**587**)	(1144)	(2746)
Unskilled	0.11	0.14	0.16	**0.58**	1
	(957)	(1187)	(1406)	(**4987**)	(8537)
Column total	0.12	0.16	0.18	0.54	1
	(1770)	(2375)	(2737)	(8228)	(15110)
Counterfactual Italians in the US					
	0.19	0.20	0.39	0.22	100

Notes: Each cell in the table shows the percentage and the number (between brackets) of individuals in each occupational category upon arrival to Argentina or the US (rows) and in the 1895 or 1900 censuses (columns). Occupations were classified based on the HISCLASS scheme. White-collar (HISCLASS 1-5), farmer (HISCLASS 8), skilled/semi-skilled (HISCLASS 6-7,9) and unskilled (HISCLASS 10-12). In the last row, I report a counterfactual occupational distribution of Italians in the US had they been exposed to the transition matrix of Italians in Argentina.
Sources: Data are from the samples linking passenger lists to the census as described in the main text.

Table 4: First-generation immigrants in Argentina and the US, linked data

(a) Home ownership

	(1)	(2)	(3)	(4)	(5)	(6)
Italian in Argentina	0.0564***	0.0487***	0.0345***	0.0315***	0.0323***	0.0389***
	(0.00455)	(0.00526)	(0.00750)	(0.00752)	(0.0105)	(0.00803)
Years since arrival	No	Yes	Yes	Yes	Yes	Yes
Port of origin	No	No	Yes	Yes	Yes	No
Literacy	No	No	No	Yes	Yes	Yes
Occupation	No	No	No	Yes	Yes	Yes
Surname	No	No	No	No	Yes	Yes
Observations	30922	30922	30922	30922	30922	30922
Mean of dep. var.	0.145	0.145	0.145	0.145	0.145	0.145
R^2	0.0219	0.0280	0.0339	0.0401	0.438	0.435

(b) Unskilled occupation

	(1)	(2)	(3)	(4)	(5)	(6)
Italian in Argentina	-0.252***	-0.237***	-0.222***	-0.216***	-0.214***	-0.223***
	(0.00561)	(0.00647)	(0.00921)	(0.00902)	(0.0127)	(0.00967)
Years since arrival	No	Yes	Yes	Yes	Yes	Yes
Port of origin	No	No	Yes	Yes	Yes	No
Literacy	No	No	No	Yes	Yes	Yes
Occupation	No	No	No	Yes	Yes	Yes
Surname	No	No	No	No	Yes	Yes
Observations	30922	30922	30922	30922	30922	30922
Mean of dep. var.	0.476	0.476	0.476	0.476	0.476	0.476
R^2	0.0762	0.0813	0.0916	0.139	0.492	0.489

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. This table compares Italians in Argentina and the US with respect to the likelihood of owning their home and the likelihood of being employed in an unskilled occupation. The mean of the dependent variable is computed among Italians in the United States. Column 1 just includes age fixed effects. Columns 2 to 6 include additional fixed effects as indicated by the table.
Sources: Data are from the samples linking passenger lists to the census as described in the main text.

Table 5: Second-generation Italians in Argentina and the US

	Home ownership				Unskilled occupation			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Italian in Argentina	0.0443***	0.106***	0.148***	0.112***	-0.100***	-0.145***	-0.0732***	-0.142***
	(0.0145)	(0.0153)	(0.0179)	(0.0153)	(0.0120)	(0.0132)	(0.0146)	(0.0130)
Including natives	No	Yes	No	Yes	No	Yes	No	Yes
Including other immigrants	No	No	Yes	Yes	No	No	Yes	Yes
Observations	7151	341420	115740	450009	7151	341420	115740	450009
Mean of dep. var.	0.325	0.402	0.405	0.404	0.187	0.233	0.198	0.224

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. This table compares children of Italian migrants in Argentina to children of Italians in the US. In columns 1 and 5, the sample is restricted to children of Italian migrants. In columns 2 and 6, I also include the children of the native born in Argentina and the US. I columns 3 and 7, I compare children of Italians to the children of other immigrants in Argentina and the US. In columns 5 and 8, I compare children of Italians to the rest of the working-age population.
Sources: Data for Argentina are from Pérez (2017) and data for the US are from Ruggles et al. (1997).

Table 6: The role of language similarity

(a) Likelihood of entering a manual occupation, Italians in Argentina and the US

	Manual job			
	(1)	(2)	(3)	(4)
Italian in Argentina	-0.0856***	-0.0773***	0.0482***	-0.0562***
	(0.00440)	(0.00384)	(0.00574)	(0.00394)
Including natives	No	Yes	No	Yes
Including other immigrants	No	No	Yes	Yes
Observations	23769	1712842	304032	1993105
Mean of dep. var.	0.911	0.911	0.911	0.911

(b) Likelihood of entering a manual occupation and English ability among immigrants in the US

	Manual job	
	(1)	(2)
Speaks English	-0.0934***	-0.0760***
	(0.00530)	(0.00210)
Including other migrants	No	Yes
Country of birth FE	No	Yes
Observations	13089	230692
Mean of dep. var.	0.901	0.874

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. Panel (a) uses the data from Argentina and the US to estimate the likelihood that Italians would be employed in manual occupations. In column 1, the sample is restricted to first-generation Italian migrants. In column 2, I also include native-born individuals in Argentina and the US. In column 3, I compare Italians to other immigrant groups in Argentina and the US. In column 4, I compare Italians to the rest of the working-age population. Panel (b) uses the US data to estimate the likelihood that an individual would be employed in a manual occupation as a function of his ability to speak English.
Sources: Argentine data are from Somoza (1967) and US data are from Ruggles et al. (1997).

Table 7: Argentine Surname Frequency (ASI) and likelihood of moving to Argentina

	(1)	(2)	(3)	(4)	(5)
ASI	0.492***	0.478***	0.460***	0.464***	0.410***
	(0.00129)	(0.00128)	(0.00148)	(0.00143)	(0.00367)
ASI X female			0.0704***		
			(0.00284)		
ASI X child				0.0700***	
				(0.00306)	
ASI X unskilled/farm					0.0327***
					(0.00407)
Individual controls	No	Yes	Yes	Yes	Yes
Observations	1308843	1308843	1308843	1308843	778318
Mean of dep. var.	0.396	0.396	0.396	0.396	0.385
R^2	0.302	0.318	0.319	0.319	0.337

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. The dependent variable is an indicator that takes a value of one if the observation belongs to the Argentine data. The Argentine Surname Index (ASI) measures the relative frequency of an individual's surname in the Argentine and US data based on the surnames of previous Italian arrivals to Argentina and the US. The regressions with individual controls include indicators for age, gender, and occupational category upon arrival. The sample is restricted to arrivals from 1882 to 1900. Columns 1 to 4 include the full sample of immigrant arrivals to Argentina and the US. In column 5, the sample is restricted to working age (18 to 60 years old) males.
Source: Passenger lists data as described in the main text.

Data Appendix (Not for Publication)

Coverage of passenger lists data

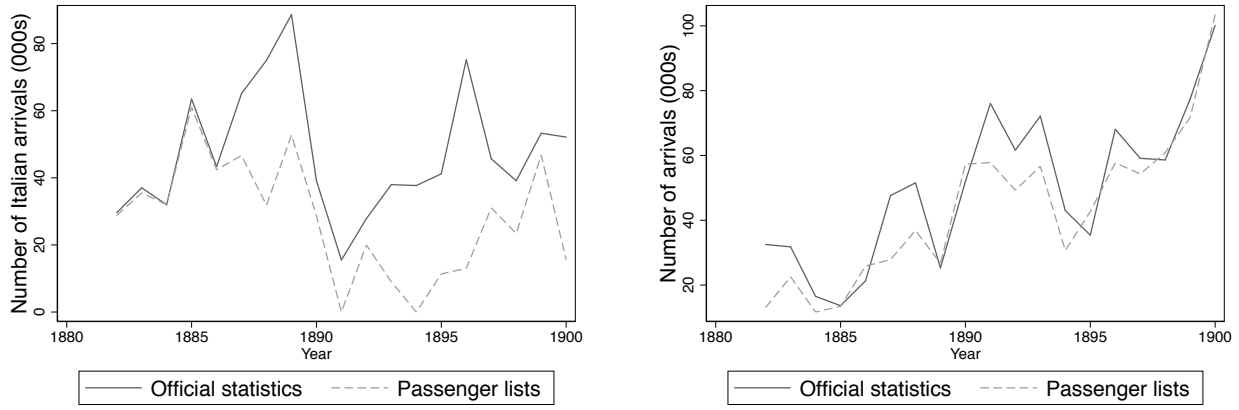
Figure A1 compares the number of yearly arrivals as computed with the passenger lists data and according to Ferenczi (1929). In the US data, the number of Italian arrivals in the passenger lists closely tracks the figures in the official statistics. In the 1882-1900 period, there were 940,000 Italian arrivals according to Ferenczi (1929), and there are 820,000 records in the passenger lists data in this period.

The Argentine passenger lists data are less complete than the US data. There were 900,000 Italian arrivals to Argentina from 1882 to 1900 according to Ferenczi (1929), but there are 530,000 records in the passenger lists data. There are a number of reasons for this incomplete coverage. First, only migrants arriving through the port of Buenos Aires are included in the data (about 75% of arrivals). Second, the digitization effort prioritized those lists that were in the worst state of preservation. Third, some of the original lists were impossible to digitize due to their state of preservation. For instance, there are no data corresponding to the years 1891 and 1894 (a total of 87,000 records). Similarly, when collapsing the data at the year-month of entry, there are no observations for about 30% of the months.

One concern is the extent to which the digitized data for Argentina is representative of Italian arrivals in this period. Figure A2 compares the fraction of males among Italians in the passenger lists and the fraction of males among all immigrants (including non-Italians) according to Ferenczi (1929). The fraction in the passenger lists data tracks closely that in the official statistics. Similarly, Figure A3 shows that the age structure by arrival decade (1881-1890 and 1891-1900) of Italians in the passenger lists data and the age structure of all migrants according to Ferenczi (1929) are also close to each other. Finally, note that the cross-sectional results (which do not use the passenger lists) yield similar results as those using the linked data.

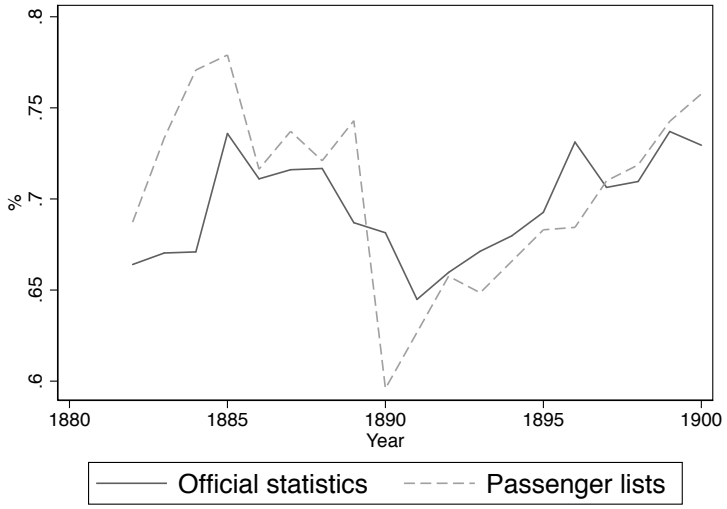
Figure A1: Coverage of passenger lists data, Argentina and the US

(a) Argentina **(b)** US



Notes: This figure shows the annual number of Italian arrivals according to the passenger lists data and the overall gender ratio (including non-Italians) according to official immigration statistics for Argentina and the US based on the data in Ferenczi (1929).

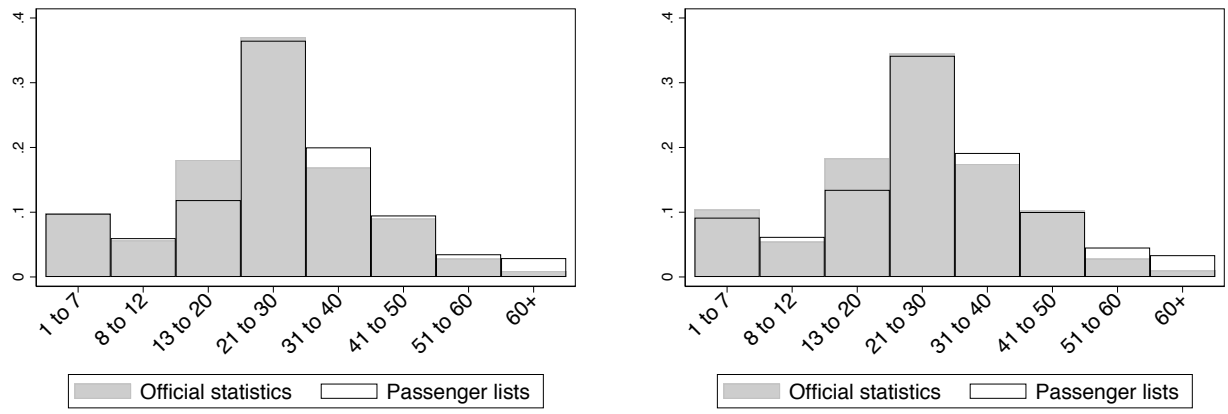
Figure A2: Fraction of males among Italian arrivals to Argentina



Notes: This figure shows the gender ratio of Italian arrivals according to the passenger lists data and according to official immigration statistics for Argentina based on the data in Ferenczi (1929).

Figure A3: Age structure of Italians in Argentine data

(a) 1881-1890 **(b)** 1891-1900



Notes: This figure shows the age structure of iItalian arrivals according to the passenger lists data and according to official immigration statistics for Argentina based on the data in Ferenczi (1929).

Table A1: Comparing the linked sample to the passenger lists, Argentine data

Variable	Cross-section (1)	Panel (2)	Difference (3)
Age	31.071	31.799	0.728***
			(0.084)
Literacy	0.725	0.715	-0.010***
			(0.004)
Occupation			
White-collar	0.025	0.029	0.004***
			(0.001)
Farmer	0.391	0.405	0.015***
			(0.004)
Skilled/semi-skilled	0.101	0.104	0.003
			(0.003)
Unskilled	0.443	0.424	-0.019***
			(0.004)

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. This table compares individuals in the passenger lists of Italian arrivals to Argentina to individuals in the linked data. Columns 1 and 2 report the average value of each variable in each of the datasets, whereas column 3 reports the difference between the cross section and the linked data.
Sources: Passenger lists data as described in the main text.

Table A2: Comparing the linked sample to the passenger lists, US data

Variable	Cross-section (1)	Panel (2)	Difference (3)
Age	31.266	29.919	-1.347***
			(0.067)
Illiterate	0.193	0.191	-0.001
			(0.003)
Literate	0.165	0.228	0.063***
			(0.003)
No literacy data	0.642	0.580	-0.062***
			(0.004)
Occupation			
White-collar	0.028	0.031	0.003**
			(0.001)
Farmer	0.212	0.217	0.005
			(0.003)
Skilled/semi-skilled	0.142	0.176	0.034***
			(0.003)
Unskilled	0.598	0.558	-0.040***
			(0.004)

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. This table compares individuals in the passenger lists of Italian arrivals to the US to individuals in the linked data. Columns 1 and 2 report the average value of each variable in each of the datasets, whereas column 3 reports the difference between the cross section and the linked data.
Sources: Passenger lists data as described in the main text.

Table A3: Comparing the linked sample to the census, Argentine data

Variable	Cross-section (1)	Panel (2)	Difference (3)
Literacy	0.702	0.741	0.039***
			(0.007)
Home ownership	0.202	0.218	0.016***
			(0.006)
Occupation			
White-collar	0.208	0.176	-0.032***
			(0.006)
Farmer	0.245	0.197	-0.049***
			(0.006)
Skilled/semi-skilled	0.282	0.343	0.062***
			(0.007)
Unskilled	0.207	0.210	0.003
			(0.006)

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. This table compares individuals in the 1895 census cross section of Argentina to individuals in the linked passenger lists to census data. Columns 1 and 2 report the average value of each variable in each of the datasets, whereas column 3 reports the difference between the cross section and the linked data.
Sources: Cross-sectional data are from Somoza (1967). Linked sample as described in the main text.

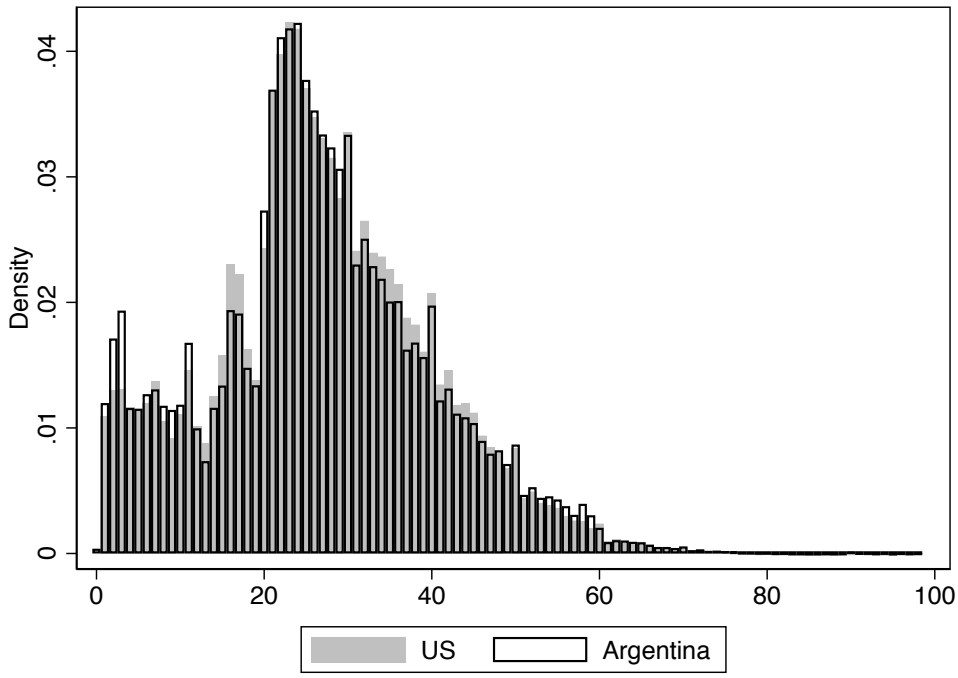
Table A4: Comparing the linked sample to the census, US data

Variable	Cross-section (1)	Panel (2)	Difference (3)
Literacy	0.593	0.625	0.032***
			(0.006)
Home ownership	0.144	0.145	0.002
			(0.004)
Occupation			
White-collar	0.114	0.117	0.003
			(0.004)
Farmer	0.153	0.156	0.004
			(0.004)
Skilled/semi-skilled	0.170	0.181	0.012**
			(0.005)
Unskilled	0.564	0.545	-0.019***
			(0.006)

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. This table compares individuals in the 1900 census cross section of the US to individuals in the linked passenger lists to census data. Columns 1 and 2 report the average value of each variable in each of the datasets, whereas column 3 reports the difference between the cross section and the linked data.
Sources: Cross-sectional data are from Ruggles et al. (1997). Linked sample as described in the main text.

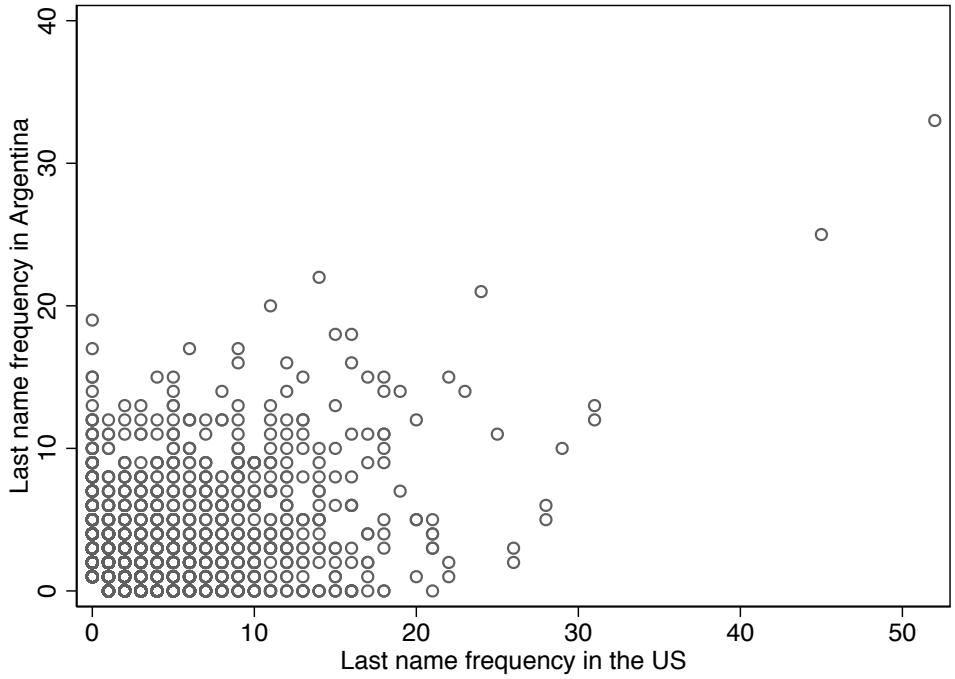
Additional Results (Not for Publication)

Figure B1: Age structure, Italian arrivals to Argentina and the US (1882-1900)



Notes: This figure plots the age distribution among Italian arrivals to Argentina and the US in the 1882-1900 period.
Source: Passenger lists data as described in the main text.

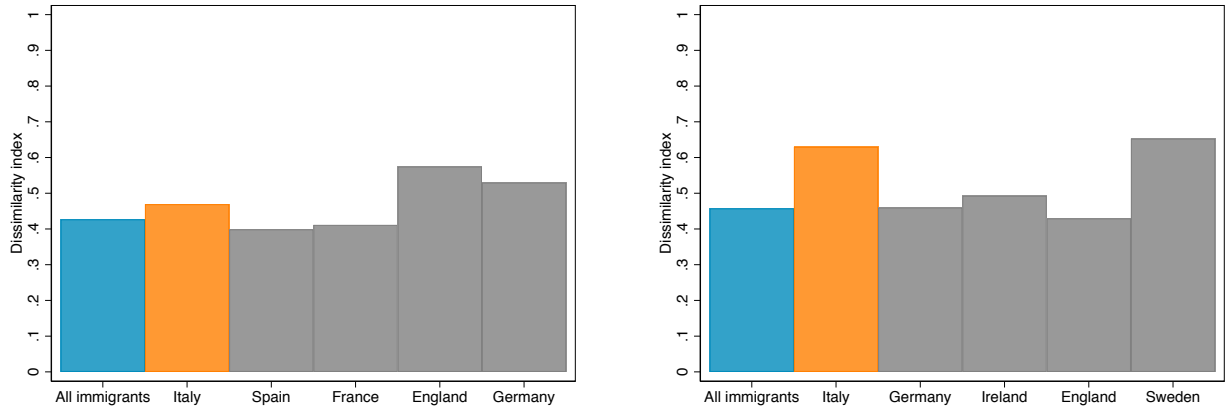
Figure B2: Overlap between Italian surnames in Argentina and the US



Notes: This figure shows the overlap between surnames of Italians arriving to Argentina and Italians arriving to the US. Each dot corresponds to a different surname.
Source: Data are from the samples linking passenger lists to the census as described in the main text.

Figure B3: Residential concentration of Italians in Argentina and the US: Dissimilarity Indexes

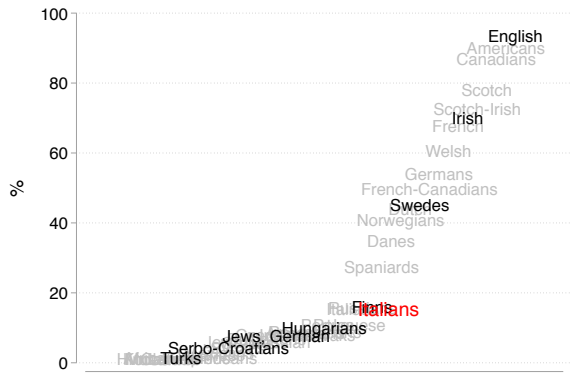
(a) Argentina **(b)** US



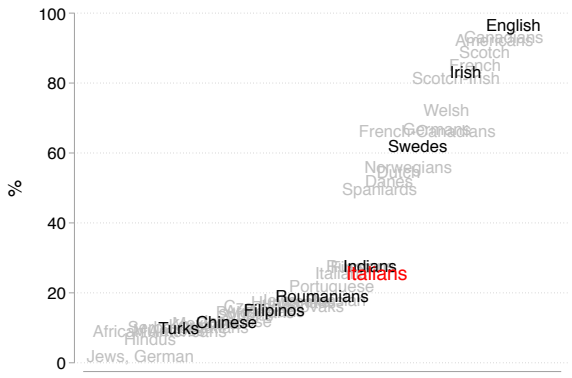
Notes: This figure shows a dissimilarity index of the residential distribution of different immigrant groups across counties (departments) in Argentina and the US. The first bar in each graph corresponds to the dissimilarity index for immigrants as a whole, while the remaining bars compute country-specific dissimilarity indices. A value of one corresponds to complete residential segregation, whereas a value of zero corresponds to complete residential integration.

Figure B4: Prejudice against Italian Immigrants: % of respondents who would:

(a) Marry someone from this group

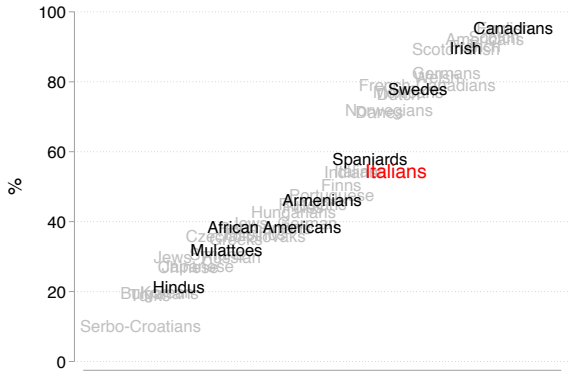


(b) Be friends with someone from this group

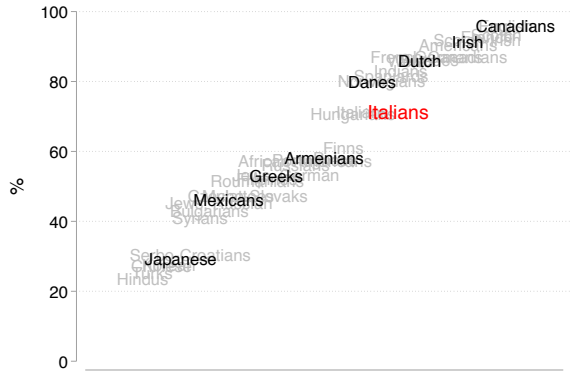


(c) Be neighbors with someone from this group

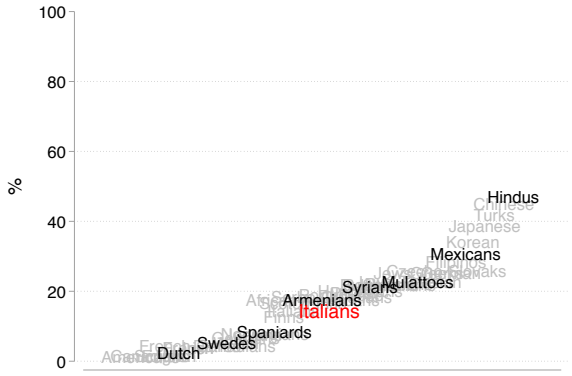
(d) Be in the same occupation



(e) Allow as citizen of my own country

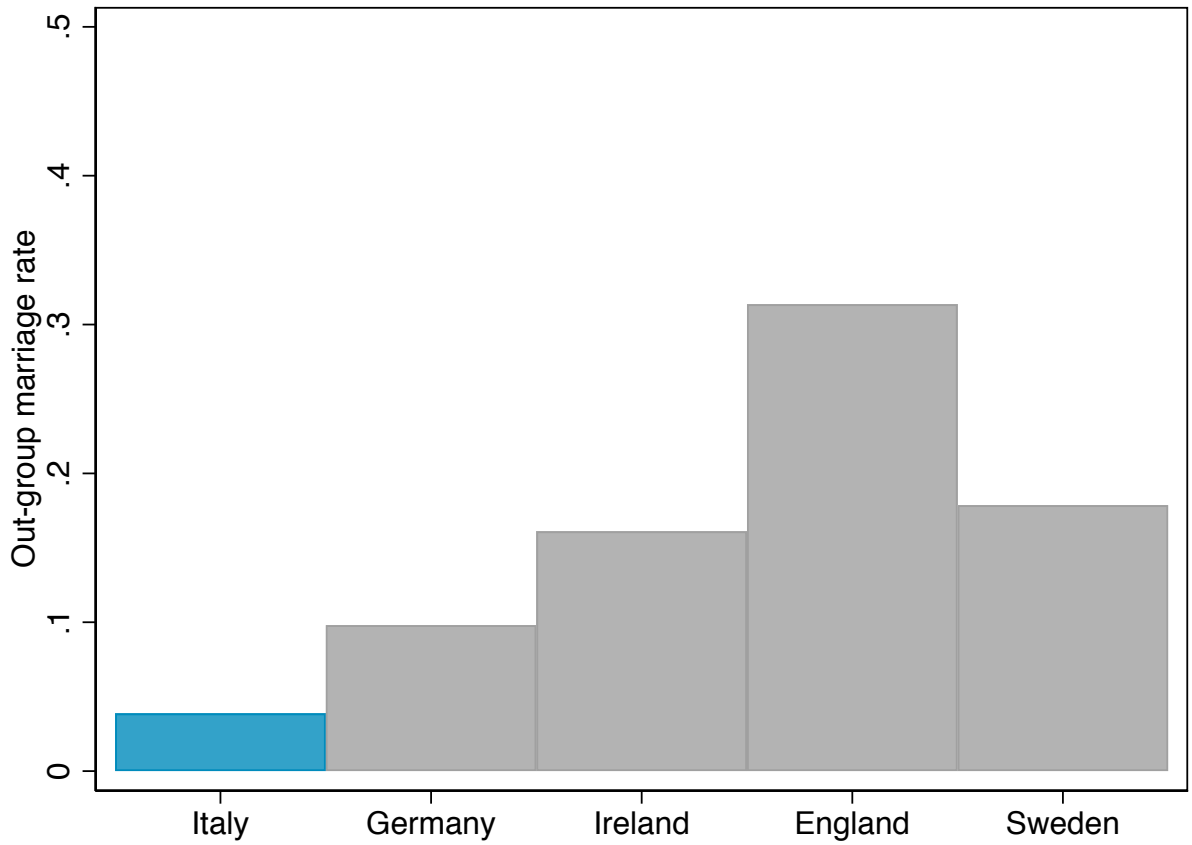


(f) Only allow as visitors to my own country



Notes: Each of the figures shows the percentage of individuals who reported being willing to accept someone from a given group as: (a) marriage partners, (b) friends, (c) neighbors, (d) workers in their same occupation, (e) citizens of their own country, and (f) only as visitors to their own country.
Source: Data are from a sample of 1725 American college students compiled by Bogardus (1928).

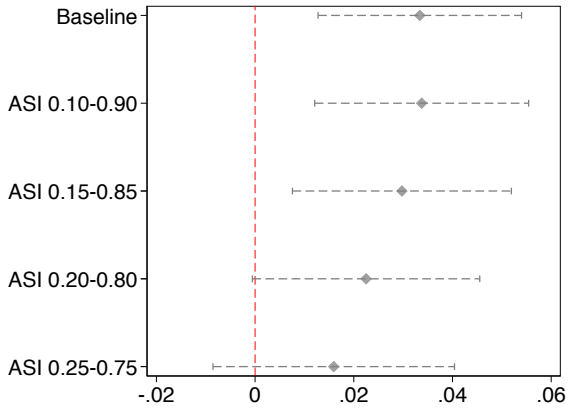
Figure B5: Intermarriage rates in the US, 1900



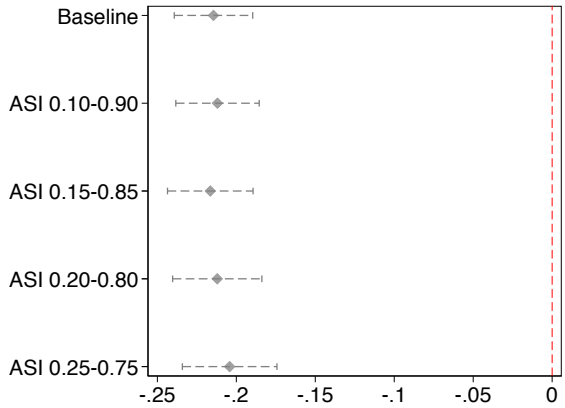
Notes: This figure shows the proportion of married immigrant men from each country of origin who are married outside of their group in 1900 US. Out-group marriage is defined as marrying someone born outside one own's country of origin and with both parents also born outside of it.
Source: Ruggles et al. (1997).

Figure B6: Surname fixed effects results using only "neutral" surnames

(a) Home ownership



(b) Unskilled occupation



Notes: This figure shows the results after progressively excluding individuals with "distinctive" surnames from the sample (that is, surnames with values of the ASI either close to zero or close to one. The specification corresponds to the one with the largest set of controls in Table 4 (including surname fixed effects).
Sources: Data are from the samples linking passenger lists to the census as described in the main text.

Table B1: First-generation immigrants in Argentina and the US, cross-sectional data, by urban/rural areas

(a) Urban areas

	Home ownership				Unskilled occupation			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Italian in Argentina	0.0789***	0.169***	0.166***	0.182***	-0.295***	-0.288***	-0.162***	-0.248***
	(0.00663)	(0.0112)	(0.0108)	(0.0100)	(0.00894)	(0.00968)	(0.0113)	(0.00917)
Including natives	No	Yes	No	Yes	No	Yes	No	Yes
Including other immigrants	No	No	Yes	Yes	No	No	Yes	Yes
Observations	13095	315764	150550	453219	13095	315764	150550	453219
Mean of dep. var.	0.117	0.117	0.117	0.117	0.468	0.468	0.468	0.468

(b) Rural areas

	Home ownership				Unskilled occupation			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Italian in Argentina	-0.0329***	0.256***	0.310***	0.271***	-0.276***	-0.270***	-0.207***	-0.260***
	(0.00991)	(0.0132)	(0.0150)	(0.0130)	(0.0120)	(0.0118)	(0.0139)	(0.0116)
Including natives	No	Yes	No	Yes	No	Yes	No	Yes
Including other immigrants	No	No	Yes	Yes	No	No	Yes	Yes
Observations	6604	542171	93559	629126	6604	542171	93559	629126
Mean of dep. var.	0.206	0.206	0.206	0.206	0.541	0.541	0.541	0.541

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. This table reports differences between Italians in Argentina and the US with respect to rates of home ownership and the likelihood of holding an unskilled occupation. The mean of the dependent variable is computed among Italians in the United States. In columns 1 and 5, the sample is restricted to first-generation Italian migrants. In columns 2 and 6, I also include native-born individuals in Argentina and the US. I columns 3 and 7, I compare Italians to other immigrant groups in Argentina and the US. In columns 4 and 8, I compare Italians to the rest of the working-age population. In panel (a), I restrict the sample to urban areas in Argentina and the US. In panel (b), I restrict the sample to rural areas.
Sources: Argentine data are from Somoza (1967) and US data are from Ruggles et al. (1997) and the 1900 full count census available though the NBER server.

Table B2: First-generation immigrants in Argentina and the US, cross-sectional data of Buenos Aires and New York city

	Home ownership				Unskilled occupation			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Italian in Argentina	0.0867***	0.0635***	0.0712***	0.0862***	-0.257***	-0.193***	-0.152***	-0.184***
	(0.00751)	(0.0153)	(0.0109)	(0.0114)	(0.0125)	(0.0148)	(0.0152)	(0.0128)
Including natives	No	Yes	No	Yes	No	Yes	No	Yes
Including other immigrants	No	No	Yes	Yes	No	No	Yes	Yes
Observations	5480	31425	30906	56851	5480	31425	30906	56851
Mean of dep. var.	0.0458	0.0458	0.0458	0.0458	0.408	0.408	0.408	0.408

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. This table reports differences between Italians in Argentina and the US with respect to rates of home ownership and the likelihood of holding an unskilled occupation. The mean of the dependent variable is computed among Italians in the United States. In columns 1 and 5, the sample is restricted to first-generation Italian migrants. In columns 2 and 6, I also include native-born individuals in Argentina and the US. I columns 3 and 7, I compare Italians to other immigrant groups in Argentina and the US. In columns 4 and 8, I compare Italians to the rest of the working-age population. The sample is restricted to the cities of Buenos Aires (Argentina) and New York (US).
Sources: Argentine data are from Somoza (1967) and US data are from Ruggles et al. (1997) and the 1900 full count census available though the NBER server.

Table B3: First-generation immigrants in Argentina and the US, excluding departures from non-Italian ports

	Owner		Unskilled	
	(1)	(2)	(3)	(4)
Italian in Argentina	0.0323***	0.0313**	-0.214***	-0.228***
	(0.0105)	(0.0124)	(0.0127)	(0.0149)
Years since arrival	Yes	Yes	Yes	Yes
Port of origin	Yes	Yes	Yes	Yes
Literacy	Yes	Yes	Yes	Yes
Occupation	Yes	Yes	Yes	Yes
Surname	Yes	Yes	Yes	Yes
Italian Ports	No	Yes	No	Yes
Observations	30922	23360	30922	23360
Mean of dep. var.	0.145	0.127	0.476	0.512
R^2	0.438	0.471	0.492	0.532

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. This table compares Italians in Argentina and the US with respect to the likelihood of owning their home and the likelihood of being employed in an unskilled occupation. The sample in columns 2 and 4 is restricted to Italians departing from Italian ports.

Table B4: First-generation northern and southern Italians in Argentina and the US

	Owner		Unskilled	
	(1)	(2)	(3)	(4)
Italian in Argentina	0.0319***		-0.225***	
	(0.00842)		(0.0101)	
Southern Italian X Argentina		0.0436***		-0.252***
		(0.0131)		(0.0157)
Northern Italian X Argentina		0.0625***		-0.264***
		(0.00660)		(0.00791)
Northern Italian X US		0.0352***		-0.0529***
		(0.00887)		(0.0106)
Years since arrival	Yes	Yes	Yes	Yes
Port of origin	Yes	Yes	Yes	Yes
Literacy	Yes	Yes	Yes	Yes
Occupation	Yes	Yes	Yes	Yes
Observations	23360	23360	23360	23360
Mean of dep. var.	0.127	0.127	0.512	0.520
R^2	0.0389	0.0388	0.153	0.153

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. This table compares Italians in Argentina and the US with respect to the likelihood of owning their home and the likelihood of being employed in an unskilled occupation. In columns 2 and 4, I split Italians into four mutually exclusive groups based on whether they departed from a Northern or a Southern port, and whether they migrated to Argentina or the US. The omitted category in columns 2 and 4 are southern Italians in the US.

Table B5: First-generation immigrants in Argentina and the US, excluding farmers and unskilled workers upon arrival

	Owner		Unskilled	
	(1)	(2)	(3)	(4)
Italian in Argentina	0.0315***	0.0254	-0.216***	-0.176***
	(0.00752)	(0.0172)	(0.00902)	(0.0198)
Years since arrival	Yes	Yes	Yes	Yes
Port of origin	Yes	Yes	Yes	Yes
Literacy	Yes	Yes	Yes	Yes
Occupation	Yes	Yes	Yes	Yes
Observations	30922	5444	30922	5444
Mean of dep. var.	0.145	0.150	0.476	0.357
R^2	0.0401	0.0534	0.139	0.137

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. This table compares Italians in Argentina and the US with respect to the likelihood of owning their home and the likelihood of being employed in an unskilled occupation. The sample in columns 2 and 4 excludes individuals who declared farming or an unskilled occupation upon arrival.

Table B6: First-generation immigrants in Argentina and the US, surname × port of origin fixed effects

	Owner		Unskilled	
	(1)	(2)	(3)	(4)
Italian in Argentina	0.0323***	0.0255*	-0.214***	-0.219***
	(0.0105)	(0.0136)	(0.0127)	(0.0161)
Years since arrival	Yes	Yes	Yes	Yes
Literacy	Yes	Yes	Yes	Yes
Occupation	Yes	Yes	Yes	Yes
Surname	Yes	Yes	Yes	Yes
Surname X Port	No	Yes	No	Yes
Observations	30922	30922	30922	30922
Mean of dep. var.	0.145	0.145	0.476	0.476
R^2	0.438	0.622	0.492	0.671

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. This table compares Italians in Argentina and the US with respect to the likelihood of owning their home and the likelihood of being employed in an unskilled occupation. In columns 2 and 4, I include port of origin × surname fixed effects.

Table B7: First-generation immigrants in Argentina and the US, first and last name fixed effects

	Owner		Unskilled	
	(1)	(2)	(3)	(4)
Italian in Argentina	0.0323***	0.0378***	-0.214***	-0.215***
	(0.0105)	(0.0121)	(0.0127)	(0.0145)
Years since arrival	Yes	Yes	Yes	Yes
Literacy	Yes	Yes	Yes	Yes
Occupation	Yes	Yes	Yes	Yes
Port	Yes	Yes	Yes	Yes
Surname	Yes	Yes	Yes	Yes
First name	No	Yes	No	Yes
Observations	30922	30922	30922	30922
Mean of dep. var.	0.145	0.145	0.476	0.476
R^2	0.438	0.472	0.492	0.526

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. This table compares Italians in Argentina and the US with respect to the likelihood of owning their home and the likelihood of being employed in an unskilled occupation. In columns 2 and 4, I include first name in addition to surname fixed effects.

Table B8: First-generation immigrants in Argentina and the US, Altonji et al. (2005) bounds

	Owner		Unskilled	
	(1)	(2)	(3)	(4)
Italian in Argentina	0.0349***	0.0320***	-0.222***	-0.216***
	(0.00748)	(0.00750)	(0.00920)	(0.00900)
Controls	No	Yes	No	Yes
Observations	31025	31025	31025	31025
Mean of dep. var.	0.145	0.145	0.475	0.475
$\frac{\beta_C}{\beta_{NC}-\beta_C}$		11.11		37.56

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. This table compares Italians in Argentina and the US with respect to the likelihood of owning their home and the likelihood of being employed in an unskilled occupation. Columns 1 and 3 control for age fixed effects, years since migration and port of origin fixed effects. In columns 2 and 4, I control for the occupational category declared upon arrival and for literacy.

Table B9: First-generation immigrants in Argentina and the US, excluding short-term arrivals

	Owner	Unskilled
	(1)	(2)
Italian in Argentina	0.0302**	-0.221***
	(0.0151)	(0.0166)
Years since arrival	Yes	Yes
Port of origin	Yes	Yes
Literacy	Yes	Yes
Occupation	Yes	Yes
Surname	Yes	Yes
Observations	21788	21788
Mean of dep. var.	0.174	0.443
R^2	0.486	0.538

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. This table compares Italians in Argentina and the US with respect to the likelihood of owning their home and the likelihood of being employed in an unskilled occupation. Sample is restricted to Italians who had spent five year or more at the destination by the time of the census.
Sources: Data are from the samples linking passenger lists to the census as described in the main text.

Table B10: Differences by US region of destination

	Home ownership		Unskilled occupation	
	(1)	(2)	(3)	(4)
US Northeast X Italian	-0.0792***	-0.245***	0.290***	0.316***
	(0.00585)	(0.00888)	(0.00764)	(0.00784)
US Midwest X Italian	0.0483***	-0.228***	0.278***	0.283***
	(0.00989)	(0.0140)	(0.0129)	(0.0123)
US South X Italian	-0.0490***	-0.239***	0.226***	0.197***
	(0.0125)	(0.0174)	(0.0163)	(0.0154)
US West X Italian	0.0903***	-0.139***	0.250***	0.258***
	(0.0107)	(0.0152)	(0.0140)	(0.0135)
Including natives	No	Yes	No	Yes
Observations	19699	857935	19699	857935
Mean of dep. var.	0.195	0.195	0.209	0.209

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. This table repeats the analysis in Table 1 but splitting the US indicator into four variables: US Northeast, US Midwest, US South and US West. The sample in columns 1 and 3 is restricted to Italian immigrants. In columns 2 and 4, the sample also includes natives. Omitted category are Italians in Argentina.
Sources: Argentine data are from Somoza (1967) and US data are from Ruggles et al. (1997).

Table B11: The role of relative human capital

	Home ownership		Unskilled occupation	
	(1)	(2)	(3)	(4)
Italian in Argentina	0.259***	0.174***	-0.274***	-0.177***
	(0.00841)	(0.00841)	(0.00740)	(0.00738)
Including natives	Yes	Yes	Yes	Yes
Literacy	No	Yes	No	Yes
Observations	857935	857935	857935	857935
Mean of dep. var.	0.144	0.144	0.490	0.490

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. This table repeats the analysis in Table 1 but adding literacy (as measured in the destination country) as a control variable.
Sources: Argentine data are from Somoza (1967) and US data are from Ruggles et al. (1997).

Table B12: Argentine Surname Index and likelihood of moving to Argentina, standardized surnames

	(1)	(2)	(3)	(4)	(5)
ASI (NYSIIS)	0.471***	0.456***	0.437***	0.440***	0.393***
	(0.00147)	(0.00146)	(0.00169)	(0.00163)	(0.00420)
ASI X female			0.0737***		
			(0.00327)		
ASI X child				0.0776***	
				(0.00353)	
ASI X unskilled/farm					0.0232***
					(0.00466)
Individual controls	No	Yes	Yes	Yes	Yes
Observations	1308843	1308843	1308843	1308843	778318
Mean of dep. var.	0.396	0.396	0.396	0.396	0.385
R^2	0.281	0.298	0.298	0.298	0.319

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. The dependent variable is an indicator that takes a value of one if the observations belongs to the Argentine data. The Argentina Surname Index (ASI) measures the relative frequency of an individual's surname in the Argentine and US data based on the *standardized* surnames of previous Italian arrivals to Argentina and the US. Columns 1 to 4 include the full sample of immigrant arrivals to Argentina and the US. In column 5, the sample is restricted to working age (18 to 60 years old) males.
Source: Passenger lists data as described in the main text.