Southern (American) Hospitality:

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

Santiago Pérez*

January 24, 2021

Abstract

Argentina and the US were the principal destinations for Italian immigrants during the Age of Mass Migration. I assemble data following Italians from passenger lists to censuses in Argentina and the US, enabling me to compare the economic outcomes of migrants with similar pre-migration characteristics but who moved to different countries. Italians assimilated faster in Argentina, and this advantage was unlikely to be due to selection. A higher human capital relative to natives and the Italian-Spanish similarity largely explain Italians' advantage in Argentina. These findings highlight the importance of the fit between migrants' characteristics and those of the receiving country.

Keywords: immigration; assimilation; Argentina; United States

JEL Codes: J61; J62; N30

*Corresponding author: Santiago Pérez, Department of Economics, University of California, Davis, 1 Shields Ave, Davis, CA 95616, and NBER (e-mail: seperez@ucdavis.edu). I thank Enrique Pérez and María Fabiana Vaccaro for their help collecting the data. I thank the editor, Hans-Joachim Voth, for his most valuable guidance, and two anonymous referees for their very useful feedback. I have also benefitted from feedback from Ran Abramitzky, Leah Boustan, Federico Curci, Hisham Foad, 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; ASSA Meetings; University of Milan; Universidad de San Andrés and Universidad de los Andes. A longer version of this paper with additional results circulated as NBER Working Paper 26127.

The extent to which immigrants successfully integrate into the societies that host them varies widely across receiving countries. For instance, the average immigrant in the US has lower earnings upon arrival than in Australia or Canada but experiences faster earnings growth as she spends time in the destination (Antecol *et al.*, 2006). These contrasting experiences may be due to differences in the *selection* of migrants who move to different destinations, or to differences in how similar migrants fit socially and economically into different countries.

To understand the source of these differences, one would like to observe the integration experience of immigrants with similar pre-migration characteristics but who moved to different destinations. A challenge in doing so, however, is that the conventional sources used to study immigrant assimilation rarely include information on the background characteristics of immigrants beyond their country of origin and year of arrival. Moreover, some of the characteristics included in these sources (such as schooling) are potentially affected by the migration experience itself.

This paper studies the selection and assimilation of Italian migrants to Argentina and the US during the Age of Mass Migration (1850–1913). Nearly seven million Italians moved to Argentina or the US during this period: By the 1920s, Italians accounted for 40% of the immigrant population in Argentina and for 10% of the immigrant population in the US.[1] Much like some immigrants today, Italians were perceived by sectors of US society as being unlikely to assimilate.[2] Indeed, concerns about "new immigrants" from Southern and Eastern Europe (among which Italians were the largest group) were a main driver of the imposition of country-of-origin quotas in the 1920s (Goldin, 1994). In contrast, by most historical accounts, Italians had a rapid and successful integration in Argentina (Klein, 1983; Baily, 1983, 2004).

Beyond its intrinsic importance as a major migration episode in world history, a key advantage of this setting is the possibility to observe detailed information on migrants' backgrounds and their outcomes at different host countries. Specifically, I have assembled individual-level data following Italian migrants from passenger lists to population censuses. In these data, I observe the arrival years, ports of origin and pre-migration occupations 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 conventional sources– the degree to which heterogeneity in immigrant outcomes across different

[1] Immigrants constituted 25% of the Argentine and 14% of the US populations by the early 1920s.
[2] See for instance LaGumina (1999).

host countries can be explained by differences in selection.

I start by using the passenger-list data to compare the characteristics of Italians who moved to Argentina or the US. The main difference between both groups was the higher proportion of immigrants departing from northern Italian ports among arrivals to Argentina. In contrast, demographic characteristics and pre-migration occupations had only a weak association with destination choices: 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. This pattern suggests that Italians' destination choices were mostly driven by the strength of the networks connecting them to each destination (rather than by characteristics related to their own human capital).

My main analysis compares the economic outcomes of Italians in Argentina and the US. Italians in late-19th-century Argentina were 5.7 percentage points more likely to own their home and 25 percentage points less likely to hold an unskilled job than Italians in late-19th-century US. When comparing Italians who departed from the same port, these gaps are reduced by 30% and 10%, respectively. This reduction reflects the better average outcomes of Italians who departed from northern ports (who were overrepresented in the flow to Argentina). However, other individual-level characteristics such as pre-migration occupation or literacy explain very little of the gap in outcomes. The stability of the estimates after the inclusion of these characteristics suggests a limited scope for remaining selection bias (Altonji *et al.*, 2005; Oster, 2017).

What explains the differences in outcomes? I find evidence supporting two main explanations. First, despite Italian immigrants in Argentina and the US had similar levels of human capital, those in Argentina benefited from a higher level of human capital *relative to the native-born*: When focusing on *relative* gaps between Italians and the native-born in each destination country, simply controlling for literacy substantially reduces the advantage of Italians in Argentina. Second, the smaller linguistic distance between Italian and Spanish enabled those Italians who moved to Argentina to enter a broader range of occupations. Indeed, Italians in Argentina had a significantly smaller advantage when compared to Italians in the US who reported speaking English. Taken together, these two explanations account for 60% of the Argentina-US gap in Italians' home-ownership rates, and for 40% of the gap in the likelihood of holding an unskilled job. Overall, these results highlight the importance of the *match* between immigrant characteristics and those of the

destination country.

This paper contributes to our understanding of immigrant assimilation during the Age of Mass Migration. While several papers have focused on specific receiving countries, no quantitative studies have looked at the *comparative* performance of immigrants across destinations.[3] Italian migration to Argentina and the US is an especially relevant case, as it deals with the main sending country and the two largest destinations in this period.

More broadly, this paper contributes to the literature comparing the economic performance of migrants across different receiving countries.[4] Specifically, by accounting for a much richer set of pre-migration characteristics than in existing studies, I am able to shed light on the relative roles of selection and host-country conditions in explaining cross-country differences in immigrant outcomes. Finally, this paper also relates to the literature on "place effects" (see, for instance, Chetty *et al.* (2016) and Chetty and Hendren (2018)). Unlike most of such literature, my analysis focuses on the consequences of exposure to different countries (rather than to different neighbourhoods), and on exposure during adulthood rather than childhood.

1 Data

I use two sources of individual-level data: passenger lists of immigrant arrivals and population censuses. The passenger lists include the name, age, sex, literacy, occupation, arrival date and port of origin of Italian migrants to Argentina or the US.[5] The Argentine lists span 1882 to 1920 and include 1,020,000 Italians who arrived through the port of Buenos Aires.[6] These data were collected by Argentina's National Direction of Immigration and have been digitalised by *Centro de Estudios Migratorios Latinoamericanos* and *Fondazione Rodolfo Agnelli*. The US passenger lists are based on US Customs Service records and are available from the National Archives. They include 845,000 passengers who arrived between 1855 and 1900 and who identified their origin as Italy (or an Italian region).

I used the information on names, country of birth and age to link males in these lists to

[3] See for instance Abramitzky *et al.* (2014) and Ferrie (1994, 1997) on the US, and Pérez (2017) on Argentina.

[4] See Borjas (1991), Duleep and Regets (1992), Antecol *et al.* (2006), Algan *et al.* (2010), Bauer *et al.* (2011) and Kaushal *et al.* (2016).

[5] Other than port of origin, the Argentine data lack any systematic information on regional origins.

[6] I discuss the coverage of these data in Online Appendix section A.1.

population censuses of Argentina or the US. The 1895 census is the only Argentine census for which such linking is possible, since the previous census was conducted before the passenger lists started being systematically collected (and there are no surviving records of the next census). To assemble the US sample, I linked Italian passengers arriving after 1882 to the 1900 census. The baseline linked samples include about 13,000 individuals in Argentina and 17,000 in the US. I provide details on the linking procedure and sensitivity checks in Online Appendix section A.2.

2 Understanding Destination Choices

Argentina and the US had each received nearly one million Italians by 1900. Why did some Italians choose to go to Argentina whereas others choose the US? Although the passenger-list data do not enable me to directly measure migrants' motives, they enable me to observe *who* went to each destination. This, in turn, can shed light on the likely reasons why migrants selected one destination or the other.[7]

Although by 1900 Argentina was among the top five countries based on per-capita income (Taylor, 2018), unskilled wages were 25% lower than in the US (Williamson, 1995). Hence, a first hypothesis is that Italians deciding between Argentina and the US might have faced a trade-off between higher short-term wages and better prospects for long-term upward mobility. As a result, Argentina would have attracted immigrants that placed a higher value on long-term integration.

An implication of this hypothesis is that Argentina should have received a greater proportion of families aiming to establish themselves permanently in the country (as opposed to migrants traveling on their own). To test whether this was the case, I investigate whether Argentina had received more women and children (a usual proxy for the prevalence of family migration) by 1900. The first two columns in Panel (a) of Table 1 show the correlation between these characteristics and the likelihood of moving to Argentina. Italians moving to Argentina were slightly more likely to be female and below the age of 16 (Column 1). This pattern, however, reverses once I compare Italians who departed from the same port (Column 2). Overall, there is little evidence of a meaningfully higher prevalence of family migration in the flow to Argentina; such prevalence is similar to that in

[7] Unfortunately, direct information on migrants' motives for selecting destinations is lacking. As discussed in Baily (2004), 'we have such evidence for almost none of the migrants who went to Buenos Aires and New York at the turn of the past century.'

the US in the raw data, and is in fact *smaller* after accounting for differences in migrants' broad regional origins.

A second hypothesis is that Italians sorted on the basis of relative returns to skill (Grogger and Hanson, 2011). If returns to skill were higher in Argentina than in the US, then Argentina would have attracted more skilled Italians (which in turn would explain Italians' better outcomes there). The main implication of this hypothesis is that Italians who moved to Argentina should have been more skilled than those who went to the US.

Panel (b) of Table 1 focuses on the pre-migration occupations of males aged 18 to 60 upon arrival. Italians who went to Argentina were overrepresented among white-collar workers (although the proportion of white-collar workers was fairly small in both flows, at around 3%; see Table B1 in the Online Appendix), and underrepresented in the skilled/semi-skilled category. The most salient difference is that Italians who migrated to Argentina were more likely to report farming and less likely to report unskilled jobs compared to their counterparts in the US (Columns 3 and 4). However, the distinction between farm and general labourers is unlikely to have been very informative in this context: As late as 1911, 60% of the Italian workforce was still in agriculture.[8] Overall, occupations are limited in their ability to predict destination choices: Simultaneously adding indicators for all occupational categories explains only 5% of the variation in destination choices (Column 6 of Panel (b)).

I next compare the literacy of Italians moving to Argentina or the US. I measure literacy in the census cross-sections because the US passenger lists are missing information on this variable for 60% of the observations.[9] The data show a 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 (Somoza, 1967; Ruggles *et al.*, 1997). This difference is much smaller than that between southern and northern Italians who remained in Italy: By 1901, only 30% of southerners were literate, compared to 65% of northerners (Klein, 1983).

[8] 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 '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.' When pooling unskilled and farm workers into a single category, there is little correlation between membership in this category and the likelihood of moving to Argentina (see Column 5 in Table 1 and Table B1 in the Online Appendix).

[9] A concern with this approach is that the census might exaggerate initial differences if Italians were more likely to accumulate skills in one of the countries.

A third hypothesis emphasises the early settlement patterns of "pioneer" migrants as the main driver of destination choices.[10] According to this hypothesis, most individuals would have been unable to migrate to a given destination unless they were part of a migration chain linking them to it.[11] Hence, the early decisions of pioneer immigrants generated path dependence in destination choices, whereby immigrants with similar regional origins tended to migrate to similar destinations.[12] An implication of this hypothesis is that we should observe clustering of immigrants across destinations depending on the areas of Italy they hailed from.

Consistent with this hypothesis, Column 3 in Panel (a) shows a strong positive correlation between departing from a northern Italian port and the likelihood of moving to Argentina: 40% of the variation in destination choices can be accounted for on the basis of this single variable. A challenge, however, in establishing whether this pattern was driven by immigrant networks is that the passenger lists lack direct information on the strength of the "connections" that migrants might have had at each potential destination. To deal with this limitation, I constructed a surname-based proxy for the size of migrants' networks. Specifically, for immigrant i with surname s arriving in year t, I compute the "Argentina Surname Index" (ASI) as:

$$ASI_{ist} = \frac{\frac{\#Italians\,with\,surname\ s\,in\,Argentina}{\#Italians\,in\,Argentina}}{\frac{\#Italians\,with\,surname\ s\,in\,Argentina}{\#Italians\,in\,Argentina} + \frac{\#Italians\,with\,surname\ s\,in\,US}{\#Italians\,in\,US}} \tag{1}$$

where the number of Italians in each receiving country is based on arrivals up to year $t-1$. This measure takes a value of one if the immigrant has a surname that, up to year $t-1$, can be found only among arrivals to Argentina, and a value of zero if it can be found only among US arrivals.[13]

[10] A number of scholars highlight the importance of this mechanism. See, for instance, Gould (1980), Moretti (1999), and Spitzer and Zimran (2019).

[11] This was because, as discussed in Moretti (1999), 'many of [them] were illiterate and had relatively little knowledge of the world beyond the village.'

[12] This argument is summarised in Gould (1980): 'An interesting implication is that a strong locational pattern in the settlements of migrants from a particular part of Europe may reflect nothing more than the momentum arising from an original choice which was itself determined by some quite minor, adventitious circumstances, or indeed quite random.' Hence, 'purely incidental connections were then reinforced by the process of feedback to build enduring locational patterns.'

[13] This measure is based on Fryer Jr and Levitt (2004). Because ASI_{ist} is undefined for immigrants whose surnames appear for the first time at time t, I assign a value of 0.5 to those surnames. The stock of Italian surnames in 1882 Argentina is based on the 1869 Argentine census. The US stock is based on the surnames of immigrants arriving from 1855 to 1881 in the passenger lists.

Figure 1 shows the empirical distribution of ASI among immigrants whose surnames had appeared at least once in the data by their arrival year. The figure shows a bimodal distribution with peaks at zero and one: The modal immigrant was "connected" to only one of the two destinations. Moreover, migrants tended to move to the destination in which they had the largest network: Going from a distinctively "American" to a distinctively "Argentine" surname is associated with a 63-percentage points increase in the probability of moving to Argentina (Column 4 of Table 1). This index predicts destination choices as well when considering migrants who departed from the *same* port, suggesting that it does not simply capture differences in the propensity to migrate to Argentina across broad regions of Italy (Column 5).[14]

If migrants' choice between moving to Argentina or the US was driven primarily by the presence of networks, we should also observe that such networks predict immigrants' final locations *within* receiving countries.[15] To test this prediction, I use the fact that the US passenger lists include information on migrants' intended destination within the country, usually at the city level.[16] I then use this information to estimate a model of migrants' intended destinations within the US, including a surname-based measure of networks (analogous to the one discussed above) as an independent variable.

Assume that migrant i arriving in year t obtains the following utility from choosing destination d:

$$U_{itd} = \alpha_t + \alpha_d + \beta Network_{itd} + \gamma_d X_i + \epsilon_{itd} \tag{2}$$

where α_t are arrival-year fixed effects, α_d are destination fixed effects, and X_i are individual-level characteristics. The variable of interest is $Network_{itd}$, which captures the fraction of arrivals with the same surname as immigrant i (and who arrived prior to year t) who intended to move to

[14] As discussed in Baily (2004) and Devoto (2006), Genoese shipping companies were important in directing early northern Italian migration to Argentina. Devoto (2006) describes how early Italian migration to Argentina occurred in waves, starting with Genoese seaman, who were then followed by artisans (particularly those associated with the maritime industry), to finally include farmers and unskilled laborers from the interior regions.

[15] The historical literature strongly suggests this possibility. For instance, MacDonald and MacDonald (1964) describes how 'In Middletown (Connecticut), the greater part of the large Italian population came from the Sicilian town of Melilli in Syracuse province, and concentrated in one neighborhood.' Baily (2004) highlights the presence of 'village-based clusters' in Buenos Aires.

[16] Unfortunately, the Argentine data lack such information. One alternative would be to use the linked data, but the drawback of such approach is that I would be able to observe destinations only among those immigrants whom I can match to an observation in the census.

destination d. If we assume that ϵ_{itd} follows an Extreme Value Type I distribution, then this model corresponds to the standard conditional logit due to McFadden (1974).

Table B2 in the Online Appendix shows that $Network_{itd}$ predicts destination choices *within* the US as well. Indeed –and like the results above– the magnitude of this association is similar after including individual-level characteristics such as gender, age and pre-migration occupation. Overall, the evidence in this section suggests that Italians' destination choices were largely driven by the strength of the networks connecting them to each destination.

3 The Economic Outcomes of Italians in Argentina and the US

By most historical accounts, Italians had a more successful integration in Argentina than in the US (Baily, 2004; Klein, 1983). How much of this advantage can be explained by differences in the pre-migration characteristics of the Italians who went to each country? To answer this question, I estimate:

$$y_{ic} = \alpha + \beta Argentina_{ic} + \gamma X_{ic} + \epsilon_{ic} \tag{3}$$

where y_{ic} is an outcome of immigrant i in country c, and X_{ic} are pre-migration characteristics. The coefficient of interest is β, which measures the advantage/disadvantage of Italians in Argentina relative to those in the US. Throughout this section, I restrict the sample to Italian males aged 18 to 60 and focus on two outcomes that can be consistently measured in the Argentine and US censuses: the likelihood of homeownership and the likelihood of holding an unskilled occupation.[17]

The key empirical challenge in this context is to measure the consequences of migrants' destination choices as accurately as possible in the presence of selection. Specifically, the concern is that ϵ_{ic} may be correlated with destination choices, leading to biased estimates of β. For instance, as discussed above, Italians sorted across destinations on the basis of their regional origins (which could be correlated with outcomes at the destination if migrants from certain regions were on average

[17] I coded occupations using the Historical International Standard Classification of Occupations (HISCO), which I then mapped into occupational categories using HISCLASS (Leeuwen *et al.*, 2002). Unskilled jobs are those in HISCLASS categories 10 to 12. The 1895 Argentine census asked "Do you own real estate property?" The 1900 US census asked "Is the person's home owned or rented?" Unfortunately, unlike the 1900 US census, the 1895 Argentine census did not include a question on naturalisation. Thus, I am not able to investigate differences in the extent of political integration. Luconi (2015) argues that Italians had an easier political adjustment and earlier involvement with politics in Argentina than in the US.

more skilled).

The richness of my data enables me to address this selection to a much larger extent than in existing comparative studies of immigrant assimilation. First, to account for differences in regional origins, I estimate models that include port-of-origin fixed effects. Alternatively, I include *surname* fixed effects, which in the Italian context enables me to absorb a finer level of geography (Spitzer and Zimran, 2018). Second, I can directly absorb migrants' occupations prior to moving. Finally, when focusing on the likelihood of holding an unskilled occupation, I can estimate equation 3 in first differences, enabling me to absorb unobservable fixed effects at the individual level.

Table 2 shows the results of estimating equation 3. Panel (a) focuses on the likelihood of homeownership, whereas Panel (b) focuses on the likelihood of holding an unskilled job. In the first column of each panel, X_{ic} includes only age fixed effects: Italians in Argentina were 5.7 percentage points more likely to own their home and 25 percentage points less likely to hold an unskilled job than similarly aged Italians in the US.

In the second column, I add years-since-arrival indicators. Adding these indicators reduces the advantage of Italians in Argentina with respect to both outcomes (reflecting a longer average stay in Argentina). Figure 2 shows the relationship between time spent in each of the countries and outcomes at the destination. In both countries, a longer stay was associated with a higher likelihood of homeownership and a lower likelihood of holding an unskilled job. However, the figure suggests little convergence between Italians in Argentina and those in the US: Home-ownership rates start from similarly low levels in both countries but grow at a higher rate in Argentina, whereas the likelihood of holding an unskilled job starts from a higher level in the US and remains similarly higher.

In the third column, I test if the different mix between northern and southern Italians in Argentina and the US could explain the differences in outcomes. To do so, I expand X_{ic} to include port-of-origin fixed effects. Adding this variable leads to a 10% decline in the *Argentina* coefficient when focusing on the likelihood of holding an unskilled job, but to a larger (30%) decline when focusing on homeownership.

In the fourth column, I consider the possibility that Italians in Argentina had better outcomes because of higher pre-migration skills. Specifically, 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 little impact on the estimated coefficients. This pattern is not surprising, given the relative balancing in these characteristics documented in Section 2.

Column 5 shows that the results are similar when I include surname fixed effects, thus comparing immigrants with the same surname but who moved to different destinations.[18] There are two reasons why surnames provide useful information in this context. First, as discussed above, Italian surnames are informative of regional origins. Second, surnames are informative of family linkages among individuals with rare surnames (Güell *et al.*, 2014).

I further exploit these features of surnames in Figure B1, where I re-estimate equation 3 while progressively excluding individuals with common surnames from the sample. The idea behind this exercise is that, by restricting the sample to those with rare surnames, I am increasingly likely to narrow the comparison to related individuals (as, for instance, in Abramitzky *et al.* (2012) and Collins and Wanamaker (2014)). If anything, the gaps between Italians in Argentina and the US become *wider* in these restricted samples.

In Column 6, I instead implement a coarsened exact matching approach (Iacus *et al.*, 2012).[19] To implement this approach, I first use pre-migration characteristics to select a "control" individual (among Italian migrants to the US) for each Italian migrant to Argentina. In the second step, I estimate equation 3 using the matched treatment-control sample. Using this approach, I continue to find an advantage for Italian migrants in Argentina with respect to both homeownership and the likelihood of holding an unskilled occupation.

Finally, because the passenger lists include information on occupations, when focusing on the likelihood of holding an unskilled job I can estimate equation 3 in first differences. Using this approach, which enables me to net out individual-level time invariant characteristics, I continue to find a lower likelihood of working in unskilled jobs among Italians in Argentina (Table B3 in the Online Appendix).

Although these results enable me to account for selection into destinations to a much greater

[18] Because of transcription errors in the census and passenger lists, I use a phonetically equivalent version of surnames based on NYSIIS (Taft, 1970). I preserve the last letter of the original surname because they are a strong predictor of regional origins among Italian surnames.

[19] Coarsened exact matching is a non-parametric procedure that is used to generate treatment and control groups that are balanced on the basis of baseline covariates. I provide a detailed description of this approach in Online Appendix section **??**.

degree than with cross-sectional data, I cannot fully rule out the possibility that the differences between Italians in Argentina and those in the US were driven by differences in unobservable characteristics of those who moved to each country. To assess the scope for such unobservables to explain the results, I implement the approaches proposed in Altonji *et al.* (2005) and Oster (2017). Specifically, I estimate models that include years-since-migration and port of origin fixed effects, and assess the sensitivity of the results to adding controls for immigrants' skills. Table B4 in the Online Appendix shows the degree of selection on unobservables –relative to selection on observables– necessary to overturn the results: The amount of selection on unobservables would have needed to be at least 4.5 times greater than the selection on observables to rationalise the difference in homeownership, and at least 21 times greater to rationalise the difference in the likelihood of holding an unskilled job. Overall, I conclude that it is unlikely that the large differences I observe could be plausibly explained by selection alone.

4 What Explains the Differences in Outcomes?

The previous section suggests a likely limited role for selection in explaining the differences in outcomes of Italians in Argentina and the US. Which factors can account for these differences?

A first potential explanation is that the observed gaps reflected broader differences between the Argentine and US economies rather than Italian-specific differences. For instance, Klein (1983) argues that the preponderance of small artisan shops in Argentine manufacturing offered more opportunities for skilled blue-collar jobs than the more industrialised US economy. Similarly, access to housing might have differed between Argentina and the US for reasons such as the relative cost of housing, access to mortgages, etc.

To test this hypothesis, I focus on *relative* differences between Italian immigrants and the native-born populations of Argentina and the US (rather than on *absolute* differences as in the previous section). To do so, I expand the sample to include the native-born and estimate:

$$y_{ic} = \alpha_0 + \beta_1 Italian_{ic} + \beta_2 Argentina + \beta_3 Italian_{ic} \times Argentina_{ic} + \gamma X_{ic} + \epsilon_{ic} \qquad (4)$$

The coefficient of interest is β_3, which measures the difference between the outcomes of Italians in

Argentina and the US relative to natives in each host country.

Table 3 shows that this mechanism cannot explain the results. The top row in this table focuses on *absolute differences* between Italians in Argentina and the US (i.e., β in equation 3), whereas the bottom row focuses on *relative differences* (i.e., β_3 in equation 4). The gap in homeownership is actually five times *wider* when considered relative to the native-born population (bottom versus top rows in Column 1), whereas the gap in the likelihood of holding an unskilled occupation is only slightly narrower (bottom versus top rows in Column 6).

Next, I consider three potential explanations: (1) differences in relative human capital between Italians and the native-born population, (2) the role of language, and (3) the role of location choices within the host countries.

Although Italians in Argentina and the US had similar levels of human capital upon arrival (as captured by their occupations and literacy), Italians in Argentina had higher levels of human capital *relative* to the native-born population: 89% of native-born males aged 18 to 60 were literate in 1900 US, compared to only 54% in 1895 Argentina. To quantify the role of differences in relative human capital, I investigate how the relative gaps in outcomes (i.e., β_3 in equation 4) change as I include literacy as a control variable. Although adding such variable makes little difference when explaining *absolute* differences between Italians in Argentina and the US (top row, Columns 1 versus 2 and Columns 6 versus 7), it does play an important role when focusing on *relative* differences: The gap in homeownership decreases by 33% (from 28 to 19 percentage points), whereas the gap in the likelihood of holding an unskilled occupation narrows by 45% (from 23 to 13 percentage points).

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 than in the US. A challenge in testing this hypothesis is that, although the US census includes information on English proficiency, the Argentine census does not include a comparable question. To address this limitation, I perform an exercise that enables me to place an upper bound on the likely importance of this channel. Namely, I assume that *all* Italians in Argentina were proficient in Spanish. Combining this assumption with direct information on English proficiency from the US census, I estimate equations 3 and 4 adding an indicator that takes a value of one if the person spoke the host-country language. Columns 3 and 8 in Table 3 suggest the likely importance of this channel: the absolute gap in homeownership is 40% narrower than in the baseline (from 5.7 to 3.3 percentage points), whereas the absolute gap

in the likelihood of holding an unskilled occupation decreases by 30% (from 25 to 19 percentage points).

I next consider the role of location choices within the host countries. Italians in the US tended to settle in the "older" regions of the country and predominantly in cities: By 1900, 72% of them 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 mobility, as upward mobility tended to be higher in "younger" and smaller places.

To test this hypothesis, I estimate equations 3 and 4 adding a variable that takes a value of one if the individual lived in an urban area of Argentina or the US. Columns 4 and 9 of Table 3 show that said variable plays no role in explaining either the absolute or relative gaps between Italians in Argentina and the US. Overall, this finding suggests that the choice of destination country had a stronger influence on outcomes than migrants' location choices *within* receiving countries.[20]

Finally, I consider the combined role of the four explanations considered so far–namely: selection, differences in relative human capital, language, and location choices. To do so, in Columns 5 and 10 I simultaneously include controls for: (1) pre-migration characteristics, (2) language ability, and (3) urban status.[21] Taken together, these factors account for 40-65% of the gap in homeownership, and for 30-50% of the gap in the likelihood of holding an unskilled occupation.

4.1 Possible Reasons for Remaining Gaps

Differences in regional origins, a higher human capital relative to the native-born, and the linguistic proximity of Italian and Spanish explain a substantial portion of the relative advantage of Italians in Argentina. A potential explanation for the remaining gaps is that Italians might have suffered more nativist prejudice in the US than in Argentina. Although my data do not enable me to conclusively prove this hypothesis, it is mostly consistent with the qualitative evidence. First, 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 and share a common Latin culture. Moreover, the elites who governed Argentina during this period had a very positive

[20] Relatedly, the lower likelihood of working in unskilled occupations is not driven solely by a higher propensity to work in farming. Indeed, Italians in Argentina were also more likely to work in white-collar and skilled blue-collar jobs (see Table B5 in the Online Appendix).

[21] When focusing on *relative* differences, I include indicator variables for the native-born to account for the fact that variables such as port of origin and pre-migration occupation are not defined for this group.

view of European immigration, which they considered a source of "civilisation." This positive view manifested in a legal framework that was friendly towards European immigration, from the enactment of the country's constitution in 1853 (which explicitly welcomed Europeans) to the maintenance of nearly open doors until the 1930s (Sánchez-Alonso, 2013).

In the US, in contrast, Italians and other "new immigrants" from Eastern and Southern Europe were the target of anti-immigrant sentiments by the turn of the 20th century. Indeed, the quota acts of 1921 and 1924 were written with the explicit goal of reducing the number of such migrants, among which Italians were the largest group (Goldin, 1994). Nativist prejudice against Italians was rooted in long-standing anti-Catholic sentiments in the US (Fouka *et al.*, 2018; Higham, 2002), which had already contributed to a backlash against Irish immigration in the 1850s (Collins and Zimran, 2018). Manifestations of this prejudice were likely consequential for the outcomes considered in this study. For instance, Hillier (2003) discusses how, even before the formal emergence of redlining in the 1930s, private lenders in the US avoided neighbourhoods with large concentrations of African Americans and "new immigrants." This practice may have negatively impacted Italian homeownership rates.

5 Conclusions

Seven million Italians moved to Argentina and 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 in host-country conditions. Using data linking Italian immigrants from passenger lists to population censuses, I showed that differences in the selection of migrants moving to each destination are unlikely to explain the differences in outcomes.

These results highlight the importance of the *match* between migrants' characteristics and those of the destination country. An implication of this finding is that receiving countries might benefit from an increased focus on potential fit (rather than on general human-capital characteristics) when screening immigrants. Finally, despite being a very large group (nearly 12% of the total population by the 1920s), Italians were able to successfully integrate into the economy of Argentina. This success suggests that the size of an immigrant group may not be *per se* an important driver of assimilation.

University of California, Davis and NBER

References

Abramitzky, R., Boustan, L.P. and Eriksson, K. (2012). "Europe's tired, poor, huddled masses: Self-selection and economic outcomes in the age of mass migration", *American Economic Review*, vol. 102(5), pp. 1832–1856.

Abramitzky, R., Boustan, L.P. and Eriksson, K. (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*, vol. 122(3).

Algan, Y., Dustmann, C., Glitz, A. and Manning, A. (2010). "The economic situation of first and second-generation immigrants in france, germany and the united kingdom", *The Economic Journal*, vol. 120(542), pp. F4–F30.

Altonji, J.G., Elder, T.E. and Taber, C.R. (2005). "Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools", *Journal of political economy*, vol. 113(1), pp. 151–184.

Antecol, H., Kuhn, P. and Trejo, S.J. (2006). "Assimilation via prices or quantities? sources of immigrant earnings growth in australia, canada, and the united states", *Journal of human Resources*, vol. 41(4), pp. 821–840.

Baily, S.L. (1983). "The adjustment of italian immigrants in buenos aires and new york, 1870-1914", *The American Historical Review*, pp. 281–305.

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

Bauer, T.K., Cobb-Clark, D.A., Hildebrand, V.A. and Sinning, M.G. (2011). "A comparative analysis of the nativity wealth gap", *Economic Inquiry*, vol. 49(4), pp. 989–1007.

Borjas, G.J. (1991). "Immigration policy, national origin, and immigrant skills: A comparison of canada and the united states", National Bureau of Economic Research.

Chetty, R. and Hendren, N. (2018). "The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects", *The Quarterly Journal of Economics*, vol. 133(3), pp. 1107–1162.

Chetty, R., Hendren, N. and Katz, L.F. (2016). "The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment", *The American Economic Review*, vol. 106(4), pp. 855–902.

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

Collins, W.J. and Wanamaker, M.H. (2014). "Selection and economic gains in the great migration of african americans: New evidence from linked census data", *American Economic Journal: Applied Economics*, vol. 6(1), pp. 220–252.

Collins, W.J. and Zimran, A. (2018). "The economic assimilation of irish famine migrants to the united states", National Bureau of Economic Research.

Devoto, F. (2006). *Historia de los italianos en la Argentina*, Editorial Biblos.

Duleep, H.O. and Regets, M.C. (1992). "Some evidence on the effect of admission criteria on immigrant assimilation", *Immigration, language and ethnic issues: Canada and the United States*, pp. 410–439.

Ferrie, J.P. (1994). "The wealth accumulation of antebellum european immigrants to the us, 1840–60", *The Journal of Economic History*, vol. 54(01), pp. 1–33.

Ferrie, J.P. (1997). "The entry into the us labor market of antebellum european immigrants, 1840–1860", *Explorations in Economic History*, vol. 34(3), pp. 295–330.

Fouka, V., Mazumder, S. and Tabellini, M. (2018). "From immigrants to americans: Race and assimilation during the great migration", *Harvard Business School BGIE Unit Working Paper*, (19-018).

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

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*pp. 223–258, University of Chicago Press.

Gould, J.D. (1980). "European international emigration: The role of diffusion and feedback", *Journal of European Economic History*, vol. 9(2), p. 267.

Grogger, J. and Hanson, G.H. (2011). "Income maximization and the selection and sorting of international migrants", *Journal of Development Economics*, vol. 95(1), pp. 42–57.

Güell, M., Rodríguez Mora, J.V. and Telmer, C.I. (2014). "The informational content of surnames, the evolution of intergenerational mobility, and assortative mating", *The Review of Economic Studies*, vol. 82(2), pp. 693–735.

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

Hillier, A.E. (2003). "Redlining and the home owners' loan corporation", *Journal of Urban History*, vol. 29(4), pp. 394–420.

Iacus, S.M., King, G. and Porro, G. (2012). "Causal inference without balance checking: Coarsened exact matching", *Political analysis*, vol. 20(1), pp. 1–24.

Kaushal, N., Lu, Y., Denier, N., Wang, J.S.H. and Trejo, S.J. (2016). "Immigrant employment and earnings growth in canada and the usa: evidence from longitudinal data", *Journal of population economics*, vol. 29(4), pp. 1249–1277.

Klein, H.S. (1983). "The integration of italian immigrants into the united states and argentina: A comparative analysis", *The American Historical Review*, pp. 306–329.

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

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

Luconi, S. (2015). "Italian immigrants' political adjustment in the americas: The case of the united states and argentina", *IdeAs. Idées d'Amériques*, (6).

MacDonald, J.S. and MacDonald, L.D. (1964). "Chain migration ethnic neighborhood formation and social networks", *The Milbank Memorial Fund Quarterly*, vol. 42(1), pp. 82–97.

McFadden, D. (1974). "The measurement of urban travel demand", *Journal of public economics*, vol. 3(4), pp. 303–328.

Moretti, E. (1999). "Social networks and migrations: Italy 1876–1913", *International Migration Review*, vol. 33(3), pp. 640–657.

Oster, E. (2017). "Unobservable selection and coefficient stability: Theory and evidence", *Journal of Business & Economic Statistics*, pp. 1–18.

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*, vol. 77(4), pp. 971–1006.

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

Sánchez-Alonso, B. (2013). "Making sense of immigration policy: A rgentina, 1870–1930", *The Economic History Review*, vol. 66(2), pp. 601–627.

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.

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

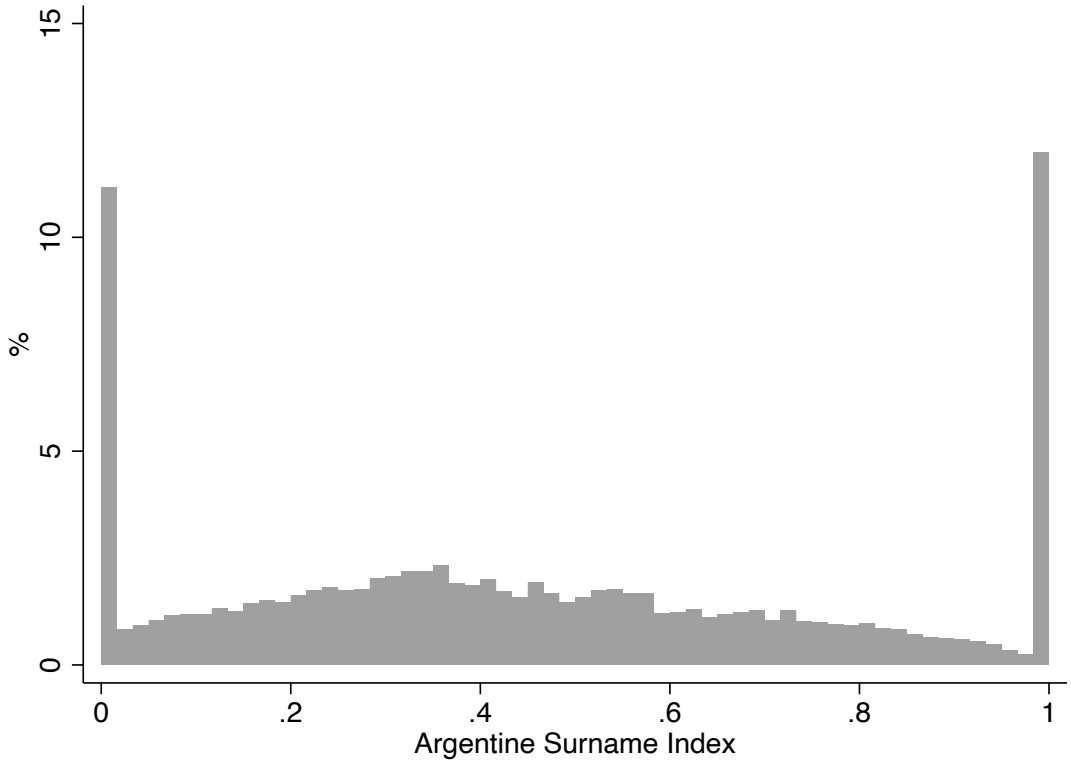
Spitzer, Y. and Zimran, A. (2019). "Like an ink blot on paper", .

Taft, R. (1970). "Name search techniques: New york state identification and intelligence system", *Albany: Special Rep*, vol. 1.

Taylor, A.M. (2018). "The argentina paradox: microexplanations and macropuzzles", *Latin American Economic Review*, vol. 27(1), p. 3.

Williamson, J.G. (1995). "The evolution of global labor markets since 1830: background evidence and hypotheses", *Explorations in Economic History*, vol. 32(2), pp. 141–196.

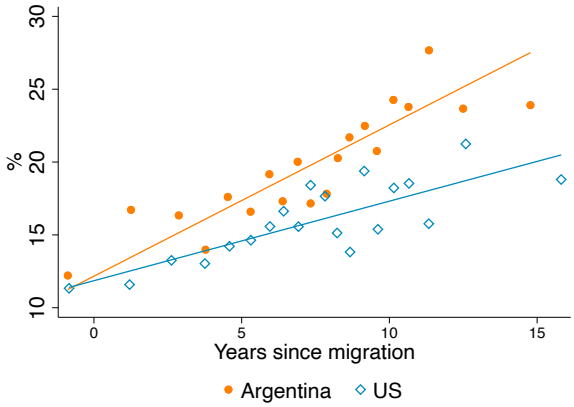
Figure 1: *Empirical Distribution of Argentine Surname Index, 1882-1900 Arrivals*



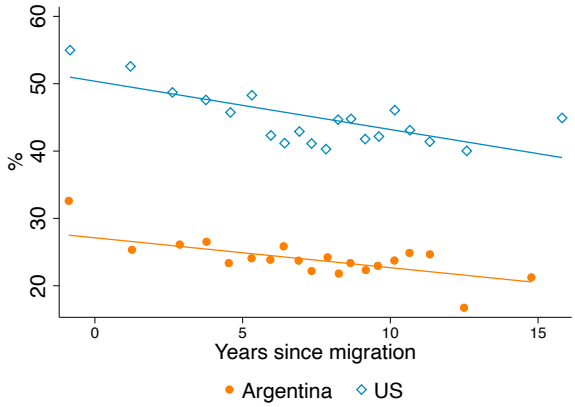
Notes: This figure shows the empirical distribution of the *Argentine Surname Index (ASI)* among arrivals to Argentina and the US from 1882 to 1900. A value of one indicates a surname that was only held by Italians who had previously moved to Argentina, whereas a value of zero indicates a surname that was only held by Italians who had previously moved to the US. The sample is restricted to migrants whose surnames had shown up at least once by their arrival year.
Source: Passenger lists data as described in the main text.

Figure 2: *Main Economic Outcomes, by Years Since Migration*

(a) Homeownership



(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.

Table 1: *Correlates of Likelihood of Moving to Argentina*

(a) Full Sample

	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.0332***	-0.00197***				-0.00250***
	(0.000974)	(0.000674)				(0.000653)
Age less than 16	0.00745***	-0.0123***				-0.0106***
	(0.00107)	(0.000737)				(0.000713)
Northern port			0.633***			
			(0.000668)			
ASI				0.632***	0.315***	0.315***
				(0.00139)	(0.00105)	(0.00105)
Port of Origin	No	Yes	No	No	Yes	Yes
Observations	1335334	1335334	1335334	1335334	1335334	1335334
R^2	0.000975	0.524	0.403	0.134	0.554	0.554

(b) Working-age Males

	(1)	(2)	(3)	(4)	(5)	(6)
White-collar	0.0552***					0.131***
	(0.00322)					(0.00318)
Farmer		0.233***				0.240***
		(0.00119)				(0.00124)
Skilled/Semi-skilled			-0.0783***			0.00935***
			(0.00164)			(0.00167)
Unskilled				-0.166***		
				(0.00110)		
Unskilled/Farmer					0.0539***	
					(0.00151)	
Observations	752385	752385	752385	752385	752385	752385
R^2	0.000390	0.0483	0.00302	0.0293	0.00170	0.0505

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 passenger lists data. The sample in panel (a) includes all Italian arrivals to Argentina or the US from 1882 to 1900. The sample in panel (b) is limited to males aged 18 to 60 upon arrival. *Northern port* is an indicator that takes a value of one if the migrant departed from the ports of Genoa or Trieste. 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.
Source: Passenger lists data as described in the main text.

Table 2: *The Economic Outcomes of Italians in Argentina and the US*

(a) Homeownership

	(1)	(2)	(3)	(4)	(5)	(6)
Italian in Argentina	0.0575***	0.0486***	0.0347***	0.0315***	0.0337***	0.0278***
	(0.00457)	(0.00525)	(0.00750)	(0.00752)	(0.0106)	(0.00696)
Years since arrival	No	Yes	Yes	Yes	Yes	No
Port of origin	No	No	Yes	Yes	Yes	No
Literacy	No	No	No	Yes	Yes	No
Occupation	No	No	No	Yes	Yes	No
Surname	No	No	No	No	Yes	No
Coarsened Exact Matching	No	No	No	No	No	Yes
Observations	30620	30620	30620	30620	30620	11273
Mean of dep. var.	0.143	0.143	0.143	0.143	0.143	0.143
R^2	0.0220	0.0279	0.0337	0.0400	0.439	0.00141

(b) Unskilled occupation

	(1)	(2)	(3)	(4)	(5)	(6)
Italian in Argentina	-0.252***	-0.237***	-0.222***	-0.216***	-0.214***	-0.168***
	(0.00563)	(0.00647)	(0.00922)	(0.00903)	(0.0127)	(0.00854)
Years since arrival	No	Yes	Yes	Yes	Yes	No
Port of origin	No	No	Yes	Yes	Yes	No
Literacy	No	No	No	Yes	Yes	No
Occupation	No	No	No	Yes	Yes	No
Surname	No	No	No	No	Yes	No
Coarsened Exact Matching	No	No	No	No	No	Yes
Observations	30620	30620	30620	30620	30620	11273
Mean of dep. var.	0.476	0.476	0.476	0.476	0.476	0.476
R^2	0.0768	0.0819	0.0922	0.140	0.494	0.0333

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 (panel (a)) and the likelihood of being employed in an unskilled occupation (panel (b)). The mean of the dependent variable is computed among Italians in the United States. Column 1 just includes age fixed effects. Columns 2 to 5 include additional fixed effects as indicated by the table. Column 6 implements a coarsened matching approach as described in the main text. The data in column 6 are weighted using coarsened exact matching weights.
Sources: Data are from the samples linking passenger lists to the census as described in the main text.

Table 3: *What Explains the Differences in Outcomes?*

	Home ownership					Unskilled Occupation				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Absolute Differences (N=30620)	0.057***	0.052***	0.033***	0.057***	0.020*	−0.252***	−0.234***	−0.186***	−0.253***	−0.180***
	(0.00457)	(0.00457)	(0.00505)	(0.00456)	(0.01081)	(0.00563)	(0.00551)	(0.00617)	(0.00562)	(0.01301)
Relative Differences (N=868856)	0.277***	0.193***	0.250***	0.292***	0.161***	−0.230***	−0.132***	−0.165***	−0.221***	−0.109***
	(0.00691)	(0.00693)	(0.00752)	(0.00677)	(0.01446)	(0.00610)	(0.00609)	(0.00664)	(0.00605)	(0.01289)
Literacy	No	Yes	No	No	Yes	No	Yes	No	No	Yes
Language	No	No	Yes	No	Yes	No	No	Yes	No	Yes
Urban/Rural	No	No	No	Yes	Yes	No	No	No	Yes	Yes
Selection	No	No	No	No	Yes	No	No	No	No	Yes

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 (columns 1 to 5) and the likelihood of being employed in an unskilled occupation (columns 6 to 10). In the top row ("Absolute Differences"), the sample is restricted to Italian males aged 18 to 60. In the bottom row ("Relative Differences"), the sample also include similarly-aged native born individuals in both Argentina and the US. Columns 1 and 6 include only age fixed effects. In columns 2 and 7, I add an indicator variable that takes a value of one if the individual is literate. In columns 3 and 8, I add an indicator variable that takes a value of one if the individual is proficient in the language of the host country (as described in the main text). In columns 4 and 9, I add an indicator that takes a value of one if the person resided in an urban area of Argentina or the US. In columns 5 and 10, I simultaneously include all the controls in columns 1 to 4 plus those in column 5 of Table 2 (which account for differences in pre-migration characteristics between Italians in Argentina and those in the US). In these columns, I include indicator variables for the native born to account for the fact that variables such as origin port and pre-migration occupation are not defined for this group.

Sources: Data on Italian migrants are from the samples linking passenger lists to the census as described in the main text. Data on natives are from the census cross sections of 1895 Argentina and 1900 US.

ONLINE APPENDIX:

Southern (American) Hospitality:

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

Santiago Pérez

January 24, 2021

A Data Appendix

A.1 Coverage of passenger lists data

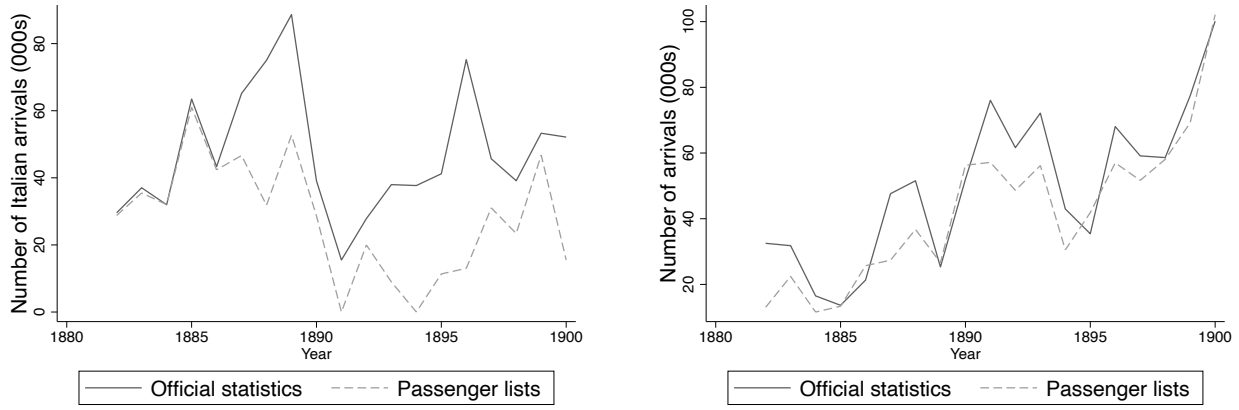
Figure A1 compares the number of yearly arrivals as computed with the passenger lists data and according to Ferenczi (1929). In the US data, the number of Italian arrivals in the passenger lists closely tracks the figures in the official statistics. In the 1882-1900 period, there were 940,000 Italian 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 according to de Inmigración (1925)). Second, the digitisation effort prioritised those lists that were in the worst state of preservation. Third, some of the original lists were impossible to digitise 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 digitised data for Argentina are 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 data) yield similar results as those using the linked data.

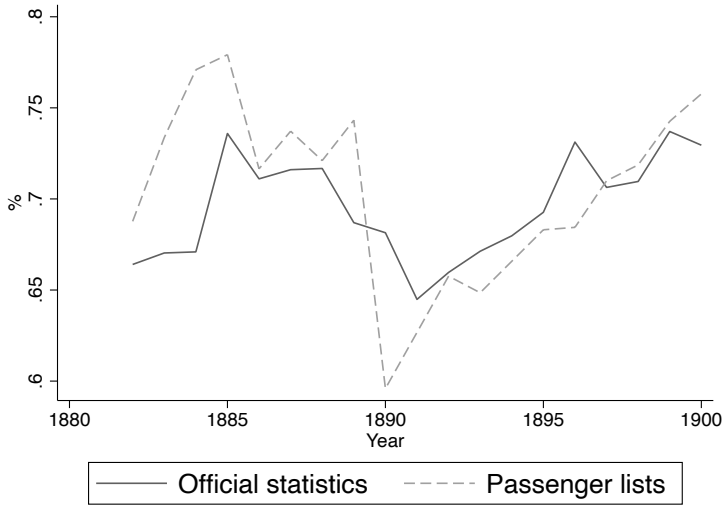
Figure A1: Coverage of passenger lists data, Argentina and the US

(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).

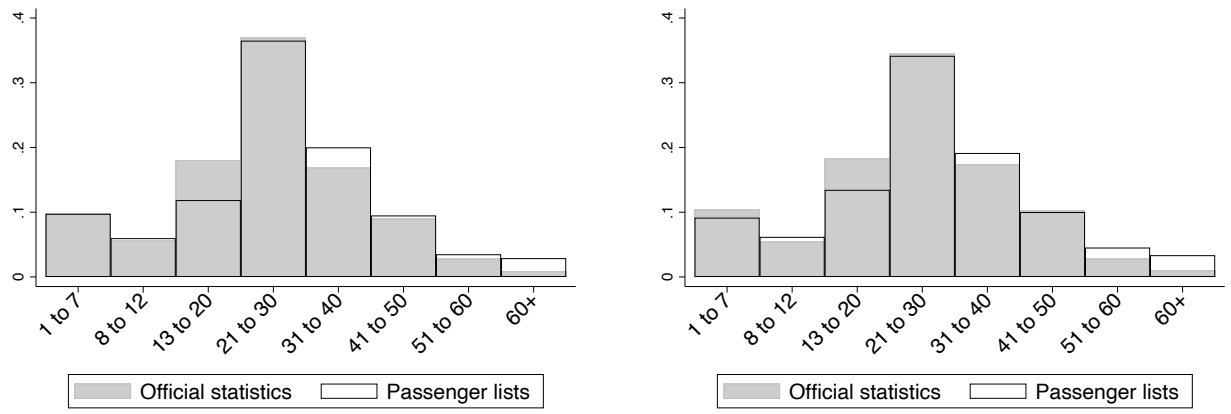
Figure A2: Fraction of males among Italian arrivals to Argentina



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).

A.2 Linking Algorithm and Data Digitisation

I used the information on names, country of birth and age to link males in the passenger lists to the 1895 Argentine census and the 1900 US census.[1] 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.[2]

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 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.

Both the US passenger lists and the US 1900 census are fully digitised, including the information

[1] The 1895 census is the only Argentine census for which such linking is possible. Argentine censuses in this period took place in 1869, 1895 and 1914. The 1869 is too early in time since the passenger list data are only available starting in 1882. Unfortunately, there are no surviving individual-level records of the 1914 census.

[2] 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.

on occupations and other economic outcomes. The Argentine passenger lists are also fully digitised, but only the *indexes* of the 1895 census are. Hence, after linking the data, I manually digitised the economic information in the 1895 Argentine census (using the original manuscripts available in the genealogy website familysearch.org).

A.3 Sensitivity of Results to Linking Algorithm

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 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).[3]

To further address this possibility, Figure A4 progressively excludes 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.

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.

[3]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.

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 likelihood 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 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, 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 job. 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 2.3 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. I address this concern in two main ways. First, note that the results using cross-sectional data (which do not rely upon linking) are consistent with the results that use the linked data (see Table A5). Second, in the last row of Figure A4 I reweight the data to account for selection into the linked sample based on observable characteristics upon arrival.[4] The results are similar to those in the baseline sample, suggesting

[4]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.

that selection into the linked samples (at least with respect to observable characteristics) is not driving the results.

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)
Literate	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	31.263	29.945	-1.319***
			(0.068)
No literacy data	0.642	0.572	-0.070***
			(0.004)
Illiterate	0.193	0.195	0.002
			(0.003)
Literate	0.165	0.233	0.068***
			(0.003)
Occupation			
White-collar	0.028	0.031	0.003**
			(0.001)
Farmer	0.212	0.218	0.006*
			(0.003)
Skilled/semi-skilled	0.142	0.177	0.035***
			(0.003)
Unskilled	0.598	0.557	-0.041***
			(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	0.664	0.693	0.029***
			(0.007)
Home ownership	0.195	0.218	0.023***
			(0.006)
Occupation			
White-collar	0.208	0.176	-0.032***
			(0.006)
Farmer	0.243	0.197	-0.046***
			(0.006)
Skilled/semi-skilled	0.287	0.343	0.056***
			(0.007)
Unskilled	0.209	0.210	0.001
			(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)

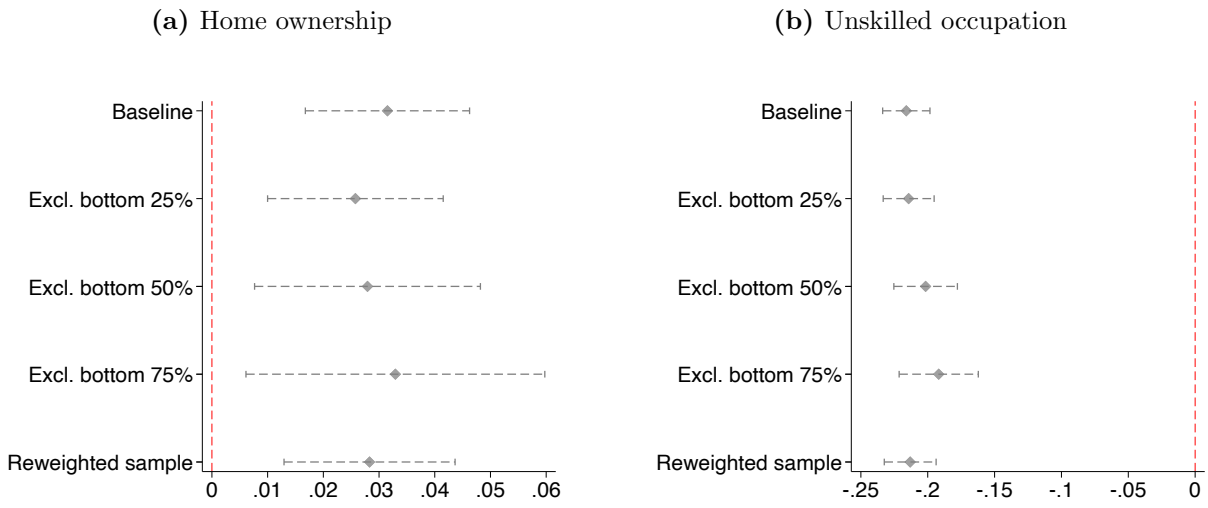
Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. This table compares individuals in the 1900 census cross section of the US to individuals in the linked passenger lists to census data. Columns 1 and 2 report the average value of each variable in each of the datasets, whereas column 3 reports the difference 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.

Table A5: 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).

Figure A4: Robustness to linking procedure



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.

B Additional Results

B.1 Coarsened Exact Matching

This subsection describes the coarsened exact matching approach (Iacus *et al.*, 2012) that is implemented at the end of Section 3. The goal of this approach is to identify a "control" individual (i.e. a migrant to the US) for each individual in the "treatment" group (i.e. for each migrant to Argentina).

The matching is performed using the same individual-level characteristics that are used as control variables in column 4 of Table 2: age, years since arrival, origin port, pre-migration occupational category and literacy. In the first step, I split individuals into mutually exclusive strata based on the joint values of these variables.[5] Then, for each migrant to Argentina I select a migrant to the US within their same stratum to serve as control individual; if there are no such individuals, the observation is discarded from the sample. After having created the treatment and control groups, I simply estimate (using the matched sample):

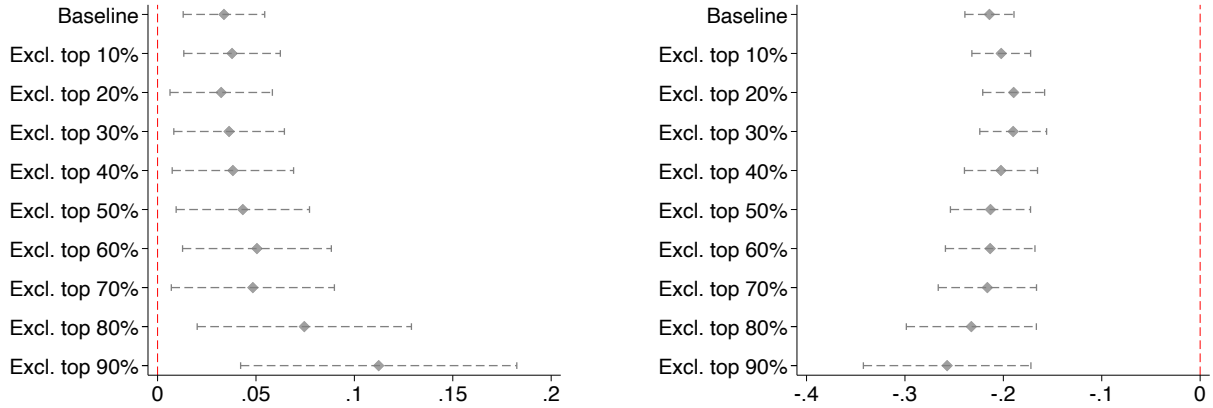
$$y_{ic} = \alpha + \beta Argentina_{ic} + \epsilon_{ic} \tag{1}$$

Finally, since the number of treatment and control individuals within each strata is not necessarily equal, the data are weighted according to the size of each strata. This procedure yields the estimates reported in column 6 of Table 2.

[5] For these purposes, I coarsen continuous variables (age and years since arrival) using Sturges' rule (Iacus *et al.*, 2012). Sturges' rule is the default coarsening rule in the *cem* command in Stata. For categorical variables (origin port, pre-migration occupational category and literacy), I use a variable's exact values. This procedure splits the data into 5,291 mutually exclusive strata.

Figure B1: 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 **??** (including surname fixed effects).
Source: Data are from the samples linking passenger lists to the census as described in the main text.

Table B1: Selection of Italian immigrants in Argentina and the US

Variable	Argentina (1)	US (2)	Difference (3)
i. Demographic (N=1,335,705)			
Age upon arrival	26.719	26.975	-0.256***
Age less than 16	0.207	0.198	0.008***
Female	0.277	0.249	0.027***
ii. Occupation (N= 752,385)			
White-collar	0.035	0.028	0.007***
Farmer	0.429	0.221	0.208***
Skilled/Semi-skilled	0.110	0.148	-0.039***
Unskilled	0.426	0.603	-0.176***
Unskilled/Farmer	0.855	0.824	0.032***

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 B2: Multinomial Logit Model of Destination Choices within the US

	(1)	(2)	(3)
Network	0.494***	0.492***	0.487***
	(0.0321)	(0.0321)	(0.0322)
Observations	540846	540846	540846
Port of origin FE	No	Yes	Yes
Individual-level characteristics	No	No	Yes

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. This table shows the result of estimating a conditional logit model of migrants intended destinations within the US. The table focuses on the top 20 intended destinations in the 1882-1900 period, relative to the "other" category. The *Network* variable measures the relative frequency of an individual's surname across different intended destinations. 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.
Source: US passenger lists data as described in the main text.

Table B3: Likelihood of Holding an Unskilled Occupation, First Differences

	(1)	(2)	(3)	(4)
Italian in Argentina	-0.116***	-0.130***	-0.303***	-0.235***
	(0.00803)	(0.0181)	(0.00708)	(0.0161)
Grouping Farmers/Unskilled upon Arrival	No	No	Yes	Yes
Years since arrival	No	Yes	No	Yes
Port of origin	No	Yes	No	Yes
Literacy	No	Yes	No	Yes
Surname	No	Yes	No	Yes
Observations	30620	30620	30620	30620
R^2	0.0138	0.455	0.0729	0.479

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. The dependent variable in this table is the difference between the likelihood of holding an unskilled occupation in the host country minus the likelihood of holding it upon arrival. In columns 3 and 4, individuals declaring farming or an unskilled occupation upon arrival are grouped into a single "unskilled/farmer" category. The odd columns include just age fixed effects. The even columns 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 B4: The Economic Outcomes of Italians in Argentina and the US, Altonji *et al.* (2005) and Oster (2017) bounds

	Home ownership		Unskilled occupation	
	(1)	(2)	(3)	(4)
Italian in Argentina	0.0347***	0.0315***	-0.222***	-0.216***
	(0.00750)	(0.00752)	(0.00922)	(0.00903)
Controls	No	Yes	No	Yes
Observations	30620	30620	30620	30620
$\frac{\beta_C}{\beta_{NC}-\beta_C}$		10.02		36.61
Delta		4.476		21.26

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. The last two rows show the degree of selection on unobservables, relative to the selection on observables, necessary to overturn the results. The second to last row shows the ratio between the estimated coefficients in the restricted and unrestricted models, following the approach in Altonji *et al.* (2005). The last row follows the approach in Oster (2017), using a value of R_{max} equal to 1.3 times the R^2 in the regression with the full set of controls.

Table B5: The Occupations of Italians in Argentina and the US

	White Collar		Farmer		Skilled Blue Collar		Unskilled	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Italian in Argentina	0.0623***	0.0506***	0.0638***	0.0347***	0.187***	0.190***	-0.252***	-0.214***
	(0.00412)	(0.00937)	(0.00446)	(0.0102)	(0.00508)	(0.0115)	(0.00563)	(0.0127)
Years since arrival	No	Yes	No	Yes	No	Yes	No	Yes
Port of origin	No	Yes	No	Yes	No	Yes	No	Yes
Literacy	No	Yes	No	Yes	No	Yes	No	Yes
Occupation	No	Yes	No	Yes	No	Yes	No	Yes
Surname	No	Yes	No	Yes	No	Yes	No	Yes
Observations	30620	30620	30620	30620	30620	30620	30620	30620
Mean of dep. var.	0.101	0.101	0.137	0.137	0.158	0.158	0.476	0.476
R^2	0.0188	0.454	0.00895	0.443	0.0543	0.482	0.0768	0.494

Notes: $***p < 0.01$, $**p < 0.05$, $*p < 0.1$. This table compares the occupations of Italians in Argentina and the US. The mean of the dependent variable is computed among Italians in the United States. The odd columns include just age fixed effects. The even columns 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.

References

Abramitzky, R., Boustan, L.P., Eriksson, K. and Feigenbaum, James J.and Pérez, S. (2018). "Automated linking of historical data", NBER WP 25825.

Abramitzky, R., Mill, R. and Pérez, S. (2019). "Linking individuals across historical sources: A fully automated approach*", *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, vol. 0(0), pp. 1–18, doi:10.1080/01615440.2018.1543034.

Alexander, R., Ward, Z. *et al.* (2018). "Age at arrival and assimilation during the age of mass migration", *Journal of Economic History*.

Altonji, J.G., Elder, T.E. and Taber, C.R. (2005). "Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools", *Journal of political economy*, vol. 113(1), pp. 151–184.

Bailey, M., Cole, C., Henderson, M. and Massey, C. (2017). "How well do automated linking methods perform in historical samples? evidence from new ground truth", Working Paper.

Biavaschi, C., Giulietti, C. and Siddique, Z. (2017). "The economic payoff of name americanization", *Journal of Labor Economics*, vol. 35(4), pp. 1089–1116.

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

de Inmigración, D.G. (1925). "Resumen estadístico del movimiento migratorio en la república argentina. años 1857–1924", .

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

Iacus, S.M., King, G. and Porro, G. (2012). "Causal inference without balance checking: Coarsened exact matching", *Political analysis*, vol. 20(1), pp. 1–24.

Oster, E. (2017). "Unobservable selection and coefficient stability: Theory and evidence", *Journal of Business & Economic Statistics*, pp. 1–18.

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*, vol. 77(4), pp. 971–1006.

Ruggles, S., Sobek, M., Fitch, C.A., Hall, P.K. and Ronnander, C. (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.

Winkler, W.E. (1990). "String comparator metrics and enhanced decision rules in the fellegi-sunter model of record linkage.", .