Mass Emigration and Human Capital over a Century: Evidence from the Galician Diaspora

Martín Fernández-Sánchez*

JOB MARKET PAPER
Click here for the latest version

September 22, 2020

Abstract

This article examines the effects of emigration on human capital accumulation at origin throughout a century. The context is the Galician diaspora, a massive episode with the equivalent of 60% of the region 1900's population emigrating to Latin America during 1900-30. I construct a database of all Galician municipalities combining newly-digitized historical data with contemporary census and survey data and exploit two sources of plausibly-exogenous variation for identification: pioneer emigration caused by extreme rainfall and changes in economic growth in the main migrant destinations. I find that while emigration depressed literacy rates at origin in the short run, its impact became positive after one decade and led to gains in human capital that still persist one hundred years later. I provide evidence of two novel mechanisms on how emigrants raise human capital in the long run. Galician emigrants funded associations that financed the construction of schools in their hometowns and diffused norms conducive to a persistent change in beliefs about the value of education and effort.

Keywords: Mass emigration, Human capital, Migrants' associations, Values, Long-run.
JEL Classification: I25, J11, O15, N33, N34.

*Paris School of Economics. E-mail: m.fernandez-sanchez@psemail.eu. Contact: +33 782 386 262
I am grateful to Hillel Rapoport, Andrew Clark, Oliver Vanden Eynde, Paola Giuliano and Katia Zhuravskaya for very helpful discussions. I would also like to thank for their valuable comments: Pol Antràs, Sam Bazzi, George Borjas, Denis Cogneau, Gemma Dipoppa, Sofia Dromundo, Katherine Eriksson, Elisa Facchetti, James Feigenbaum, James Fenske, Jesús Fernández Huertas Moraga, Santiago Garriga, Andrea Ibba, Sylvie Lambert, François Libois, Elisa Macci, Karen Maccours, Bob Margo, David Margolis, Clara Martínez-Toledano, Luigi Minale, Jose Montalbán, Petra Moser, Nathan Nunn, Lorenzo Neri, Daniele Passerman, Tatiana Rosá, Sandra Sequeira, Marco Tabellini, Clémence Tricaud, Felipe Valencia Caicedo, Daniel Waldenström, Fabian Waldinger, and seminar participants at the Paris School of Economics, Harvard University, Boston University, University of Valencia, Carlos III University, Queen Mary University, University of Glasgow, the Ifo and RWI institues, the Virtual Economic History Seminar, and the International Conference on Migration and Development. I am also thankful to the staff of Biblioteca Nacional de España, Biblioteca Xeral Universidade de Santiago and Biblioteca del INE for their help in finding historical data and to Séamus O'Donovan who helped digitizing some of it. I gratefully acknowledge financial support from the Labex OSE.

1 Introduction

During the Age of Mass Migration (1850-1930), over 40 million Europeans crossed the Atlantic towards the Americas (Hatton & Williamson, 1998). This unique historical event led to profound transformations in both continents whose effects can still be seen today. Recent research has examined the long-term consequences of migration for receiving or host countries, showing that it promoted economic development and higher levels of human capital (Rocha, Ferraz and Soares, 2017; Droller, 2018; Sequeira, Nunn and Quian, 2020). The implications for the sending communities however remain much less unexplored (Karadja and Prawitz, 2019; Anderson, Karadja and Prawitz 2020). Specifically, there is still no empirical evidence on the long-run effects of emigration on human capital accumulation at origin.

A long-term perspective is especially relevant in the context of human capital accumulation since the impact of emigration and the mechanisms involved may change over time. As underlined by the empirical and theoretical literature (McKenzie and Rapoport, 2010; Dustman and Weiss, 2007), the type of selection into emigration, the incentives to acquire education and the extent of return migration are all key factors that play a role at different stages. Moreover, migrants may also have a long-run impact on their communities of origin through more subtle mechanisms. For instance, they may contribute to the financing of public goods or the shaping of norms and values that encourage investments in human capital of the future generations.

This paper examines the impact of mass emigration on human capital accumulation at origin thorough a century. I study one of the greatest emigration episodes of modern history in Europe: the Galician diaspora. Over the 1900-1930 period, more than 1.1 million people left this Spanish region towards Latin America, a figure representing 58% of its 1900 population. In relative terms, Galicia had higher emigration rates than those in any other European country at the time. The magnitude of this historical event, together with the similarities between Galicia in 1900-1930 and other developing countries today, make it an exceptional setting to study the short and long-run effects of mass emigration.

In this work, I combine newly digitized data from multiple historical sources (e.g., population censuses, embarkation lists and migrants' associations) with modern administrative and survey data to construct a unique database of Galician municipalities from 1860 to date. I use data on absent men as a novel proxy for emigration and instrument this via a measure of migrant networks (pioneer emigration caused by extreme rainfall) and a pull factor (economic growth at destination).

Using this empirical strategy, I analyze the drivers of emigration at the local level, its short and medium-run effects on literacy rates (1900-1930), and its long-run impact on different measures of educational attainment (1930s-2010s). In a nutshell, I find that while emigration depressed literacy rates in the short run, one decade later the effect became positive and led to gains in human capital that still persist around one hundred years later. Using data on migrants' associations and their investments at origin as well as census and survey data, I provide evidence that this long-term persistence is partially due to migrants' financing of new schools in their hometowns and a change in beliefs about the value of education and effort.

A key novelty of my study is to propose and validate a new proxy for historical emigration based on the concept of missing residents.[1] The population censuses of various countries (e.g., Spain, Italy, and Portugal) gathered information of all household members, including those temporarily absent.[2] I perform an exhaustive number of checks that validate the share of missing men as a good proxy for emigration. For instance, I show that the numbers of missing men and sea departures in Galicia are extremely similar, and follow the exact same trend between 1877 and 1930. I also find no evidence that these missing men could reside somewhere else in Galicia or Spain.[3]

Analyzing the effects of historical emigration on human capital accumulation is a challenge for a number of reasons. First, historical data can suffer from substantial measurement error, leading to attenuation bias. Second, there may be confounding factors affecting both migration flows and investments in education. Last, changes in education may influence migration decisions leading to reverse causality. To deal with these issues, I implement an empirical strategy in which I instrument the share of missing men by two sources of plausibly exogenous variation. In the historical analysis using decadal panel data between 1900 and 1930, I instrument the share of missing men in a given year by a time-varying pull factor and its interaction with a proxy for pioneer migrant networks (measured four decades before the onset of mass emigration). In addition, I control for municipality time-invariant and aggregate time-varying unobserved factors that capture most of the variation in missing men and literacy. When examining the long-run effects of emigration using a cross-section of municipalities, I instrument the average share of missing men during

[1] The study of European mass emigration has traditionally been limited by the lack of local level data. In Spain, for instance, official migration records only exist at the province level (49) and for a handful of years. Information on missing residents however is available for all municipalities and census years since 1877.

[2] Household members who did not sleep in the dwelling in the past two nights prior to the census were registered as missing regardless of the duration or reason of their absence. Since most Galician migrants were young men departing alone, household heads would report them as missing. Note that the census took place on December 31st precisely to minimize absences others than those due to emigration.

[3] One additional advantage of relying on data of missing men is that it captures clandestine emigration (i.e., overlooked by official statistics) that at the time represented 20% of recorded departures in Spain (Sánchez-Alonso, 1995).

1900-1930 by the proxy for migrant networks (1860) and control for other historical factors.

The main ingredient of my identification strategy is a proxy for pioneer migrant networks.[4] I rely on the fact that the first emigration waves (1840-1860) were largely driven by extreme rainfall that accentuated agricultural crises (Rodríguez Galdo, 1977; Hernández Borge, 1986; Eiras Roel, 1991). These pioneers fostered new departures in the following decades and were a key force during the mass emigration era (1900-1930). Since most pioneers were men, I rely on the adult sex ratio in 1860 across municipalities as proxy for the size of migrant networks.[5] As a pull factor, I use the variation in economic growth across destinations during each decade. In particular, I construct a measure of GDP per capita growth in the average destination country for every municipality using historical records on the destination of over 17,000 migrants during 1900-1930.

To assess the validity of my identification strategy, I perform several robustness checks. First, I show that the adult sex ratio in 1860 is uncorrelated with the evolution of relevant factors between 1860 and 1900, such as population or literacy rates. Second, I show that economic growth at destination and origin are completely uncorrelated and pay special attention to the concern that other local economic shocks (e.g., international trade and weather shocks) may confound my results. Third, I use alternative weights to construct my pull factor and exploit different measures of migrant networks.[6] Finally, I control flexibly for a large set of historical controls and perform multiple sample restrictions. Overall, the results remain robust to all these checks.

I begin my analysis by studying the drivers of emigration at the local level. Using both a decadal panel of municipalities between 1900 and 1930 and a cross-section, I find that the proxy for migrant networks is the major determinant of emigration. The adult sex ratio in 1860 explains 20% of the variation in the average share of missing men during 1900-1930. Higher economic growth at destination is associated with increases in the share of missing men, especially in areas with a larger migrant network. This mediating effect of migrant networks is consistent with other findings in the literature (McKenzie and Rapoport, 2010; Brum, 2019). I also find suggestive evidence that migrants were credit constrained, as rainfall shocks correlate negatively with changes in missing men. Overall, the instruments are highly significant and robust across all the specifications.

I then analyze the short and medium-run effects of emigration on literacy rates during the mass emigration era (1900-1930). Exploiting variation in the pull and network instruments, I find that

[4] I refer to migrant networks as a group of migrants from the same origin community and residing in the same place.

[5] I cannot rely on missing men in 1860 because this information is only available from 1877 onwards.

[6] For instance, I use the distance to the closest transatlantic port in a similar fashion to Karadja and Prawitz (2019), but taking into account the road network of Galicia in the nineteenth century. This variable reflects the diffusion of information and transportation costs. Additionally, I use the raw data of rainfall variability during 1840-1860.

emigration depressed literacy rates in the short run. On average, a one standard deviation rise in the share of missing men led to a 51% of a standard deviation drop in literacy rates. This negative impact however switches sign one decade later. Introducing the share of missing men one decade earlier as the main explanatory variable, I find that the medium-run effect of migration is positive and significant for men and zero for women. I discuss potential mechanisms that may explain these effects. First, I present data on migrants' characteristics and suggest that the initial drop in literacy can be partly explained by the departure of people who were more educated than the average. Second, the medium-run effects could be accounted for greater investments in children's education and return migration. Using literacy information by cohort, I find suggestive evidence that i) emigration increased literacy rates for both boys and girls, ii) there was no negative selection into return migration and iii) men were much more likely to return than women.

I next examine the long-run effects of mass emigration on human capital accumulation. Instrumenting the average share of missing men (1900-1930) with the migrant network measured in 1860, I find that historical emigration led to persistent gains in human capital (1981-2011). A one standard deviation rise in the average share of missing men (i.e., 7 percentage points) is associated with a 50% of a standard deviation rise in the share of individuals with completed secondary education in 2011 (i.e., 5 percentage points). In line with the medium-run effect in the historical analysis, the impact for women is delayed and becomes positive and significant only in the 2000s. I find similar positive effects, but smaller in magnitude, when looking at the share of college-educated individuals. To see whether these effects can be driven by internal migration, I separate the sample between individuals who were born in the municipality and those who moved there. There turn out to be no significant difference in the results for these two groups. I also examine other contemporary outcomes and find no impact on the number of schools or libraries, but a positive impact on the number of local associations with an educational goal.

After establishing that mass emigration led to gains in human capital in the long run, I explore two potential mechanisms for this long-term persistence: the role of migrants' investments at origin and social remittances. Galician migrants founded local associations with the primary goal of financing the construction of schools in their hometowns, and founded over 300 new schools between 1910 and 1940 (14% of the stock of public schools in 1908). I collected and assembled data on hundreds of these migrants' associations with detailed information on the country and year of foundation, the Galician municipality to which they are related, and the exact location and year of opening of the schools they financed. I then merge this database with education data by birthplace

and cohort from the 1981 census. A cohort analysis reveals that migrants' schools had a significant and positive impact, increasing the rates of primary education completion for the children young enough to benefit from them. Consistent with historical narratives, I find that migrants' associations seem to have encouraged further migration. However, it is important to note that since most associations financed several schools, the net effect on educational attainment is positive.

To better understand the motivations of migrants to found associations and invest in their hometowns, I leverage historical data from different sources and perform a cross-sectional analysis. I first find that the size of the migrant community seems to have played a crucial role. A one standard deviation increase in the average share of missing men (1900-1930) is associated with one additional migrant association and 1.6 more migrant schools in the municipality of origin. Migrants were more likely to found an association and finance a new school if there was already a greater need for them to begin with. I also find that rainfall shocks (i.e., negative income shocks) are negatively and strongly correlated with the number of associations and schools founded, suggesting that migrants faced a trade-off between sending individual and collective remittances.

Another channel by which migrants could have fostered higher investments in human capital is social remittances related to education. Anecdotal evidence suggests that migrants realized the importance of education once in Latin America and transmitted these values in their letters, through the role of migrants' association and directly after their return. To test this hypothesis, I merge the historical data with contemporary survey data on beliefs and values. Using the proxy for migrant networks (1860) as instrument, I find that historical emigration (1900-1930) led to greater value being put on education and effort today. This result is robust to controlling for other historical and contemporary factors. Additionally, I find suggestive evidence that these effects are stronger in municipalities that were connected to Latin American countries where education was more valuable. I find no changes in other individual beliefs such as the degree of satisfaction with public education or the willingness to pay higher taxes to invest more in public goods.

My paper relates to a number of different literatures. First, it contributes to the growing research exploring the long-run effects of migration. Recent work has documented positive long-run effects on economic development and human capital accumulation in the areas receiving migrants (Rocha, Ferraz and Soares, 2017; Droller, 2018; Sequeira, Nunn and Quian, 2020). Regarding the impacts at origin, the literature has found positive effects on innovation (Anderson, Karadja and Prawitz, 2020), and mixed results in terms of economic development and human capital formation (Dinkelman and Mariotti ,2016; Testa, 2020). I provide novel evidence that, despite a short-run

loss, the Galician mass emigration led to gains in human capital that persisted across several generations. To understand the variety of results, it is crucial to consider the context of each study. Dinkelman and Mariotti (2016), for instance, examine the circular migration from Tanzania to South African mines, which for political reasons ended abruptly forcing all migrants to return home. This is a setting particularly beneficial for a positive impact on human capital formation given that all migrants had to return and they kept all their earnings. In the case of Testa (2020), he analyzes an episode of forced migration (Germans expelled from Czechoslovakia after WWII) implying that parts of the population were expelled from their hometowns and not allowed to come back. In this regard, the Galician diaspora provides a more general setting to understand the effects of emigration since the decisions to emigrate, return, or send remittances, were not influenced by institutional factors or wars. Moreover, it was similar to other migration flows from Southern Europe during the age of Mass Migration and to other migration flows today, characterized by male migrants often leaving their family behind, and thus, maintaining strong links with their communities of origin in clear contrast with episodes of forced migration (Becker and Ferrara, 2019).

Second, I contribute to the literature on the effects of migration on the education of the left-behind and the debate on the brain drain from a historical perspective.[7] The research on the Age of Mass Migration has pointed to both negative (Abramitzky, Boustan and Eriksson, 2012; Connor, 2019) and positive (Spitzer and Zinram, 2018) selection of migrants depending on the context. My results suggest that Galician migrants were positively selected and although their departure reduced literacy rates in the short run, the high levels of return migration (in line with other countries, Bandiera, Rasul and Viarengo, 2013) and greater investments on children's education led to net gains in human capital. This is the first evidence of positively selected migration followed by brain gain in the context of the Age of Mass Migration.

My paper also relates to a recent literature on migration and culture. For instance, Knudsen (2019) shows that migrants from Scandinavian countries were on average more individualistic and, as a result, their departure raised the level of collectivism at origin.[8] Becker et al. (2020) show that Polish forced migrants subsequently became less materialistic and invested more in education. I contribute to this literature by examining how emigration affected beliefs about the value of ed-

[7] See for instance McKenzie and Rapoport (2006, 2010), Antman (2012), Batista, Lacuesta and Vicente (2011), Cortes, Gibson and McKenzie (2011, 2012), and Shresta (2016). Docquier and Rapoport (2012) review the existing literature.

[8] In a related work, Bazzi, Fiszbein and Gebresilasse (2020) show that individuals living in areas that were the US frontier for longer periods of time display higher levels of individualism.

ucation and effort. Rather than the result of migrant selection or forced migration, my findings seem to suggest that an exposure to an environment where education is highly valued may make migrants change their norms and transmit these to their communities of origin.

Finally, some work has examined the role of non-governmental institutions that provide public goods and contribute to local development, investments in human capital and cultural change. The evidence ranges from religious institutions centuries ago (Valencia-Caicedo, 2019; Calvi et al., 2019) to contemporary migrant associations (Chavet et al., 2015). I document that Galician migrants established hundreds of associations with the aim of investing in their home communities, what could be understood as a pioneer form of development aid. I analyze the factors influencing the origin of these associations, their decision to construct schools in their villages, and their impact on educational attainment and further migration.

The remainder of the paper is structured as follows. Section 2 provides some historical background. Sections 3 and 4 describe the data and empirical strategy. Section 5 and 6 discuss the results and mechanisms. Finally, Section 7 concludes by drawing some policy implications.

2 Historical background

2.1 Galician mass emigration

Galician mass emigration in perspective. The Galician diaspora is one of the largest emigration episodes in the twentieth century. Between 1900 and 1930, over 1.1 million Galicians left this Spanish region, which represents 58% of its initial 1900 population (Vázquez-González, 1999, p. 375).[9] As shown in Table 1, Galicia had higher gross emigration rates than any other European country at the time. Return migration was relatively high, close to 70% during 1900-1930. Based on these figures, Galicia accounts for around 45% of Spanish net migration to Latin America (Vázquez-González, 1999, p.379). Galician mass migration was concentrated in 1900-1930, stopping abruptly in the early 1930s after Latin American countries underwent severe economic crises and introduced migratory restrictions. The yearly flows and returns are displayed in Figure A2.

Causes of migration. The first emigration waves to Latin America took place in 1840-1860 and it is

[9] This estimate is a lower bound since official statistics do not capture clandestine migration that could exceed 20% (Sánchez-Alonso, 1995). For the 1835-1930 period, Vázquez-González (1999, p. 374-75) estimates between 1.3 and 1.6 million departures while Eiras Roel (1991) a figure of 1.7 million (for 1840-1930). In terms of return migration, the official statistics document 600 thousands in 1900-1930 although the data on returns may be even more underestimated than that of departures.

argued to have been caused by agricultural and rural textile industry crises (Bertrand, 1997; Eiras Roel, 1991; Hernandez, 1999). As shown in Figure 1, rainfall levels were extraordinary high and volatile in those decades, what could have been detrimental for the cereal harvests and the flax used in textile. The effect of these shocks would have been accentuated by an archaic agriculture (Álvarez Campos, 1993; Farías, 2010) and a high demographic pressure (Eiras Roel, 1991, Vázquez-González, 1993). While emigration kept increasing steadily during the following decades, it remained relatively low. In the twentieth century, however, it reached unprecedented levels fostered by the rapid economic growth in Latin America, transportation innovations[10] and the contribution of migrant networks (Álvarez Campos, 1993; Farias, 2010).[11] Importantly, Galicia already had a tradition of seasonal migration to the North of Portugal and Castille during the XVIII and XIX centuries (Farías, 2010).

Destinations. The main destinations were Rio de la Plata (58%), Cuba (32%), and to a lower extent Brazil (9%). Due to the strong influence of migrant networks, most migrants from a given area tended to go to the same destination.[12] Yet, as illustrated by Figure 3, there was a large geographical heterogeneity in the relative importance of each country. Section B in the Appendix provides more details about migrant destinations.

Migrants' profile. Galician migrants were mostly men (over 70%), young (50% aged 16 to 30) and rural farmers or craftmen (90%) (Vázque-González, 1999).[13] Migration was a family strategy under which one or more members would go to Latin America temporarily to complement household income and to reduce the burden at home. Once at destination, they located almost exclusively in large urban centers and worked in services (e.g., small stores, catering, retail, storage, etc.). Migrants usually starting working for a relative or acquaintance until having enough savings to set their own business and eventually return after a few years (Álvarez-Campos, 1993; Vázquez-González, 1999).

[10] In the 1870s, sailing from Galicia to Cuba and Argentine took around one month and half and two months respectively. The spread of steamships in the 1880s reduced drastically travel times, making it possible to reach Cuba in 19 days and Argentine in 25 days (Vázquez-González, 1993).

[11] Other factors stressed in the literature are the difficult access to land due to the "foros" regime and the heritage system (Betrand, 1997), the high tax burden (Eiras Roel, 1989), and military desertion (Álvarez Campos, 1993). See the Appendix for a brief summary of the economic history of Galicia.

[12] Vázquez-González (1999, pp.882-886) and Seixas-Núñez and Vazquez (2004) show that this pattern is also present across parishes within the same municipality.

[13] These figures are a probably a lower bound since clandestine migration was characterized by young men from rural areas (Vázquez-González, 1999, p. 383).

Migrant associations. In the mid-nineteenth century Spanish migrants created associations of mutual insurance providing health care to their members.[14] These were followed by regional associations in the 1870s and local associations in the twentieth century.[15] Between 1900 and 1930 hundreds of Galician local associations were created thanks to the large inflow of migrants and their concentration in the same cities in Latin America.[16] The grounds for local associations were often an informal gathering, for example, to celebrate a feast in honor of the village's saint (Fernández-Santiago, 2008). Through these associations migrants raised funds to invest in public goods in their hometowns such as roads, schools, etc. While the membership of local associations started to decline in the 1930s (until their extinction in the following decades), regional Galician centers have persisted until nowadays. Most migrant associations had an educational goal and among their main objectives was the construction of schools in their hometowns. Over 300 hundred schools were founded by these migrant associations in the first half of the twentieth century, representing around 7% of the stock of public schools at the time. Figure A22 and Figure A23 in the Appendix show the foundation of associations and schools over time and their location.

2.2 Education in Galicia

Until the first decades of the XX, the education system in Galicia presented a large deficit of material and human resources, low attendance rates, poor educational outcomes and large gender inequalities (De Gabriel, 2006). While the education laws required a minimum number of schools depending on municipalities' population, the highly dispersed settlements within Galicia municipalities led to the closest school often being located several kilometers away and to a ratio of schools per children lower than the national average.[17] The difficult access to schools was aggravated by the abrupt geography and poor transportation infrastructure. The conditions in most establishments were penurious; muddy ground, lack of light and ventilation, shared materials,

[14] Some examples are "Sociedad de Beneficiencia de Naturales de Cataluña", founded in La Habana (Cuba) in 1840, "Sociedad Española de Beneficiencia" and "Asociación Española de Socorros Mútuos" founded in Buenos Aires (Argentine) in 1852 and 1857 respectively. The first Galician association of this kind appeared in 1871 in La Habana: "Sociedad de Beneficiencia de Naturales de Galicia" (Fernández-Santiago, 2008).

[15] The Centros Gallegos were founded simultaneously in La Habana, Buenos Aires and Montevideo in 1879. In Buenos Aires there were also associations from other regions: Centro Aragonés (1885), Centre Catalá (1886), Centro Balear (1905) and Centro Asturiano (1913). The first local associations of Galician migrants appeared in 1904 ("Alianza Arensana de Instrucción" in Cuba and "La Concordia" in Argentine).

[16] Vázquez-González (1999) documents 665 of them, among which over 400 had an educational goal.

[17] Moyano law (1853) required at least one school in municipalities with more than 500 inhabitants and two schools if population exceeded 2,000. While in 1910 Galician municipalities had on average around 6,000 inhabitants (compared to 2,100 in Spain), people often lived in parishes of less than 500 people (Martínez-Dominguez, 2000, p. 49). The average number of parishes in Galician municipalities is 12.

etc. Teachers lacked qualification and in many municipalities this role was done by a priest.[18] The curriculum consisted in Catholic doctrine (which involved learning how to read), writing, basic arithmetic and domestic tasks in the case of girls (Martínez-Domínguez, 2000, p. 50).

The poor quality offered, the obstacles to access schools, and the low value parents put in education made school enrolment very low and attendance irregular, peaking in winter when children were less demanded for agricultural tasks (De Gabriel, 1970). The lower enrolment of girls compared to boys has been argued to be the result of parents prioritizing boys' education given that they would be the future household heads and because it could be more profitable in the prospects of migration (Vázquez-González, 2002; Martínez-Domínguez, 2002).[19]

3 Data

This paper draws on newly digitized data from multiple historical sources. A summary of all the sources can be found in Table B1 and Table B2 in the Appendix.

Population censuses. The main data comes from the seven population censuses of Spain between 1860 and 1930.[20] I digitized all the information available for the 321 Galician municipalities; the number of present, missing, transient and literate individuals by sex. Household members were classified as present or missing depending on whether they had been in the municipality the two nights prior to the census day. I use the share of missing men in a municipality as proxy for migration and discuss its validity in the next section. Transients are individuals found in a different municipality than the one where they regularly reside in, typically representing less than one percent of the population. Household heads reported literacy information for all present members. I compute literacy rates as the share of present individuals knowing how to read and write.

Missing men as proxy for migration. Due to the absence of migration data at the local level, I use the share of missing men as proxy for the stock of emigrants. This measure presents the advantage of being available for all municipalities and census years in a comparable manner. Moreover, it

[18] In 1855, around 51% of teachers in public schools in Galicia had a degree compared to 81% in Spain (Martínez-Domínguez, 2002). Moreover, until the 1870s most schools had only one teacher and a single group of students with different ages.

[19] Note that the supply of school played no role in gender disparities as most establishments were mixed. Even once attending school, girls received fewer hours of instruction in literacy and numeracy compared to boys (Martínez-Domínguez, 2002).

[20] The Spanish population censuses were conducted in 1860, 1877, 1887, 1900, 1910, 1920 and 1930. Figure A3 in the Appendix provides an illustration of the original documents.

captures clandestine migration that at the time could exceed 20% (Sánchez-Alonso, 1995). Note that the census was conducted the 31^{st} of December precisely to minimize the number of missing and transient individuals as most people were expected be home those days.[21] I conduct a number of descriptive analysis to examine the validity of this measure. First, the figures of missing men and net migration in Galicia are almost identical and follow the exact same trend over the 1877-1930 period (Figure 5). Similarly, the share of missing men and gross emigration rates are strongly correlated across the Galician provinces in different years (Figure 6), but also across the 47 Spanish provinces (Figure A4) and even the Italian provinces (Figure A5). Second, the transient population in Galicia remained constant and below one percent over the period, what supports the idea that missing men were not in neighboring municipalities or elsewhere in Spain (Figure A6).[22] Galicians where the least likely to live in another Spanish region (3% compare to an average of 10%). Additionally, the number of drafted men in Galicia was fairly stable in absolute and relative terms suggesting that the draft played no role driving the trends in missing men (Figure A7).[23] Note that I do not rely on the share of missing women for a number of reasons. First, as shown in Figure A4, when emigration flows are low the figures on missing residents and sea departures are not strongly correlated. Between 1900 and 1930 the share of missing women ranged only between 2% and 3%, which suggests it may not accurately capture female departures but rather other transitory movements out of the household. Moreover, there were tighter restrictions to women's migration at the time, which made them more likely to emigrate to join family members at destination and hence with a more permanent nature (Sánchez-Alonso, 1991). As a result, female migrants were more likely to not be reported as missing in their previous household unit. I will show nonetheless that the results of the analyses are very similar using the share of missing residents instead of the share of missing men.

Sex ratios as proxy for migrant networks. Since the 1860 population census does not contain information on missing residents, I rely on the adult sex ratio as proxy for the intensity of the first migration waves to Latin America (1840-1860). The pioneers left Galicia mostly as a response to

[21] *"La fecha fijada por la ley para llevar a efecto la inscripción de habitantes es la más apropiada para obetener la población en su mayor estado de reposo, o sea para reducir a una cifra mínima el número de ausentes y el de transeuntes"* (Spanish population census 1920, p. 68).

[22] It is unlikely that missing men reflects internal migration since it was mostly concentrated between the months of May and September (Palmás, 1984).

[23] Given that there were no consequences of having missing household members, household heads should have no incentives to miss-report. Individuals were warned that they would be subject to a fine if they provided inaccurate information.(XX)

the extreme weather conditions of those decades and their impact on agriculture. Figure 2 shows that there is a strong correlation between the adult sex ratio and the intensity of rainfall variability in 1830-1860.

Economic conditions at destination as pull factor. Galician migrants were attracted by the economic conditions in Latin America. I use data from the last release of the Maddison Project Database (2018) to construct a measure of GDP per capita growth in the average destination. To infer the main destinations across Galician municipalities I digitized information compiled by Vázquez- González (1999) of over 17 thousand people emigrating during 1900-1930.[24] This information comes from a variety of sources: embarkation lists, notarial documents and desertion records. As a robustness, I use official data of sea departures at the province level in 1885-1895, the earliest period this data is available. Section B in the Appendix describes how the pull factor measures were constructed. Figure C1 shows the variation in economic growth at destination across municipalities and over time.

Migrant associations and schools. I constructed a database of over 400 Galician migrant associations in Latin America with information on the country and year of their foundation, the Galician municipality to which they relate, their intervention focus, the schools they constructed in Galicia and the opening year of such schools. The sources used are detailed in Table B1 in the Appendix.

Contemporary data. I rely on two alternative data sources to obtain measures of educational attainment: the population censuses (1981-2011) and the regional household surveys (2007-2018). To trace internal mobility, I use administrative data from the population registers ("padrón" of 1998 and 2003). The surveys "Opinión Pública y Política Fiscal" by the research center CIS provide me with information on attitudes towards education, value of effort and preferences over redistribution (2011-2018). Additionally, I use data on schools and enrollment at the municipality level from the official statistics of 1975 and 2018 and information about the universe of local associations (around 32 thousand in 2018) and public libraries.

Alternative proxy for migrant networks. I use the walking distance to the closest transatlantic

[24]The data prior to 1900 is much more limited but there seems to be a high degree of persistence in the main destinations. For instance, the correlation between the share of migrants going to Rio de la Plata in 1900-1930 and in 1850-1860 across districts is 0.54.

port as an alternative proxy for the size of migrant networks. Due to the poor transportation infrastructure of Galicia, being close to a port was crucial for the diffusion of information and for the establishments of stronger migrant networks.[25] For each municipality, I calculate the shortest distance to a port following the road network of the nineteenth century based on an official historical map that I geo-referenced (Figure E1). Section D of the Appendix provides further details.

Rainfall shocks at origin as push factor. I use data from the European Gridded Seasonal Precipitation Reconstructions (Pauling et al., 2006) to construct measures of weather shocks. The fine resolution of the data-set (0.5° x 0.5°) yields 24 cells for Galicia. I focus on rainfall during the Fall season given that maize was the main crop cultivated in the region and it was harvested in that period. For every decade, I compute the standard deviation of rainfall as shown in Figure D1.[26]

Supplementary historical data. I exploit supplementary data coming from multiple sources described in Table B1 in the Appendix. These include newly digitized data at the municipality level on wealth and taxation (Madoz, 1845), schools and enrollment (1875, 1908), and statistics of livestock across districts (1865, 1897, 1917).

4 Empirical strategy

The objective of this paper is to examine the impact of mass emigration on human capital accumulation at origin over a century. Fist, I use a decadal panel of Galician municipalities in 1900-1930 to understand the determinants of emigration (Section 5.2) and analyze its short and medium-run impact on literacy rates (Section 5.3). I then look at the long-run effects of historical emigration (1900-1930) on different education-related outcomes using a cross-section of all municipalities (Section 5.4). Finally, I explore different mechanisms exploiting additional historical data as well as contemporary administrative and survey data across individuals, cohorts and municipalities (Section 6). Below I describe the different empirical strategies and the main identification challenges in each of them.

[25] Vázquez-González (1999) documents the areas of activity of several migration hooks in the 1850s. These agents operated in municipalities surrounding the ports of Coruña, Vigo and Vilargacía and expanded their activities inland over time.

[26] This data has been recently used as push factor in other works such as Sequeira, Nunn, Qian, 2020. Section C in the Appendix provides more details.

4.1 Short and medium-run effects [Panel analysis]

Estimating equations. To examine the dynamic relationship between changes in missing men and literacy rates, I exploit a panel of 321 municipalities during four periods (1900, 1910, 1920, 1930). I refer to the short-run impact as the contemporary relationship and to medium-run impact as the one observed one decade later. The main estimating equations are:

Short-run effect:

$$Literacy_{it} = \alpha_i + \alpha_t + \phi ShMissing_{it} + X_{ijt-1}\Gamma + \epsilon_{it} \quad (1)$$

Medium-run effect:

$$Literacy_{it} = \alpha_i + \alpha_t + \psi ShMissing_{it-1} + X_{ijt-2}\Theta + v_{it} \quad (2)$$

where $Literacy_{it}$ is the share of people knowing how to read and write in municipality i and year t, θ_i and θ_t represent municipality and year fixed effects, $ShMissing_{it}$ is the share of missing men in year t and ϵ_{it} stands for robust standards errors clustered at the municipality level. The vector X_{ijt-1} represents lagged controls (i.e., population, literacy rates and GDP per capita in the province). $ShMissing_{it-1}$ represents the share of missing men mesaured one decade before.

Estimating equations (1) or (2) by OLS may give rise to three concerns. First, measurement error in the share of missing men and literacy rates would lead to attenuation bias and less precise estimates. Note that the share of missing men is an imperfect proxy of the stock of emigrants and other scholars have identified inconsistencies in literacy information in the census (De Gabriel, 2006). Second, there are potential confounders of migration that could bias the results in different ways (e.g., a negative income shock in a municipality could affect both investments in education and migrant flows). Third, the analysis could suffer from reserve causation if a raise in literacy rates affected the diffusion of information about the prospects abroad (Sánchez-Alonso, 1995) or if it deterred migration by revitalizing the local economy. To deal with these issues, I implement an instrumental variables strategy exploiting two sources of plausibly exogenous variation.

Instrumental variables. I instrument the share of missing men with a pull factor (average GDP per capita growth at destination) and its interaction with a proxy for migrant networks (the adult sex ratio in 1860). The intuition is that faster economic growth at destination would attract more migrants, but its impact could be amplified by migrant networks. Recent empirical work has

stressed the prominent role of networks during the Age of Mass Migration (Spitzer & Zimran, 2018; Brum, 2019) but also in the Galician case (Vázquez-González, 1999). While the proxy for the migrant network is constant over time, the pull factor varies according to the economic growth at the average destination. The strategy exploits thus variation provided by these two forces at the local level. The first stage equation is:

$$ShMissing_{it} = \theta_i + \theta_t + \delta\ Pull_{it} + \rho\ Pull_{it} \times Network_i + X_{ijt-1}\Pi + u_{it} \quad (3)$$

where i indexes municipalities and t years. The variable $Pull_{it}$ represents the GDP per capita growth at the average destination country, $Network_i$ represents the adult sex ratio in 1860, X_{ijt-1} a set of lagged controls and u_{it} standard errors clustered at the municipality level. The pull factor is constructed using information on the destination of over 17 thousand Galician migrants between 1900 and 1930. Formally:

$$Pull_{it} = \sum_c w_{ic} \times \Delta y_{ct} \quad (4)$$

where w_{ic} is the share of migrants from municipality i going to country c during 1900-30 and Δy_{ct} the GDP per capita growth of country c in the previous decade. As shown in Figure C1, there is a lot of geographical and time variation in this measure because municipalities were connected to different countries (Figure 3) and their economic growth was not correlated (Figure 4).

Identifying assumption. The identification requires that economic growth at destination is not correlated with any time-varying factor that could affect literacy rates differently depending on the initial sex ratio in 1860. Note that the identification does not require the sex ratio in 1860 to be exogenous as the effect of any time-invariant variable would be captured by the municipality fixed effects. The strategy would be flawed for instance if (i) economic growth at destination and origin were correlated and (ii) if the adult sex ratio in 1860 was correlated with economic activity.[27] None of these hypotheses seems to be supported by the data. Figure A9 shows that the evolution of GDPpc at destination is uncorrelated with changes in economic activity at origin. This result is not surprising given that Latin America represents less than 10% of Spanish trade in 1900-1930 and that Galicia was not particularly open to international trade in that period (see **??**). Goldsmith-Pinkham, Sorking and Swift (2019) highlight that the exogeneity of the time-invariant

[27] This would imply that the instrument predicts higher emigration (when high economic growth in Latin America and having a lower sex ratio) as well as an changes in income levels at origin that could in turn affect literacy rates.

component is crucial in a shift-share like instrument. The fact that the first migration wave (1840-1860) was mainly caused by extreme rainfall gives support to the idea that the adult sex ratio in 1860 may be exogenous, or at least conditional on covariates. I test formally the relationship between the adult sex ratio in 1860 and several pre-migration variables (1850-1860) showing that it is not correlated with educational inputs or the evolution of population or literacy rates between 1860 and 1900 (Figure A10). Recent work by Jaeger, Ruist and Stuthler (2019) argue that shift-share instruments may confound the shor tand long-run effects of immigration. This concern is less likely to apply in my analysis given the decade nature of my data. Similarly, Christian and Barret (2017) caution about the use of instruments with a time varying factor in an interaction as there could serial correlation at the aggregate level. My strategy does not suffer from this concern given that literacy rates experience an upward trend over the period while missing men display an inverted-u pattern. Note that I will perform a series of robustness checks to rule out that my results are driven by confounders.

4.2 Cross-sectional long-run analysis

Estimating equation. To examine the long-run effects of historical emigration on contemporary outcomes I rely on a cross-section of all Galician municipalities. The main explanatory variable is the average share of missing men in 1900-1930 that I instrument with the adult sex ratio in 1860, a proxy for pioneer migration. The estimating equation is:

2^{nd} Stage:

$$E_i = \alpha + \beta\, \overline{ShMissing_i} + Z_i\Omega + \omega_i \quad (5)$$

1^{st} Stage:

$$\overline{ShMissing_i} = \lambda + \gamma\, Network_i + Z_i\Phi + e_i \quad (6)$$

where E_i represents different educational outcomes, Z_i historical / contemporary controls, and $Network_i$ is the adult sex ratio in 1860. The standard errors ω_i and e_i are clustered at the district level (n=47).

Identifying assumption. The identification strategy requires that the adult sex ratio in 1860 is uncorrelated with other factors affecting human capital accumulation over the twentieth century. Although this assumption cannot be tested directly the balance test performed in Table A4 shows that the adult sex ratio in 1860 is not correlated with relevant covariates such as wealth per capita

(1855), local taxes per capita (1855), the supply of schools (1875), school enrollment rates (1875), changes in population or in literacy (1860-1900). There is a small (yet significant correlation) between the sex ratio and literacy rates in 1860, however it becomes insignificant once province FE are introduced. Controlling for different historical factors barely affects the estimates of the first stage nor the second stage. In the regressions I will control for population size, literacy rates, wealth per capita and local taxes in 1855-1860 although this inclusion makes no difference in the results.

4.3 Cohort analysis

Estimating equation. To evaluate the impact of migrant associations and the schools they built in their hometowns I perform a cohort analysis. The identification is very similar to Duflo (2001) with the addition of considering several cohorts given that associations and schools opened at different years in 1900-1940. I exploit data from the 1981 population census with information on the number of individuals by birth cohort (5-year groups) and educational attainment across municipalities. I restrict the sample to individuals born between 1902 and 1927 (aged 54 to 79 in 1981). The estimating equation is:

$$y_{ci} = \theta_c + \theta_i + \alpha\ Association_{ci} + \beta\ School_{ci} + v_{ci} \quad (7)$$

where y_{ci} represents the outcome of individuals born in municipality i (living there in 1981) from birth cohort c , θ_c and θ_i represent cohort and municipality fixed effects respectively, $Assoc_{ci}$ and $School_{ci}$ represent the number of associations / schools openened in municipality i when the birth cohort c was below 10 years old or younger.[28]

To explore the impact of migrant associations on net migration, I rely on the ratio between size of a cohort c and that of cohort $c-1$. Given that individuals are observed when aged 54 to 79, all emigrants coming back to their home municipalities would be included. To measure educational attainment I use the share of individuals who have completed at least primary education (e.g., 50% of men and 27% of women).

Identifying assumption. A causal interpretation of equation (7) relies on the assumption that there should be no time-varying local factors correlated with the opening of associations or schools

[28]This measure is takes into account the relative exposure of individuals within a cohort window. For instance, consider the cohort aged 6-10 in 1920; if the first migrant school opens in 1920, then all the individuals in that cohort group would be exposed and $School_{1910-14,i} = 1$, if the school opened in 1921, individuals aged 10 in 1920 would be untreated while younger ones fully treated, meaning $School_{1910-14,i} = 4/5$.

and affecting investments on children. In particular, there should be no changes in mortality or migration for the cohorts affected. Moreover, the timing of the opening of associations and schools should be uncorrelated with the expected impact, that is assumed to be constant across municipalities. To examine the validity of these assumptions I will estimate equation (2) treating different combinations of cohorts as exposed and I will implement a more flexible dynamic event study (Abraham and Sun, 2018). Note that an important caveat is that migration flows may influence the delay between the foundation of an association and the opening of a school, however, there is no simple way of tackling this issue as migration is a necessary condition for migrant's investments.

5 Results

5.1 Descriptive evidence: Evolution of literacy rates and the share of missing men

This section describes briefly the trends in missing men and literacy rates between 1860 and 1940. During the second half of nineteenth century, emigration affected mostly western coastal municipalities, which saw their share of missing men reach between 10% to 20% by 1900. During the first two decades of the twentieth century, the share of missing men raised significantly in the whole region, especially inland and in northern coastal municipalities. By 1920, the typical Galician municipality had over 18% of men were missing the day of the census. The decade of the 1920s saw a slight decrease in the share of missing men in most municipalities and with the stop of migrant flows during the 1930s the the share of missing men came back to levels of 1900.[29]

Literacy rates experienced a rapid increase between 1860 and 1940, especially after 1910 (see Figure 9). The initial gap between men and women literacy rates narrowed over time and almost disappeared by 1940. During the nineteenth century, literacy rates were not significantly different across municipalities that would experience higher/lower emigration later on. Yet, after 1900, a gap starts to emerge with high emigration areas seeing a faster increase in literacy rates. Figure 10 shows the evolution of the gap in literacy rates between municipalities above and below the average share of missing men (1900-1930). The gap peaks in 1920 at 2.5 percentage points and falls afterward. The similar evolution of the share of missing men and the gap in literacy rates between areas differently exposed motivates a deeper investigation of the potential relationship between emigration and human capital accumulation.

[29] Figure 8 in the Appendix describes the spatial distribution of missing men in each census year.

5.2 Explaining the share of missing men, 1900-1930

Before examining the short and medium-run effects of emigration on literacy rates, in this section I discuss the main determinants of the share of missing men.

Panel Analysis. Table 2 shows the results of estimating equation (3), a panel regression of the share of missing men in year t on different lagged covariates. All the columns include municipality and year fixed effects capturing the influence of time-invariant local factors and common trends across all municipalities. The table displays beta coefficients to make interpretation easier. In column (1), the share of missing men is regressed on the main instruments, namely economic growth in the average destination and its interaction with the adult sex ratio in 1860. At the average level of the adult sex ratio (i.e., setting the interaction effect to zero), an increase in GDP per capita growth at destination in one standard deviation is associated with a raise in the share of missing men in 7% of a standard deviation (e.g., 1 percentage point). Increasing the adult sex ratio in one standard deviation amplifies this effect by more than half (i.e., an additional 4 percent) leading to an increase in the share of missing men of 1.5 percentage points. This effect is non-negligible, especially given that the fixed effects already capture a large part of the variation in the share of missing men. The coefficients are very robust to the inclusion of other lagged controls even the lagged share of missing men (columns 2 to 4). The last column displays the result including all covariates. There is a high degree of persistence in missing men. On average, one standard deviation increase in the lagged share of missing men is associated with a 24% of a standard deviation increase in the contemporary share. Rainfall shocks are negatively related with migration, suggesting that migrants may have faced credit constraints. Province-level factors such as population and literacy rates are also strong drivers of aggregate changes in missing men.

Cross-Sectional Analysis. Table 3 shows the results of estimating equation (6), a cross-section regression of the average share of missing men (1900-1930) on the adult sex ratio (1860) and different pre-migration covariates. The table displays beta coefficients to make interpretation easier. Column (1) shows that the instrument, the adult sex ratio in 1860, is a key predictor of the share of missing men explaining close to 20% of its variation. Columns (2) to (4) add geographic, socio-demographic and economic variables. The coefficient of the adult sex ratio remains very robust across the different specifications. On average, an increase in the adult sex ratio (1860) in one standard deviation is associated with a raise in the average share of missing men (1900-1930)

in 30% of a standard deviation (i.e., a 2.8-3.5 percentage points relative to a mean of 14 percent). Column (5) includes all the covariates at once. The results show that bigger municipalities and those far away from a port had significantly lower shares of missing men during 1900-1930. Distance to the port can be interpreted as reflecting transportation costs. Regarding initial literacy rates (1860), there is a positive correlation in the case of men but a negative one in the case of women. These findings suggest that men literacy rates facilitated migration, either by allowing high educated individuals to migrate or by contributing to the diffusion of information (as hypothesized by Sánchez-Alonso, 2000). Land inequality and local taxes are negatively associated with the share of missing men (although the latter is not statistically significant), supporting the idea that migrants may have been credit constrained. Regarding transportation infrastructure, access to a train network and to the sea are both strongly and negatively associated with the share of missing men but imprecisely estimated, potentially given that the regression already controls for distance to the port and to province capitals. Although the role of population density has been stressed in other works (Eiras Roel, 1989; Vázquez-González, 1993), I do not find any evidence that population pressure in 1860 influenced the extent of missing men later on.

Robustness Checks: The significance of the instrument and the point estimates remain very robust across different specifications in the panel and cross-sectional analysis. Table A3 shows the relationship between each instrument and the share of missing men in different years.

5.3 Short and medium-run effects of emigration on literacy rates, 1900-1930

In this section, I describe the short and medium-run effects of mass emigration on literacy rates. I rely on a decadal panel of municipalitites between 1900-1930 and instrument the share of missing men (in a given year t) by the economic growth at migrants' destination in the precedent decade (between t-10 and t) and its interaction with the adult sex ratio in 1860.

Main Result Graphically. I first plot the coefficients from estimating variants of equation (1) by 2SLS with different lags in the share of missing men or literacy rates. As shown in Figure 11, the estimates of t-10 and t-20 suggest that the share of missing men in a given year has not a strong impact on the literacy rates one or two decades before. This exercise serves as a placebo test of the instrumental variables strategy. The estimate of t shows that the contemporary effect of missing men on literacy rates is large and negative implying a reduction in literacy rates in the short run.

In the medium run the impact of missing men is positive and significant implying a net gain in literacy rates one decade later (t+10) which persists two decades after the migrant shock (t+20).

Main Result in Detail. Table 4 summarizes the results on the short and medium-run impact of missing men on literacy rates (i.e., OLS and 2SLS estimates of equations 1 and 2). The coefficient in column (1) shows that taking into account municipality and year fixed effects there is no correlation between the share of missing men and literacy rates. Using the share of missing men in the previous census (10 years before) yields a small positive and significant correlation as shown in column (2). The 2SLS estimates suggest that on average one standard deviation increase in the share of missing men in year t decreases literacy rates in t by 51% of a standard deviation but increase literacy rates 10 year later by around 35% of a standard deviation. The 95% confident intervals around this estimate imply that a migration shock raising the share of missing men from 11% to 20% could increase literacy rates one decade later by around 0-6 percentage points (i.e., an increase of 0%-17%).

Note that there is a substantial gap between 2SLS and OLS estimates, although the coefficients are not statistically different from each other. The short-run OLS estimate is upward biased, which could be the result of positive shocks affecting simultaneously literacy and emigration. There is recent evidence documenting that economic growth can indeed foster emigration (Clemens and Mendola, 2020). Similarly, there could be reverse causation to the extent that increases in literacy rates facilitate the diffusion of information about the option to emigrate. Sanzhez-Alonso (1995) examines this hypothesis using data across Spanish provinces at the beginning of the twentieth century and finds that higher literacy rates are associated with more emigration. Notice that these two sources of bias would tend to disappear in the medium run to the extent that other areas experience similar positive shocks. Interestingly, the medium-run OLS estimate is downward biased. One potential explanation is that, despite an eventual short-run positive shock, areas experiencing higher emigration are generally on less favorable development paths with gaps that tend to widen over longer horizons. It is crucial to point out that both the short-run and medium-run OLS estimates are biased toward zero. This could be the result of measurement error in the the share of missing men. As discussed in the Data section, the share of missing men is a valid but imperfect proxy for emigration and the instrumental variable strategy would tackle this attenuation bias.

Gender Differences. The impact of emigration on literacy rates differs by gender as shown in

Figure 12. While there is a short-run negative effect for both men and women, the medium-run gains are only present for men. In the short-run, an increase in the share of missing men in one standard deviation (i.e., 9 percentage points) leads to a decrease in male and female literacy rates of 33 and 55% of a standard deviation respectively. One decade later, male literacy rates are 66% of a standard deviation higher (i.e., around 6 percentage points) while no significant effect is found for women. The fact that the coefficient for women is close to zero suggests that the short-run negative effects dissipate but there are no gain arising in the medium run. Although these findings are consistent with the idea that families may invest more in the education of individuals who have higher returns to schooling (Vanden Eynde, 2016; Shrestha, 2017), in the Mechanism section I provide suggestive evidence that this is not the case, as emigration contributed to equally higher investments on both boys and girls' education. The more plausible explanation for the gender disparities in the medium run might be differences in the degree of return migration and selection into return.

Heterogeneous Results. I explore the potential heterogeneity of the short and medium-run effects depending on migrants' destinations (1900-1930), population size and initial literacy rates before the mass emigration. The results are summarized in Figure A13. First, I divide the sample between municipalities with a share of migrants to Rio de la Plata above and below the median. The results suggesting that at least in the medium run, migrants' destination did not influence the effects on literacy rates. I then divide the sample between municipalities with a population above (High population) and below (Low population) the median in 1860. In the case of men, the negative short-run effects are entirely driven by areas that had relatively higher population. In the medium run, however, the gains in literacy rates are of similar magnitude everywhere. In the case of women, the estimates tend to be larger in municipalities with higher population in 1860 but the coefficients are not statistically different from one another. Finally, I divide the sample between municipalities with literacy rates above (High literacy) and below (Low literacy) the median in 1860 (gender-specific). For men, the short-run and medium-run effect are larger and only significant in municipalities with relatively higher male literacy rates in 1860. The coefficients however are not statistically different. In the case of women, the differences are smaller, potentially because in 1860 female literacy rates were extremely low in all municipalities.

Robustness Checks. I perform a number of tests to assess the robustness of my main results

and address potential threats to the identification:

Alternative instruments. The main results remain very similar by using alternative pull factors constructed using district weights for the period 1900-1930 or province weights for the years 1885-1895 as shown in Figure A14. Being able to replicate the findings using weights before 1900 shows that the selection of migrants to certain countries is not confounding the results. The estimates are also very similar when using the walking distance to the port or directly as network instrument .

International trade. One potential concern is that there may be a relationship between economic growth at destination and Galician trade with that country. Controlling for trade with destination countries, its interaction with the walking distance to the port (columns 1 and 2 of Table A8), or excluding municipalities closer to the ports (columns 1 and 2 of Figure A15) has no effect on the main results. Note that Latin America represented a small fraction of Spanish exports and imports, and Galicia was not particularly open to international trade in 1900-1930.

Remittances and economic shocks. Given that I exploit an instrument that captures the size of the network, one may worry that the effects of remittances get amplified and therefore influence the results on literacy rates. To partially deal with this concern, I add different measures of economic activity at origin (e.g., GDP per capita in the province, GDP per capita growth in the province, interaction between GDP per capita in the province interacted with distance to the closest port, etc.). I also introduce measures of rainfall shocks at origin that by changing the level of remittances (or their use) could alter the impact of emigration on literacy rates. Overall, including these variables has no effect on the estimated coefficients as shown in Table A8.

Extensive controls. To make sure that none of the potential (observable) confounders of the adult sex ratio (1860) is mediating the results, I introduce the pre-migration covariates (1855-1860) interacted with time dummies. The results show that the instrument is unaffected by the inclusion of these variables and that the main findings remain robust.

Sample restrictions. I analyze whether the estimates are sensitive to outliers by excluding municipalities with extreme values of population, missing men, literacy rates or changes in literacy rates (see Figure A15). I also exclude provinces one by one to check if the effects differ geographically. The coefficients remain very similar across the different specifications.

Missing women. I replicate the analysis replacing the share of missing men with the share of missing population (i.e., men and women absent the day of the census) and the share missing women alone. The results are displayed in table Table A9. As shown in columns (1) to (6), using the share of missing population produces very similar estimates of the short and medium-run

effects. The strength of the first stage is however significantly lower due to the noise introduced by incorporating missing women into the explanatory variable. Columns (7) and (8) look at the impact of the share of missing women on women's literacy rates but in this case the instruments are not fitted for the specification, producing an F-stat of 1.28 and 4.35 respectively.

5.4 Long-run effects of emigration on human capital, 1930s-2000s

In this section, I describe the long-run effects of mass emigration on different measures of human capital between 1930 and 2011. I estimate equation (5) with a cross-section of all Galician municipalities instrumenting the average share of missing men (1900-1930) with the adult sex ratio in 1860 (i.e., a proxy for pioneer emigration).

MAIN RESULT. Table 5 summarizes the results on the long-run effects of emigration showing that it contributed to gains in average levels of human capital that have persisted until today. Focusing on 2011, the year of the last census, the estimate from column (4) suggests that one standard deviation increase in the average share of missing men (1900-1930) led to a higher share of individuals with completed secondary education by 46% of a standard deviation (i.e., 6 percentage points relative to a mean of 59 percent). Similar to the results on the medium-run effect of missing men in the panel analysis, the OLS coefficients of the long-run effects are downward biased and less precisely estimated than 2SLS. This result supports the idea that areas that experienced more emigration (on aggregate) had characteristics that lower their economic potential and limited human capital accumulation over the long run. As discussed in section 5.2, the average share of missing men (1900-1930) is positively associated with more isolated areas (e.g., further away from a province capital, less likely to have a train connection, higher average altitude, etc.) and prone to greater rainfall variability which could have had a negative and persistent influence on educational investments. [30]

I further explore the impact on different levels of education for the cohorts born between 1930 and 1980 using data from the 2001 population census. As shown in Table A10, mass emigration seem to have affected human capital accumulation from the bottom of the distribution up. First, increasing literacy rates in the period while emigration was taking place (1900-1930), then raising primary completion rates for cohorts born in the 1930s and thereafter, secondary completion rates

[30] Controlling for these factors in the OLS regression reduces the downward bias slightly but the coefficients remain quite small, pointing to the relevance of attenuation bias.

for those born in the 1930s and 1940s, and finally tertiary education rates for those born in the 1940s and 1950s. Interestingly, there is a zero impact of missing men on tertiary completion for individuals born in the 1960s and a large negative effect among those born in the 1970s. These findings may suggest that emigration could have fostered these generations to attend college and, as a consequence, to temporarily leave their municipalities of origin and decrease the overall share of individuals with completed tertiary education. Overall, my results contrast remarkably with the findings of Testa (2020), which reports a negative long-run effect of forced migration on human capital at origin six decades later. Yet, as mentioned previously, episodes of forced migration and voluntary migration are not comparable due to the particularities of the former.

Gender Differences. Figure 13 shows that the long-run effects of emigration differ by gender but converge over time. While in 1981, there is no evidence of a significant impact for women, the share of missing men (1900-1930) positively affected secondary completion among men. These results are in line with the findings for the historical period (1900-1930), where men experienced gains in literacy rates in the medium run while women did not. In 1991, the coefficient for women is positive and marginally significant and in 2001 and 2011 becomes much larger in size and highly significant. Importantly, if one restricts the sample to women with less than secondary (illiterate, no studies completed, primary completed), there is a positive effect in 1981 for the share who had completed at least primary suggesting again that emigration raised educational attainment bottom-up and with a lag for women writh respect to men.

Robustness Checks. I perform a number of checks to assess the robustness of the long-run analysis. The results are summarized in Figure A16. The baseline regression controls for population (1860), men and female literacy rates (1860), wealth per capita (1855) and local taxes per capita (1855). The baseline point estimate is represented by the horizontal line. I then control for population growth between 1860 and 1900, male and female literacy rates growth between 1860 and 1900, income shocks between 1900 and 1930 (rainfall variability), land inequality (1860), and educational measures (enrollment and school supply in 1908). The estimated coefficient remains very robust to all these additional controls. Next, I control for covariates capturing spatial variation such as latitude and longitude as well as other variables related to the size of municipalities and population dispersion. Although the point estimate falls slightly when controlling for latitude and municipality's size, it remains positive and significant, and it is not significantly different from the

baseline estimate. Last, I perform a number of sample exclusions to ensure that my results are not driven by outliers or only one province. I exclude municipalities with a walking distance to the closest port below 20km, with extreme levels of population (1900), literacy rates (1900) or missing men (1900-1930), and exclude provinces one by one. When excluding the province of Coruña the point estimate increases slightly because the effect is smaller in this province and it accounts for almost one third of all municipalities. Overall, the results remain fairly robust to all these sample restrictions.

OTHER OUTCOMES. I also explore the potential long-run effects of mass emigration on other educational outcomes. In particular, I look at the number of public and private schools, public libraries, and local associations per thousand inhabitants in 2018. As displayed in Table A11, the average share of missing men (1900-1930) did not seem to have increased the supply of public or private schools. These results suggest that despite large investments in schools made by migrants, in the long run the State may have tried to compensate other areas. It is impossible to determine, however, the extent to which migrants might have crowded-out the State. In terms of the availability of public libraries, I find no significant difference between areas that experience more or less migration despite anecdotal evidence that migrants financed the opening of several libraries. Using the registry of all Galician local associations (approximately 20,000), I find that the average share of missing men (1900-1930) reduced the number of local associations per inhabitant, probably due to its effects on population density. However, when looking at the share of associations having an educational goal, I find a large positive impact. This higher prevalence of educational local associations might be the legacy of the hundreds of migrants' associations with that aim that were created one century ago. I discuss this aspect further in the next section.

6 Mechanisms

6.1 Selection into emigration, return migration and incentives

In the previous section, I show that the emigration has a short-run negative impact on literacy of both males and females but just temporary. One decade later emigration leads to net gains in male literacy rates while female literacy rates recovered their baseline level. In this section, I discuss some mechanisms that can explain these results.

Age composition of migrants. The short-run negative effect can be partially explained by the departure of young individuals and the fact that literacy rates in the population decreased with age (especially among women). According to the statistics compiled by Vázquez-González (1999), around 54% of migrants were younger than 25 years old (Table A1). Moreover, at any given point in time, literacy rates were higher among younger individuals (Table A2). For instance, in 1900 close to 20% of women in their 20s were literate compared to 13% in their 40s. By 1930, literacy rates were 62% and 41% among women in their 20s and 40s respectively. As shown in Table A2, this age differential was pronounced among women but smaller among men.

Historical evidence on selection. The degree of selection in literacy among migrants could amplify or soften this age composition effect, however, the scarce evidence available points to no selection conditional on age. The (aggregate) migration statistics suggest that Galician (as well as Spanish) migrants were positively selected but the pattern changes once age is taken into account. For instance, the detailed information compiled by Vázquez-González (1999) shows that 87% of male migrants in the 1920s were literate, compared to 1930 men's literacy rates of 61% on average and 81% among young men (i.e., in their 20s). Table A1 gathers the only available information about Galician migrants' literacy coming from the Official Migration Statistics and the data of Vázquez-González (1999). Official Migration Statistics may over-estimate the degree of literacy since embarkation lists often classified individuals knowing how to sign as literate. An additional reason is that literacy rates would probably be lower among clandestine migrants. Despite the caveats of official statistics, they seem to provide strong evidence that Galician migrants were not negatively selected. We lack information about literacy among returnees, yet the data for Spain shows that migrants and returnees had similar literacy rates suggesting no selection into return migration for the country as a whole.

Descriptive evidence with data by cohort. To explore more in depth the role played by selection into migration, return migration and investments into children's education, I exploit data by cohort from the population censuses. This information is only available at the province level (in every census year) and for the 23 largest municipalities in 1900 and 1910.[31]

The census data for Galicia shows that the size of young cohorts experienced a large decrease

[31] Despite this caveat, the distribution of the share of missing men in this sub-sample is very similar to the full sample.

during the mass emigration era. As shown in Figure A18, in 1900 there were around 100 thousand men aged 11 to 15 in Galicia, 20 years later the size of this cohort had decreased to 41 thousand (i.e., a 55% decline). In the case of women, the size of the cohort goes from 100 to 67 thousand (i.e., a 33% decline). Comparing Galicia with the rest of Spain, we observe that between 1920 and 1940 the size of the cohort barely falls in Galicia (7% drop) compared to the rest of the country (20% drop).[32] This points to the compensating effect of return migrants. Note that according to the official statistics as many as 60% of Galician migrants eventually return, a figure similar to that from other Southern European countries (Bandiera et al., 2013). The roe of return migration is a crucial difference from other settings of forced migration in which individuals are not allowed to come back to their communities of origin (e.g., Testa, 2020).

The same pattern emerges using the data across districts between 1900 and 1910. There is a strong negative relationship between the reduction in size of the cohort born between 1885 and 1889 and the increase in the share of missing men in 1900-1910 (Figure A19). Similarly, changes in average literacy rates in this cohort are also correlated with changes in missing men, suggesting that young individuals were investing more in education in high migration areas (Figure A20). Although this descriptive evidence is only suggestive, it goes in line with the historical evidence highlighting that many migrants invested in education before departing (Vázquez-González, 1999; Gabriel, 2013) and with other with other studies showing that remittances and the option to emigrate can increase parental investments on children's educatino (Dinkelman & Mariotti, 2016). Moreover, the fact that there is an increase in both boys and girls literacy suggets that reason behind the null medium-run effect on women's education is not a lack of investments on them (Shresta, 2008), but rather related to the role of return migration and selection into return.

6.2 Migrants' associations and schools, 1900-1940

Background. In clear contrast with episodes of forced migration (Becker and Ferrara, 2019), Galician migrants maintained strong links with their communities of origin and founded hundreds of associations with the aim of financing investments in their hometowns. Most of these associations had an educational goal, such as purchasing materials or funding the construction of new schools in their hometowns. Migrants believed that these institutions were the best way in which they could contribute to the development of Galicia in the long run (q. 7, 8, 10, 19, 20, 21). The schools constructed were of outstanding quality for the standards of the time: large, bright, fully

[32] The evolution of the cohort aged 6 to 10 in 1900 displays a very similar pattern during these decades.

equipped, etc. Migrants even stressed the importance of hygiene and vaccination (q. 16, 24) and offered free resources to poor students (q. 28, 29).[33] Moreover, these schools provided high quality education manifested by novel teaching practices (q. 30, 31) and a different curriculum (e.g., no religious education, courses on geography, accounting, farming). Some critics argued that migrants' schools fostered even more migration and that children were not educated into the Galician tradition (q. 32, 33, 34, 35). The predominant view however seems to be that of praising these investments and aspiring to have migrants' schools in every single municipality (q. 7, 22, 23). As shown in Figure A22, most associations were founded between 1910 and and 1924 while most schools opened between 1920 and 1934. Figure A23 and Figure A24 show the location of these schools over time and the strong overlap between these investments and the extent of missing.

Foundation of migrants' associations and schools. To better understand the determinants behind the foundation of migrants' associations and schools, I perform cross-sectional regressions with key historical variables. As in the long-run analysis, I instrument the share of missing men (1900-1930) with the adult sex ratio in 1860. Table 6 summarizes the results. Column (1) shows that most variables are not significant predictors of the existence of at least one association. Interestingly, the share of missing is not a relevant factor at this margin. Both higher taxes per capita (1855) and the availability of more schools (1875) reduce the likelihood that one association was created in the municipality. One of the reasons why it may be so difficult to predict the opening of an association is because some of them originated in a municipality but operated in several, and hence, have been assigned to all those locations. In Column (2), the dependent variable is the number of associations. The share of missing men (1900-1930) seems to be a strong determinant of migrants' associations, although the coefficient is not significant (p-value=0.13). An increase in the share of missing men (1900-1930) in on standard deviation translates into one more migrant associations on average. Larger municipalities and those closer to a transatlantic port also had significantly more migrants' associations while municipalities with higher taxes per capita (1855) and more schools (1875) significantly fewer. An increase in children's enrollment rates (1875) in one standard deviation is associated with an increase of 0.3 units in the number of associations. In columns (3) and (4) the dependent variable is the number of schools founded by migrants. The last column includes the number of migrants' associations as an additional variable. Focusing on column (4), the major determinants of the number of migrants' schools are the share of missing

[33] Figure A21 shows some of them.

men (1900-1930), the number of migrants' associations, local shocks (i.e., rainfall variability) and literacy rates. Even after controlling for the number of associations, an increase in the share of missing men (1900-1930) in one standard deviation leads to 1.6 more schools on average. This results suggest that the size of the migrant community abroad was crucial for the financing of new schools. Importantly, economic forces also played a role. Land inequality (measured by the share of landless peasants) and local economic shocks (rainfall variability) are negatively correlated with the number of new schools. For instance, one standard deviation increase in rainfall variability is associated with 0.75 less schools on average. The large and negative coefficient of rainfall variability may indicate that migrants faced a trade-off between individual and collective remittances. If families were in need of resources to face these shocks, it would have been more difficult that migrants could collect money to finance new interventions.

Impact of migrants' associations and schools. Table 7 presents the results of the cohort analysis. In columns (1) to (3) the dependent variable is the share of individuals with completed primary education while in columns (4) to (6) the ratio between the size of cohort c and $c-1$. Column (1) shows that the opening of an association does not seem to have any impact on children's primary achievement. Yet, this effect may be confounded by the fact that most associations financed a new school 5 to 10 years later. A new migrant school, in fact, seems to increase primary completion among the children exposed (column 2). When both variables are considered together, as in column (3), we see that foundation of associations does have a negative effect on educational attainment while the opening of a school a positive one. The point estimate of column (3) implies that a new school increases the primary completion rate by 1 percentage point among children younger than 10. This effect is modest (i.e., 4% of a standard deviation) but one needs to consider that often several schools opened in the same municipality.

Looking at the size of the cohorts (columns 4-6), a similar pattern emerges. The foundation of a migrant association reduces the size of young cohorts while the opening of a school has the reverse effect. As shown in column (6), the creation of an association is associated with a fall of 3.3 percentage points in the size of the cohort of children younger than 10, although the effect is imprecisely estimated. A new migrant school, however, increases the size of the cohort by 1.7 percentage points (i.e., 3% of a standard deviation). Note that this effect may be somewhat mechanical if the foundation of an association is preceded by a large migration flow that in turn reduces fertility. However, there are significant differences between men and women what goes

against this hypothesis.

The heterogeneous effects by gender may shed some light on these results. I find that the positive impact of a new migrant school on educational attainment is very similar for both men and women. The negative effect of the foundation of an association however is considerably larger for women (the point estimate is twice as large although not statistically different). This result is consistent with the hypothesis that migrants' association may have encouraged family migration (and/or reunification) therefore affecting more women than men.

6.3 Beliefs about education and effort

Background. The work of historians suggests that Galician migrants realized the value of education once at Latin America and tried to transmit these norms (Peña Saavedra, 1983; De Gabriel, 2012). There is plenty of anecdotal evidence in this direction, some of which is shown in Section G. Migrants acknowledged their lack of education and the importance of it (q. 1, 2, 3), not just in the face of migration (q. 3, 4, 6, 8) but for the development of Galicia (q. 2, 3, 5, 7, 8). The diffusion of these values for education would take place through migrants' associations and their investments (q. 7, 8, 10, 19, 20 , 21, 22, 23, 24), migrants' missives, magazines and other publications (q. 9, 13, 14) as well as with the return of migrants and the transmission from parents to children (q. 11, 12, 13). Migrants not only encouraged directing funds towards educational resources (q. 18), but stressed the importance of regularly attending school (q. 12, 14, 15) and urged public authorities to enforce compulsory education laws (q. 16, 17). Moreover, migrants worked extremely hard at destination, saving most of their earnings, what could have also contributed to change their mindset about the value of effort (q. 26, 25, 27).

Empirical analysis. To test whether emigration could have led to a change in values about the importance of education, I combine my historical data with contemporary survey data based on the municipality of respondents. I use several waves of the Encuestas Opinión Pública where individuals report in a scale from 0 to 10 whether success in life is due to one's education/effort (rather than luck/contacts). I estimate the following equation by 2SLS:

$$V_{ij} = \alpha + \gamma \overline{ShMissing_i} + X_i\Pi + Z_j\Psi + u_{ij} \quad (8)$$

where V_{ij} is the value individual i (in municipality j) puts on education and effort, $\overline{ShMissing_i}$ is the average share of missing men (1900-1930) in municipality j, X_i and Z_j stand for individual

and municipality level controls controls and u_{ij} represents standard errors that I cluster at the municipality level. I instrument $\overline{ShMissing_i}$ with the proxy for pioneer migrant networks (i.e., adult sex ratio in 1860).

The results, displayed in Table 8, show that individuals living in municipalities that experienced higher historical emigration put greater value on education and effort. On average, one standard deviation increase in the share of missing men (1900-1930) is associated with a 22% standard deviation increase in the reported valuation (e.g., around 0.5 points in a 0-10 scale with an average of 5). The results are very robust to the inclusion of individual and historical controls.[34] I do not find any impact on other related dimensions such as their satisfaction with public education, their willingness to pay more taxes in exchange of more public goods, nor their sense of how much society benefits from taxes (Table A12). I do find, however, a small reduction in the share of people declaring that we spend too little on education or that we pay too little taxes.

7 Conclusion

This paper examines the impact of mass emigration on human capital accumulation at origin over a century. I show that the Galician mass emigration reduced literacy rates in the short run but its impact became positive after one decade and led to gains in human capital that persist even one hundred years later. I provide empirical evidence that migrants' associations and the schools they financed in their hometowns fostered human capital formation and that migration contributed to shaping social norms at origin about the value of education and effort.

My findings have important implications for both research on migration and public policy. First, they highlight the relevance of taking a long-term view, as the impact of migration and the individuals affected may change over time. For instance, while in the first few decades emigration only affect men's educational attainment, in the long run both men and women benefited similarly. They also stress the need to consider other mechanisms beyond monetary remittances, such as migrants' financing of public goods in their hometowns and social remittances. My results suggest that fostering migrants engagement with their communities of origin (e.g., through local associations, information campaigns, fundraising) can be an effective way to promote human capital accumulation and economic development at origin. Innovative public policies could be implemented to strengthen these links and encourage the transmission of collective remittances

[34] One would expect that controlling for individuals' education the effect of the share of missing men (1900-1930) should decrease. One reason why it may not be the case is that educational attainment is measured in categories and potentially with error (e.g., attended vs. completed studies).

targeting specific goals. Finally, my findings shed new light on the impact of migration on the adoption of new social norms, a channel that may lead to persistent effects in the long run. Future work should examine in more depth the interplay between these different mechanisms, and explore the long-run impact of emigration on other social and economic aspects in the communities of origin.

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* 102:1832–56.

ANDERSSON, D., MOUNIR, K., AND PRAWITZ, E. 2017. Mass migration, cheap labor, and innovation. *working paper* .

ANDREWS, I., STOCK, J. H., AND SUN, L. 2019. Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics* 11:727–753.

ANTMAN, F. M. 2012. Gender, educational attainment, and the impact of parental migration on children left behind. *Journal of Population Economics* 25:1187–1214.

BADA, X. 2014. Mexican hometown associations in Chicagoacán: From local to transnational civic engagement. Rutgers University Press.

BATISTA, C., LACUESTA, A., AND VICENTE, P. C. 2012. Testing the 'brain gain'hypothesis: Micro evidence from cape verde. *Journal of Development Economics* 97:32–45.

BAZZI, S., FISZBEIN, M., AND GEBRESILASSE, M. forthcoming. Frontier culture: The roots and persistence of" rugged individualism" in the united states. Technical report.

BECKER, S. O. AND FERRARA, A. 2019. Consequences of forced migration: A survey of recent findings. *Labour Economics* 59:1–16.

BECKER, S. O., GROSFELD, I., GROSJEAN, P., VOIGTLANDER, N., AND ZHURAVSKAYA, E. 2020. Forced migration and human capital: evidence from post-wwii population transfers. *American Economic Review* 110:1430–63.

BERTRAND, J. R. 1999. La emigración gallega,¿ sistema original? *SEMATA, Ciencias Sociais e Humanidades* 11:439–455.

BOLT, J., INKLAAR, R., DE JONG, H., AND VAN ZANDEN, J. Maddison project database, version 2018. *Rebasing 'Maddison': new income comparisons and the shape of long-run economic development* .

BRUM, M. 2019. Italian migration to the united states: The role of pioneers' locations. *working paper* .

CALVI, R., MANTOVANELLI, F., AND HOEHN-VELASCO, L. forthcoming. The protestant legacy: Missions and human capital in india. *Journal of Human Resources* .

CAMPOS ALVAREZ, J. R. 1993. La emigración gallega a américa, 1880-1930. integración y retorno. *Minius* pp. 133–145.

CHAUVET, L., GUBERT, F., MERCIER, M., AND MESPLÉ-SOMPS, S. 2015. Migrants' home town associations and local development in mali. *The Scandinavian Journal of Economics* 117:686–722.

CHRISTIAN, P. AND BARRETT, C. B. 2017. Revisiting the effect of food aid on conflict: A methodological caution. The World Bank.

CLEMENS, M. A. AND MENDOLA, M. 2020. Migration from developing countries: Selection, income elasticity, and simpson's paradox. *IZA Discussion Paper* .

CONNOR, D. S. 2019. The cream of the crop? geography, networks, and irish migrant selection in the age of mass migration. *The Journal of Economic History* 79:139–175.

DÍEZ-MINGUELA, A., MARTINEZ-GALARRAGA, J., AND TIRADO-FABREGAT, D. A. 2016. Why did spanish regions not converge before the civil war? agglomeration economies and (regional) growth revisited. *Revista de Historia Economica-Journal of Iberian and Latin American Economic History* 34:417–448.

DINKELMAN, T. AND MARIOTTI, M. 2016. The long-run effects of labor migration on human capital formation in communities of origin. *American Economic Journal: Applied Economics* 8:1–35.

DOCQUIER, F. AND RAPOPORT, H. 2012. Globalization, brain drain, and development. *Journal of Economic Literature* 50:681–730.

DROLLER, F. 2018. Migration, population composition and long run economic development: Evidence from settlements in the pampas. *The Economic Journal* 128:2321–2352.

DUFLO, E. 2001. Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment. *American economic review* 91:795–813.

DUSTMANN, C. AND WEISS, Y. 2007. Return migration: theory and empirical evidence from the uk. *British Journal of Industrial Relations* 45:236–256.

EIRAS ROEL, A. 1991. La emigración gallega a américa: panorama general. *In* La emigración española a Ultramar, 1492-1914, pp. 17–40. Asociación de Historia Moderna.

EYNDE, O. V. 2016. Military service and human capital accumulation: Evidence from colonial punjab. *Journal of Human Resources* 51:1003–1035.

FARÍAS, R. 2010. La inmigración gallega en el sur del gran buenos aires:(1869-1960).

FERNÁNDEZ-SANTIAGO, M. X. 2008. Asociacionismo gallego en américa, 1871-1960. *ESPAÑOLA A AMÉRICA* p. 199.

FOX, J. AND BADA, X. 2008. Migrant organization and hometown impacts in rural mexico. *Journal of Agrarian Change* 8:435–461.

GABRIEL, N. D. 2006. Ler e escribir en galicia. a alfabetización dos galegos e das galegas nos séculos xix e xx.

GABRIEL, N. D. 2013. El proceso de alfabetización en galicia: un intento de explicación y comprensión. *Historia de la Educación* 32:289–313.

GOLDSMITH-PINKHAM, P., SORKIN, I., AND SWIFT, H. 2020. Bartik instruments: What, when, why, and how. Technical report.

GROSJEAN, P. AND KHATTAR, R. 2018. It's raining men! hallelujah? the long-run consequences of male-biased sex ratios. *The Review of Economic Studies* .

HATTON, T. J. AND WILLIAMSON, J. G. 1998. The age of mass migration: Causes and economic impact. Oxford University Press on Demand.

HERNÁNDEZ BORGE, J. 1986. Notas sobre la emigración gallega en la segunda mitad del siglo xix.

JAEGER, D. A., RUIST, J., AND STUHLER, J. 2018. Shift-share instruments and the impact of immigration. Technical report, National Bureau of Economic Research.

KARADJA, M. AND PRAWITZ, E. 2019. Exit, voice, and political change: Evidence from swedish mass migration to the united states. *Journal of Political Economy* 127:000–000.

KELLY, M. 2019. The standard errors of persistence. *CEPR Discussion Paper No. DP13783* .

KNUDSEN, A. S. B. 2020. Those who stayed: Individualism, self-selection and cultural change during the age of mass migration. *working paper* .

MCKENZIE, D. AND RAPOPORT, H. 2010. Self-selection patterns in mexico-us migration: the role of migration networks. *The Review of Economics and Statistics* 92:811–821.

MCKENZIE, D. AND RAPOPORT, H. 2011. Can migration reduce educational attainment? evidence from mexico. *Journal of Population Economics* 24:1331–1358.

NÚÑEZ-SEIXAS, X. M. 1994. Migraciones y política. las remesas invisibles. algunas notas sobre la influencia socio-política de la emigración transoceánica en galicia,(1890-1930). *Estudios Migratorios Latinoamericanos* 9:301–346.

NÚÑEZ-SEIXAS, X. M. 1998a. Emigrantes, caciques e indianos: o influxo sociopolítico da emigración transoceánica en Galicia (1900-1930). Xerais de Galicia.

NÚÑEZ-SEIXAS, X. M. 1998b. Retornados e inadaptados: el" americano" gallego, entre mito y realidad (1880-1930). *Revista de Indias* 58:555–593.

NÚÑEZ-SEIXAS, X. M. 2001. La Galicia austral: la inmigración gallega en la Argentina, volume 3. Editorial Biblos.

NÚÑEZ-SEIXAS, X. M. AND VÁZQUEZ-SOUTELO, R. 2005. As cartas do destino: unha familia galega entre dous mundos, 1919-1971. Editorial Galaxia.

PALMÁS, R. 1976. La emigración. *FABRA, Gustavo, et al. Los gallegos. Madrid, Istmo* .

PEÑA-SAAVEDRA, V. 1991. Las escuelas de americanos y las sociedades de instrucción en galicia. *PhD thesis, Universidade de Santiago de Compostela, Santiago de Compostela* 1-4.

PEÑA-SAAVEDRA, V. 1995. As sociedades galegas de instrucción: proxecto educativo e realizacións escolares. *Estudios migratorios* pp. 8–83.

PEÑA-SAAVEDRA, V. 2010. Universidad de Vigo.

RAPOPORT, H., SARDOSCHAU, S., AND SILVE, A. 2020. International migration and cultural convergence. Technical report.

ROCHA, R., FERRAZ, C., AND SOARES, R. R. 2017. Human capital persistence and development. *American Economic Journal: Applied Economics* 9:105–36.

ROSÉS, J. R., MARTÍNEZ-GALARRAGA, J., AND TIRADO, D. A. 2010. The upswing of regional income inequality in spain (1860–1930). *Explorations in Economic History* 47:244–257.

SÁNCHEZ-ALONSO, B. 1995. Las causas de la emigración española, 1880-1930. Alianza Editorial.

SÁNCHEZ-ALONSO, B. 2000. Those who left and those who stayed behind: explaining emigration from the regions of spain, 1880–1914. *The Journal of Economic History* 60:730–755.

SEQUEIRA, S., NUNN, N., AND QIAN, N. 2020. Immigrants and the making of america. *Review of Economic Studies* .

SHRESTHA, S. A. 2017. No man left behind: Effects of emigration prospects on educational and labour outcomes of non-migrants. *The Economic Journal* 127:495–521.

SHRESTHA, S. A. AND PALANISWAMY, N. 2017. Sibling rivalry and gender gap: intrahousehold substitution of male and female educational investments from male migration prospects. *Journal of Population Economics* 30:1355–1380.

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* 134:226–247.

STOLZ, Y. AND BATEN, J. 2012. Brain drain in the age of mass migration: Does relative inequality explain migrant selectivity? *Explorations in Economic History* 49:205–220.

TESTA, P. forthcoming. Migration from developing countries: Selection, income elasticity, and simpson's paradox. *The Economic Journal* .

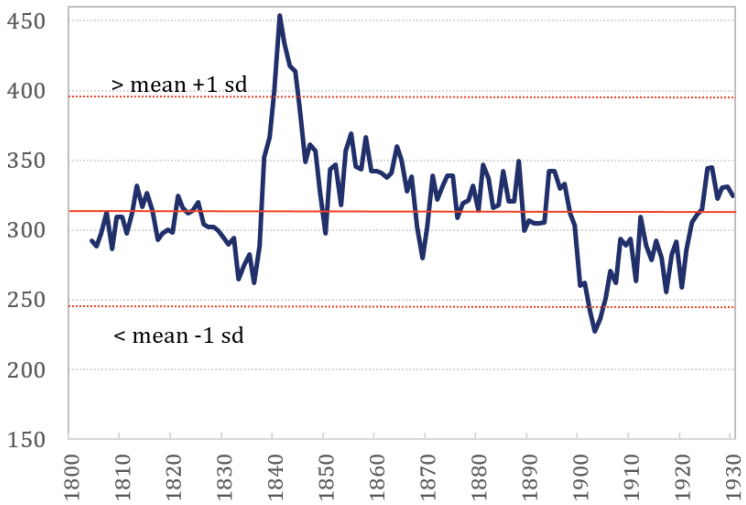
VALENCIA CAICEDO, F. 2019. The mission: Human capital transmission, economic persistence, and culture in south america. *The Quarterly Journal of Economics* 134:507–556.

VÁZQUEZ-GONZÁLEZ, A. 1999. La emigración gallega a américa, 1830-1930. *Memoria de doctorado inédita, Facultade de Ciencias Económicas e Empresariais, Universidade de Santiago de Compostela, Santiago de Compostela* 1-4.

VIANA, V. 2017. La emigración a las Américas por el Puerto de Vilagarcía de Arousa. Teofilo Edicións.

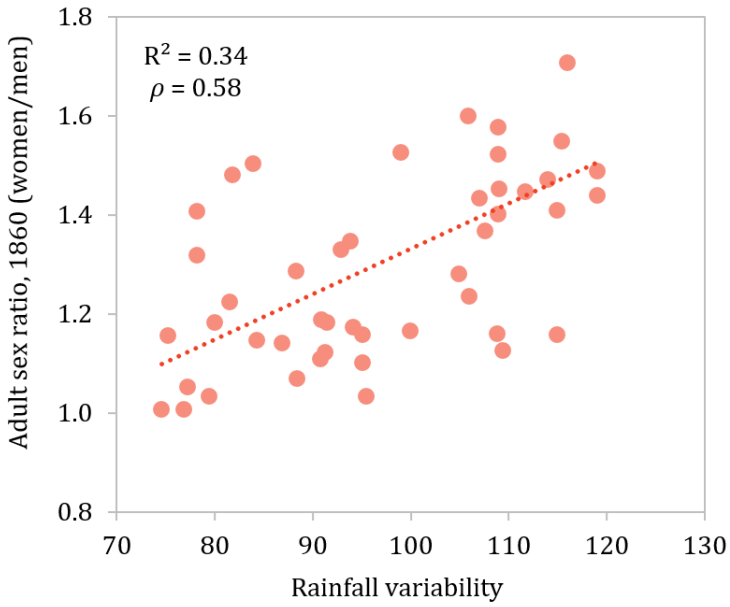
Figures

FIGURE 1: Rainfall in Galicia, 1800-1930



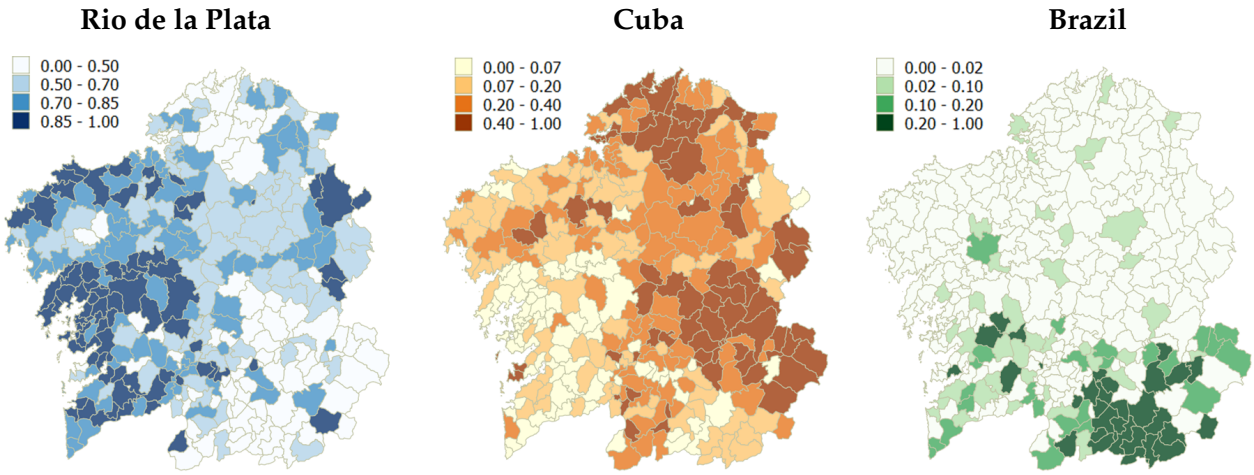
Source: Pauling et al. (2006) European Gridded Seasonal Precipitation Reconstructions.
Note: The figure shows the 5-year moving average of precipitation in Galicia during the Fall season. Return to page 7.

FIGURE 2: Rainfall shocks and adult sex ratio (1860)



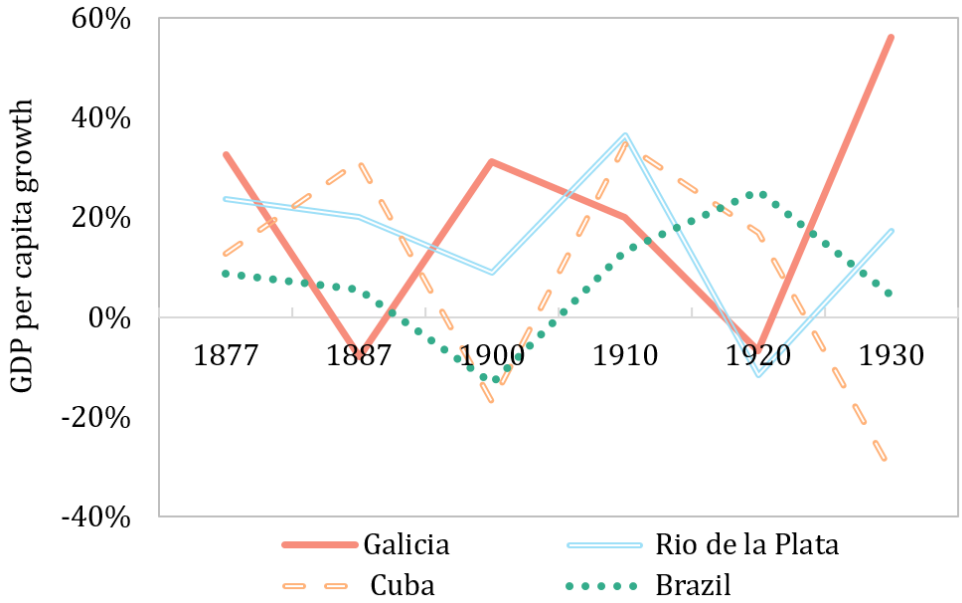
Note: Each dot represents a district. Rainfall information is available at the 0.5°x 0.5°(24 cells in Galicia). The measure of rainfall variability represents the standard deviation of precipitation in the Fall season between 1830 and 1860. Return to page 11.

FIGURE 3: Main migrant destinations across Galician municipalities, 1900-1930



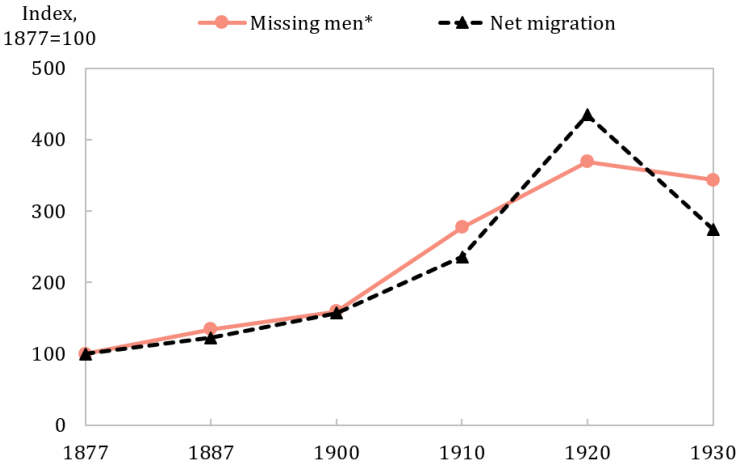
Source: Adapted from the records compiled by Vázquez-González (1999).
Note: The map shows the share of migrants going to each of the three main destinations between 1900-1930. Rio de la Plata stands for Argentine and Uruguay. Return to page 8 or 14.

FIGURE 4: GDP per capita growth in the main destinations, 1877-1930



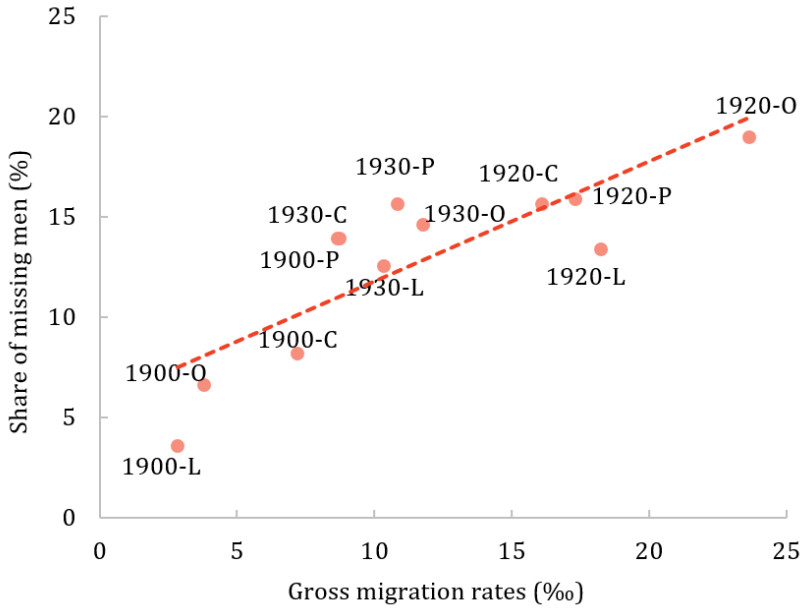
Source: Maddison Project Database 2018 for Latin American countries and Spain, Rosés et al. (2010) for Galicia (relative to Spain). *Note*: The figure displays GDP per capita growth by decade. Return to page 8 or 14 .

FIGURE 5: Emigration and missing men in Galicia, 1877-1930



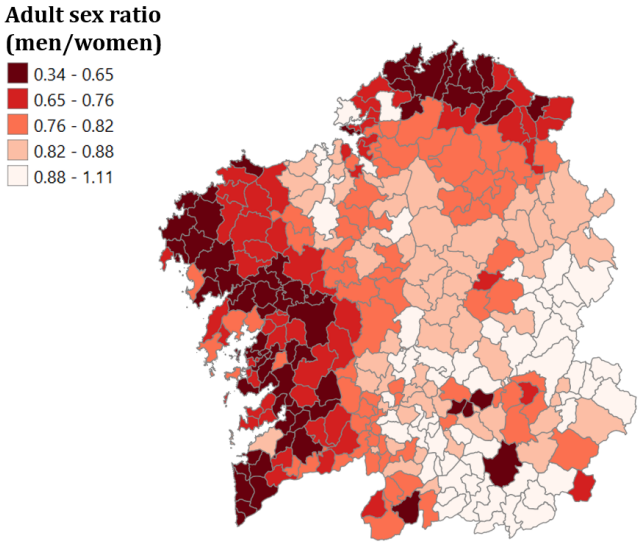
Source: Vázquez-González (1999) and Spanish population census (1877-1930).
Note: "Net migration" stands for sea departures minus arrivals in the last 15 years. "Missing" stands for residents absent the day of the census. Transients were subtracted from the missing to avoid double counting. Return to page 10.

FIGURE 6: Emigration and missing men across Galician provinces, 1900-1920



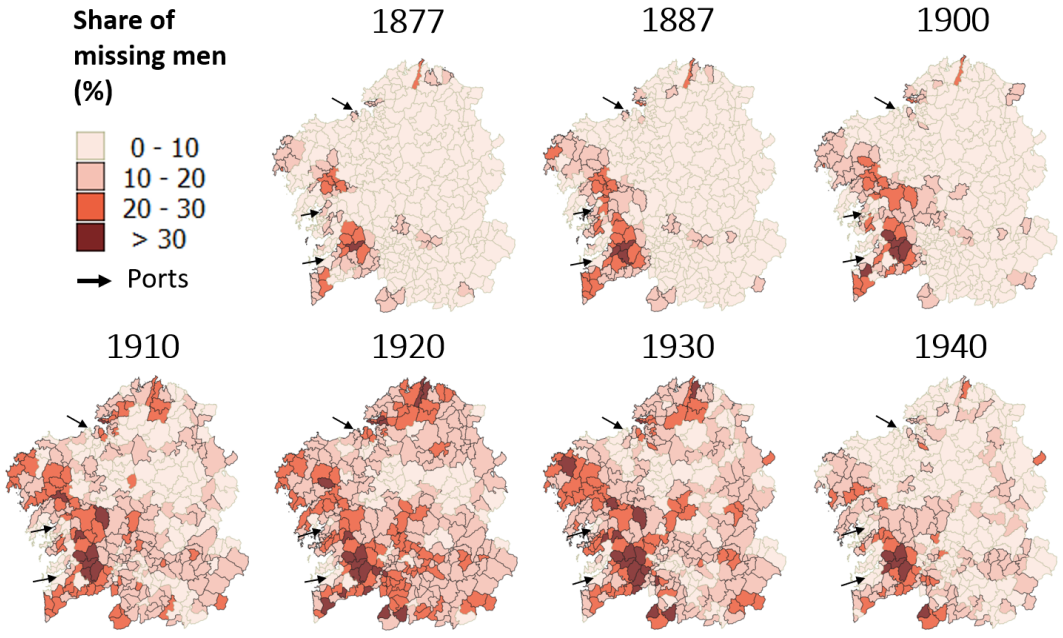
Source: Vázquez-González (1999) and Spanish population census (1900-1920).
Note: "Migration" stands for sea departures. "Missing" stands for residents absent the day of the census. Transients were subtracted from the missing to avoid double counting. C: Coruña, L: Lugo, O: Ourense, P: Pontevedra. Return to page 10.

FIGURE 7: Adult sex ratio in 1860



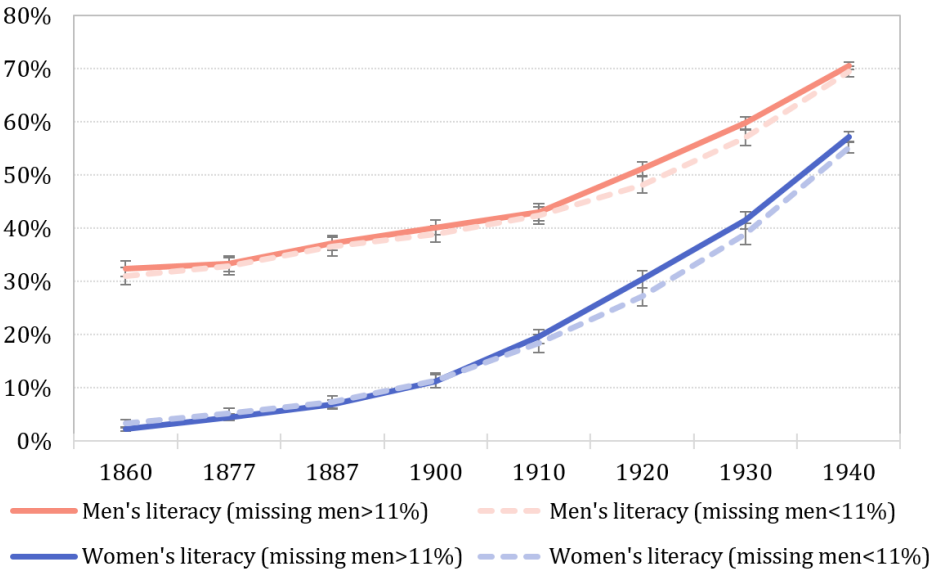
Note: The figure shows the adult sex ratio in 1860 computed as the ratio between men and women aged 30 to 60. Return to page 11.

FIGURE 8: Share of missing men across Galician municipalities (%), 1877-1930



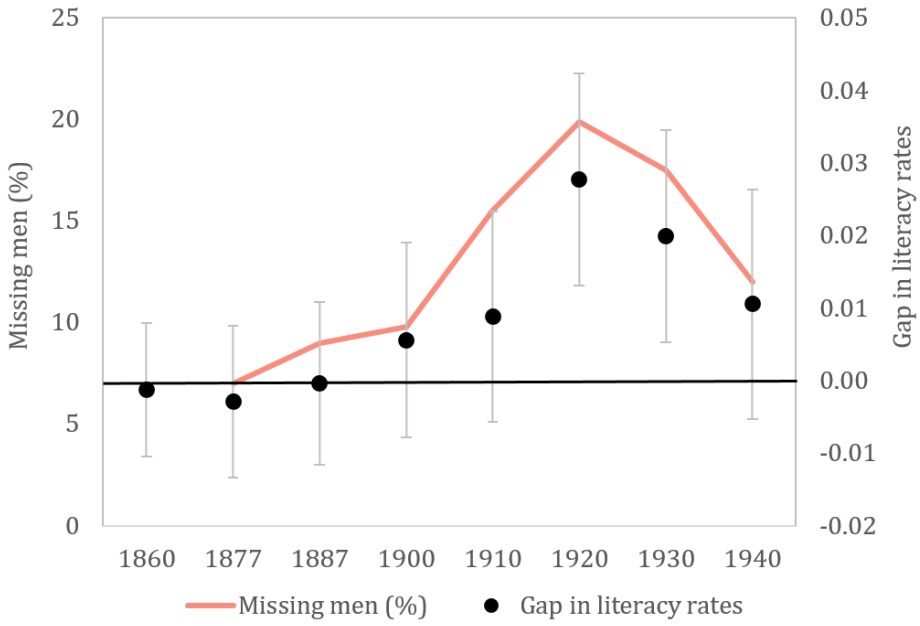
Source: Spanish population census (1877- 1930).
Note: The figure shows the share of missing men the day of the census in different years. The arrows indicate the ports shipping boats to Latin America. Return to page 17.

FIGURE 9: Evolution of literacy rates in Galicia, 1860-1940



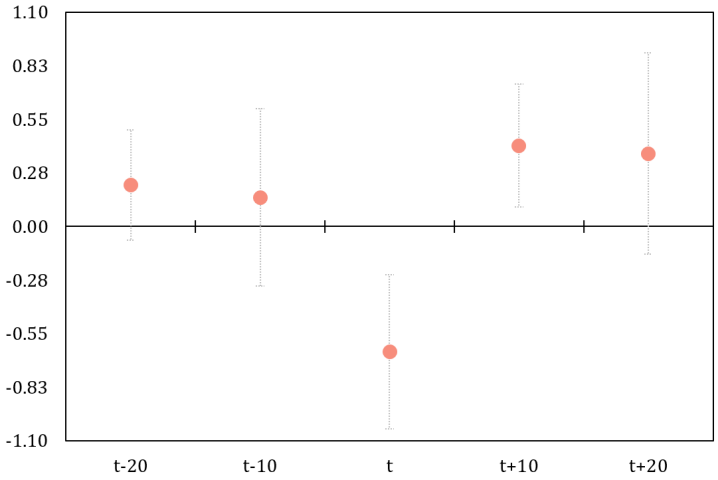
Source: Spanish population census (1860- 1940). *Note*: The solid (/dashed) lines represent the average for municipalities with an average share of missing men (1900-1930) above (/below) the median. Return to page 17.

FIGURE 10: Gap in literacy rates and share of missing men, 1860-1940



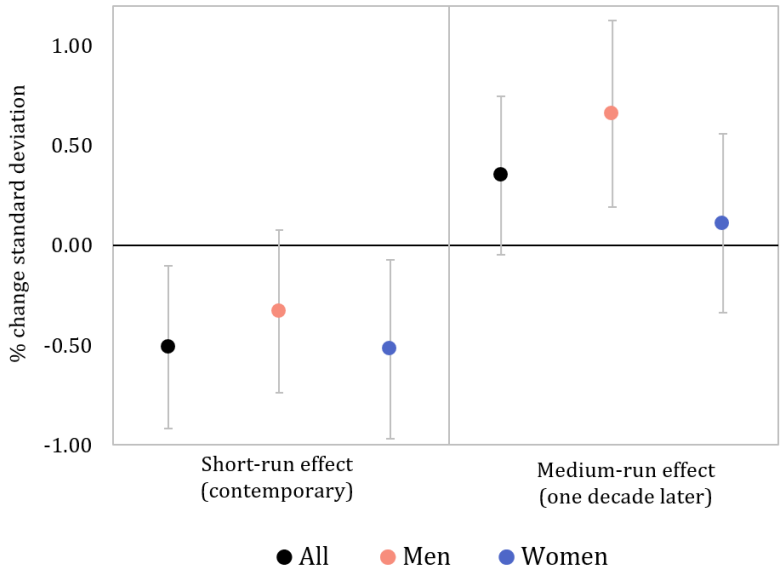
Source: Spanish population census (1860- 1940).
Note: The figure shows the share of missing men in Galicia (left) and the gap in literacy rates between municipalities with an average share of missing men (1900-1930) above and below the median (right). Error bars represent 95% confidence intervals of the difference in means. Return to page 17.

FIGURE 11: Average impact of missing men on literacy rates, 1877-1930



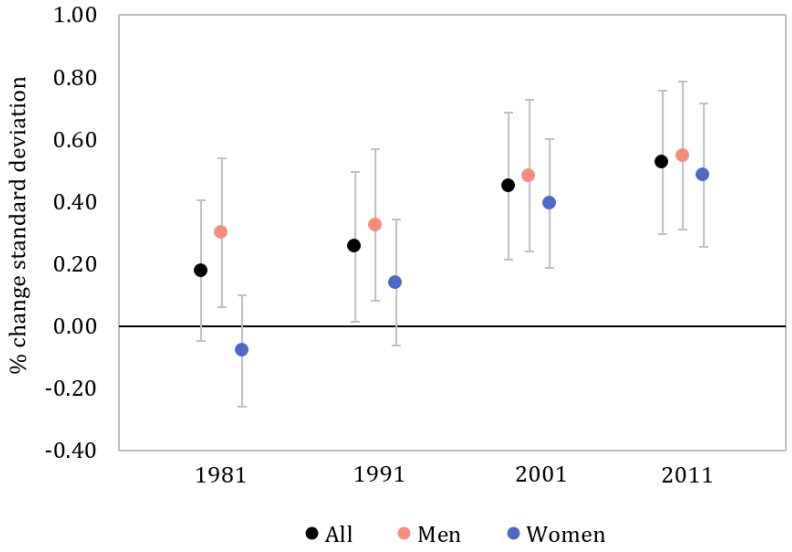
Note: The figure plots β coefficients from estimating variants of equation (1) with different lags of the share of missing men or literacy rates. For instance, the estimate of *t-k* is obtained from estimating the impact of the share of missing men in year *t* on literacy rates in year *t*-k in a panel of 4 periods (decades). The estimate of $t + k$ is obtained from estimating the impact of missing men in *t-k* on literacy rates in *t* in a panel of 4 periods (decades). The share of missing men (in year *t*) is instrumented by the growth in GDPpc at the average migrant destination (between *t*-10 and *t*) and its interaction with the adult sex ratio in 1860. Dotted lines indicate 95% confidence intervals around the point estimates. Return to page 20.

FIGURE 12: Impact of missing men on literacy rates, 1900-1930



Note: The figure shows β coefficients from estimating equations (1) and (2) by 2SLS. The short-run effect suggests that one standard deviation increase in the share of missing men (in year *t*) led to a drop in literacy rates in about 51% of a standard deviation The medium-run effect suggests that one standard deviation increase in the share of missing men (in year *t*) led to an increase in literacy rates (10 years later) in about 35% of a standard deviation The share of missing men (in year *t*) is instrumented by the growth in GDPpc at the average migrant destination and its interaction with the adult sex ratio in 1860. Dotted lines indicate 95% confidence intervals around the point estimates. Return to page 20.

FIGURE 13: Long-run impact of missing men on the share of secondary education



Note: The figure plots β coefficients from estimating equation (4) by 2SLS in different years. The estimates of the long-run effect suggest that one standard deviation increase in the average share of missing men (1900-1930) led to an increase in the share of individuals with completed secondary education in about 50% of a standard deviation in 2011. The average share of missing men (1900-1930) is instrumented by the adult sex ratio in 1860. Return to page 22.

Tables

TABLE 1: Departures from Europe to the Americas, 1850-1930

	Departures (m.),	Gross migration rates (departures/1000 people)						
	1815-1930	1861-1870	1871-1880	1881-1890	1891-1900	1901-1910	1911-1920	1921-1930
England	11.4	2.80	4.00	5.60	3.60	5.50	7.60	2.70
Italy	9.9	1.10	3.50	5.00	10.80	16.30	3.40	
Ireland	7.3	14.60	6.60	14.20	8.90	7.00	6.80	5.90
Austria-Hungary	5.0	–	–	1.10	1.60	4.80	6.10	1.40
Germany	4.8	–	1.50	2.90	1.00	0.50	0.40	1.00
Spain	4.4	–	–	3.40	3.40	7.00	10.60	4.00
Portugal	1.8	1.90	2.90	4.30	5.60	6.50	13.90	5.30
Galicia	**1.5**	**1.40**	**4.16**	**5.82**	**7.14**	**14.62**	**25.38**	**14.52**
Sweden	1.2	3.10	2.40	7.00	4.10	4.20	3.10	1.80
Norway	0.8	5.80	4.70	9.50	4.50	8.30	4.20	3.10
Finland	0.4	–	–	1.30	2.30	5.50	6.40	2.10
France	0.4	0.20	0.20	0.30	0.10	0.10	0.20	–
Denmark	0.4	–	2.10	3.90	2.20	2.80	3.20	1.70
Switzerland	0.3	–	1.30	3.20	1.40	1.4	1.70	1.40
Netherlands	0.2	0.60	0.50	1.20	0.50	0.50	0.40	0.50
Belgium	0.2	–	–	0.90	0.40	0.60	1.00	0.30

Source: Baines (1991), Sánchez-Alonso (1995) and Vázquez-González (1999).
Note: Gross migration rates represent yearly departures per thousand people at baseline. For instance, a rate of 10 means that every year there were departures equivalent to 1% of the baseline population. Return to page 7.

TABLE 2: Explaining the evolution of missing men, 1887-1930 [Panel first stage]

	Share of missing men (t)				
	(1)	(2)	(3)	(4)	(5)
Average growth at destination	0.07***	0.07**	0.09***	0.07**	0.06*
	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]
Average growth at destination x Adult sex ratio (1860)	0.04***	0.04***	0.03**	0.04***	0.04***
	[0.01]	[0.01]	[0.01]	[0.01]	[0.02]
Population in the province (10 years before)		-0.27***			-0.51***
		[0.10]			[0.14]
Literacy rate in the province (10 years before)		0.37**			0.58***
		[0.18]			[0.22]
Missing men (10 years before)			0.24***		0.24***
			[0.03]		[0.04]
Rainfall variability (previous decade)				-0.04	-0.14***
				[0.04]	[0.05]
Municipality and year FE	Yes	Yes	Yes	Yes	Yes
KP F -statistic	9.29	9.48	9.95	8.97	7.80
Mean dependent variable	0.13	0.13	0.13	0.13	0.13
Standard dependent variable	0.09	0.09	0.09	0.09	0.09
Observations (municipalities=322, years=5)	1610	1610	1610	1610	1610

Note: The table shows β coefficients from regressing the average share of missing men (in year t) on different lagged controls (equation 3). As an illustration, column (1) shows that one standard deviation increase in the average growth at destination is associated with a 7% standard deviation raise in the share of missing men on average. Increasing the adult sex ratio (1860) by one standard deviation magnifies this effect by 4% more (i.e., a 11% of standard deviation increase in the share of missing men). Clustered standard errors at the municipality level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels. Return to page 18.

TABLE 3: Explaining the intensity of missing men (1900-1930) [Cross-section first stage]

	Average share of missing men, 1900-1930				
	(1)	(2)	(3)	(4)	(5)
Adult sex ratio (w/m ,1860)	0.34***	0.31***	0.32***	0.35***	0.28***
	[0.07]	[0.07]	[0.06]	[0.06]	[0.06]
Altitude		0.21**			0.12
		[0.10]			[0.11]
Municipality size		-0.24***			-0.20**
		[0.09]			[0.08]
Population settlements		0.04			0.03
		[0.10]			[0.09]
Access to sea		-0.37*			-0.29
		[0.20]			[0.18]
Access to train		-0.09			-0.19
		[0.16]			[0.16]
Distance to port		-0.31**			-0.29**
		[0.13]			[0.13]
Distance to capital		0.09			0.09
		[0.09]			[0.08]
Population (1860)			-0.02		0.01
			[0.08]		[0.07]
Men's literacy (1860)			0.20***		0.17***
			[0.06]		[0.06]
Women's literacy (1860)			-0.11*		-0.11**
			[0.06]		[0.05]
Wealth per capita (1855)				-0.01	0.02
				[0.10]	[0.09]
Local contributions per capita (1855)				0.20*	0.09
				[0.09]	[0.09]
Land inequality (1860)				0.18**	0.20***
				[0.08]	[0.08]
KP F -statistic	23.64	21.93	26.18	33.58	22.47
Mean dependent variable	0.14	0.14	0.14	0.14	0.14
Standard deviation dependent variable	0.10	0.10	0.10	0.10	0.10
Adjusted R-squared	0.19	0.21	0.21	0.20	0.24
Observations	321	321	321	321	321

Note: The table shows β coefficients from regressing the average share of missing men (1900-30) on different covariates (equation 6). As an illustration, column (1) shows that one standard deviation increase in the adult sex ratio (1860) is associated with a 34% standard deviation raise in the average share of missing men (1900-1930). Clustered standard errors at the district level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels. Return to page 18.

TABLE 4: Effect of missing men on literacy rates, 1900-1930

	Average literacy rates (t)			
	OLS		2SLS	
	(1)	(2)	(3)	(4)
Share of missing men (t)	0.01		-0.51**	
	[0.03]		[0.21]	
Share of missing men (10 years before)		0.07**		0.35*
		[0.03]		[0.21]
Municipality and year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
KP F -statistic	-	-	9.92	15.80
Mean dependent variable	0.35	0.35	0.35	0.35
Standard deviation dependent variable	0.13	0.13	0.13	0.13
Mean share of missing men	0.14	0.11	0.14	0.11
Standard deviation share of missing men	0.08	0.09	0.08	0.09
Observations (municipalities=321, years=4)	1284	1284	1284	1284

Note: The table shows β coefficients from estimating equation (1) and (2) by OLS and 2SLS. As an illustration, column (3) shows that one standard deviation increase in the share of missing men (in year t) leads to a fall in average literacy rates (in year t) in 51% of a standard deviation. Column (4) shows that one standard deviation increase in the share of missing men (in year t) led to an increase in average literacy rates (10 years later) in 35% of a standard deviation. Clustered standard errors at the municipality level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels. Return to page 20.

TABLE 5: Long-run effect of missing men on human capital, 1981-2011

	Share of individuals with secondary education			
	1981	1991	2001	2011
	(1)	(2)	(3)	(4)
	Panel A: 2SLS			
Share of missing men (1900-1930)	0.14	0.21*	0.40***	0.46***
	[0.10]	[0.12]	[0.11]	[0.13]
	Panel B: OLS			
Share of missing men (1900-1930)	-0.09**	-0.09*	-0.02	0.04
	[0.05]	[0.05]	[0.05]	[0.06]
Controls	Yes	Yes	Yes	Yes
Kleibergen Paap F-statistic	50.79	50.79	50.79	50.79
Mean dependent variable	0.05	0.24	0.36	0.59
Standard deviation dependent variable	0.03	0.07	0.10	0.11
Mean missing men 1900-1930	0.14	0.14	0.14	0.14
Standard deviation missing men 1900-1930	0.07	0.07	0.07	0.07
Observations	314	314	314	314

Note: The table shows β coefficients of estimating equation (4) by OLS and 2SLS. As an illustration, column (4) in Panel A shows that one standard deviation increase in the average share of missing men (1900-1930) leads to an increase in the share of individuals with completed secondary education (in 2011) in 46% of a standard deviation. Historical controls include population size (1900), male and female literacy rates (1900) and wealth and local taxes (1855) at the municipality level. Clustered standard errors at the district level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels. Return to page 22.

TABLE 6: Explaining the creation of migrant associations and schools, 1910-1960

	At least one association	Number of associations	Number of migrant schools	
Covariates standardized	(1)	(2)	(3)	(4)
Share of missing men (1900-1930)	0.11	1.07	2.25**	1.67**
	[0.14]	[0.72]	[0.90]	[0.68]
Altitude	-0.08*	-0.35*	-0.53**	-0.34*
	[0.04]	[0.19]	[0.23]	[0.18]
Municipality size	0.07	0.90***	0.88**	0.39
	[0.05]	[0.34]	[0.41]	[0.26]
Population settlements	0.07*	0.19	0.19	0.08
	[0.04]	[0.27]	[0.31]	[0.24]
Access to sea	-0.13	0.24	0.54	0.41
	[0.10]	[0.47]	[0.51]	[0.41]
Access to train	0.10	0.47	0.58	0.32
	[0.08]	[0.34]	[0.44]	[0.35]
Distance to port	-0.07	-0.70**	-0.35	0.03
	[0.06]	[0.29]	[0.40]	[0.30]
Distance to capital	0.05	0.09	0.11	0.06
	[0.04]	[0.17]	[0.23]	[0.18]
Population (1860)	0.01	-0.03	-0.21	-0.20
	[0.03]	[0.13]	[0.18]	[0.14]
Men's literacy (1860)	0.01	-0.10	-0.45*	-0.40**
	[0.04]	[0.19]	[0.24]	[0.18]
Women's literacy (1860)	0.01	0.21	0.44*	0.33**
	[0.03]	[0.17]	[0.23]	[0.16]
Wealth per capita (1855)	0.05	0.17	-0.08	-0.17
	[0.03]	[0.16]	[0.22]	[0.17]
Local contributions per capita (1855)	-0.11***	-0.48**	-0.33	-0.07
	[0.03]	[0.19]	[0.24]	[0.18]
Land inequality (1860)	-0.05	-0.04	-0.34	-0.32*
	[0.04]	[0.17]	[0.22]	[0.17]
Rainfall variability (1900-1930)	0.01	-0.44	-1.00***	-0.75***
	[0.07]	[0.33]	[0.35]	[0.28]
Schools per 1000 people (1875)	-0.10***	-0.50***	-0.45**	-0.18
	[0.03]	[0.14]	[0.19]	[0.14]
Children enrollment rates (1875)	0.01	0.30*	0.31	0.14
	[0.03]	[0.18]	[0.21]	[0.16]
Migrants to Rio de la Plata (1900-1930)	0.02	-0.08	-0.27*	-0.23*
	[0.03]	[0.16]	[0.16]	[0.12]
Number of migrant associations				0.54***
				[0.08]
Mean of dependent variable	0.50	1.91	0.92	0.92
Adjusted R-squared	0.18	0.23	0.19	0.48
Observations	321	321	321	321

Note: The table shows the result of estimating equation (8) instrumenting the average share of missing men (1900-1930) by the adult sex ratio in 1860. Covariates are standardized to ease interpretation. As an illustration, column (4) shows that one standard deviation increase in the share of missing men (1900-1930) leads to an increase of 1.67 migrant schools on average. Clustered standard errors at the district level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels. Return to page 25.

TABLE 7: Impact of migrants' associations and schools (1910-1960) on education and mobility

	Educational attainment (share of people with completed primary)			**Emigration** (size of cohort c relative to $c-1$)		
	(1)	(2)	(3)	(4)	(5)	(6)
Opening of a migrant association	-0.001		-0.007*	-0.023		-0.033
	[0.003]		[0.004]	[0.022]		[0.023]
Opening of a migrant school		0.007***	0.01***		0.009	0.017**
		[0.002]	[0.003]		[0.008]	[0.009]
Mean of dependent variable	0.30	0.30	0.30	1.18	1.18	1.18
Standard deviation dependent variable	0.27	0.27	0.27	0.55	0.55	0.55
Mean of explanatory variable	1.04	0.60	-	1.04	0.60	-
Standard deviation explanatory variable	1.67	1.49	-	1.67	1.49	-
Municipality and cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations (municipalities=312, cohorts=8)	2496	2496	2496	2496	2496	2496

Note: The table shows the results of estimating equation (7) by OLS. As an illustration, column (3) shows that the foundation of a new migrant association when individuals are younger than 10, reduces their likelihood of completing primary education by 0.07% on average. Similarly, the opening of a new school financed by migrants increased their chances to complete primary education by 1%. Note that most association financed several school having therefore a larger positive impact on educational attainment. Column (6) shows that the foundation of a new migrant association when individuals are younger than 10, reduces the size of their cohort by 3.3%. Similarly, the opening of a new school financed by migrants increased the size of their cohort by 1.7%. These results suggest that migrants' associations fostered more migration, but if they financed at least two schools, the net impact is zero. All specifications include cohort and municipality fixed effects. The sample comprises all 5-year cohort groups between 1907 and 1947. Clustered standard errors at the municipality level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels. Return to page 26.

TABLE 8: Long-run effects of missing men on beliefs about education and effort

	Value of education and effort (0-10 scale)			
	(1)	(2)	(3)	(4)
Share of missing men (1900-1930)	0.21	0.23**	0.21	0.22*
	[0.16]	[0.11]	[0.14]	[0.12]
Individual controls	Yes	No	No	Yes
Historical controls	No	Yes	No	Yes
Educational attainment	No	No	Yes	Yes
Mean of dependent variable	4.98	4.98	4.98	4.98
Standard deviation dependent variable	2.33	2.33	2.33	2.33
Mean share of missing men (1900-1930)	0.11	0.11	0.11	0.11
Standard deviation share of missing men (1900-1930)	0.06	0.06	0.06	0.06
Observations	1335	1335	1335	1335

Note: The table shows β coefficients of estimating equation (8) by 2SLS. As an illustration, column (4) shows that one standard deviation increase in the average share of missing men (1900-1930) leads to a 22% of standard deviation increase in reported value for education and effort (e.g., 0.5 points in a 0-10 scale). The average share of missing men (1900-1930) is instrumented by the proxy for migrant networks (i.e., adult sex ratio in 1860). Individual controls include sex, age, and marital status. Historical controls include population size (1900), male and female literacy rates (1900), wealth and local taxes (1855) at the municipality level. Clustered standard errors at the municipality level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels. Return to page 28.

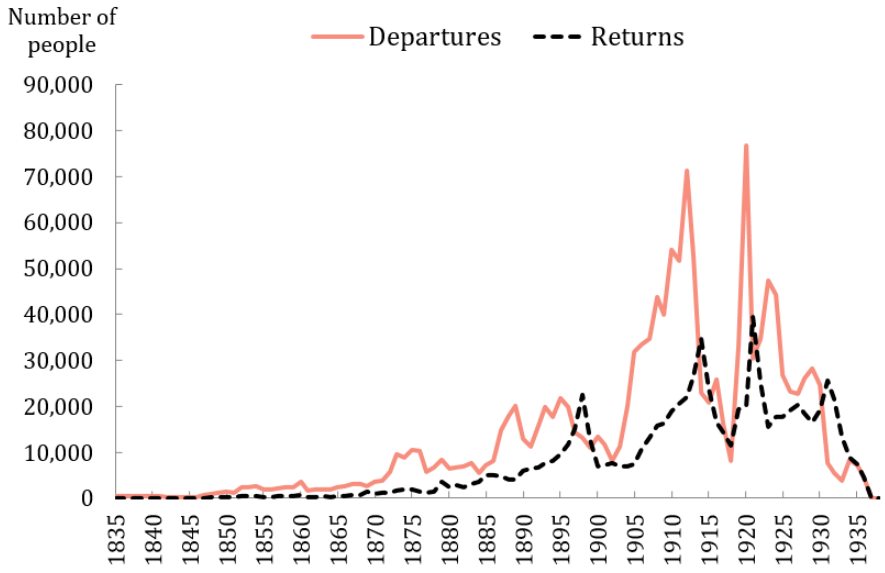
Online Appendix

A Additional Figures and Tables

FIGURE A1: Location of Galicia (Spain)

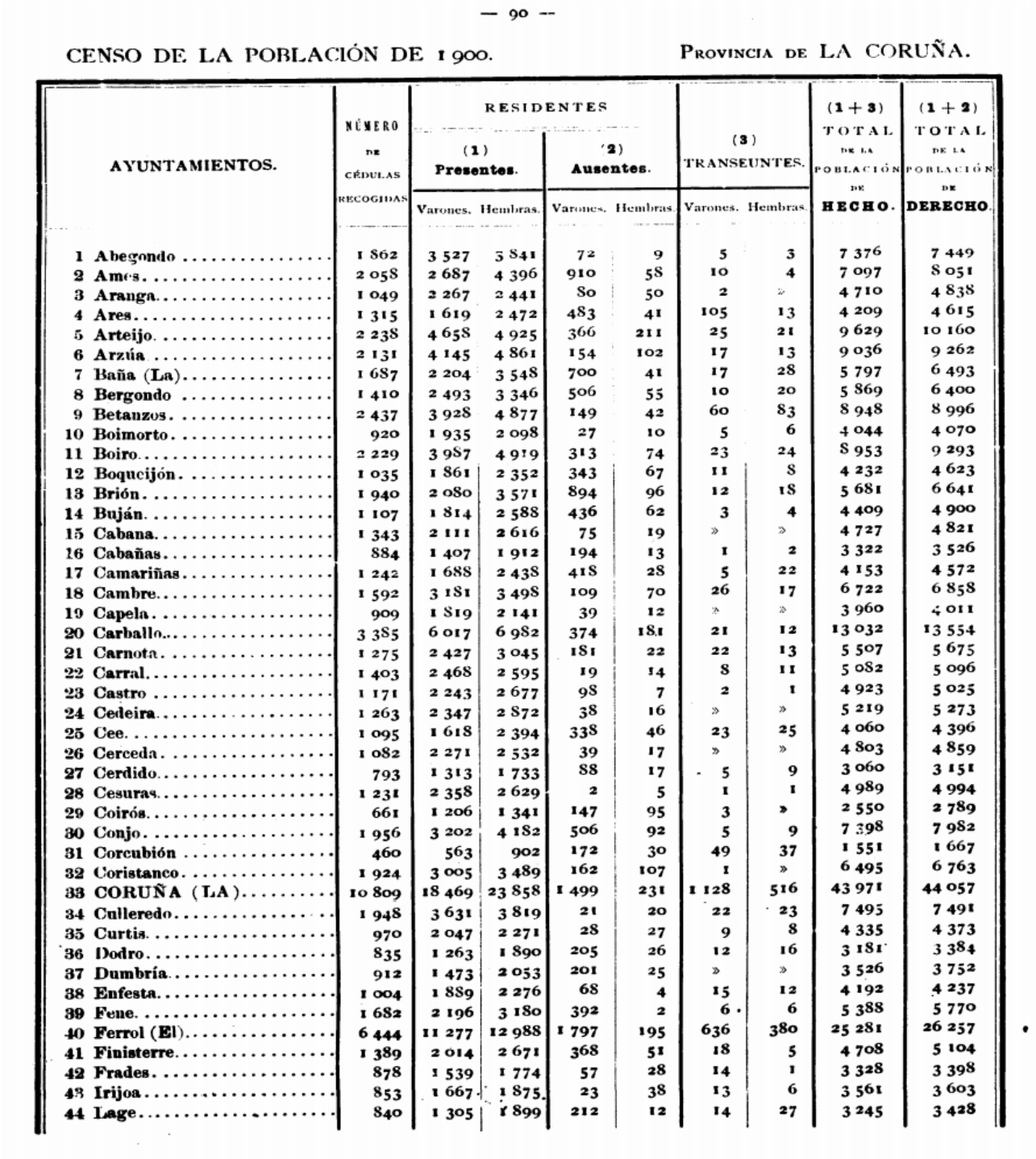
Note: Return to page 8.

FIGURE A2: Yearly departures and returns of Galicians, 1835-1935



Source: Vázquez-González (1999) and Official Migration Statistics.
Note: The figure displays the number of departures (returns) of Galicians to (from) Latin America every year. These migration flows are probably an underestimation because clandestine migration was not recorded. Return to page 8.

FIGURE A3: Example of original census sources



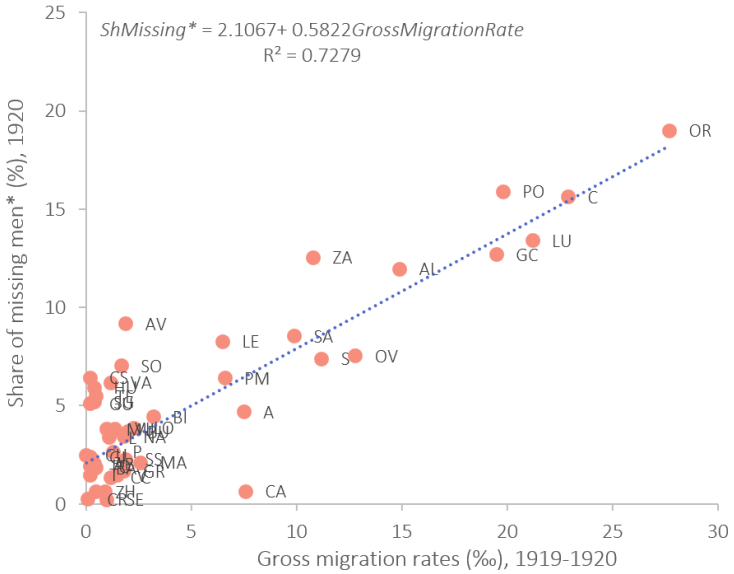
— 90 —

CENSO DE LA POBLACIÓN DE 1 900. PROVINCIA DE LA CORUÑA.

AYUNTAMIENTOS.	NÚMERO DE CÉDULAS RECOGIDAS	RESIDENTES (1) Presentes. Varones.	RESIDENTES (1) Presentes. Hembras.	RESIDENTES (2) Ausentes. Varones.	RESIDENTES (2) Ausentes. Hembras.	(3) TRANSEUNTES. Varones.	(3) TRANSEUNTES. Hembras.	(1 + 3) TOTAL DE LA POBLACIÓN DE HECHO.	(1 + 2) TOTAL DE LA POBLACIÓN DE DERECHO.
1 Abegondo	1 862	3 527	3 841	72	9	5	3	7 376	7 449
2 Ames	2 058	2 687	4 396	910	58	10	4	7 097	8 051
3 Aranga	1 049	2 267	2 441	80	50	2	»	4 710	4 838
4 Ares	1 315	1 619	2 472	483	41	105	13	4 209	4 615
5 Arteijo	2 238	4 658	4 925	366	211	25	21	9 629	10 160
6 Arzúa	2 131	4 145	4 861	154	102	17	13	9 036	9 262
7 Baña (La)	1 687	2 204	3 548	700	41	17	28	5 797	6 493
8 Bergondo	1 410	2 493	3 346	506	55	10	20	5 869	6 400
9 Betanzos	2 437	3 928	4 877	149	42	60	83	8 948	8 996
10 Boimorto	920	1 935	2 098	27	10	5	6	4 044	4 070
11 Boiro	2 229	3 987	4 919	313	74	23	24	8 953	9 293
12 Boqucijón	1 035	1 861	2 352	343	67	11	8	4 232	4 623
13 Brión	1 940	2 080	3 571	894	96	12	18	5 681	6 641
14 Buján	1 107	1 814	2 588	436	62	3	4	4 409	4 900
15 Cabana	1 343	2 111	2 616	75	19	»	»	4 727	4 821
16 Cabañas	884	1 407	1 912	194	13	1	2	3 322	3 526
17 Camariñas	1 242	1 688	2 438	418	28	5	22	4 153	4 572
18 Cambre	1 592	3 181	3 498	109	70	26	17	6 722	6 858
19 Capela	909	1 819	2 141	39	12	»	»	3 960	4 011
20 Carballo	3 385	6 017	6 982	374	181	21	12	13 032	13 554
21 Carnota	1 275	2 427	3 045	181	22	22	13	5 507	5 675
22 Carral	1 403	2 468	2 595	19	14	8	11	5 082	5 096
23 Castro	1 171	2 243	2 677	98	7	2	1	4 923	5 025
24 Cedeira	1 263	2 347	2 872	38	16	»	»	5 219	5 273
25 Cee	1 095	1 618	2 394	338	46	23	25	4 060	4 396
26 Cerceda	1 082	2 271	2 532	39	17	»	»	4 803	4 859
27 Cerdido	793	1 313	1 733	88	17	5	9	3 060	3 151
28 Cesuras	1 231	2 358	2 629	2	5	1	1	4 989	4 994
29 Coirós	661	1 206	1 341	147	95	3	»	2 550	2 789
30 Conjo	1 956	3 202	4 182	506	92	5	9	7 398	7 982
31 Corcubión	460	563	902	172	30	49	37	1 551	1 667
32 Coristanco	1 924	3 005	3 489	162	107	1	»	6 495	6 763
33 CORUÑA (LA)	10 809	18 469	23 858	1 499	231	1 128	516	43 971	44 057
34 Culleredo	1 948	3 631	3 819	21	20	22	23	7 495	7 491
35 Curtis	970	2 047	2 271	28	27	9	8	4 335	4 373
36 Dodro	835	1 263	1 890	205	26	12	16	3 181	3 384
37 Dumbría	912	1 473	2 053	201	25	»	»	3 526	3 752
38 Enfesta	1 004	1 889	2 276	68	4	15	12	4 192	4 237
39 Fene	1 682	2 196	3 180	392	2	6	6	5 388	5 770
40 Ferrol (El)	6 444	11 277	12 988	1 797	195	636	380	25 281	26 257
41 Finisterre	1 389	2 014	2 671	368	51	18	5	4 708	5 104
42 Frades	878	1 539	1 774	57	28	14	1	3 328	3 398
43 Irijoa	853	1 667	1 875	23	38	13	6	3 561	3 603
44 Lage	840	1 305	1 899	212	12	14	27	3 245	3 428

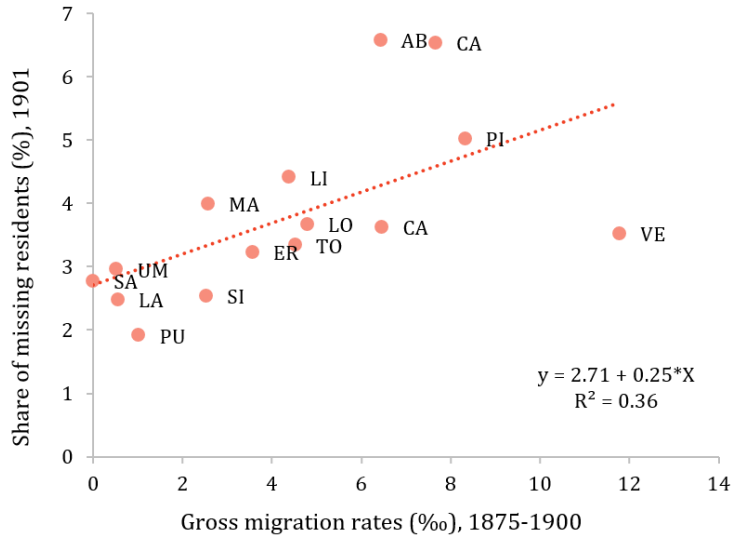
Note: The image shows a typical page in the census chapters. Each row represents a municipality. The population is classified between present (1), missing (2) and transients (3). Return to page 10.

FIGURE A4: Emigration and missing men across Spanish provinces, 1920



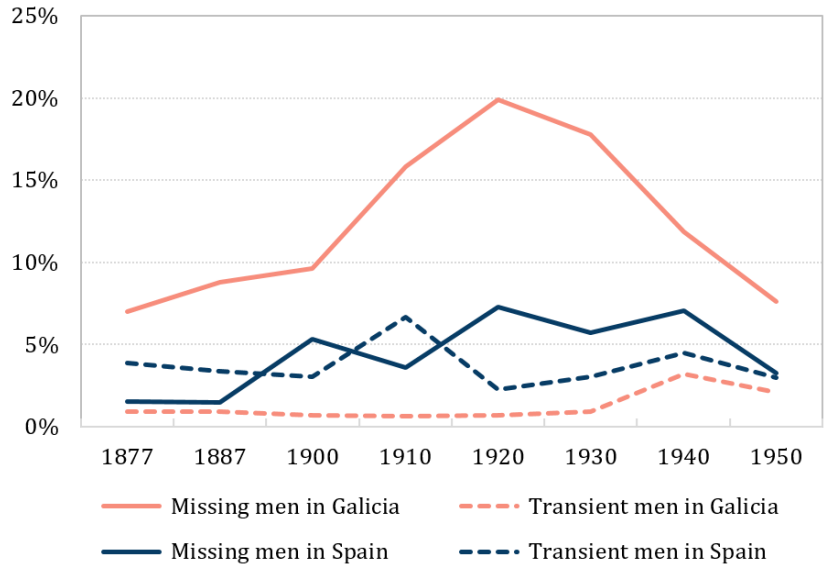
Source: Sánchez-Alonso (1995) and Spanish population census (1920).
Note: "Migration" stands for sea departures. "Missing" stands for residents absent the day of the census. Transients (passers-by) were subtracted from the missing to avoid double counting. Return to page 10.

FIGURE A5: Emigration and missing residents across Italian provinces, 1901



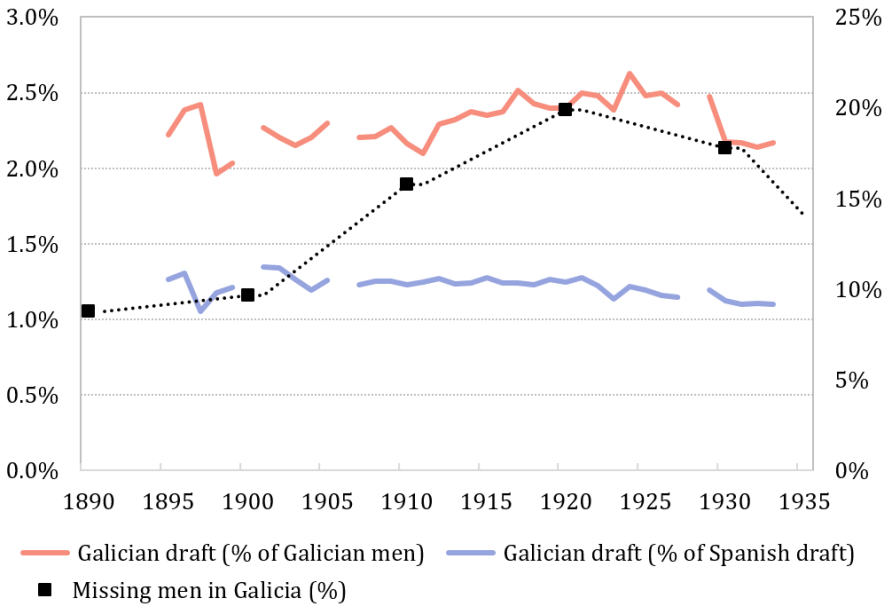
Source: Italian population census (1901) and migration statistics.
Note: PI (Piemonte), LI (Liguria), LO (Lombardia), VE (Veneto), ER (Emilia Ducati/Romagna), TO (Toscana), MA(Marche), UM(Umbria), LA (Lazio), AB (Abruzzi e Molise), CA (Campania), PU (Puglie), BA (Basilicata), CA (Calabrie), SI (Sicilia), SA (Sardegna). Return to page 10.

FIGURE A6: Missing and transient men in Galicia and Spain, 1877-1950



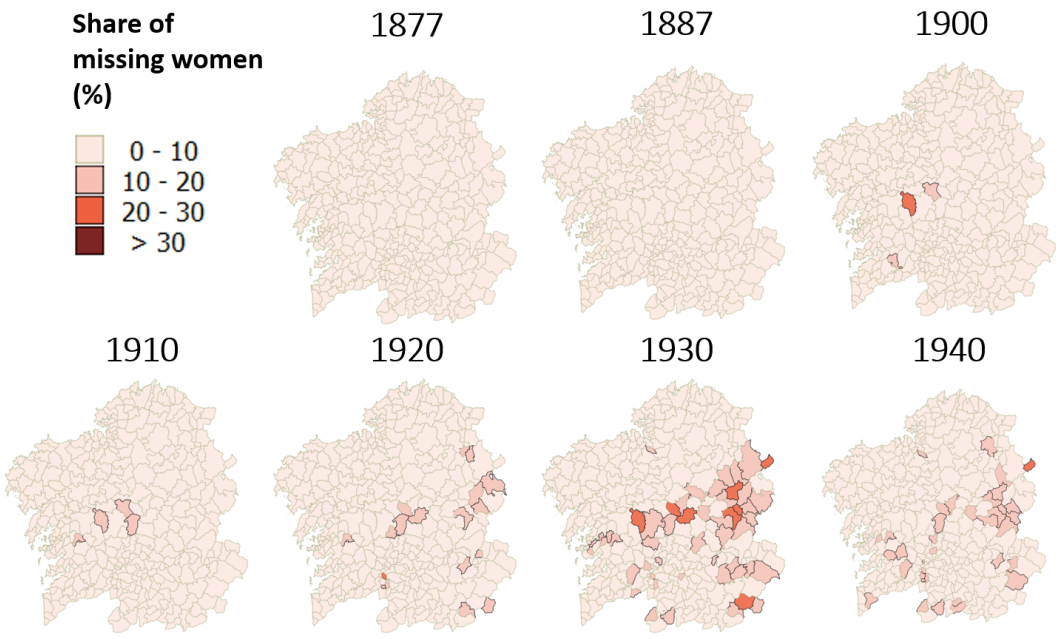
Note: The figure shows the number of missing (i.e., absent the census day) and transient (i.e., in a municipality other than where they regularly reside) men as share of the total male population in Galicia and Spain. Return to page 10.

FIGURE A7: Military draft in Galicia, 1890-1934



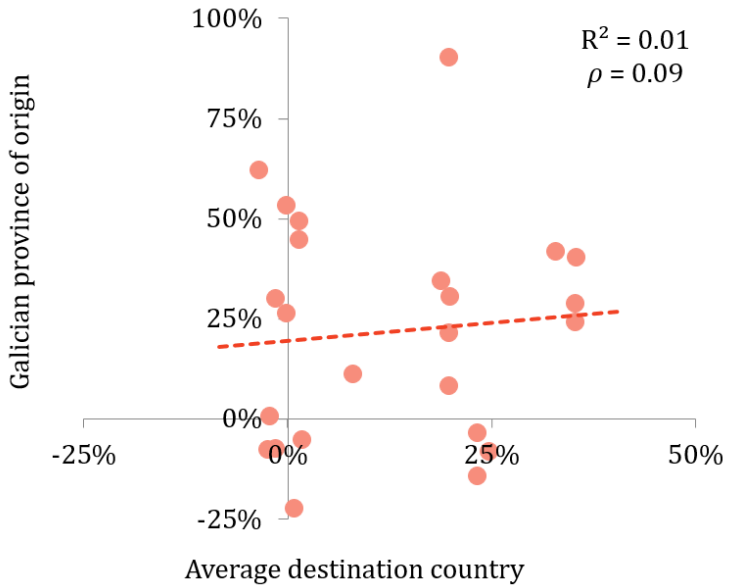
Note: The figure shows the number of Galician men drafted into the military as a share of all Galician men (left axis) and as a share of all Spanish men drafted (right axis). Return to page 10.

FIGURE A8: Share of missing women across Galician municipalities (%), 1877-1930



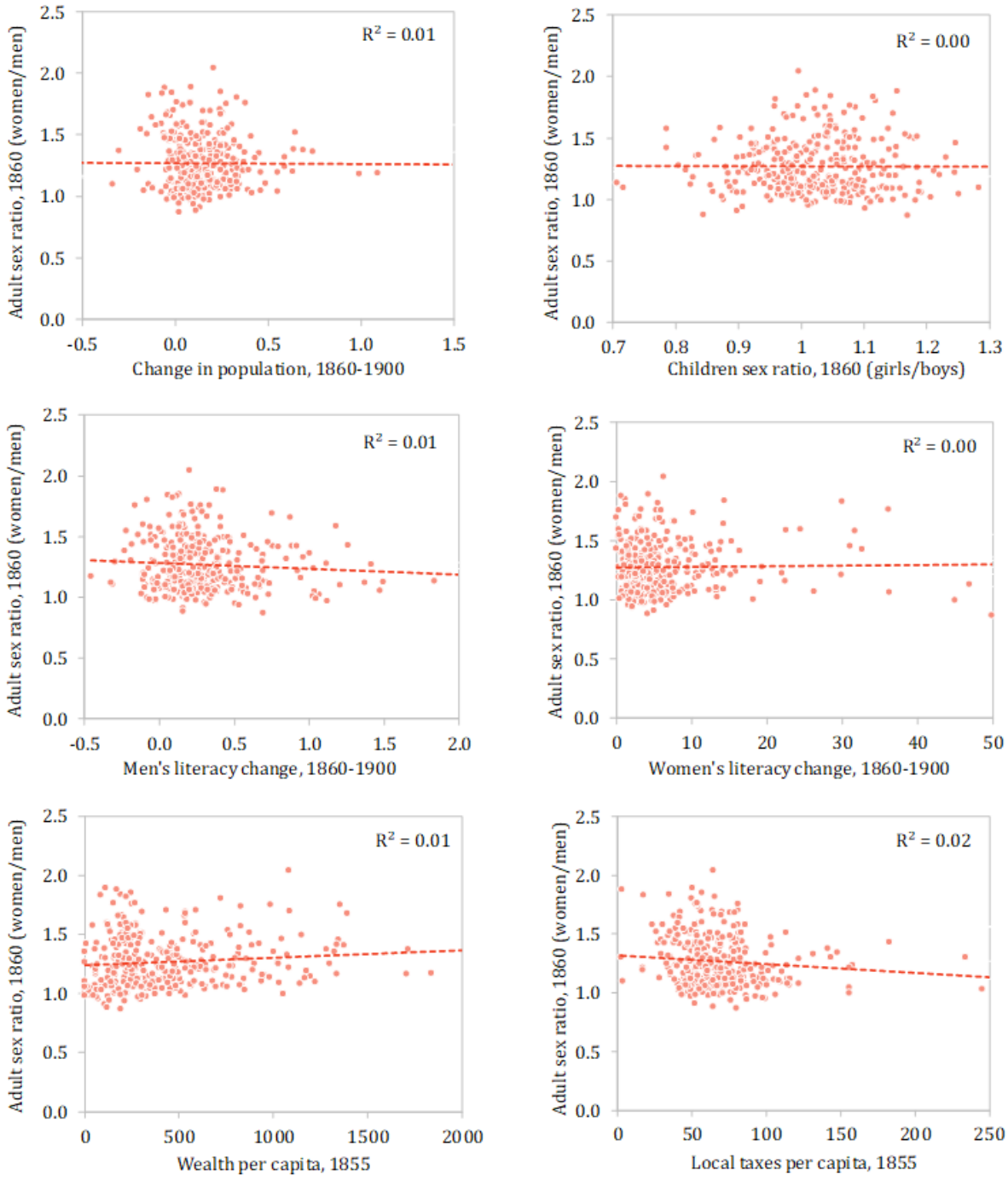
Source: Spanish population census (1877- 1930).
Note: The figure shows the share of women reported as missing in the population censuses of different years. The stronger the intensity of the color the higher the share.

FIGURE A9: Economic growth at destination and origin, 1860-1930



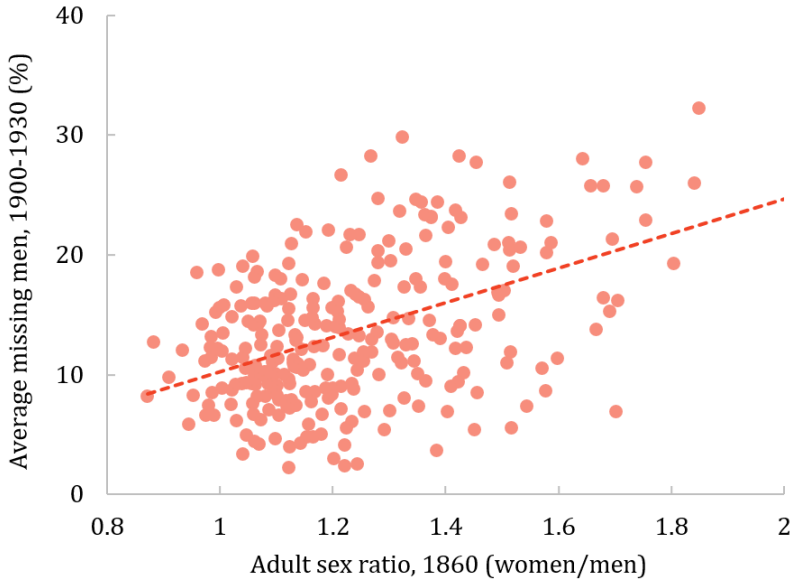
Source: Maddison (2018) for GDP in Latin America, Díez-Minguela et al. (2018) for GDP in Galician provinces, Vázquez-González (1999) for destinations by province.
Note: The x-asis represents the deviations from the GDPpc trend in the average destination country and the y-axis the deviations from the GDPpc trend in the corresponding Galician province. The correlation between both series is 0.09. Return to page 14.

FIGURE A10: Relationship between adult sex ratio (1860) and other covariates



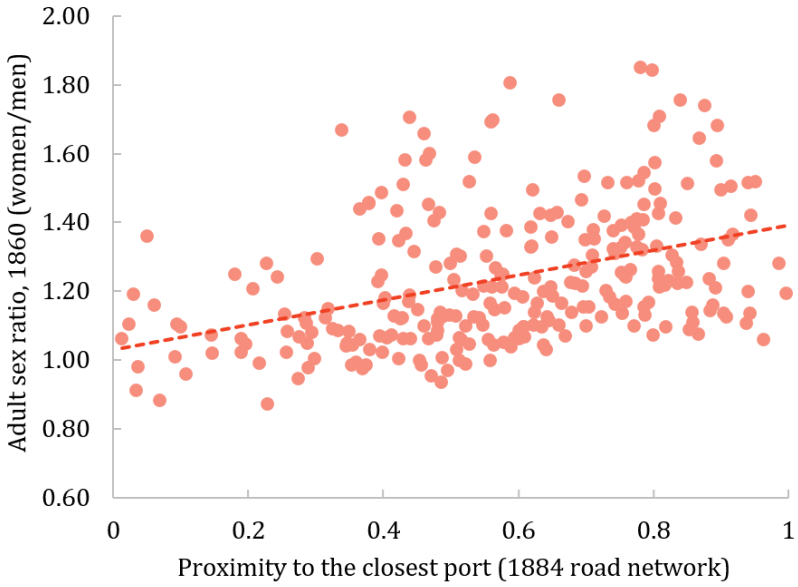
Note: The different figures show the relationship between the adult sex ratio in 1860 (i.e., proxy for pioneer migrant networks) and other covariates measured before the mass emigration era (1900-1930). Return to page 14.

FIGURE A11: Sex ratio in 1860 and average share of missing men in 1900-1930



Note: The figure shows the relationship between the adult sex ratio in 1860 (i.e, proxy of pioneer migrant networks) and the average share of missing men between 1900 and 1930 across all municipalities. Return to page 14.

FIGURE A12: Proximity to the port and average share of missing men in 1900-1930



Note: The figure shows the relationship between the walking distance to the closest port and the adult sex ratio in 1860 (i.e, proxy for pioneer migrant networks) across all municipalities. Note that the measure of distance has been normalized to range between 0 and 1. Return to page 14.

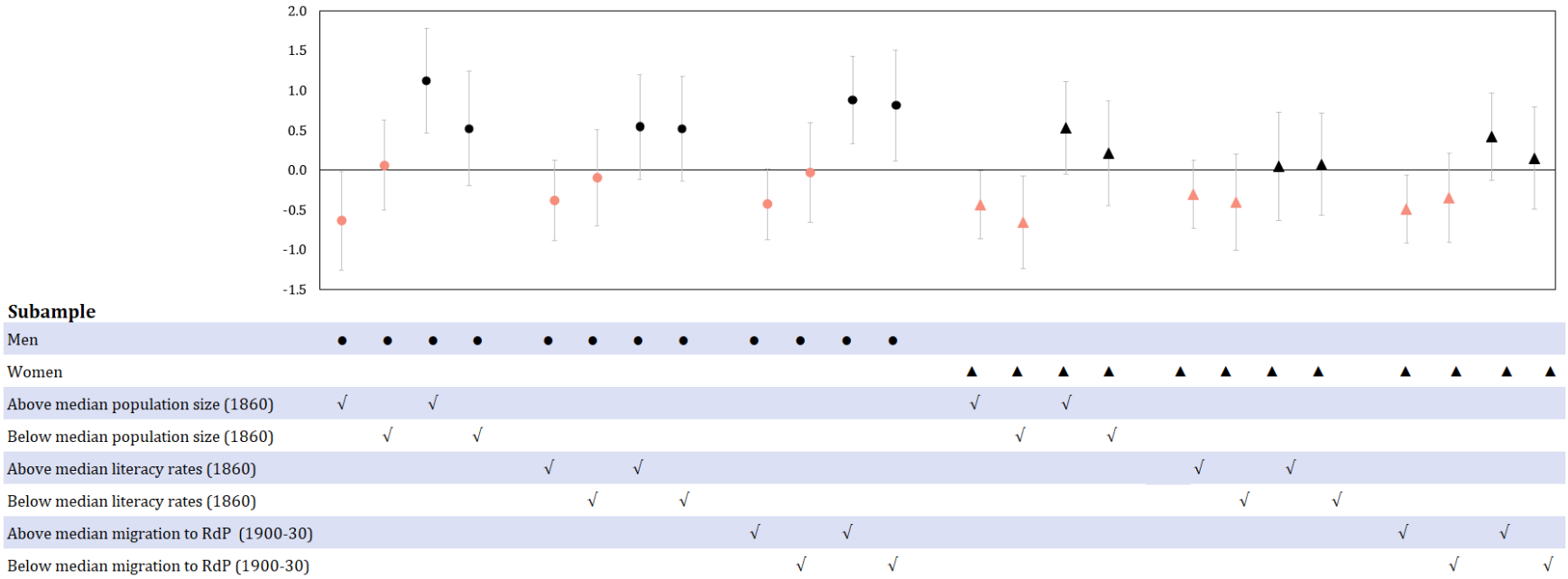
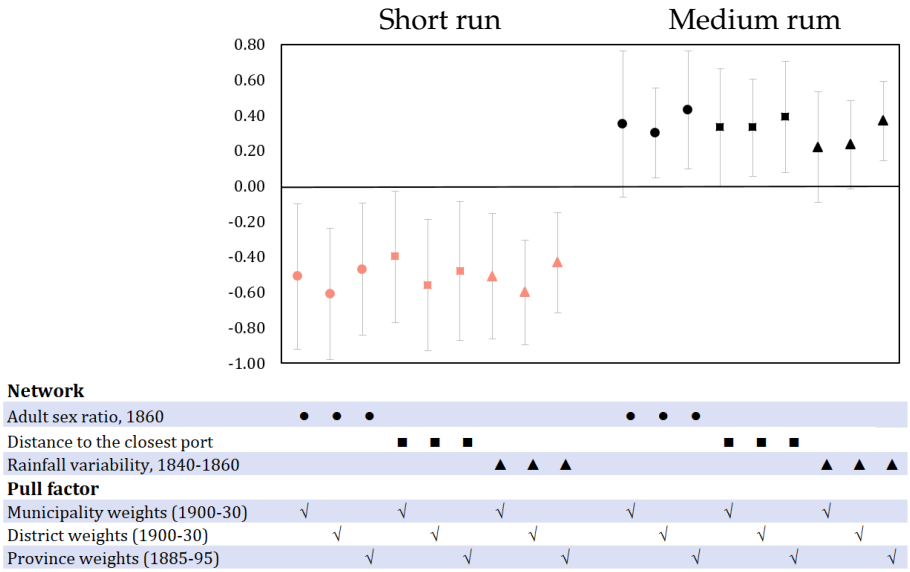


FIGURE A13: Heterogeneity in the short and medium-run effects of missing men

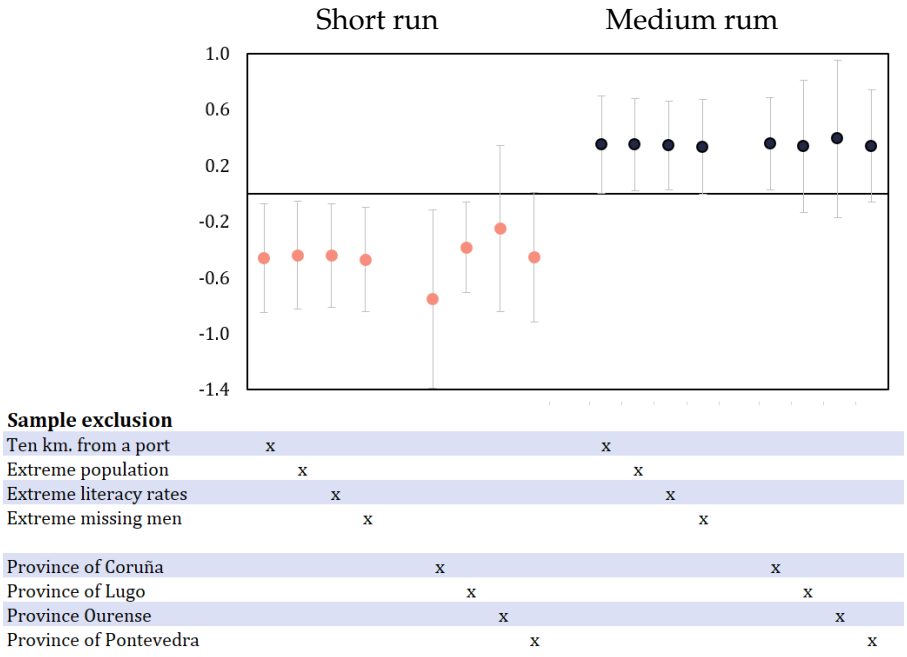
Note: The figure shows the results of estimating equations (1) and (2) for different subsamples. Red circles indicate β coefficient for equation (1), the short-run effect of missing men in literacy rates, while black circles β coefficient for equation (2), the medium run effect of missing men on literacy rates. The share of missing men (in year t) is instrumented with the pull factor (in year t) and its interaction with a proxy for migrant networks. The brackets indicate confidence interval at the 95% level. Return to page 21

FIGURE A14: Short and medium-run effects with alternative instruments

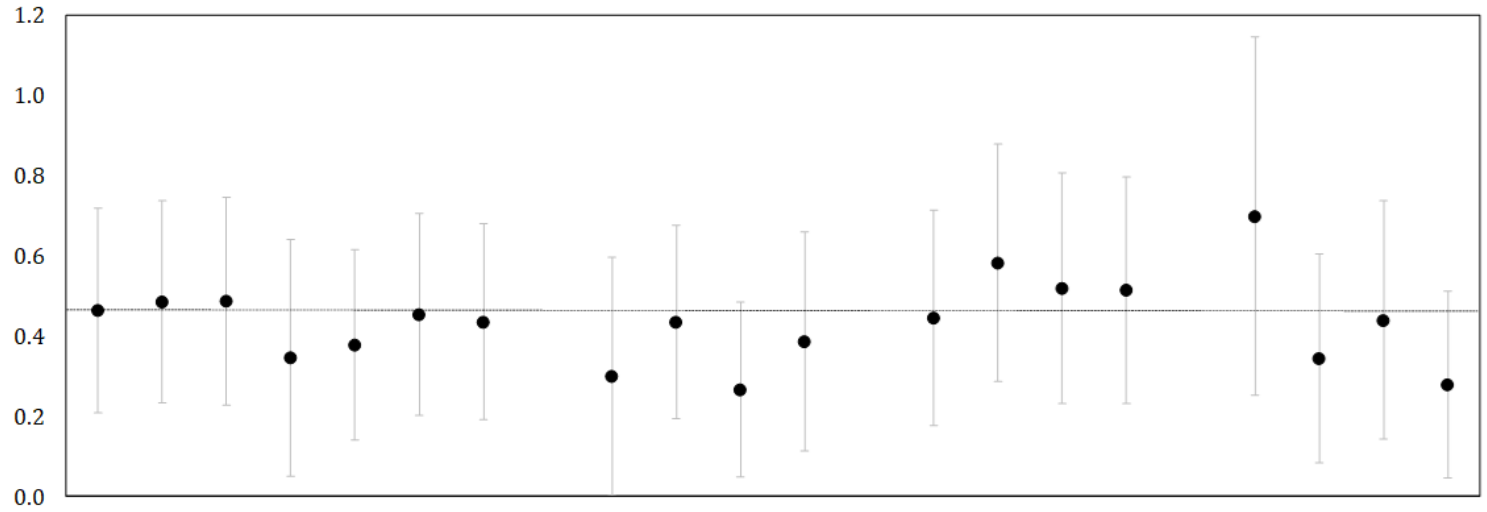


Note: The figure shows the results of estimating equations (1) and (2) with alternative instruments for the pull and network variables. Red circles indicate the β coefficient for equation (1), the short-run effect of missing men in literacy rates, while black circles the β coefficient for equation (2), the medium run effect of missing men on literacy rates. The brackets indicate confidence interval at the 95% level. Return to page 21.

FIGURE A15: Robustness checks of short and medium-run effects



Note: The figure shows the results of estimating equations (1) and (2) with different sub-samples. Red circles indicate β coefficient for equation (1), the short-run effect of missing men in literacy rates, while black circles β coefficient for equation (2), the medium run effect of missing men on literacy rates. The brackets indicate confidence interval at the 95% level. Return to page 21.

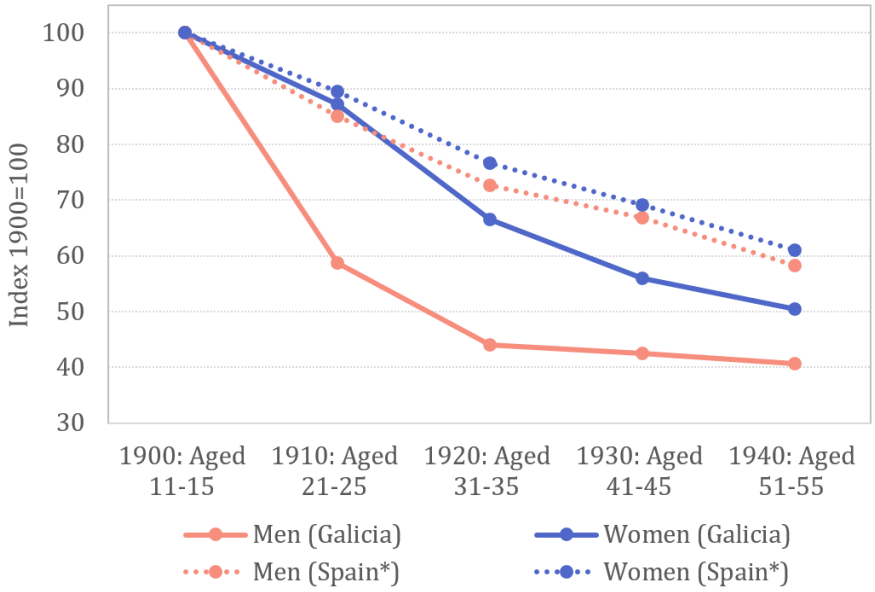


Additional controls				
Baseline	x			
Population growth (1860-1900)	x			
Literacy growth (1860-1900)	x			
Rainfall shocks (1900-1930)	x			
Land inequality (1860)	x			
Enrollment rates (1908)	x			
Schools availability (1908)	x			
Latitude		x		
Longitude		x		
Municipality's size		x		
Population settlements		x		
Sample exclusion				
Twenty km. from a port			x	
Extreme missing men			x	
Extreme population			x	
Extreme literacy rates			x	
Province of Coruña				x
Province of Lugo				x
Province Ourense				x
Province of Pontevedra				x

FIGURE A16: Robustness checks of long-run effect of missing men (1900-30) on educational attainment capital (2011)

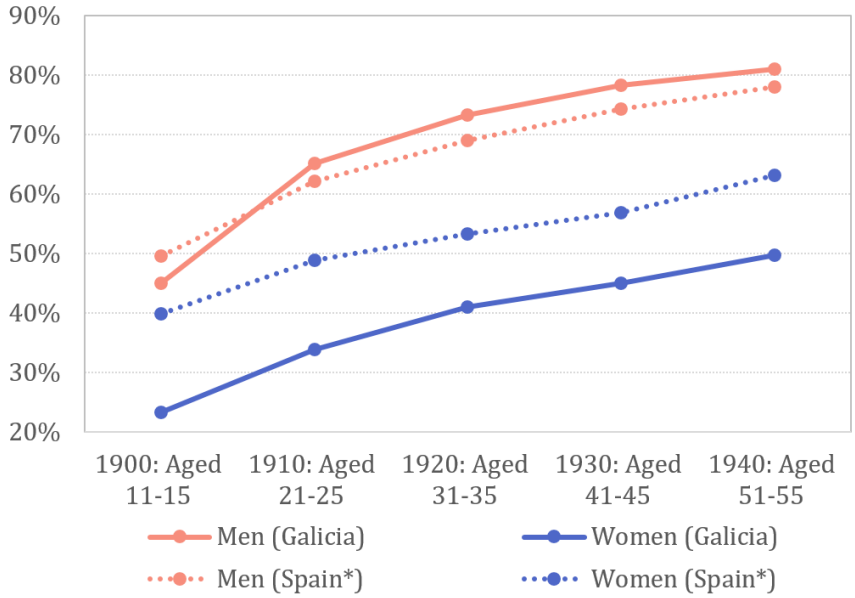
Note: The figure shows the results of estimating equation (4) by 2SLS with different controls and sub-samples . The dependent variable is the share of individuals with completed secondary education in 2011 and the main explanatory variable the average share of missing men (1900-1930). The point estimates are β coefficients and the horizontal line represent the baseline estimate. The average share of missing men (1900-1930) is instrumented with the proxy for migrant networks (adult sex ratio in 1860). Baseline controls include population (1860), literacy rates (1860), wealth per capita (1855) and local contributions per capita (1855) at municipality level. The brackets indicate confidence interval at the 95% level based on clustered standard errors at the district level. Return to page 23.

FIGURE A17: Evolution of the size of the cohort born in 1885-1899



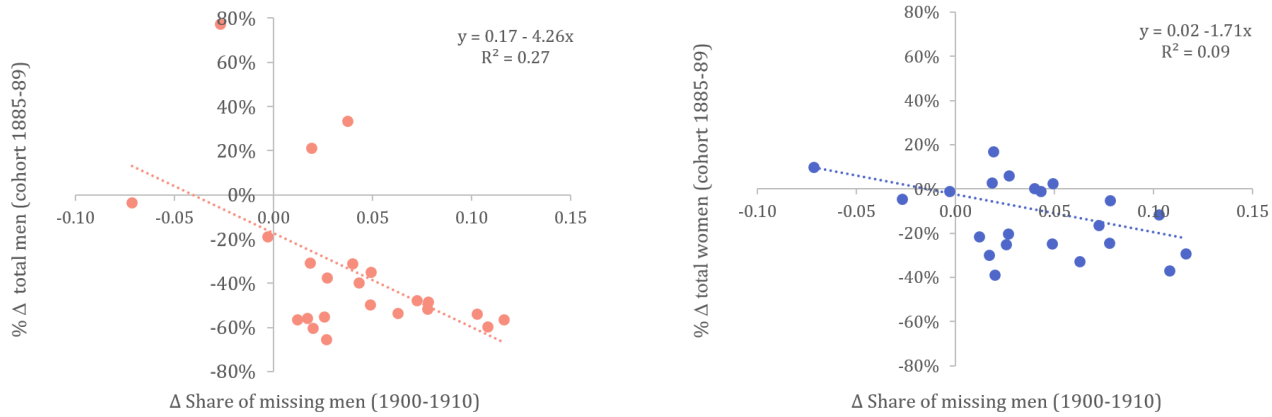
Note: The figure shows the evolution of the cohort of individuals born in 1885-1889. The size of the cohort has been normalized to 100 in 1900. For instance, the figure shows that the size of this cohort of Galician men decreased by 40% between 1900 and 1910. The figures for Spain do not include Galicia. Return to page 25.

FIGURE A18: Evolution of literacy rates of the cohort born in 1885-1889



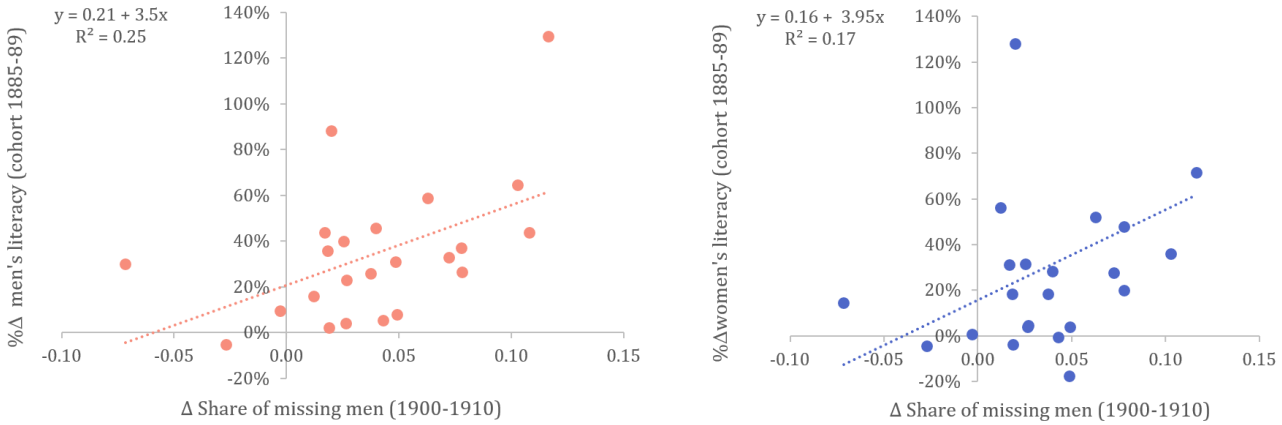
Note: The figure shows the evolution of literacy rates for the cohort of individuals born in 1885-1889. For instance, the figure shows that the literacy rate of this cohort of Galician men increased from 45% to 66% between 1900 and 1910. The figures for Spain do not include Galicia. Return to page 25.

FIGURE A19: Missing men and cohort size changes, 1900-1910



Note: The figure shows the relationship between changes in the share of missing men (1900-10) and changes in the size of the cohort born in 1885-1889 (1900-10) across the 23 largest Galician districts. For example, a value of -50% means that the size of the cohort halved between 1900 and 1910. Return to page 21.

FIGURE A20: Missing men and literacy rates changes, 1900-1910

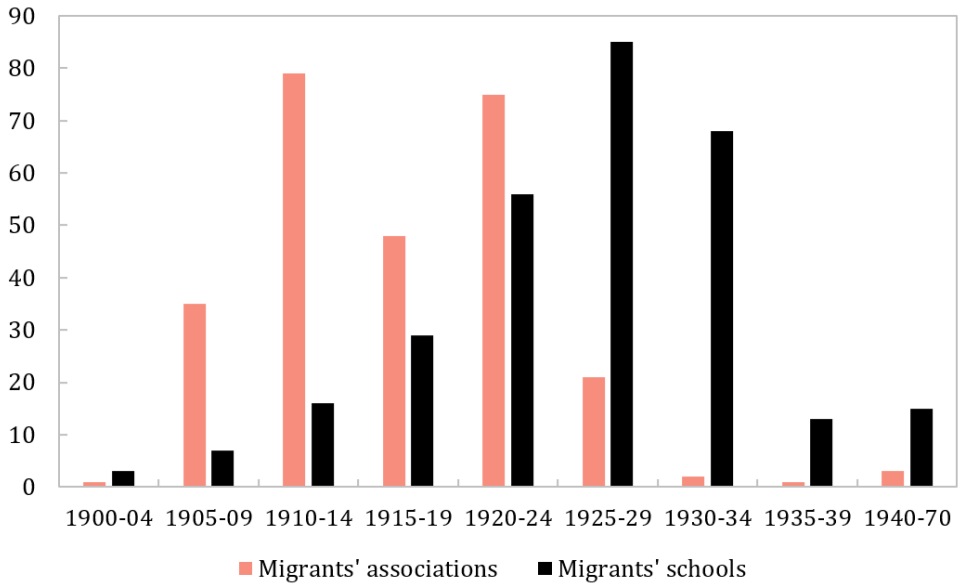


Note: The figure shows the relationship between changes in the share of missing men (1900-10) and changes in literacy rates of the cohort born in 1885-1889 (1900-10) across the 23 largest Galician districts. For instance, a value of 20% means that the average literacy rate of the cohort increased by 20% between 1900 and 1910. Return to page 21.

FIGURE A21: Examples of schools financed by Galician migrants in their hometowns

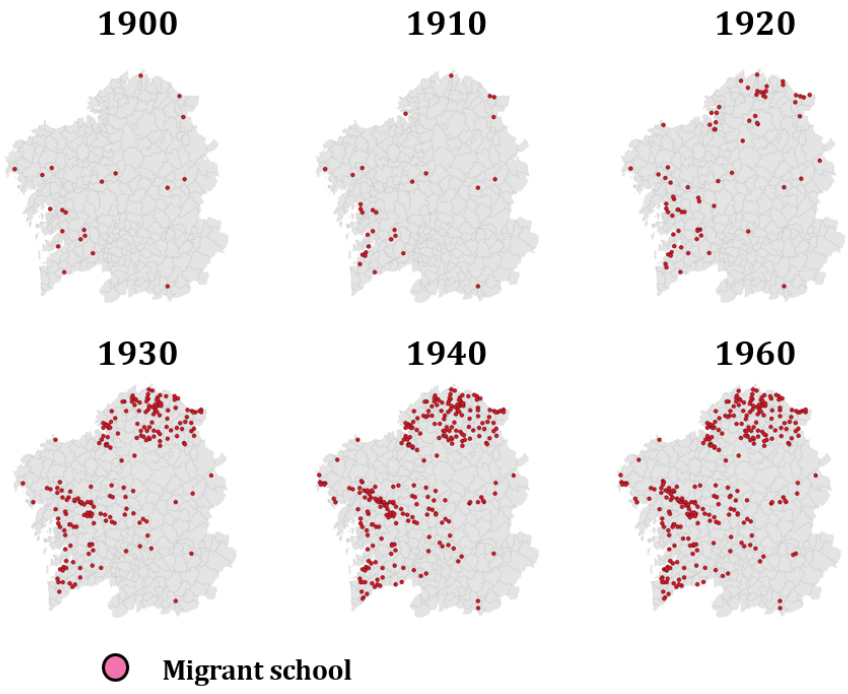
Note: Pictures extracted from the project "Luces de Alén Mar". Information of around 300 migrants' schools is available here: http://mapas.consellodacultura.gal/escolas/. Return to page 25.

FIGURE A22: Foundation of migrants' associations and schools, 1900-1960



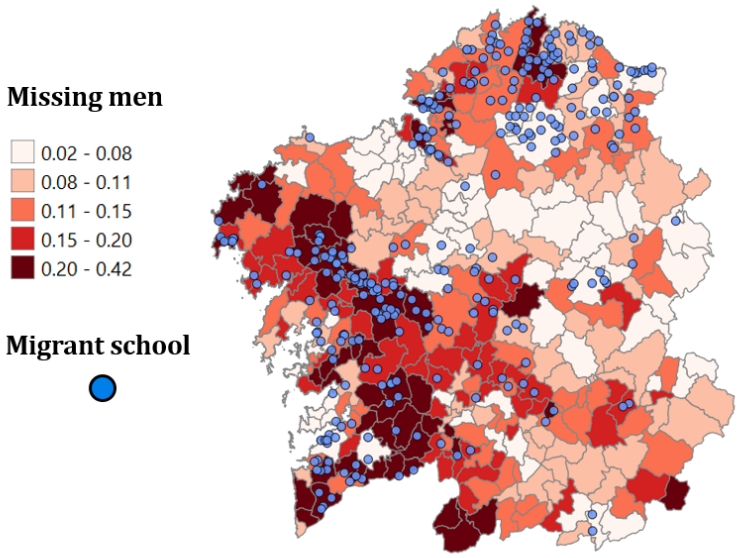
Note: The figure shows the number of Galician migrants' associations (with an educational goal) and the new schools financed by migrants in 5-year intervals. Return to page 25.

FIGURE A23: Location of migrants' schools over time, 1900-1960



Note: The figure shows the location of all schools financed by migrants at different points in time. Return to page 25.

FIGURE A24: Location of migrants' schools and intensity of missing men (1900-1930)



Note: The figure shows the average share of missing men between 1900 and 1930 and the location of all the migrants' schools founded in Galicia until 1960. Return to page 25.

Tables

TABLE A1: Characteristics of Galician migrants

	Sex, age, civil status, literacy			Occupation	
	1840-1900	1900-1930		1890-1900	1900-1930
Share of men	81%	67%	Share of farmers	94%	90%
Share younger than 10	7%	6%	Share of merchants	5%	6%
Share aged 10-15	25%	7%	Other occupations	1%	4%
Share aged 15-25	29%	40%		**Literacy**	
Share older than 25	39%	46%		1918	1925-1930
Share of single	65%	70%	Literacy of emigrants	83%	84%
Share of literate	51%	82%	Literacy of returnees	-	82%

Source: Left panel [Sex, age, civil, status, literacy] comes from Vázquez-González (1999). He compiled information of around 20 thousand migrants. Right panel [Occupation/Literacy], comes from the Official Emigration Statistics (see data sources in the Appendix). The data refers to Galician ports. *Note*: Return to page 19 or 23.

TABLE A2: Literacy rates in Galicia across cohorts and time

	Men					Women				
Ages	1900	1910	1920	1930	1940	1900	1910	1920	1930	1940
11-20	49%	57%	64%	77%	89%	24%	38%	51%	70%	85%
21-30	57%	65%	74%	81%	92%	20%	32%	47%	62%	81%
31-40	59%	62%	71%	79%	87%	16%	24%	37%	52%	69%
41-50	57%	62%	68%	76%	64%	13%	20%	28%	41%	60%
51-60	52%	58%	65%	70%	78%	9%	15%	21%	30%	47%
61-70	48%	52%	61%	66%	73%	8%	12%	17%	22%	34%
71-80	49%	51%	56%	63%	68%	9%	13%	14%	20%	29%
Average	54%	59%	67%	76%	86%	16%	26%	37%	50%	67%
Standard deviation	8%	9%	9%	12%	11%	6%	10%	15%	21%	24%

Source: Spanish population census (1900-1940). *Note*: The table shows the literacy rates of men and women in Galicia in different years and for different age groups. By following cells of a same color diagonally one can look at the evolution of literacy rates of a given cohort over time. Return to page 23.

TABLE A3: OLS regressions with different instruments and various years

	Share of missing men							
	1877	1887	1900	1910	1920	1930	1877-1930	1900-1930
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Adult sex ratio, 1860	0.34***	0.40***	0.30***	0.24***	0.17***	0.24***	0.26***	0.19***
(women/men)	[0.04]	[0.04]	[0.04]	[0.03]	[0.04]	[0.04]	[0.00]	[0.00]
Adjusted R-squared	0.52	0.51	0.30	0.18	0.11	0.19	0.46	0.32
Rainfall variability, 1830-1860	0.09***	0.13***	0.10***	0.08***	0.24	0.06***	0.46***	0.33***
	[0.02]	[0.03]	[0.03]	[0.03]	[0.02]	[0.03]	[0.00]	[0.00]
Adjusted R-squared	0.17	0.21	0.16	0.10	0.05	0.07	0.35	0.24
Distance to closest port	0.09***	0.13***	0.10***	0.08***	0.24	0.06***	0.46***	0.33***
	[0.02]	[0.03]	[0.03]	[0.03]	[0.02]	[0.03]	[0.00]	[0.00]
Adjusted R-squared	0.10	0.14	0.08	0.04	0.00	0.02	0.30	0.19
Average growth at destination (municipality level)	0.40***	-0.24	0.26***	0.10	0.06	0.06	0.04***	0.07***
	[0.13]	[0.19]	[0.08]	[0.10]	[0.06]	[0.06]	[0.02]	[0.02]
Adjusted R-squared	0.04	0.02	0.04	0.00	0.00	0.00	0.75	0.74
Average growth at destination (district level)	0.67***	-0.54	0.50***	0.09	0.04	0.18***	0.07***	0.12***
	[0.19]	[0.40]	[0.13]	[0.17]	[0.09]	[0.10]	[0.02]	[0.02]
Adjusted R-squared	0.06	0.04	0.09	0.00	0.00	0.03	0.75	0.75
Average growth at destination (province level)	1.04***	-0.42***	0.85***	-0.13	0.07	0.25***	0.06***	0.10***
	[0.27]	[0.17]	[0.18]	[0.15]	[0.13]	[0.08]	[0.02]	[0.03]
Adjusted R-squared	0.11	0.06	0.18	0	0	0.06	0.75	0.74
Municipality ans year FE	No	No	No	No	No	No	Yes*	Yes*
Observations	322	322	322	322	322	322	1932	1288

Note: The table shows β coefficients of regressing the share of missing men (in year t) on different covariates. In columns (7) and (8), I include municipality and year fixed effects depending on the level of variation of the variable of interest. Clustered standard errors at the district (columns 1-6) and municipality (columns 7-8) level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels. Return to page 19.

TABLE A4: Relationship between adult sex ratio in 1860 and other covariates

	Adult sex ratio (women/men, 1860)			
	No controls		Within province	
	β Coefficient	p-value	β Coefficient	p-value
	(1)	(2)	(3)	(4)
Altitude	-0.51	0.00	-0.26	0.00
Access to sea	0.38	0.00	0.16	0.00
Access to road	-0.28	0.00	-0.18	0.00
Access to train	-0.02	0.78	-0.07	0.13
Distance to capital	-0.04	0.52	0.02	0.72
Distance to port	-0.51	0.00	-0.15	0.06
Municipality size	-0.18	0.00	-0.14	0.01
Number of parishes	-0.20	0.00	-0.19	0.00
Population (1860)	0.18	0.00	0.02	0.61
Share of children (1860)	-0.19	0.00	-0.21	0.00
Men's literacy (1860)	0.18	0.00	0.06	0.22
Women's literacy (1860)	0.13	0.02	0.07	0.14
Wealth per capita (1855)	0.03	0.62	-0.02	0.80
Local taxes per capita (1855)	-0.13	0.02	-0.13	0.01
School enrollment rates (1875)	0.03	0.64	0.03	0.62
Schools per thousand people (1875)	0.10	0.17	0.11	0.05
Population change (1860-1900)	0.02	0.72	-0.04	0.34
Men's literacy change (1860-1900)	0.05	0.37	0.09	0.04
Women's literacy change (1860-1900)	0.04	0.52	0.05	0.28
Observations	321		321	

Note: The table shows β coefficients and p-values of multiple regressions of the adult sex ratio (1860) on covariates. As an illustration, the first row and column shows that an increase in the municipality's average altitude in one standard deviation is associated with a 51% standard-deviation decrease in the adult sex ratio (women/men) in 1860. Columns (3) and (4) show the results of similar regressions but including province fixed effects. Return to page 21.

TABLE A5: Impact of missing men on literacy rates by main destination, 1900-1930

	Literacy rates (t)							
	Men				Women			
	High RdP (1900-30)		Low RdP (1900-30)		High RdP (1900-30)		Low RdP (1900-30)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share of missing men (t)	-0.38		-0.10		-0.31		-0.40	
	[0.26]		[0.31]		[0.22]		[0.31]	
Share of missing men (10 years before)		0.54*		0.52		0.05		0.07
		[0.34]		[0.34]		[0.35]		[0.33]
Municipality and year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dependent variable	0.46	0.46	0.44	0.44	0.21	0.21	0.21	0.21
Standard deviation dependent variable	0.12	0.12	0.13	0.13	0.15	0.15	0.16	0.16
Mean share of missing men	0.13	0.13	0.12	0.12	0.13	0.13	0.12	0.12
Satandard deviation share of missing men	0.09	0.09	0.08	0.08	0.09	0.09	0.08	0.08
Observations (municipalities=321, years=4)	644	644	640	640	644	644	640	640

Note: The table shows β coefficients from estimating equations (1) and (2) by 2SLS. The share of missing men (in year t is instrumented by the economic growth at destination (in the decade preceding t) and its interaction with the proxy for pioneer migrant networks (adult sex ratio in 1860). Clustered standard errors at the municipality level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels. Return to page 21.

TABLE A6: Impact of missing men on literacy rates by initial population, 1900-1930

	Literacy rates (t)							
	Men				Women			
	High population 1860		Low population 1860		High population 1860		Low population 1860	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share of missing men (t)	-0.43**		-0.03		-0.49**		-0.35	
	[0.23]		[0.32]		[0.22]		[0.29]	
Share of missing men (10 years before)		0.88***		0.81**		0.42		0.15
		[0.28]		[0.36]		[0.28]		[0.33]
Municipality and year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dependent variable	0.47	0.47	0.43	0.43	0.20	0.20	0.16	0.16
Standard deviation dependent variable	0.12	0.12	0.13	0.13	0.16	0.16	0.15	0.15
Mean share of missing men	0.13	0.11	0.13	0.10	0.13	0.11	0.13	0.10
Satandard deviation share of missing men	0.09	0.08	0.09	0.08	0.09	0.08	0.09	0.08
Observations (municipalities=321, years=4)	644	644	640	640	644	644	640	640

Note: The table shows β coefficients from estimating equations (1) and (2) by 2SLS. The share of missing men (in year t is instrumented by the economic growth at destination (in the decade preceding t) and its interaction with the proxy for pioneer migrant networks (adult sex ratio in 1860). Clustered standard errors at the municipality level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels. Return to page 21.

TABLE A7: Impact of missing men on literacy rates by initial literacy, 1900-1930

	Literacy rates (t)							
	Men				Women			
	High literacy 1860		Low literacy 1860		High literacy 1860		Low literacy 1860	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share of missing men (t)	-0.64**		0.06		-0.44**		-0.66**	
	[0.32]		[0.29]		[0.22]		[0.30]	
Share of missing men (10 years before)		1.12***		0.52		0.53*		0.21
		[0.34]		[0.37]		[0.30]		[0.34]
Municipality and year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dependent variable	0.51	0.51	0.40	0.40	0.26	0.26	0.17	0.17
Standard deviation dependent variable	0.11	0.11	0.12	0.12	0.16	0.16	0.14	0.14
Mean share of missing men	0.14	0.12	0.12	0.10	0.14	0.12	0.12	0.10
Satandard deviation share of missing men	0.09	0.09	0.08	0.08	0.09	0.08	0.09	0.08
Observations (municipalities=321, years=4)	644	644	640	640	644	644	640	640

Note: The table shows β coefficients from estimating equations (1) and (2) by 2SLS. The share of missing men (in year t is instrumented by the economic growth at destination (in the decade preceding t) and its interaction with the proxy for pioneer migrant networks (adult sex ratio in 1860). Clustered standard errors at the municipality level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels. Return to page 21.

TABLE A8: Impact of missing men on literacy rates controlling for economic variables

	Men's literacy rates (t)					
	(1)	(2)	(3)	(4)	(5)	(6)
Share of missing men (t)	-0.11		-0.03		-0.22	
	[0.22]		[0.24]		[0.28]	
Share of missing men (10 years before)		0.57***		0.86***		0.64***
		[0.19]		[0.31]		[0.21]
Municipality and year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean dependent variable	0.48	0.48	0.48	0.48	0.48	0.48
Standard deviation dependent variable	0.12	0.12	0.12	0.12	0.12	0.12
Mean share of missing men	0.14	0.11	0.14	0.11	0.14	0.11
Standard deviation share of missing men	0.08	0.09	0.08	0.09	0.08	0.09
Observations (municipalities=321, years=4)	1284	1284	1284	1284	1284	1284

Note: Column (1) controls for GDP per capita in the province and its interaction with the distance to the capital. Column (2) controls for rainfall variability during the precedent decade. Column (3) controls for the size of exports and imports to Latin America and its interaction with the distance to the closest port. Clustered standard errors at the municipality level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels. Return to page 21.

TABLE A9: Impact of missing population and missing women on literacy rates

	Literacy rates (t)							
	All		Men		Women			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share of missing population (t)	-0.53**		-0.31		-0.58**			
	[0.25]		[0.24]		[0.28]			
Share of missing population (10 years before)		0.33**		0.62***		0.13		
		[0.17]		[0.18]		[0.19]		
Share of missing women (t)							-0.19	
							[0.37]	
Share of missing women (10 years before)								0.18
								[0.27]
Municipality and year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
KP F -statistic	5.33	14.81	5.33	14.81	5.33	14.81	1.28	4.35
Mean dependent variable	0.35	0.35	0.48	0.48	0.25	0.25	0.25	0.25
Standard deviation dependent variable	0.13	0.13	0.12	0.12	0.15	0.15	0.15	0.15
Mean share of missing	0.09	0.07	0.09	0.07	0.09	0.07	0.04	0.02
Standard deviation share of missing	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.03
Observations (n=321, t=4)	1284	1284	1284	1284	1284	1284	1284	1284

Note: The table shows β coefficients from estimating equations (1) and (2) by 2SLS but using the share of missing population or women as main explanatory variables. The share of missing population or women in year t is instrumented by the economic growth at destination (in the decade preceding t) and its interaction with the proxy for pioneer migrant networks (adult sex ratio in 1860). Clustered standard errors at the municipality level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels. Return to page 21.

TABLE A10: Long-run impact of missing men on human capital by birth cohorts, 1930s-1970s

	Birth cohorts				
	1930s	1940s	1950s	1960s	1970s
	(1)	(2)	(3)	(4)	(5)
	Panel A: Completed primary education (%)				
Average share of missing men (1900-1930)	0.35***	0.37***	0.35***	0.31***	0.29**
	[0.13]	[0.12]	[0.12]	[0.13]	[0.14]
Mean dependent variable	0.54	0.73	0.87	0.95	0.98
Standard deviation dependent variable	0.21	0.18	0.12	0.06	0.03
	Panel B: Completed secondary education (%)				
Average share of missing men (1900-1930)	0.30***	0.21***	0.03	-0.04	0.04
	[0.11]	[0.11]	[0.12]	[0.13]	[0.14]
Mean dependent variable	0.10	0.25	0.49	0.75	0.88
Standard deviation dependent variable	0.08	0.14	0.15	0.12	0.08
	Panel C: Completed tertiary education (%)				
Average share of missing men (1900-1930)	0.17	0.21*	0.21*	-0.00	-0.26***
	[0.12]	[0.13]	[0.13]	[0.13]	[0.12]
Mean dependent variable	0.02	0.04	0.07	0.12	0.14
Standard deviation dependent variable	0.02	0.03	0.05	0.06	0.05
Controls	Yes	Yes	Yes	Yes	Yes
KP F-statistic	50.79	50.79	50.79	50.79	50.79
Mean average share of missing men (1900-30)	0.14	0.14	0.14	0.14	0.14
Standard deviation average missing men (1900-30)	0.07	0.07	0.07	0.07	0.07
Observations	315	315	315	315	315

Note: The table shows β coefficients of estimating equation (5) by 2SLS using as dependent variable different measures of human capital by birth cohort (measured in 2001). The average share of missing men (1900-1930) is instrumented by the adult sex ratio in 1860. Historical controls include population size (1900), male and female literacy rates (1900) and wealth and local taxes (1855) at the municipality level. Clustered standard errors at the district level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels. Return to page 23.

TABLE A11: Long-run effect of missing men (1900-1930) on various modern outcomes (2018)

	Public schools per 1000 people	Private schools per 1000 people	Public libraries per 1000 people	Local associations per 1000 people	Educational associations (%)
	(1)	(2)	(3)	(4)	(5)
Average share of missing men (1900-1930)	0.02	0.01	0.05	-0.35***	0.44***
	[0.03]	[0.01]	[0.11]	[0.10]	[0.13]
Controls	Yes	Yes	Yes	Yes	Yes
KP F-statistic	50.79	50.79	50.79	50.79	50.79
Mean dependent variable	0.64	0.03	0.27	13.39	0.11
Standard deviation dependent variable	1.10	0.18	0.30	16.60	0.08
Mean missing men 1900-1930	0.14	0.14	0.14	0.14	0.14
Standard deviation missing men 1900-1930	0.07	0.07	0.07	0.07	0.07
Observations	315	315	315	315	315

Note: The table shows β coefficients of different cross-sectional regressions. The average share of missing men (1900-1930) is instrumented by the adult sex ratio in 1860. Historical controls include population size (1900), male and female literacy rates (1900) and wealth and local taxes (1855) at the municipality level. Clustered standard errors at the district level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels. Return to page 23.

TABLE A12: Long-run impact of missing men on beliefs about taxation and public goods (2007-2018)

	Satisfied with public education	Too little spend on public education	More taxes for better public goods (0-10)	We benefit little from taxes	We pay little taxes
	(1)	(2)	(3)	(4)	(5)
Average share of missing men (1900-1930)	0.00	-0.08*	0.02	0.00	-0.08
	[0.01]	[0.05]	[0.08]	[0.05]	[0.06]
Individual controls	Yes	Yes	Yes	Yes	Yes
Historical controls	Yes	Yes	Yes	Yes	Yes
Educational attainment	Yes	Yes	Yes	Yes	Yes
Mean of dependent variable	0.45	0.55	4.98	4.98	0.61
Standard deviation dependent variable	0.50	0.50	2.33	0.60	0.49
Mean share of missing men (1900-1930)	0.11	0.11	0.11	0.11	0.11
Standard deviation share of missing men (1900-1930)	0.06	0.06	0.06	0.06	0.06
Observations	1862	1834	1264	1918	2066

Note: The table shows β coefficients of different cross-sectional regressions. The average share of missing men (1900-1930) is instrumented by the adult sex ratio in 1860. Historical controls include population size (1900), male and female literacy rates (1900) and wealth and local taxes (1855) at the municipality level. Clustered standard errors at the district level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels. Clustered standard errors at the municipality level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels. Return to page 28.

B Summary of data sources

TABLE B1: Main historical sources

Description	Source
Main demographic data (e.g., literate, missing) 1860, 1877, 1887, 1900, 1910, 1920, 1930	Spanish population censuses http://www.ine.es/inebaseweb/libros.do?tntp=71807&L=1
Migrants' associations and schools 1900-1960	Peña Saavedra (1991), Nuñhez Seixas (1992), Naranjo Orovio (1987), Costa Rico (1989), Llorden Miñambres (1996), Díaz Sal (1975)
Galician migrant destinations 1900-1930	Vázquez-González (1991): desertion records, embarkation lists and notary documents.
Pre-migration data (e.g., wealth, taxes) 1855	Diccionario geográfico estadístico histórico de España y sus posesiones de Ultramar https://archive.org/details/diccionariogeogr10madouoft/page/n4
Educational statistics 1874, 1908, 1974	Estadística de la matrícula escolar del año académico 1874-1875, Censo escolar (1908), Estadística educativa de España (1974)
Agricultulture statistics 1865, 1891, 1917, 1962	Censo de la ganadería de España (1865), Avance de la riqueza pecuaria en 1891, Estudio de la ganadería en España (1917), Censo Agrario (1962) https://www.ige.eu/web/mostrar_paxina.jsp?paxina=002006&idioma=es
Rainfall data 1800-1930	Pauling et al. 2006 European Gridded Seasonal Precipitation Reconstructions https://www.ncdc.noaa.gov/paleo-search/study/6342

TABLE B2: Main contemporary sources

Description	Source
Demographic and educational data 1981, 1991, 2001, 2011	Spanish population censuses http://www.ine.es/dyngs/INEbase/en/categoria.htm?c=Estadistica_P&cid=1254734710990
Internal mobility 1998, 2003	Spanish population registry (Padrón) http://www.ine.es/dyngs/INEbase/en/categoria.htm?c=Estadistica_P&cid=1254734710990
Survey data on beliefs 2007-2018	Public Opinion and Fiscal Policy Survey, Centro de Estudios Sociológicos http://www.cis.es/cis/opencms/EN/2_bancodatos/catalogoencuestas.html
Register of schools and libraries 2018	http://abertos.xunta.gal/catalogo/ensino formacion/ /dataset/0257/centros educativos galicia https://rbgalicia.xunta.gal/es/rede/directorio mapa?field_biblios_prov_concello_tid=624
Register of local associations 2018	https://datos.gob.es/en/catalogo/a12002994 registro de asociaciones

Return to section 3 Data.

C Destinations and economic growth in Latin America

Galician migrants headed almost exclusively to three destinations: Rio de la Plata,[35] Cuba and Brazil. While these countries accounted for 97% of migrants in 1850-1930, their relative importance varied substantially from some areas to others and over time.

Province level data: The official migration statistics provide information on the destination of Galician migrants only for the 4 provinces and the periods 1885-95 and 1911-34.
#Name of sources #Overview of destinations and returns across provinces

Municipality level data: The only information of migrant destinations across municipalities comes from different historical sources compiled by Vázquez-González (1999). The sources include embarkation lists, desertion records, and notarial documents for the periods 1851-1860, 1901-1912 and 1920-1930. I aggregated the original data at the municipality and district level as displayed in Table C3. The data of 1851-1860 is not exploited given its limited coverage, nevertheless, the destination shares of 1851-1860 and 1900-1930 are highly correlated across districts.

[35] I use the term Rio de la Plata to refer to Argentina and Uruguay, as migrants could easily transit from one country to to the other.

TABLE C1: Official statistics of departures by province and country

	1911-1934					1911-1934			
	Total	Rio de la Plata	Cuba	Brazil	Other	% Rio de la Plata	% Cuba	% Brazil	% Other
Coruña	177,734	110,169	57,316	2,954	7,295	0.62	0.32	0.02	0.04
Lugo	141,213	75,798	61,966	1,273	2,176	0.54	0.44	0.01	0.02
Ourense	151,475	67,332	59,759	19,173	5,211	0.44	0.39	0.13	0.03
Pontevedra	151,586	100,714	22,648	23,409	4,815	0.66	0.15	0.15	0.03
Total departures	622,008	354,013	201,689	46,809	19,497	0.57	0.32	0.08	0.03
	1885-1895					1885-1895			
	Total	Rio de la Plata	Cuba	Brazil	Other	% Rio de la Plata	% Cuba	% Brazil	% Other
Coruña	69,853	26,168	35,846	5,717	2,122	0.37	0.51	0.08	0.03
Lugo	18,314	4,739	12,809	390	376	0.26	0.70	0.02	0.02
Ourense	21,235	3,446	6,999	9,877	913	0.16	0.33	0.47	0.04
Pontevedra	58,689	33,262	9,299	13,963	2,165	0.57	0.16	0.24	0.04
Total departures	168,091	67,615	64,953	29,947	5,576	0.40	0.39	0.18	0.03

Source: Return to page 11.

TABLE C2: Official statistics of returns by province and country

	1916-1934					1916-1934			
	Total	Rio de la Plata	Cuba	Brazil	Other	% Rio de la Plata	% Cuba	% Brazil	% Other
Coruña	106,589	37,799	54,884	1,522	12,384	0.35	0.51	0.01	0.12
Lugo	65,258	20,579	41,721	575	2,383	0.32	0.64	0.01	0.04
Ourense	75,107	22,388	42,470	4,796	5,453	0.30	0.57	0.06	0.07
Pontevedra	80,582	44,696	21,321	9,151	5,414	0.55	0.26	0.11	0.07
Total returns	327,536	125,462	160,396	16,044	25,634	0.38	0.49	0.05	0.08
	1887-1995					1887-1995			
	Total	Rio de la Plata	Cuba	Brazil	Other	% Rio de la Plata	% Cuba	% Brazil	% Other
Coruña	23,415	8,131	13,944	749	591	0.35	0.60	0.03	0.03
Lugo	5,299	725	4,353	68	153	0.14	0.82	0.01	0.03
Ourense	4,200	890	2,048	1,154	108	0.21	0.49	0.27	0.03
Pontevedra	20,429	12,482	3,683	3,733	531	0.61	0.18	0.18	0.03
Total returns	53,343	22,228	24,028	5,704	1,383	0.42	0.45	0.11	0.03

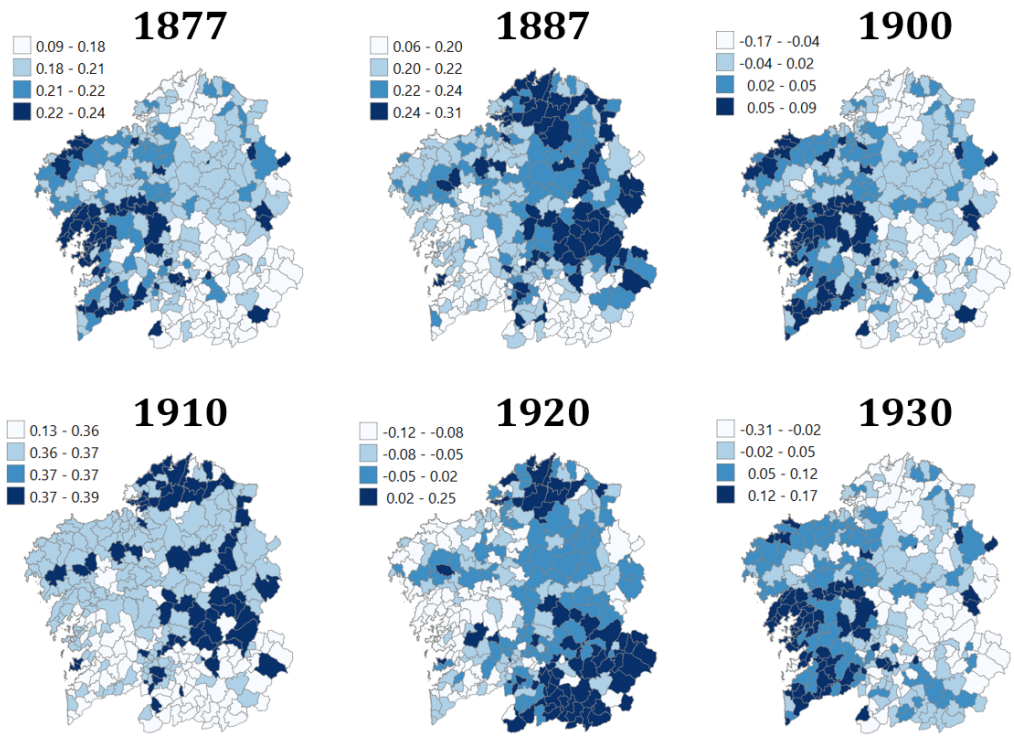
Source: Return to page 11.

TABLE C3: Main migrant destinations across Galician districts, 1900-1930

District name	Total	Rio de la Plata	Cuba	Brazil	% Rio de la Plata	% Cuba	% Brazil
Allariz	130	51	28	51	0.26	0.37	0.37
Arzua	581	478	100	3	0.18	0.81	0.00
Bande	105	62	26	17	0.25	0.59	0.16
Becerrea	346	269	76	1	0.22	0.78	0.00
Betanzos	711	571	140	0	0.24	0.76	0.00
Caldas	429	410	14	5	0.04	0.96	0.00
Cambados	887	848	31	8	0.03	0.96	0.00
Canhiza	114	95	6	13	0.08	0.79	0.13
Carballinho	193	130	46	17	0.34	0.58	0.08
Carballo	749	654	95	0	0.12	0.88	0.00
Celanova	105	62	37	6	0.44	0.51	0.05
Chantada	810	532	273	5	0.34	0.65	0.01
Corcubion	731	612	117	2	0.16	0.84	0.00
Corunha	1,059	801	258	0	0.24	0.76	0.00
Estrada	175	145	19	11	0.08	0.90	0.02
Ferrol	432	261	155	16	0.37	0.59	0.04
Fonsagrada	391	281	108	2	0.28	0.72	0.01
Xinzo	142	40	32	70	0.33	0.24	0.42
Lalin	737	533	190	14	0.25	0.73	0.02
Lugo	823	503	312	8	0.38	0.61	0.01
Mondonhedo	528	378	147	3	0.28	0.72	0.01
Monforte	231	88	142	1	0.61	0.38	0.00
Muros	184	146	36	2	0.21	0.78	0.01
Negreira	260	165	95	0	0.33	0.67	0.00
Noia	369	350	18	1	0.18	0.82	0.00
Ordes	341	231	108	2	0.32	0.68	0.01
Ourense	335	207	103	25	0.37	0.56	0.07
Ortiguera	235	66	169	0	0.72	0.28	0.00
Padron	545	503	34	8	0.05	0.95	0.01
Pontevedra	245	205	34	6	0.21	0.77	0.01
Pobra de Trives	149	63	61	25	0.51	0.35	0.14
Ponte Caldelas	145	101	4	40	0.04	0.70	0.26
Ponteareas	161	148	9	4	0.10	0.86	0.05
Pontedeume	449	240	204	5	0.51	0.48	0.01
Quiroga	132	56	73	3	0.56	0.42	0.02
Redondela	102	84	9	9	0.09	0.76	0.15
Ribadavia	130	96	30	4	0.58	0.40	0.02
Ribadeo	271	167	102	2	0.37	0.62	0.01
Santiago	571	444	36	91	0.07	0.78	0.15
Sarria	415	253	158	4	0.38	0.61	0.01
Tui	288	248	13	27	0.07	0.83	0.10
Valdeorras	253	93	134	26	0.54	0.36	0.10
Verin	455	128	47	280	0.14	0.27	0.59
Viana do Bolo	144	81	31	32	0.30	0.50	0.20
Vigo	359	308	30	21	0.15	0.81	0.04
Vilalba	484	269	207	8	0.43	0.55	0.02
Viveiro	453	303	146	4	0.33	0.66	0.01
Total	17,884	12,759	4,243	882	0.24	0.71	0.05

Economic growth at destination: The Maddison Project Database 2018 provides data of GDP per capita in the main Latin American countries over the period 1850-1930. To construct a measure of economic growth at the average destination I weight each country with the municipality level data from Vázquez-González (1999). I rely on the 1900-1930 data because it is the most exhaustive and highly correlated with that of previous periods. I construct alternative measures with data aggregated at the distrcit level (1900-1930) and the province level data (1885-1895) from the official migration statistics.

FIGURE C1: Economic growth at the average migrant destination, 1877-1930



Note: the figure displays GDP per capita growth at the average migrant destination by decade using as weights the main destinations in 1900-1930. For example, if 80% of migrants from a municipality headed to Rio de la Plata and 20% to Cuba, the GDP per capita growth of each country is weighted accordingly. Return to page 8 or 14.

D Weather shocks at origin

In 1860-1930, over 80% of Galicians worked in subsistence agriculture. As a result, weather conditions could have a large impact on households' earnings as well as on migration decisions. I rely on historical rainfall data to construct measures of local economic shocks in this period.[36]

Weather data: I use data from the European Gridded Seasonal Precipitation Reconstructions (Pauling et al. 2006) which contains yearly information of precipitation at a high-degree resolution from 1500 on wards. Its fine resolution (0.5° x 0.5°) yields a grid of 24 cells in Galicia with an average approximate size of 50km^2 (53km x 43km).

Rainfall shocks: I construct a measure of rainfall variability in the harvest season by computing the standard deviation of precipitation in each grid cell and decade:

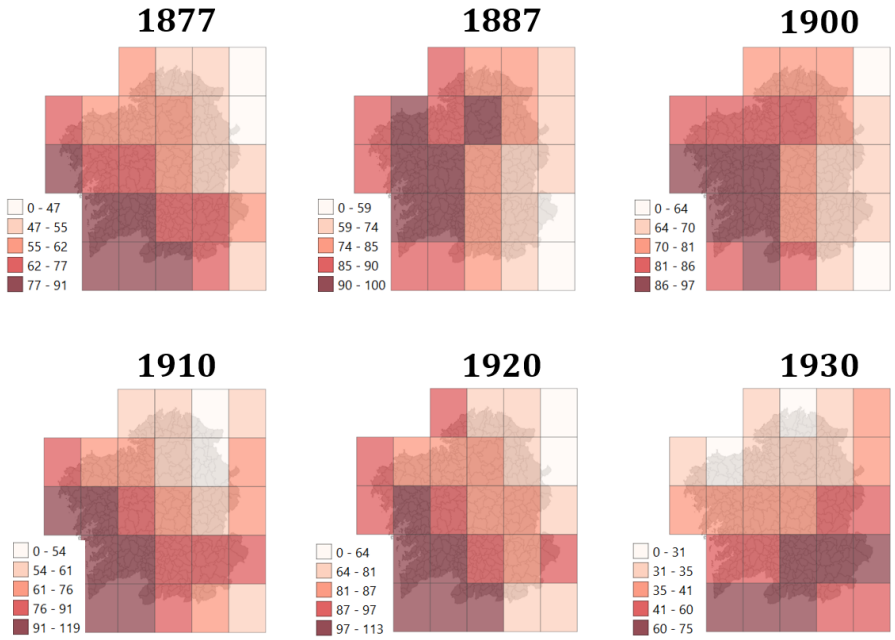
$$SDR_{gt} = \sqrt{\frac{1}{10}\sum_{i=t-9}^{t}(r_{gi} - \bar{r}_{gt})^2}$$

where r_{gi} represents total precipitation in cell g and year i during the Fall and $\bar{r}_{gt}$ the average total precipitation in cell g during the decade preceding t ($\bar{r}_{gt} = \frac{1}{10}\sum_{i=t-9}^{t} r_{gi}$).
Note that since I will use this measure in a panel setting, the municipality fixed effects will capture the fact that some areas have higher variability than others on average. SDR_{gt} should then be interpreted as changes with respect to the average variability of a location.
In the cross-sectional analysis I compute the standard deviation difference with respect to the mean in each grid cell and decade as follows:

$$ZR_{gt} = \frac{1}{10}\sum_{i=t-9}^{t}\frac{|r_{gi} - \bar{r}_{gt}|}{SDR_{gt}}\,.$$

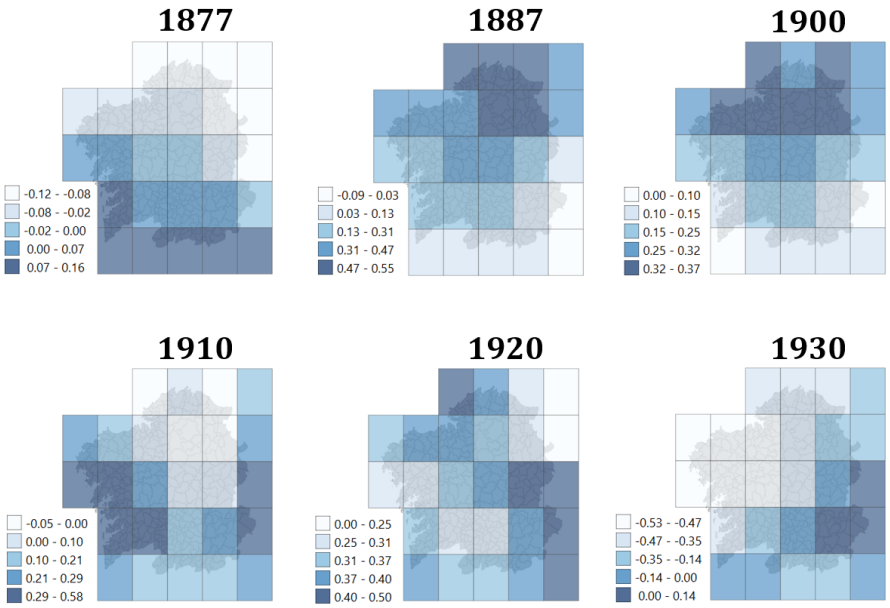
[36] While both temperature and rainfall are important determinants for agricultural yields, there is much more variation in rainfall than temperatures across the different areas of Galicia.

FIGURE D1: Rainfall variability in the previous decade, 1887-1930



Note: the figure shows the variability in rainfall in every decade (standard deviation). Darker colours represent higher variability.

FIGURE D2: Rainfall variability in the previous decade (relative), 1887-1930



Note: the figure shows the variability in rainfall in every decade relative to the standard in each location (z-score), using as reference the average precipitation during 1860-1930. Darker colours represent higher variability.

Alternative data & validation: Since maize is harvested at the end of September/ beginning of October my weather shocks measure relies on rainfall variability in the Fall. Yet, a priory it is uncertain which periods or shock measures could be more relevant to explain migration. To give additional support to this choice I rely on data from the last release of CRU TS (CRU TS v. 4.03) which contains monthly information of rainfall from 1901 onward at the same resolution (0.5° x °). This monthly data allows me to compute multiple measures of rainfall shocks (e.g., variability/excessive/insufficient rainfall in each month, group of months, etc.). I therefore constructed 240 different variables of rainfall shocks and implemented a Lasso procedure particularly adapted for panel settings, the Cluster-Lasso (Belloni et al., 2016). Using this procedure I found that the variables that better explain changes in the share of missing men in 1910-1930 were rainfall variability in October and excessive precipitation in October (e.g., number of months in the decade with rainfall higher than one standard deviation above the mean).

E Distance to the port and migrant networks

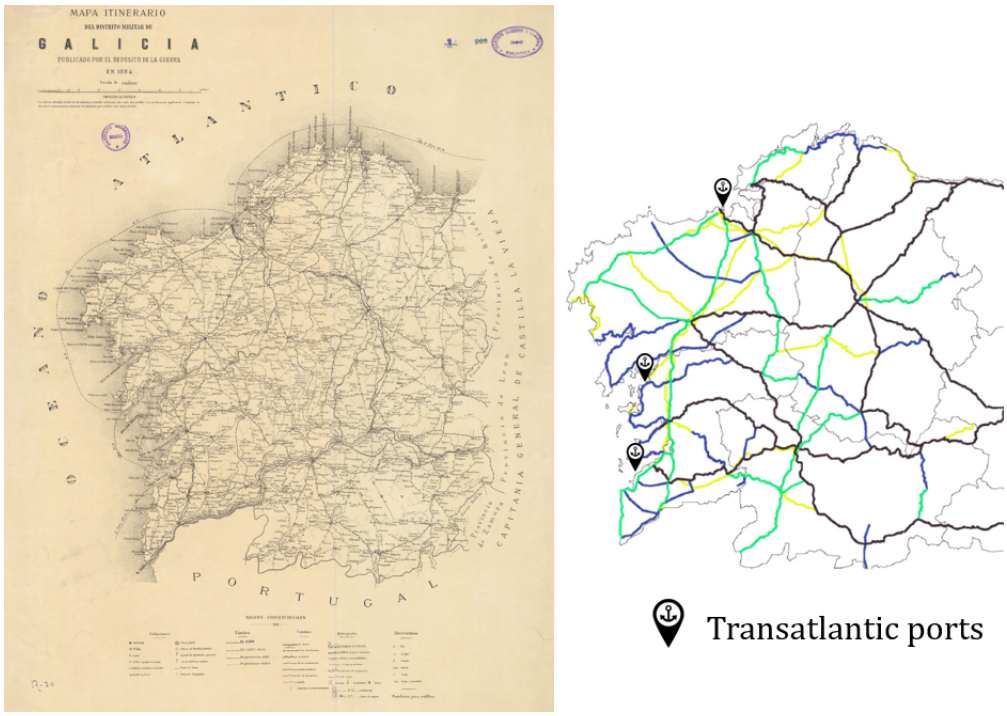
Background: The first migration wave in the 1850s affected mostly areas closer to the transatlantic ports. Local agents working for shipping lines took an active role spreading information and facilitating migration in rural municipalities with no prior experience, expanding their catchment areas inland over time.[37] However, the diffusion of information was hampered by the poor transportation infrastructure that had historically isolated municipalities from each other. In the second half of the nineteenth century, the distance to a transatlantic port was a major determinant of migrant flows, not only due to the information channel but also to the high transportation costs implied by individuals having to walk until the ports.

Computing the distance to the closest port: The most direct way to measure the distance to a port is to estimate the linear shortest path.[38] Yet, in a context with numerous natural barriers (e.g., mountains, rivers, etc.) as in Galicia, there can be an important mist-match between "as the crow flies" distances and distances following actual roads. To account for this I manually draw and geo-referenced the network of the primary walking paths in the 1880 using historical maps of Galicia. Using this network I can estimate the walking distance in km to the closest transatlantic port. To test the validity of this measure, I also rely on the shortest walking path to these ports today based on Google maps.

[37] These agents were important to familiarize individuals with the opportunities abroad, provide them with the necessary documents (e.g., passports, tickets, etc.) and even finance the trip.

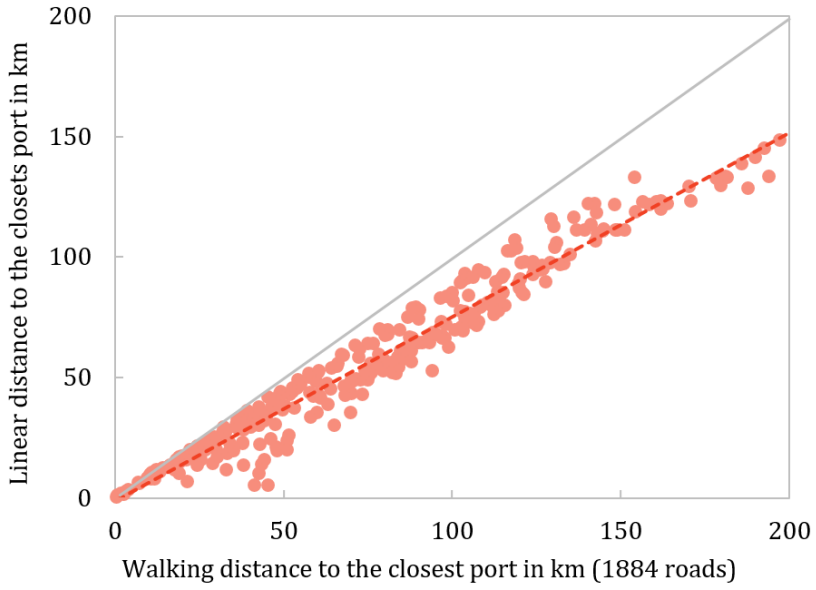
[38] One example is Karadja & Prawitz (2019) where they use the great-circle haversine formula to take into account the shape of the Earth.

FIGURE E1: Transatlantic ports and primary roads in Galicia in 1884



Note: The symbol of an anchor indicates the ports shipping boats to the Americas. From North to South: Coruña, Carril/Vilagarcía and Vigo.

FIGURE E2: Comparison between linear and walking distance



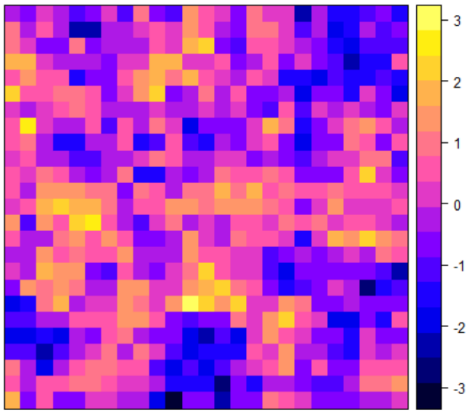
Note: The figure shows the linear distance (in km) to the closets transatlantic port and the actual walking distance (in km) following the road network in 1884.

F Spatially correlated errors (Kelly, 2019)

In a recent working paper, Kelly (2019) highlights that spatially correlated errors can lead to distorted significance levels in spatial regressions. In particular, he shows that randomly generated noise variables with a certain degree of spatial correlation can be significant predictors of historical and contemporary (spatially correlated) variables. Revisiting 27 papers published in top journals, he concludes that: *"only about one quarter of the persistence results that we examine are robust after we take account of the possibility that their regressions might be fitting spatial noise."*

The paper proposes a simple procedure to investigate the role played by spatial noise in influencing the significance of regression results. In a nutshell, it consists in generating random noise variables with spatial correlation and using them as explanatory or dependent variable in the analysis. Figure F1 is an example of a spatially correlated random noise variable in a square grid of 35x35 cells.[39] I generate one thousand grids like Figure F1 and mapped them to my database of Galician municipalities to perform the tests suggested.

FIGURE F1: Randon noise variable with spatial correlation



Note: The figure shows a random variable in a 35x35 grid, with a spatial correlation of range 2.

I first regress the share of individuals with secondary education in 1981 (the dependent variable in Figure 13 and Table 5) sequentially on these thousand random noise variables with spatial correlation. I obtain a significant coefficient at the 5%, 1% and 0.1% level in 12.9% of the cases, 4.6% and 0.7% respectively. These point estimates are extremely small and heavily centered around zero (95% of them range between -0.004 and 0.004) while my ('true') estimated effect is 0.14 (Table 5) .

[39] I use these dimensions so that they correspond approximately to the size of municipalities. I also try generating alternative maps with different ranges of spatial correlation and obtaining similar results.

Secondly, I replicate equation (5) a thousand times using as dependent variable with a randomly generated spatially correlated noise variable and instrumenting the share of missing men (1900-1930) with the adult sex ratio in 1860. The results are displayed in Figure F2 in comparison with those obtained by Kelly (2019) for other studies. Overall, despite a substantial spatial correlation in the explanatory variable and instrument, the robustness of my analysis fares relatively well compared to previous work. It is important to stress once again that this comparison only refers to the degree of potential distortion in significance levels, and not to the size of the estimates effects, that in my case are very large. The point estimates in Table 5 range from 0.14 to 0.46. Having noise as a dependent variable and performing my instrumental variables strategy one could obtain a point estimate higher than 0.14 in less than 17.5% of the cases and as large as 0.46 in less than 2.3% of the cases.

FIGURE F2: Ability of persistent regressions to explain random noise

Note: The figure replicates Figure 11 in Kelly (2019) adding the results for this paper. It represents the fraction of simulations where the persistence variable explained spatial noise at a significance level of 0.001 or 0.0001.

G Anecdotal evidence

This section gathers direct anecdotal evidence from Galician migrants about the importance of education, migrants' associations and schools, and other aspects. The original quotes in Spanish/Galician can be found here.

Migrants realization of the importance of education [Return to page 27].

1. "Each of them (migrants) recalls the poor education received and stands up, proclaiming that they want that future generations are nurtured with the power of the most useful and needed knowledge to go around the world." - El emigrado, A Estrada, no. 7-7-1923

2. "It is crucial that we realize the deficient education we received, and acknowledging how difficult it is out there without it, that we transmit it to our brother and sisters so that can enjoy better lives."- «El Progreso de Coles». Pro galicia. Vol. I, n. 2, 1912 (p.19)

3. " [...] understanding the need to educate people at origin [...] not only to facilitate basic knowledge to those migrating, but to diffuse the power of education back home, the only way to free ourselves from exploitation and elites' abuses [...]" - Álvarez Gallego (1906, p.313)

4. "[...] In Cuba they never asked me if I knew how to pray, they always asked if I knew how read and write and since I said yes, I got better and better jobs." - Pereiro (2012:3)

5. "Schools in Galicia are the result of love and patriotism from good heart people who know the power of education to conquer good, wealth and freedom. And that is what (migrants) want for their brothers and sisters, [...] they dream that in the land they yearn for, a more educated society will lead to political and economic progress." - Nova Galicia, n. 472 de 1915

6. "[...] With what will be taught there (migrants' schools) people will have enough to fare better than those migrating in the past without knowing how to read, write or count."- «Mi opinión». El Faro de Veiga, Colegio San Adrián, 1914 (pp.38-39)

7. "Schools are the best investment, with more immediate and future outcome, that we migrants can do for the economic revival of Galicia - [...]" - Alonso Rios, 1929

8. “[...] We Galician migrants, thanks to the educational associations we are supporting with so much effort, we aim to bring to our hometowns the highest education possible, educating our children and building schools not only to prepare future generations for migration, but with the aim to take advantage of all the wealth our region has. [...]” - José Loureiro García: «Siempre adelante», Galicia. Revista Regional Ilustrada. Órgano de la Colonia Gallega y Sociedades Regionales de Cuba, A Habana, no. 40, 30 de setembro de 1922, s.p.

   Explicitly transmitting higher value for education.[Return to page 27].

9. “ [...] The driving force of social transformation in our country will be knowledge and the education of our people through newspapers, magazines and books, conferences and meetings, but most importantly, schools.” - Constantino Horta: « Cultura y regionalismo II», Ecos del Eume, Pontedeume, no. 181, 26 de maio de 1912

10. “It has been two years since the foundation of this association [...] aiming at contructing and financing [...] schools with the best conditions and all the pedagogical innovations needed [...] If we want a free country and responsible citizens, we have to start by disseminating education among the young generations and making our brothers and sisters understand their rights [...]”-Junta directiva Sociedad de Instrucción Alianza de Vilamarín (1912, Habana)

11. “–¿Emigration? - asks a friend. But, you think emigration is evil? All the money Galicians make in the Americas comes back here, fueling our industry. And it is not just money what migrants make circulate between us, it is the spirit of progress and tolerance [...], Galician migrants are modernizing Galicia. Do not be a fool. Emigration is good.” - Julio Camba: «La emigración es un bien», Eco de Galicia, A Habana, no. 115, 21 de marzo de 1920, (p.8)

12. “In Cuba my father worked in the rail company because he knew how to read and write, and he always said that made him improve [...] That is why he was obsessed that we could not miss a single day of school. [...]” - Extraído de la vida de Eurisna Campos Cudeiro (Amoeiro, OU).

13. “In their letters, often written by others’ hands, what do they (migrants) tell you? [...] Do not they advise you, with their sad regret, that you should go to school? So yes, my beloved neighbors, I encourage you, leaving the laws aside, that you should send your kids to school as much as possible.”- «Un bando que honra a un pueblo y a su alcalde», Galicia. Revista del Centro Gallego, Bos Aires, no. 216, decembro de 1930, (p.6)

14. “[...] Greetings to my beloved brothers and sisters, I hope the work is not to harsh, and to Pepe and all the others at home, make sure Aquilino studies well.” - Cayetano Borrajo (La Habana, 21-7-2018)

15. “I know sometimes you have no option but missing class because you need to help your mothers at home, but do please ask them to let you come to school, to not miss class; tell them that we do not live only on present and palpable things.” - Discurso de clausura do curso académico pronunciado pola mestra da escola de nenas de Ares, creada pola Sociedad Instructiva Redes-Caamouco de La Habana - Memoria de la Sociedad Instructiva Redes-Caamouco, 1922

16. “I would like to ask public authorities to enforce the educational law that obliges children to attend school, and that kids are examined by doctors to prevent the spread of contagious diseases .” - Conclusión do Primeiro Congreso da Federación de Sociedades Gallegas Agrarias y Culturales, Buenos Aires, El Tea, Puenteareas, 1922.

17. “CHAPTER ONE NAME AND AIM OF THE ASSOCIATION [...] Article 5 We urge the corresponding public authorities to make education compulsory between the ages 7 to 14, setting penalties for parents who do not comply to these rules.” - Estatutos de la Sociedad Pro-Escuelas en Paradela y sus Contornos. Ayuntamiento de Meis en Pontevedra, Buenos Aires, Imp. F. Vilanova, 1912, (pp. 3-4)

18. “WE RECOMMEND to all families that, [...] either individually or collectively, try by all means to avoid the payments to the church and elites and in turn use that money to purchase materials for schools, to pay teachers and everything that could improve education [...]” - Teo. Órgano Oficial de la Federación Residentes de Teo en Sud América, Bos Aires, no. 41, 15 de setembro de 1913, (p.1)

Praising the role of migrants’ associations.

19. “From all the exceptional actions that the Galicians who are missing did, the greatest without doubt, has been the creation of associations that gathered funds for the construction of schools, and for the diffusion of culture in their home country.” - «Una gran obra de cultura para Galicia», Vida Gallega, Vigo, 1912.

20. “I remember that schools, for long time, were critized as some said they encouraged emigration. They said they were centers to train future migrants. Yet, the reality was that Galicia lived in misery and backwardness. The first and immediate objective of a migrant was to contribute to improving the village or parish where they grew up, among other things, because that village or parish was the only place of their own, the only part of Galicia they knew.” - M. Meilán en: Samuelle, C.,80:1993

21. “This institutions (migrants’ associations) will rise the culture of Galicia to an unprecedented height, and improve the doncidionts of our migrants, facilitating their labor market insertion in the Americas, and educating the people in our fields, rising their moral and public wealth.” - Alianza Aresana de Instrucción, Vida Gallega, 1911.

22. “The wave continues, the enthusiasm is everywhere, educational migrant associations are spreading so fast that very soon will be no Galician municipality without a school constructed and financed by our brothers emigrated to the Americas.” - Sueva, Habana, 1910.

23. “Galicia, for its peculiar geography, is in real need of diffusing basic education. We need to found many schools, loads of them -good and less good-, to bring education to all little towns in the mountains and the valleys [...] War against illiteracy, that’s what is most pressing thing; Galicia without a single illiterate, that should be the aim of all our aspirations.” - Nicasio Pajares: « El magno problema», Fomento de la Instrucción Gallega. Órgano de la Sociedad Pro escuela en Bandeira, no. 4, 15 de novembro de 1909, (pp. 2-3)

24. “[...] I will never have enough words to celebrate the workers from Ares, my beloved colleagues, whom with their modest earnings, barely enough to sustain themselves in cities as expensive as Havane, they have managed to found and finance a school for boys and soon one for girls in their hometowns, equipping them with all materials, and with all the hygiene conditions and modern necessities [...]” - Manuel Curros Enríquez: «Bien por los aresanos!», Galicia, A Habana, no. 30, 25 de xullo de 1905, (p. 8)

Migrants hard work and high savings.

25. “Most of the people were foreigners who came to earn some money and return. Me too, I thought I would spend here more than three or four years, I thought I would come back soon [...] But while I was in the Americas, everything was saving. It was a really different life.”-Buenos Aires, febrero 1994. Interviews from Samuelle-Lamela (2000)

26. “[...] Galicians migrated to earn some money and had no option but saving a lot. Generally they could come here to marry spend three or four years and come back. You could identify Galicians by their skin color, so pale because of spending all their time working behind the counter.” - Ricardo Flores, Buenos Aires 1994. Interviews from Samuelle-Lamela (2000)

27. “I went to work in a restaurant, from 5 in the morning to 9 at night, only resting one day pwer week and one hour for lunch and another for dinner.” - Montevideo, agosto 1988. Interviews from Samuelle-Lamela (2000)

About conditions in migrants’ schools.

28. “This is the sshool ”Unión Hispano Americana Valle Miñor´´. We have more 300 kids from 7 to 15 years old. Education is completely free; and the poor, who are more than 160, are offered free books, clothes [...] During the most favorable season, we offer night lectures, having 60 young adults attending. These classes are also free of charge.” - «Escuela Americana», Boletín de la Unión Hispano Americana Pro-Valle Miñor, Bos Aires, no. 40, xaneiro de 1913, (p.1)

29. “[...] Education will be free o charge for everybody who needs it, we will give them books, ink, paper, and all necessary materials. To enroll in this school the only conditions is to be 6 years or older and [...] All students that without justified reason (such as sickness, bad weather, impassable roads) do not attend for a week will have their parents notified and will be expelled from the school. Everybody who is not from Mera is allowed to come here, send their kids, brothers or anyone in need of education.” - Reglamento para el interior del Colegio de la Sociedad de Instrucción PUENTE DE MERA

30. “The basics (of teaching practices), according to the aspirations of the founders of or association “Ferrol y su Comarcar”, will be to use the most intuitive and practical methods possible to transmit the basic knowledge in modern life [...]” - Sociedad Ferrol y su Comarca (1911, p.5-6)

31. “Making education rational, practical, harmonic and progressive, training our students to think, reflect, and to do research on their own [...] Considering the skills and conditions of every kid, making our lessons clear and precise, using all kinds of intuitive methods, relying on materials in our class, [...], so that education is entertaining, interesting, varied

and likeable, making kids put to test the knowledge they have acquired [...]" - Boletín de la Unión Hispano–Americana pro Valle Miñor (1914)

Criticism of migrants' schools because they might foster more migration.

32. "It helpful and interesting the endeavor of Galician migrants in America financing schools in Galicia. But, would it not be more fruitful and a better example if they led the State alone to build the schools, being its responsibility? [...] There are also some of those schools that, with good intentions, end up being dangerous and lethal, since they are centers where Galician identity is lost and were they instill on children the desire to emigrate, instead of transmitting them love for what is ours." - Antón Villar Ponte: «Cartas da vila. A civilización», A Nosa Terra, A Coruña, no. 96, 5 de agosto de 1919, (p.5)

33. "Meanwhile in our country, nothing is done for the "Galician" school, migrants, with a wrong approach but good intentions, waste our money in financing schools that have a bad orientation for Galicia. Those schools supported with money from the Americas do not only make children lose their Galician identity, they even guide students towards migrating, they even adopt history and geography books from these countries." - «A creación da escola galega, como toda laboura galeguizante debe contar conorganizacións que cordinen e metodicen a acción de todol-os galeguistas», A Fouce, Buenos Aires, 1931

34. "[...] Galician in the Americas finance schools, that is true. There, Galician kids learn what is good and evil, what they should like and dislike, they learn geography, language, history, all of that without the spirit of Galicia. The schools financed by migrants have the same orientation as the schools financed by the State and thus collaborate to the loss of Galician identity [...]" - Xan de Pazos: «Algunhas refreusiós ós galegos da América», A Fouce. Periódico Galego, Bos Aires, no. 3, xuño de 1926, (p.1)

35. "The problem is that too much money is invested in the construction of these schools, to the detriment of the main aim which is teaching and education [...]" - M. Nodar Magán: «Un consejo», El Emigrado, A Estrada, no. 22, 16 de xuño de 1921, (p.1)

Chained migration and importance of contacts.

36. "Those standing out were the ones emigrating to Cuba at the end of the nineteenth century and coming back looking wealthy. That instilled in others the desire to get rich themselves.

Above all, they were different for their appearance: they came back with a car, dressed like the elite and wearing many golden things."

37. "I had three seasons that motivated me to leave. We are a numerous family struggling financially because only my father had a formal salary, while the rest of us worked in the field [...] Something else that made me dream was when I saw other migrants coming back from the Americas with modern cars, and they arrived and refurbished their houses, and people said "Look the *Brazilian* has arrived, full of money!" And I was young and with that hope though "What am I doing here?" [...] And thirdly I already had there two older brothers and one uncle. They had already settled and had money. They sent me an invitation letter and I went as a mechanic [...]"-Luis Vilariño Garrido (Brazil)

38. "In my time it was very simple. Almost always there was a pionneer, he would get a job, settled, and just after, he would begin calling others back home: brothers, friends, cousins, neighbors [...] relationships between people from the same place were very important, because besides helping each other, it was the way to access the labor market." - Manuel Meilán, interviewed by Samuelle-Lamela (2000)

39. "Almost everybody had some relative here (in Latin America), and in that moment of high migration there was plenty of work here, and I think they believed that relatives from Spain would work harder that having employees from there." - Montevideo, octubre 1986. Interviews from Samuelle-Lamela (2000)

40. "Far from home, the first thing they did was to look for a neighbor, recreating their parish of origin there, and that's how they founded local associations. Those who succeded came back to Galicia and brought other neighbors to work in some business (in Latin America). That is how there were bakers from an area, builders from another, taxi drivers from Chantada, barman from Ourense, choffers from Bergantiños, butchers from Val Miñor [...]" Montevideo, 10 de octubre 1986. Interviews from Samuelle-Lamela (2000)

41. "My grandfather and uncle had come earlier and they always told me: from Argentine to paradise."- Buenos Aires, marzo 1994. Interviews from Samuelle-Lamela (2000)

Other.

42. "The only trade between Galicia and the colonies is precisely migration to the latter [...]"- Spain general correspondence, FO 72/702, 25/3/1946. [cited by Vázquez-González (1999)]

43. "Excluding a minority that comes back to Spain within a year, most of them stay two or up to five years and then come back [...] The idea that there is short-lived circular migration to the republics of Plata is simply ridiculous."- Consejo Superior de la Emigración, 1916. [cited by Vázquez-González (1999)]

44. "From Galicia only men emigrate. Galician women very rarely come to Portugal, they stay home raising and educating their children."- As Farpas, Ortigâo, R. 1950. [cited by Vázquez-González (1999, pp.107)]

H Brief description of the Galicia economy, 1800-1930

During the XIX century and until 1920, Galicia was the poorest region of Spain. In this period, income per capita levels were around half of the country's average (Rosés, Martínez-Galarraga and Tirado, 2010), the vast majority of the population worked in the primary sector (close to 90%) and urbanization levels were very low (less than 10% of municipalities had a population above 10 thousand inhabitants).

As described by Alonso-Álvarez (2010), most land was owned by landlords (´´fidalgos") but exploited by peasants under long-term concessions ("contratos forales"). While these agreements made peasants *de facto* quasi-proprietaries, they were obliged to pay a yearly in-kind contribution ("foros") on top of other taxes to the Church and local elites. This property system influenced the type of plots (small and geographically dispersed) and farming practices (extensive use of the plough and little investments in machinery or fertilizers). The main crops were maize, wheat and rye, whose production was devoted to household consumption and the local market. Livestock was also very common, contributing to producing natural fertilizer and being a crucial link in agriculture production. Communal lands, such as forests and meadows, were exploited for grazing, to obtain wood and other inputs (e.g, gorse for compost). Overall, Galician agriculture was very inefficient and harvests barely supported subsistence level.

The low agricultural productivity forced households to engage in supplementary activities such as rural domestic industries (e.g., fish salting, mule driving, blacksmithing, weaving and tanning) and seasonal migration to the harvest in Castilla and Portugal (Farías, 2010; Palmás, 1976). These activities did not require specialization and could be carried out during periods of less intense agricultural work. The output produced by rural industries could be sold in other parts of Spain through the seasonal migrations (Alonso-Álvarez, 2010).

A central aspect of the Galician economy was the high degree of population dispersion and the poor transportation infrastructure. Population was scattered into several settlements ("parroquias") that constituted a local market, with an average of twelve settlements per municipality. As a result, peasants did not need good roads for their daily activities but relied instead on a comprehensive network of ground walking paths (Barreiro Gil, 2001). Yet, the deficient transportation infrastructure became an obstacle for establishing an integrated regional market and contributed to the economic backwardness of Galicia (Correa, 2008; Barreiro Gil, 2001). The bad state of roads, especially after adverse rainfall events, increased transportation costs making it unprofitable to

sell products far away (Correa, 2008) and leading to a system in which rural markets operated in autarchy. The exceptionally high number of fairs in Galicia (more than 4000), rather than economic dynamism, exemplifies the atomization of the market (Barreiro Gil, 1983; Correa, 2008). The maintenance of paths depended on municipalities, but since neighbors already supported a high tax burden, it was very difficult to increase taxes (Barreiro Gil, 2001). Note that although dismantled by the nineteen century, institutional barriers such as tolls in crossing points had also contributed to the isolation of municipalities and of Galician from the rest of the country.

Thorough the nineteenth century, the Galician economy evolved very slowly. While the State progressively replaced the Church and landlords as main tax collector, fiscal pressure did not decrease. The modest textile industry that had expanded during the first decades of the century went bankrupt during the 1830s. Other agricultural crises would occur in the following decades. The canning industry emerged in the mid-century, driving other related sectors (e.g., fishing, oil, tin, and machinery) but was marginal in terms of employment (Alonso-Álvarez, 2010). Cattle was the main and almost sole Galician export, raising rapidly in 1860s but collapsing after the closure of the British market at the end of the nineteenth century (Barreiro Gil, 1983). Most railroads opened during the 1870s and 1880s, connecting Galicia with the rest of Spain. The access to railroads was crucial helping redirecting the fishing and cattle sectors towards the domestic market during the twentieth century (Barreiro Gil, 1983). It also contributed to a more integrated regional market and to the reduction of price disparities.

In the 1900-1930 period, Galician agriculture experience several changes that made it more productivity. First, most farmers became full owners, being thus redeemed from paying rents (Alonso-Álvarez, 2010). Getting access to the land was crucial for the adoption of technological innovations such as better seeds, chemical fertilizers, machinery, etc. This innovations spread very rapidly thanks to agrarian unions and the access to credit (Fernández-Prieto, 1997). For instance, by the 1930s the thrasher had completely replace the traditional threshing mallets. One of the greatest changes occurred in the sector of ranching thanks to the introduction of hygiene measures, disease control and new breads of cattle (Alonso-Álvarez, 2010). The agricultural sector became more integrated in the Spanish market with cattle and wood being the main products. New industries related to fishing activities expanded during the 1920s, including shipyards, sawmills, metallurgy, etc. These economic changes fostered an incipient urbanization and encouraged people to move from rural areas to the emerging cities. Nonetheless, Galicia remained being one the Spanish regions more underdeveloped with income per capita levels of 60% of the average in 1930.

Appendix References

ALONSO ÁLVAREZ, L. 2010. La economía de galicia, una panorámica, c. 1750-2010. *Historia contemporánea* .

BANDIERA, O., RASUL, I., AND VIARENGO, M. 2013. The making of modern america: Migratory flows in the age of mass migration. *Journal of Development Economics* 102:23–47.

BARREIRO GIL, M. J. 1983. La generalización de la produccion de mercancias y la modernizacion productiva de la agricultura en galicia, 1876–1976. *Revista de Historia Economica-Journal of Iberian and Latin American Economic History* 1:133–146.

BARREIRO GIL, M. J. 2001. El atraso económico y el sistema de transporte terrestre en la galicia del siglo xix. *Revista Galega de Economía* 10:0.

CORREA, H. S. 2008. Obstáculos estructurales al desarrollo de los intercambios comerciales en la galicia de la edad moderna. *Ohm: Obradoiro de Historia Moderna* .

COSTA RICO, A. 1989. Escolas e mestres. a educación en galicia: da restauración á segunda república. *Santiago de Compostela: Xunta de Galicia* .

DIEZ-MINGUELA, A., MARTINEZ-GALARRAGA, J., AND TIRADO-FABREGAT, D. A. 2018. Regional income inequality in spain 1860–2015, pp. 81–103. *In* Regional Inequality in Spain. Springer.

FERNÁNDEZ-PRIETO, L. 1997. Selección de innovaciones en una agricultura atlántica de pequeñas explotaciones. galicia 1900-1936. la adopción de las trilladoras mecánicas.

KELLY, M. 2019. The standard errors of persistence.

LLORDÉN MIÑAMBRES, M. 1998. Las asociaciones de inmigrantes españoles en américa. algunas respuestas a los desequilibrios y carencias de la emigración a ultramar. *Exils et migrations ibériques au XXe siècle* 2:79–130.

NARANJO OROVIO, C. 1988. Del campo a la bodega: Recuerdos de gallegos en cuba (siglo xx). *Sada, Ediciós do Castro* .

NÚÑEZ SEIXAS, X. M. 1998. Retornados e inadaptados: el" americano" gallego, entre mito y realidad (1880-1930). *Revista de Indias* 58:555–593.

PALMÁS, R. 1976. La emigración. *FABRA, Gustavo, et al. Los gallegos. Madrid, Istmo* .

PAULING, A., LUTERBACHER, J., CASTY, C., AND WANNER, H. 2006. Five hundred years of gridded high-resolution precipitation reconstructions over europe and the connection to large-scale circulation. *Climate dynamics* 26:387–405.

PEÑA SAAVEDRA, V. 1991. Éxodo, organización comunitaria e intervención escolar: la impronta educativa de la emigración transoceánica en Galicia. Xunta de Galicia, Secretaría Xeral para as Relacións coas Comunidades Galegas.

RICO, A. C. 1989. Escolas e mestres: a educación en Galicia: da Restauración á IIa República. Servicio Central de Publicacións, Consellería da Presidencia e ....

ROSÉS, J. R., MARTÍNEZ-GALARRAGA, J., AND TIRADO, D. A. 2010. The upswing of regional income inequality in spain (1860–1930). *Explorations in Economic History* 47:244–257.

SAMUELLE LAMELA, C. 2000. La emigración gallega al Río de la Plata. PhD thesis, Universidad Complutense de Madrid.

SÁNCHEZ ALONSO, B. 1995. Las causas de la emigración española, 1880-1930. Alianza Editorial Sa.

VÁZQUEZ GONZÁLEZ, A. 1999. La emigración gallega a américa, 1830-1930. *Memoria de doctorado inédita, Facultade de Ciencias Económicas e Empresariais, Universidade de Santiago de Compostela, Santiago de Compostela* 2.