Migration, Population Composition and Long-run Economic Development: Evidence from Settlements in the Pampas*

Federico Droller†

August 2015

Abstract

This paper analyzes the impact of population composition on long-run economic development, by studying the European mass migration to Argentina in the late nineteenth century. I use an instrumental variables approach that assigns immigrants across counties by interacting two sources of exogenous variation: the availability of land for settlement and the arrival of Europeans over time. Counties with historically higher shares of European population currently have higher per capita GDP, higher education rates and a greater proportion of skilled workers. I show that the effect is linked to the process of industrialization and the level of human capital.

Keywords: Economic growth and development, Human capital, Literacy, Industrialization, Migration.
JEL Codes: O11, O14, O15, O47, N16, N66.

*I am grateful to Pedro Dal Bó, Ross Levine and David Weil for their guidance and advice. Ernesto Dal Bó, Vernon Henderson, Oded Galor, Blaise Melly, Florencia Borrescio Higa, and participants at LACEA-PEG 2012, Utrecht's Deep Causes of Economic Development Conference 2014 and other seminars provided valuable comments and suggestions.

†Department of Economics, FAE - University of Santiago de Chile. Av. Lib. B. O'Higgins 3363, Estación Central, 9170022. Santiago, Chile. Phone: +56 2 2718 0841. federico.droller@usach.cl.

1 Introduction

Why have some regions achieved higher levels of development than others? The literature has long demonstrated the importance of this question, and many different explanations have been proposed. In this paper, I provide novel evidence for the importance of the composition of the population for long-run economic development. Over the course of history, colonialism and large population movements have changed the composition of the world's population (Putterman and Weil, 2010). Empirically assessing the effect of the population composition and disentangling it from other confounding factors is a challenge for most studies.

In order to identify the causal effect of historical population composition on long-run development, I exploit the history of the settlement of the *Pampas*, the fertile plains of Argentina.[1] The case of Argentina offers a contained setting, focusing on a single country with common national institutions and similar geographic endowments. The process of population and settlement in the *Pampas*, was greatly influenced by the mass migration of Europeans, although the exposure of different counties to the arrival of immigrants was not equal. Within the *Pampas,* the composition of population, as measured by the share of the European-born population, varies considerably.

To overcome the problem of endogenous sorting of migrants, I use an instrumental variables (IV) approach. The IV is constructed from a simple model of settlement and demographic growth and exploits variation over time in the incorporation of land for settlement, interacted with variation in the immigration of Europeans. Thus, the IV is a synthetic measure of the share of the European-born population in any given county.

The literature has proposed a number of explanations to shed light on why some regions have attained higher levels of development. Most of the work has highlighted three determinants: the geographic characteristics (such as land, climate or disease

[1] I will refer to this geographical region as fertile plains or the Pampas interchangeably.

environment); the enduring political and economic institutions from the colonial period; and the human capital brought by colonizer or immigrants (Engerman and Sokoloff, 1997, 2002; Acemoglu et al., 2001, 2002; Acemoglu, Gallego and Robinson, 2014; Bruhn and Gallego, 2012; Gallup, Sachs, and Mellinger, 1998, 2000; Gallup and Sachs, 2001; Sachs and Malaney, 2002; Glaeser et al., 2004; Easterly and Levine, 2012).[2] However, with the available evidence, it is difficult to disentangle the effect attributable to each hypothesis (or other potential explanations), which renders the discussion inconclusive (Easterly and Levine, 2012).

In this paper, I establish the causal effect of the historical composition of the population on different measures of past and current economic development. In order to isolate the effect of population composition from other confounding factors, I focus my analysis on the fertile plains of Argentina. As mentioned above, counties within the fertile plains share the same national institutions and very similar geographic characteristics, and they have faced the same history.

I show that the composition of the population, as measured by the share of the European-born population, explains long-run development when macro-institutions and geography are held constant. Having had a higher share of European population in 1914 raised per capita GDP, the share of the population with higher education and the share of skilled labor, by the end of the twentieth century.

The empirical analysis exploits the particular setting of the fertile plains of Ar-

[2]Engerman and Sokoloff (1997, 2002) analyze the effect of initial endowments and their distribution on inequality, political power and institutions. In a similar line of research, Acemoglu, Johnson and Robinson (2001, 2002) focus on the importance of colonial institutions for economic development, where these early institutions are persistent, and depending on their inclusive or extractive nature, are conducive or not to growth. Sachs and coauthors emphasize that levels of development are strongly correlated with geographical and ecological variables (Gallup, Sachs, and Mellinger, 1998, 2000; Gallup and Sachs, 2001; Sachs and Malaney, 2002). Glaeser et al. (2004) argue that European settlers brought human capital, and these past differences in human capital across countries explain a greater part of the current differences in economic growth, a point also stressed by Easterly and Levine (2012). On the other hand, Acemoglu, Gallego and Robinson (2014) argue that once institutions and human capital are taken as endogenous variables and instrumented for, only institutions have a robust impact on long-run development.

gentina, an area originally occupied by the native indigenous population, over which the Argentine government struggled to gain power. The availability of land to potential settlers in the fertile plains varied over time, depending on the level of civil and international conflicts and on the success of subsequent military campaigns that aimed to conquer the plains.[3]

European migration to Argentina was restricted throughout the colonial period and only started years after independence (1816), with peaks by the end of the nineteenth century and before the First World War. Between 1857 and 1914, close to 5.5 million Europeans migrated to Argentina.[4] The fertile plains were shocked in varying intensity by European immigrants. The shock to the population was far from negligible, as different areas experienced different intensities of immigration, and the percent of European population in each county after the shock ranged from 0% to 50% in 1914.

In order to overcome the endogenous sorting of Europeans, I construct a synthetic measure of the share of Europeans in each county. For this, I estimate the flow of the European population and the growth of the Argentine population for each county, from 1857 until 1885. I adjust the population for mortality and fertility and compute the stock of Argentinians and Europeans by 1885. With these synthetic stocks of Europeans and Argentinians, I compute the share of the European-born population.

Settlements in a given region became feasible only after the government took control of this region. Government's expansion over the fertile plains was unrelated to the arrival of Europeans or to social pressure to increase the amount of productive land.[5] The date of conquest of a region and the beginning of settlements is, thus, exogenous

[3]The process of settling the *Pampas* drastically contrasted with what happened in the US: while, in Argentina, settlers arrived after the government conquered the land, in the US, colonizers preceded the military.

[4]The Argentine government started recording statistics for immigration in 1857 and in 1914 conducted a census.

[5]For a complete explanation on the historic conditions that motivated the conquest of the fertile plains, see Section 2.

to the analysis and determines the exposure of a region to the arrival of Europeans. Interacted with the time-varying migration from Europe, this synthetic measure captures the exogenous share of Europeans.[6]

Using this synthetic measure of the share of the European population as an instrumental variable for the actual share of the European population, I compare counties in the fertile plains and estimate that an increase of 11% (one standard deviation (s.d.)) in the share of the European population raises per capita GDP by 92% in the long-run (0.83 standard deviations). Similar results hold for education and skilled labor: there is a higher share of the population with more education, as well as a higher share of skilled workers in 2001, in areas that had a higher degree of European immigration in 1914. The absolute size of this effect is considerable but in line with results found by Easterly and Levine (2012).

Next, I address the question of how Europeans affected development. I argue that Europeans brought human capital that was complementary to industrial activities, developed most of the industries and provided the skilled and unskilled labor. Immigrants were significantly involved in the process of industrialization, owning 80% of the industrial establishments by 1895. Using industrial census data, I find that measures of industrial development –such as the number of industrial establishments, the employment of high- and low-skilled industrial workers and the investment in energy– were substantially higher by 1935 in regions where the intensity of immigration was greater at the beginning of the twentieth century. Squicciarini and Voigtlaender (2015) show how population in the *upper-tail of knowledge* was an important driver of development during the first industrial revolution in France. Although I am not able to distinguish average human capital from upper-tail knowledge with the available data, evidence on the ownership of industrial establishments shows the important role that European

[6]By 1885, all of the fertile plains have been conquered, and no more variation from incorporated land can be exploited.

immigrants played in Argentina's industrial development.

Moreover, when human capital is proxied by the literacy rates, results show that these same counties had higher literacy in 1895 and 1914. The evidence suggests that immigrants not only contributed with their higher literacy, but also generated a positive externality by raising the level of human capital in the population as a whole. Human capital is an important factor in the process of economic growth (Galor and Weil, 1999, 2000 and Galor et al., 2009), as it is directly related to technological progress, increases productivity and contributes to the rapid growth of per capita GDP (Squicciarini and Voigtlaender, 2015).

The results I present in this paper show the importance of people themselves for economic development. The setting I exploit allows me to abstract from the classical institutional view, as well as from the hypotheses of the direct effect of geographic endowments on development. These results point to the importance of people, and how they matter, for reasons related to their knowledge and skills.

This paper is organized as follows. Section 2 reviews the conquest of the fertile plains and the European immigration to Argentina. I provide an historical account of the reasons behind military campaigns to the *Pampas*; the timing of these campaigns; and the process by which the plains were settled. Section 3 describes the data and their sources. Section 4 develops the empirical strategy and presents the results: I first present OLS estimates and then develop a demographic model that leads to the instrumental variable approach. I show the causal effect of immigrants on long-run development. Next, in Section 5 I present the channels of persistence: industrialization and human capital. In Section 6, I perform a series of robustness check, and consider alternative explanations: the direct effect of exportable crops and land inequality. Section 7 concludes.

2 The History of the Fertile Plains

It was not until end of the nineteenth century that the Argentinian government gained political power over the whole territory that is Argentina. During colonial times and after independence from the Spanish Empire in 1816, most of the fertile plains where settled by several indigenous tribes that did not recognize the Argentinian government. Relationships between Argentinians and the indigenous tribes were characterized by mistrust and violence. By the time of independence, the situation was such that Argentinians used to dispute land and wild livestock with the indigenous tribes, while the indigenous people organized assaults on settlements and cities, stealing livestock, goods and kidnapping people. Indigenous peoples' raids on cities and military excursions into indigenous settlements, both ending in destruction and deaths, were common. The Argentinian government and main tribes often agreed on peace treaties, but the government never recognized that area as an independent state or the indigenous people as legal owners of the land (Luna, 1993).

The threat to Argentinian settlements was not the government's only concern regarding the national territory. For Argentina to consolidate as a nation, it was necessary to delimit its frontiers, which became necessary to occupy Patagonia, an area also claimed by neighboring country Chile (Lacoste, 2002). But it was not until the end of the civil war in 1862 that a unified national government developed systematic plans to conquer the rest of the territory, starting in 1870 and continuing until 1885. Previous to 1870, military campaigns had developed with many years of interruption and loss of domain, especially during episodes of civil war and the war against Paraguay. Walther (1964) provides detailed information on the military campaigns and their effect on how the internal *frontier* between Argentinians and the indigenous tribes changed over time. Figures 2-3 depict maps showing the frontier between Argentina and the indigenous tribes in 1779, 1823, 1826, 1828, 1852, 1860, 1864 and 1876. Gains of terri-

tory by the Argentinian army and losses of domain over these years were a consequence of the government's limited resources for engaging in the multiple military conflicts it faced (Luna, 1993).

Which territories were conquered first and which were taken later strongly relates to military concerns: in particular, the distance to the main cities where armies were organizing and the distance to the indigenous settlements. The correlation between the timing of the conquest and the fertility of the land is 0.015%, while the correlation between the timing of the conquest and the distance to the capital city of Buenos Aires is 0.186%. It does not appear to be the case that the government based its actions on acquiring better or more productive land. Moreover, the lack of navigable rivers in the unconquered areas of the fertile plains, mostly inland areas, explains the pattern of conquest expanding from the main cities.

The end of the civil war and the reunification and pacification of the country started a period of European migration to Argentina in the second half of the nineteenth century. Immigrants were granted the same legal rights as Argentinians, without need to naturalize or acquire citizenship. The flow of immigrants to Argentina resembles the flow of immigrants to the USA, Canada and Australia. Figure 4 shows the time series of immigration and the net immigration of Europeans to Argentina. The series starts in 1857, when the national government started recording statistics on the arrival of immigrants to its ports. The flow of migration is far from constant; nor is a monotonic function of time.

Immigrants settled in cities, urban areas and the countryside and were employed both as skilled labor and unskilled labor. Occupations were diverse, ranging from farmers to construction workers, merchants and craftsmen. As of 1895, 41 percent of the European immigrants (males, aged 15 or above) were living in urban areas, while 32 percent devoted their time to farming and 28 percent to non-farm skilled labor.

The ultimate conquest of the *Pampas* was possible between 1870 and 1885, once military resources were no longer used in civil or international wars. At the same time, the peace achieved in Argentina and the economic conditions in Europe motivated Europeans to migrate to Argentina. Between independence and the reunification of the country, a period of almost fifty years, civil war prevented many Europeans from migrating to Argentina.[7] Although the decision to conquer the plains was unrelated to the immigration patterns, the timing of the expansion of the frontier over the plains overlaps with the arrival of the first European immigrants to the country, as shown in Figure 5. There may be considerations about Europeans migrating to Argentina because of the growing availability of land, but the data don't point to this conclusion. The correlation between the time series of immigration and the amount of land in the fertile plains under the political power of the government over time is close to 0.5, and a regression of immigration on the amount of land yields an R-squared of 20%. Temporary and permanent workers migrated mostly to the fertile plains; some of them returned to Europe after the harvest in the southern hemisphere (right before the harvest in the northern hemisphere), and others settled down and brought the rest of their families to Argentina. Progress and well being among immigrants was not immediate, but ultimately not hard to achieve.

Settlers decided which territories to contest based on private information, job opportunities, luck or other reasons. The endogenous choice by settlers on where to go within the plains motivates the use of an IV in my empirical analyses.

[7] In contrast to the US, which experienced large migration from northern Europe over this period.

3 Data and Summary Statistics

This study combines recent data on economic development (per capita GDP, higher education rate and share of skilled workers) with historical data on economic and social conditions. The unit of observation is at the county level. The sample covers the four provinces that comprise the fertile plains: Buenos Aires, Santa Fe, Córdoba and Entre Ríos. The southwest section of the fertile plains lies in the state of La Pampa, which is not included in the sample. It was not until 1952 that La Pampa became a province; before that, it was a *national territory* –i.e., a territory ruled by the national government, with appointed officials and no state constitution. Statistical information is not as exhaustive for national territories as it is for states. Moreover, the state of La Pampa changed all the county boundaries over the period considered in this study. Working with four states allows me to control for unobservable fixed variables at the state level. Though county boundaries have changed in some counties over time, it is possible to match older counties with new counties. New counties were founded mainly on previously unoccupied land, but there were cases where old counties split into two or more counties. When a new county can not be linked to an old county, the observation is dropped from the sample. There are 197 counties in these provinces, 31 of which are new counties not linked to an old county. Among the remaining 166 counties, 25 are capital cities or large urban areas and five are counties without current information on economic outcomes. Excluding capital cities and large urban areas, the final sample has 136 counties in four states.[8]

Historical information on population comes from three sources: the 1895 and 1914 Argentinian censuses and the Argentine Office of Migration. Both censuses contain detailed county-level information on population characteristics and economic activities.

[8]Capital cities are Buenos Aires, La Plata, the city of Santa Fe, the city of Córdoba and Parana. Large urban areas refer to the *Greater Buenos Aires*, *Greater La Plata* and Rosario.

I digitalize data on all variables used from the censuses: total population, foreign-born population and population living in urban areas. Moreover, the 1914 census includes an agricultural and livestock census, which was used to construct a variable on the economic activities performed at the county level. Somoza and Lattes (1967) computerized representative samples of historical 1895 census microdata, from which individual-level data on nationality, age, sex and occupation and literacy are obtained. The share of the European-born population is computed as the number of European-born divided by the total number of adults in the population.[9] Since 1857, the Argentine Office of Migration has recorded all non-Argentine incoming and outgoing people. Detailed data on the number of immigrants and their countries of origin from 1857 to 1914 were digitalized for this study.

Data on the territory under the political power of the Argentine government comes from Walther (1964). Walther's detailed description of the military campaigns are summarized with a series of maps, that show, for different years the actual *frontier* between the territory under the Argentinian government and the native tribes' territory. Walther's work is based on military and historical documents. I complement these maps with Gallo (1983) and Tell (2008), which provide more detailed information for the states of Córdoba and Santa Fe.

The Argentinian Statistical Office (INDEC) computes GDP at the national and province level, but not at the county level. In 1994, INDEC conducted the National Economic Census (CNE), censusing all businesses at the county level, except for the agricultural sector, recording the value of production, costs, investment, etc.. Per capita GDP is constructed by combining CNE's gross product data with yearly agri-

[9]To have a more accurate measure of the European population, special care is needed when considering younger cohorts, for the children of Europeans were considered by law, and counted as, Argentinians. I computed the share of Europeans among the population older than 14 years of age. Given data availability, those younger than six years of age are also included in 1914, an issue that affects the denominator, since there were few, if any, Europeans that young.

cultural output estimates from the Ministry of Agriculture (see Appendix). For the provinces of Buenos Aires and Santa Fe, provincial-statistical offices compute GDP at the county level. For these two states, the correlation between CNE's gross product and province GDP at the county level is 95%; the correlation between CNE's gross product augmented by the agricultural output estimates and province GDP is also 95%. The regression of provinces' GDP on the CNE's gross product, augmented by agricultural output, has an R^2 of 90.34. I will use CNE's gross product augmented by agricultural output as a proxy for GDP at the county level.

Further, I will use data from the 1935 Industrial Census, which documents different measures of industrial development at the county level.

Data on higher education rates and the share of skilled workers is from the 2001 Population Census and is publicly available from the Argentine Statistical Office. Finally, geo-referenced data on the quality of the soil come from the National Institute for Agriculture and Livestock Technology (INTA) (Cruzate et al., 1990). INTA provides geo-referenced detailed data on the quality of the soil and elaborates an index that assigns a greater value to better soils. This index of land quality refers to the geographical conditions of the soil (such us ground composition and rain) and not to the technologies used for cultivation. I combine the geo-referenced data provided by INTA with the county boundaries and compute an area weighted average of the land-quality index. Geographical information on average rainfall and temperature comes from Worldclim;[10] data on elevation from the National Oceanic and Atmospheric Administration (NOAA) and U.S. National Geophysical Data Center; and data on ruggedness of the terrain from Nunn and Puga (2012). All the geographical variables are geo-referenced data, which I combined with county boundaries to compute county averages. The availability of railroads in a given county is computed as the average railroad density in a radius of five

[10]See http://www.worldclim.org/formats.

km, and data on railroads come from *ATLAS de Suelos de la República Argentina.*[11]

Table 1 shows the summary statistics for the variables used in this study. As a measure of the composition of the population, I construct the share of European population, defined as the fraction of the European-born population in 1895 and 1914. The average share of European population is 23% with a standard deviation of 11%, with counties ranging from less than 1% to 48% of its population of European origin in 1914. Average GDP per capita is slightly above 6.700 dollars, with the bottom 25% of the counties having less than 3.560 dollars and the top 25% having a per capita GDP above 9.000 dollars. On average, 10.4% of the population aged 25 and older has completed more than 12 years of education (completed secondary school and started or finished tertiary or university degrees). Of those individuals reporting an occupation in 2001, on average, 18% work in high-skilled jobs.

4 Estimation Strategy and Results

The empirical analysis compares various measures of recent development (log per capita GDP (1994), higher education rates (2001) and the share of skilled workers (2001)) across counties with past population composition (1914). I estimate the following equation:

$$y_i = \alpha + \beta SE_i + X_i\gamma + \eta_p + \epsilon_i, \tag{1}$$

where y_i is the dependent variable in county i and SE_i is the share of the European-born population in county i in 1914. X_i are controls for county i characteristics in 1914, including geographic and socio-economic characteristics, and controls for distance to the city of Buenos Aires and for the availability of railroads. η_s are state fixed effects.[12]

[11] See Cruzate et al. (1990).

[12] The city of Buenos Aires is the capital city of the country, the main port of entry for traded goods and immigrants, and the most densely populated city. Proximity to this politically and economically

Figure 1 shows the positive correlation between per capita GDP in 1994 and the share of European-born population in 1914. Table 2 documents OLS results from the regression of per capita GDP in 1994 on the share of European population in 1914 (equation (1)). In the different columns, the set of controls is expanded sequentially: column 1 shows the correlation between per capita GDP and the share of European population; column 2 adds state fixed effects; column 3 adds geographical controls (rain, temperature, elevation, ruggedness and land quality); column 4 adds controls for the distance to the city of Buenos Aires and availability of railroads; and column 5 adds controls for socio-economic variables (the share of productive land used for agriculture, population density and urbanization rate). The OLS regressions show a robust association between the share of European population in 1914 and recent income per capita. Following columns 4 and 5, the preferred specifications, one standard deviation in the share of Europeans increases per capita GDP by a 0.66 and a 0.56 standard deviation, respectively. Socio-economic covariates could, themselves, be an outcome of the share of Europeans, which leads to a concern about "bad controls" and a bias coefficient of interest. In this case, column 4 is not affected by such a problem.

In Table 3, I consider two other measures of current development: in Panel A, the share of population with higher education and, in Panel B, the share of population with high-skilled occupations, both measured in 2001. Results show that the share of Europeans positively correlates with these two variables. In accordance with the results for GDP, close to eighty years after the arrival of European immigrants, differences in economic performances and development across counties correlate with the pattern of settlement at the beginning of the twentieth century.

Although results are encouraging, the evidence presented in Tables 2 and 3 should be taken with caution. If European immigrants selected themselves into the different

relevant city may have independent effects on development.

counties for reasons that correlate with prospects of development, then results may be biased.[13] In order to deal with this potential problem, I use instrumental variables to account for the possible endogeneity in the selection of Europeans to the different counties.

4.1 Instrumental Variable Approach

To account for the possible endogeneity in where European immigrants settled once they arrived in Argentina, I construct a synthetic measure of the share of immigrants in each county and use it as an instrumental variable for the actual share of immigrants in a given county.

The IV exploits two sources of variation: first, the movement of the internal frontier between Argentina and the indigenous tribes, which determines the incorporation of land available for settlement; and second, the variation in the net-immigration of Europeans to Argentina. As will be discussed below, I use a simple model of settlement and demographic growth to interact both sources of variation and estimate, for each county, the flow of the European population and the growth of the Argentinian population, from 1857 until 1885. I adjust the population for mortality and fertility and compute the stock of Argentinians and Europeans by 1885. I then use these synthetic stocks of Europeans and Argentinians to compute the synthetic share of the European-born population. After 1885, all of the fertile plains were controlled by the government, and no more variation can be exploited from the incorporation of land for settlement. This motivates my use of 1885 as the end year for the construction of the synthetic

[13]Immigrants may have had information in hand to choose one destination in favor of another; for example, previously settled immigrants may have sent letters or gone back to the home country to attract the rest of the family to the newly settled area across the ocean. Even differences in infrastructure, access to railroad or size of the cities in the plains may have played a role for immigrants when deciding where to settle. Moreover, job opportunities listed at the port of entry or references from other immigrants may have also been important choice variables.

measure of Argentinian and European populations.

Using historical information on the Argentinian government's military campaigns, I am able to assign to each county a year in which (at least half of) the land was available to settlers. Historical records compiled by Walther (1964) trace the evolution of the area under the political power of the Argentine government over time, until the government gained power over the whole territory. With a series of maps, Walther documents the end result of military excursions and the boundary between the Argentinian government and the indigenous tribes that resulted from these expeditions (Figures 2-3 are two examples of these maps that show the internal frontier). By 1885, the Argentinian government controlled all of the fertile plains. I assume that no land was conquered or lost until the next military campaign, an assumption very close to the actual events. I overlap the county boundaries on these maps and establish the date on which the county was considered to be on the Argentinian side.[14]

The second source of variation comes from the time series of immigration to Argentina. The migration pattern to Argentina resembles that of the USA and the rest of the world: Europeans escaped famine and wars, looking for better prospects. In particular, the time series of migration to Argentina and the USA have a correlation of 0.795.[15]

In an ideal experiment, regions (counties) are equal in all respects, and are randomly shocked with European population in different intensities. In such an experiment, one can analyze long-run economic and social development, and see whether any differences can be explained by the share of European population, the only variable that varies across regions. The empirical setting thet I am analyzing very closely approximates

[14]The date a county *enters* Argentina should not be confused with the date on which a county is officially founded, usually years after it was under Argentinian power.

[15]Data on USA migration are from Historical Statistics of the United States, Millennial Edition On Line, edited by Susan B. Carter, Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright, Cambridge University Press 2006. http://hsus.cambridge.org/HSUSWeb/toc/tableToc.do?id=Ad1-2.

my ideal experiment, in that it consists of geographically uniform counties that were shocked by different degrees of immigration. The key difference is that Europeans were not randomly distributed, as they chose where to settle. The IV I construct corrects for this by randomly assigning Europeans to counties, exploiting the variation in the timing of seizure of land from the indigenous tribes and the arrival of Europeans, as mentioned above.

For the flow of European population to be exogenous in my analysis, the reasons that Europeans had to migrate to Argentina should have been unrelated to the success or failure of the military campaigns in conquering new land. Moreover, it should also be the case that the government's decision to conquer these vast tracks of land was independent of the arrival of European immigrants to the country. History shows that this appears to be the case; as discussed above, military and safety issues prompted the government to take control of this region, starting years before the first wave of European immigrants arrived. To my understanding, the expansion over the fertile plains was unrelated to the arrival of Europeans, and there was no social pressure within the country or need for new productive land. The military campaigns in the fertile plains ended by 1885, when slightly less than 900,000 immigrants had arrived in Argentina, in comparison to the approximately three million net-immigrants that arrived by 1914. At the time of the military campaigns, civil unrest within the country prevented many Europeans from migrating to Argentina. That is why the flow of migration started after the reunification and pacification of the country in 1860, and was particularly strong years after the conquest of the plains had ended. Finally, for the identifying assumption to be correct, the synthetic share of European immigration has to affect the dependent variable (per capita GDP, higher education, etc.) only through the actual share of European immigration, while having no effect through other variables.

The information used for the IV is the number of Europeans that arrived each year, the number of counties within the Argentinian boundary each year and the initial number of Argentinians in a given county. I combine the data, which are available each year, in a simple model of settlement and demographic growth.

Each year between 1857 and 1885, European immigrants are distributed across counties in the fertile plains that the Argentinian government controls. To make the distribution of immigrants as random as possible, Europeans are distributed uniformly across counties–i.e., all *available* counties get the same fraction of immigrants. Matters such as county size, the distance to the city of Buenos Aires, the amount of unoccupied land or the ease of access by railroad are not considered for the distribution of immigrants. Indeed, these matters were important for the actual decision of where to settle and were taken into consideration by immigrants; therefore, for the IV to be valid, this exercise considers a distribution of immigrants independent of all these essential aspects.

Once Europeans are distributed, I assume that they never move again. Europeans die at the yearly rate δ and reproduce at rate ρ, and children born to Europeans in Argentina are considered European. Argentinians, on the other hand, are initially present in counties that came under the political power of the Argentine government by 1857, but not in counties conquered after 1857. If the initial number of Argentinians can be regarded as exogenous to the analysis, the Argentinian population by 1857 is estimated from the 1869 census, adjusted by the population growth rate to the year 1857. On the other hand, since the initial population is not a random variable and may depend on observed and unobserved characteristics, I also consider the initial stock of Argentinians to be the average number for Argentinians in 1857. The estimation below is robust to both assumptions. Argentinians die and reproduce at the same rates as Europeans, δ and ρ, respectively, and there is a fraction ϕ of Argentinians who decide to

move to a new county each year. I distribute the *movers* uniformly across all counties that belong to Argentina.

The mortality rate, the fertility rate and the fraction of Argentinians that move each year are computed from the 1869, 1895 and 1914 censuses. The mortality rate is computed to be equal to 2.2; the fertility rate is computed to be equal to 5.3%; and the moving rate for Argentinians, ϕ, is computed to be equal to 1.95%.[16] The first stage and the analyses in the coming section are robust to changes in the parameters of the demographic model, as well as to changes in the assumption on the initial Argentinian population. All of these possibilities will be considered as robustness checks in Section 6.

The synthetic number of Europeans in each county in 1885 is defined as:

$$CE_i = \sum_{t=1857}^{1885} \frac{1}{N_t}(1-\delta+\rho)^{1885-t} e_t \cdot \mathbb{1}_i\{t \geq D_i\}. \tag{2}$$

The synthetic number of Argentinians in each county in 1885 is defined as:

$$CA_i = \hat{A}_{i1857}(1-\delta+\rho-\phi)^{28} + \sum_{t=1857}^{1885} \frac{1}{N_t}(1-\delta+\rho-\phi)^{1885-t}\phi a_t \cdot \mathbb{1}_i\{t \geq D_i\}, \tag{3}$$

where CE_i and CA_i are the synthetic number of Europeans and Argentinians, respectively, in county i in 1885. e_t is the number of Europeans that arrive in year t, and a_t is the number of Argentinians that move to a different county in year t. $\hat{A}_{i1857}$ is the initial number of Argentinians in a given county. The two assumptions on A_{i1857} are: $\hat{A}_{i1857} = A_{i1857}$, the actual population; and $\hat{A}_{i1857} = \bar{A}_{1857}$ if $A_{i1857} > 0$ and $\hat{A}_{i1857} = 0$ otherwise.[17] $\mathbb{1}_i\{\cdot\}$ is an indicator of whether county i belongs to Argentina, and D_i is

[16] See the Appendix for a detailed explanation of how to compute these values. Values for the fertility and mortality rates do not differ from those listed on historical records. There is no official statistic on the Argentinian moving rate

[17] As a robustness check, I will also consider the case in which all counties are assigned the same initial stock of Argentinians, $\hat{A}_{i1857} = \bar{A}_{1857}$.

the year in which county i started to be under the political power of the Argentine government (year in which land seizure occurred). $N_t = \sum_i n_{it}$ is the number of counties under Argentinian political control at time t, and n_{it} equals 1 if county i belongs to Argentina at time t, 0 otherwise.

The synthetic share of the European-born population is defined as $CSE_i = CE_i/(CE_i + CA_i)$ and is used as IV for the actual share of the European population. Variation in both CE_i and CA_i will induce variation in the constructed share. CE_i varies across counties i, depending on the year in which county i came under the political power of the Argentine government, D_i, and also on the number of immigrants, e_t, that arrived at time t. Variation in CA_i depends on D_i, the number of Argentinians moving, and ϕa_t.[18]

Walther's account of the conquest of the plains can be summarize in eight waves of land seizures in the following years: 1779, 1823, 1826, 1860, 1864, 1869, 1876 and 1884. Not all counties were conquered at the same time: 66 counties already existed at the time of the independence of the Spanish colonial government, while six were conquered in 1860, seven in 1864, eleven in 1869, eleven in 1876 and five in 1885.

4.2 The long-run effect of European immigration

I run the following specification for the first stage:

$$SE_i = \alpha + \psi CSE_i + X_i\gamma + \eta_p + \epsilon_i, \tag{4}$$

where CSE_i is the constructed share of European immigration. I consider both assumptions on $\hat{A}_{i1857}$ and run a first stage for each case, IV1 and IV2. IV1 assumes that $\hat{A}_{i1857} = A_{i1857}$, the actual population, while IV2 assumes that $\hat{A}_{i1857} = \bar{A}_{1857}$ if

[18] Using the actual Argentinian population by 1857, instead of the average population, will also introduce variation in CA_i.

$A_{i1857} > 0$ and $\hat{A}_{i1857} = 0$ otherwise.

Figure 6 shows the correlation between the share of the European population and the synthetic share of the European population. Figure 7 shows the conditional correlation when control variables and fixed effects are included. Figures 6 and 7 show a strong positive correlation between the two variables across specifications and assumptions on the initial population.

Table 4 shows the first-stage regression, equation (4). Columns 1 and 2 show results for IV1 and columns 3 and 4 for IV2. The specification on columns 1 and 3 includes geographical variables and controls for the distance to the city of Buenos Aires and availability of railroads, while the specification on columns 2 and 4 adds socio-economic controls. In all four columns the coefficient on the synthetic share of immigration remains positive and significant, confirming the result presented in Figures 6 and 7. F-test of the coefficient ψ shows a strong first stage, and weak identification is ruled out by the Kleibergen-Paap test.

Tables 5 and 6 show results for three measures of current development, where the synthetic share of immigration is used as the instrumental variable for the actual share of European population. I report results for the two specifications discussed above: a first specification that includes geographic variables and controls for the distance to the city of Buenos Aires and availability of railroads, and a second specification that adds socio-economic controls. In Table 5, the dependent variable is log per capita GDP in 1994. Columns 1 and 3 show the OLS results from Table 3 for comparison. The IV estimates of the share of Europeans are positive and significant, with the estimates showing the long-run effect of the share of European population on per capita GDP. IV estimates from column 2 indicate that an increase of one standard deviation in the share of European population in 1914 led to a 0.92 standard deviation increase in per

capita GDP in 1994.[19] The IV estimates are higher than the OLS estimates, suggesting a negative bias in the selection of Europeans to counties. This difference may also be due to measurement error in the share of the European population. For a county such as *Río Cuarto*, with a 20% share of Europeans, increasing the share to 25% would raise per capita GDP from 6,912 to 9,495 dollars. That is certainly an economically significant effect.

Table 6 reports the IV estimates of the effect of the share of European population in 1914 on the population with higher education (Panel A) and the share of workers in high-skilled occupations in 2001 (Panel B). Results show a positive and significant effect of the composition of the population on these variables. IV estimates from column 2 indicate that an increase of one standard deviation in the share of European population in 1914 raises the share of population with higher education by a 0.32 standard deviation and the share of workers in high-skilled occupations by a 0.62 standard deviation.[20] In accordance with the result for GDP, these other measures of recent development show a similar picture: economic growth and development arose in counties that experienced a higher intensity of immigration. How and why Europeans contributed to a divergence in the paths of economic development is the subject of the next section, in which I examine the channels through which development arose and persisted over time.

[19] One standard deviation in the share of Europeans equals 0.11 (11%), a 50% increase in the share of Europeans for an average county.

[20] Results for high-skilled occupations are significant in all specifications, while results for higher education are significant in all specifications but not in column 2.

5 The effect of European immigration: the channels of persistence

Why did Europeans affect economic outcomes close to a century after their arrival? How did their initial effect on the economy propagate and persist over time? The literature on economic growth and development provides many answer to these questions. This section provides evidence for two particular channels: the process of industrialization and literacy. The process of industrialization and literacy rates are linked together and represent different aspects of human capital. Measures of human capital in the study period are not easy to compute, as years of schooling or the population with graduate education were not surveyed in Argentina at that time. Therefore, although literacy rates can be regarded as a noisy measure of human capital, literacy still captures how instructed the population was. Industrialization, on the other hand, captures the development of economic activities in the society for which a certain level of human capital is needed. Even though we cannot observe the knowledge and skills that these industrial entrepreneurs and workers had, we can observe the result of their effort: the industries they forged; the industrial output; and whether they work in skilled occupations.

Industrialization has been widely understood as an important factor in a country's development. Countries that industrialized earlier rank higher today in development, per capita income and living standards. Since the Industrial Revolution, higher standards of development have been linked to the degree of industrialization of an economy, with terms such as *industrialized nation*, *developed nation* and *advanced economy* used interchangeably. In the case of Argentina, industrialization occurred in some counties more than in others, and cities that developed more were also cities that experienced higher industrialization at the beginning of the twentieth century. Why industrializa-

tion arose in the first place is an open question, but from the industrial censuses in 1895, 1914 and 1935, we know that the process of industrialization was tightly linked to immigrants and their ability and willingness to set up and operate industrial establishments. In this sense, industrialization operated as a vehicle that propagated development over time, and long-term differences across regions emerged between more and less industrialized counties.

Table 7 examines the nationality of the owners and workers of industrial establishments in Argentina in 1895, 1913 and 1935. In 1895, 81% of these establishments were owned by foreigners, while 59% of their workers were immigrants. Close to twenty years later, in 1913, 65% of the industrial establishments were run by foreigners, and workers of foreign origin made up 49% of the employment. Industry at that time was centered mainly around the production of garment, food, wooden, metal and chemical products, and construction. Table 6 also shows that in 1935, 58% of the industrial establishments were still under the ownership of foreign citizens.

Below, I investigate the relationship between the structure of the industrial sector in 1935 and the share of Europeans in 1914.[21] The 1935 industrial census keeps records of information at the establishment level and the county level. The census provides four measures of industrial development: the number of establishments per person; the share of skilled workers in the population; the per capita value of production; and the investment in energy per person.[22] In Table 8, I examine the effect of the share of European population in 1914 on these variables, using the different specifications described above. IV estimates are shown in all cases, together with OLS estimates for comparison. Panel A shows results for the value of industrial production, Panel B for the share of skilled workers, Panel C for the number of factories per person and Panel

[21] 1935 is the first industrial census for which data at the county level are available.

[22] Value of production is in 1935 peso currency and energy is measured in horse power. For the per-person variables, I consider the 1914 population, since it is the closest population census.

D for the investment in energy in horse power per person. Following column 2, the share of Europeans has a positive and significant effect on all industrial variables. An increase of one standard deviation in the share of European population in 1914 raises the value of industrial production by 0.49 s.d., the share of skilled workers by 0.57 s.d., the number of factories per person by 0.82 s.d. and the energy in horse power per person by 0.38 s.d. in 1935. For a county such as *Río Cuarto*, having a share of Europeans of 25% instead of 20% would have raised the value of industrial production in 1935 by 28%.

In sum, consistent with the results presented in the previous section, the greater development of counties with a high share of European population can be traced back to the process of industrialization at the beginning of the twentieth century. While I cannot observe the level of human capital or the know-how of a particular industrial activity, I do observe that industrial development was closely linked to the intensity of European immigration.[23]

Glaeser et al. (2004) find evidence for human capital as a channel for growth and better political institutions, and Easterly and Levine (2012) point out that human capital was an important intermediate channel through which colonial settlement affected development in the long-run.

The role of immigrants in the economy is also evident by the type of work, industry and occupation they chose. The 1895 and 1914 census reports the percent of the workforce, aged 14 and above, that was of European origin for all occupations in the economy.[24] While Europeans represented 36.4% and 42.9% of the workforce in 1895

[23]In a recent study, Franck and Galor (2015) show that while industrialization was initially conducive to economic development, it has had an adverse effect on standards of living in the long-run. This study concludes that the "characteristics that permitted the onset of industrialization, rather than the adoption of industrial technology per se, have been the source of prosperity among the currently developed economies". The evidence for the *Pampas* is consistent with this conclusion, namely that some unobserved characteristic(s) of the European immigrant, motivated the adoption of industrial activities, and were closely linked to the process of development.

[24]This table refers to the country's total workforce and is not limited to the four provinces that

and 1914, they represented 42.4% and 44.4%, respectively, of the workforce employed in industry and craftsmanship in those years. More importantly, Europeans made up 86.9% and 45.3% of those employed in science-oriented jobs and 54.6% and 46.3% of those occupied in health, in 1895 and 1914, respectively. Europeans were over-represented in most of the high-skilled occupations–in particular, in occupations such as industry and craftsmanship, health and science–playing an important role in key occupations that are relevant for technological advancement and economic growth.

An even more accurate measure of human capital is the level of literacy of the population. The level of human capital at the end of the nineteenth century and beginning of the twentieth century was altered by the inflow European immigrants.[25] Differences in literacy rates were greater among Europeans than between Europeans and Argentinians. Table 9 examines literacy rates in 1895 and 1914 by nationality, as reported by immigrants in Argentina: while the Argentinian population was, on average, 51.7% and 63.2% literate in 1895 and 1914, respectively, Germans had a literacy rate above 87% for both years. Immigrants from Italy, Spain and France had a literacy rate of 59.6%, 67.4% and 79.3% in 1914, respectively. When weighted by population, on average, Europeans were 67% literate in 1895 and 64.2% literate in 1914, while the population as a whole was 57% literate in 1895 and 63% literate in 1914. Europeans migrating to the *Pampas* were, on average, more literate than locals, but the difference appears smaller in later years. What was the effect, if any, of a population with higher human capital on development?

In Table 10, I examine the relationship between the literacy rate at the county level and the share of the European population. Panel A shows IV estimates for the effect of

comprise the *Pampas*.

[25] A growing literature focuses on the effect of European migration during the Era of Mass Migration on human capital; see Abramitzky et al. (2012) and Lafortune et al. (2014) for the case of the U.S. and de Carvalho Filho and Monasterio (2012) and Rocha, Ferraz and Soares (2015) for the case of Brazil.

the share of the European population in 1895 on literacy rates in that year, and Panel B shows the effect of the share of European population in 1914 on 1914 literacy rates. Both Panels A and B show that the effect is positive and significant. Following column 2, an increase of one standard deviation in the share of European population increases literacy rates by 1.24 s.d. in 1895 and 0.14 s.d in 1914, respectively.

Table 9 and 10 raise the question: What explains this difference in literacy rates across counties? Can this difference be explained by a composition effect–that is, by replacing a less literate Argentinian with a more literate European? Or is the effect of immigration on literacy the consequence of an increase in the acquisition of human capital? As Table 8 documents, in 1895 Europeans were, on average, 15% more literate than Argentinians, while in 1914, this difference had gone down to 1%. In Table 9, Panel A, column 2, the IV estimate of 0.71 implies an effect on literacy far greater than the original 15% difference, and the same is true for the IV estimate in Panel B, column 2. The composition effect can explain part of but not the whole difference in literacy rates across counties. Beyond the composition effect, immigration has a positive externality on literacy rates of the rest of the population. There are several potential explanations for this: it may be that Europeans provided more education to their offspring; that Europeans demanded more schools in the places were they settled and that, afterwards, the schools provided education to all citizens; that the Argentinian government provided education to the newly arrived immigrants; or that economic progress generated a demand for more skilled labor, providing higher incentives to acquire human capital.

To assess these possibilities, I investigate whether more education was provided in areas with a higher share of European immigrants. Since the mid-eighteenth century, the Argentine government had built schools throughout the country, offering free public education to all individuals of school-age (6 to 14 years old). These schools were mostly

in urban areas or highly densely populated areas. Private schools were also present and offered religious learning and/or were present in areas where public schools were not readily available. Given the government's active policy of public education, it is plausible that counties with a higher share of Europeans provided more publicly financed education in order to assimilate and introduce immigrants into Argentinian society. However, census data show a different story: areas with a higher share of European immigrants are associated with a higher number of private schools and a lower number of public schools (per school-age population).

Census data on schools in 1914 list schools' location and the school-aged population in each county, from which I construct the number of schools per 1000 school-aged children. On average, there were 5.3 public schools and 0.85 private schools in each county per 1000 school-ages individuals, with standard deviations of 2.32 and 0.71, respectively. Table 10, Panels C and D show IV estimates for the effect of the share of the European-born population on the number of public schools and private schools, respectively. The share of the European population had a negative and significant effect on the number of public schools in 1914; following column 2, an increase of one standard deviation in the share of European population in 1914 reduces the number of public schools by 0.66 standard deviation, a magnitude equivalent to eliminating 1.5 public schools. Panel D shows IV estimates for the regression of the number of private schools on the share of Europeans. Results show a positive effect of immigrants, though not significant in all specifications, on the quantity of schools. Following column 2, an increase of one standard deviation in the share of immigrants increases by 0.28 s.d. the number of private schools per school-aged population. In general, the evidence points to literacy rates being higher in areas with more Europeans–not because of educational policies pursued by the national government, but most likely because of the individual decisions of the citizens in these counties.

6 Robustness Checks

The results in this paper are robust to a series of variations on the assumptions of the demographic model, as well as to alternative explanations for the divergence in economic growth.

In Tables 11 and 12, I consider four variations to the demographic model presented in Section 4; in particular, I set the values of the parameters arbitrarily high (double): Panel A considers the moving rate, ϕ, equal to 4%; in Panel B, the fertility rate, ρ, is set equal to 10%; in Panel C, the mortality rate, δ, is set equal to 6%; and in Panel D, $\phi = 4\%$, $\rho = 10\%$ and $\delta = 6\%$ simultaneously. Panels A-D in Tables 11 and 12 show that the results remain consistent; changes in the assumptions of the model do not alter the IV estimate; and the long-run effect of the share of European population on per capita GDP, the share of population with higher education and the share of workers in high-skilled occupations are robust to these changes.

Panel E in Tables 11 and 12 considers an alternative explanation for the divergence in the paths of economic development. Wheat was at that time one of the most valuable export crops and, therefore, growing it was an important economic activity. In order to consider the production of wheat in my estimation, I add to the regression equation a variable that measures the share of land dedicated to wheat in 1914. Results show that the share of land dedicated to wheat does not have an effect on any of the recent measures of economic development, even though wheat-growing was an important activity, and wheat later became the biggest export product of Argentina. Wheat does not consistently explain the divergent paths of economic development across counties; nor does it invalidate the proposed hypothesis of the relevance of the composition of the population.

In a second robustness check, I explore whether inequality in the distribution of land across individuals can have a direct, likely negative, effect on development. I use

1914 census data on the size of plots and the distribution of land in counties, and I compute a measure of land inequality: a land-gini. In Panel E, Tables 11 and 12, I repeat the IV estimation, including land-gini as an independent variable. Results show that land inequality has a negative effect on development, although the coefficient is not statistically significant. The inclusion of the land-gini variable does not change the previous interpretation of the effect of the European immigration.

Results in Tables 11 and 12 reinforce the conclusion that the composition of the population explains long-run development. In sum, the regressions shown in the previous sections are robust to the inclusion of other potential explanatory variables and changes in the parameters of the model.

7 Conclusion

The period between 1850 and the First World War saw an unprecedented flow of European immigrants to Argentina, mostly to the rural and urban areas across the fertile plains. Areas where the European immigration accounted for a greater share of the total population in 1914 managed to develop more than areas with fewer Europeans, as measured by the level of per capita GDP, close to one hundred years later.

Why were these areas able to develop more than areas where Europeans represented a smaller share of the population? As I discuss in this paper, the *Pampas* provides an excellent empirical setting to answer this question, for it is an area that shares common political institutions and uniform geographical conditions across counties. Therefore, it is possible to abstract from the usual competing hypotheses of differences in geographic endowments or institutions. Still, the place where migrants settled is endogenous and, thus, presents an empirical challenge. In order to overcome this issue, I construct an IV based on a synthetic distribution of immigrants across counties. To do so, I exploit

the variation in the arrival of Europeans interacted with variation in the availability of land for settlement. This synthetic measure of the population's composition captures the exogenous share of Europeans in each county, not the share that would result from the immigrants' decisions. IV estimates show that there is a positive causal effect of the composition of the population on long-term development. Counties where the share of the European-born population was higher in 1914 achieved higher per capita GDP and currently have a more educated population with a greater share of skilled workers.

Motivated by this strong result, I then analyze the role played by immigration and, in particular, whether immigrants' human capital can be linked to the process of economic development. First, I argue that European immigrants were more engaged in industrial production: they started most of the industrial activities and provided most of the industrial skilled and unskilled labor. I show that counties with a higher share of European population outperformed in various measures of industrial development in 1935. In these counties, the number of industries and workers and the value of industrial output and investment exceeded other counties, in which the share of European population was relatively smaller.

Second, I exploit variation across nationalities in literacy rates in 1895 and 1914 and show that counties' average literacy rate was higher in the counties with a higher share of European-born. Higher literacy rates cannot be explained by differences in literacy across nationalities alone; the effect goes beyond what can be attributed to a composition effect of Europeans. Nor can higher literacy rates be explained by an effort of the national government to educate and assimilate immigrants. In fact, data on the number of public schools show that the opposite is true–that the number of public schools per student was smaller in counties in which Europeans accounted for a higher share of the population. I argue that there was a positive externality on the society, raising literacy rates of the population as a whole.

This paper shows the importance of people themselves for long-run economic development. The setting I exploit allows me to abstract from differences in political institutions and geographic characteristics and to focus on the effect of the composition of the population on economic development. The results I present show the importance of people's knowledge and skills for the process of economic development and the persistence of this effect in the long-run.

References

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

[2] Acemoglu, D., F. Gallego, and J. A. Robinson, "Institutions, Human Capital and Development," *Annual Reviews of Economics*, 6, pp. 875-912 (2014).

[3] Acemoglu, D., S. Johnson, and J. A. Robinson, "The colonial origins of comparative development: an empirical investigation," *American Economic Review*, 91: 1369–401 (2001).

[4] Acemoglu, D., S. Johnson, and J. A. Robinson, "Reversal of fortunes: geography and institutions in the making of the modern world income distribution," *Quarterly Journal of Economics*, 117: 1231–94 (2002).

[5] Bruhn, M. and F. Gallego, "Good, bad, and ugly colonial activities: Do they matter for economic development?" *The Review of Economics and Statistics*, 94(2):433–461 (2012).

[6] Cruzate, G., L. Gomez, M. J. Pizarro, P. Mercuri, S. Banchero, G. Angueira, D. Prieto, J. Lopez and G. Ibarraza, "Suelos de la República Argentina, Versión 1.0." Versión digital corregida y ajustada mediante técnicas actuales de ingeniería SIG en base a la información original vectorizada a partir de los mapas de suelos provinciales que integran el *Atlas de Suelos de la República Argentina*, INTA (1990).

[7] de Carvalho Filho, I., and L. M. Monasterio, "Immigration and the origins of regional inequality: Government-sponsored European migration to Southern Brazil before World War I," *Regional Science and Urban Economics*, Volume 42, Issue 5, Pages 761–904, (2012).

[8] Dirección General de Inmigración, *Resumen Estadístico del Movimiento Migratorio en la República Argentina. Años 1857 - 1924.* Buenos Aires: Talleres Graficos del Ministerio de Agricultura de la Nacion (1925).

[9] Easterly, W., and R. Levine, "The European Origins of Economic Development," NBER, Working Paper, 2012.

[10] Engerman, S.L., and K. L. Sokoloff, "Factor endowments, institutions, and differential paths of growth among New World economies: a view from economic historians of the United States." In *How Latin America Fell Behind*, ed. S Harber, pp. 260–304. Stanford, CA: Stanford Univ. Press (1997).

[11] Engerman, S.L., and K. L. Sokoloff, "Factor endowments, inequality, and paths of development among New World economies," NBER, Working Paper 9259, 2002.

[12] Franck, Raphael and Oded Galor, "Is Industrialization Conducive to Long-Run Prosperity?," CESifo Working Paper Series No. 5354, 2015.

[13] Gallo, E., *La Pampa gringa: la colonización agrícola en Santa Fe (1870-1895).* Buenos Aires: Editorial Sudamericana, 1983.

[14] Gallup, John Luke, and Jeffrey Sachs D. with Andrew D. Mellinger, "Geography and Economic Development," *Annual World Bank Conference on Development Economics 1998 (April)*, The World Bank: Washington, DC (1998).

[15] Gallup, John Luke and Jeffrey D. Sachs, "The Economic Burden of Malaria," *The Supplement to The American Journal of Tropical Medicine & Hygiene* Vol. 64, no. 1, 2, January/February (2001).

[16] Galor, O. and David N. Weil, "From Malthusian Stagnation to Modern Growth," *American Economic Review*, 89, 150-154 (1999).

[17] Galor, O. and David N. Weil, "Population, Technology and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond," *American Economic Review*, 90, 806-828 (2000).

[18] Galor, O., O. Moav and D. Vollrath, "Inequality in land ownership, the emergence of human capital promoting institutions and the great divergence," *Review of Economic Studies* 76: 143–79 (2009).

[19] Glaeser, E.L., R. La Porta, F. Lopez-De-Silanes and A. Shleifer, "Do institutions cause growth?" *Journal of Economic Growth*, 9: 271–303 (2004).

[20] Lacoste, Pablo, *La guerra de los mapas entre Argentina y Chile: una mirada desde Chile,* Historia (Santiago), ISSN 0717–7194, vol. 35, pages 211–249 (2002).

[21] Lafortune, J., Tessada, J., and Gazmuri, A. (2014), "Lured in or crowded out? Estimating the impact of immigration on natives' education using early XXth century U.S. immigration," *mimeo.*

[22] Luna, Félix, *Breve Historia de los Argentinos,* Buenos Aires: Planeta (1993).

[23] Mellinger, Andrew D., Jeffrey D. Sachs and John L. Gallup, "Climate, Coastal Proximity, and Development," *Oxford Handbook of Economic Geography.* Clark, Gordon L., Maryann P. Feldman, and Meric S. Gertler, eds. Oxford University Press (2000).

[24] National Oceanic and Atmospheric Administration (NOAA) and U.S. National Geophysical Data Center, TerrainBase, release 1.0 (CD-ROM), Boulder, Colorado.

[25] Nunn, Nathan and D. Puga, "Ruggedness: The blessing of bad geography in Africa," *Review of Economics and Statistics*, 94(1), 20-36 (2012).

[26] Putterman, L. and D. Weil, "Post-1500 Population Flows and the Long-run Determinants of Economic Growth and Inequality," *Quarterly Journal of Economics* 125(4): 1627-1682 (2010).

[27] República Argentina, *Censo Industrial de 1935.* Buenos Aires: Dirección General de Estadística de la Nación (1938).

[28] República Argentina, *Segundo Censo de la República Argentina, Mayo 10 de 1895.* Buenos Aires: Taller Tipográfico de la Penitenciaria Nacional (1898).

[29] República Argentina, *Tercer Censo Nacional de la República Argentina, Levantado el 1^{ro} de Junio de 1914.* Buenos Aires: Talleres Gráficos de L. J. Rosso y Cía. (1916).

[30] Rocha, R., Ferraz, Claudio and Rodrigo R. Soares, "Human Capital Persistence and Development," *IZA Discussion Papers* 9101, (2015).

[31] Sachs, Jeffrey D. and Pia Malaney, "The Economic and Social Burden of Malaria," *Nature Insight*, Vol. 415, no. 6872, Feb. 7 (2002).

[32] Somoza, J. L., and A. E. Lattes, *Muestras de los dos primeros censos nacionales de población, 1869 y 1895.* Buenos Aires: Instituto Torcuato Di Tella, Centro de Investigaciones Sociales, Documento de Trabajo no 46 (1967).

[33] Squicciarini, M., and N. Voigtlaender, "Human Capital and Industrialization: Evidence from the Age of Enlightenment," *Quarterly Journal of Economics*, forthcoming (2015).

[34] Tell, S., *Córdoba rural, una sociedad campesina (1750-1850).* Buenos Aires: Prometeo Libros, Asociación Argentina de Historia Económica (2008).

[35] Walther, J. C., *La Conquista del Desierto.* Buenos Aires, Círculo Militar (1964).

Figure 1: Correlation between Ln per capita GDP in 1994 and the share of the European-born population in 1914.

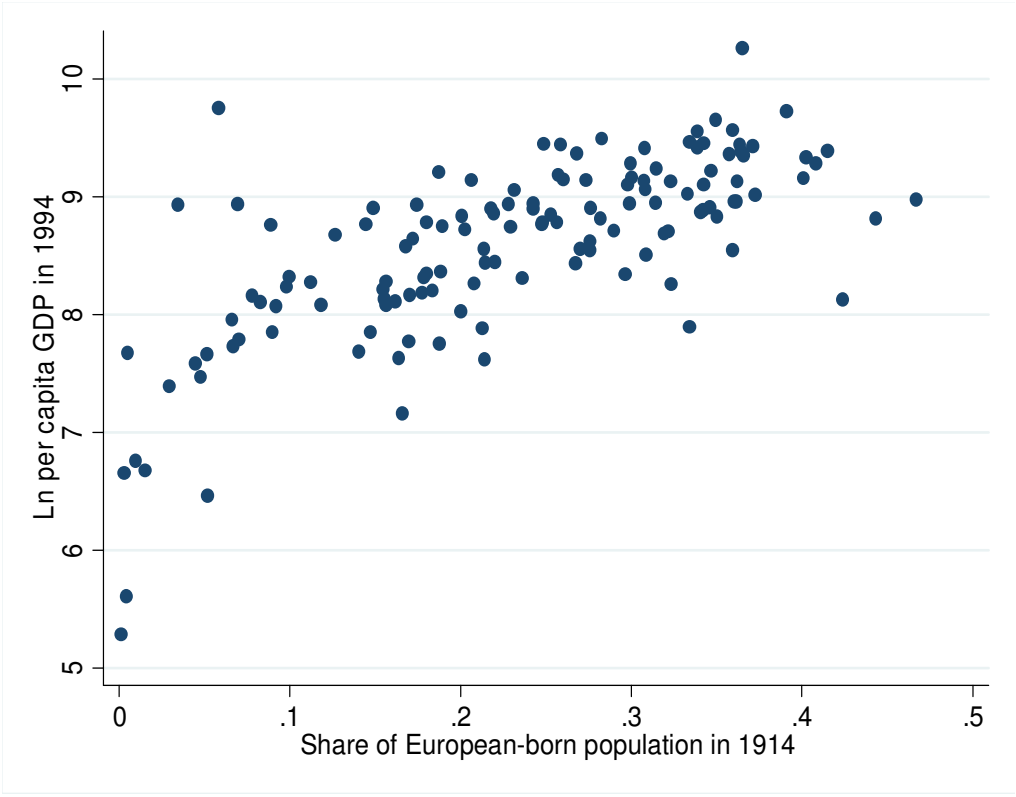

Figure 2: Frontier between Argentina and the indigenous tribes, 1810 – 1828.

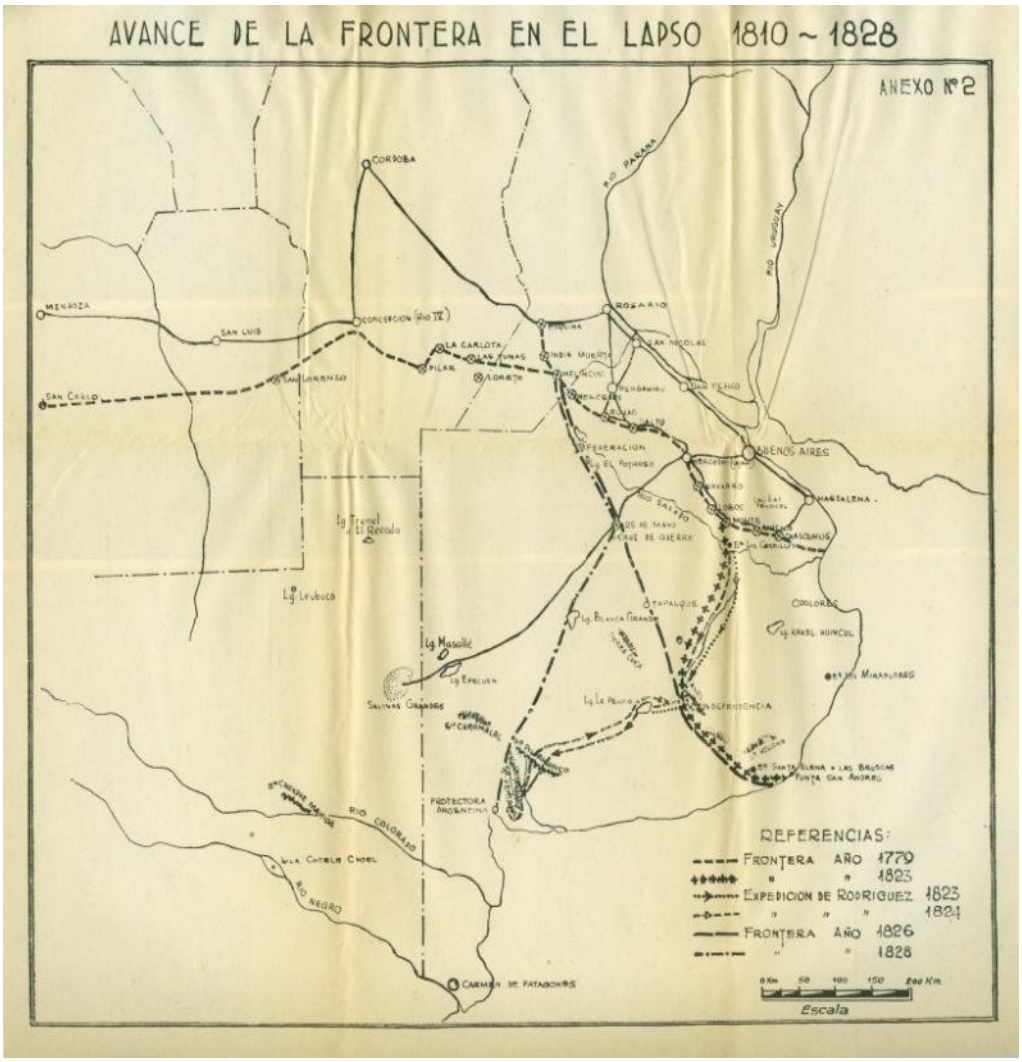

Figure 3: Frontier between Argentina and the indigenous tribes, 1852 - 1876.

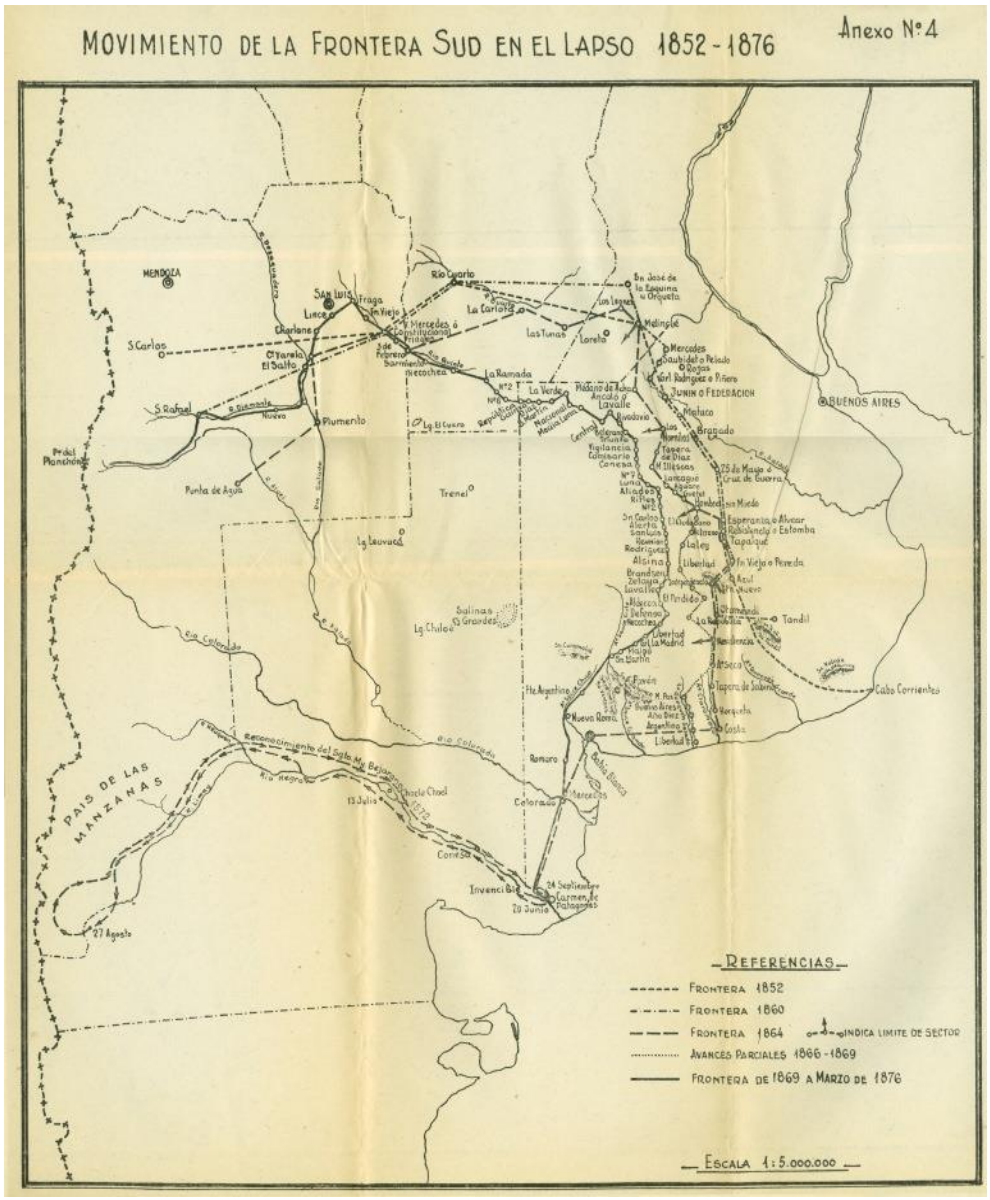

Figure 4: Immigration Time Series, 1857 - 1914.

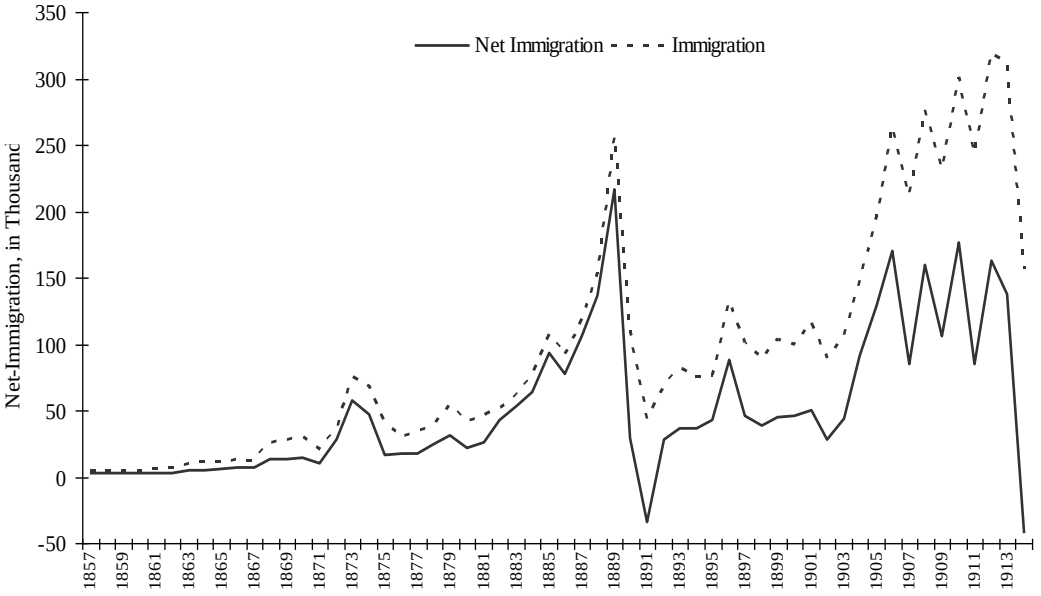

Figure 5: Cumulative Net-Immigration and Area for Settlement

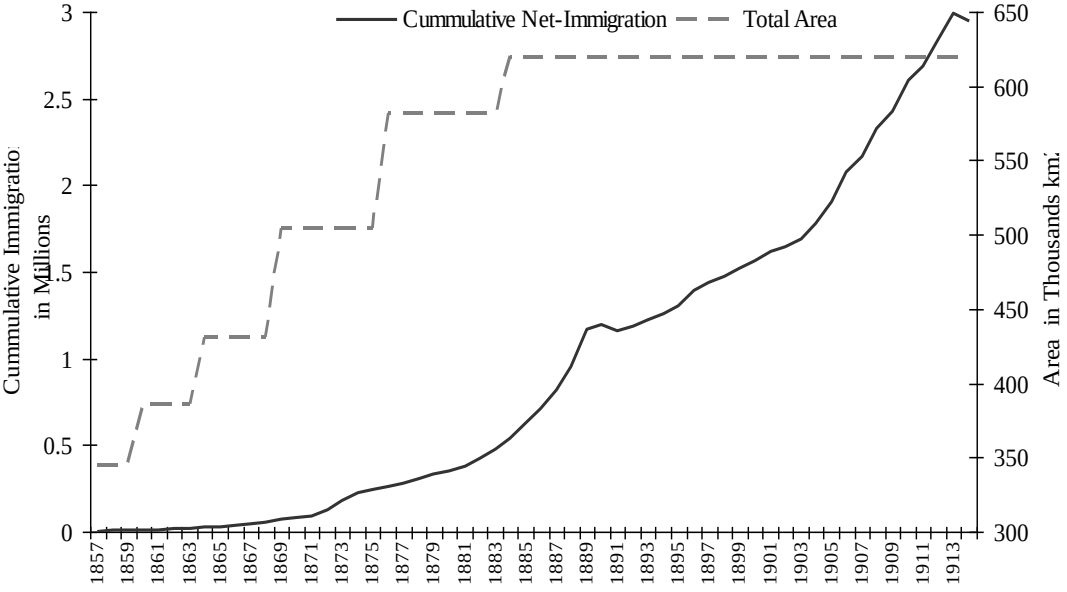

Figure 6: Correlation between the share of the European-born population and the synthetic share of European population

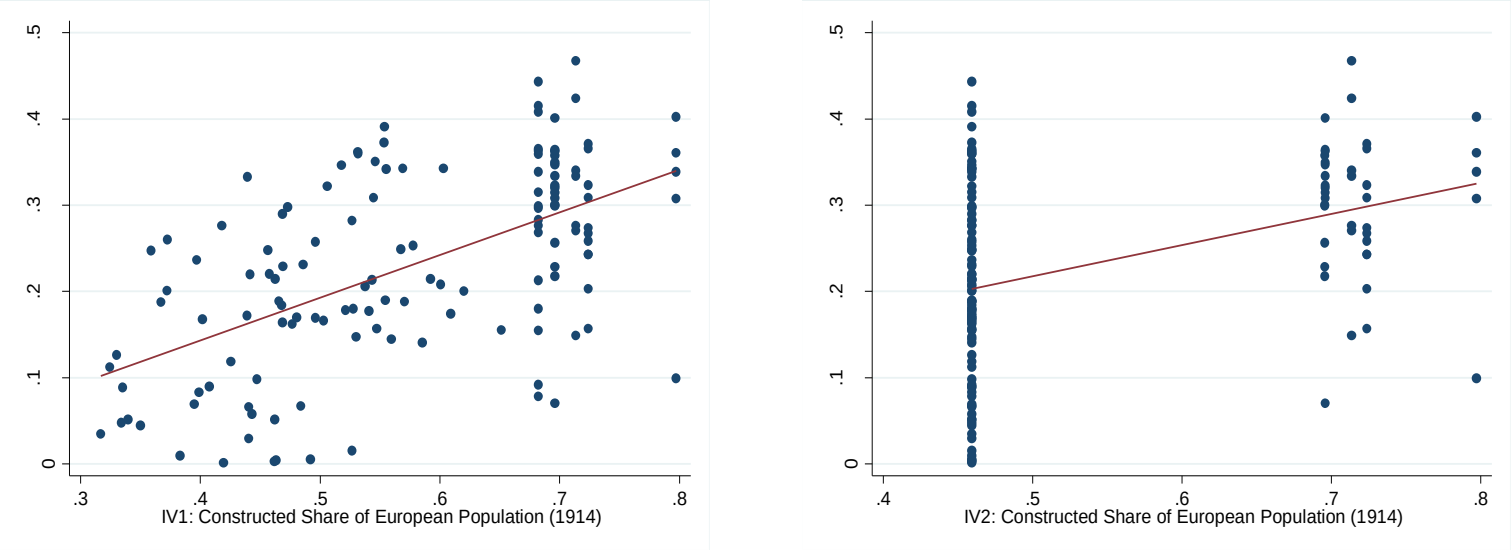

Figure 7: Correlation between the share of the European-born population and the synthetic share of European population, conditional on control variables and state fixed effects.

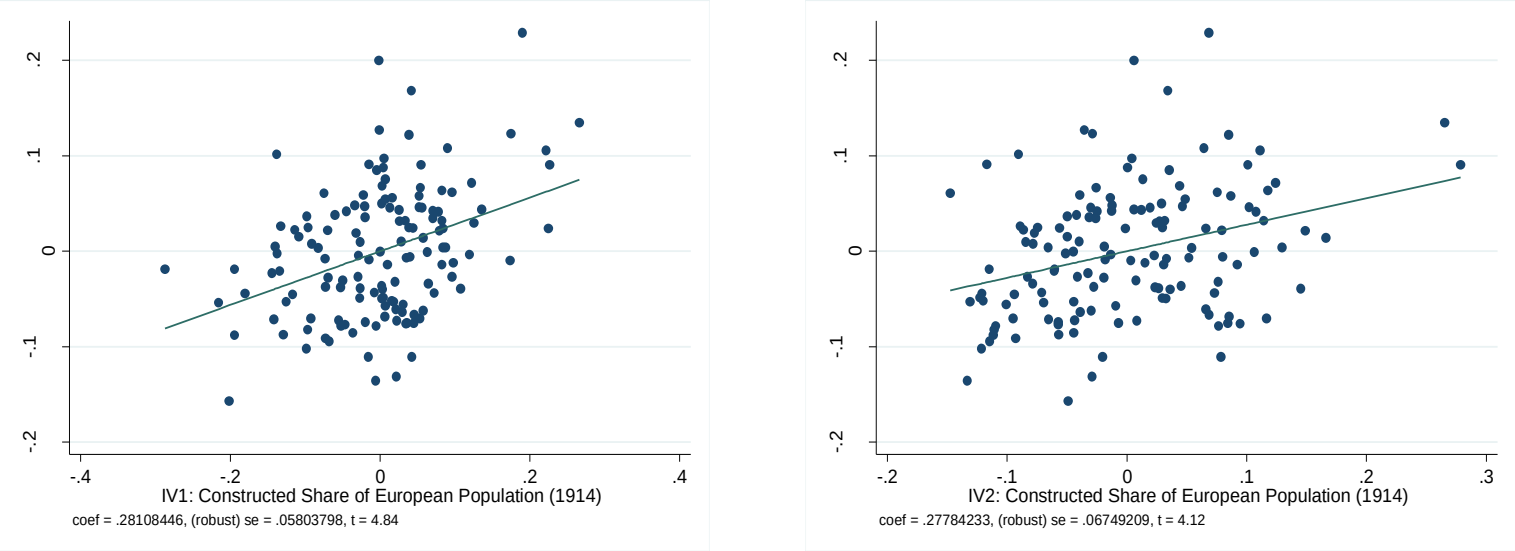

Table 1: Summary Statistics

Variable	Mean	Standard Deviation	50th Percentile
Share of the European-born population, 1914	0.23	0.11	0.16
Share of the European-born population, 1895	0.36	0.20	0.39
GDP per capita, 1994	6754	4190	3560
Ln GDP per capita, 1994	8.59	0.78	8.18
Share of Pop. with Higher Education, 2001	0.10	0.02	0.09
Share of High-skilled Workers, 2001	0.18	0.04	0.15
Ln Industrial Output per capita, 1935	4.40	1.14	3.87
Skilled Workers per-1000 Individuals, 1935	1.99	2.06	0.89
Number of Factories per-1000 Individuals, 1935	3.69	2.16	2.16
Energy in H.P. per capita, 1935	0.10	0.14	0.05
Literacy Rate, 1895	0.55	0.11	0.56
Literacy Rate, 1914	0.63	0.05	0.58
Number of Private Schools per-1000 school-aged p	0.85	0.71	0.35
Number of Puclic Schools per-1000 school-aged po	5.33	2.32	3.63
Distance to the City of Buenos Aires	5.71	0.65	5.29
Land Quality Index	45.43	16.39	33.06
Railroad Density in 5km.	4.71	2.77	2.90
Percent of Land used for Agriculture	0.28	0.23	0.07
Population Density	6.67	5.53	2.78
Urban Rate	0.33	0.18	0.22
Number of observations: 136			

Table 2: Share of European Population and per capita GDP

	OLS	**OLS**	**OLS**	**OLS**	**OLS**
	(1)	(2)	(3)	(4)	(5)
Dependent Variable:	*Ln per capita GDP 1994*				
Share of the European-born	4.862***	5.668***	4.823***	4.560***	3.914***
population in 1914	(0.610)	(0.632)	(0.744)	(0.852)	(0.796)
Distance to BA City				0.098	0.079
				(0.163)	(0.151)
Land Quality			0.008**	0.005	0.004
			(0.004)	(0.004)	(0.004)
Railroad Density				0.040	0.052*
				(0.028)	(0.029)
Percent of Land used for					0.644**
Agriculture in 1914					(0.293)
Population Density in 1914					-0.028**
					(0.011)
Urban Rate in 1914					0.557
					(0.341)
Constant	7.467***	7.437***	8.864***	8.591***	8.363***
	(0.167)	(0.227)	(0.997)	(1.316)	(1.238)
Province Fixed Effects	N	Y	Y	Y	Y
Geographic Controls	N	N	Y	Y	Y
Socio-economic controls	N	N	N	N	Y
R-squared	0.495	0.522	0.608	0.616	0.638
Observations	136	136	136	136	136

Note: OLS regressions with robust standard errors in parentheses. Dependent variable in all columns is ln per capita GDP in 1994. Column 1 shows the correlation between per-capita GDP and the share of European population, column 2 adds state fixed effects, column 3 adds geographical controls (rain, temperature, elevation, ruggedness and land quality), column 4 adds controls for the distance to the city of Buenos Aires and availability of railroads and column 5 adds controls for socio-economic variables (the share of productive land used for agriculture, population density and urbanization rate). *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Share of European Population and Development

	OLS	**OLS**	**OLS**	**OLS**	**OLS**
	(1)	(2)	(3)	(4)	(5)
Panel A. *Dependent Variable Share of Population with Higher Education 2001*					
Share of the European-born	0.059***	0.060***	0.063***	0.054***	0.037*
population in 1914	(0.016)	(0.017)	(0.021)	(0.020)	(0.020)
R-squared	0.134	0.180	0.384	0.456	0.550
Panel B. *Dependent Variable. Share of Population with High-skilled Occupations 2001*					
Share of the European-born	0.297***	0.270***	0.226***	0.208***	0.146***
population in 1914	(0.025)	(0.028)	(0.036)	(0.035)	(0.042)
R-squared	0.608	0.637	0.700	0.717	0.767
Province Fixed Effects	N	Y	Y	Y	Y
Geographic Controls	N	N	Y	Y	Y
Distance to BA and Railroad	N	N	N	Y	Y
Socio-economic controls	N	N	N	N	Y
Observations	136	136	136	136	136

Note: OLS regressions with robust standard errors in parentheses. Dependent variable in Panel A is the share of popuation with higher education in 2001, in Panel B the dependent variable is the share of population with high-skilled occupations in 2001. Column 1 shows the correlation between the dependent variable and the share of European population, column 2 adds state fixed effects, column 3 adds geographical controls (rain, temperature, elevation, ruggedness and land quality), column 4 adds controls for the distance to the city of Buenos Aires and availability of railroads and column 5 adds controls for socio-economic variables (the share of productive land used for agriculture, population density and urbanization rate). *** p<0.01, ** p<0.05, * p<0.1.

Table 4: First Stage - Share of European Population

	IV 1	**IV 1**	**IV 2**	**IV 2**
	(1)	(2)	(3)	(4)
Dependent Variable:	*Share of European Population in 1914*			
Synthetic Share of the	0.281***	0.251***	0.278***	0.218***
European population in 1885	(0.058)	(0.042)	(0.067)	(0.061)
Distance to BA City	0.012	0.024*	0.003	0.020
	(0.020)	(0.012)	(0.022)	(0.014)
Land Quality	0.002***	0.000	0.002***	0.001
	(0.000)	(0.000)	(0.001)	(0.000)
Railroad Density	0.008**	0.003	0.006	0.001
	(0.004)	(0.003)	(0.004)	(0.003)
Percent of Land used for		0.190***		0.200***
Agriculture in 1914		(0.024)		(0.028)
Population Density in 1914		0.003**		0.003*
		(0.001)		(0.001)
Urban Rate in 1914		0.087**		0.068
		(0.042)		(0.045)
Constant				
Province Fixed Effects	Y	Y	Y	Y
Geographic Controls	Y	Y	Y	Y
Socio-economic controls	N	Y	N	Y
F-stat.	23.46	36.16	16.95	12.80
Kleibergen-Paap p-value	0.0000	0.0000	0.0001	0.0005
R-squared	0.688	0.816	0.673	0.799
Observations	136	136	136	136

Note: First stage regressions with robust standard errors in parentheses. Dependent variable in all columns is the actual share of the European-born population. All regressions include geographical control variables and province fixed effects. Regressions in columns 2 and 4 also include socio-economic controls. In columns 1 and 2 the IV is constructed using the actual number of Argentinians in 1869. In columns 3 and 4 the IV is constructed using the average number of Argentinians in 1869 for those provinces with Argentinian population in 1857. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Share of European Population and per capita GDP

	Specification 1			Specification 2		
	OLS	**IV 1**	**IV 2**	**OLS**	**IV 1**	**IV 2**
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:	*Ln per capita GDP 1994*					
Share of the European-born	4.560***	6.350***	8.951***	3.914***	5.778***	9.032***
population in 1914	(0.852)	(1.474)	(2.139)	(0.796)	(1.702)	(2.632)
Distance to BA City	0.098	0.023	-0.086	0.079	-0.014	-0.175
	(0.163)	(0.155)	(0.161)	(0.151)	(0.168)	(0.184)
Land Quality	0.005	0.002	-0.003	0.004	0.004	0.003
	(0.004)	(0.005)	(0.006)	(0.004)	(0.004)	(0.004)
Railroad Density	0.040	0.023	-0.002	0.052*	0.046	0.034
	(0.028)	(0.029)	(0.035)	(0.029)	(0.028)	(0.029)
Percent of Land used for				0.644**	0.228	-0.500
Agriculture in 1914				(0.293)	(0.389)	(0.641)
Population Density in 1914				-0.028**	-0.032***	-0.038**
				(0.011)	(0.012)	(0.015)
Urban Rate in 1914				0.557	0.415	0.166
				(0.341)	(0.337)	(0.438)
Constant	8.591***	7.936***	6.984***	8.363***	7.982***	7.316***
	(1.316)	(1.503)	(1.850)	(1.238)	(1.374)	(1.784)
Province Fixed Effects	Y	Y	Y	Y	Y	Y
Geographic Controls	Y	Y	Y	Y	Y	Y
Socio-economic controls	N	N	N	Y	Y	Y
R-squared	0.616			0.638		
Observations	136	136	136	136	136	136

Note: OLS and IV regressions with robust standard errors in parentheses. Dependent variable in all columns is ln per-capita GDP in 1994. All regressions include geographical control variables and province fixed effects. Regressions in columns 4-6 also include socio-economic controls. In column 2 and 4 the IV is constructed using the actual number of Argentinians in 1869. In columns 3 and 6 the IV is constructed using the average number of Argentinians in 1869 for those provinces with Argentinian population in 1857. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Share of European Population and Development

	Specification 1			Specification 2		
	OLS	**IV 1**	**IV 2**	**OLS**	**IV 1**	**IV 2**
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. *Dependent Variable Share of Population with Higher Education 2001*						
Share of the European-born	0.054***	0.051	0.125**	0.037*	0.092**	0.162**
population in 1914	(0.020)	(0.044)	(0.056)	(0.020)	(0.046)	(0.078)
R-squared	0.456			0.550		
Panel B. *Dependent Variable. Share of Population with High-skilled Occupations 2001*						
Share of the European-born	0.208***	0.235***	0.405***	0.146***	0.210***	0.385***
population in 1914	(0.035)	(0.075)	(0.093)	(0.042)	(0.073)	(0.120)
R-squared	0.717			0.767		
Province Fixed Effects	Y	Y	Y	Y	Y	Y
Geographic Controls	Y	Y	Y	Y	Y	Y
Socio-economic controls	N	N	N	Y	Y	Y
Observations	136	136	136	136	136	136

Note: OLS and IV regressions with robust standard errors in parentheses. Dependent variable in Panel A is the share of popuation with higher education in 2001, in Panel B the dependent variable is the share of population with high-skilled occupations in 2001. All regressions include geographical control variables and province fixed effects. Regressions in columns 4-6 also include socio-economic controls. In column 2 and 4 the IV is constructed using the actual number of Argentinians in 1869. In columns 3 and 6 the IV is constructed using the average number of Argentinians in 1869 for those provinces with Argentinian population in 1857. *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Ownership and Industrial Workers

	year	Share of Foreigners
Ownership	1895	0.81
	1913	0.65
	1935	0.58
Workers	1895	0.59
	1913	0.49

Source: authors calculations based on 1935 industrial census.

Table 8: Share of European Population and Industrial Development

	Specification 1			Specification 2		
	OLS	**IV 1**	**IV 2**	**OLS**	**IV 1**	**IV 2**
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. *Dependent Variable Ln Value of Industrial Production 1935*						
Share of the European-born	5.177***	4.980*	11.593***	4.799***	6.978**	15.640***
population in 1914	(1.415)	(2.564)	(3.892)	(1.459)	(2.759)	(5.395)
R-squared	0.257			0.284		
Panel B. *Dependent Variable. Skilled Workers per-1000 Individuals 1935*						
Share of the European-born	11.222**	10.342	9.602	8.880***	18.007**	17.319**
population in 1914	(4.625)	(6.838)	(6.154)	(3.323)	(7.108)	(7.944)
R-squared	0.191			0.356		
Panel C. *Dependent Variable Factories per-1000 Individuals 1935*						
Share of the European-born	8.582***	15.675***	22.502***	9.201***	23.537***	31.941***
population in 1914	(2.422)	(4.873)	(7.647)	(2.680)	(6.430)	(11.378)
R-squared	0.424			0.489		
Panel D. *Dependent Variable Energy in H.P. per capita 1935*						
Share of the European-born	0.395***	0.479*	0.764	0.319**	0.877**	1.247*
population in 1914	(0.129)	(0.283)	(0.534)	(0.149)	(0.375)	(0.745)
R-squared	0.153			0.213		
Province Fixed Effects	Y	Y	Y	Y	Y	Y
Geographic Controls	Y	Y	Y	Y	Y	Y
Socio-economic controls	N	N	N	Y	Y	Y
Observations	136	136	136	136	136	136

Note: OLS and IV regressions with robust standard errors in parentheses. All regressions include geographical control variables and province fixed effects. Regressions in columns 4-6 also include socio-economic controls. In column 2 and 4 the IV is constructed using the actual number of Argentinians in 1869. In columns 3 and 6 the IV is constructed using the average number of Argentinians in 1869 for those provinces with Argentinian population in 1857. *** p<0.01, ** p<0.05, * p<0.1.

Table 9: Literacy Rates by Country of Birth

Nationality	Literacy rate in 1895	Literacy rate in 1914
Argentina	51.7%	63.2%
Average European	67.0%	64.2%
Average Population	57.0%	63.3%
Austria	-	69.2%
France	76.0%	79.3%
Germany	89.0%	88.2%
Great Britain	89.0%	90.9%
Italy	57.0%	59.6%
Spain	69.0%	67.4%
Switzerland	-	86.9%

Source: authors calculations based on 1895 and 1914 censuses.

Table 10: Share of European Population and Human Capital

	Specification 1			Specification 2		
	OLS	**IV 1**	**IV 2**	**OLS**	**IV 1**	**IV 2**
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. *Dependent Variable Share of Literate Population 1895*						
Share of the European-born	0.309***	0.710***	1.026***	0.285***	0.769***	1.104***
population in 1895	(0.065)	(0.199)	(0.308)	(0.062)	(0.209)	(0.351)
R-squared	0.529			0.557		
Observations	126	126	126	126	126	126
Panel B. *Dependent Variable. Share of Literate Population 1914*						
Share of the European-born	-0.039**	0.059*	0.084*	-0.032*	0.069*	0.108*
population in 1914	(0.017)	(0.035)	(0.046)	(0.017)	(0.038)	(0.063)
R-squared	0.961			0.963		
Observations	136	136	136	136	136	136
Panel C. *Dependent Variable Public Schools per 1,000 school-aged Population 1914*						
Share of the European-born	-15.504***	-13.676***	-24.101***	-13.466***	-13.293***	-26.465***
population in 1914	(2.323)	(4.588)	(6.614)	(2.636)	(5.132)	(8.648)
R-squared	0.536			0.543		
Observations	136	136	136	136	136	136
Panel D. *Dependent Variable Private Schools per 1,000 school-aged Population 1914*						
Share of the European-born	3.131***	1.802	5.659**	3.021***	1.626	6.399*
population in 1914	(0.716)	(1.738)	(2.395)	(1.023)	(2.025)	(3.322)
R-squared	0.304			0.308		
Observations	126	126	126	126	126	126
Province Fixed Effects	Y	Y	Y	Y	Y	Y
Geographic Controls	Y	Y	Y	Y	Y	Y
Socio-economic controls	N	N	N	Y	Y	Y

Note: OLS and IV regressions with robust standard errors in parentheses. All regressions include geographical control variables and province fixed effects. Regressions in columns 4-6 also include socio-economic controls. In column 2 and 4 the IV is constructed using the actual number of Argentinians in 1869. In columns 3 and 6 the IV is constructed using the average number of Argentinians in 1869 for those provinces with Argentinian population in 1857. *** p<0.01, ** p<0.05, * p<0.1.

Table 11: Robustness Check

Dependent Variable:	Ln per capita GDP 1994		Share of Population with Higher Education 2001		Share of Population with High-skilled Occupations 2001	
	IV 1	**IV 2**	**IV 1**	**IV 2**	**IV 1**	**IV 2**
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. *Assumption:*			*Moving Rate 4%*			
Share of the European-born population in 1914	4.550	8.951***	-0.011	0.125**	0.318*	0.405***
	(3.388)	(2.139)	(0.101)	(0.056)	(0.164)	(0.093)
Panel B. *Assumption:*			*Fertility Rate 10%*			
Share of the European-born population in 1914	6.824***	8.951***	0.063	0.125**	0.245***	0.405***
	(1.459)	(2.139)	(0.043)	(0.056)	(0.075)	(0.093)
Panel C. *Assumption:*			*Mortality Rate 6%*			
Share of the European-born population in 1914	5.999***	8.951***	0.038	0.125**	0.240***	0.405***
	(1.492)	(2.139)	(0.045)	(0.056)	(0.074)	(0.093)
Panel D. *Assumption:*			*Moving Rate 4%, Fertility Rate 10% and Mortality Rate 6%*			
Share of the European-born population in 1914	4.725	8.951***	0.002	0.125**	0.292**	0.405***
	(3.083)	(2.139)	(0.091)	(0.056)	(0.146)	(0.093)
Panel E.			*Including share of Land sown with Wheat*			
Share of the European-born population in 1914	6.117***	9.757***	0.032	0.126	0.161*	0.344***
	(1.708)	(3.210)	(0.050)	(0.079)	(0.089)	(0.119)
Share of land sown with Wheat in 1914	0.328	-0.696	0.026	-0.000	0.104***	0.052
	(0.678)	(1.161)	(0.018)	(0.027)	(0.034)	(0.040)
Panel F.			*Including Land-gini*			
Share of the European-born population in 1914	5.817***	8.694***	0.060	0.142**	0.214***	0.402***
	(1.555)	(2.386)	(0.048)	(0.067)	(0.080)	(0.103)
Land-gini in 1914	-0.839*	-0.396	0.014	0.027	-0.033	-0.004
	(0.487)	(0.659)	(0.014)	(0.018)	(0.024)	(0.033)
Province Fixed Effects	Y	Y	Y	Y	Y	Y
Geographic Controls	Y	Y	Y	Y	Y	Y
Socio-economic controls	N	N	N	N	N	N
Observations	136	136	136	136	136	136

Note: IV regressions with robust standard errors in parentheses. All regressions include geographical control variables and province fixed effects. In columns 1, 3 and 5 the IV is constructed using the actual number of Argentinians in 1869. In columns 2, 4 and 6 the IV is constructed using the average number of Argentinians in 1869 for those provinces with Argentinian population in 1857. *** p<0.01, ** p<0.05, * p<0.1.

Table 12: Robustness Check

Dependent Variable:	Ln per capita GDP 1994		Share of Population with Higher Education 2001		Share of Population with High-skilled Occupations 2001	
	IV 1	**IV 2**	**IV 1**	**IV 2**	**IV 1**	**IV 2**
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. *Assumption:*			*Moving Rate 4%*			
Share of the European-born population in 1914	3.664	9.032***	0.079	0.162**	0.434	0.385***
	(5.414)	(2.632)	(0.140)	(0.078)	(0.289)	(0.120)
Panel B. *Assumption:*			*Fertility Rate 10%*			
Share of the European-born population in 1914	6.330***	9.032***	0.099**	0.162**	0.215***	0.385***
	(1.677)	(2.632)	(0.046)	(0.078)	(0.074)	(0.120)
Panel C. *Assumption:*			*Mortality Rate 6%*			
Share of the European-born population in 1914	5.368***	9.032***	0.083*	0.162**	0.221***	0.385***
	(1.767)	(2.632)	(0.047)	(0.078)	(0.075)	(0.120)
Panel D. *Assumption:*			*Moving Rate 4%, Fertility Rate 10% and Mortality Rate 6%*			
Share of the European-born population in 1914	3.956	9.032***	0.086	0.162**	0.367*	0.385***
	(4.568)	(2.632)	(0.119)	(0.078)	(0.217)	(0.120)
Panel E.			*Including share of Land sown with Wheat*			
Share of the European-born population in 1914	5.738***	9.020***	0.080*	0.131*	0.184**	0.320***
	(1.724)	(2.737)	(0.044)	(0.071)	(0.076)	(0.113)
Share of land sown with Wheat in 1914	0.126	0.014	0.037**	0.035**	0.080***	0.075**
	(0.521)	(0.609)	(0.015)	(0.016)	(0.029)	(0.030)
Panel F.			*Including Land-gini*			
Share of the European-born population in 1914	5.643***	8.777***	0.091**	0.159**	0.207***	0.378***
	(1.631)	(2.512)	(0.045)	(0.074)	(0.072)	(0.114)
Land-gini in 1914	-1.089*	-1.467**	-0.004	-0.013	-0.022	-0.043
	(0.593)	(0.688)	(0.014)	(0.017)	(0.027)	(0.032)
Province Fixed Effects	Y	Y	Y	Y	Y	Y
Geographic Controls	Y	Y	Y	Y	Y	Y
Socio-economic controls	Y	Y	Y	Y	Y	Y
Observations	136	136	136	136	136	136

Note: IV regressions with robust standard errors in parentheses. All regressions include geographical control variables socio-economic controls and province fixed effects. In columns 1, 3 and 5 the IV is constructed using the actual number of Argentinians in 1869. In columns 2, 4 and 6 the IV is constructed using the average number of Argentinians in 1869 for those provinces with Argentinian population in 1857. *** p<0.01, ** p<0.05, *

8 Appendix

GDP at the county level

The Argentinian Statistical Office (INDEC) computes GDP at the national- and *provincial*-level, but not at the county level. Provincial Statistical Offices may or may not compute GDP at the county level, as they vary in their methodology and the accuracy of their procedures. Only two *provinces*, Buenos Aires and Santa Fe, out of the four *provinces* that this study covers (Buenos Aires, Santa Fe, Córdoba and Entre Ríos), compute GDP at the county level. During 1994, INDEC conducted the National Economic Census ("Censo Nacional Económico" or CNE), censusing all businesses at the county level. The census gathered information on production, employment, revenue, costs, investment, etc. The unit of observation was the *premises*, the physical space used for an economic activity. The statistical unit of observation was the firm. All businesses in Oil and Natural Gas, Mining, Manufacturing Industries, Electricity, Gas and Water, Retail and Wholesale, Financial Intermediation, Communication, Eterprise Service Providers and Personal Service Providers, were censused.

The CNE records the value of the gross product at the county level, for the abovementioned sectors. This number by itself proxies for GDP at the county level. Indeed, for those *provinces* that compute GDP at the county level, the correlation between the two measures is 94.68% for all counties in Buenos Aires and Santa Fe, and 95.95 % for the sample of counties used in this study (106 observations).

Since agriculture is not being censused and because of its importance for this area, not accounting for its value may be particularly problematic. The Ministry of Agrigulture (MAGyP) performs yearly agricultural output estimates at the county level. Adding the value of the agricultural output to the CNE gross product accounts for this relevant sector, although with one drawback. Since I observe only the value of the agricultural output, I am not discounting for intermediate goods and services used by the agricultural sector, thus overstating the value of the agriculture.

For each *province*, I consider the major agricultural products (wheat, soybean, corn, etc.)

in 1994, accounting for 84% to 96% of the sown area. For each county, I compute the value of agricultural output as the sum of each crop times its price (prices are from FAOstat). The CNE gross product augmented by the agricultural output is used as a proxy for the GDP. For those *provinces* that compute GDP at the county level, the correlation between the two variables is 94.65% for all counties in Buenos Aires and Santa Fe, and 95.05% for the sample of counties used in this study (106 observations).

The correlation between the two proxies for GDP with and without agricultural output is 99.40% for all counties and 96.56% for the sample of counties used in this study (106 observations). The regression of actual GDP on the proxy for GDP has an R2 of 90.34 (adding *province* fixed effects does not change this result) a coefficient of 1.04 with a standard error of 0.033.

Fertility, Mortality and Moving Rate

For the mortality rate, I compare the stock of Europeans in 1914 with the flow of Europeans from 1857 to 1914, and assuming that Europeans die at a constant rate δ, I solve for δ such that $\sum_{t=1857}^{1914}(1-\delta)^{1914-t} \cdot x_t = X_{1914}$, where x_t is the number of Europeans that arrived at time t, and X_{1914} is the stock of Europeans in 1914.

For the fertility rate, I proceed as follows: given the Argentinian population from the 1869 and 1914 censuses, and given that children of Europeans are considered Argentinians, I solve for ρ such that:

$w_{1870} = (1-\delta+\rho) \cdot w_{1869} + \rho x_{1869}$,

$w_{1871} = (1-\delta+\rho) \cdot w_{1870} + \rho x_{1870} = (1-\delta+\rho)^2 \cdot w_{1869} + (1-\delta+\rho) \cdot \rho x_{1870} + \rho x_{1869}$,

$\vdots$

$w_{1914} = (1-\delta+\rho)^{1914-1869} \cdot w_{1869} + \sum_{t=1869}^{1914-1}(1-\delta+\rho)^{1914-1-t} \cdot \rho x_t$,

where w_t is the number of Argentinians at time t.

The moving rate I compute is an underestimate of the true moving rate, for there is information only on province of birth and not county of birth. Using individual-level data from the 1895 census, I estimate the fraction of Argentinians living in a different province

than the one in which they were born (since there is no county-level information). Define $\pi_{i,a}$ as the fraction of people aged a born in county i, who still live in county i.

$$\pi_{i,a} = \frac{p^i_{i,a}}{\sum_j p^i_{j,a}},$$

where $p^i_{i,a}$ is the number of people born in county i who live in county i, and $p^i_{j,a}$ is the number of people born in county i who live in county j.

Then,

$$\pi^i_{i,a} = (1 - \phi_a)^a.$$

I will compute ϕ_a for all ages and then compute the average ϕ weighting by the fraction of people in each cohort.

$$\phi = \sum_{i=1}^{I} \sum_{a=1}^{99} \frac{p_{i,a}}{\sum_i \sum_a p_{i,a}} \cdot (1 - \pi^{1/a}_{i,a}),$$

where $p_{i,a} / \sum_i \sum_a p_{i,a}$ is the fraction of a years old in the population.