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

Federico Droller†

November 2013

**Abstract**

This paper analyzes the impact of population composition on long-run economic development by studying the European migration to Argentina in the late nineteenth century. I use an instrumental variable 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. Results show that historical population composition caused differences in current economic outcomes. Counties with historically higher shares of European population currently have higher per-capita GDP. Consistent with this result 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: N16, N66, O11, O15.

*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 seminar participants at LACEA-PEG 2012 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 some regions achieved higher standards of development raised a number of different explanations. 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 large population movements and colonialism contributed to change the composition of the population in the world (Putterman and Weil 2010). Empirically assessing the effect of 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 in the *Pampas* of Argentina. 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 fertile plains, the *Pampas*, was greatly influenced by the arrival of European immigrants, although the exposure of counties to the arrival of European immigrants was not equal. Within the *Pampas* the composition of population, as measured by the share of European born population, varies considerably across counties.

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 net-immigration of Europeans. Thus, the IV is synthetic measure of the share of European born population in a given county.

As shown by a large body of literature, some historic events are important determinants of current economic development (Nunn 2009). In particular, the process of settlement and population that countries followed during and after the colonial period was critical for subsequent economic development (Acemoglu et al. 2001, 2002; Banerjee and Iyer 2005; Engerman and Sokoloff 1997, 2002; La Porta et al. 1998, 1999). Europeans established themselves in different places, with different climates, many with large native population, others with scarce or no natives. Depending on these circumstances the composition of the

population varied considerably (Putterman and Weil 2010). In general, as noted by Easterly and Levine (2009) places with more European settlements in the past tend to outperform in various measures of development in the present. A number of alternative hypothesis were proposed to explain why some regions attained higher standards of development, although in light of the empirical evidence the discussion is not yet conclusive.

One hypothesis relates to the kind of political and economic institutions that countries have. Engerman and Sokoloff (1997, 2002) analyze the effect of initial endowments and its distribution on inequality, political power and institutions. They perform a comparison of European colonies and show that areas with unequal land holdings and concentrated political power created rent-seeking institutions that where less conducive to economic growth. In a similar line of research Acemoglu, Johnson and Robinson focus on the importance of colonial institutions for economic development (Acemoglu et al. 2001, 2002), where these early institutions are persistent and depending on their inclusive or extractive nature, are conducive or not to growth. However, with the available empirical evidence, it is difficult to differentiate these theories from other explanations that point to the *direct* effect of (geographical) endowments on economic development.[1] In particular, Sachs and coauthors emphasize that levels of development are strongly correlated with geographical and ecological variables (Gallup, Sachs, and Mellinger 1998 and 2000, Gallup and Sachs 2001 and Sachs and Malaney 2002).[2] Glaeser et al. (2004) propose another explanation that highlights a different aspect of population: knowledge and know-how, or human capital in a broader sense. They argue that human capital was brought by European settlers, and these past differences in human capital across countries explain a greater part of current differences in economic growth, a point also stressed by Easterly and Levine (2009). This paper contributes to this last line of research and provides strong empirical evidence for the importance of population composition and its characteristics for economic development.

[1] See J. Diamond's (1999): *Guns, Germs and Steel: The Fates of Human Societies* for an elaborated work on this subject.

[2] Geographical variables refer to, for example, latitude, disease ecology or distance from the coast.

In this paper I will establish the causal effect of historic population composition, measured as the share of European born immigrants in 1914, on three current measures of development: GDP, education and skilled labor. After establishing these long-run effects I explore characteristics of the population and economic life that are related to economic development. In particular I analyze the process of industrialization and the level of literacy as channels through which initial differences in the composition of population had a persistent effect over time.

The empirical analysis exploits a particular historical setting in the fertile plains of Argentina, the *Pampas*, an area originally occupied by native indigenous population, over which the Argentine government struggled to gain power. The availability of the fertile plains to potential settlers varied over time depending on the civil and international conflicts and on the success of military campaigns to conquer the plains.[3] European migration to Argentina was restrictive over the colonial period and only started years after independence, 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, otherwise an area with geographically similar characteristics and common political institutions, were shocked in varying intensity by European immigrants. The shock to the population was far from being negligible, areas experienced different intensities of immigration and the percent of European population in each county after the shock ranged from 0% to 50%.

Using a predicted measure of the exogenous share of European population as an instrumental variable for the actual share of European population, I compare counties in the fertile plains and estimate that an increase of 11% (one standard-deviation) in the share of European population raises per-capita GDP by 60% in the long-run (0.77 standard deviations). Similar results hold for education and skilled labor: areas with higher share of European immigration in 1914 have a higher share of population with higher education, as well as

[3]The process of settling the *Pampas* drastically contrasted to 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 the government conducted a census.

skilled workers in 2001.

Consistent with these results I argue that characteristics of the population related to the level of human capital played an important role for development. In particular, immigrants were closely related to 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 establishment, the employment of high- and low-skilled industrial workers and the investment in energy where substantially higher in regions where the intensity of immigration was higher.

Further, I show that this same areas where Europeans accounted for a higher share of the population had higher literacy rates. The evidence suggests that immigrants not only contributed with their higher literacy, but generated a positive externality by raising the level of human capital in the population as a whole in the early 1900's.

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: European immigrants are associated with greater industrialization and higher literacy for the population at large. Moreover, I show that initial difference in the composition of the population can have a long-lasting effect on the level of development.

This paper is organized as follows, Section 2 reviews the conquest of the fertile plains and the European immigration to Argentina. I provide a historical account of the reasons that motivated military campaigns to the *Pampas* and the timing of these campaigns. Further, I describe the process by which the plains were settled. Section 3 describes the data, its sources and describes how geo-referenced data was computed for this study. The next section develops the empirical strategy and presents the results. In the beginning of section 4 I show OLS estimates and in section 4.1 I proceed to develop the instrumental variable approach. I implement my IV and show the causal effect of migrants on long-run development in section

4.2. Next in section 4.3 I show two channels of persistence: industrialization and human capital. In section 4.4 I perform a series of robustness checks: I consider variations to the parameters of the demographic model, I consider the direct effect of exportable crops and land inequality and also I weight the observations by the population of the county. Section 5 concludes.

# 2 The History of the Fertile Plains

## 2.1 The Conquest of the Plains: the Desert

It was not until end of the nineteenth century that the Argentinean government gained political power over the whole territory that nowadays 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 Argentinean government. Relationships between Argentineans and indigenous tribes were characterized by mistrust and violence. By the time of independence the situation was such that Argentineans used to dispute land and wild livestock to the indigenous tribes, while indigenous people organized assaults into settlements and cities, stealing livestock, goods and kidnapping people. Indigenous raids attacking cities and military excursions into indigenous settlements, both ending in destruction and deaths, were common. The Argentinean government and main tribes often agreed on peace treaties, but the Argentinean government never recognized that area as an independent state, nor did it recognize indigenous people as legal owners of the land.

The threat of indigenous tribes over Argentinean settlements was not the only concern of the government regarding the national territory. For Argentina to consolidate as a nation it was necessary to delimit its frontiers, which turned 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 until 1885. Previous to 1870, military campaigns

developed with many years of interruption and loss of domain, in particular during episodes of civil war and the war against Paraguay. Detailed information on the military campaigns and its effect on how the internal *frontier* between Argentineans and the indigenous tribes changed over time has been documented by Walther (1964). 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 territory by the Argentinean army and losses of domain over these years were a consequence of the limited resources the government had for the multiple military conflicts it faced (Luna 1993).

## 2.2 Settlement of the Fertile Plains

The end of the civil war and the re-unification 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 Argentineans, 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 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 it is a monotonic function of time.

Immigrants settled in cities, urban areas and in the countryside, and were occupied both as skilled labor or unskilled labor. Activities 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 1895, once military resources were not longer used in civil or international wars. At the same time, the peace achieved in the country and the economic conditions in Europe motivated Europeans to migrate to Argentina. Between independence and the reunification of the country, a period

close to fifty years, civil war prevented many Europeans of migrating to Argentina.[5] 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. Although concerns might be raised on Europeans migrating to Argentina because of the growing availability of land, the data doesn'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 and R-squared of 20%. Temporary and permanent workers migrated mostly to the fertile plains, some of then coming back to Europe after the harvest in the southern hemisphere (right before the harvest in the northern hemisphere) and some of them settling down and bringing the rest of their families over time. Progress and well being among immigrants was not immediate, but not hard to achieve.

# 3 Data and Summary Statistics

This study combines current data on economic development (per-capita GDP, higher education rate and share of skilled workers) with historical data on economic and social conditions (population density, productive uses of land, etc.). The unit of observation is at the county level. The sample covers the four provinces that hold the fertile plains: Buenos Aires, Santa Fe, Córdoba and Entre Ríos. The southwest section of the fertile plains lays 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 being 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 of time considered in this study. Working with four states allows me to control for unobservable fixed variables at the state level.

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

Though county boundaries have changed in some counties over time, it is possible to match older counties to new counties. New counties were mostly founded 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, where 31 are new counties not linked to an old county. From the remaining 166 counties, 25 are capital cities or large urban areas and 5 are counties without current information on economic outcomes. Excluding capital cities and large urban areas, the final sample has 136 counties in four states.

Historical information on population comes from three sources: the 1895 and 1914 Argentinean censuses and the Argentine Office of Migration. Both censuses contain detailed information at the county level on population characteristics and economic activities. I digitalized 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 European born population is computed as the number of European born population divided by the total number of adults in the population.[6]

The Argentine Office of Migration records since 1857 all non-Argentine incoming and outgoing population. Detailed data on the number of migrants and country of origin since 1857 until 1914 was 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

[6]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, Argentineans. I computed the share of Europeans among the population older than 14 years of age. Given data availability, those younger than 6 years of age are also included, an issue that affects the denominator, since there were few, if any, Europeans that young.

with a series of maps that show for different years the actual *frontier* between the territory under the Argentinean 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) who provide more detailed information for the states of Córdoba and Santa Fe.

The Argentinean 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) censing all business 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 agricultural 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 with 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 province's 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 the number of industrial establishments, the value of the production, the number of workers and the investment in energy production at the county level.

Data on higher education rates and 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 comes 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 the geographical conditions of the soil (like 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 the average rain and temperature comes from Worldclim,[7] 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 5 km, data on railroads comes from *ATLAS de Suelos de la República Argentina.*[8]

Table 1 shows the summary statistics for the variables used in this study. As a measure of the intensity of European immigration I construct the share of European population, defined as the fraction of European born population in 1914. The average (and median) share of European population is 23% (16%) and a standard deviation of 11%, with counties ranging from less than 1% to 47% of its population of European origin. Average GDP per capita is slightly above 6.700 dollars, where the bottom 25% of the counties have less than 3.560 dollars and top 25% of the counties have a per-capita GDP above 9.000 dollars. On average 10.4% of the population 25 years of age and older have 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 current development (log per-capita GDP, higher education rates and the share of skilled workers) across counties with different population composition in the past. I start by running a regression of the dependent variable

[7] See http://www.worldclim.org/formats.
[8] See Cruzate et al. (1990).

on the share of European population and also include other control variables:

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

Where $y_i$ is the dependent variable in county $i$, $SE_i$ is a measure of the composition of the population defined as the share of European population in county $i$ in 1914, $X_i$ are controls for county $i$ characteristics in 1914, and $\eta_s$ are state fixed effects. County characteristics include population density, share of the population living in urban areas (2000 or more inhabitants), share of productive land used for agriculture, land-quality and (log) distance to the city of Buenos Aires.[9] I also control for geographical characteristics (mean temperature, rainfall and ruggedness) and for the availability of railroads.

Table 2 documents OLS results from the regression of log per-capita GDP in 1994 on the share of European population in 1914, equation (1). Column 1 only controls for state fixed effects, column 2 adds controls for the distance to the city of Buenos Aires, density of railroads, the share of productive land used for agriculture, population density and urbanization rate. Column 3 adds geographical controls (rain, temperature, elevation, ruggedness and land quality). The basic OLS regression shows that the share of Europeans in 1914 has a positive and significant coefficient. In column 3 distance to Buenos Aires has a coefficient statistically not different from zero and density of railroads has a positive coefficient. Land quality has a positive (though not different from zero) effect on development, and the share of productive land used for agriculture enters positively. Population density enters negatively, while urbanization has a positive coefficient but not statistically different from zero.

Following column 3, the preferred specification, a one standard deviation in the share of Europeans increases per-capita GDP by 0.55 standard deviation. As this result shows, European immigration positively correlate with economic development in the long-run, close to eighty years after the arrival of European immigrants differences in economic performances

[9]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 political and economic relevant city may have independent effects on development.

can be found across counties depending on the pattern of settlement. The evidence presented in table 2 is based on correlations, and its interpretation has to be taken with caution. If European immigrants selected themselves into the counties depending on an omitted characteristic or an unobservable variable, the results would be biased. To deal with this potential problem I will use instrumental variables to account for the possible endogeneity in the selection of Europeans to the different counties.

## 4.1 Instrumental Variable Approach

European migration to the different counties in the fertile plains may not have been random. 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 went 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. To account for the possible endogeneity in where European immigrants settled once they arrived to Argentina, I will 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 will exploit two sources of variation: a) changes in the internal frontier between Argentina and the indigenous tribes; and b) changes in immigration to Argentina between 1857 and 1914. As will be discussed below, a simple demographic model will exploit the variation in both, available land for settlement and arrival of immigrants, to allocate immigrants (depending on the year of arrival) and Argentineans to counties and construct an exogenous share of European population.

### 4.1.1 The History of the Instrument

Using historical information on the military campaigns followed by the Argentine government, 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 in the past, until the government gained power over the whole territory. Walther documents the end result of military excursions and the boundary that resulted of these expeditions between the Argentine government and the indigenous tribes, in a series of maps, Figures 2-3 being two examples of it. By 1884 the Argentine government controlled the rest of the fertile plains. I assume that no land is conquered or lost until the next military campaign, an assumption very close to the actual events. I overlap county boundaries to these maps and establish the date in which the boundary moved such that a county started to be on the Argentinean side.[10]

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, and also were looking for better prospects of living. In particular comparing the time series of migration to Argentina and the USA the correlation is 0.795.[11]

An ideal experimental setting would consist of regions (counties) that are equal in all respects, and have a given number of Argentinean population. These regions are then randomly shocked with European population in different intensities. I could analyze economic and social development in these regions in the long run, and see whether there are differences to be explained by the share European population, the only variable that varies across regions. The actual empirical setting I am analyzing approximates very closely my ideal experiment: it consists of regions that are geographically uniform, had an initial stock of Argentinean population and were shocked by European population in different degrees. The key difference is that Europeans were not randomly distributed as they choose where to settle. The IV I construct is based on a synthetic distribution of Europeans across counties,

[10]The date a county *enters* Argentina has not to be confused with the date in which a county is officially founded, usually years after it was under the Argentinean power.

[11]Data on USA migration 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.

using variation in the timing of seizure of land from the indigenous tribes interacted with the arrival of Europeans. In particular, for the shock of European population to be exogenous in my analysis I need that Europeans decided to migrate to Argentina for reasons unrelated to the success or failure of the military campaigns in conquering new land, and that the decision by the government 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 power of this region, starting years before the first wave of European immigrants arrived; the military campaigns in the fertile plains ended by 1884, when slightly less than 900,000 immigrants had arrived to Argentina, in comparison to circa 3million net-immigrants immigrants that arrived by 1914. 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.

### 4.1.2 The Instrument

The information used for the IV is the number of Europeans that arrived each year, the number of counties available each year and the initial number of Argentineans in a given county. This information needs to be combine and summarize in a cross-section of counties in 1914. Next I will continue defining the process of population growth that will be used for the construction of the IV.

For the construction of the IV, starting in 1857 Europeans will be distributed uniformly across counties. The quantity of immigrants each county is assigned varies by year of arrival, according to the time series. Europeans arrive every year and move uniformly to any county that is under the political power of Argentina, and once they settled they never move again. Europeans die at rate $\delta$ and reproduce at rate $\rho$, although children born to Europeans in

Argentina are considered as Argentineans.[12]

Argentineans, on the other hand, are initially present in counties under the political power of the Argentine government by 1857, but not in counties conquered after 1857. The initial Argentinean population in 1857 comes from the 1869 census, adjusted by the population growth rate to the year 1857. Argentineans die at rate $\delta$ and reproduce at rate $\rho$. There is a fraction $\phi$ of Argentineans that each year decides to move to a new county. I assume they move equally to all the counties that belong to Argentina.

The mortality rate, the fertility rate and the fraction of Argentineans 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 Argentineans, $\phi$, is computed to be equal to 1.95%.[13] The first stage and the analyses in the coming section are robust to changes in the parameters of the demographic model, as well as changes in the assumption on the initial Argentinean population. All these possibilities will be considered as robustness checks in Section 4.4.

The constructed number of Europeans in each county in 1914 is defined as:

$$CE_i = \sum_{t=1857}^{1914} \frac{1}{N_t}(1-\delta)^{1914-t} e_t \cdot \mathbb{1}_i\{t \geq D_i\}. \tag{2}$$

The constructed number of Argentineans in each county in 1914 is defined as:

$$CA_i = CA_{i1857}(1-\delta+\rho-\phi)^{57} + \sum_{t=1857}^{1914} \frac{1}{N_t}(1-\delta+\rho-\phi)^{1914-t}(\phi a_t + \rho e_t) \cdot \mathbb{1}_i\{t \geq D_i\}, \tag{3}$$

where $CE_i$ and $CA_i$ are the constructed number of Europeans and Argentineans in county $i$ in 1914, respectively. $e_t$ is the number of Europeans that arrived in year $t$, and $a_t$ is the number of Argentineans that move to a different county in year $t$. $CA_{i1857}$ is the initial number of Argentineans in a given county. $\mathbb{1}_i\{\cdot\}$ is an indicator whether county $i$ belongs to

[12] From 1857 until 1914.

[13] See the Appendix for a detailed explanation on how to compute these values. Values for the fertility and mortality rates do not differ from those listed on historical records.

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

The constructed share of Europeans population is defined as $CSE_i = CE_i/(CE_i + CA_i)$, and is used as IV for the actual share of 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 $i$ started to be 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 $CA_i$ not only depends on $D_i$, the number of Argentineans moving, $\phi a_t$, and the children of Europeans, $\rho e_t$, but also on the initial stock of Argentinean population, $CA_{i1857}$. $CA_{i1857}$ is not a random variable and may depend on observed and unobserved characteristics, to account for this potential problem as robustness checks I will consider other assumptions and show that results do not depend on any particular assumption. In particular, in Section 4.4 I assume that initial stock of Argentineans, $CA_{i1857}$, is the same for all counties with positive initial population: $CA_{i1857} = \bar{CA}_{1857}$ if $CA_{i1857} > 0$, and $CA_{i1857} = 0$ otherwise. I will also consider the case in which all counties are assigned the same initial stock of Argentineans, $CA_{i1857} = CA_{1857}$.

As mentioned above, Walthers account of the conquest of the plains can be summarize in 8 waves of land incorporation for the following years: 1779, 1823, 1826, 1860, 1864, 1869, 1876 and 1884. Figure 5 shows the distribution of counties over time: 66 counties already existed at the independence, while six were conquered in 1860, seven in 1864, eleven in 1869, eleven in 1876 and five in 1884.

## 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.

Figure 6 shows the first-stage relation between the share of European population and the constructed share of European population. Figure 7 shows the first-stage correlation when control variables and fixed effects are included. Both figures show a strong positive correlation between the two variables.

Table 3 shows the first-stage regression, equation (4). In column 1 controls for $X_i$ and no geographical controls are included, column 2 adds geographical controls, while in column 3 standard errors are clustered at the year of incorporation, $D_i$. The coefficient on the constructed share of immigration remains positive and significant across specifications, confirming the result presented in figures 6 and 7. An F-test of the coefficient $\psi$ shows a strong first-stage with a statistic greater than 30 for the full specification in column 3, and weak identification is ruled out by the Kleibergen-Paap test of 34.1.

Table 4-6 show results for three different dependent variables, where the constructed share of immigration is used as instrumental variable for the actual share of European population. I report results for three specifications discussed above: not including geographical controls (column 1), controlling for all variables (column 2) and clustering standard errors at the year of conquest level, $D_i$, (columns 3). In table 4 columns 1-3 the dependent variable is log per-capita GDP in 1994. The coefficient on the share of Europeans in columns 1-3 shows a long-run effect of the share of European population on per-capita GDP, with a significant coefficient. One standard deviation in the share of European population increases per-capita GDP by 0.77 standard deviations.[14] The point estimate of 5.49 is slightly higher than the OLS estimate of 3.91, suggesting a negative bias in the selection of Europeans to counties and/or measurement error. The effect of having relatively more European has an important effect in the long-run, an increase in the share of Europeans of 5% raises per-capita GDP by one third of a standard deviation. For a county like *Río Cuarto* with a share of Europeans of 20%, increasing the share to 25% would raise per-capita GDP from 6912 dollars to 9097.

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

Certainly an economically significant effect.

Columns 1-3 of Table 5 examine census data on higher education in 2001. Results also show a positive and significant effect of European immigration on this variable. One standard deviation in the share of European immigration raises the share of population with higher education by 0.49 s.d., an effect significant at the 5% level. Table 6 columns 1-3 repeats the analyzes for the share of workers in high skilled occupations. Results show a positive effect: one standard deviation in the share of European immigration raises the share of workers in high skilled occupations by 0.51 standard deviations, a result significant at the 1% level. The results in tables 4-6 show an important causal effect of European immigration over the long-run: Europeans affected the degree of economic development as measured by GDP, higher education and skilled workers. The intensity of European migration appears to have created a divergence in the paths of economic development across counties. Next I will be examine the channels through which development diverged and persisted over time in the next section.

## 4.3 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? To answer these questions I will next investigate two channels through which European immigration created differences in the paths of economic development over time: the process of Industrialization and literacy rates. Both channels are linked together and show two different aspects of Human Capital. Measures of human capital back in time are not easy to compute, years of schooling or population with graduate education were not surveyed at that time in Argentina. Therefore, although literacy rates can be regarded as a noisy measure of human capital, literacy still captures how instructed the population is. Industrialization, on the other hand, captures the development of economic activities in the society for which certain level of

human capital is needed. Even though we can not observe the knowledge and skills these industrial entrepreneurs had, we observe the result of their effort: the industries they forged, the industrial output and the investment.

### 4.3.1 Industrialization

Industrialization has been widely understood as an important factor in a country's development, countries that industrialized earlier rank higher in todays development, per-capita income and living standards. Since the Industrial Revolution higher standards of development have been link to the degree of industrialization of the economy, where terms as *industrialized nation* or *developed nation* and *advanced economy* have been used interchangeably. In the case of Argentina, industrialization arose in some counties more than in others, and cities that developed more were also cities that experienced higher industrialization in the beginning of the twentieth century. Why industrialization arose in the first place is an open question, but from the industrial census 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 operates as a vehicle that propagates development over time, and long-term differences across regions emerge 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 the workers employed 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 mostly centered around the production of garment, food, wooden, metal and chemical products, and construction. Table 6 also shows that still in 1935, 58% of the industrial establishments were 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 twenty years earlier.[15] The 1935 industrial census records information at the establishment level and at the county level. The census provides me with four measures in industrial development: the number of establishments per person, the share of skilled workers in the population, the per-capita value of production and investment in energy per person.[16] [17] In table 8 I examine the effect of the share of European immigration on these variables, using the instrumental variable described above for the share of European population, columns 1-4 show the second stage results. The share of European population has a positive and significant effect on all industrial variables. Following columns 1-4, one standard deviation (SD) in the share of European population raises the value of industrial production by 0.66 SD, the share of skilled workers by 0.85 SD, the number of factories per person by 1.04 SD and the energy in horse power per person by 0.64 SD. For a county like *Río Cuarto*, having a share of Europeans of 25% instead of 20% would have raised the value of industrial production in 1935 by 41%.

Tables 7 and 8 show the importance of the European population in the process of industrialization, in 1895, 1914 and 1935 the fraction of industrial firms owned by Europeans was above 50%, industrial workers were mostly of European origin and counties that happened to have a greater share of their population of European origin experienced greater industrial output and assigned more resources to industry: workers and investment in energy production.

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 we do not observe in the population the level of human capital or the knowhow of a particular industrial activity, we do observe that industrial development was closely linked to the intensity of European immigration.

[15] 1935 is the first industrial census for which data at the county level is available.
[16] Value of production is in 1935 peso currency and energy is measured in horse power.
[17] For the per person variables I consider the 1914 population, since it is the closest population census.

### 4.3.2 Literacy rates in 1914

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 contributed to the rapid growth of per-capita GDP. Contemporary differences in human capital have been shown to affect development at the macro- and micro-level, but evidence pointing to the effect of historic differences in human capital on development in the long-run is scarce. Glaeser et al. (2004) find evidence for human capital as a channel for growth and better political institutions and Easterly and Levine (2009) point out that human capital was an important intermediating channel through which colonial settlement affected development in the long-run. The measure of human capital that history provides me with is the literacy rate as of 1914. Next I analyze how the composition of the population affected early levels of human capital.

The level of human capital at the end of the nineteenth century, beginning of the twentieth century was altered by the inflow European immigrants. Differences in literacy rates were greater within Europeans than between Europeans and Argentinians. Table 9 examines literacy rates in 1914 by nationality, as reported by immigrants in Argentina: while the Argentinean population is on average 63.2% literate, Germans are 88.2% literate and immigrants from Italy, Spain and France are 59.6%, 67.4% and 79.3% respectively. When weighted by population, on average Europeans are 64.2% literate and the population as a whole is 63.3% literate. Europeans migrating to the *Pampas* were on average more literate than locals, but the difference appears to be small. 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 in 1914 at the county level and the share of European population. Column 1 shows IV estimates, once the endogenous distribution of immigrants is accounted for, the share of European immigration has a positive and significant effect on literacy rates, the coefficient of the IV regression is 0.07. This coefficient implies that one standard deviation in the share of European population rises

literacy rates by 0.15 SD. Continuing with our example on *Río Cuarto*, if the share of Europeans would have been 5% higher, the literacy rate would have been 0.35% higher, raising from 57.1% to 57.5%.

The question that tables 9 and 10 raise is what explains this difference in literacy rates across counties? Can this difference be explained by a composition effect, namely by substituting a less literate Argentinean by a more literate European? Or is the effect of immigration on literacy the consequence of an increase in the acquisition of human capital? As documented in table 9 on average Europeans are 1.1% more literate than Argentineans, implying that switching 1% European population for 1% Argentinean population will automatically raise literacy by 1.1%. The effect of 7% shown in table 10 column 1 is far greater than 1.1%. The composition effect can explain part but not the whole difference in literacy rates across counties. Beyond the composition effect, immigration has a positive externality on literacy rates on the rest of the population. There are several potential explanations for this: it may be that Europeans provide more education to their offspring, it may also be related to Europeans demanding more schools in the places were they settled and afterward schools provide education to all citizens, or the Argentinean government providing education to the newly arrived immigrant, or it may also be the case were economic progress generated a demand for more skilled labor, providing higher incentives to acquire human capital. In accordance to the results provided in the previous section, places were Europeans accounted for a higher share of the population had higher literacy rates in 1914, partly due to more literate immigrants and partly due to a positive externality on the rest of the population (their children and others). In the next section I will investigate if more education was provided in areas with a higher share of European immigrants.

**European Immigration and Human Capital formation in 1914**

Complementary to higher literacy rates, it is relevant to analyze whether more education was provided in areas with higher shares of European immigrants. Were counties with a

higher share of European population more literate because of school availability? Did the Argentinean government promote education in areas with more Europeans to assimilate them to the native population? Are counties with higher literacy the results of public financed education, or the result of private financed education? I will proceed to answer these questions using 1914 census data on the number of students and the number of public and private schools.

Since mid-eighteenth century schools were built through the country by the government, offering free public education to all individuals in 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 active policy of public education pursued by the government, it is plausible that counties with a higher share of Europeans experienced more public financed education in order to assimilate and introduced immigrants into the Argentinean society. However, census data shows 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 schooling age population).

Census data on schools in 1914 lists schools' locations and the school-age population in each county, from which I construct the number of schools per 1000 school-age children, on average there are 5.3 public schools and 0.85 private schools in each county per 1000 school-age population, with a standard deviation of 2.32 and 0.71, respectively. Using the same IV empirical strategy as in the previous sections, in table 10, I regress the number of public schools and private schools on the share of European immigrants, including controls for county characteristics.

Column 2 shows IV estimates from regressing the number of public schools per school-age population on the share of European population. The share of European population has a negative and significant effect on the number of public schools, one standard deviation in the share of European population reduces the number of public schools by 0.61 standard de-

viations, a magnitude equivalent reducing 1.5 public schools. Column 3 shows IV estimates, but this time for the regression of the number of private schools on the share of Europeans. Results show a positive, although not significant, effect of immigrants on the quantity of schools, one standard deviation in the share of immigrants increases by 0.38 SD the number of private schools per school-age population.

The findings on column 2 show that government educational policy was not targeted to areas where Europeans concentrated, quite the opposite, an increase by 0.11 in the share of Europeans is associated with a reduction of 1.5 public schools. On the other hand, the share of Europeans has a positive but not significant effect on the number of private schools. The evidence points to literacy rates being higher in areas with more Europeans not because of educational policies pursued by the national government, but because of individual decisions of the citizens of these counties.

## 4.4 Robustness Checks

The results in this paper are robust to a series of variations in the specification and construction of the Instrumental Variable. Next I consider changes on the assumptions of the demographic model, as well as alternative explanations for the divergence in economic growth.

In table 11 I consider 6 variations to the demographic model presented in section 4: First, I consider in column 1 IV results when the initial stock of Argentineans is fixed among counties with $W_{i,1857} > 0$, and 0 otherwise, namely I assume $W_{i,1857} = 6269$, the average number per county of Argentineans in 1857. Alternatively, in column 2 I consider the case where all counties have an average initial number of Argentineans equal to 3600. Changes to the assumption on the initial stock of Argentineans does not affect the results.

Further, I also consider different values for the parameters of the model, in particular I set the values of the parameters arbitrarily high (double): in column 3 the moving rate $\phi$ equals 6%, in column 4 the fertility rate $\rho$ equals 10%, in column 5 the mortality rate $\delta$ equals

6% and in column 6 $\phi = 6\%$, $\rho = 10\%$ and $\delta = 6\%$ simultaneously. Columns 3-6 in table 10 show that results remain consistent with my main results, changes in the assumptions of the model do not alter the effect on per-capita GDP (results for all the other variables considered in this study are also robust to these changes).

In table 12 I consider alternative explanations to the divergence in the paths of economic development. Following the literature I explore whether land inequality and production of wheat, a highly valuable crop, can have a direct effect on development and explain the observed differences between counties. Column 1 shows results from an IV estimation, where I added to the estimation the share of land dedicated to the production of wheat in 1914. Wheat was at that time one of the most valuable crops and one therefore an important economic activity. Results show that the share of land dedicated to wheat does not have an effect on current per-capita GDP, nor does it alter the effect of the share of Europeans encountered before.

1914 census data on the size of plots and distribution of land in counties is used to compute a measure of land inequality: a land-gini. Columns 2 shows that land inequality has a long-term negative effect on development, although it does not change the previous interpretation on the effect of European immigration. The IV results in column 2 reinforces the conclusion drawn above.

Finally, in column 3, I repeat the main IV regression of the paper but weighting by the population of the county as of 1914. Relative differences in the population size of a county may be relevant to assess the effect of the population composition on development. As column 3 shows, weighting for the population does not change the results.

In sum, the regressions shown in the previous sections are robust to the inclusion of other potential relevant variables, changes in the parameters of the model and weighting by population.

# 5 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 fewer share of the population? As I discuss in this paper, the *Pampas* provide an excellent empirical setting to answer this question, for it is an area with common political institutions and uniform geographical conditions across counties. Therefore it is possible to abstract from the usual competing hypothesis of differences in geographic endowments or institutions. Still, the rate of migration 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 exogenous variation in the arrival of Europeans interacted with variation in the availability of land for settlement. This synthetic measure of the population composition captures the exogenous share of Europeans in each county, not the share that would result from the underlying characteristics of the area. IV results show that there is a positive causal effect of the composition of the population on long term development. Counties where the share of European born population was higher in 1914 achieved higher per-capita GDP, and have a more educated population with a greater share of skilled workers in the present.

Motivated by this strong result I proceed to understand the role played by immigration and, in particular, whether immigrants' human capital can be linkedXXX 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 for most of the industrial skilled- and unskilled labor. Using census data on the number and characteristics of industrial establishments across the fertile plains, I show that counties

with a higher share of European population outperform in various measures of industrial development. In these counties the number of industries and workers, the value of industrial output and investment exceeds other counties where the European population was relatively smaller.

Second, I exploit variation across nationalities in literacy rates and show that counties' average literacy rate is higher in these same counties where the share of European born population is higher. 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 Europeans. Nor can higher literacy rates be explained by an effort of the national government to educate and assimilate immigrants, data on the number of public schools and students show that the opposite is true, the number of public schools per student was smaller in counties were 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 study benefits from an empirical setting, the fertile plains of Argentina, that allows to abstract from differences in political institutions and geographic characteristics, and focus on the effect of the composition of the population on economic development. The results I present show the importance of people themselves for the process of economic development and the persistence of this effect in the long-run.

# References

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

[2] 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).

[3] Banerjee, A., and L. Iyer, "History, institutions and economic performance: the legacy of colonial land tenure systems in India," *American Economic Review* 95: 1190–213 (2005).

[4] 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).

[5] Diamond, J. M., *Guns, germs, and steel: The fates of human societies*. New York: W.W. Norton & Co. (1999).

[6] 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).

[7] Easterly, W., and R. Levine, "The European Origins of Economic Development," mimeo, Brown University, 2009.

[8] 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).

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

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

[11] 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).

[12] 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).

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

[14] 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).

[15] 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).

[16] 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).

[17] 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).

[18] La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer and Robert W. Vishny, "Law and Finance," *Journal of Political Economy,* December, 106(6), pp. 1113-55 (1998).

[19] La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer and Robert W. Vishny, "The Quality of Government," *Journal of Law, Economics and Organization,* April 15(1), pp. 222-79 (1999).

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

[21] 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).

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

[23] Nunn, Nathan, "The Importance of History for Economic Development," *The Annual Review of Economics*, 1:65–92 (2009).

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

[25] 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).

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

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

[28] 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).

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

[30] 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).

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

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

Figure 1: Correlation between log per-capita GDP in 1994
and the share of European population in 1914, in Argentina.

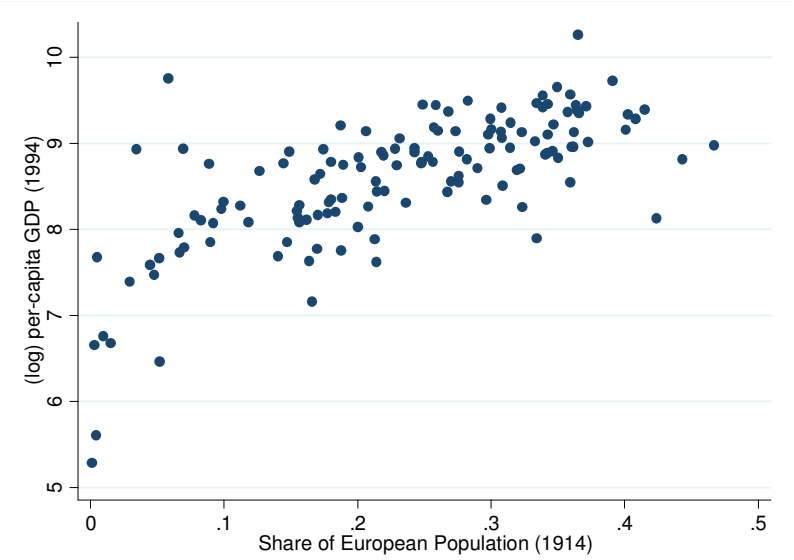

Figure 2

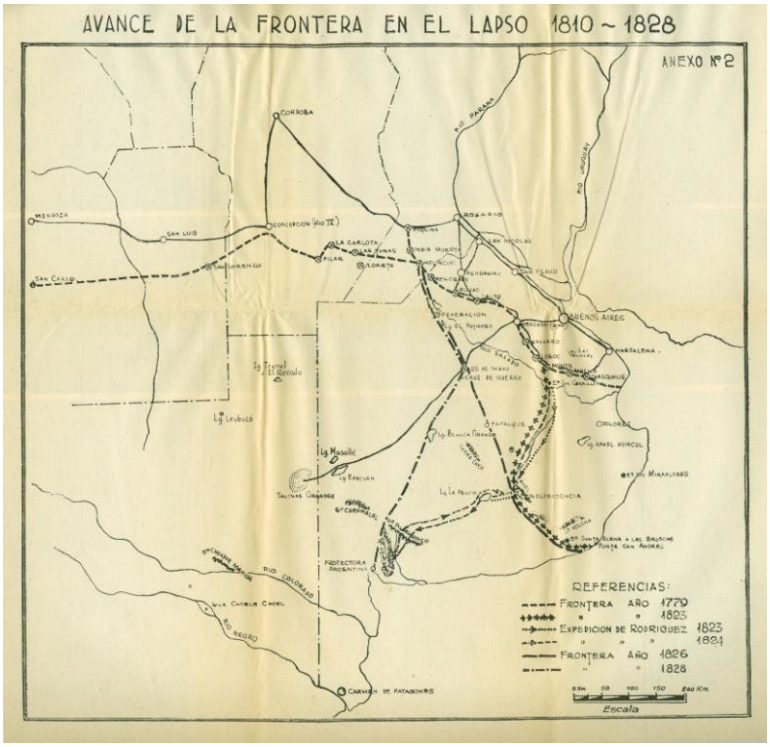


Figure 3

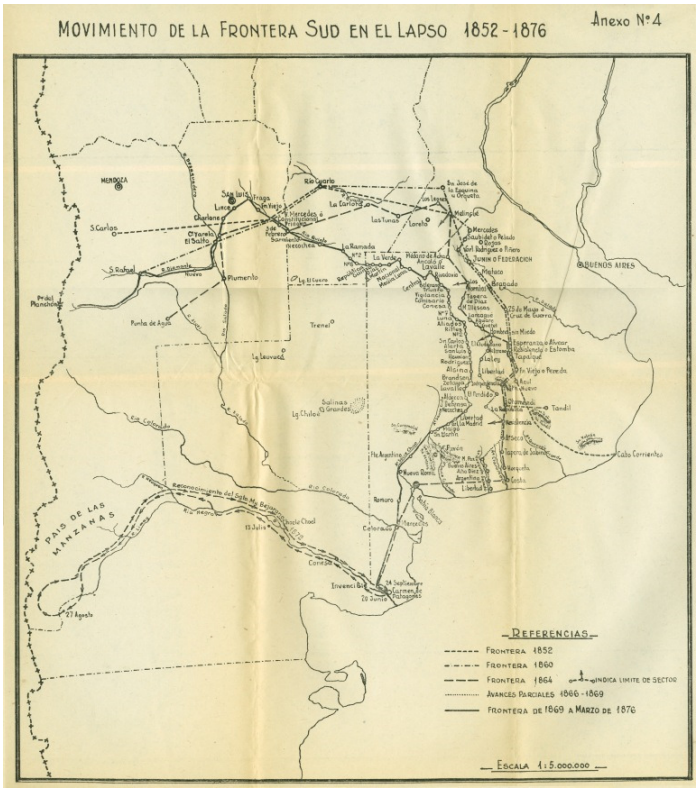

Figure 4: Immigration Time Series.

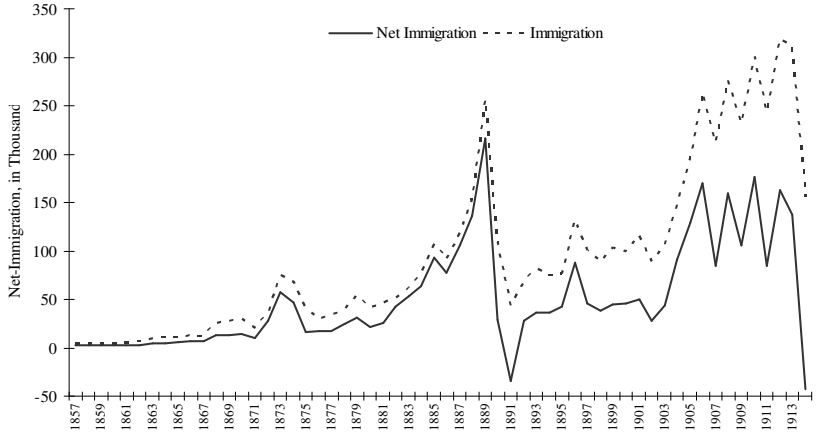


Figure 5: Cumulative Net-Immigration and Area for settlement.

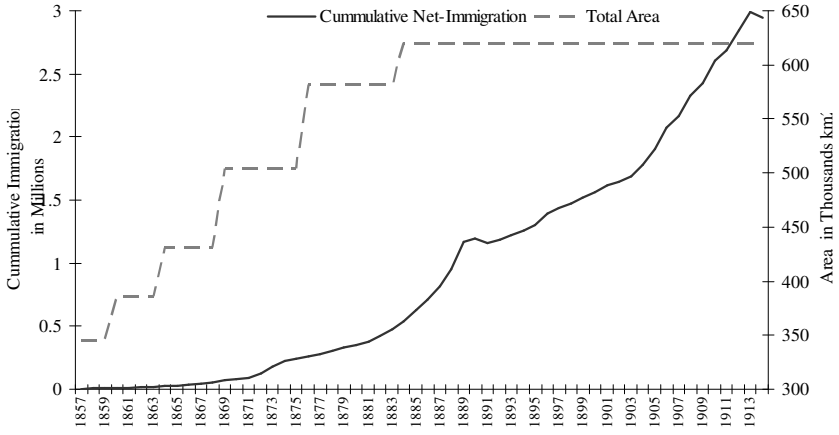

Figure 6: 1st Stage correlation between the share of European population and the constructed share of European immigration.

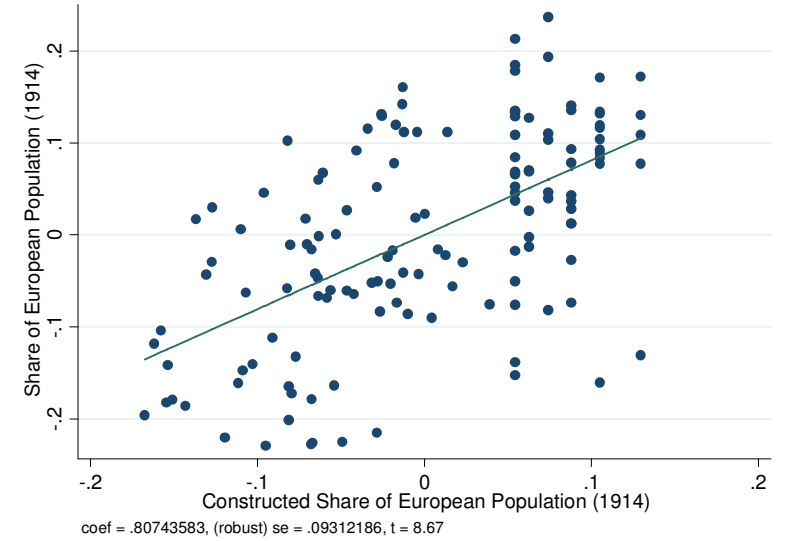


Figure 7: 1st Stage correlation between the share of European population and the constructed share of European immigration, control variables and fixed effects included.

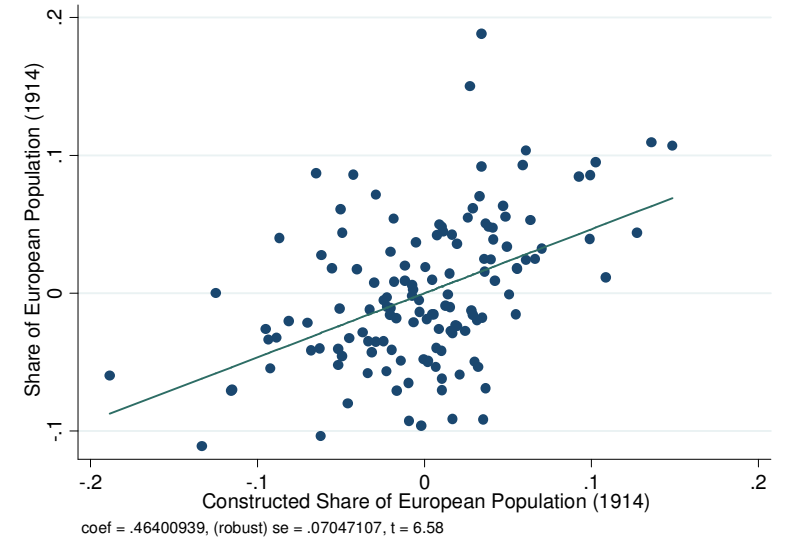

Table 1: Summary Statistics

| Variable | Mean | Standard Deviation | 50th Percentile |
|---|---|---|---|
| **Share of European population, 1914** | **0.23** | **0.11** | **0.16** |
| GDP per-capita, 1994 | 6754 | 4190 | 3560 |
| log GDP per-capita, 1994 | 8.59 | 0.78 | 8.18 |
| Share of pop. w/higher education, 2001 | 0.1 | 0.02 | 0.09 |
| Share of skilled workers, 2001 | 0.18 | 0.04 | 0.15 |
| log industrial output per-capita, 1935 | 4.4 | 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 | 0.1 | 0.14 | 0.05 |
| Literacy rate, 1914 | 0.63 | 0.05 | 0.58 |
| Number of private schools per-1000 school age pop. | 0.85 | 0.71 | 0.35 |
| Number of puclic schools per-1000 school age pop. | 5.33 | 2.32 | 3.63 |
| Number of secondary schools per-1000 individ. 2007 | 0.89 | 0.45 | 0.63 |
| 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: OLS

| Dependent Variable: | log per capita GDP, 1994 | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| **European population / total population, 1914** | **5.668*** (0.632)** | **4.403*** (0.732)** | **3.914*** (0.796)** |
| Distance to BA City | | -0.010 | 0.079 |
| | | (0.114) | (0.151) |
| Land Quality | | | 0.004 |
| | | | (0.004) |
| Railroad Density | | 0.069*** | 0.052* |
| | | (0.026) | (0.029) |
| Percent of Land used for Agriculture in 1914 | | 0.715*** | 0.644** |
| | | (0.248) | (0.293) |
| Population Density in 1914 | | -0.037*** | -0.028** |
| | | (0.009) | (0.011) |
| Urban Rate in 1914 | | 0.684** | 0.557 |
| | | (0.335) | (0.341) |
| Geographic Controls | no | no | yes |
| Province Fixed Effects | yes | yes | yes |
| Observations | 136 | 136 | 136 |
| R-squared | 0.507 | 0.561 | 0.596 |

*Note:* Ordinary least squares regressions with robust standard errors in parentheses. Dependent variable in all columns is log per-capita GDP in 1994. In column 1 only province fixed effects are included. Column 2 includes all control variables except for the geographical controls. In column 3 all control variables are included. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: First Stage

| Dependent Variable: | European population / total population | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| **Constructed European population / total population** | **0.450*** (0.084)** | **0.464*** (0.070)** | **0.464*** (0.069)** |
| Distance to BA City | 0.040*** | 0.018 | 0.018 |
| | (0.010) | (0.012) | (0.012) |
| Land Quality | | 0.000 | 0.000 |
| | | (0.000) | (0.000) |
| Railroad Density | 0.001 | 0.002 | 0.002 |
| | (0.003) | (0.003) | (0.003) |
| Percent of Land used for Agriculture | 0.221*** | 0.184*** | 0.184*** |
| in 1914 | (0.027) | (0.024) | (0.025) |
| Population Density in 1914 | 0.003* | 0.003** | 0.003*** |
| | (0.001) | (0.001) | (0.001) |
| Urban Rate in 1914 | 0.122*** | 0.086** | 0.086*** |
| | (0.033) | (0.040) | (0.021) |
| Geographic Controls | no | yes | yes |
| Province Fixed Effects | yes | yes | yes |
| Cluster SE at year of conquest | no | no | yes |
| Observations | 136 | 136 | 136 |
| Adjusted R-squared | 0.768 | 0.805 | 0.805 |

*Note:* Ordinary least squares regressions with robust standard errors in parentheses. Dependent variable in all columns is the Share of European Population in 1914. In column 1 includes all the control variables except for the geographical controls. In column 2 all control variables are included and in column 3 standard errors are clustered at the year of incorporation. *** p<0.01, ** p<0.05, * p<0.1.

Table 4: IV Results

| Dependent Variable: | log per capita GDP, 1994 | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| **European population / total population** | **5.564*** (1.451)** | **5.493*** (1.514)** | **5.493*** (0.688)** |
| Distance to BA City | -0.085 | 0.000 | 0.000 |
| | (0.149) | (0.162) | (0.050) |
| Land Quality | | 0.004 | 0.004 |
| | | (0.004) | (0.002) |
| Railroad Density | 0.067*** | 0.047* | 0.047** |
| | (0.025) | (0.028) | (0.014) |
| Percent of Land used for Agriculture | 0.419 | 0.291 | 0.291 |
| in 1914 | (0.395) | (0.360) | (0.309) |
| Population Density in 1914 | -0.038*** | -0.031*** | -0.031*** |
| | (0.009) | (0.011) | (0.005) |
| Urban Rate in 1914 | 0.541 | 0.437 | 0.437 |
| | (0.344) | (0.330) | (0.374) |
| Geographic Controls | no | yes | yes |
| Province Fixed Effects | yes | yes | yes |
| Cluster SE at year of conquest | no | no | yes |
| Observations | 136 | 136 | 136 |
| Adjusted R-squared | 0.553 | 0.583 | 0.432 |

*Note:* Instrumental Variable regressions with robust standard errors in parentheses. Dependent variable in all columns is log per-capita GDP in 1994. In column 1 includes all the control variables except for the geographical controls. In column 2 all control variables are included and in column 3 standard errors are clustered at the year of incorporation. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: IV Results

| Dependent Variable: | share of population with higher education, 2001 | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| **European population / total population** | **0.074*** | **0.089**** | **0.089**** |
| | **(0.044)** | **(0.041)** | **(0.037)** |
| Distance to BA City | -0.004 | -0.006* | -0.006* |
| | (0.004) | (0.003) | (0.003) |
| Land Quality | | -0.000** | -0.000** |
| | | (0.000) | (0.000) |
| Railroad Density | 0.002** | 0.002*** | 0.002*** |
| | (0.001) | (0.001) | (0.001) |
| Percent of Land used for Agriculture | -0.020 | -0.012 | -0.012 |
| in 1914 | (0.013) | (0.011) | (0.010) |
| Population Density in 1914 | 0.000 | -0.000 | -0.000 |
| | (0.000) | (0.000) | (0.001) |
| Urban Rate in 1914 | 0.026** | 0.034*** | 0.034** |
| | (0.011) | (0.009) | (0.012) |
| Geographic Controls | no | yes | yes |
| Province Fixed Effects | yes | yes | yes |
| Cluster SE at year of conquest | no | no | yes |
| Observations | 136 | 136 | 136 |
| Adjusted R-squared | 0.295 | 0.472 | 0.224 |

*Note:* Instrumental Variable regressions with robust standard errors in parentheses. Dependent variable in all columns is the share of population age 25 and above with higher education in 2001. In column 1 includes all the control variables except for the geographical controls. In column 2 all control variables are included and in column 3 standard errors are clustered at the year of incorporation. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: IV Results

| Dependent Variable: | share of population with high skilled occupations, 2001 | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| **European population / total population** | **0.174***** | **0.184***** | **0.184** |
| | **(0.067)** | **(0.066)** | **(0.105)** |
| Distance to BA City | 0.006 | 0.002 | 0.002 |
| | (0.006) | (0.006) | (0.007) |
| Land Quality | | -0.000 | -0.000 |
| | | (0.000) | (0.000) |
| Railroad Density | 0.003*** | 0.004*** | 0.004*** |
| | (0.001) | (0.001) | (0.001) |
| Percent of Land used for Agriculture | 0.035* | 0.036* | 0.036 |
| in 1914 | (0.021) | (0.019) | (0.040) |
| Population Density in 1914 | -0.002*** | -0.002*** | -0.002** |
| | (0.001) | (0.001) | (0.001) |
| Urban Rate in 1914 | 0.065*** | 0.064*** | 0.064*** |
| | (0.018) | (0.015) | (0.016) |
| Geographic Controls | no | yes | yes |
| Province Fixed Effects | yes | yes | yes |
| Cluster SE at year of conquest | no | no | yes |
| Observations | 136 | 136 | 136 |
| Adjusted R-squared | 0.675 | 0.738 | 0.484 |

*Note:* Instrumental Variable regressions with robust standard errors in parentheses. Dependent variable in all columns is the share workers in high-skilled occupation in 2001. In column 1 includes all the control variables except for the geographical controls. In column 2 all control variables are included and in column 3 standard errors are clustered at the year of incorporation. *** 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 |

Table 8: IV Results

| Dependent Variable: | log value of industrial production | skilled workers per-1000 individuals | factories per-1000 individuals | energy in h.p. per person |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| **European population / total population** | **6.885*** (2.498)** | **16.025*** (6.091)** | **20.381*** (5.527)** | **0.817** (0.323)** |
| Distance to BA City | -0.235 | -0.847 | -0.725 | 0.034 |
| | (0.299) | (0.714) | (0.547) | (0.030) |
| Land Quality | -0.022*** | -0.056*** | -0.001 | -0.003*** |
| | (0.008) | (0.019) | (0.015) | (0.001) |
| Railroad Density | 0.000 | -0.189* | 0.012 | 0.009 |
| | (0.053) | (0.112) | (0.076) | (0.007) |
| Percent of Land used for Agriculture in 1914 | -0.870 | -2.956** | -4.402*** | -0.187* |
| | (0.672) | (1.326) | (1.548) | (0.110) |
| Population Density in 1914 | 0.035* | 0.217*** | 0.019 | 0.007*** |
| | (0.020) | (0.047) | (0.046) | (0.003) |
| Urban Rate in 1914 | 0.215 | -0.914 | 1.653 | 0.006 |
| | (0.790) | (1.238) | (1.057) | (0.070) |
| Geographic Controls | yes | yes | yes | yes |
| Province Fixed Effects | yes | yes | yes | yes |
| Observations | 136 | 136 | 136 | 136 |
| Adjusted R-squared | 0.190 | 0.243 | 0.344 | 0.084 |

*Note:* Instrumental Variable regressions with robust standard errors in parentheses. Dependent variables in columns 1-4 are the value of industrial production, the number of skilled workers per 1000 individuals, the number of factories per 1000 individuals and the energy in h.p. per person. Each column includes all the control variables. *** p<0.01, ** p<0.05, * p<0.1.

Table 9: Literacy Rates by Contry of Birth

| Nationality | Literacy rate |
|---|---|
| Argentina | 63.2% |
| Average European | 64.2% |
| Average Population | 63.3% |
| Austria | 69.2% |
| France | 79.3% |
| Germany | 88.2% |
| Great Britain | 90.9% |
| Italy | 59.6% |
| Spain | 67.4% |
| Switzerland | 86.9% |

Table 10: IV Results

| Dependent Variable: | share of literate population | Public Schools x 1000 school-age population | Private Schools x 1000 school-age population |
|---|---|---|---|
| | (1) | (2) | (3) |
| **European population / total population** | **0.070**** | **-12.817***** | **2.430** |
| | **(0.035)** | **(4.536)** | **(1.850)** |
| Distance to BA City | -0.011*** | -1.522*** | -0.119 |
| | (0.003) | (0.436) | (0.169) |
| Land Quality | -0.000* | -0.004 | -0.003 |
| | (0.000) | (0.012) | (0.004) |
| Railroad Density | 0.000 | 0.032 | 0.019 |
| | (0.001) | (0.087) | (0.033) |
| Percent of Land used for Agriculture | -0.021** | -1.142 | 0.185 |
| in 1914 | (0.009) | (1.268) | (0.494) |
| Population Density in 1914 | -0.001*** | -0.015 | -0.006 |
| | (0.000) | (0.031) | (0.014) |
| Urban Rate in 1914 | 0.001 | -0.783 | 0.372 |
| | (0.007) | (1.232) | (0.364) |
| Geographic Controls | yes | yes | yes |
| Province Fixed Effects | yes | yes | yes |
| Observations | 136 | 136 | 136 |
| Adjusted R-squared | 0.945 | 0.490 | 0.226 |

*Note:* Instrumental Variable regressions with robust standard errors in parentheses. Dependent variables in columns 1-3 are the share of literate population in 1914, the number of public schools per 1000 school-age population and the number of private schools per 1000 school-age population. Each column includes all the control variables. *** p<0.01, ** p<0.05, * p<0.1.

Table 11: Robustness Checks

| Dependent Variable: | log per capita GDP, 1994 | | | | | |
|---|---|---|---|---|---|---|
| Assumptions: | If initial Arg.>0, Arg_0=6300 | Arg_0=3600 for all counties | moving rate =6% | fertility rate =10% | mortality rate =6% | assumptions (3), (4) and (5) |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **European population / total population** | **7.249*** (1.826)** | **5.387** (2.501)** | **4.810*** (1.774)** | **7.025*** (1.663)** | **5.492*** (1.666)** | **5.300*** (1.571)** |
| Distance to BA City | -0.087 | 0.006 | 0.034 | -0.076 | 0.000 | 0.010 |
| | (0.172) | (0.192) | (0.173) | (0.152) | (0.169) | (0.164) |
| Land Quality | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 | 0.004 |
| | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) |
| Railroad Density | 0.040 | 0.047* | 0.049* | 0.041 | 0.047* | 0.047* |
| | (0.029) | (0.027) | (0.028) | (0.028) | (0.028) | (0.028) |
| Percent of Land used for Agriculture in 1914 | -0.101 | 0.315 | 0.444 | -0.051 | 0.291 | 0.335 |
| | (0.430) | (0.585) | (0.392) | (0.452) | (0.372) | (0.366) |
| Population Density in 1914 | -0.034*** | -0.031** | -0.030*** | -0.034*** | -0.031*** | -0.031*** |
| | (0.012) | (0.013) | (0.011) | (0.012) | (0.011) | (0.011) |
| Urban Rate in 1914 | 0.303 | 0.445 | 0.489 | 0.320 | 0.437 | 0.451 |
| | (0.381) | (0.359) | (0.326) | (0.367) | (0.331) | (0.329) |
| Geographic Controls | yes | yes | yes | yes | yes | yes |
| Province Fixed Effects | yes | yes | yes | yes | yes | yes |
| Observations | 136 | 136 | 136 | 136 | 136 | 136 |
| Adjusted R-squared | 0.538 | 0.585 | 0.592 | 0.546 | 0.583 | 0.586 |

*Note:* Instrumental Variable regressions with robust standard errors in parentheses. Dependent variables in columns 1-6 are different assumptions for the construction of the IV. Each column includes all the control variables. *** p<0.01, ** p<0.05, * p<0.1.

Table 12: Robustness Checks

| Dependent Variable: | log per capita GDP, 1994 | | |
|---|---|---|---|
| Assumptions: | Wheat | Land-gini | Obs. Weighted by population |
| | (1) | (2) | (3) |
| **European population / total population** | **5.451*** (1.533)** | **5.389*** (1.444)** | **4.067*** (1.389)** |
| Percent of Land used for Wheat | 0.136 (0.516) | | |
| Land-gini in 1914 | | -1.058* (0.590) | |
| Distance to BA City | -0.004 (0.164) | 0.059 (0.158) | 0.036 (0.141) |
| Land Quality | 0.004 (0.004) | 0.003 (0.004) | 0.004 (0.003) |
| Railroad Density | 0.045 (0.028) | 0.042 (0.029) | 0.037 (0.025) |
| Percent of Land used for Agriculture in 1914 | 0.257 (0.349) | -0.051 (0.452) | 0.379 (0.275) |
| Population Density in 1914 | -0.030*** (0.011) | -0.027** (0.012) | -0.026*** (0.010) |
| Urban Rate in 1914 | 0.428 (0.325) | 0.490 (0.325) | 0.407 (0.318) |
| Geographic Controls | yes | yes | yes |
| Province Fixed Effects | yes | yes | yes |
| Observations | 136 | 136 | 136 |
| Adjusted R-squared | 0.580 | 0.592 | 0.582 |

*Note:* Instrumental Variable regressions with robust standard errors in parentheses. Dependent variables in columns 1-3 is log per-capita GDP in 1914. In column 1 the percent of land used for wheat production is included as a regressor. In column 2 the land gini is included as a regressor. In column 3 observations are weighted by the population. Each column includes all the control variables. *** p<0.01, ** p<0.05, * p<0.1.