THE ECONOMIC JOURNAL

The Economic Journal, **128** (*September*), 2321–2352. Doi: 10.1111/ecoj.12505 © 2017 Royal Economic Society. Published by John Wiley & Sons, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

MIGRATION, POPULATION COMPOSITION AND LONG RUN ECONOMIC DEVELOPMENT: EVIDENCE FROM SETTLEMENTS IN THE PAMPAS*

Federico Droller

This article analyses the impact of population composition on long run economic development, by studying European migration to Argentina during the Age of Mass Migration (1850–1914). I use an instrumental variables (IV) approach that assigns immigrants to counties by interacting two sources of variation: the availability of land for settlement and the arrival of Europeans over time. Counties with historically higher shares of European population in 1914 have higher *per capita* GDP 80 years later. I show that this long run effect is linked to the higher level of human capital that immigrants brought to Argentina. I show that Europeans raised literacy rates in the receiving counties, and that high-skilled Europeans played an important role in the onset of industrialisation, owned most of the industrial establishments, and provided the majority of the industrial labour force.

Between the late nineteenth century and the First World War, over 40 million Europeans migrated to the New World (Hatton and Williamson, 1992). The United States, Argentina and Canada were among the main destinations. Immigrants significantly contributed to the labour force in the receiving countries and affected political and economic life there (Hatton and Williamson, 1992; Abramitzky *et al.*, 2014; Abramitzky and Boustan, 2017). The arrival of more than five million Europeans to Argentina during the Age of Mass Migration (1850–914) dramatically changed the composition of the local population. Over this period of time, Argentina saw its population grow from 1.9 to 7.9 million and its GDP increase tenfold (Ferreres, 2005).

Over the course of history, colonialism and large population movements have changed the composition of the world's population (Putterman and Weil, 2010). These episodes also significantly affected both the short run and long-run development of different countries. A vast literature on the deep determinants of economic growth provides evidence for enduring political and economic institutions from the colonial period. Engerman and Sokoloff (1997, 2002), La Porta *et al.* (1998, 1999) and Acemoglu *et al.* (2001, 2002) show, for example, how institutions influenced economic development, a point more recently stressed by Banerjee and Iyer (2005) and Bruhn and Gallego (2012). Using a different approach, Glaeser *et al.* (2004) argue in favour of human capital as an explanation for economic growth and institutional

* Corresponding author: Federico Droller, Department of Economics, FAE – Universidad de Santiago de Chile, Av. Lib. B. O'Higgins 3363, Estación Central, 9170022. Santiago, Chile. Email: federi.codroller@usach.cl.

I am grateful to the Editor, Hans-Joachim Voth, and three anonymous referees for helpful observations. I thank 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. Financial support from the Universidad de Santiago DICYT 031462D is gratefully acknowledged.

improvement. Acemoglu *et al.* (2014) add to the debate, reconsidering whether human capital or institutions are fundamental determinants of economic growth, and providing evidence in favour of the latter. Easterly and Levine (2016) also argue for the place that human capital holds in explanations for divergent paths of economic growth.

In this article, I show the effect of historical population composition on long run development, using the fertile plains of Argentina as an empirical setting. These fertile plains, also known as the Pampas, represent an excellent setting to study the effect of migration on economic development for several reasons. First, the area experienced a significant inflow of migrants relative to the local population over a short period of time, from 1869 to 1914 (Taylor, 1995). Second, and more importantly, counties across the fertile plains share common national institutions and homogeneous geographic endowments. This makes it possible not only to assess the effect of change in the composition of the population, but also to rule out alternative explanations.[1]

The empirical analysis is based on instrumental variables (IV) approach to overcome the problem of the endogenous sorting of migrants. 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. The model uses the actual inflow of Europeans each year and distributes them randomly over the land available for settlement. The availability of land to potential settlers varies over time, depending on levels of civil and international conflict and on the success of subsequent military campaigns that aimed to conquer the plains.[2] The identification assumption is that both the arrival of Europeans and the availability of land for settlement are exogenous to the later development of any one county other than through the IV. As described in detail in subsection 2.2, the timing of land of incorporation in the fertile plains was carried out for reasons unrelated the economic potential of the land. In particular, conflict with indigenous tribes living on the 'unoccupied' territory prompted the conquest, while the availability of military resources dictated when military campaigns were organised.

The model also incorporates changes in both populations, Europeans and Argentinians, due to mortality, fertility and moving rates across counties, in order to construct the synthetic stocks of Europeans and Argentinians. I then compute a synthetic share of the European-born population as of 1914 and use it as an IV. This approach uses only the part of the variation in the share of Europeans across counties that is exogenously generated by the model.

Results show a significant and positive long run effect of the share of the European population on economic development. In this respect, for an average county, with a 20% share of Europeans in 1914, increasing the share to 25% would raise *per capita* GDP from US $6,754 to US $9,278. I conduct a series of robustness checks to show that results are stable across different specifications in the parameters of the model and that

[1] Throughout the article, I use the terms fertile plains and Pampas interchangeably.

[2] The process of settling the Pampas differed notably from that which occurred in the US. While in Argentina settlers arrived after the government conquered the land, in the US., colonisers preceded the military.

they are not driven by distance to the capital city, importance of agricultural activities, land distribution, or the date of conquest of a county.

In the second part of the article, I analyse the characteristics of arriving Europeans to shed light on the mechanism linking the composition of the population and long run economic development, and find that European migrants had a higher human capital profile than the native Argentinians. I start by showing that Europeans arriving in Argentina were on average more literate than locals and that there were also major differences in literacy rates among Europeans themselves, depending on their country of origin. County-level average literacy rates changed substantially following the arrival of immigrants; not only was there a direct effect (of migrant inflow on average literacy) but an indirect effect as well, since European arrival was also related to the creation of private schools.

Next, I use a second measure of human capital: skills. Mokyr (2005), and more recently Squicciarini and Voigtländer (2015), have shown that distinguishing between average human capital and upper-tail knowledge is crucial for identifying the importance of human capital for economic development. Hornung (2014, p. 85) points to the relation between skilled-worker immigration and productivity, claiming that skilled immigrants 'raise the overall level and composition of skills and thus might foster growth and development of receiving countries'. In Argentina, immigrants provided valuable human capital complementary to industrial activities (Rocchi, 2006). Using census data on occupations, I divide Europeans into high-skilled (professional) and low-skilled groups and show that over the long run, high-skilled Europeans have a significantly higher effect than low-skilled Europeans on economic development. To understand the mechanism better, I also analyse the short run effects of skilled immigrants, focusing on their effect on industrialisation. The industrial sector around the early twentieth century was largely run by immigrants: a vast majority of industrial establishments were owned by Europeans and industrial workers were also mostly European. County-level information from the 1935 Industrial Census supports the aggregate data: high-skilled Europeans contributed far more to the onset of industrialisation than low-skilled Europeans. An analysis by country of origin further supports this argument. The groups that most contributed were from Great Britain, Germany, Austria, Switzerland and France, as compared to migrants from Italy and Spain. English, German and French migrants were more literate and arrived from countries with higher levels of industrial development compared to their Italian and Spanish counterparts (Bairoch, 1982). Further, cross-county differences in the level of industrialisation in 1935 are tightly linked to current economic development.

This article sheds light on one channel through which the European portion of the population shaped long run economic development, human capital. In interpreting my results, I do not argue that this was the only channel. Europeans affected development in many ways besides bringing their knowledge and skills; they provided a labour force, for example. I do, however, argue that immigration to the fertile plains deeply changed the human capital profile of the population, with considerable consequences for the onset of industrialisation. My results contribute to continued debate in the literature on the relevance of human capital for economic growth (Becker and Woessmann, 2009; Galor, 2011; Hornung, 2014; Squicciarini and Voigtländer, 2015) and industrialisation (Becker *et al.*, 2011; Franck and Galor,

2015). This article focuses its analysis on the fertile plains, allowing a look within a region that holds institutions and geographic endowments constant, and showing new evidence for the importance of human capital for long run development. In this contained setting, human capital can be identified and its positive long run effects estimated. This article is also related to a broader literature examining the deep determinants of economic development (Nunn, 2009; Spolaore and Wacziarg, 2013) as well as recent work analysing the effect of migration on the receiving economy (Abramitzky *et al.*, 2012 2014; Abramitzky and Boustan, 2017).

This article is structured as follows. Section 1 provides the historical background of land conquest in the fertile plains, and describes migrant inflow to Argentina during the Age of Mass Migration. Section 2 describes the data, while subsections 3.1 and 3.2 present the empirical strategy, the construction of the IV and the main results. Subsection 3.3 provides evidence on the role of human capital in the onset of the industrialisation process and provides an analysis by migrant country of origin. Section 4 concludes.

1. Background and Institutional Setting

1.1. *The Expansion of the Frontier*

Argentina gained independence from the Spanish Empire in 1816 but it was not until the end of the century that the government gained political power over the entire (current) territory. Delimitation of the national territory and its borders was a main concern for the Argentinian government. Most of the fertile plains were, in fact, already settled by indigenous tribes, and conflict was frequent. The government would organise military excursions to dispute land, as the area was neither recognised as an independent state nor were the indigenous tribes seen as legal owners (Cortés Conde, 1968; Lacoste, 2002; Mandrini, 2008). It was not, however, until the end of the civil war in 1862, and the creation of a unified national government, that systematic plans to conquer the rest of the territory were developed. The final military campaigns started in 1870 and continued through 1885. Walther (1964) describes in detail these military operations and their effect on the expansion of the Argentinian frontier over time.[3]

The order in which different territories were conquered is strongly linked to military as opposed to economic concerns, particularly distance between the main cities and fortresses, where armies were organised, and indigenous settlements. Moreover, it was not the population or citizens themselves who moved the frontier, nor a need for new land to populate that motivated the military campaigns (Cortés Conde, 1968). Figure 1 shows a map of the fertile plains where each county is coloured according to the year in which it was conquered. From here, two facts are easy to identify. First, navigable rivers to the north of Buenos Aires were known since colonial times, and thus explain early settlements in this area. Second, the old colonial road to Peru, running northeast from Buenos Aires, explains settlements along this line, together with the city of Cordoba,

[3] Campaigns took place in 1779, 1823, 1826, 1828, 1852, 1864 and 1876. Figures A1–A2 in the online Appendix depict maps showing changes in the frontier for selected years from 1779 to 1876 (Walther, 1964). In particular, these show that there were both gains and losses over this time period. For example, while in 1828 almost two-thirds of the Buenos Aires Province was conquered, most of the territory was lost again by 1860, and it was only in 1880 that full control was regained (Cortés Conde, 1968).

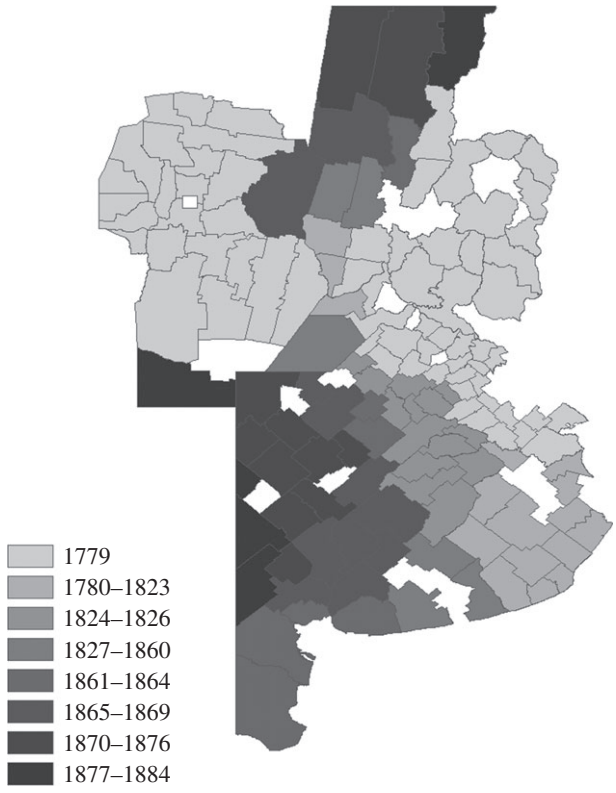

Fig. 1. *Counties By Year of Conquest*

Notes. The map shows the four provinces under study, which contain the fertile plains. Each county is marked with a different colour, according to the year of conquest. White corresponds to counties conquered in 1779 or earlier, different shades of grey to counties conquered between 1860 and 1875 and black corresponds to counties conquered in 1885. See main text for data sources and details.

founded in 1573. Thus, conquest expanded from the main cities and fortresses, in part due to the lack of navigable rivers in the fertile plains. Conquest of the Pampas was finally possible between 1870 and 1885.

1.2. *The Age of Mass Migration*: *The Arrival of Europeans*

The end of the civil war and the unification of all provinces under a national government opened the country to European migration. The flow of arrivals in Argentina resembles those to the United States, Canada and Australia, although the United States also experienced a large migration from northern European countries before 1870. Argentina was among the largest recipients of European migrants, receiving more than five million Europeans between 1857 and 1914. The country quadrupled in population, from about 1.9 million in 1869 to 7.9 million in 1914.[4]

[4] The migration pattern to Argentina resembles that of the US; the correlation between both is 0.795. Data from http://hsus.cambridge.org.

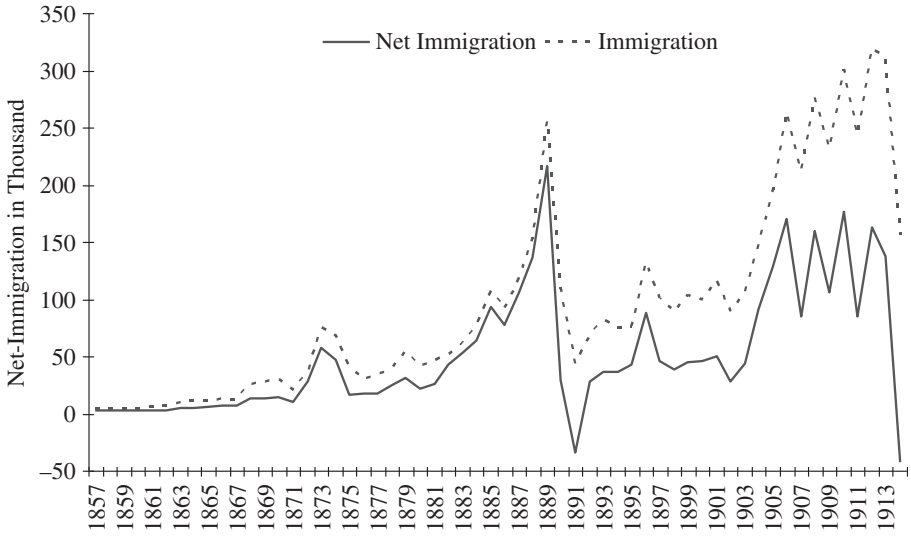

Fig. 2. *Immigration Time Series, 1857–914*

Notes. Time series of immigration and net-immigration to Argentina, 1857–914. See main text for data sources and details.

Foreign-born residents represented 7.6% of the population in 1869, rising to 30% in 1914. Figure 2 shows the time series of immigration and net immigration of Europeans to Argentina. In particular, it shows that the arrival of immigrants is not a monotonic function of time, with a peak around 1889 and a steady, though not constant, increase after 1900. The majority of immigrants came from Italy, Spain, and France, but also from Great Britain, Germany, Austria and Switzerland. Immigrants were mostly working-age males: 60% were between the ages of 20 and 50, while only 30% of all immigrants were female.

2. Data and Descriptive Statistics

This study combines data on current economic development with historical data on migration and economic and sociodemographic characteristics of counties. The data cover 136 counties in four provinces in the fertile plains: Buenos Aires, Santa Fe, Cordoba and Entre Rios. Online Appendix A provides detailed descriptions of all variables and sources. In this Section, I briefly describe the main variables used in my analyses.

2.1. *Measures of Current Economic Development*

The main outcome variable is the current level of economic development, measured as *per capita* gross domestic product (GDP). Other outcome variables include the share of adults with post-secondary education and the share of skilled workers. County-level *per capita* GDP is constructed by combining gross product data from the 1994 National Economic Census (NEC) with agricultural output estimates from Ministry of

Table 1

Summary Statistics

Variable	Mean	SD	50th Percentile	Number of observations
Share of the European-born population, 1914	0.23	0.11	0.24	136
Share of the European-born population, 1895	0.36	0.20	0.39	126
GDP *per capita*, 1994	6,754	4,190	6,418	136
Ln GDP *per capita*, 1994	8.59	0.78	8.77	136
Ln industrial output *per capita*, 1994	5.40	1.45	5.52	136
Share of pop. with higher education, 2001	0.10	0.02	0.10	136
Share of high-skilled workers, 2001	0.25	0.04	0.25	136
Ln industrial output *per capita*, 1935	4.40	1.14	4.38	136
Skilled workers per 1,000 individuals, 1935	1.99	2.06	1.40	136
Number of industrial workers in the population, 1935	0.02	0.02	0.01	136
Number of factories per 1,000 individuals, 1935	3.69	2.16	3.31	136
Energy in H.P. *per capita*, 1935	0.10	0.14	0.07	136
Literacy rate, 1895	0.63	0.05	0.65	136
Literacy rate, 1914	0.55	0.11	0.56	126
Number of public schools per 1,000 school-aged pop. 1914	5.33	2.32	4.78	136
Number of private schools per 1,000 school-aged pop. 1914	0.85	0.71	0.74	136
Distance to the city of Buenos Aires	5.71	0.65	5.84	136
Land quality index	45.43	16.39	45.10	136
Railroad density in 5 kilometres	4.71	2.77	4.14	136
Percent of land used for agriculture, 1914	0.28	0.23	0.26	136
Population density, 1914	6.67	5.53	5.22	136
Urban rate, 1914	0.33	0.18	0.32	136

Source. Share of European population, total population, urban population, land used for agriculture, number of schools and literacy from 1895 and 1914 General Censuses of Argentina. Industrial variables taken from 1935 Industrial census. 1994 GDP constructed from the National Economic Census (see Appendix for construction of this variable). Education and Occupations in 2001 from 2001 National Census. Land quality from INTA (see online Appendix A for source and construction of this variable). Distance to the City of Buenos Aires and railroad density constructed from GIS data (see online Appendix A for details).

Agriculture reports.[5] A more detailed description of the construction of this variable can be found in the online Appendix. Data on higher education rates and the share of skilled workers comes from the 2001 Population Census. Descriptive statistics in Table 1 show that average *per capita* GDP in the sample is slightly above US $6,700, whereas for the bottom 25% of counties it is under $3,560 and for the top 25% it is over $9,000. On average, 10.4% of the population aged 25 and over has completed more than 12 years of education (completed secondary school and started or finished a post-secondary education or university degree). Of those individuals reporting an occupation in 2001, 18% work in high-skilled jobs.[6]

2.2. *Historical Data and Immigration*

Historical information on population comes from the 1869, 1895 and 1914 Population Censuses. These censuses records contain county-level information on demographics

[5] The Argentinian Statistical Office computes *per capita* GDP at the national and province level, but not at the county level.

[6] See the online Appendix for a definition of high-skilled jobs.

and economic activity. Information from the 1914 census on the number of individuals by country of birth, at the county level, is used to construct the main explanatory variable: the share of European-born population, which is computed as the number of European-born individuals divided by the total population, in 1914. Individual-level information on nationality, age, sex, occupation and literacy are obtained from 1895 census microdata samples from Somoza and Lattes (1967), and is used to analyse skills and occupations among the population. Descriptive statistics in Table 1 show that Europeans represent a sizeable fraction of the population: the average share of European population is 23% (the standard deviation is 11%). In addition, there is substantial variation across counties in the share of Europeans, ranging from under 1% to 48% of the 1914 population who were born in Europe.

Based on newly digitised historical records from the Argentinian Migration Office, I build a data set containing the number of immigrants arriving each year from 1857 to 1914, and their countries of origin (Dirección General de Inmigración, 1925).

Further, based on Walther's (1964) work on military campaigns and Cacopardo (1967), Cortés Conde, 1968, Gallo (1983) and Tell (2008), I construct the internal frontier of the territory under the political power of the Argentinian government for each year in the sample, using the latter to define whether or not land was available for settlement between 1857 and 1885.[7] By 1885, the Argentinian government controlled all of the fertile plains. Given that military campaigns were carried out in different years, I assume the frontier remains constant until the next military campaign. As the frontier expanded (or contracted), new counties opened (or closed) for settlement. Finally, I overlap county boundaries on these maps and establish the date on which each county was considered to be part of Argentinian territory.[8] The conquest of the plains can be summarised in eight waves of land seizures, from 1779 to 1885. Sixty-six counties already existed at the time of independence from Spanish colonial rule, 16 were incorporated in 1823, 14 in 1826, six in 1860, seven in 1864, 11 in 1869, 11 in 1875, and five in 1885; Figure 1 summarises this information, where each county has a colour that reflects the year in which it was incorporated.

The 1914 Education Census provides county-level information on the number of primary schools and whether they were public or privately funded. Finally, in the second part of the article, I use data from the 1935 Industrial Census, which provides county-level measures of industrial development, including the value of industrial production, the number of skilled workers, the number of factories, and energy used, so as to shed light on the mechanism linking the share of Europeans and economic growth.

2.3. *Other Control Variables*

I construct a county-level land quality index that is a weighted average of soil quality in about 1 kilometre cells within current county boundaries. Soil quality comes from Cruzate *et al.* (1990), who construct a geo-referenced data set based on soil

[7] I supplement this information with maps digitised by the National Library of Argentina (http://trapalanda.bn.gov.ar/).

[8] The date a county was incorporated should not be confused with the date in which the county's foundation actually occurred, which was usually years later.

characteristics (including a quality index). This index refers to the geological conditions of the soil, such as ground composition and rain, and not to the technologies used for cultivation. Other geographical control variables include average rainfall and temperature (Worldclim), elevation (from the National Oceanic and Atmospheric Administration, and the US National Geophysical Data Center), and ruggedness of the terrain from Nunn and Puga (2012). In all cases, I combine cell-level information with county boundaries to construct county-level averages. Other control variables included in the empirical analyses are the share of land used in agriculture, population density, and the urbanisation rate, all constructed from the 1914 Census. See the Appendix for further details.

The availability of railroads in a given county is computed as the average railroad density within a 5 kilometre radius in each county, based on data from Cruzate *et al.* (1990). The (log) distance to the capital city of Buenos Aires is also included as a control variable.

Counties in the fertile plains provide an extremely useful setting in which to understand the effect of migration on long run growth. Similar geographic endowments, common institutions, and a varying number of European immigrants, allows to isolate the effect of migration and population composition from other competing explanations. Migration to Argentina over this period of time dramatically altered the composition of the population, and had a long-term impact on economic development.

3. The European Origin of Development

3.1. *Association*

In this Section, I present evidence on the association between the share of Europeans in the late nineteenth century and the level of economic development today. Figure 3 shows

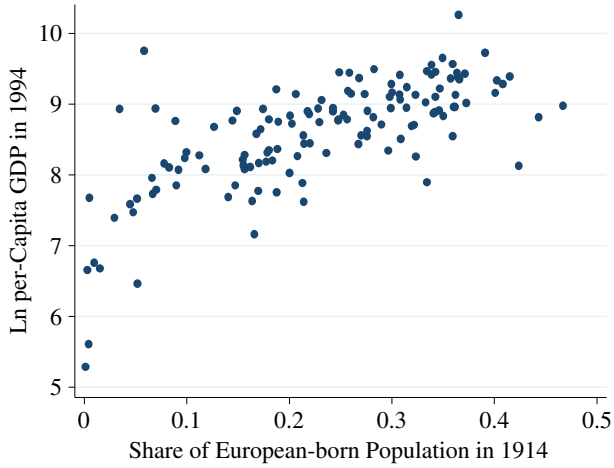

Fig. 3. *Correlation Between Ln Per Capita GDP in 1994 and the Share of the European-born Population in 1914*

Notes. Scatter plot between ln *per capita* GDP in 1994 and the share of European-born population in 1914. The unit of observation is the county. See main text for data sources and details. Colour figure can be viewed at www.wileyonlinelibrary.com.

Table 2

Share of European Population and Per Capita GDP

	OLS (1)	OLS (2)	OLS (3)	OLS (4)	OLS (5)
Dependent variable:			Ln *per capita* GDP 1994		
Share of the European-born population in 1914	4.862*** (0.610)	5.668*** (0.632)	4.823*** (0.744)	4.560*** (0.852)	3.914*** (0.796)
Distance to Buenos Aires city				0.098 (0.163)	0.079 (0.151)
Land quality			0.008** (0.004)	0.005 (0.004)	0.004 (0.004)
Railroad density				0.040 (0.028)	0.052* (0.029)
Percent of land used for agriculture in 1914					0.644** (0.293)
Population density in 1914					−0.028** (0.011)
Urban rate in 1914					0.557 (0.341)
Constant	7.467*** (0.167)	7.437*** (0.227)	8.864*** (0.997)	8.591*** (1.316)	8.363*** (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^2	0.495	0.522	0.608	0.616	0.638
Observations	136	136	136	136	136

Notes. OLS regressions with robust standard errors corrected for heteroscedasticity (White) 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 urbanisation rate). See main text for data sources and details. ***p < 0.01, **p < 0.05, *p < 0.1.

the correlation between *per capita* GDP in 1994 and the share of the European-born population in 1914 for counties in the fertile plains. This Figure shows that those counties where Europeans accounted for a higher share of the population in 1914, currently have a higher level of economic development.[9] I estimate equations of the form:

$$y_i = \alpha + \beta E_i + X_i\gamma + \eta_p + \epsilon_i, \quad (1)$$

where y_i represents measures for economic development in a county i (log *per capita* GDP in 1994, higher education rates in 2001, and the share of skilled workers in 2001); E_i is the share of European-born population in the county in 1914; and X_i is the set of county-level controls in 1914, including land quality, distance to the city of Buenos Aires, and availability of railroads; and η_p are state fixed effects.[10] Column (1) in Table 2 documents a strong association between the share of Europeans in 1914 and *per capita* GDP today.

[9] See Figure A3 in the online Appendix for maps showing the spatial distribution of these two variables across counties.

[10] 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.

The following columns test the robustness of this association by adding, sequentially, a set of control variables. These include province fixed effects (column (2)) to account for time-invariant province-level differences, geographic controls (column (3)) to account for geographical differences across counties, and socio-economic county-level controls. In particular, these include distance to the capital city of Buenos Aires and railroad density (column (4)), and the share of productive land allocated to agriculture, population density, and urbanisation rate (column (5)). Qualitative results are not affected by the inclusion of these controls. These results suggest that a one standard-deviation change in the historical share of Europeans increases current *per capita* GDP by between 0.56 and 0.66 standard deviations. The preferred specification is defined as the one in column (4) and is the most conservative.[11]

Table 3 presents two other measures of current economic development as outcome variables: a share of the population with higher education in panel (*a*) and share of the population with high-skilled occupations in panel (*b*). Results show that the share of

Table 3

Share of European Population and Development

	OLS (1)	OLS (2)	OLS (3)	OLS (4)	OLS (5)
Panel (*a*): dependent variable: share of population with higher education 2001					
Share of the European-born population in 1914	0.059***	0.060***	0.063***	0.054***	0.037*
	(0.016)	(0.017)	(0.021)	(0.020)	(0.020)
R^2	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 population in 1914	0.236***	0.221***	0.221***	0.204***	0.135***
	(0.025)	(0.026)	(0.030)	(0.030)	(0.036)
R^2	0.527	0.574	0.665	0.693	0.749
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

Notes. OLS regressions with robust standard errors corrected for heteroscedasticity (White) in parentheses. Dependent variable in panel (*a*) is the share of population 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 urbanisation rate). See main text for data sources and details. ***p < 0.01, **p < 0.05, *p < 0.1.

[11] In particular, it does not include control variables for the share of land in agriculture, population density, and urbanisation rate. These variables may be regarded as less than ideal control variables, as they could also be outcome variables themselves. In this respect, adding these right-hand side variables to the regression could potentially bias the result. Results are consequently shown with and without this set of variables.

Europeans is positively associated with these variables and is robust to all of the specification changes described above.

3.2. *Instrumental Variables Approach*

The results above show that differences in current economic development are associated with the pattern of migrant settlement at the beginning of the twentieth century. In an ideal experiment, counties are equal in all respects and are randomly shocked with European population in different intensities. In such a setting, one can analyse long run economic development and observe whether any county-level differences can be explained by the original share of European-born population, the only variable that differs across counties. The empirical setting of this article closely resembles this ideal experiment, since the institutional setting of the counties in the fertile plains is the same and the geographical variation is arguably small. The key difference is, of course, that Europeans were not randomly distributed across counties.

In this Section, I use an IV approach to address the potential endogeneity in the selection of European migrants to different counties. I construct a synthetic measure of the share of immigrants in each county and use it as an IV for the actual share of immigrants in a given county. The IV exploits two sources of variation: first, the expansion of the frontier, which determines the incorporation of land available for settlement; and second, variation in the number of European migrants arriving in Argentina.[12]

I use a simple model of settlement and demographic growth to interact both sources of variation and estimate, for each county, the flow of Europeans and the growth of the Argentinian population from 1857 to 1885.[13] I then use the synthetic stocks of European and Argentinian population to compute a synthetic share of European-born population in each county in 1895.[14]

The identification assumption is that both the arrival of Europeans and the availability of land for settlement are exogenous to the later development of any one county. While it is plausible that the aggregate number of arriving Europeans to the port of Buenos Aires is exogenous to any one county, I discuss in further detail below the second identifying assumption: that the timing of conquest of the fertile plains does not correlate with later development of any one county, other than through the IV.

Two key characteristics of the process of conquest support this assumption. First, military campaigns to conquer the fertile plains took place when no other conflict was taking place. More specifically, a succession of conflicts, from civil wars after independence to international conflicts (e.g. the Triple Frontier war) consumed all military resources, leaving provincial governments (before the unification of the country) and the national

[12] I use net immigration, defined as the number of arriving Europeans minus the number of Europeans leaving each year, to proxy the number of long-term migrants arriving in Argentina.

[13] After 1885, the fertile plains were entirely controlled by the government, and no further variation can be exploited from the incorporation of land for settlement.

[14] Throughout the article, I treat Argentinians as different from newly arrived Europeans. By 1850, Argentinians have both Spanish (from the colonisation period) and indigenous ancestors but almost no other European ancestry.

government unable to expand the inner frontier simultaneously.[15] In general, the years of conquest of the fertile plains did not overlap with the years of civil and international conflicts (Cacopardo, 1967; Perry, 1972). Moreover, the periods of conquest usually followed after the end of other conflicts. The lack of navigable rivers in the unconquered fertile plains limited approaches by foot or horseback, which had to be initiated from the inner frontier-line (Morse, 1962; Cortés Conde, 1968).[16] In this respect, the timing in which each county was conquered was determined by its distance from the inner frontier, controlled by a line of fortresses, and not by distance to the city of Buenos Aires. Military campaigns started from different points along the frontier and proceeded towards the different tribes' settlements (Cacopardo, 1967; Walther, 1967; Cortés Conde, 1968). What is more, changes in leadership shaped the course of policy towards the expansion of the inner frontier. In 1874, the Minister of War, A. Alsina, favoured a defensive strategy, including the digging of a trench and the building of a line of fortresses, so as to avoid indigenous horse raids. After his sudden death in 1877 at the age of 48, J. A. Roca was appointed as Alsina's successor. Roca held different views with respect to how handle the inner frontier, favouring instead an offensive strategy, which would eventually lead to the final conquest of the territory (Perry, 1972).

Second, it should be noted that conquest of the fertile plains did not see the incorporation of counties with higher land quality first, followed by areas of lower land quality. Such a pattern would imply a relationship between land quality and year of conquest, in turn potentially affecting long run development. Because the timing of the conquest precedes the onset of industrialisation (in Argentina), the only relevant characteristic in terms of the economic potential of the fertile plains is that of land quality, determining its potential use for agriculture or ranching. The lack of a correlation between the timing of the conquest with the quality of the land (correlation of 0.015), together with the uniform geographic characteristics of the fertile plains, reinforces the idea that land quality is unrelated to the year in which each county was incorporated.

3.2.1. *A simple model of settlement*

For each year between 1857 and 1885, the actual number of European immigrants arriving at the port of Buenos Aires is distributed across counties in the fertile plains. To make the distribution as random as possible, Europeans are distributed uniformly across counties, i.e. all available counties get the same fraction of immigrants. County size, distance to the city of Buenos Aires, amount of unoccupied land, and ease of access by railroad are not considered in the distribution mechanism. European immigrants did choose to reside in geographically or economically favoured areas; therefore, for the IV to be valid, this exercise considers an immigrant settlement pattern that is independent of these aspects.

Once Europeans are settled within a county, I assume that they never move (in a robustness check I consider the possibility of Europeans moving to another county). Europeans die at a yearly rate δ and reproduce at rate ρ, and children born to

[15] For example, in 1827, the Governor of Buenos Aires requested that the national government push the frontier further out but a campaign could not be organised due to an internal uprising (Cacopardo, 1967).

[16] The line of fortresses changed over the years but by the late 1870s ran northeast from the city of Mendoza to the city of San Luis, then southeast to the city of Trenque Lauquen and south to the cities of Bahia Blanca and Patagones (see Figures A1 and A2 in the online Appendix).

Europeans are considered Europeans.[17] Argentinians are initially present in counties controlled by the central government by 1857 but not those conquered afterwards (the assumption is that new counties were initially 'empty').

If the initial number of Argentinians in a given county is assumed to be exogenous, then the latter is simply calculated from the 1869 census and adjusted by population growth. Since, however, one may worry that the initial population in a county is not a random variable, I also consider the initial stock of Argentinians in the model as the county-level average 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 who move to a new county each year. I distribute the movers and newborns uniformly across all available counties. The mortality rate ($\delta = 2.2$), the fertility rate ($\rho = 5.3$), and the moving rate ($\phi = 1.95$) are computed from the 1869, 1895, and 1914 censuses.[18] Under these assumptions, the synthetic number of Europeans in each county in 1885 is as follows:

$$\text{syntheticstock of Europeans}_i = \sum_{t=1857}^{1885} \frac{1}{N_t}(1-\delta+\rho)^{1885-t} e_t \times \mathbb{1}_i\{t \geq D_i\}. \tag{2}$$

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

$$\begin{aligned}\text{syntheticstock of Argentinians}_i = {} & \hat{A}_{i1857}(1-\delta+\rho-\phi)^{28} + \\ & \sum_{t=1857}^{1885} \frac{1}{N_t}(1-\delta+\rho-\phi)^{1885-t}\phi a_t \times \mathbb{1}_i\{t \geq D_i\},\end{aligned} \tag{3}$$

where synthetic stock of *Europeans*$_i$ and synthetic stock of *Argentinians*$_i$ represent 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. As discussed above, there are two assumptions on $\hat{A}_{i1857}$: $\hat{A}_{i1857} = A_{i1857}$, the actual population, and $\hat{A}_{i1857} = \bar{A}_{1857}$ if $A_{i1857} > 0$ and $\hat{A}_{i1857} = 0$ otherwise. $\mathbb{1}_i\{t \geq D_i\}$ is an indicator of whether county i is part of the Argentinian territory, where D_i is the year in which county i was under the political control of the central government. N_t is the number of counties under political control of the government at time t.

The synthetic share of Europeans is defined as follows:

$$\text{synth E}_i = \frac{\text{synth stock of } Eu_i}{\text{synth stock of } Eu_i + \text{synth stock of } Arg_i}, \tag{4}$$

and is used as an IV for the actual share of the European population. Variation in the synth stock of Eu_i and the synth stock of Arg_i will induce variation in the constructed share. The synth stock of Eu_i varies across counties depending on the year in which

[17] Although any child born in Argentina to foreign parents is considered Argentinian by law, I treat this population as European since they were raised by European parents, absorbing their culture, learning their skills, etc.

[18] See the online Appendix for a detailed explanation of how these values were computed. The first stage and analyses in the Section below are robust to changes in the parameters of the model, as well as to changes in the assumption on the initial level of the Argentinian population. These possibilities are considered as robustness checks in Table 6.

county i came under political control of the Argentine government (D_i) and also on the number of immigrants that arrive at time t (e_t). Variation in the synth stock of Arg_i also depends on D_i, and the number of Argentinians in 1857.[19]

3.2.2. *Results*

I use the synthetic share of Europeans as an instrument for the actual share of Europeans in a county. Columns (1) and (3) in Table 4 report the first-stage estimation, showing that the synthetic share is a strong predictor of the actual share. Weak identification is ruled out by the Kleibergen-Paap test.[20]

The second stage is presented in columns (2) and (4), where geographic variables and controls for distance to the city of Buenos Aires and railroad availability are included, and socio-economic controls are further added in columns (6) and (8). The IV estimates from column (4) indicates 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, implying that for an average county, with a 20% share of Europeans, increasing the share to 25% would raise *per capita* GDP from US $6,754 to $9,278. This shows a significant and positive long run effect of the share of European population on *per capita* GDP.[21] The IV estimates are higher than the OLS ones, suggesting negative bias in the selection of Europeans into these counties.[22]

Estimates for the other two measures of economic development show the same result: a positive and significant effect of the composition of the population on the share of the population with higher education (Table 5, panel (*a*)) and on the share of workers with high-skilled occupations (Table 5, panel (*b*)).[23]

Although context specific, these results are closely related to Easterly and Levine's (2016) similar finding that the European share of the population during colonisation is strongly related to economic development today. They estimate the effect of the share of Europeans on current income to be around 3.5. In this respect, my estimates are in line with those found in other studies.

3.2.3. *Robustness*

I assess the robustness of the main findings, presenting variations in the assumptions of the demographic model, as well as to alternative specifications in the estimation. The first variation in the model of settlement and demographic growth allows Europeans to move as well: first, I let them move at the same rate as Argentinians (Table 6, panel (*a*), columns (1) and (2)) and, as a second robustness check, I double the moving rate of Europeans (Table 6, panel (*a*), columns (3) and (4)). In both cases, results do not vary qualitatively.

[19] See Figures A2 and A3 in the online Appendix for the spatial distribution of the synthetic share of Europeans across counties in the fertile plains.

[20] Figure A5 in the online Appendix shows the correlation between the share of European population and the synthetic share, and Figure A6 repeats this graph when control variables and fixed effects are included. These Figures show a strong association between both variables across specifications.

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

[22] This difference may also arise if there is measurement error in the share of European population.

[23] For example, estimates from column (1) indicate that an increase of one standard deviation (SD) in the share of European population in 1914 raises the share of population with higher education by 0.32 SD and the share of high-skilled workers by 0.62 SD.

Table 4

Share of European Population and Per Capita *GDP*

	Specification 1				Specification 2			
	IV 1		IV 2		IV 1		IV 2	
	1 Stage (1)	2 Stage (2)	1 Stage (3)	2 Stage (4)	1 Stage (5)	2 Stage (6)	1 Stage (7)	2 Stage (8)
Dependent variable:	Ln *per capita* GDP 1994							
Share of the European-born population in 1914		6.350*** (1.474)		8.951*** (2.139)		5.778*** (1.702)		9.032*** (2.632)
Synthetic share of the European-born population in 1914	0.281*** (0.058)		0.278*** (0.067)		0.251*** (0.042)		0.218*** (0.061)	
Distance to Buenos Aires city	0.012 (0.020)	0.023 (0.155)	0.003 (0.022)	−0.086 (0.161)	0.024* (0.012)	−0.014 (0.168)	0.020 (0.014)	−0.175 (0.184)
Land quality	0.002*** (0.000)	0.002 (0.005)	0.002*** (0.001)	−0.003 (0.006)	0.000 (0.000)	0.004 (0.004)	0.001 (0.000)	0.003 (0.004)
Railroad density	0.008** (0.004)	0.023 (0.029)	0.006 (0.004)	−0.002 (0.035)	0.003 (0.003)	0.046 (0.028)	0.001 (0.003)	0.034 (0.029)
Percent of land used for agriculture in 1914					0.190*** (0.024)	0.228 (0.389)	0.200*** (0.028)	−0.500 (0.641)
Population density in 1914					0.003** (0.001)	−0.032*** (0.012)	0.003* (0.001)	−0.038** (0.015)
Urban rate in 1914					0.087** (0.042)	0.415 (0.337)	0.068 (0.045)	0.166 (0.438)
Constant	0.340* (0.182)	7.936*** (1.503)	0.472** (0.188)	6.984*** (1.850)	0.180 (0.167)	7.982*** (1.374)	0.297* (0.171)	7.316*** (1.784)
Province fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
Geographic controls	Y	Y	Y	Y	Y	Y	Y	Y
Socio-economic controls	N	N	N	N	Y	Y	Y	Y
Kleibergen-Paap stat.	23.46	–	16.95	–	36.16	–	12.80	–
R^2	0.688	–	0.673	–	0.816	–	0.799	–
Observations	136	136	136	136	136	136	136	136

Notes. First and second-stage regressions with robust standard errors corrected for heteroscedasticity (White) in parentheses. Dependent variable in the second stage is ln *per capita* GDP in 1994. All regressions include geographical control variables and province fixed effects. Regressions in columns (5)–(8) also include socio-economic controls. In columns (2) and (6) the IV is constructed using the actual number of Argentinians in 1869. In columns (4) and (8) the IV is constructed using the average number of Argentinians in 1869 for those provinces with Argentinian population in 1857. See main text for data sources and details. ***p < 0.01, **p < 0.05, *p < 0.1.

Table 5

Share of European Population and Development

	Specification 1		Specification 2	
	IV 1 (1)	IV 2 (2)	IV 1 (3)	IV 2 (4)
Panel (*a*): dependent variable: share of population with higher education 2001				
Share of the European-born population in 1914	0.051	0.125**	0.092**	0.162**
	(0.044)	(0.056)	(0.046)	(0.078)
Panel (*b*): dependent variable: share of pop. with high-skilled occupations 2001				
Share of the European-born population in 1914	0.175**	0.289***	0.182***	0.280***
	(0.069)	(0.075)	(0.064)	(0.094)
Province fixed effects	Y	Y	Y	Y
Geographic controls	Y	Y	Y	Y
Socio-economic controls	N	N	Y	Y
Observations	136	136	136	136

Notes. IV regressions with robust standard errors corrected for heteroscedasticity (White) in parentheses. Dependent variable in panel (*a*) is the share of population 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 (3) and (4) also include socio-economic controls. In columns (1) and (3) the IV is constructed using the actual number of Argentinians in 1869. In columns (2) and (4) the IV is constructed using the average number of Argentinians in 1869 for those provinces with Argentinian population in 1857. See main text for data sources and details. ***p < 0.01, **p < 0.05, *p < 0.1.

In Table 6, panel (*b*), I also consider variations to the fertility and mortality rates in the demographic model. In particular, I let the fertility rate be 10% for Europeans and continue to use 5.3% for Argentinians (columns (1) and (2)), and then replicate this exercise by switching the fertility rate between them (columns (3) and (4)). I conduct a similar exercise with the mortality rate, raising it to 6% for Europeans, compared to 2.2% for Argentinians, and again, I also consider the opposite case (panel (*c*)). Finally, in panel (*d*), I consider that Europeans move as well, raising both fertility and mortality for Europeans only (columns (1) and (2)). I then consider Argentinians to move at 4%, simultaneously raising fertility and mortality for Argentinians only (columns (3) and 4). In all these robustness checks, the result remains qualitatively unchanged.

I also present an alternative definition of the synthetic share of Europeans; columns (2) and (4) in Table 6 use the average number of Argentinians in 1857 in the computation, as discussed in the previous Section. I find that changing this assumption has little effect on the results. Finally, I consider alternative hypotheses that might explain divergence in the paths of economic development. Wheat was one of the most valuable export crops (Ferreres, 2005) and by 1900 made up half of all exports. I therefore include the share of land dedicated to wheat in 1914 as a control variable in panel (*e*), columns (1) and (2). In panel (*e*), columns (3) and (4), I explore whether inequality in the distribution of land across individuals can have a (likely negative) effect on development. Based on 1914 census data on plot size, I construct a county-level land-Gini variable and include it as a separate control in the regression. I both cases, I find that the coefficients on the share of Europeans are very similar to the baseline results, and remain significant. Results for land inequality suggest that there is indeed a negative effect on development (in line with Galor *et al.*, 2009), although the coefficient is not statistically significant.

Table 6

Robustness Checks

	Ln *per capita* GDP 1994			
Dependent variable:	IV 1 (1)	IV 2 (2)	IV 1 (3)	IV 2 (4)
Panel (*a*): assumption:	Moving rate 1.95% for EU and Argentinians		Moving rate 4% for EU and 1.95 for Argentinians	
Share of the European-born population in 1914	6.446*** (1.404)	8.811*** (2.007)	6.558*** (1.354)	8.951*** (2.139)
Panel (*b*): assumption:	Fertility rate 10% for EU		Fertility rate 10% for Argentinians	
Share of the European-born population in 1914	5.412*** (1.923)	9.691*** (3.056)	7.227*** (1.340)	8.645*** (1.889)
Panel (*c*): assumption:	Mortality rate 6% for EU		Mortality rate 6% for Argentinians	
Share of the European-born population in 1914	6.740*** (1.343)	8.697*** (1.906)	5.146*** (1.946)	9.770*** (3.086)
Panel (*d*): assumption:	Moving rate 1.95%, fertility rate 10% and mortality rate 6% for EU		Moving rate 4%, fertility rate 10% and mortality rate 6% for Argentinians	
Share of the European-born population in 1914	6.337*** (1.440)	8.874*** (2.065)	6.225*** (1.568)	9.158*** (2.342)
Panel (*e*): assumption:	Including share of land sown with wheat in 1914		Including share of Land-Gini in 1914	
Share of the European-born population in 1914	6.117*** (1.708)	9.757*** (3.210)	5.817*** (1.555)	8.694*** (2.386)
Share of land sown with wheat in 1914	0.328 (0.678)	−0.696 (1.161)	–	–
Land-Gini in 1914	–	–	−0.839* (0.487)	−0.396 (0.659)
Panel (*f*): assumption:	Including year of conquest		Clustered standard errors at the year of conquest-level	
Share of the European-born population in 1914	5.322*** (2.020)	10.665*** (3.852)	6.350 p-value = 0.0075	6.350 p-value = 0.0075
Year of conquest	0.002 (0.003)	−0.002 (0.004)	–	–
Share of the European-born population in 1914			Clustered standard errors at the provincial-level	
			p-value = 0.1700	p-value = 0.0575
Share of the European-born population in 1914			Clustered standard errors at the provincial-land-quality-level	
			p-value = 0.0525	p-value = 0.0125
Province fixed effects	Y	Y	Y	Y
Geographic controls	Y	Y	Y	Y
Socio-economic controls	N	N	N	N
Observations	136	136	136	136

Notes. IV regressions with robust standard errors corrected for heteroscedasticity (White) in parentheses. All regressions include geographical control variables and province fixed effects. See text for a description of all robustness checks. In columns (1) and (3) the IV is constructed using the actual number of Argentinians in 1869. In columns (2) 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.

Tables 2 and 4 above show that the main result is robust to the inclusion of a series of control variables that are plausibly correlated with later development, such as distance to the city of Buenos Aires, land quality and the availability of railroads. Next, I consider if there is any direct effect of the year in which a county was conquered on later development. To take this into account, I include a separate control variable for the year of conquest in the main regression. The IV estimates in Table 6, panel (*f*), columns (1) and (2), show that there is no qualitative difference in the estimation of the main result, the coefficient on year of conquest is not statistically significant. This is to be expected as the year of conquest affects development through the IV and, as described in subsection 2.2, the timing of conquest was determined by factors unrelated to later development.

In panel (*f*), columns (3) and (4), I correct standard errors with wild bootstrapped cluster-robust standard errors. In particular, I use Webb six-point weight distribution (Webb, 2014; Cameron and Miller, 2015), I report the p-value for the significance of the coefficient. First, I cluster at the year of conquest level, totalling seven clusters; results show rejection of the null hypothesis for the variable of interest being zero. Next, I cluster at the province level, totalling 4 clusters. In column (4), the coefficient remains significant, while in column (3) it is not. This result is not surprising, given the very few number of clusters (4 clusters) of different sizes (cluster sizes are 82, 24, 17, and 13). MacKinnon and Webb (2017) show that for cases with very few clusters (under 5) and different cluster sizes, even the wild bootstrap cluster errors (with Webb six-point distribution) can under-reject the null hypothesis. For this reason, I consider a special case where I augment the number of clusters by splitting province-clusters by quality of land. In particular, I divide land into two categories – above and below median land-quality – and generate a total of seven clusters.[24] Standard errors for the augmented province clusters show that the null hypothesis for the coefficient of interest being equal to zero is rejected.

3.3. *The European Origin of Industrialisation in the Pampas: a Human Capital View*

How did Europeans affect economic outcomes almost a century after their arrival? Why does the immigrant share of Europeans matter? Glaeser *et al.* (2004) argue that Europeans who settled in the Americas brought their human capital and thus affected long-run development. A growing literature focuses on the effect of European inflows during the Age of Mass Migration on human capital: Abramitzky *et al.* (2012) and Lafortune *et al.* (2014) study the case of the US, de Carvalho and Monasterio (2012) and Rocha *et al.* (2015) analyse the case of Brazil. In this Section, I show that migrant inflow during the Age of Mass Migration changed the human capital profile of the population in the fertile plains.

3.3.1. *Literacy*

To understand the change in the human capital profile brought about by the arrival of Europeans, I analyse literacy rates, a measure frequently used in the literature

[24] Buenos Aires, Santa Fe and Cordoba each have two clusters. Entre Rios, having only below median land-quality, has one cluster.

(Squicciarini and Voigtländer, 2015). The average literacy rate in 1895 was 49.2% for Argentinians and 64% for Europeans. The European average masks, however, a very different human capital profile by nationality. Respectively, 89% and 87% of immigrants from Great Britain and Germany were literate, compared to 57% of Italians and 69% of Spaniards. By 1914, the gap had narrowed, respectively, 63.2% and 64.2% for Argentinians and Europeans. The literacy rate for Germans was 88%, while immigrants from Italy, Spain and France had rates of 59.6%, 67.4% and 79.3%, respectively.[25] At the county level, IV estimates in Table 7 show a positive and significant effect of the share of European population on 1895 (panel (*a*)) and 1914 (panel (*b*)) literacy rates. Taken together, the coefficients in Table 7, panels (*a*) and (*b*), imply an effect on literacy rates that is far greater than what could be

Table 7

Share of European Population and Human Capital

	Specification 1		Specification 2	
	IV 1 (1)	IV 2 (2)	IV 1 (3)	IV 2 (4)
Panel (*a*): dependent variable: share of literate population 1895				
Share of the European-born population in 1895	0.710***	1.026***	0.769***	1.104***
	(0.199)	(0.308)	(0.209)	(0.351)
R^2				
Observations	126	126	126	126
Panel (*b*): dependent variable: share of literate population 1914				
Share of the European-born population in 1914	0.059*	0.084*	0.069*	0.108*
	(0.035)	(0.046)	(0.038)	(0.063)
R^2				
Observations	136	136	136	136
Panel (*c*): dependent variable: public schools per 1,000 school-aged population 1914				
Share of the European-born population in 1914	−13.676***	−24.101***	−13.293***	−26.465***
	(4.588)	(6.614)	(5.132)	(8.648)
R^2				
Observations	136	136	136	136
Panel (*d*): dependent variable: private schools per 1,000 school-aged population 1914				
Share of the European-born population in 1914	1.802	5.659**	1.626	6.399*
	(1.738)	(2.395)	(2.025)	(3.322)
R^2				
Observations	126	126	126	126
Province fixed effects	Y	Y	Y	Y
Geographic controls	Y	Y	Y	Y
Socio-economic controls	N	N	Y	Y

Notes. OLS and IV regressions with robust standard errors corrected for heteroscedasticity (White) in parentheses. All regressions include geographical control variables and province fixed effects. Regressions in columns (4)–(6) also include socio-economic controls. In columns (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. See main text for data sources and details. ***p < 0.01, **p < 0.05, *p < 0.1.

[25] See Table A1 in the online Appendix for more details.

explained by a composition effect. While Europeans were on average 15% more literate than Argentinians in 1895, the IV estimate is 0.71, implying that the composition effect (that is, replacing a less literate Argentinian with a more literate European) can explain part, but not all, of the difference in literacy rates across counties. These results are consistent with Glaeser *et al.* (2004) and Easterly and Levine (2016); the arrival of Europeans in the fertile plains raised the human capital profile of the population. To support this argument further, I investigate whether more education was provided in areas with a higher share of European immigrants. In the mid-eighteenth century, the Argentine government built schools throughout the country, offering free public education to school-aged children (6–14 years old). These schools were mostly in urban or densely populated areas, and private and religious schools were also available. Given the government's active policy of public education, it is plausible that in counties with a higher share of Europeans, more publicly financed education was provided in order to assimilate immigrants into Argentinian society. However, empirical results show a different story: areas with higher shares of European immigrants are associated with a higher number of private schools and a lower number of public schools (per school-aged population).

Based on 1914 Census data on schools, I construct the number of schools per 1,000 school-aged children in each county. On average, there were 5.3 public schools and 0.85 private schools in each county per 1,000 school-aged individuals, with standard deviations of 2.32 and 0.71, respectively. Table 7, 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 (1), an increase of one standard deviation in the share of European population in 1914 reduces the number of public schools by 0.66 standard deviations, 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 private schools. Following column (1), an increase of one standard deviation in the share of immigrants increases by 0.28 standard deviations 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.

3.3.2. *Skills*

In the context of the massive arrival of Europeans to the fertile plains, it is important to study another aspect of the human capital they brought: their skills. The European migrants were not only more literate but they also possessed a different set of skills that ended up being conducive to economic growth.

Upon arrival, immigrants reported their occupation, together with their age, gender, marital status and country of origin to the migration office. Table 8 summarises immigrant occupations in the eight available categories, for all immigrants between 1876 and 1897, by country of origin. For all nationalities, 'farmer' is the most common occupation, ranging from 71% of Italians to 34% of immigrants from Great Britain,

Table 8

Immigration 1876–97 by Country of Birth

Occupations	Italy	Spain	France	UK	German-speak. Countries	Belgium	Others	Total
Farmers	0.71	0.50	0.60	0.34	0.49	0.66	0.40	0.63
Construction workers	0.02	0.03	0.02	0.03	0.01	0.01	0.04	0.02
Craftmen	0.03	0.04	0.08	0.05	0.05	0.06	0.16	0.04
Artists	0.01	0.02	0.02	0.04	0.02	0.01	0.09	0.02
Merchants	0.01	0.01	0.03	0.08	0.07	0.05	0.06	0.02
Day workers	0.11	0.16	0.07	0.07	0.08	0.04	0.05	0.11
Other professions	0.03	0.08	0.08	0.20	0.11	0.07	0.06	0.05
No profession	0.09	0.15	0.11	0.19	0.18	0.10	0.14	0.11

Source. Main occupation categories for immigrants by country of birth. Data from Immigrations' Office Annual Reports 1857–1924. See main text for data source and details on occupation categories.

and between 50% and 60% for immigrants from Spain, France, and German-speaking countries. Notably, only 3% of Italians, compared to respectively 8%, 5%, and 5% of immigrants from France, Great Britain, and German-speaking countries declared 'craftsman' as their occupation. 'Merchant' was a common occupation among immigrants from German-speaking countries and Great Britain, 7% and 8% respectively, while only 1% of Italians and Spaniards declared the same. 'Other Professions' were popularly reported among British and German-speaking immigrants; unfortunately there are no additional details, although excluding the categories of farmer, merchant, day-worker, or 'without profession', reduces the list of possibilities. The reported information on occupations upon arrival is useful but to analyse the importance of immigrant skills further I turn to county level census data. Micro-samples from the 1895 Census allow categorising occupations into high and low-skilled, and analysis of whether the share of European immigrants working in high- and low-skilled occupations contributed differentially to the process of economic development.[26]

Data restrictions do not allow computation of the actual share of Europeans by skills in 1914, so I resort to estimating a reduced-form equation:

$$y_i = \alpha + \beta_1 \times \text{synth HSE}_i + \beta_2 \times \text{synth LSE}_i + X_i\gamma + \eta_p + \epsilon_i, \quad (5)$$

where synth HSE and synth LSE are the synthetic share of high and low-skilled Europeans. These synthetic high and low-skilled Europeans are constructed in a similar fashion to those in the previous Section: here I use the number of high and low-skilled workers in the base year in each county and apply national growth rates by country of birth to predict the share of high and low-skilled immigrants in each county in 1914.[27] I use (log *per capita*) GDP in 1994 as the outcome variable and, similar to the

[26] See the online Appendix for definitions and constructions of these variables.

[27] In particular, following Boustan *et al.* (2013), I construct a synthetic share of high and low-skilled Europeans in each county, that is not a function of the changing composition of the area. This allows isolating change in the local composition that is driven by national shifts over time. I use 1895 as the base year and take the number of high- and low-skilled workers as given, where workers are males aged 14 and above. Next, I apply national growth rates by country of birth, where the growth rates are computed from the 1895 and 1914 censuses. Using this 1914 predicted (working) population, I compute the synthetic share of high- and low-skilled immigrants.

main specification in the previous Section, I include control variables for geography, railroad availability and distance to the city of Buenos Aires.

Table 9, column (1) shows that the point estimate for high-skilled Europeans is higher than the estimate for low-skilled Europeans, although the difference is not statistically different from 0 (see the Wald test at the bottom of the Table).[28] Following Hornung's (2014) and Squicciarini and Voigtländer's (2015) argument on the relation between skills and industrialisation, I proceed to estimate the same equation as above, using industrial GDP in 1994 as the dependent variable. In this case, column (2) shows a coefficient of 23.02 for high-skilled European workers, and a coefficient of 1.35 for low-skilled workers, in both cases statistically significant. Moreover, these results show that not all immigrants contributed equally; the coefficients are statistically different from each other (p-value: 0.035). A 1% increase in the share of high-skilled workers in 1914 increases current *per capita* industrial GDP by 26%, and a 1% increase in the share of low-skilled workers increases *per capita* industrial GDP by only 1.3%.

To delve more deeply into this last result, I explore the origin of this industrialisation process by analysing aggregate data from industrial censuses in 1895, 1914 and 1935. The biggest industries at the time were the meat industry, the flour industry, and the production of sugar and wine. On a smaller scale, there was a growing number of small shops offering an array of consumer goods, tools, and farm implements. Bread, beer, candles, soap, bricks, tiles, furniture, garments, etc. were locally produced, and often competed with imported alternatives (Rocchi, 2006). Table 10 shows that immigrants were over-represented both in the ownership of industrial establishments and in the industrial workforce: for example, in 1895, 81% of the establishments were owned by foreigners, while 56% of their workers were immigrants.[29] Close to 20 years later, in 1913, 65% of industrial establishments were run by foreigners, and workers of foreign origin made up 49% of the labour force. Table 10 shows that even as late as in 1935, 58% of industrial establishments were still under the ownership of foreign citizens.

The first census with disaggregated, county-level data on industrial establishments was carried out in 1935. From this census, I compute five measures of industrial development: the number of establishments per person, the *per capita* value of production, the share of skilled workers in the population, the share of industrial workers in the population, and the investment in energy per person. Table 9, columns (3)–(7) show the result for the reduced form model, where I regress these five different 1935 industrial outcomes on the synthetic share of high and low-skilled Europeans, including a set of control variables. Two striking facts stand out from that

[28] Note that the number of observations drops to 126; in this set of regressions information comes from the 1895 census, where data for fewer counties is available.

[29] In 1895, out of a total of 288 flour mills in Buenos Aires, Santa Fe, Entre Rios and Cordoba, 206 belonged to European owners. The situation in the meat processing industry was similar; out of 39 establishments, 23 belonged to immigrants. Likewise, there were 49 beer factories owned by immigrants. In comparison, 12 factories were owned by Argentinians, employing 722 immigrants and 235 Argentinians. There were 13 establishments producing gas, eight owned by Argentinians, and five by immigrants, employing 1,465 immigrants and 108 Argentinians. Finally, out of the 44,100 commercial establishments, only one-quarter were owned by Argentinians, together employing around 72,000 Argentinians and 97,000 immigrants (1895 and 1914 Censuses and Rocchi, 2006).

Table 9

Share of High-skilled and Low-skilled Europeans and Industrial Development

Dependent variable:	GDP 1994 (1)	Industrial GDP 1994 (2)	Industrial production 1935 (3)	Share of skilled ind. workers 1935 (4)	Share of ind. workers 1935 (5)	Num. of factories 1935 (6)	Energy in H.P. 1935 (7)
Synthetic share of high-skilled Europeans in 1914	8.331 (8.703)	42.147** (18.449)	34.408** (13.729)	84.969** (42.675)	0.999*** (0.375)	82.943** (32.975)	2.590 (2.012)
Synthetic share of low-skilled Europeans in 1914	1.291*** (0.493)	2.470*** (0.891)	2.292*** (0.719)	5.879*** (2.114)	0.051** (0.022)	5.243*** (1.407)	0.172** (0.069)
p-value for the test: high-skilled = low-skilled	0.422	0.035	0.022	0.068	0.014	0.020	0.230
Province fixed effects	Y	Y	Y	Y	Y	Y	Y
Geographic controls	Y	Y	Y	Y	Y	Y	Y
Socio-economic controls	N	N	N	N	N	N	N
Adjusted R^2	0.462	0.414	0.172	0.125	0.114	0.443	0.063
Observations	126	126	126	126	126	126	126

Notes. Reduced-form regressions using the synthetic share of high and low-skilled immigrants as of 1914. Dependent variable in columns (1) and (2) is *per capita* GDP and Industrial *per capita* GDP in 1994. Dependent variables in columns (3)–(7) are industrial variables from 1935 Industrial census. All regressions include geographical control variables and province fixed effects. Regressions in columns (4)–(6) also include socio-economic controls. Regressions with robust standard errors corrected for heteroscedasticity (White) in parentheses. See main text for data sources and details on construction of the IV. ***p < 0.01, **p < 0.05, *p < 0.1.

estimation. First, high and low-skilled Europeans positively affect all the measures of industrialisation considered (with the exception of high-skilled workers and investment in energy). Moreover, the point estimate for high-skilled Europeans is always greater than the estimate for low-skilled Europeans, as shown by the p-value of a Wald test for equality of coefficients in the Table (again, with the exception of high-skilled workers and investment in energy). An increase of one percentage point in the share of high-skilled Europeans in 1914 raises the value of industrial production in 1935 by 41%, and the number of establishments by 130%, while the same increase in the share of low-skilled workers increases industrial production by 2% and the number of establishments by 5%.

Hornung's (2014) results show how long-term industrial productivity is positively affected by the technological knowledge introduced by the Huguenots. The findings in this Section similarly show how human capital, as measured by immigrant skills, was one of the channels through which European immigration affected the onset of industrialisation in the Pampas. The arrival of highly skilled Europeans benefited the

Table 10

Ownership of Industrial Establishments 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. Ownership of industrial establishments and industrial workers from 1935 National Industrial census.

Table 11

Industrial Development in 1935 and GDP in 1994

	(1)	(2)	(3)	(4)	(5)
Panel (*a*): dependent variable:			Ln *per capita* GDP 1994		
Ln value of industrial production 1935	0.238***				
	(0.067)				
Share of skilled workers 1935		0.093**			
		(0.039)			
Skilled workers per 1,000 individuals 1935			7.619**		
			(3.527)		
Factories per 1,000 individuals, 1935				0.107***	
				(0.037)	
Energy in H.P. *per capita* 1935					0.371
					(0.467)
Panel (*b*): dependent variable:			Ln *per capita* industrial GDP 1994		
Ln value of industrial production 1935	0.529***				
	(0.112)				
Share of skilled workers 1935		0.230***			
		(0.043)			
Skilled workers per 1,000 individuals 1935			20.196***		
			(4.048)		
Factories per 1,000 individuals, 1935				0.279***	
				(0.057)	
Energy in H.P. *per capita* 1935					2.473**
					(1.111)
Province fixed effects	Y	Y	Y	Y	Y
Geographic controls	Y	Y	Y	Y	Y
Socio-economic controls	N	N	N	N	N
Observations	136	136	136	136	136

Notes. This Table shows the correlation between GDP in 1994 and past measures on industrialisation. Dependent variable in panel (*a*) is ln *per capita* GDP in 1994, and in panel (*b*) is ln *per capita* Industrial GDP in 1994. As independent variables there are different measures of industrialisation in 1935. See main text for data sources and details on these variables. OLS regressions with robust standard errors corrected for heteroscedasticity (White) in parentheses. All regressions include geographical control variables and province fixed effects. ***p < 0.01, **p < 0.05, *p < 0.1.

counties where they settled; industries arose and their positive impact persisted over time.

Finally, I show that the level of industrial development in 1935 is strongly associated to present-day GDP. Table 11 shows results from regressing log *per capita* GDP in 1994 on five measures of industrial development in 1935, including all the control variables considered in the estimations above. Columns (1)–(5) show that the coefficients are both positive and significant; counties that had higher industrial development in 1935 also had higher GDP in 1994. Panel (*b*) repeats this exercise with industrial GDP in 1994 as the dependent variable; results point to the same conclusion as in panel (*a*). It would be interesting to go further and estimate a model in which the historical share of Europeans affects the level of industrial development, which in turn explains current *per capita* GDP, in a 3SLS form. Unfortunately, although human capital is an important channel through which Europeans contributed to economic development, as has been shown above, it may not be the only channel. The share of Europeans may affect development through other channels in addition to early literacy and industrialisation. Consequently, a model in which the share of Europeans affects the level of industrial development (or literacy), in turn explaining current *per capita* GDP, could potentially not satisfy the exclusion restriction for the instrument. For this reason, I examine separately the impact of the share of Europeans on literacy or industrialisation, as well as on the overall effect of the share of Europeans on *per capita* GDP in 1994.

3.3.3. *Nationality*

I use the composition of the migrant inflow to analyse the long-term effects of Europeans by nationality also.[30] As an exploratory exercise, I decompose the effect of the share of Europeans by country of birth. I consider the share of Italians and the share of Spaniards etc., encompassing all major immigrant groups, with the aim of capturing differential effects by country of origin so as to further understand how the share of Europeans contributed to economic growth. To this end, in Table 12 I regress the main dependent variables on the share of immigrants coming from Italy, Spain, France, Great Britain and the German-speaking countries (Germany, Austria, and Switzerland).[31] Column (1) shows OLS results when current log *per capita* GDP is the dependent variable. Columns (2)–(4) show results for the 1935 industrial variables: coefficients for northern Europeans (Great Britain, France, and the German-speaking countries) are generally higher than coefficients for southern Europeans (Italy and Spain). Panel (*b*) repeats this exercise, using instead instruments for the share of each nationality. As in the previous analysis, I use synthetic instruments based on the 1895 census micro-samples: here I take the number of males aged 14 and older by country of origin as the base population and apply population growth rates by country of origin.[32]

[30] Data availability prevents a county-level analysis by occupation and nationality.

[31] I pool together German-speaking countries, as in Abramitzky *et al.* (2012), given that each group's individual size is relatively small.

[32] As before, population growth rates are computed comparing the stock of population by nationality between the 1895 and 1914 censuses.

Table 12

Share of Europeans by Country of Birth and Development

Dependent variable:	GDP 1994 (1)	Industrial GDP 1994 (2)	Industrial production 1935 (3)	Share of skilled ind. workers 1935 (4)	Share of ind. workers 1935 (5)	Num. of factories 1935 (6)	Energy in H.P. 1935 (7)
Panel (*a*): OLS results							
Share of immigrants in 1914 from:							
Germany, Austria, Switz. 1914	13.668**	39.051***	21.635**	51.503**	0.561*	19.779	3.890***
	(5.370)	(9.952)	(10.529)	(22.460)	(0.311)	(26.787)	(1.391)
Great Britain 1914	50.506***	105.482***	112.859***	238.725**	1.982*	105.057	6.568*
	(18.618)	(36.103)	(31.912)	(107.658)	(1.118)	(78.614)	(3.928)
France 1914	12.660***	22.877***	17.225**	1.610	0.045	31.558***	−0.496
	(3.882)	(7.088)	(7.229)	(14.875)	(0.185)	(11.993)	(1.007)
Italy 1914	4.287***	8.534***	6.149***	14.639**	0.134**	11.460***	0.569**
	(1.237)	(2.779)	(2.123)	(5.697)	(0.063)	(4.329)	(0.274)
Spain 1914	4.257***	−0.953	2.902	5.932	0.061	4.789	0.023
	(1.084)	(2.237)	(1.999)	(4.347)	(0.054)	(4.716)	(0.272)
Observations	136	136	136	136	136	136	136
Panel (*b*): IV results							
Germany, Austria, Switz. 1914	14.124**	40.427***	39.948***	68.987**	1.019*	54.320*	6.061***
	(6.689)	(11.340)	(13.571)	(34.848)	(0.567)	(30.047)	(2.192)
Great Britain 1914	72.626***	116.419**	63.441*	100.323	0.888	181.362**	−0.377
	(22.950)	(46.188)	(38.071)	(92.039)	(1.114)	(73.199)	(4.170)
France 1914	11.054***	18.566***	15.062**	14.173	0.022	13.841	−0.539
	(3.042)	(5.672)	(6.546)	(19.575)	(0.187)	(19.683)	(1.074)
Italy 1914	5.215***	14.171***	12.027***	38.703***	0.323***	23.913***	1.382***
	(1.875)	(3.803)	(3.094)	(8.078)	(0.090)	(8.606)	(0.362)
Spain 1914	2.947	1.085	1.995	21.760*	0.188	14.891	0.194
	(2.347)	(5.461)	(4.389)	(11.279)	(0.127)	(10.655)	(0.685)
Observations	126	126	126	126	126	126	126
Province fixed effects	Y	Y	Y	Y	Y	Y	Y
Geographic controls	Y	Y	Y	Y	Y	Y	Y
Socio-economic controls	N	N	N	N	N	N	N

Notes. Panel (*a*) shows OLS results, panel (*b*) shows IV results. Column (1) shows results for the regression of ln *per capita* GDP in 1994 on the share of population born in Germany, Austria, Switzerland, Great Britain, France, Italy and Spain in 1914; and column (2) shows results for ln *per capita* industrial GDP as the dependent variable. Columns (3)–(7) repeat this exercice for different measures of industrialisation in 1935. See main text for data sources and details. OLS and IV regressions with robust standard errors corrected for heteroscedasticity (White) in parentheses. All regressions include geographical control variables and province fixed effects. ***p < 0.01, **p < 0.05, *p,< 0.1.

The synthetic population is then used to construct a share for each nationality as of 1914.[33] Column (1) shows IV results with log *per capita* GDP as the dependent variable, where the coefficients are 72.6, 14.1, 11, 5.2 and 2.9 for Great Britain, German-speaking countries, France, Italy and Spain, respectively. A t-test for the equality of coefficients rejects the null hypothesis that the coefficient for Italy or Spain are equal to the coefficient of Great Britain and France but it does not reject the equality with the coefficient for German-speaking countries. Columns (3)–(7) show results for the 1935 industrial variables. In all cases, IV estimates remain qualitatively similar to OLS estimates and the point estimates for northern Europeans are higher than the estimates for southern Europeans.

To support the hypothesis that human capital drives the relation between the historical composition of the population and current economic development, I analyse several measures of average characteristics across European nationalities.[34] I start by comparing average literacy rates in Europe with the literacy of European immigrants arriving in Argentina. Based on Cipolla (1969) and Crafts (1996), literacy rates in 1870 were 76%, 80%, 69%, 32% and 28% for Great Britain, Germany, France, Italy and Spain respectively. Using the 1895 census micro-sample, I compute literacy rates for all immigrants aged 14 and older, and find that 89%, 87%, 76%, 57% and 69% of immigrants from Great Britain, Germany, France, Italy and Spain were, respectively, literate. That is to say, immigrants from northern European countries living in Argentina had higher literacy rates than immigrants from southern European countries and these rates were even higher than the averages in their home countries. Data on enrolment rates collected by Benavot and Riddle (1988) for the period 1870–900 show a similar picture; enrolment rates were 54%, 66%, and 77% for the countries of Great Britain, German-speaking countries and France, and 35% and 49% for Italy and Spain.

Moreover, I compare Bairoch's (1982) data on *per capita* volume of industrial production with the point estimates in Table 10: immigrants from countries with higher levels of industrial output have a greater effect on industrial development in Argentina. For the period 1860–1913, northern European countries had on average an (industrial-production) index of 37.4 (Great Britain in 1900 = 100), whereas southern European countries had an index of 13. By the year 1900, Great Britain was the leading country (value of 100), followed by Switzerland and Germany with an index close to 50, France with an index of 40, and Italy and Spain with indices of 17 and 19 respectively. Table 12, column (3) shows that immigrants from Great Britain, the German-speaking countries and France have IV point estimates of 63.4, 39.9 and 15.1 respectively, while immigrants from Italy and Spain have estimates of 12.0 and 2.0, respectively. Arguably, higher industrialisation levels in the home countries increase the likelihood that migrants were more familiar with industries and the work patterns associated with them. Taken together, these results

[33] The IV developed in subsection 3.2, for the share of the European-born population, cannot be used to generate instruments by country of origin. The previous IV exploits variation in the arrival of Europeans and, although the arrival of immigrants from different countries did not occur simultaneously, there is not enough variation in the arrival-by-country-of-origin data to create separate instruments. I present results for the first-stage in the online Appendix, Table A2.

[34] The following measures of human capital are country-level averages. Sample size in the 1895 census micro-sample for Argentina does not allow to break down the analysis by nationality and characteristics at the county level.

provide suggestive evidence of the composition of the population as a determinant of long run development. Population composition changed dramatically due to mass migration, resulting in a higher profile of human capital, which in turn favoured the onset of an industrialisation process in the fertile plains. As Rocchi (2006, p. 20) notes, 'With immigration, Argentina could profit from the benefits of human capital provided by recent arrivals, an issue especially important for manufacturing – many factories, indeed, could not have opened had the skills of the immigrant population not been available'.

4. Conclusion

During the Age of Mass Migration, an unprecedented flow of Europeans migrated to the fertile plains in Argentina and dramatically changed the composition of the population. Counties where European immigration accounted for a larger share of the total population in 1914 achieved a higher *per capita* GDP in 1994. This positive long-term effect can be linked to the way Europeans changed the human capital profile of the population: their literacy and skills were significant for long-run development. These characteristics and the role played by Europeans were crucial to the onset of industrialisation in Argentina.

The fertile plains of the Pampas provides an excellent empirical setting to analyse these effects, for it is an area that shares common political institutions and uniform geographical conditions. A cross-county analysis therefore allows one to abstract from the main competing hypotheses posed by the literature.

However, where migrants settled was not a random process and presents an empirical challenge. To overcome this issue, I construct an IV based on a synthetic distribution of immigrants across counties that exploits variation in the arrival of Europeans, interacted with variation in the availability of land for settlement. The IV uses only the exogenously generated variation in the share of Europeans across counties. IV estimates show important long run effects of European migration: counties home to a higher share of Europeans currently have higher levels of *per capita* GDP, a more educated population, and a greater share of skilled workers. In addition, the identified effects are large.

The analysis of the European human capital profile is insightful; Europeans arrived with higher literacy rates than the native population, as well as helped raise the overall literacy of the population. Moreover, I show that there is a differential effect between high and low-skilled immigrants. A one percentage point increase in the share of high-skilled Europeans increases *per capita* industrial GDP by 26%, whereas a same increase in the share of low-skilled workers raises industrial GDP by only 1.3%. Importantly, Europeans played a fundamental role during the onset of industrialisation. Immigrants were over-represented both in the ownership of industrial establishments and in the industrial workforce; the share of high skilled Europeans in 1914 explains a significant part of industrial development, both in the short and long run. An analysis broken down by nationality of the migrants reveals the higher contribution of northern Europeans, as compared to southern Europeans. This can be linked to industrialisation levels in home countries using Bairoch's (1982) data: migrants from Great Britain, Germany, Austria, Switzerland and France came from the most industrialised countries in Europe and were likely more familiar with industries, less so the case for migrants from Italy and Spain.

This article contributes to the growing literature on the long-run determinants of growth, and to the literature on the effects of migration. In particular, the results I present show the importance of human capital, in a setting with shared institutions and uniform geography, for the process of economic development and the persistence of this effect in the long run. Importantly, I do not argue that human capital is the only channel through which migrants affected long-term development, but rather that in this setting, human capital differences are key, as they combined profitably with industrial development. As immigration is at the forefront of policy concerns, this article provides evidence on the positive long-term impact of large-scale immigration.

FAE - Universidad de Santiago de Chile

Submitted: 10 August 2015
Accepted: 20 January 2017

Additional Supporting Information may be found in the online version of this article:

Appendix A. Source Data and Variables.
Data S1.

References

Abramitzky, R. and Boustan, L.P. (2017). 'Immigration in American Economic History', *Journal of Economic Literature*, vol. 55(4), pp. 1311–45.
Abramitzky, R., Boustan, L.P. and Eriksson, K. (2014). 'A nation of immigrants: assimilation and economic outcomes in the age of mass migration', *Journal of Political Economy*, vol. 122(3), pp. 467–506.
Abramitzky, R., Boustan, L.P. and Eriksson, K. (2012). 'Europe's tired, poor huddled masses: self-selection and economic outcomes in the age of mass migration', *American Economic Review*, vol. 102(5), pp. 1832–56.
Acemoglu, D., Gallego, F. and Robinson, J.A. (2014). 'Institutions, human capital and development', *Annual Reviews of Economics*, vol. 6, pp. 875–912.
Acemoglu, D., Johnson, S. and Robinson, J.A. (2001). 'The colonial origins of comparative development: an empirical investigation', *American Economic Review*, vol. 91(5), pp. 1369–401.
Acemoglu, D., Johnson, S. and Robinson, J.A. (2002). 'Reversal of fortunes: geography and institutions in the making of the modern world income distribution', *Quarterly Journal of Economics*, vol. 117(4), pp. 1231–94.
Banerjee, A. and Iyer, L. (2005). 'History, institutions and economic performance: the legacy of colonial land tenure systems in India', *American Economic Review*, vol. 95(4), pp. 1190–213.
Bairoch, P. (1982). 'International industrialisation levels from 1750 to 1980,' *Journal of European Economic History*, 11(2), 269–333.
Becker, S.O., Hornung, E. and Woessmann, L. (2011). 'Education and catch-up in the industrial revolution', *American Economic Journal: Macroeconomics*, vol. 3(3), pp. 92–126.
Becker, S.O. and Woessmann, L. (2009). 'Was Weber wrong? A human capital theory of protestant economic history', *Quarterly Journal of Economics*, vol. 124(2), pp. 531–596.
Benavot, A. and Riddle, P. (1988). 'The expansion of primary education, 1870–1940: trends and issues', *Sociology of Education*, vol. 61(3), pp. 191-210.
Boustan, L., Ferreiro, F., Winkler, H. and Zolt, E. (2013). 'Income inequality and local government in the United States, 1970–2000', *Review of Economics and Statistics*, vol. 95(4), pp. 1291–1302.
Bruhn, M. and Gallego, F. (2012). 'Good, bad, and ugly colonial activities: do they matter for economic development?', *The Review of Economics and Statistics*, vol. 94(2), pp. 433–61.
Cacopardo, M.C. (1967). *República Argentina: Cambios en Los Límites Nacionales, Provinciales y Departmentales, A Través de Los Censos Nacionales de Población*, Buenos Aires: Instituto Torcuato Di Tella.
Cameron, A.C. and Miller, D.L. (2015). 'A practitioner's guide to cluster-robust inference', *Journal of Human Resources*, vol. 50(2), pp. 317–72.
Cipolla, C.M. (1969). *Literacy and Development in the West*, Baltimore, MD: Penguin Books.

Cortés Conde, R. (1968). 'Algunos rasgos de la expansión territorial en argentina en la segunda mitad del siglo XIX,' *Desarrollo Económico*, vol. 8(29), pp. 3–29.
Crafts, N.F.R. (1996). 'The human development index: some historical comparisons', Working Papers in Economic History, LSE, Number 33.
Cruzate, G., Gomez, L., Pizarro, M.J., Mercuri, P., Banchero, S., Angueira, G., Prieto, D., Lopez, J. and Ibarraza, G. (1990). '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.
de Carvalho, F.I. and Monasterio, L.M. (2012). 'Immigration and the origins of regional inequality: government-sponsored European migration to Southern Brazil before World War I', *Regional Science and Urban Economics*, vol. 42(5), pp. 761–904.
Dirección General de Inmigración. (1925). *Resumen Estadístico del Movimiento Migratorio en la República Argentina. Años 1857–1924*, Buenos Aires: Talleres Graficos del Ministerio de Agricultura de la Nacion.
Easterly, W. and Levine, R. (2016). 'The European origins of economic development', *Journal of Economic Growth*, vol. 21(3), pp. 225–57.
Engerman, S.L. and Sokoloff, K.L. (1997). 'Factor endowments, institutions, and differential paths of growth among New World economies: a view from economic historians of the United States', in (S. Harber, eds), *How Latin America Fell Behind*, pp. 260–304, Stanford, CA: Stanford University Press.
Engerman, S.L. and Sokoloff, K.L. (2002). 'Factor endowments, inequality, and paths of development among New World economies', NBER, Working Paper 9259.
Ferreres, O. (2005). *Dos siglos de economía argentina (1810–2004)*, Fundación Norte y Sur, Argentina.
Franck, R. and Galor, O. (2015). 'Is industrialization conducive to long-run prosperity?', CESifo Working Paper No. 5354.
Gallo, E. (1983). *La Pampa gringa: la colonización agrícola en Santa Fe (1870–1895)*, Buenos Aires: Editorial Sudamericana.
Galor, O. (2011). *Unified Growth Theory*, Princeton, NJ: Princeton University Press.
Galor, O., Moav, O. and Vollrath, D. (2009). 'Inequality in land ownership, the emergence of human capital promoting institutions and the great divergence', *Review of Economic Studies*, vol. 76(1), pp. 143–79.
Glaeser, E.L., La Porta, R., Lopez-De-Silanes, F. and Shleifer, A. (2004). 'Do institutions cause growth?', *Journal of Economic Growth*, vol. 9(3), pp. 271–303.
Hatton, T. and Williamson, J. (1992). 'What drove the mass migrations from Europe in the late nineteenth century', NBER, Historical Paper No. 43.
Hornung, E. (2014). 'Immigration and the diffusion of technology: the Huguenot Diaspora in Prussia', *American Economic Review*, vol. 104(1), pp. 84–122.
Lacoste, P. (2002). 'La guerra de los mapas entre Argentina y Chile: una mirada desde Chile, Santiago', *Historia*, vol. 35, pp. 211–249.
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 US immigration', Working Paper, PUC University, Chile.
La Porta, R., Lopez-de-Silanes, F., Shleifer, A. and Vishny, R.W. (1998). 'Law and finance', *Journal of Political Economy*, vol. 106(6), pp. 1113–55.
La Porta, R., Lopez-de-Silanes, F., Shleifer, A. and Vishny, R.W. (1999). 'The quality of government', *Journal of Law, Economics and Organization* vol. 15(1), pp. 222–79.
Mokyr, J. (2005). 'Long-term economic growth and the history of technology', *Handbook of Economic Growth*, vol. 1, Amsterdam: Elsevier.
Mandrini, R. (2008). *La Argentina aborigen. De los primeros pobladores a 1910*, Buenos Aires: Colección Biblioteca Básica de Historia, Siglo XXI Editores.
Morse, R. M. (1962). 'Latin American cities: aspects of function and structure', *Comparative Studies in Society and History*, vol. 4(4), pp. 473–93.
MacKinnon, J.G. and Webb, M.D. (2017). 'Wild bootstrap inference for wildly different cluster sizes', *Journal of Applied Econometrics.*, vol. 32(2), pp. 233–54.
Nunn, N. (2009). 'The importance of history for economic development', *Annual Review of Economics*, vol. 1, pp. 65–92.
Nunn, N. and Puga, D. (2012). 'Ruggedness: the blessing of bad geography in Africa', *Review of Economics and Statistics*, vol. 94(1), pp. 20–36.
Perry, R. (1972). 'Warfare on the Pampas in the 1870s', *Military Affairs*, vol. 36(2), pp. 52–58.
Putterman, L. and Weil, D. (2010) 'Post-1500 population flows and the long-run determinants of economic growth and inequality', *Quarterly Journal of Economics*, vol. 125(4), pp. 1627–1682.
República Argentina. (1938). *Censo Industrial de 1935*, Buenos Aires: Dirección General de Estadística de la Nación.
República Argentina. (1872). *Primer Censo de la República Argentina, 15, 16 y 17 de Setiembre de 1869*, Buenos Aires: Imprenta del Porvenir.

República Argentina. (1898). *Segundo Censo de la República Argentina, Mayo 10 de 1895*, Buenos Aires: Taller Tipográfico de la Penitenciaria Nacional.
República Argentina. (1916). *Tercer Censo Nacional de la República Argentina, Levantado el 1ro de Junio de 1914*, Buenos Aires: Talleres Gráficos de L. J. Rosso y Cía.
Rocchi, F. (2006). *Chimneys in the Dessert*, Rewwood City, CA: Stanford University Press.
Rocha, R., Ferraz, C. and R.R. Soares (2015). 'Human capital persistence and development', IZA Discussion Paper 9101.
Somoza, J.L. and Lattes, A.E. (1967). *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.
Squicciarini, M. and Voigtländer, N. (2015). 'Human capital and industrialization: evidence from the age of enlightenment', *Quarterly Journal of Economics*, vol. 130 (4), pp. 1825–83.
Spolaore, E. and Wacziarg, R. (2013). 'How deep are the roots of economic development?', *Journal of Economic Literature*, vol. 51(2), pp. 325–69.
Taylor, A. (1995). 'Peopling the pampa: on the impact of mass migration to the river plate, 1870–1914', NBER, Historical Working Paper No. 68.
Tell, S. (2008). *Córdoba rural, una sociedad campesina (1750–1850)*. Buenos Aires: Prometeo Libros, Asociación Argentina de Historia Económica.
Walther, J.C. (1964). *La Conquista del Desierto*, Buenos Aires: Círculo Militar.
Webb, M. (2014). 'Reworking wild bootstrap based inference for clustered errors', Queens Economics Department Working Paper, No. 1315.