TEXTO PARA DISCUSSÃO

No. 640

Human Capital Persistence and Development

Rudi Rocha
Claudio Ferraz
Rodrigo R. Soares

DEPARTAMENTO DE ECONOMIA
www.econ.puc-rio.br

Human Capital Persistence and Development*

Rudi Rocha† Claudio Ferraz‡ Rodrigo R. Soares§

June 2015

Abstract

This paper examines the role of human capital persistence in explaining long-term development. We exploit variation induced by a state-sponsored settlement policy that attracted a pool of immigrants with higher levels of schooling to particular regions of Brazil in the late 19^{th} and early 20^{th} century. We show that municipalities that received settlements experienced increases in schooling that persisted over time. One century after the policy, localities that received state-sponsored settlements had higher levels of schooling and income per capita. We provide evidence that long-run effects were driven by persistently higher supply and use of educational inputs and shifts in the structure of occupations towards skill-intensive sectors.

JEL: O15, O18, N36

Keywords: Human capital, Education, Immigration, Development

*We are grateful to Ran Abramitzky, Juliano Assunção, Maria Silvia Bassanezi, Ciro Biderman, Adalberto Cardoso, Ernesto Dal Bó, Irineu de Carvalho Filho, Fred Finan, Francisco Gallego, Gustavo Gonzaga, Joana Naritomi, Naércio Menezes-Filho, André Portela, Pablo Querubin, Eustáquio Reis, and seminar participants at Bocconi University, Paris School of Economics, PUC-Rio, Sao Paulo School of Economics - FGV, Stanford University, Universidad Carlos III de Madrid, University of Essex, the LACEA 2010 Meeting, the SBE 2010 Meeting, and the 1^{st} PUC-Chile Economic History Clio Lab Workshop for comments, data, and suggestions. Financial support from the EH Clio lab at PUC-Chile is gratefully acknowledged.

†UFRJ; Instituto de Economia, Av. Pasteur 250, 22290-240, Rio de Janeiro, RJ, Brazil; e-mail: *rudi.rocha* at *ie.ufrj.br*

‡PUC-Rio and BREAD; Departmento de Economia, Pontifícia Universidade Católica do Rio de Janeiro; Rua Marquês de São Vicente 225, Rio de Janeiro, RJ, 22453-900, Brazil; e-mail: *cferraz* at *econ.puc-rio.br*

§Sao Paulo School of Economics - FGV and IZA; Rua Itapeva 474, São Paulo, SP, 01332-000, Brazil; e-mail: *rodrigo.reis.soares* at *fgv.br*

1 Introduction

Human capital plays a key role in explaining income differences across regions of the world, but identifying the role of human capital as a determinant of long-term development remains a challenge.[1] Many authors argue that differences in education across countries were determined by variation in institutional settings that also drove development. According to this view, countries with less inequality and inclusive institutions provided mass public schooling, which led to development, while unequal countries kept literacy levels low, which hindered long-term growth (North and Thomas, 1973, Acemoglu et al., 2002, Acemoglu et al., 2014, Engerman et al., 2012, Galor et al., 2009). Others claim that characteristics associated with specific groups – such as ethnicity, culture, or religion – drove both higher levels of schooling and long-term development (Easterly and Levine, 2012, Fernandez, 2011, Peri, 2012). These arguments, and the associated evidence, go against the view that human capital alone can determine development, assigning to it instead the role of a mechanism in the causal link between institutional quality – or culture, or work ethics – and long-term growth.

This paper reexamines the role of human capital, and of its persistence through time, as a determinant of long-term development. We exploit a particular historical episode where, after the international ban on slave trade in 1850, and in the midst of a massive inflow of European immigrants to Brazil, immigrants with relatively more education were channeled into specific localities through deliberate government policies. In the end of the 19^{th} century and beginning of the 20^{th} century, public authorities established a number of official settlement colonies throughout the state of São Paulo. This policy had goals involving occupation of territory, food production, "whitening" of the population, and was driven by a centralized decision at the state level. The settlements were established typically near previously existing rural villages and were occupied by relatively high-skill European immigrants of various nationalities. Simultaneously, European immigrants were also spontaneously arriving in virtually every other area of the state. Given the ambivalent motivations, the distribution of state-sponsored settlements across the state ended up being roughly uniform.

We exploit three unique features of this historical episode to isolate the effect of human capital on long-term development. First, differently from the US where immigrants were, on average, less educated than the native population, the age of mass migration in Latin

[1] See Hall and Jones (1999), Glaeser et al. (2004), Acemoglu et al. (2014) for cross-country evidence and Acemoglu and Dell (2010) and Gennaioli et al. (2013) for cross-regional evidence.

America, and specially in Brazil, brought immigrants that were more educated than natives, creating a positive human capital shock (see Bandiera et al. (2015)). Second, settlements were established before Brazil's industrialization period. Thus, the choice of settlement location and migration decisions in an agrarian economy were not correlated with expected returns from education. And third, the areas where state-sponsored settlements were created were not different from other areas, but for the fact that they received immigrants with higher levels of schooling. The exogenous increase in schooling represented by this natural experiment allows us to disentangle human capital from other factors commonly highlighted in the literature on the impacts of European colonization.

Our main empirical strategy compares schooling and other outcomes across municipalities that received a state-sponsored settlement and municipalities that did not receive a settlement but had similar pre-existing characteristics. To measure the effects of state-sponsored settlements we put together data from the 1872, 1920, 1940, and 2000 Brazilian censuses, as well as a vast array of auxiliary information from various historical sources, such as state statistical yearbooks and provincial presidential reports. We construct variables that measure geographic, transportation, and socioeconomic conditions within the state from the end of the 19^{th} century to the year 2000.

We begin by documenting that in 1872, before the establishment of the first state-sponsored settlements, localities that were to receive settlements in the future were very similar to other areas in the state. We then show that the creation of settlements attracted more educated immigrants to settlement municipalities. In 1920, a few years after the establishment of the last settlements, the literacy rate in settlement municipalities was 8 percentage points (or 27 percent) higher than elsewhere in the state, despite an only marginally higher share of immigrants. At that moment, apart from the educational composition of the population, areas that had received settlements were not different in any other respect. We show that this initial shock to human capital persisted not only until 1940, but all the way up to 2000. Approximately one century after the establishment of many of the original settlements, individuals living in the respective municipalities had on average 0.52 more year of schooling when compared to individuals living in other municipalities.

This persistent difference in schooling also translated into differential development paths. Our estimates suggest that average income per capita in 2000 was 15 percent higher in municipalities that had received a state-sponsored settlement when compared to other municipalities. Our hypothesis is that a significant part of this difference was driven by increases in human capital dating back to the settlements, as current income per capita is strongly associated with literacy rates in 1920. This suggests that settlements could be

used as an instrument in a regression relating income per capita in 2000 to literacy rates in 1920. When we implement this strategy our estimate indicates that a 10 percentage point increase in literacy rates in 1920 is associated with a 19 percent higher income per capita in 2000. When we control for years of schooling in 2000, this coefficient is reduced by 85 percent and ceases to be statistically significant, supporting the idea that most of the effect of the settlements on long-term development worked through increased schooling and its persistence over time.

We also implement a series of robustness tests. First, all results are robust to controlling for geographic and pre-settlement (1872) socioeconomic characteristics. Second, the results are robust to using within municipality variation and controlling for municipal fixed-effects (when possible). And finally, the results are also robust to controlling for proxies of culture and social norms associated with nationality (e.g. Italian versus Portuguese) and religion (e.g. Catholic versus Protestant). Additionally, we show that our results are not driven by other potential effects of the settlements. We find no short-term effect of settlements on population density, land inequality, agricultural productivity, land prices, or wages.

In sum, we show that the initial shock represented by the educational composition of immigrants persisted through time and led, in the long-run, to higher income per capita. In order to shed light on the persistence of this initial shock and on how it translated into higher education and income per capita in the long-run, we analyze the evolution in the supply and use of educational inputs and in the structure of employment through time. We show that in 1920, 1940, and 2000, enrollment rates and number of teachers per school aged children were higher in locations associated with settlements. In these same locations, employment shifted away from agriculture and towards manufacturing and, to a lesser extent, services between 1920 and 1940. In 2000, this shift in the structure of employment ended up reflected entirely on a larger services sector. None of these differences in the use of educational inputs and in the structure of employment were present in 1872, before the settlements were established. The evidence suggests that local communities that developed from historical settlements demanded higher educational investments and, over time, shifted economic activity to skill-intensive sectors.

Our paper speaks to three sets of literature. First, it relates to a large literature that measures the impact of population composition on long-run economic development (e.g. Putterman and Weil (2010) and Easterly and Levine (2012)). Distinctively from the existing work, we use micro-data from a single country to exploit a natural experiment that distributed heterogeneous types of immigrants across localities. We also provide evidence

on the specific channels through which the composition of immigrants affected long-term development.

Second, the paper contributes to a growing literature that documents the persistence of human capital through time. Huillery (2009) shows that colonial investments in education in West Africa had effects that persist until today: areas that received higher investments in education in the colonial past still display higher educational investments and higher levels of schooling. Similarly, but in a different setting, Wantchekon et al. (2015) show that individuals selected to attend school in colonial Benin, during the establishment of the first missionary schools, benefited in terms of income and socioeconomic conditions. In addition, in the long-run, the entire villages where historical missionary schools were established displayed improved outcomes. Nunn (2011) also finds persistent effects of protestant missionary schools on educational levels in Africa. Differently from these papers, our results address the mechanisms behind the persistence of human capital and its impact on income per capita.[2]

Finally, our work sheds light on the role of human capital as a determinant of structural change and long-term development. A growing number of papers has focused on the importance of education for productivity improvements through the absorption of new technologies. Hornung (2014), for example, looks at the forced migration of skilled Huguenots from France to Prussia in the late 17^{th} century and shows that firms in areas receiving immigrants experienced increased productivity. The author argues that skilled Huguenots brought industry specific knowledge that was complementary to new technologies. Fourie and von Fintel (2014) look at Huguenots that left wine producing regions of France to migrate to South Africa in this same historical episode and show, using tax records, that they became more productive wine makers than previously established farmers and that this difference in productivity persisted through time. Becker et al. (2011), also looks at Prussia and use occupational data to show that historical variations in schooling across counties, supposedly related to cultural factors, determined the ability of the different regions to take advantage of the technological innovations brought by the industrial revolution. Squicciarini and Voigtlaender (2015) present evidence that "upper-tail knowledge" was an important driver of city growth during the first industrial revolution in France, mainly through increased productivity in industrial technologies. For a more re-

[2]For a related literature that focuses on Latin America, see Summerhill (2010), de Carvalho Filho and Colistete (2010), de Carvalho Filho and Monasterio (2011), Droller (2013), and Valencia Caicedo (2014). Our paper differs from this literature by having a clear identification strategy based on the location of state-sponsored settlements, by isolating the role of human capital, and by analyzing the specific channels behind its historical persistence and its effect on long-term development.

cent period and looking at cross-country data, Ciccone and Papaioannou (2009) show that, during the 1980s and 1990s, countries with higher initial levels of education experienced higher growth in industries intensive in new technologies. Many authors also find a positive relationship between human capital and growth exploiting cross-sectional variation within countries, though mostly focused on contemporary data from developed regions (for example, Glaeser et al., 1995, Shapiro, 2006, and Iranzo and Peri, 2009). By documenting the persistently higher use of educational inputs and the shift in the structure of occupations towards skill-intensive sectors, we are able to present in a single setting, and in an unified and sequential way, several results that appeared fragmented in the previous literature. We show that: (i) the initial human capital shock persisted through time via increased investments in education, as in Huillery (2009); (ii) higher human capital was associated, as industrialization progressed over time, with shifts in the structure of employment towards skill-intensive sectors, as in Becker et al. (2011); and (iii) together, these changes ended up reflected on higher income per capita in the long-run.

The remainder of the paper is structured as follows. Section 2 outlines the historical background and discusses the establishment and the likely impacts of state-sponsored settlements in the state of São Paulo. Section 3 presents the data and descriptive statistics. Section 4 describes the empirical strategy. Section 5 presents the main econometric results, discusses causality channels, and performs robustness exercises. Section 6 closes the paper with some concluding remarks.

2 Historical Background

2.1 Historical Context and the Policy of State-sponsored Settlements

São Paulo has long been among the richest and most industrialized regions not only in Brazil but in all developing world. The transition from agrarian producer to industrial and financial center spanned several decades, culminating with its consolidation as Brazil's wealthiest state in the second half of the 20^{th} century. This process was fueled by the 19^{th} century expansion in coffee production and the construction of an integrated railway system, which connected the seaport of Santos to the fertile and sparsely occupied inland areas (Furtado, 1989). Coffee became then a major driver of infrastructure construction, urban development, and accumulation of wealth, and quickly took over the agricultural frontier (Martins, 1973 and Summerhill, 2003).

The labor demanded during the first years of the coffee expansion was primarily supplied by reallocation of slaves within the country. But the ban on the international slave trade in 1850 (Law Eusébio de Queiroz) and the abolishment of slavery in 1888 (Law Áurea) precluded the use of slave labor as an input in the continued growth of the state. Already by the mid-1850s, rising popular pressure for the end of slavery and increasing shortages of workers and food brought international immigration to the center of the political agenda.[3] At that time, the view that European immigrants were more cultured and productive than afro-descendants and native workers was widespread among the Brazilian elite (Petrone, 1982). In the second half of the 19^{th} century, openness to foreign immigration in general, and to the model of state-sponsored colonies in particular, gained increased support in legislative debates (Martins, 1973).

State-sponsored settlements had two main objectives: food production for the growing urban areas and attraction of "higher" quality immigrants, through the prospect of land ownership, to uninhabited regions. The geographic distribution of settlements across the state reflected this dual policy motivation (Petrone, 1982). On the one hand, the role as food suppliers led to the creation of official colonies in areas with growing demand. On the other hand, in many other cases, unclaimed public lands used in the establishment of settlements were located in marginal or backward areas.

The administrative procedures associated with the creation of official settlements can be described as follows. After legislative approval, public employees searched for locations mainly among unclaimed public lands, but also among private lands and foreclosure opportunities. Given the alternatives available, the precise locality was chosen based on two criteria: reasonable soil fertility and proximity to transportation. Plots of land were then outlined and classified into three different types depending on size and location within settlement. Rural plots were the largest while suburban and urban plots were smaller and more central. The government was responsible for building the basic infrastructure around an administrative office at the center of the colony.

In the 1870s, the government launched a first wave of state-sponsored settlements in São Paulo. Official settlements became a major attraction in the official Brazilian propaganda

[3]The first experiences of state-sponsored international immigration in Brazil date back to the early 19^{th} century, when the imperial government promoted the peopling of strategic and sparsely occupied regions with European families. The first official settlements were the colonies of Nova Friburgo (1818) in the state of Rio de Janeiro and São Leopoldo (1824) in the state of Rio Grande do Sul. This initial wave of colonization also included four areas in the state of São Paulo: two in the surroundings of the state capital (Santo Amaro and Parelheiros, in the 1820s) and two in the coast (Iguape and Cananéia, in the 1850s). Due to economic isolation, the initiatives in São Paulo failed and the farthest colonies of Parelheiros, Iguape, and Cananéia were largely abandoned (Furtado, 1989, Petrone, 1982, Paiva, 1993, and Siriani, 2005).

disseminated in European cities (Petrone, 1982). The plots of land were offered by Brazilian representatives in Europe and at the Immigration Station Houses, the lodgings where immigrants were first allocated to when arriving in Brazil. Immigrants did not receive the land freely, but had to pay for it in installments starting after the first harvest. During this initial period, some scant public support was provided: food, seeds, agricultural tools, and payments for improvements in infrastructure. Public employees were in charge of coordinating and regulating the developments within official settlements. After all installments were paid for, the settler could claim the property title for his plot. In general, a settlement was emancipated from public administration once all of its plots had been entirely paid for.

The whole process seems to have been kept under serious political accountability. Administrative records and detailed documentation on why some areas were selected were often presented in Annual Provincial Presidential Reports, in Messages to the Legislative Assembly, and in other administrative documents.[4] Over most of the period, state and central governments also paid for or partly subsidized travel expenses, including both international and internal displacements. In addition, through the Immigration Station Houses, local authorities coordinated the job matching between workers and farmers, and the allocation of immigrants to colonial settlements.

In the 1880s, with the imminence of the abolishment of slavery, openness to immigration and the policy of colonial settlements gained renewed strength. The political reforms that came with the 1891 constitution, resulting from the end of the empire and the rise of the republic in 1889, decentralized immigration policy and the management of unclaimed public lands to the states (Iotti, 2001). São Paulo then achieved prominence as main receiving region and immigration to the state reached a massive scale, absorbing a major share of the total inflow of European families arriving in Brazil. Figure 1 presents the cumulative inflow of immigrants into the state from 1872 onwards, and the share of São Paulo in the yearly inflow of immigrants to Brazil. Between 1872 and 1920, roughly 1.8 million immigrants entered São Paulo, corresponding to more than 200 percent of the initial population of the state in 1872 (837,354 inhabitants). This inflow amounted to 53 percent of the total number of immigrants received by Brazil in the period, but reached over 80 percent in specific years at the turn from the 19^{th} to the 20^{th} century. Yet only a small frac-

[4]For example, the 1877 Annual Report of the Secretariat of Agriculture of São Paulo explains in detail the creation of the settlement colonies of São Caetano, São Bernardo, Glória, and Santana; the 1885 Annual Provincial Presidential Report explains the choice of locality for the settlements of Canas and Cascalho; and the 1907 Annual Provincial Presidential Report details the procedures for the creation of the settlements of Nova Europa and Gavião Peixoto.

tion of immigrants ended up in state-sponsored colonies. The vast majority of immigrants arriving at the Immigration Station Houses were rapidly locked into long-term contracts offered by coffee farmers.[5]

The specific sequence of creation of state-sponsored settlements in São Paulo can be outlined as follows. In the 1870s, 4 settlements were created around the state capital: Santana and Glória, nowadays neighborhoods of the capital, and São Bernardo and São Caetano, nowadays independent municipalities in the metropolitan area of the capital. From the mid-1880s to the late 1890s, 13 official settlements were created: 4 in the economically backward Vale do Paraíba, 2 in the central area of the state, 4 in the expanding coffee region of the Northwest, 2 in the Midwest (where the main agricultural products were cotton and sugarcane), and 1 in the coastal Vale do Ribeira. In the early 1900s, 9 additional colonies were founded in still isolated areas of the Northwest.[6] Finally, following a renewed interest in populating lagging regions, the federal government established 2 official settlements in the state: one at the Southwestern border with the state of Paraná and another at the border with the state of Rio de Janeiro (Iotti, 2001).[7]

In total, 28 official settlements were created and consolidated from 1872 to 1920. In terms of the 1920 administrative division of the state, settlements were distributed across 10 percent of the existing municipalities (20 municipalities had received a settlement by 1920). Different motivations behind the colonization policy led to a more or less uniform distribution of official colonies across the various regions in the state. This pattern is discussed in detail in the next session.

[5]Data on number of immigrants arriving during the initial establishment of the settlements does not exist in a systematic way, but historical sources provide information for specific cases. In particular, the 1910s Annual Statistical Reports provide data for 12 settlements created between 1905 and 1911, for which we have the total number of arrivals between 1912 and 1920. The arrival into these 12 settlements during this period represented 1.5 percent of the total number of immigrants arriving in São Paulo. It is difficult though to read this number as more than a simple ballpark figure for the share of immigrants going to the settlements. First, we do not have data on the remaining settlements. Second, entry into these 12 settlements may have been substantially higher in the first years immediately after foundation (1905-1911) when compared to the period for which we have data (1912-1920). And third, entry may have been lower during other periods in the 1872-1920 interval.

[6]The state also created the colony of Conde do Pinhal in the coastal municipality of Ubatuba. However, the first settlers immediately abandoned the place and the project did not take-off. Official justifications for the failure are presented in the 1907 and 1908 Annual Reports of the Secretariat of Agriculture. For this reason, this colony is not included in the analysis.

[7]There were a few more federal settlement attempts in Vale do Ribeira after the 1929 crisis, but these projects were closer to regulated land occupations than to active immigration promoting initiatives (Paiva, 1993). For this reason, these cases are not included in the analysis.

2.2 Characteristics of the State-sponsored Settlements

Table 1 enumerates the state-sponsored settlements included in our analysis, presenting basic information on dates of foundation and – when available – emancipation, initial area, location (in terms of the original and current political division of municipalities), predominant immigrant nationalities, as well as the main sources of information used. The 28 colonies were founded between 1877 and 1911, 17 of them in the end of 19^{th} century and the remaining 11 in the first years of the 20^{th} century. Emancipation came, on average, 10 years after foundation. The size of settlements varied considerably: those created between the late 1870s and the early 1880s had areas mostly between 1,000 and 3,000 hectares, while those created in the 1910s were somewhat larger, roughly between 3,000 and 5,000 hectares. Still, in both periods, a few settlements were created with much larger areas, reaching over 10,000 hectares. Through time, settlements sometimes expanded geographically as the number of settlers increased. Information from the 1910s Annual Statistical Reports of São Paulo shows that, a few years after initial operation, the average population of settlements reached typically from 1,000 to 3,000 inhabitants. The most common nationalities were Italian and German, but Spanish, Portuguese, and Brazilian, among others, were also present (Bassanezi et al., 2008).[8]

Detailed data on immigrants initially arriving during the establishment of the settlements does not exist in a systematic way, but historical sources provide information for some specific cases: the 1888 Annual Statistical Report of São Paulo discusses in detail three settlements created in the 1870s; the 1898 Annual Report of the Secretariat of Agriculture presents data on another 6 settlements, mostly created in the 1880s and 1890s; and the 1910s Annual Statistical Reports mention 12 additional settlements created between 1905 and 1911. We calculate literacy rates for these cases and compare them to the average literacy across municipalities in the state of São Paulo at two points in time: 1890 and 1920 (using data from the respective censuses). Figure 2 presents this information. The average literacy rate across these settlements was 40 percent, as compared to 12 percent across municipalities in the state of São Paulo in 1890, and 23 percent in 1920. The settlements surveyed in the first two reports had average literacy rates of 32 percent, while those mentioned in the reports from the 1910s had literacy rates of 47 percent, indicating some improvement in the pool of immigrants over time. Overall, the population within settlements had substantially higher literacy rates when compared to the remainder of the state.

[8] Brazilians were allowed to purchase plots of land within settlements under certain conditions.

While we are unaware of any study that identified this positive selection statistically and explained its origins, some hypotheses naturally arise. The state-sponsored settlements were populated in two ways. First, Brazilian representatives in European cities could select settlers and negotiate available plots of land. There is evidence that representatives searched for more skilled individuals. This is illustrated by the 1907 Provincial Presidential Report (p.353), where it is recorded that "*the inflows of immigrants that have entered into the state, brought by the Immigration Commissariat in Antwerp and originating from northern Europe, indicate that we managed to find the* ***convenient element*** *for occupying the official colonies*" (translated by the authors, emphasis added). Other examples come from the purposeful selection of Russian families for the settlement of Nova Odessa and Belgium families for the settlement of Rodrigo Silva, reported, respectively, in the 1905 Annual Report of the Secretariat of Agriculture and the 1888 Provincial Presidential Report. Second, available plots of land within colonies were also offered at the Immigration Station Houses. Thus, upon arrival in São Paulo, immigrants sometimes had a choice between official settlements and job opportunities in coffee farms. As the 1906 Provincial Presidential Report (p.43) mentions, "*the immigration trends will increase due to the official propaganda the government is disseminating abroad, and whose efficacy has been already observed in the spontaneous arrival of immigrants searching for the official colonies*" (translated by the authors). Official selection abroad and self-selection upon arrival may have channeled more educated immigrants into official settlements.

Despite the positive selection of immigrants, life conditions in the state-sponsored settlements were not particularly good. Official documents explicitly recommended that settlements be placed in areas close to transportation and with reasonable soil quality.[9] However, the process of choice of location was not homogeneous and, as a result, soil quality varied considerably. On the one hand, some sources document that engineers responsible for the settlements' location weighed in favor of soil fertility when deciding where to establish colonies (see Martins, 1973). On the other hand, selection of location was conditional on alternatives available among unclaimed public lands and unused private or foreclosure lands purchased by the government. These were areas unlikely to be highly productive or particularly suitable for coffee, the main crop at the time. In fact, some official reports mention explicitly the low quality of soils in certain settlements (for example, see discussion in Martins (1973) and the 1899 Annual Report of the Secretariat

[9] Martins (1973) details this point when describing the procedures behind the choice of location for the colony of São Caetano. The recommendation on transportation and soil fertility is explicit in many administrative reports and presidential speeches that justified the choice of location of official settlements. Examples can be found in the 1877 and 1885 Provincial Presidential Reports, and in the 1885 and 1892 Annual Reports of the Secretariat of Agriculture.

of Agriculture).

The 1910s Annual Statistical Reports of São Paulo present detailed information on the economic conditions within 11 settlements, most of them created between 1905 and 1911. We use this information to compare the agricultural productivity within colonies to the average productivity in the state. Productivity is measured as the ratio between the value of production (agricultural, extractive, and animal) and total population. To smooth production and population fluctuations within colonies, we calculate the average of this ratio for the years between 1915 and 1920 (when data are available). We then compare this productivity to the per capita value of agricultural production in the state, reported in the 1920 census. Figure 3 presents these numbers. The annual per capita production within colonies (median around 225 Réis) was systematically lower than the state level production (average of 531 Réis). If anything, this suggests that soil quality may have provided relatively poor levels of subsistence within settlements. But it is difficult to tell whether this lower productivity reflected the difficulties typically faced in maintaining the colonies in the first years immediately after occupation, or indeed lower productivity in the long-run. In any case, there is no evidence suggesting that the quality of the soil in state-sponsored settlements was above that observed elsewhere in the state.

Examples of the challenges faced in the first years after occupation can be found in the 1898 and the 1905 Annual Reports of the Secretariat of Agriculture, where the cases of the colonies of Campos Salles, Nova Odessa, and Jorge Tibiriçá are discussed. Still, relatively few immigrants left the colonies. Figure 4 presents the data available on arrivals and departures for 11 state-sponsored settlements, obtained from the 1910s Annual Statistical Reports of São Paulo. For each settlement, we present cumulative arrivals, departures, and net growth from 1912 to 1918 (as shares of the 1912 population).[10] The 11 settlements grew substantially during this period, on average by 66 percent. In 73 percent of the cases, outflows were considerably low, below 20 percent of the initial population. In the remaining 3 colonies, outflows were substantial, reaching close to or above 100 percent in the cases of Bandeirantes and Monção. But, even in these cases, population grew by, respectively, 29 and 242 percent. So the extreme inflows and outflows in these two colonies seem to be related to the then recent creation of the settlements – dated from 1908 and 1910 – and to the still unstable initial occupation, rather than to a dynamics typical of state-sponsored colonies.

[10]This cumulative number excludes 1917, since the Annual Statistical Report of São Paulo was not available for this year.

2.3 Potential Effects of State-sponsored Settlements

In principle, the establishment of state-sponsored settlements in São Paulo could represent a many-folded intervention. Without further historical information, one could picture such event as affecting human capital levels, culture, land concentration, property rights, and initial agglomeration. All of these factors have received substantial attention from specific literatures within development economics.

The evidence to be presented in the next sections shows that the main distinguishing feature of the settlements was the attraction of a better pool of immigrants, with higher educational levels. The literature on growth and development has repeatedly shown that initial levels of human capital are correlated with subsequent growth, be it across countries or across regions or cities within a country (see, for example, Barro, 1991 and Glaeser et al., 1995). The theoretical link between human capital and growth dates back to the early endogenous growth models, such as Lucas (1988), where there are internal (individual improvements) and external (positive externalities) effects of human capital on productivity. Moretti (2004), for example, argues that knowledge spillovers increase aggregate productivity over and above the direct effect of human capital on individual wages.

Still, it may not be immediately obvious why the initial human capital composition would have persisted over long periods of time. The literature on the history of schooling in Brazil documents that immigrants from more educated groups actively demanded more public investments in schooling, which were typically consolidated through community ethnic schools (Kreutz, 2000 and Maschio, 2005). Based on descriptive historical information, this literature claims that Brazilian states that adopted immigration as a systematic policy to attract labor experienced a process of substantial expansion in schooling. According to Maschio (2005), ethnic schools benefited not only the children of immigrants but also children of Brazilian families who previously inhabited the areas. She documents the specific demands of immigrants and respective government responses, culminating with the creation of community schools, in the case of some state-sponsored colonies in Paraná. It seems natural to think that similar processes could as well have operated whenever more educated immigrants were concentrated in other state-sponsored settlements.

We also consider some alternative hypotheses in our empirical exercise and discuss them briefly here. First, settlements may have changed the distribution of land within settled regions. Land was partitioned into relatively small plots to be allocated to settlers. This process could have led to a reduction in land inequality. It might also have affected property rights, since lands used to establish colonies were sometimes from unclaimed public

areas otherwise potentially subject to dispute through violence. Both land inequality and property rights have been identified as important determinants of long-term development (see, for example, Acemoglu et al., 2005, Galor et al., 2009, and North, 2009). Other relevant aspects have to do with social norms, culture, and religion. As an embodied asset, human capital can be confounded with other embodied characteristics. Settlers may have brought not only productive skills, but also social capital, work ethics, and trust. These factors have a long tradition in the literature on economic development, going back as further as Weber (1930) and including more recent work such as Landes (1999) and Fernandez (2011). Finally, settlements may have represented an initial agglomeration shock, which could then have had long-run impacts through scale effects as suggested by the regional economics literature (see, for example, Ciccone et al., 1996, Combes et al., 2010, and Baldwin and Martin, 2004). This process could have been further intensified by the development of transportation infrastructure, as in Donaldson (2010). We try to assess the relevance of each of these alternative hypotheses while conducting our empirical exercises. But human capital, its persistence through time, and its long-term effect on income per capita and the structure of production remain our main focus throughout the paper.

3 Data

We use three sets of historical data. First, we build a data set of state-sponsored settlements in São Paulo based on historical reports. We identify, locate, and characterize all official colonies created between 1870 and 1920. Second, we merge this information with historical municipality-level data drawn from the population censuses. Third, we complement these data with geographical variables and information on transportation infrastructure. We describe the construction of each data set below.

State-sponsored Settlements

There are scant statistics and literature on the state-sponsored settlements in São Paulo. In order to gather data and qualitative information, we combine the few comprehensive analysis found in Martins (1973), Petrone (1982), Gadelha (1982), and Iotti (2001) with case studies and historical documentation from various sources. In particular, we make use of data from the Annual Reports of the Secretariat of Agriculture of São Paulo, published between 1882 and 1920, and from the Annual Statistical Reports of São Paulo, published in 1888, between 1898 and 1920, and also in 1940. Additional qualitative information on

the procedures for the selection of settlers and settled regions was found in the Public Archives of the State of São Paulo, and in the Provincial Presidential Reports and Messages to the Legislative Assembly of São Paulo.[11]

These sources of information combined provide us with a list containing the location and date of foundation of each settlement, displayed in Table 1. We make use of this list to build our variable of interest, an indicator for settlements created in a given municipality and year. In addition, the Annual Statistical Reports of São Paulo contain data on demographics and socioeconomic conditions for some settlements for selected years. Although the information for earlier settlements is scarce, life conditions within those created in the later waves are relatively well documented. This allows us to examine the economic and demographic patterns within settlements that had not yet been emancipated by the 1910s.

Census Data

The official settlements were created between 1877 and 1911. Given this timing, our analysis is based on municipality level data drawn from the population censuses of 1872, 1920, 1940, and 2000. The 1872 census allows us to examine pre-existing socioeconomic characteristics right before the implementation of the first settlements. The 1920 census is used to build outcome variables for the period just after the establishment of the last colonies. The 1940 and 2000 censuses allow us to analyze, respectively, the mid and long-term impacts of the settlements.

The splitting of municipalities over time generated an increasingly fragmented administrative division. While there were 88 municipalities in 1872, the number of municipalities reached 645 in 2000.[12] Hence we build two different samples for our analysis. The sample used in our main exercises maintains the 1920 border definition and simply aggregates the municipalities in 1940 (270 municipalities) and 2000 (645 municipalities) to the 220 municipalities that existed in 1920.[13] We also match the original data from the 88 municipalities in 1872 to the 202 municipalities that existed in 1920 in order to account for pre-settlement characteristics.[14]

[11] It is worth mentioning that many official settlements became independent municipalities during the first half of the 20^{th} century. Some of them have their histories published online. Official websites of municipalities that have origins closely connected to the settlements also provided useful information.

[12] More precisely, there were 89 municipalities in 1872. However, the independent municipality of Santo Amaro in 1872 is considered part of the capital city, since it was incorporated by its administration in 1935.

[13] A similar procedure has been used in the US (see Hornbeck, 2012).

[14] Under this procedure, one municipality in 1872 may be matched to several municipalities in 1920. Thus, in our empirical analysis, we cluster the standard-errors of the regressions at the level of the original 1872

The second data set merges the layers of data from different census years using the 88 municipalities from the 1872 administrative organization. In this case, the 202 municipalities from later censuses are merged to match the 1872 census boundaries, making the various rounds of data geographically comparable. Although the sample is smaller, this data set allows us to estimate our model using municipal fixed-effects.

We select a comprehensive set of variables from the 1872 census to characterize pre-existing conditions at the municipality level. This includes the shares of slaves, foreigners, literate inhabitants, and children attending school, labor force by sector, and population density. From the 1920 census, we use the following variables: share of literate inhabitants, number of schools and teachers per children aged between 7 and 14, share of foreigners (total and by nationality), share of small farms (up to 100 hectares), average value of farmland per hectare, coffee production in tons, population density, shares of the labor force by sector, and wages in construction and agriculture.

Both the 1940 and the 2000 censuses have educational variables similar to those listed for 1920, as well as a larger set of demographic characteristics including the shares of children attending school, literate individuals aged between 15-19, and, as before, labor force by sector (though the sectors contained in the data vary across years). Since the census micro data are available for 2000, we are able to calculate average years of schooling for different cohorts born during the 20^{th} century. We can also calculate years of schooling for individuals who were born in the municipality and those who are immigrants (in this particular case, meaning not born in the municipality). Finally, we calculate income per capita at the municipality level for 2000 based on micro data from the census.

Data on Geography and Railroads

We use two important sets of additional controls in our regressions. By the late 1870s, many regions in the state of São Paulo were still geographically isolated and had low population density. This was particularly true for the Northwest, which was farthest away from the capital. Thus, geographic characteristics, such as distance to the capital, latitude, and longitude are likely to have been important determinants of the timing of the economic expansion across São Paulo (longitude, in this case, is closely related to distance to the coast). We include these variables as controls in our analysis, as well as average elevation and indicators for the types of soil present in a given municipality (the latter two account for potential heterogeneity in land quality and agricultural productivity).

division.

Geographic characteristics are available from Ipeadata (distance to the capital, latitude, longitude, and elevation) and Embrapa Solos (types of soil). Finally, we include a control for access to transportation infrastructure. Railroads were closely associated with the expansion of the agricultural frontier and urbanization in the state (Dean, 1969). As mentioned before, the location of the state-sponsored settlements was often associated with proximity to transportation. We thus include in our analysis a control for the initial presence of the railroad system (based on historical information on the date of construction of each train station in the state of São Paulo).[15] In the Appendix we provide the definition of each variable included in our analysis.

3.1 Descriptive Statistics

Table 2 presents summary statistics based on the administrative division from 1920, with 202 municipalities. Official settlements were created in 20 of these municipalities, representing 10 percent of the sample. Panel A reports summary statistics for geographic variables. Since the data are restricted to the state of São Paulo, we observe limited variation in latitude, longitude, and elevation. Latosol is the predominant type of soil, present in 55 percent of the municipalities, while argisol is present in 37 percent of the sample. These are the main soil types in the state of São Paulo and are both suitable for coffee production.

Panels B to E present characteristics of the sample by census year. It is worth noting that average literacy rate increased by only 10 percentage points between 1872 and 1920, while population size doubled, the average share of foreigners increased from 1 to 13 percent, and railroad coverage increased from 2 to nearly 70 percent. Average literacy continued to grow in subsequent years, reaching 43 percent in 1940, though the share of children attending school still lagged at 34 percent. Illiteracy was still present up to 2000, but at that point reflecting mostly the lower literacy rates of older cohorts. In 2000, the share of children attending school was 96 percent, while average years of schooling reached 5.5.

Table 3 compares the 1872 characteristics of municipalities that would receive settlements with those of other municipalities in the state. Overall, demographic and geographic variables, as well as the structure of employment, were very similar across the two groups. Differences across municipalities that would receive settlements and other municipalities are typically very small quantitatively and appear as statistically significant for only 1

[15] Data available from www.estacoesferroviarias.com.br.

of the 16 variables considered (and, even in this case, at the 10 percent level). The evidence indicates that the distribution of state-sponsored settlements across regions of São Paulo – in terms of demographic and geographic characteristics, and structure of economic activity – was approximately balanced. This supports the idea that non-settlement municipalities serve as an appropriate comparison group for settlement municipalities.

4 Empirical Strategy

The paper focuses on the effect of state-sponsored settlements on the evolution of educational outcomes over time and on long-term development. Therefore, our main empirical strategy compares schooling and other outcomes across municipalities that received a state-sponsored settlement and municipalities that did not receive a settlement but had similar pre-1920 characteristics. We examine the short, middle, and long-run effects of the settlements using municipality-level data from the 1920, 1940, and 2000 censuses and the following regression:

$$y_i = \alpha + \beta S_i + \Gamma' X_i + \Pi' W_i + \epsilon_i \quad (1)$$

where y_i is the outcome of interest in municipality i measured in 1920 for the short-run effects, 1940 for the middle-run, and 2000 for the long-run; S_i is a variable indicating whether the municipality received a state-sponsored settlement between 1872 and 1920; and ϵ_i is an error term. The regression also includes X_i, a vector of geographic controls (distance to the capital city, latitude, longitude – which is similar to distance to the coast in this context –, altitude, and dummies for the presence of four different types of soil – argisol, cambisol, spondosol, and latosol), and W_i a vector of socioeconomic characteristics measured at the baseline in 1872 (discussed in detail in the presentation of the results). The regression is estimated with the 1920 administrative division, which contains 202 municipalities. Because these municipalities were originally part of the less fragmented 1872 division and we control for pre-determined characteristics based on this administrative organization, we report standard errors that are clustered at the more aggregated 1872 division. We use identical specifications in complementary exercises to analyze: (i) whether there were other dimensions of heterogeneity across municipalities that received and did not receive settlements in 1920, immediately after the initial establishment of settlements; (ii) whether other factors, such as nationality or religion of immigrants, intervened in the relationship between settlements and educational outcomes; and (iii) whether increased schooling over time ended up reflected on a change in the occupational structure of the economy (in the short, medium, and long-run).

One potential concern in relation to the specification from equation (1) is that there might be unobservable characteristics that could have affected the decision to place settlements in specific locations, and that were also correlated with patterns of skilled migration. If this were true, we would observe individuals with more education located in municipalities with settlements, but this correlation would have been driven by unobserved characteristics that might directly affect long-term educational outcomes. One way to control for unobservable factors that are fixed over time is to use panel data and control for municipal fixed-effects. We use an alternative data set based on the 1872 municipality boundaries and pool data from the different censuses to run specifications that estimate year-specific differences between municipalities with and without settlements, relative to the base year of 1872. This specification includes municipality and year fixed-effects and each β is interpreted as the average difference in the outcome variable between settlement and non-settlement municipalities in a given year relative to 1872 (this approach is similar to that adopted by Hornbeck (2010) and Hornbeck and Naidu (2014)). The specific model that we estimate is:

$$y_{it} = \alpha + \beta_t S_i + \gamma_t X_i + \pi_t W_i + \eta_i + \mu_t + \epsilon_{it} \tag{2}$$

where i indexes a municipality and t indicates the census year. The variables η_i and μ_t represent municipality and year fixed-effects and ϵ_{it} is an error term. X_i is a vector of geographic controls and W_i a vector of socioeconomic characteristics measured at the baseline in 1872, both of which are interacted with year effects to allow for differential trends across settlement and non-settlement municipalities. The sample follows the 1872 administrative division and contains 88 municipalities. This exercise is limited to the analysis of literacy rates, the only schooling outcome that can be compared across the different censuses.

Finally, we assess whether the shock to human capital induced by the creation of state-sponsored settlements affected local economic development in the long-run. We use income per capita as a measure of economic development and regress income in 2000 on the settlement indicator from the 1920 municipality boundaries. Assuming that the main channel through which settlements affected long-run development was the persistence of education, we also present estimates of the long-run effect of increases in literacy rates in 1920 using settlements as an instrumental variable. Specifically, we estimate the following equations using 2SLS:

$$\begin{aligned} y_{i2000} &= \alpha + \beta \hat{H}_{i1920} + \Gamma' X_i + \Pi' W_i + \epsilon_i \\ H_{i1920} &= \mu + \pi S_i + \Gamma' X_i + \Pi' W_i + \varepsilon_i \end{aligned}$$

where i stands for municipality, y_{i2000} is the log of income per capita in 2000, H_i is the literacy rate in 1920, S_i is an indicator variable for whether the municipality received a state-sponsored settlement between 1872 and 1920, X_i is the same vector of geographic controls discussed before, and W_i is the vector of socioeconomic characteristics measured at the baseline in 1872.

The identification of the long-run effect rests on two assumptions: first, state-sponsored settlements affected education in the short and medium-run; and, second, the long-run effect of settlements on income worked only through increased schooling in 1920. The strength of the first stage is shown below when we present the short-run effects of the state-sponsored settlements on literacy rates. While we cannot test the exclusion restriction directly, we provide evidence that other mechanisms that could potentially affect long-run development – such as land inequality, culture, and initial shocks to agricultural productivity – do not seem to be affected by the settlements. Also, we provide indicative evidence that the long-term development effect of the 1920 literacy rate worked mostly through a persistent increase in years of schooling.

5 Results

5.1 Baseline Differences in Characteristics

Our empirical strategy compares schooling outcomes and inputs over time across municipalities that received a state-sponsored settlement and municipalities that did not receive a settlement but had similar pre-1920 characteristics. So we start by testing whether initial characteristics in 1872 were indeed similar across municipalities that would receive a settlement and other municipalities. Table 3 already introduced this discussion and showed that municipalities with and without settlements had similar characteristics in 1872, before the establishment of the settlements. The initial literacy rate and the percentage of children attending school were very similar across these localities (19 and 17 percent literacy rates, and 15 and 16 percent of school-aged children attending school), as well as the population density, the percentage of foreigners, and the percentage of slaves. Geographic characteristics – such as distance to the capital, latitude, and altitude – and the types of economic activities and occupations were also similar. Two important differences that can be observed in Table 3, although not statistically significant, are the presences of the soil latosol and of a railway.

Table 4 reproduces the comparison from Table 3 in a multivariate setting. It reports the results of regressions estimating the ”determinants of settlement presence," where the dummy for settlement is regressed on all 1872 characteristics from Table 3. Columns 1 to 3 include the sets of variables related to geography, socioeconomic characteristics, and occupations separately, each at a time, in the regression. In general, pre-determined characteristics have limited, if any, predictive power for settlement location. Once we add all variables simultaneously (column 4), only the presence of a railway is a statistically significant correlate of the future establishment of a state-sponsored settlement (and only at the 10 percent level). Still, since some of these 1872 characteristics could be correlated with future schooling outcomes, we control for all of them in our benchmark specification.

5.2 The Educational Effects of State-sponsored Settlements

Table 5 reports the results from estimating equation (1) using the literacy rate in 1920 as the dependent variable. In column 1 we report the bivariate relationship between literacy rate and the settlement indicator, while in column 2 we control for geographic characteristics. The coefficients on the settlement variable in columns 1 and 2 are very similar and show that 1920 literacy rates were higher in municipalities that received a state-sponsored settlement. The estimate represents a 10 percentage point difference in the literacy rate between settlement and non-settlement municipalities (or a 33 percent difference, given the average of 0.30 in 1920). In column 3, we control for the 1872 socioeconomic characteristics. The estimated coefficient on the settlements drops to 8 percentage points, but remains highly significant. Thus, while the share of slaves and the presence of a railway in 1872 were correlated with literacy rates in 1920, our results are robust to controlling for these initial differences across municipalities. The relationship we uncover between settlements and literacy rates is not driven by geographic characteristics or previously existing socioeconomic conditions associated with these areas.

As a robustness test of the previous results, columns 4 to 6 show the same set of regressions using a continuous variable as an indicator of the intensity of treatment: the population of the initial settlement relative to the existing population in the municipality in 1872. This is a very noisy statistic for the size of settlements since initial population is not always measured at the same moment. In addition, we do not have this information for some of the settlements and therefore loose some observations under this specification. Despite these limitations, all specifications show that an increase in the relative size of the settlement is associated with an increase in the literacy rate in the municipality in 1920. The

magnitude of the coefficient declines when we control for pre-settlement characteristics in 1872 (column 6), but the results are still significant at the 10 percent level.

We next document the persistence over time of the initial shock to human capital. Table 6 shows results from estimating equation (1) for literacy rate in 1940 and literacy rate and years of schooling in 2000 (years of schooling are not available in the 1920 and 1940 censuses). We use two measures of literacy rate, one for the entire population (above age 5) and another for those aged 15 to 19. The second measure captures the effect on individuals who were not directly affected by the policies implemented before 1920. Column 1 shows that, in 1940, municipalities that had a settlement established between 1872 and 1920 had literacy rates 10 percentage points higher when compared to municipalities that did not have a settlement (this represents a 23 percent difference based on the average literacy rate of 0.43 in 1940). This effect is 13 percentage points if we focus on individuals aged 15-19 (column 2), suggesting that the effect is not driven solely by the older cohorts who arrived with the establishment of the settlements or soon thereafter. The fact that individuals aged 15-19 were more educated in 1940 suggests persistence of policies at the municipality level or intergenerational transmission of human capital.

In columns 3 to 5, we show the long-run effects of state-sponsored settlements on education using data from the 2000 census. By the year of 2000, roughly 90 percent of São Paulo's population was literate, with relatively little variation across municipalities. Because literacy at that point was close to universal for younger cohorts, we find a small effect of settlements on overall literacy (column 3) and no effect for individuals aged 15-19 (column 4). So we also look at the effect of the settlements on years of schooling in 2000. We find that municipalities that received a state-sponsored settlement before 1920 had 0.52 more year of schooling when compared to other municipalities (column 5). Given the average years of schooling of 5.5 in 2000, this represents a 10 percent increase in long-run educational attainment.

In Table 7, we repeat the same regression for years of schooling from column 5 in Table 6, but restrict the sample to individuals born in the respective municipality and also decompose the result by cohort of birth. In column 1, where we consider all cohorts simultaneously, the coefficient is substantially larger than that from Table 5, indicating that most of the quantitative effect of settlements on schooling is indeed driven by individuals who were educated in the municipalities. Individuals born in settlement municipalities have, on average, 0.88 more year of schooling when compared to individuals born in non-settlement municipalities. Looking at birth cohorts from 1920-29 to 1960-69 (columns 2 to 6), which were likely to have already completed their schooling by 2000, we see a positive

and statistically significant effect that increases over time. Individuals born in settlement municipalities between 1960 and 1969 had on average 1 more year of schooling when compared to individuals born elsewhere. The analogous number for the cohort born between 1920 and 1929, soon after the establishment of the settlements, was 0.66 year of schooling. But comparing the point estimates with the mean of the dependent variable – displayed in the table – one can see that the proportional effect of settlements on years of schooling declined over time, from 24 percent for the 1920-29 cohort to 15 percent for the 1960-69 cohort.

One potential drawback of the estimates presented above is that there may be unobservable characteristics that affect the decision of location of settlements and that are also correlated with patterns of skilled migration. In Table 8, we present results from an alternative specification that controls for unobservable factors that are fixed over time. We construct a panel of municipalities covering 1872, 1920, and 1940, use municipal fixed-effects, and allow coefficients capturing the effect of settlements to vary over time. The literacy rate, which is the only educational outcome that can be compared across censuses, is used as dependent variable. We restrict the analysis to 1872, 1920, and 1940, since by 2000 differences across municipalities were not anymore reflected on literacy but on completed years of schooling (see Table 6). The coefficients displayed in column 1 suggest that in 1920 there was a 4.4 percentage point increase in literacy rates relative to the base year of 1872 in localities with settlements. This effect increased to 9.7 percentage points in 1940. To assess whether the results are robust to trends in initial characteristics that could be correlated with the choice of location of the settlements, column 2 presents results with year effects interacted with baseline geographic characteristics, and column 3 shows the results including year effects interacted with both geographic and demographic characteristics. While the effect in 1920 remains similar and close to 4 percentage points, the effect in 1940 decreases a bit, reaching 6.8 percentage points in the more saturated specification in column 3, but remains statistically significant. Overall, these results are consistent with the cross-section estimates presented in Tables 5 and 6 showing that state-sponsored settlements led to a significant increase in literacy rates in the short and medium run.

Tables 5 to 8 show that the settlements represented a positive shock to educational outcomes that lasted until 2000. We next document how the initial shock to human capital persisted through time. Table 9 presents a series of cross-section specifications for 1920, 1940, and 2000 where we regress three different measures of schooling inputs on the settlement indicator. Because the specific data collected by the census vary over time, the three measures are not available for all years. In column 1, we show that settlement mu-

nicipalities had more schools per child in 1920 than non-settlement municipalities. The effect on number of schools, however, was only significant in the short-run. Column 2, on the other hand, shows that settlement municipalities had more teachers per child in 1920 and that this effect persisted until 2000 (coefficient 11.2 in column 2, Panel C). Finally, in column 3, we show that, consistently with the positive shock to the number of schools and teachers, a larger percentage of children in settlement municipalities were enrolled in school. Similarly to literacy, we find a stronger effect on the percentage of children enrolled in school in 1940, since by 2000 basic education was close to universal in the state of São Paulo (data on school enrollment were not available from the 1920 census).

5.3 Alternative Hypotheses

The previous results show a persistent increase in educational levels in municipalities that received state-sponsored settlements. We next investigate whether this increase was associated with the initial composition of immigrants or with other characteristics of settlements that might have affected long-term development and, indirectly, educational outcomes. Table 10 reports a series of results from estimating equation (1) using a large set of 1920 dependent variables that, in principle, could have been affected by the settlements. All regressions use the same specification from Table 5 and control for geographic and pre-1872 socioeconomic characteristics.

Columns 1 and 2 report the initial effect of the settlements on the presence of foreigners and the percentage of literate foreigners. Municipalities with settlements had 2.5 percentage points more foreigners when compared to municipalities without settlements (average of 13 percent in 1920). Conditioning on the presence of foreigners, column 2 shows that the settlements increased the percentage of literate foreigners in almost 5 percentage points (an increase of 11 percent over a base of 0.44). Thus settlements brought slightly more foreigners to the respective areas, but particularly more educated foreigners.[16] In columns 3 to 8, we examine other potential effects of the settlements, associated with alternative theoretical channels that have been highlighted in the literature (see Section 2.3). We find no evidence that state-sponsored settlements affected population density, the percentage of small farms (proxy for land inequality), or agricultural productivity in

[16]These results should be interpreted with caution. In the 1920 census, the question that allows identification of foreigners is related to nationality, not locality of birth as in most current censuses. So it is possible that foreign born individuals who by 1920 had acquired Brazilian citizenship would be identified as Brazilian nationals rather than foreigners. Without specific assumptions about the identification of these individuals – and its difference across settlements and non-settlement municipalities – it is difficult to tell the type of bias that would be implied by this measurement error.

the short-run (proxied by coffee productivity). We also find no significant effect on land prices, which could capture investments in rural infrastructure, nor on wages of either construction or agricultural workers. Despite the fact that settlements attracted a selected pool of immigrants, this seems to have had no immediate effect on any form of income or on the value of land. These results are consistent with the idea that state-sponsored settlements brought more educated foreigners to localities where the higher level of human capital had no immediate economic return, at a moment when coffee production was still the dominant and most profitable economic activity (Dean, 1969).

Another potential limitation in our results is that settlements may be proxying for certain groups of immigrants, and it may be the culture or social norms associated with these groups that affect schooling outcomes and development, rather than the other way around. Table 11 assesses this possibility by introducing controls related to the nationality of immigrants and religion in our benchmark specification (column 3 from Table 5). In column 1, we control for the 1920 percentage of foreigners by major national groups: German, Spaniard, Italian, Portuguese, and Japanese. Some of these nationalities appear as positively and significantly associated with higher literacy rates, but the introduction of these additional controls does not interfere with the estimated coefficient on state-sponsored settlements. In column 2, we control for the 1920 percentage of Catholic individuals. Again the coefficient of interest is unaffected, despite the fact that the percentage of Catholics appears as positively associated with literacy rates. Finally, column 3 controls simultaneously for nationality and religion. The estimated effect of settlements on 1920 literacy rates remains strongly significant and, in fact, increases by 12 percent when compared to Table 5. Columns 4 and 5 repeat the same exercise from column 3 using as dependent variables, respectively, the literacy rate in 1940 and average years of schooling in 2000. Results remain almost identical to those from Table 6. In short, Table 11 suggests that our previous results are not driven by cultural factors associated with the nationality or religion of immigrants.

5.4 Human Capital, Long-term Development, and Structural Transformation

In the previous sections, we showed how the establishment of state-sponsored settlements improved the educational outcomes of the local population, both in the short and in the long-run. We next examine whether the establishment of settlements, and the persistent increase in human capital, also induced a differential development path. We begin by

analyzing the effect of settlements on income per capita in 2000. We then provide evidence that settlements induced a structural transformation in economic activity and, in the long-run, attracted more educated immigrants from inside Brazil. The long-term growth in income per capita is likely to have been a consequence of both the increase in the education level of the native population and the attraction of more educated domestic immigrants.

Table 12 presents estimates of the effect of state-sponsored settlements on income per capita in 2000. Column 1 shows that income per capita in 2000 was 15 percent higher in municipalities that received a state-sponsored settlement when compared to other municipalities. Our hypothesis is that a significant part of this difference is driven by increases in human capital dating back to the settlements. So in column 2 we regress current income per capita on the literacy rate in 1920 and find a strong correlation between the two variables. But there are unobserved factors at the municipality level correlated with both literacy and long-term development that preclude a causal interpretation of this result. If, as argued in the previous sections, there is indeed an exogenous dimension of variation in education that was driven by the establishment of settlements, and settlements did not have any other direct effect on long-term growth, then we can use the presence of settlements as an instrument for literacy rates in 1920. We present the result from this exercise in column 3. The effect is consistent with column 2, where we showed a strong association between literacy rates in 1920 and income per capita in 2000, with the IV coefficient being somewhat larger than the respective OLS estimate. Still, standard errors are such that we cannot reject that the OLS and IV coefficients are identical. The coefficient from column 3 suggests that a 10 percentage point increase in the 1920 literacy rate would be associated with an income per capita in 2000 approximately 19 percent higher. Finally, in column 4, we control for years of schooling in 2000 and the coefficient on the 1920 literacy rate (instrumented with the presence of settlements) drops from 1.9 to 0.27 and ceases to be statistically significant. This pattern suggests that most of the association between settlements and development was indeed driven by improvements in educational outcomes.

In Table 13, we assess specifically how the increased schooling translated into higher income per capita in the long-run. In particular, we analyze whether the higher level of education allowed individuals to explore more skill-intensive economic activities. Panel A shows that, already in 1920, municipalities with settlements had lower employment shares in agriculture and higher shares in manufacturing and services. At this early point, coffee production was still the leading and more profitable economic activity in the state of São Paulo and accounted for close to 70 percent of total Brazilian exports (Abreu and Bevilaqua, 1996). So, as Table 10 indicated, this initial shift in employment towards man-

ufacturing and services was not reflected on higher urbanization, higher wages, or higher value of land. It was simply a direct results of the different composition of immigrants arriving at these locations. Manufacturing was then still incipient and marginal within the local economy, mostly supplying basic goods demanded by the coffee economy, such as bricks, tiles, cement, glass, and plumbing (Dean, 1969). This process of structural transformation deepened and started becoming indeed an economic advantage between 1920 and 1940, when industrialization took off in the state and the capital previously accumulated with coffee production was invested in manufacturing (Dean, 1969). As Panel B shows, by 1940 there was a 9 percentage point difference in the share of employment in manufacturing – and a 7 percentage point difference in services – across municipalities that had received a settlement before 1920 and other municipalities. Finally, while by 2000 the whole state of São Paulo had gone through a process of structural transformation, municipalities that had received a settlement before 1920 still had 5 percentage points more employment in services – and less in agriculture – when compared to other areas in the state. Interestingly, the higher share of employment in services in 2000 was mainly concentrated in high skill sectors, such as banking and finance, education, health, and retail (see Appendix Table A2).

Once the process of structural transformation and income growth took place, one should expect settlement municipalities to become poles of attraction for immigrants from inside Brazil. If this was indeed the case, long-term growth would be partly attributable as well to agglomeration effects.[17] In Table 14, we look at the 2000 census and repeat the same cohort based exercise from Table 7, but looking at immigrants (now meaning individuals who were not born in the municipality where they lived in 2000, rather than foreign born). We estimate the effect of the settlements on years of schooling across different cohorts for individuals who were not born in the settlement municipalities where they live. Two hypotheses underlie this exercise and our interpretation of its results: (i) individuals not born in the settlement municipality where they live are more likely to have been born in municipalities without settlements; and (ii) older cohorts of individuals not born in the municipality where they leave are more likely to have migrated to this municipality longer ago in the past (as compared to younger cohorts of immigrants). Looking across all cohorts, in column 1, we see that immigrants to settlement municipalities have, on average, 0.54 more year of schooling when compared to immigrants to other municipalities. Columns 2 to 6 present the effect of the settlements on the characteristics of immigrants

[17] For instance, Ciccone et al. (1996), Ciccone (2002), Combes et al. (2010) show that employment density positively affects labor productivity at the regional level in the US, Europe and France, respectively. See also Rosenthal and Strange (2004) and Duranton and Puga (2004) for reviews of the empirical evidence and a discussion of the likely mechanisms.

across cohorts. Older cohorts of immigrants, born between 1920 and 1939, were not significantly different between settlement and non-settlement municipalities. But starting with the cohorts born between 1940 and 1949, immigrants to settlement municipalities are significantly more educated than immigrants to other locations. This effect increases in magnitude for younger cohorts and continues to be strongly significant. Though substantially smaller than the effect detected for natives in Table 7, the relative difference between natives and immigrants in settlement municipalities shrinks in magnitude for the later cohorts, as can be seen by comparing the analogous columns across Tables 7 and 14. In addition, contrary to the results from Table 7, the relative size of the coefficient (as compared to the mean of the dependent variable) does not decline with time. This pattern suggests that the initial shock to schooling represented by state-sponsored settlements also attracted, over time, more educated individuals from within Brazil. We interpret this evidence to be consistent with human capital agglomeration effects.

6 Concluding Remarks

This paper looks at a unique episode where, in the midst of a mass inflow of European immigrants to the Brazilian state of São Paulo, those relatively more educated were channeled into specific areas. In the end of the 19^{th} and beginning of the 20^{th} century, public authorities established a number of official settlement colonies throughout São Paulo. Simultaneously, European immigrants were also arriving in virtually every other area of the state. We exploit a key feature of this episode to document the persistence of human capital through time and to isolate its role as a determinant of long-term development: state-sponsored settlements were not too different from other areas in the state, but for the fact that they received immigrants with higher educational levels. This natural experiment allows us to disentangle human capital from other factors typically highlighted in the literature on the impacts of European colonization and provides evidence on the effect of the composition of immigrants on long-term development.

We follow more than 120 years of Brazilian history to characterize the initial conditions in the state of São Paulo in the second half of the 19^{th} century and the dynamics of the impact of state-sponsored settlements through time. We use data from the 1872, 1920, 1940, and 2000 Brazilian censuses, as well as a vast array of auxiliary information from various historical sources. The results show that, in 1872, before the establishment of any settlement, areas that were to receive settlements in the future were very similar to other areas in the state. In 1920, immediately after the initial establishment of the settlements,

these areas had populations with literacy rates 8 percentage points higher than elsewhere, despite having only marginally higher shares of immigrants. At that moment, apart from the educational composition of the population, areas that had received settlements were not too different from other areas. Still, in 1940 and 2000, long after the settlements had officially ceased to exist, their populations continued to be more educated. In 2000, areas corresponding to previous settlements had income per capita 15 percent higher than the remainder of the state. We also show that, throughout the 20^{th} century, enrollment rates and number of teachers per school aged children were higher in locations associated with settlements. In these same locations over this period, employment shifted away from agriculture and towards manufacturing and services . None of these differences in the use of educational inputs and in the structure of employment were present before the settlements were established. The evidence suggests that local communities that developed from historical settlements demanded higher educational investments and, through time, shifted economic activity to skill-intensive sectors.

Our results are consistent with the idea that state-sponsored settlements brought more educated foreigners to localities where the higher level of human capital had no immediate economic return, since coffee production – intensive in unskilled agricultural workers – was still the dominant and most profitable economic activity. But as soon as industrialization took hold, the higher level of human capital started representing an economic advantage that ended up materializing in changes in the productive structure of the economy and increased income per capita in the long-run.

Although this paper is based on a specific historical episode, our findings provide evidence for broader interpretations that connect European colonialism to long-term development specifically through the human capital channel. We contribute to this literature by disentangling human capital from other factors that typically accompanied most historical experiences of European immigration.

References

Abreu, M. and Bevilaqua, A. (1996). Brazil as an export economy, 1880-1930. PUC-Rio, Working Paper 363.

Acemoglu, D. and Dell, M. (2010). Productivity Differences Between and Within Countries. *American Economic Journal: Macroeconomics*, 2(1):169–188.

Acemoglu, D., Gallego, F. A., and Robinson, J. A. (2014). Institutions, Human Capital, and Development. *Annual Review of Economics*, 6(1):875–912.

Acemoglu, D., Johnson, S., and Robinson, J. (2005). *Institutions as a Fundamental Cause of Development*, volume In: Aghion, P. and Durlauf, S. Handbook of Economic Growth, Vol. 1A. North-Holland.

Acemoglu, D., Johnson, S., and Robinson, J. A. (2002). Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution. *The Quarterly Journal of Economics*, 117(4):1231–1294.

Baldwin, R. and Martin, P. (2004). *Agglomeration and Regional Growth*, volume In: Henderson, V. and Thisse, J.F., Handbook of Regional and Urban Economics, V. 4, pages 2671–2711. Elsevier.

Bandiera, O., Mohnen, M., Rasul, I., and Viarengo, M. (2015). Nation-Building Through Compulsory Schooling During the Age of Mass Migration. *Mimeo, LSE*.

Barro, R. (1991). Economic Growth in a Cross Section of Countries. *The Quarterly Journal of Economics*, 106(2):407–443.

Bassanezi, M., Scott, A., Bacellar, C., and Truzzi, O. (2008). *Atlas da Imigração Internacional em São Paulo - 1850-1950*. São Paulo: Editora UNESP.

Becker, S. O., Hornung, E., and Woessmann, L. (2011). Education and catch-up in the industrial revolution. *American Economic Journal: Macroeconomics*, 3(3):92–126.

Ciccone, A. (2002). Agglomeration effects in Europe. *European Economic Review*, 46(2):213–227.

Ciccone, A., Hall, R. E., et al. (1996). Productivity and the Density of Economic Activity. *American Economic Review*, 86(1):54–70.

Ciccone, A. and Papaioannou, E. (2009). Human Capital, the Structure of Production, and Growth. *Review of Economics and Statistics*, 91(1):66–82.

Combes, P.-P., Duranton, G., Gobillon, L., and Roux, S. (2010). *Estimating Agglomeration Economies with History, Geology, and Worker Effects*, chapter In: Glaeser, E. (Ed). Agglomeration Economics, pages 15–66. University of Chicago Press.

de Carvalho Filho, I. and Colistete, R. (2010). Education Performance: Was It All Determined 100 Years Ago? Evidence From São Paulo, Brazil.

de Carvalho Filho, I. and Monasterio, L. (2011). Immigration and the Origins of Regional Inequality: Government-Sponsored European Migration to Southern Brazil Before World War I. *Regional Science and Urban Economics*. Forthcoming.

Dean, W. (1969). *The Industrialization of São Paulo, 1880-1945*. University of Texas Press.

Donaldson, D. (2010). Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. *NBER Working Paper*, 16487.

Droller, F. (2013). Migration, population composition and long-run economic development: Evidence from settlements in the pampas. *Mimeo*.

Duranton, G. and Puga, D. (2004). *Micro-foundations of Urban Agglomeration Economies*, chapter In: Henderson, J. and Thisse, J. (Eds). Handbook of Regional and Urban Economics V4, pages 2063–2117. Elsevier.

Easterly, W. and Levine, R. (2012). The european origins of economic development. Working Paper 18162, National Bureau of Economic Research.

Engerman, S. L., Mariscal, E., and Sokoloff, K. L. (2012). *The Evolution of Schooling: 1800–1925*, chapter In: Engerman, Stanley L and Sokoloff, Kenneth L (Eds). Economic Development in the Americas since 1500: Endowments and Institutions. Cambridge University Press.

Fernandez, R. (2011). *Does Culture Matter?*, volume In: Benhabib, J., Jackson, M. and A. Bisin. Handbook of Social Economics, Vol. 1A. North-Holland.

Fourie, J. and von Fintel, D. (2014). Settler skills and colonial development: the huguenot wine-makers in eighteenth-century dutch south africa. *The Economic History Review*, 67(4):932–963.

Furtado, C. (1989). *Formação Econômica do Brasil*. São Paulo: Editora Nacional.

Gadelha, R. (1982). *Os Núcleos Coloniais e o Processo de Acumulação Cafeeira (1850-1920): Contribuições ao Estudo da Colonização de São Paulo*. PhD thesis, FFLCH-USP, São Paulo.

Galor, O., Moav, O., and Vollrath, D. (2009). Inequality in Landownership, the Emergence

of Human-Capital Promoting Institutions, and the Great Divergence. *The Review of Economic Studies*, 76(1):143–179.

Gennaioli, N., Porta, R. L., de Silanes, F. L., and Shleifer, A. (2013). Human Capital and Regional Development. *The Quarterly Journal of Economics*, 128(1):105–164.

Glaeser, E., La Porta, R., Lopez-de Silanes, F., and Shleifer, A. (2004). Do Institutions Cause Growth? *Journal of Economic Growth*, 9(3):271–303.

Glaeser, E., Scheinkman, J., and Shleifer, A. (1995). Economic Growth in a Cross-Section of Cities. *Journal of Monetary Economics*, 36(1):117–143.

Hall, R. E. and Jones, C. I. (1999). Why Do Some Countries Produce So Much More Output Per Worker Than Others? *The Quarterly Journal of Economics*, 114(1):83–116.

Hornbeck, R. (2010). Barbed Wire: Property Rights and Agricultural Development. *The Quarterly Journal of Economics*, 125(2):767–810.

Hornbeck, R. (2012). The Enduring Impact of the American Dust Bowl: Short-and Long-Run Adjustments to Environmental Catastrophe. *The American Economic Review*, 102(4):1477–1507.

Hornbeck, R. and Naidu, S. (2014). When the Levee Breaks: Black Migration and Economic Development in the American South. *The American Economic Review*, 104(3):963–990.

Hornung, E. (2014). Immigration and the diffusion of technology: The huguenot diaspora in prussia. *The American Economic Review*, 104(1):84–122.

Huillery, E. (2009). History matters: The long-term impact of colonial public investments in french west africa. *American Economic Journal: Applied Economics*, 1(2):176–215.

Iotti, L. H. (2001). *Imigração e Colonização: Legislação de 1747-1915*. Caxias do Sul: EDUCS.

Iranzo, S. and Peri, G. (2009). Schooling Externalities, Technology, and Productivity: Theory and Evidence from US States. *The Review of Economics and Statistics*, 91(2):420–431.

Kreutz, L. (2000). Escolas comunitarias de imigrantes no brasil: instancias de coordenacao e estruturas de apoio. *Revista Brasileira de Educacao*, 15:159–176.

Landes, D. (1999). *The Wealth and Poverty of Nations: Why Some Are So Rich and Some So Poor*. WW Norton & Company.

Lucas, R. (1988). On the Mechanics of Economic Development. *Journal of Monetary Economics*, 22(1):3–42.

Martins, J. S. (1973). *A Imigração e a Crise do Brasil Agrario*. São Paulo: Pioneira.

Maschio, E. (2005). A experiencia escolar entre imigrantes italianos no parana. Universidade Federal do Parana.

Moretti, E. (2004). *Human Capital Externalities in Cities*, volume In: Henderson, J.V. and Thisse, J.F. Handbook of Regional and Urban Economics, V.4. Amsterdam: Elsevier.

North, D. (2009). *Institutions, Institutional Change and Economic Performance*. Cambridge University Press.

North, D. C. and Thomas, R. P. (1973). The Rise ofthe Western World. *A New Economic History, Cambridge*.

Nunn, N. (2011). Christians in colonial africa. Harvard University.

Paiva, O. C. (1993). Intervenção Governamental e Reoardenação Fundiária. Master's thesis, Unicamp.

Peri, G. (2012). The Effect of Immigration on Productivity: Evidence from US States. *Review of Economics and Statistics*, 94(1):348–358.

Petrone, M. (1982). *O Imigrante e a Pequena Propriedade*. São Paulo: Brasiliense.

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

Rosenthal, S. S. and Strange, W. C. (2004). *Evidence on the Nature and Sources of Agglomeration Economies*, chapter In: Henderson, J. and Thisse, J. (Eds). Handbook of Regional and Urban Economics V4, pages 2119–2171. Elsevier.

Shapiro, J. (2006). Smart Cities: Quality of Life, Productivity, and the Growth Effects of Human Capital. *The Review of Economics and Statistics*, 88(2):324–335.

Siriani, S. (2005). Os Descaminhos da Imigração Alemã para São Paulo no Seculo XIX - Aspectos Políticos. *Almanack Braziliense*, 2:91–100.

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

Summerhill, W. (2003). *Order Against Progress: Government, Foreign Investment, And Railroads in Brazil, 1854-1913*. Stanford University Press.

Summerhill, W. (2010). Colonial Institutions, Slavery, Inequality, and Development: Evidence from São Paulo, Brazil.

Valencia Caicedo, F. (2014). The mission: Economic persistence, human capital transmission and culture in south america. *Mimeo*.

Wantchekon, L., Klasnjaz, M., and Novta, N. (2015). Education and Human Capital Externalities: Evidence from Colonial Benin. *Quarterly Journal of Economics*, 130(2):703–757.

Weber, M. (1930). *The Protestant Ethic and the Spirit of Capitalism*. London: Allen & Unwin.

Appendix

Merging Municipal Boundaries across Censuses

Matching the 1872-1920 censuses. Each municipality of the 1920 administrative division is matched to the original municipality it belonged to in the 1872 administrative division. For instance, the municipality of the capital city as existed in 1872 was partitioned into 6 new municipalities between 1872 and 1920. Thus, all these 6 municipalities are linked to the same pre-existing conditions of the original municipality as reported in the 1872 census.

Merging the 1920 layer of municipal data with the 1940 and 2000 censuses. To match the 1920 layer to the variables from the 1940 and 2000 censuses, we proceed as follows. First, the municipalities as they existed in the 1920 division are used as minimum comparable areas in order to match the merged 1920-1872 variables with the variables from the more fragmented divisions of 1940 (270 municipalities) and 2000 (645 municipalities). Second, if a given municipality was partitioned into two or more units between 1920 and 1940 (2000), the variables from the 1940 (2000) census associated with these new units are then aggregated in order to rebuild the original municipality. The aggregation of a given variable of the 1940 (2000) census into the original 1920 minimum comparable area follows two rules: (i) the variable is aggregated as an weighted average by using the population of the administrative units of 1940 (2000) as weights. For instance, suppose municipality X in 1920 (or the minimum comparable area X) is partitioned into X1 and X2 in 1940. Thus, the share of literate individuals for this area X in 1940 is the sum of the literate individuals in X1 and X2 over the sum of the 1940 populations of X1 and X2. This first rule was applied to those municipalities created from the partition of only one original municipality, and corresponds to more than 95% of the cases. The second rule, which was applied to 4 cases, relates to a very specific way a municipality can be partitioned. This specific case arises if a given municipality in 1940 (2000) was originally made by parts of two or more municipalities in the past. In this case, we just follow the capital district of the municipality. For example, suppose municipality Y in 1940 was originally made by parts of municipalities Z and K in the past. In this case, we ask whether the capital-district of Y was originally located in the municipality Z or K. If it were in Z, we thus consider Y to be part of Z in the past and incorporate it into Z in the subsequent years. Thus, for example, the share of literate individuals for the area Z in 1940 is the sum of the literate individuals in Z and Y in 1940 over the sum of the populations of Z and Y in 1940. This second rule is based on

the facts that the capital district of a given municipality is its most populated district, and that any weighting that would consider district populations is restricted by the fact that all the data in the historical censuses are available only at the municipality level (and not at the less fragmented district level).[18]

Merging the 1872-1920-1940 censuses for the panel analysis. The panel analysis requires the baseline sample to be the more aggregate one, the 1872 administrative division. We proceed analogously to the above cases. Information from the 1920 (1940) census was mapped into the smaller set of the 1872 division. For example, in order to make the share of literate individuals for the capital city compatible across time, we first calculated the share of literate individuals for the capital city as it was recorder in the 1872 Census. The capital city was partitioned into 6 new municipalities between 1872 and 1920. Thus, in order to calculate the 1920 share of literate individuals for the capital city as it was in 1872, we divide the (i) sum of the literate individuals recorded in the 1920 Census for those 6 municipalities, by the (ii) sum of the total individuals recorded in those same 6 municipalities. The four specific cases mentioned above were aggregated accordingly.

The case of municipalities merged with others. Some municipalities have been incorporated into others over time. Santo Amaro was an independent municipality of the state of São Paulo until 1935, when it was definitively incorporated by law as a district of the capital city of São Paulo. For this specific reason, and because we are not able to follow data at the district level, we incorporate Santo Amaro into the capital city of São Paulo since the first 1872 census. There are also cases of temporary merges. In the late 1930s, some independent municipalities (eg. municipality y) were made districts of larger ones (eg. municipality X), and made again independent a few years later. We thus have historical data for the independent municipalities from all the censuses included in our analysis except from the 1940 one. In these cases, for those municipalities that were incorporated into others (eg. for y, that was incorporated into X), we imputed the value of the variables from the 1940 census calculated for the municipalities that absorbed them (eg. we assigned for y in 1940 the same literacy rate calculated for X in 1940). In doing so, we keep intact our benchmark sample based on the 1920 administrative division. None of these municipalities received a state-sponsored settlement.[19]

[18]This second rule was applied to the municipalities of Pereira Barreto (located at the far-western region of the state, near the frontier with the state of Mato Grosso do Sul), Marília, Garça and Tupã (all located near the frontier with the state of Paraná).

[19]The municipalities that were temporarily incorporated by others are Araçariguama, Bom Sucesso, Buquira, Conceição do Monte Alegre, Espírito Sto do Turvo, Igaratá, Lagoinha, Platina, Ribeirão Branco, Sta Cruz da Conceição, and Sta Isabel.

Definition of Variables

Table A1 presents the definition of the variables included in our analysis and their respective sources. We provide below some additional details.

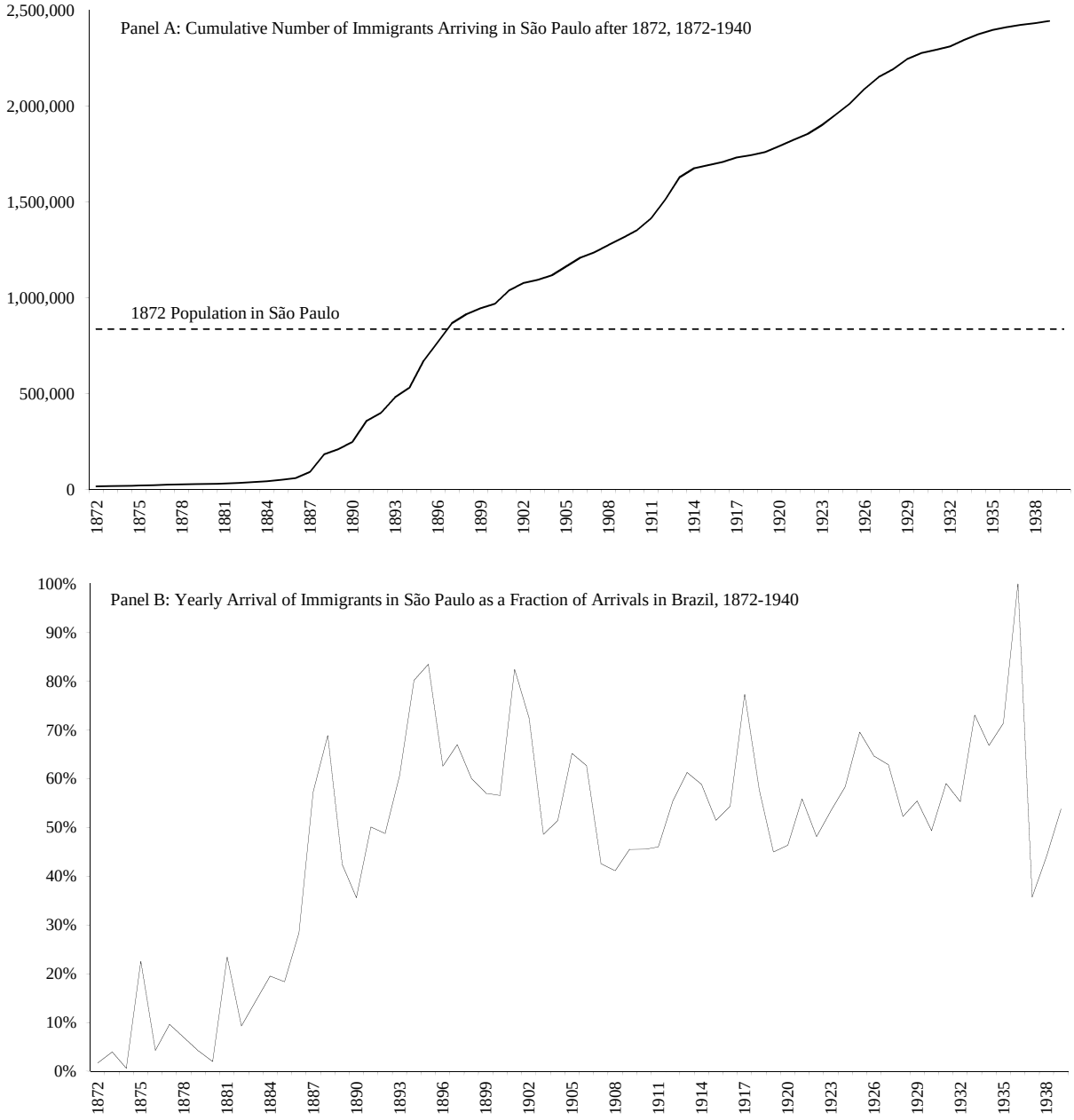
Demographic variables. All demographic variables are constructed directly from census data at the (i) *parochial* level, as originally available from the 1872 census, and then aggregated at the municipality level; (ii) at the municipality level, as originally available from the 1920 and the 1940 census; and (iii) at the individual and household level, as originally available from the 2000 census. Variables for the 1872 and 1920 layers of data are directly constructed from the respective census. For example, the share of literate individuals aged 6+ was simply built as the total number of literate individuals aged 6+ over total population aged 6+. Variables for the 1940 and 2000 layers of data were also constructed directly from the respective census, but were made compatible with the 1920 layer based on the merging rules described above. For example, suppose that a given municipality X in 1920 was partitioned into X1 and X2 just before 1940. The share of literate individuals for the area X in 1940 is simply the sum of the literate individuals in X1 and X2 over the sum of the 1940 populations of X1 and X2 - all the data needed the build the share of literate individuals in X in 1940 are therefore available in the 1940 census for the municipalities X1 and X2. Regarding more specifically the income and education variables for the 2000 layer of data: (i) per capita income in 2000 of a given unit of the 1920 municipal division level was calculated as the sum of all incomes over total population in the municipalities that make part of that given administrative unit; (ii) average years of schooling in 2000 of a given unit of the 1920 municipal division level was calculated for those aged 5+ living in the municipalities that make part of that given administrative unit, and also by specific cohorts, either restricted only for individuals born in those municipalities, and for those that were not born in those municipalities; (iii) total number of schools was calculated from the 2000 School Census originally available from INEP/MEC.

Employment and occupational structure. For the 1872 layer of employment data: (i) the share of legal professionals was calculated as the total number of judges, lawyers, prosecutors, notaries and court clerks over total population (*1000); (ii) the share of workers in the public administration was calculated directly from the number of workers in the public administration available in the census over total population (*1000); (iii) the share of workers in the agricultural sector is the total number of individuals assigned by the 1872 census under the labels of agricultural professions, fishermen and miners over total number of workers; (iv) the share of workers in manufacturing is the total number of individ-

uals assigned by the 1872 census to the labels of manual and mechanical professions over total number of workers; (v) the share of workers employed in the services and retail sectors include the remainder occupations, such as liberal professionals (eg. physicians and lawyers), artists, merchants, retailers, and local traders. We followed as close as possible these same occupational categories in the definition of the employment structure for the 1920, 1940 and 2000 layers of data.

Geographic variables and transportation. Geographic characteristics are available from Ipeadata (distance to the capital city, latitude, longitude, and elevation) and Embrapa Solos (different types of soil) at the more fragmented municipality level of the 2000 administrative division. We use this information at the 2000 administrative level to build geographical variables for the 1920 (and 1872) administrative division. Regarding distance to the capital city, latitude, longitude, and elevation: for a given administrative unit of the 1920 layer of data, we take the average of each variable across the municipalities from the 2000 administrative division that compose that unit. For example, suppose the unit X in 1920 is partitioned into X1 and X2 in 2000. The latitude (longitude) assigned to X is simply the average across the latitude (longitude) of the centroids of X1 and X2. We follow the same procedure for elevation (recorded in each municipality of the 2000 layers as the average elevation in meters) and distance to the capital city (linear distance from the city of São Paulo to the capital district of the municipality in km). Regarding types of soil we use information at the 2000 administrative division level in order to build dummy variables that indicate whether the type of soil is present in a given administrative unit of the 1920 layer of data (types of soil include latosol, argisol, cambisol and spondosol). Regarding the presence of railway, we use historical information at the train station level (location and date of construction, available from www.estacoesferroviarias.com.br) in order to build a dummy variable indicating whether the municipality is provided with railway transportation.

Figure 1: Arrival of Immigrants in the State of São Paulo, Brazil, 1872-1940



Source: Memorial do Imigrante and Censuses of 1872, 1890, 1900, 1910, 1920 and 1940.

Figure 2: Literacy Rates in Selected State Sponsored Settlements, State of São Paulo, Brazil, Late 19th Century and Early 20th Century

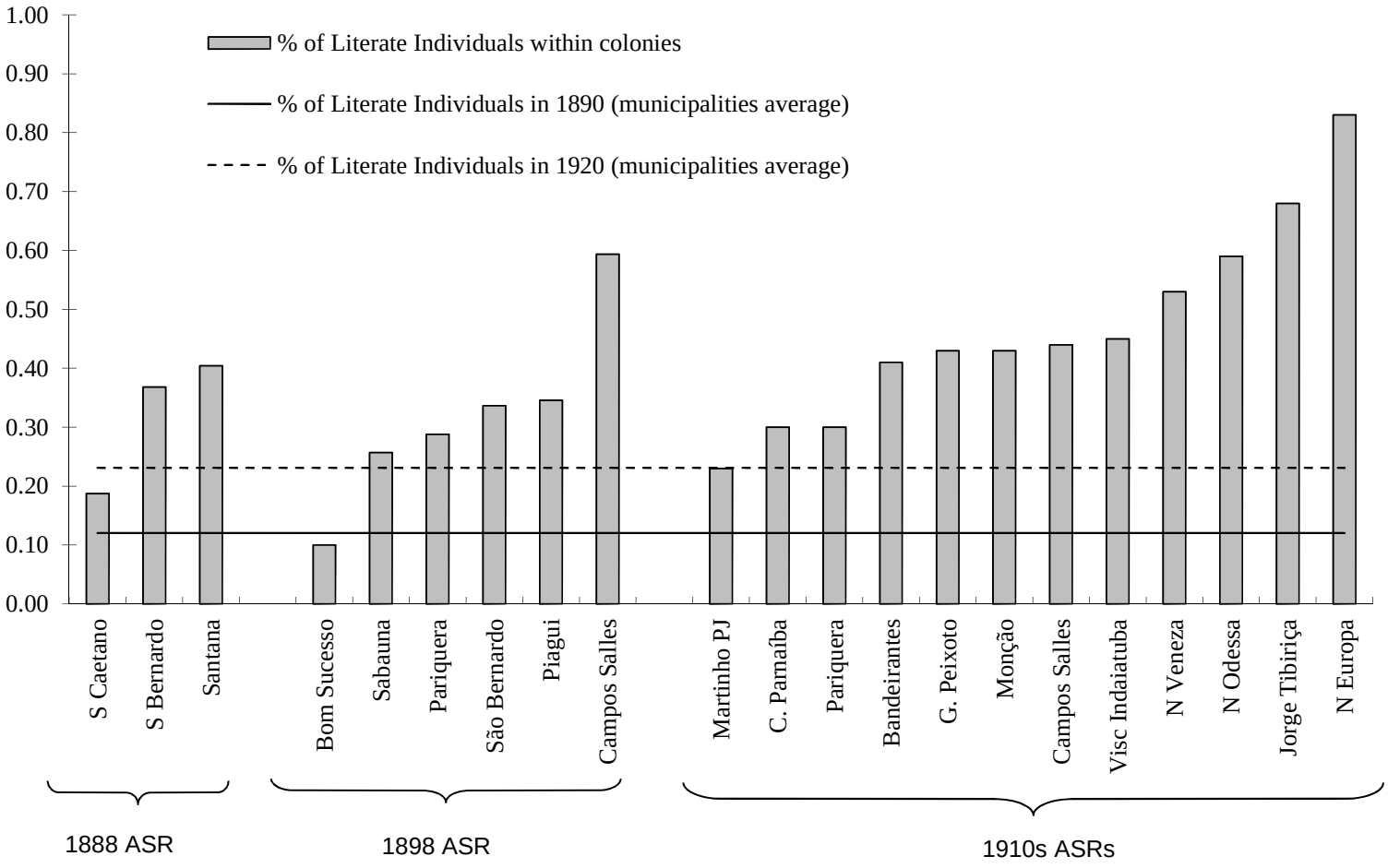
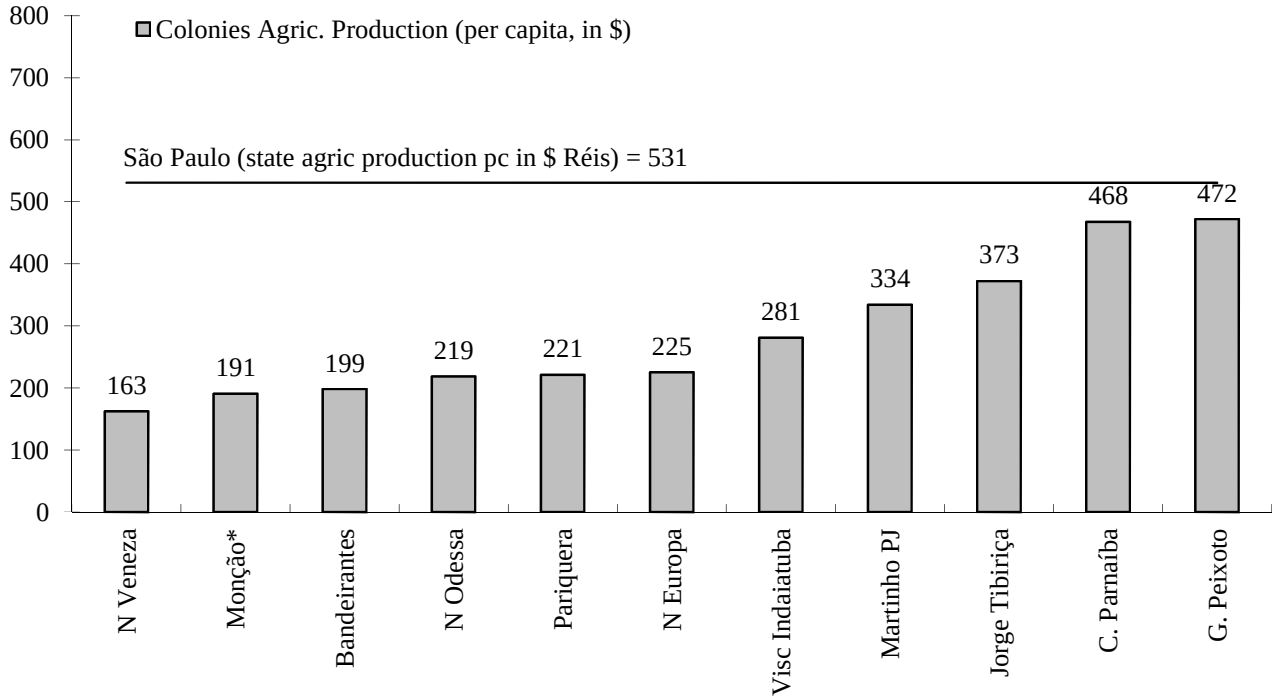
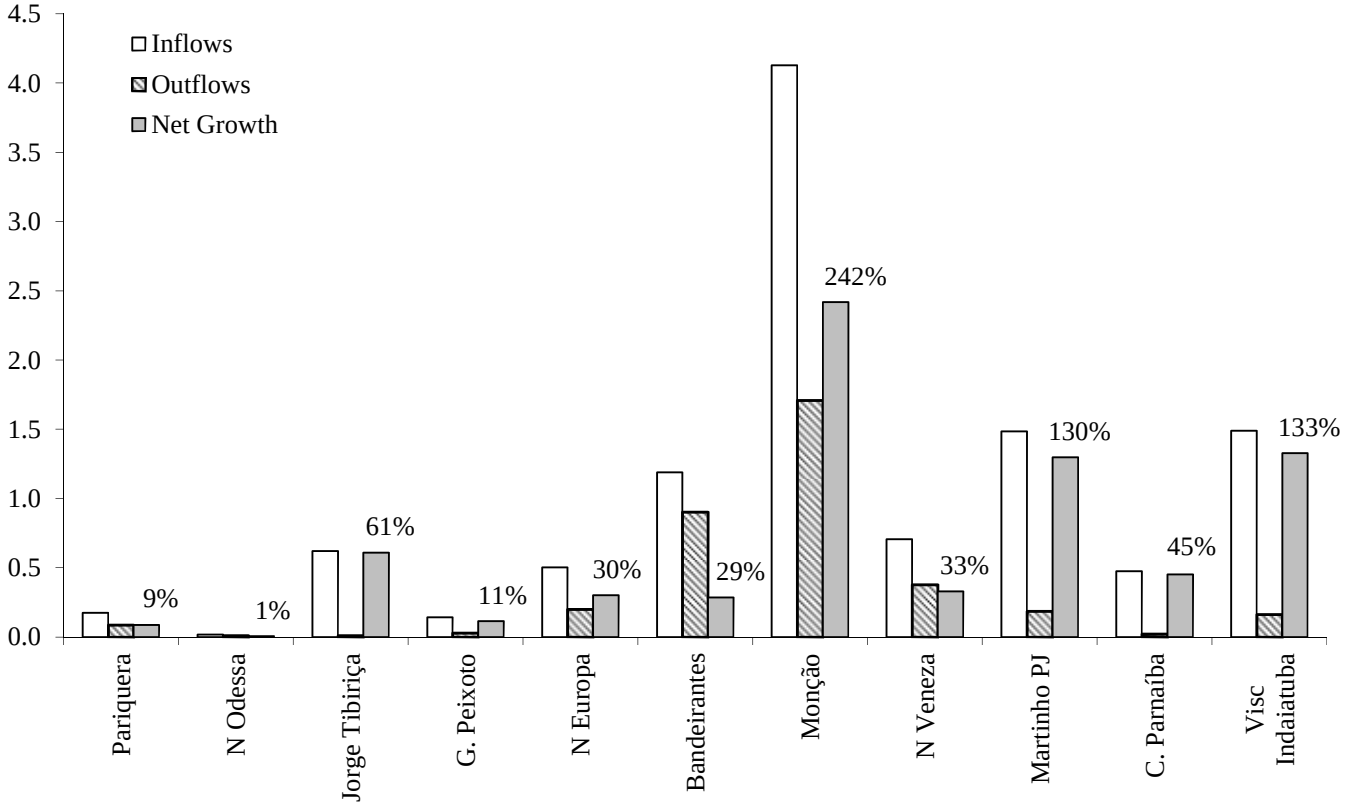


Figure 3: Productivity in Selected State Sponsored Settlements, State of São Paulo, Brazil, Early 20th Century



Source: 1920 Census and 1910s ASRs.

Figure 4: Inflow and Outflow of Immigrants and Net Growth in Selected State Sponsored Settlements, State of São Paulo, Brazil, 1912-1918 (as fractions of 1912 population)



Source: 1910s ASRs.

Table 1: Basic Information on the State Sponsored Settlements Considered in the Analysis, São Paulo, Brazil

Settlements	Year of Foundation	Year of Emancipation	Original Municipality	Current Municipality	Initial Area (hectares)	Predominant Nationalities	Sources of Information
Glória	1877	N/A	São Paulo	São Paulo	N/A	N/A	Pires (2006)
Santanna	1877	N/A	São Paulo	São Paulo	84	N/A	1877 PPR
São Bernardo	1877	1901	São Bernardo	São Bernardo	1959	Italians/Germans/Poles	1877 PPR
São Caetano	1877	N/A	São Bernardo	São Caetano	1909	Italians	Martins (1973)
Canas	1885	1893	Lorena	Canas	1110	Italians/Brazilians	1892 ARSA
Cascalho	1885	1893	Limeira	Cordeirópolis	N/A	Italians	1892 ARSA
Pariquera-açu	1887	1901	Iguape	Pariquera-açu	16603	Germans/Brazilians	1898 ARSA
Sen. Antônio Prado	1887	1893	Ribeirão Preto	Ribeirão Preto	1500 aprox	Italians	Silva (2004)
Ribeirão Pires	1887	1893	São Bernardo	Ribeirão Pires	N/A	N/A	Freitas (2008)
Rodrigo Silva	1887	1893	Porto Feliz	Porto Feliz	1601	Belgians	Souza (1978)
Boa Vista	1888	1893	Jacareí	Jacareí	N/A	Italians	
Barão de Jundiaí	1887	1893	Jundiaí	Jundiaí	514 aprox	Italians	Pereira and Fillipini (1988)
Sabaúna	1889	1901	Mogi das Cruzes	Mogi das Cruzes	N/A	Spaniards/Brazilians	1898 ARSA
Quiririm	1890	1893	Taubaté	Taubaté	N/A	Italians	Lorenzo (2002)
Piagui	1892	1901	Guaratinguetá	Guaratinguetá	1264	N/A	1892 ARSA
Bom Sucesso	1894	1899	Sorocaba	Sorocaba	N/A	Brazilians	1898 ARSA
Campos Sales	1897	1903	Campinas	Cosmópolis	3200	Swiss/Germans	Baldini (2008)
Jorge Tibiriçá	1905	1923-24	Rio Claro	Corumbataí	4356	Spaniards/Italians/Brazilians	1900-1910s ASRAs and ASRs
Nova Odessa	1905	1920-21	Campinas	Nova Odessa	4310	Russians	1900-1910s ASRAs and ASRs
Gavião Peixoto	1907	1923-24	Araraquara	Gavião Peixoto	4840	N/A	1900-1910s ASRAs and ASRs
Nova Paulicéia	1907	1920-21	Araraquara	Gavião Peixoto	4840	N/A	1900-1910s ASRAs and ASRs
Nova Europa	1907	1920-21	Ibitinga/Araraquara	Nova Europa	4840	Germans/Brazilians	1900-1910s ASRAs and ASRs
Bandeirantes	1908	N/A	S. José do Barreiro	S. J. do Barreiro	14104	Germans/Brazilians	1900-1910s ASRAs and ASRs
Monção	1910	N/A	Sta Barbara do Rio Pardo	Iaras	28873	Various	1900-1910s ASRAs and ASRs
Nova Veneza	1910	1919	Campinas	Sumaré	3395	N/A	1900-1910s ASRAs and ASRs
Conde de Parnaíba	1911	1919	Mogi-Mirim	Conchal	4622	N/A	1900-1910s ASRAs and ASRs
Martinho Prado Jr	1911	1920-21	Mogi-Guaçu	Mogi-Mirim	5729	Spaniards/Italians/Brazilians	1900-1910s ASRAs and ASRs
Visc. de Indaiatuba	1911	1923-24	Mogi-Mirim	Mogi-Mirim	2853	N/A	1900-1910s ASRAs and ASRs

Note: N/A (Non-available) indicates that we could not find a source containing the respective information of the settlement.

Table 2 - Summary Statistics for Geographic Characteristics and for Socioeconomic Variables by Census Year

Variables	Mean	Std. Dev.	Min	Max
Panel A - Geography				
Settlement	0.10	0.30	0.00	1.00
Log Distance to capital (km)	5.14	0.69	2.08	6.24
Latitude	-22.59	0.99	-25.02	-20.13
Longitude	-47.61	1.40	-51.45	-44.39
Altitude (100m)	6.02	1.87	0.01	11.98
Dummy for latosol	0.55	0.46	0.00	1.00
Dummy for argisol	0.37	0.44	0.00	1.00
Dummy for cambisol	0.11	0.30	0.00	1.00
Dummy for spondosol	0.01	0.10	0.00	1.00
Panel B - Variables in 1872				
% Literate (aged 6+)	0.20	0.11	0.05	0.45
% Children attending school	0.14	0.10	0.03	0.76
Teachers/children (*1000)	5.07	4.38	0.00	25.51
Population (in 1000)	11.15	7.20	1.57	41.75
% Foreigners	0.01	0.02	0.00	0.08
% Slaves	0.15	0.09	0.04	0.53
Dummy for Railway	0.02	0.16	0.00	1.00
% Emp agriculture	0.59	0.10	0.35	0.91
% Emp manufacturing	0.11	0.04	0.02	0.24
% Emp services and retail	0.30	0.09	0.06	0.57
Panel C - Variables in 1920				
% Literate (aged 6+)	0.30	0.10	0.10	0.70
Schools/children (*1000)	0.39	0.38	0.00	1.86
Teachers/children (*1000)	9.56	6.94	0.83	46.00
Population (in 1000)	22.6	43.1	2.9	577.6
% Foreigners	0.13	0.10	0.00	0.36
% Literate foreigners	0.44	0.16	0.00	1.00
% Small farms	0.67	0.19	0.00	0.97
Coffee production	0.03	0.04	0.00	0.24
Value of farmland	191.0	153.6	12.0	895.0
Dummy for Railway	0.69	0.46	0.00	1.00
% Emp agriculture	0.78	0.13	0.07	0.96
% Emp manufacturing	0.09	0.07	0.01	0.49
% Emp services and retail	0.12	0.07	0.03	0.55
Panel D - Variables in 1940				
% Literate (aged 5+)	0.43	0.13	0.14	0.84
% Children attending school	0.34	0.12	0.08	1.00
Schools/children (*1000)	5.63	1.84	1.61	15.42
Population (in 1000)	37.0	102.2	3.8	1326.3
% Emp agriculture	0.76	0.18	0.03	0.95
% Emp manufacturing	0.08	0.10	0.01	0.70
% Emp services and retail	0.16	0.11	0.03	0.77
Panel E - Variables in 2000				
% Literate (aged 5+)	0.89	0.03	0.79	0.93
% Children attending school	0.96	0.02	0.86	0.99
Schools/children (*1000)	6.84	4.20	2.35	29.19
Teachers/children (*1000)	82.83	21.23	19.67	152.64
Years of schooling (aged 5+)	5.54	0.66	3.53	7.11
Population (in 1000)	183.3	818.4	2.9	11100
Income per capita	307.4	80.1	120.4	597.1
% Emp agriculture	0.20	0.13	0.00	0.60
% Emp manufacturing	0.25	0.09	0.10	0.55
% Emp services and retail	0.53	0.10	0.25	0.78

Notes: In all panels the sample consists of 202 municipalities based on the 1920 municipality boundaries. Data for geographic variables originally from Ipeadata (distance to the capital, latitude, longitude and elevation) and Embrapa Solos (types of soil). Indicator for railway for each municipality and year built on data originally from www.estacoesferroviarias.com.br. The remainder variables in panels B to E correspond to socioeconomic characteristics for each municipality and year originally available from the respective census.

Table 3 - Summary Statistics for Geographic Characteristics and for Socioeconomic Variables in 1872 for Municipalities With and Without Settlements After 1872

Variables in 1872	Settlements		No Settlements		Difference
	Mean	Std. Dev.	Mean	Std. Dev.	
Log distance to capital (in km)	4.92	0.73	4.87	0.58	0.053
Latitude	-22.81	0.76	-22.94	0.90	0.126
Longitude	-47.14	1.49	-46.84	1.20	-0.299
Altitude (in 100m)	5.74	1.53	5.96	2.57	-0.217
Latosol (0/1)	0.53	0.38	0.44	0.44	0.091
% Literacy rate	0.19	0.07	0.17	0.10	0.019
% Children attending school	0.15	0.15	0.16	0.13	-0.015
% Foreigners	0.02	0.02	0.01	0.01	0.010*
% Slaves	0.20	0.10	0.17	0.10	0.025
Population density	9.72	7.49	9.47	6.88	0.249
Railway	0.11	0.32	0.01	0.12	0.091
Public administration	1.15	1.05	1.01	1.11	0.137
Legal professionals	0.95	0.57	0.86	0.94	0.087
% Emp agriculture	0.61	0.10	0.59	0.12	0.014
% Emp manufacturing	0.11	0.04	0.10	0.05	0.012
% Emp services and retail	0.28	0.09	0.31	0.11	-0.027

Notes: The sample of municipalities that would receive settlements after 1872 contains 19 observations (column Settlements); there are 69 observations in the sample of municipalities that would not receive any settlements (column No Settlements). All variables computed according to the 1872 census boundaries. Data for geographic variables originally from Ipeadata (distance to the capital, latitude, longitude and elevation) and Embrapa Solos (indicator for presence of Latosol). Indicator for railway built on data originally from www.estacoesferroviarias.com.br. The remainder variables were built on data originally available from the 1872 Census. Share of foreigners and slaves computed over total population. Share of literate over population aged 6+. Share of children attending school over total number of children aged 6-15. Population density is total population per hectare. Public administration and Legal profession refer to total number of workers in the given occupation relative to total population *1000. Share of workers in agriculture, manufacturing, services and retail computed over total number of occupied workers.

Table 4 - Baseline Differences in Characteristics: Geographic and Socioeconomic Determinants of Settlements Location

Dependent variable: Settlement after 1872	Geography	Demographics	Occupations	All Characteristics
	(1)	(2)	(3)	(4)
Log Distance to capital (in km)	-0.045			0.029
	(0.143)			(0.145)
Latitude	0.032			-0.003
	(0.082)			(0.095)
Longitude	-0.021			-0.015
	(0.049)			(0.061)
Altitude (in 100m)	-0.033			-0.024
	(0.017)*			(0.023)
Dummy for latosol	0.176			0.172
	(0.249)			(0.302)
% Literacy rate		0.015		-0.185
		(0.452)		(0.582)
% Children attending school		-0.154		-0.043
		(0.320)		(0.619)
% Foreigners		4.904		4.043
		(3.407)		(4.074)
% Slaves		0.277		0.017
		(0.499)		(0.682)
Population density		-0.005		0.001
		(0.008)		(0.010)
Railway		0.376		0.507
		(0.259)		(0.274)*
Public administration			0.013	0.021
			(0.046)	(0.059)
Legal professionals			0.076	0.073
			(0.059)	(0.081)
Agriculture			-10.626	-3.669
			(18.054)	(18.006)
Manufacturing			-9.793	-2.993
			(18.138)	(18.172)
Services and Retail			-11.001	-3.903
			(18.217)	(18.198)
Observations	88	88	88	88
R-squared	0.038	0.09	0.055	0.166

Notes: Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1. All columns report the results from OLS regressions where the dependent variable is an indicator for the presence of settlements after 1872. All variables computed according to the 1872 census boundaries. Data for geographic variables originally from Ipeadata (distance to the capital, latitude, longitude and elevation) and Embrapa Solos (indicator for Latosol). Indicator for railway built on data originally from www.estacoesferroviarias.com.br. The remainder variables were built on data originally available from the 1872 Census. Share of foreigners and slaves computed over total population. Share of literate over population aged 6+. Share of children attending school over total number of children aged 6-15. Population density is total population per hectare. Public administration and Legal profession refer to total number of workers in the given occupation over total population *1000. Share of workers in agriculture, manufacturing, services and retail computed over total number of occupied workers.

Table 5 - The Short-Term Effects of Settlements on Literacy Rates

Dependent Variable: Literacy rate (%) in 1920	No controls	Control for Geography	Control for Charact. 1872	No Controls	Control for Geography	Control for Charact. 1872
	(1)	(2)	(3)	(4)	(5)	(6)
Settlement	0.104	0.102	0.081			
	(0.030)***	(0.028)***	(0.027)***			
% Population in Settlement				0.748	0.676	0.484
				(0.310)**	(0.320)**	(0.260)*
% Foreigners 1872			-0.06			0.021
			(0.604)			(0.535)
% Slaves 1872			0.319			0.333
			(0.090)***			(0.090)***
% Literate 1872			0.063			0.077
			(0.060)			(0.061)
% Agriculture 1872			1.579			0.962
			(2.417)			(2.398)
% Manufacturing 1872			1.865			1.239
			(2.417)			(2.402)
% Services and Retail 1872			1.611			0.989
			(2.401)			(2.385)
Railway 1872			0.115			0.091
			(0.024)***			(0.026)***
Geography controls	No	Yes	Yes	No	Yes	Yes
Observations	202	202	202	199	199	199
Adjusted R-squared	0.10	0.31	0.38	0.04	0.29	0.38

Notes: Robust standard errors in parentheses, clustered at the 1872 census boundaries: *** p<0.01, ** p<0.05, * p<0.1. All columns report the results from OLS regressions where the dependent variable is the municipality literacy rate in 1920 (in %). In columns 1-3 the variable settlement is a dummy that equals one if the municipality received at least one state sponsored settlement before 1920. In columns 4-6 the variable settlement is computed as the population of the settlement relative to the existing population in the municipality, and only include 199 municipalities where we have information of the settlement population. Columns 2 and 4 include only geographic controls (distance to the capital, latitude, longitude, elevation, and indicators for different types of soil - latosol, argisol, cambisol, and spondosol). In columns 3 and 6 we add controls for the presence of railway, demographic and economic chacarteristics in 1872 (share of foreigners, share of slaves, share of literate population, share of children attending school, population density, total number of workers in public administration and legal professions relative to total population, share of workers in agriculture, manufacturing, services and retail computed over total number of occupied workers). All variables computed according to the 1920 census boundaries.

Table 6 - Mid and Long-Term Effects of Settlements on Human Capital

Dependent variable:	1940 Census		2000 Census		
	Literacy rate (aged 5+)	Literacy rate (aged 15-19)	Literacy rate (aged 5+)	Literacy rate (aged 15-19)	Years of Schooling (aged 5+)
	(1)	(2)	(3)	(4)	(5)
Settlement	0.107	0.135	0.015	0.001	0.519
	(0.034)***	(0.040)***	(0.005)***	(0.002)	(0.138)***
Geographic controls	Yes	Yes	Yes	Yes	Yes
Characteristics 1872	Yes	Yes	Yes	Yes	Yes
Observations	202	202	202	202	202
Adjusted R-squared	0.45	0.45	0.41	0.16	0.42

Notes: Robust standard errors in parentheses, clustered at the 1872 census boundaries: *** p<0.01, ** p<0.05, * p<0.1. In all columns the variable settlement is a dummy that equals one if the municipality received at least one state sponsored settlement before 1920. Dependent variables defined at the top of each column, and built based on data originally from the 1940 census (columns 1-2) and the 2000 census (columns 3-5). In columns 1-4 literacy rates calculated in % and relative to the respective population. In column 5 the dependent variable is the municiaplity average years of schooling for those aged 5+. All specifications include geographic controls (distance to the capital, latitude, longitude, elevation, and indicators for different types of soil - latosol, argisol, cambisol, and spondosol) and controls for the presence of railway and other chacarteristics in 1872 (share of foreigners, share of slaves, share of literate population, share of children attending school, population density, total number of workers in public administration and legal professions relative to total population, share of workers in agriculture, manufacturing, services and retail computed over total number of occupied workers). All variables computed according to the 1920 census boundaries.

Table 7 - Persistent Effects of Settlements on Years of Schooling, by Cohorts of Individuals Born in the Municpality

Dependent variable: Years of schooling for individuals born in municipality	All Cohorts	1920-29	1930-39	1940-49	1950-59	1960-69
	(1)	(2)	(3)	(4)	(5)	(6)
Settlement	0.883	0.655	0.759	0.941	0.927	1.022
	(0.229)***	(0.236)***	(0.225)***	(0.277)***	(0.303)***	(0.200)***
Mean Dependent Variable	6.63	2.73	3.54	4.64	6.10	7.04
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes
Characteristics 1872	Yes	Yes	Yes	Yes	Yes	Yes
Observations	202	202	202	202	202	202
Adjusted R-squared	0.50	0.39	0.46	0.45	0.43	0.50

Notes: Robust standard errors in parentheses, clustered at the 1872 census boundaries: *** p<0.01, ** p<0.05, * p<0.1. In all columns the variable settlement is a dummy that equals one if the municipality received at least one state sponsored settlement before 1920. Dependent variables are average years of schooling for each cohort of individuals born in the municipality, as defined at the top of each column, and built on microdata originally from the 2000 census. All specifications include geographic controls (distance to the capital, latitude, longitude, elevation, and indicators for different types of soil - latosol, argisol, cambisol, and spondosol) and controls for the presence of railway and other chacarteristics in 1872 (share of foreigners, share of slaves, share of literate population, share of children attending school, population density, total number of workers in public administration and legal professions relative to total population, share of workers in agriculture, manufacturing, services and retail computed over total number of occupied workers). All variables computed according to the 1920 census boundaries.

Table 8 - Mid and Long-Term Effects of Settlements on Human Capital: Panel-Data Specifications

	Dependent Variable: Literacy Rate (%)		
	(1)	(2)	(3)
Settlement × 1920	0.044	0.043	0.041
	(0.024)*	(0.024)*	(0.024)*
Settlement × 1940	0.097	0.086	0.068
	(0.037)***	(0.031)***	(0.031)**
Municipality fixed-effect	Y	Y	Y
Year fixed-effect	Y	Y	Y
Geography*year	N	Y	Y
Demographics*year	N	N	Y
Observations	264	264	264
Municipalities	88	88	88
R-squared	0.748	0.799	0.81

Notes: Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1. In all columns the dependent variable is literacy rate (in %). Dependent and independent variables for 1872, 1920 and 1940 computed according to the 1872 census boundaries. All specifications include municipality fixed-effects and year fixed-effects. Column 2 adds interaction terms between yer of census and geographic variables (distance to the capital, latitude, longitude, elevation, and indicators for different types of soil - latosol, argisol, cambisol, and spondosol). Column 3 adds interaction terms between year of census and demographic variables in 1872 (share of foreigners, share of slaves, share of literate population, share of children attending school, population density, total number of workers in public administration and legal professions relative to total population, share of workers in agriculture, manufacturing, services and retail computed over total number of occupied workers).

Table 9 - The Effects of Settlements on School Inputs and School Attendance

Dependent variable:	Schools per school aged child	Teachers per school aged child	% Children attending school
	(1)	(2)	(3)
Panel A: 1920 Census			
Settlement	0.222	5.053	NA
	(0,108)**	(1,874)***	
R-squared	0.163	0.423	
Panel B: 1940 Census			
Settlement	-0.251	NA	0.08
	(0.294)		(0,025)***
R-squared	0.202		0.381
Panel C: 2000 Census			
Settlement	-0.622	11.235	0.006
	(0.721)	(4,601)**	(0,003)*
R-squared	0.575	0.121	0.405
Geographic controls	Yes	Yes	Yes
Characteristics 1872	Yes	Yes	Yes
Observations	202	202	202

Notes: Robust standard errors in parentheses, clustered at the 1872 census boundaries: *** p<0.01, ** p<0.05, * p<0.1. In all columns and panels the variable settlement is a dummy that equals one if the municipality received at least one state sponsored settlement before 1920. Dependent variables are defined at the top of each column, and were built on data originally from the 1920 census (Panel A), 1940 census (Panel B) and the 2000 census (Panel C). All dependent variables computed relative to total population of children aged 7-14 years old. In columns 1-2 dependent variables computed as the total number of schools (1) or teachers (2) per school aged children *1000. All specifications include geographic controls (distance to the capital, latitude, longitude, elevation, and indicators for different types of soil - latosol, argisol, cambisol, and spondosol) and controls for the presence of railway and other chacarteristics in 1872 (share of foreigners, share of slaves, share of literate population, share of children attending school, population density, total number of workers in public administration and legal professions relative to total population, share of workers in agriculture, manufacturing, services and retail computed over total number of occupied workers). All variables computed according to the 1920 census boundaries.

Table 10 - The Short-Term Effects of Settlements on Other Demographic and Economic Characteristics

Dependent variable in 1920	% Foreigners	% Literate foreigners	Population density	% Small farms	Coffee productivity	Log land prices	Log wages construction	Log wages agriculture
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Settlement	0.025	0.051	14.226	0.046	-0.008	0.185	0.04	0.031
	(0.011)**	(0.027)*	(15.060)	(0.029)	(0.007)	(0.144)	(0.062)	(0.056)
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Characteristics 1872	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	202	172	202	202	202	202	148	125
R-squared	0.63	0.33	0.30	0.25	0.27	0.51	0.44	0.61

Notes: Robust standard errors in parentheses, clustered at the 1872 census boundaries: *** p<0.01, ** p<0.05, * p<0.1. In all columns the variable settlement is a dummy that equals one if the municipality received at least one state sponsored settlement before 1920. Dependent variables are defined at the top of each column, and were built on data originally from the 1920 census: share of foreigners is defined relative to total population, share of literate foreigners defined relative to total population of foreigners; population density is computed as total population per hectare; share of small farms refers to the total number of farms up to 100 hectares relative to total number of farms in the municipality; coffee productivity defined as total production in tonnes per hectare; land price and wages in Réis. All specifications include geographic controls (distance to the capital, latitude, longitude, elevation, and indicators for different types of soil - latosol, argisol, cambisol, and spondosol) and controls for the presence of railway and other chacarteristics in 1872 (share of foreigners, share of slaves, share of literate population, share of children attending school, population density, total number of workers in public administration and legal professions relative to total population, share of workers in agriculture, manufacturing, services and retail computed over total number of occupied workers). All variables computed according to the 1920 census boundaries. In column 2 we restrict the sample to municipalities that have at least 1% of foreigners in population. The samples in columns 7 and 8 are restricted to municipalities without missing observations in wages.

Table 11 - The Effects of Settlements on Literacy Rates, Conditional on National Identities and Religion

Dependent Variable:	Literacy Rate in 1920			Literacy Rate in 1940	Years of Schooling in 2000
	Control for country of origin	Control for religion	Control for country of origin and religion	Control for country of origin and religion	Control for country of origin and religion
	(1)	(2)	(3)	(4)	(5)
Settlement	0.087	0.085	0.091	0.121	0.521
	(0.027)***	(0.025)***	(0.026)***	(0.032)***	(0.149)***
% Germans	0.224		0.206	0.182	0.061
	(0.081)***		(0.080)**	(0.100)*	(0.528)
% Spaniards	0.067		0.066	0.113	0.602
	(0.070)		(0.066)	(0.066)*	(0.356)*
% Italians	0.128		0.129	0.211	0.54
	(0.055)**		(0.055)**	(0.062)***	(0.332)
% Portuguese	0.102		0.104	0.242	1.081
	(0.059)*		(0.059)*	(0.086)***	(0.378)***
% Japanese	0.019		0.02	0.034	0.462
	(0.074)		(0.070)	(0.077)	(0.411)
% Catholics		0.083	0.081	0.064	0.128
		(0.023)***	(0.024)***	(0.039)	(0.307)
Geographic controls	Yes	Yes	Yes	Yes	Yes
Characteristics 1872	Yes	Yes	Yes	Yes	Yes
Observations	202	202	202	202	202
Adj R-squared	0.46	0.45	0.48	0.44	0.37

Notes: Robust standard errors in parentheses, clustered at the 1872 census boundaries: *** p<0.01, ** p<0.05, * p<0.1. In columns 1-3 the dependent variable is literacy rate in 1920 (in %) built on data originally from the 1920 census. In column 4 the dependent variable is literacy rate in 1940 (in %) built on data originally from the 1940 census. In column 5 the dependent variable is average years of schooling in 2000, built on data from the 2000 census. In all columns the variable settlement is a dummy that equals one if the municipality received at least one state sponsored settlement before 1920. Columns 1 and 3 add the shares of individuals from different nationalities computed as the total number of individuals fom the given nationality relative to total population in the municipality in 1920. Columns 2 and 3 add the share of catholics computed as the number of catholics relative to total population in 1920. Data on nationalities and number of catholics originally from the 1920 census. All specifications include geographic controls (distance to the capital, latitude, longitude, elevation, and indicators for different types of soil - latosol, argisol, cambisol, and spondosol) and controls for the presence of railway and other chacarteristics in 1872 (share of foreigners, share of slaves, share of literate population, share of children attending school, population density, total number of workers in public administration and legal professions relative to total population, share of workers in agriculture, manufacturing, services and retail computed over total number of occupied workers). All variables computed according to the 1920 census boundaries.

Table 12 - Long-Run Effects of Settlements on Income per Capita

Dependent variable: Log income per capita in 2000	OLS	OLS	IV	IV
	(1)	(2)	(3)	(4)
Settlement	0.153			
	(0.054)***			
Literacy rate (%) in 1920		1.122	1.884	0.274
		(0.131)***	(0.434)***	(1.252)
Years of schooling in 2000				0.148
				(0.074)**
F-test excluded instrument			9.34	2.31
Geographic controls	Yes	Yes	Yes	Yes
Characteristics 1872	Yes	Yes	Yes	Yes
Observations	202	202	202	202
Adjusted R-squared	0.44	0.53		

Notes: Robust standard errors in parentheses, clustered at the 1872 census boundaries: *** p<0.01, ** p<0.05, * p<0.1. Dependent variable in all columns is log income per capita in 2000, built from data from the 2000 census. In column 1 the variable settlement is a dummy that equals one if the municipality received at least one state sponsored settlement before 1920. The first two columns report OLS regressions, while the remainder report 2nd stage regressions from IV specifications where literacy rates in 1920 are instrumented by the variable settlement. In column 4 years of schooling refers to the average years of schooling computed from the 2000 census at the 1920 census boundaries. All specifications include geographic controls (distance to the capital, latitude, longitude, elevation, and indicators for different types of soil - latosol, argisol, cambisol, and spondosol) and controls for the presence of railway and other chacarteristics in 1872 (share of foreigners, share of slaves, share of literate population, share of children attending school, population density, total number of workers in public administration and legal professions relative to total population, share of workers in agriculture, manufacturing, services and retail computed over total number of occupied workers). All variables computed according to the 1920 census boundaries.

Table 13 - The Long-Run Effects of Settlements on Employment and Structural Transformation

Dependent variable:	% Emp. Agriculture	% Emp. Manufacturing	% Emp. Retail and Services
	(1)	(2)	(3)
Panel A: 1920 Census			
Settlement	-0.121	0.072	0.048
	(0.039)***	(0.024)***	(0.017)***
Adjusted R-squared	0.457	0.390	0.503
Panel B: 1940 Census			
Settlement	-0.165	0.092	0.073
	(0.046)***	(0.027)***	(0.028)**
Adjusted R-squared	0.496	0.472	0.396
Panel C: 2000 Census			
Settlement	-0.053	0.006	0.048
	(0.019)***	(0.016)	(0.017)***
Adjusted R-squared	0.391	0.315	0.280
Observations	202	202	202
Geographic controls	Yes	Yes	Yes
Characteristics 1872	Yes	Yes	Yes

Notes: Robust standard errors in parentheses, clustered at the 1872 census boundaries: *** p<0.01, ** p<0.05, * p<0.1. Dependent variables are defined at the top of each column, and were built on data originally from the 1920 census (Panel A), the 1940 census (Panel B) and the 2000 census (Panel C). All dependent variables refer to the number of workers employed in the given industry relative to the total number of occupied individuals in the municipality. In all columns and panels the variable settlement is a dummy that equals one if the municipality received at least one state sponsored settlement before 1920. All specifications include geographic controls (distance to the capital, latitude, longitude, elevation, and indicators for different types of soil - latosol, argisol, cambisol, and spondosol) and controls for the presence of railway and other chacarteristics in 1872 (share of foreigners, share of slaves, share of literate population, share of children attending school, population density, total number of workers in public administration and legal professions relative to total population, share of workers in agriculture, manufacturing, services and retail computed over total number of occupied workers). All variables computed according to the 1920 census boundaries.

Table 14 - The Effects of Settlements on Years of Schooling, by Cohorts of Individuals Not Born in the Municipality

Dependent variable: Years of Schooling of immigrants	All Cohorts	1920-29	1930-39	1940-49	1950-59	1960-69
	(1)	(2)	(3)	(4)	(5)	(6)
Settlement	0.544	0.333	0.285	0.534	0.542	0.766
	(0.147)***	(0.174)*	(0.185)	(0.191)***	(0.198)***	(0.177)***
Mean Dependent Variable	5.79	3.02	3.61	4.56	5.85	6.51
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes
Characteristics 1872	Yes	Yes	Yes	Yes	Yes	Yes
Observations	202	202	202	202	202	202
R-squared	0.33	0.28	0.27	0.27	0.21	0.20

Notes: Robust standard errors in parentheses, clustered at the 1872 census boundaries: *** p<0.01, ** p<0.05, * p<0.1. In all columns the variable settlement is a dummy that equals one if the municipality received at least one state sponsored settlement before 1920. Dependent variables are average years of schooling for each cohort of individuals that were not born in the municipality, as defined at the top of each column, and built on microdata originally from the 2000 census. All specifications include geographic controls (distance to the capital, latitude, longitude, elevation, and indicators for different types of soil - latosol, argisol, cambisol, and spondosol) and controls for the presence of railway and other chacarteristics in 1872 (share of foreigners, share of slaves, share of literate population, share of children attending school, population density, total number of workers in public administration and legal professions relative to total population, share of workers in agriculture, manufacturing, services and retail computed over total number of occupied workers). All variables computed according to the 1920 census boundaries.

Table A1 - Definition of Variables and Original Sources of Information

Variables	Description	Original Source of Information
Socioeconomic Conditions in 1872		
% Literate (aged 6+)	Number of literate aged 6+ over total population aged 6+	1872 Census
% Children attending school	Number of children aged 6-15 attending school over total population aged 6-15	1872 Census
Teachers/children (*1000)	Number of teachers over total number of chindren aged 6-15 *1000	1872 Census
Population density	Number of individuals by total area in hectares	1872 Census
% Foreigners	Number of foreigners over total population	1872 Census
% Slaves	Number of slaves over total population	1872 Census
Public administration	Total number of workers in public administration over total population *1000	1872 Census
Legal professionals	Total number of workers in legal professions over total population *1000	1872 Census
% Emp agriculture	Share of workers in agriculture over total number of occupied workers	1872 Census
% Emp manufacturing	Share of workers in manufacturing over total number of occupied workers	1872 Census
% Emp services and retail	Share of workers in services and retail over total number of occupied workers	1872 Census
Socioeconomic Conditions in 1920		
% Literate (aged 6+)	Number of literate aged 6+ over total population aged 6+	1920 Census
Schools/children (*1000)	Number of schools over total number of children aged 6-15 *1000	1920 Census
Teachers/children (*1000)	Number of teachers over total number of chindren aged 6-15 *1000	1920 Census
% Foreigners (and by nationality)	Number of foreigners over total population (and by nationality)	1920 Census
% Literate foreigners	Number of literate foreigners over total foreign population	1920 Census
% Small farms	Number of farms up to 100 hectares over total number of farms	1920 Census
Coffee productivity	Total production of coffee (in tonnes) over total size of rural establishments (in hectares)	1920 Census
Value of farmland	Land price in 1000 Réis/hectare	1920 Census
Wages in construction and agriculture	In Réis	1920 Census
% Emp agriculture	Share of workers in agriculture over total number of occupied workers	1920 Census
% Emp manufacturing	Share of workers in manufacturing over total number of occupied workers	1920 Census
% Emp services and retail	Share of workers in services and retail over total number of occupied workers	1920 Census
Socioeconomic Conditions in 1940		
% Literate (aged 5+)	Number of literate aged 5+ over total population aged 5+	1940 Census
% Children attending school	Number of children aged 7-14 attending school over total population aged 7-14	1940 Census
Schools/children (*1000)	Number of schools over total number of children aged 7-14 *1000	1940 Census
% Emp agriculture	Share of workers in agriculture over total number of occupied workers	1940 Census
% Emp manufacturing	Share of workers in manufacturing over total number of occupied workers	1940 Census
% Emp services and retail	Share of workers in services and retail over total number of occupied workers	1940 Census
Socioeconomic Conditions in 2000		
% Literate (aged 5+)	Number of literate aged 5+ over total population aged 5+	2000 Census
% Children attending school	Number of children aged 7-14 attending school over total population aged 7-14	2000 Census
Schools/children (*1000)	Number of schools over total number of children aged 7-14 *1000	2000 School Census/INEP
Teachers/children (*1000)	Number of teachers over total number of chindren aged 7-14 *1000	2000 Census
Years of schooling (aged 5+ and by cohorts)	Average number of years of schooling for those aged 5+ (or by cohorts)	2000 Census
Income per capita	Total income over total population	2000 Census
% Emp agriculture	Share of workers in agriculture over total number of occupied workers	2000 Census
% Emp manufacturing	Share of workers in manufacturing over total number of occupied workers	2000 Census
% Emp services and retail	Share of workers in services and retail over total number of occupied workers	2000 Census
Geography and Transportation		
Railway	Dummy indicating presence of railway station in the municipality (for each year)	www.estacoesferroviarias.com.br
Distance to the capital in km (city of São Paulo)	Avg distances of the municip. of the 2000 adm. div. that constitute the original municip in 1872 (1920)	Ipeadata
Latitude/Longitude degrees	Idem, but averaging centroids	Ipeadata
Dummy for latosol	Indicates that the predominant type of soil is latosol	Embrapa
Dummy for argisol	Indicates that the predominant type of soil is argisol	Embrapa
Dummy for cambisol	Indicates that the predominant type of soil is cambisol	Embrapa
Dummy for spondosol	Indicates that the predominant type of soil is spondosol	Embrapa

Table A2 - The Effects of Settlements on Employment and Structural Transformation

	Sectors in 1920		Sectors in 2000	
	(1)		(2)	
Manufacturing:				
Construction	0.015	(0.005)***	-0.000	(0.005)
Clothing	0.039	(0.015)**	-0.000	(0.012)
Wood and furniture	0.003	(0.002)	0.002	(0.005)
Metal	0.007	(0.003)**	0.004	(0.003)
Chemical	0.001	(0.000)**	0.004	(0.002)*
Food	0.002	(0.001)**	-0.007	(0.003)**
Machines and Equip	-	-	0.003	(0.003)
Automobile	-	-	0.006	(0.005)
Plastics	-	-	-0.005	(0.005)
Retail and Services:				
Transportation	-	-	0.003	(0.003)
Banks and finance	-	-	0.014	(0.005)**
Education	-	-	0.007	(0.002)***
Health	-	-	0.008	(0.003)**
Personal services	-	-	0.002	(0.001)**
Culture	-	-	0.001	(0.001)
Political activism	-	-	-0.000	(0.000)
Automotive retail and repairing	-	-	0.006	(0.002)***
Agricultural goods	-	-	0.000	(0.001)
Food retail and supermarkets	-	-	0.003	(0.002)
Textile and clothing	-	-	0.002	(0.001)
Dept stores	-	-	0.019	(0.006)***
General repairing	-	-	0.001	(0.000)
Personal goods and hygiene retail	-	-	0.003	(0.001)***
Housing	-	-	0.002	(0.001)
Machines and equipments	-	-	0.001	(0.001)**

Notes: Robust standard errors in parentheses, clustered at the 1872 census boundaries: *** p<0.01, ** p<0.05, * p<0.1. Each coefficient reports the result of a distinct regression where the dependent variable is regressed on the variable settlement, a dummy that equals one if the municipality received at least one state sponsored settlement before 1920. Dependent variables are computed as the total number of workers in the given sector over total number of occupied individuals. Column 1 (2) reports the results for dependent variables in 1920 (2000). All specifications include geographic controls (distance to the capital, latitude, longitude, elevation, and indicators for different types of soil - latosol, argisol, cambisol, and spondosol) and controls for the presence of railway and other chacarteristics in 1872 (share of foreigners, share of slaves, share of literate population, share of children attending school, population density, total number of workers in public administration and legal professions relative to total population, share of workers in agriculture, manufacturing, services and retail computed over total number of occupied workers). All variables computed according to the 1920 census boundaries.

Departamento de Economia PUC-Rio
Pontifícia Universidade Católica do Rio de Janeiro
Rua Marques de Sâo Vicente 225 - Rio de Janeiro 22453-900, RJ
Tel.(21) 31141078 Fax (21) 31141084
www.econ.puc-rio.br
flavia@econ.puc-rio.br