

## Human Capital Persistence and Development<sup>†</sup>

By RUDI ROCHA, CLAUDIO FERRAZ, AND RODRIGO R. SOARES\*

*This paper documents the persistence of human capital over time and its association with long-term development. We exploit variation induced by a state-sponsored settlement policy that attracted immigrants with higher levels of schooling to particular regions of Brazil in the late nineteenth and early twentieth century. We show that one century after the policy, municipalities that received settlements had higher levels of schooling and higher income per capita. We provide evidence that long-run effects worked through higher supply of educational inputs and shifts in the structure of occupations toward skill-intensive sectors. (JEL I26, J22, J24, J61, N36, O15, Z13)*

The sixteenth century European colonial expansion and the ensuing migratory waves out of the Old World permanently changed human societies. While a number of studies document how these historical events shaped current patterns of economic development (Acemoglu, Johnson, and Robinson 2001; Glaeser et al. 2004), a large debate remains on what mechanisms explain this persistence. On the one hand, Acemoglu, Johnson, and Robinson (2005) suggest that persistence occurred through institutional and political channels.<sup>1</sup> On the other hand, a growing set of papers suggests that culture and human capital are important channels through which historical shocks persist through time.<sup>2</sup>

This paper documents the persistent effects of state-sponsored settlements located in the Brazilian state of São Paulo and examines the channels behind this persistence. 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

\*Rocha: Institute of Economics, Federal University of Rio de Janeiro, Av. Pasteur 250, 22290-240, Rio de Janeiro, RJ, Brazil (email: rudi.rocha@ie.ufrj.br); Ferraz: Department of Economics, Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Rua Marquês de São Vicente 225, Rio de Janeiro, RJ, 22453-900, Brazil and BREAD (email: cferraz@econ.puc-rio.br); Soares: Columbia University, 420 W. 118th Street, New York, NY, 10027 and Sao Paulo School of Economics-FGV (email: r.soares@columbia.edu). We are grateful to three anonymous referees for suggestions that greatly improved the paper. We also thank Ran Abramitzky, Juliano Assunção, Maria Sílvia 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, Nico Voigtländer, Hans-Joachim Voth, and seminar participants at Bocconi University, Paris School of Economics, PUC-Rio, São 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 First Economic History Clio Lab Workshop for comments, data, and suggestions. Ferraz gratefully acknowledges financial support from the EH Clio lab at PUC-Chile.

<sup>†</sup>Go to <https://doi.org/10.1257/app.20150532> to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

<sup>1</sup>Dell (2010) provides evidence that extractive institutions in Peru generated persistence in the protection of property rights that explain current differences in welfare.

<sup>2</sup>See, for example, Guiso, Sapienza, and Zingales (2016); and Lowes et al. (2015) on cultural persistence and Wantchekon, Klačnjaz, and Novta (2015); and Dittmar and Meisenzahl (2016) on the role of human capital.

education were channeled into specific localities through deliberate government policies. In the end of the nineteenth and beginning of the twentieth century, public authorities established a number of official settlement colonies throughout the state of São Paulo. This policy had goals involving occupation of territory and food production, 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 ended up being roughly uniform across the region.

We show that settlements attracted a selected type of immigrant and that, analogously to the results from the literature on institutions (such as Acemoglu, Johnson, and Robinson 2001), this selection persisted through time and was associated with higher income per capita in the long run. But different from this literature, we exploit variation across municipalities within the same country. Hence, our findings cannot be explained by differences in formal political institutions. Our findings cannot be accounted for either by the dimensions of culture and social capital that we can observe in the data (nationality, religion, or other cultural traits possibly associated with the incidence of entrepreneurship). In terms of proximate determinants, we provide evidence that the local shock induced by state-sponsored settlements persisted and ended up reflected on local economic development through higher supply of schooling inputs and faster structural change.

We exploit three unique features of this historical episode to characterize the persistence of the initial compositional shock through time and its effects on long-term development. First, differently from the United States where immigrants were, on average, less educated than the native population (see Bandiera et al. 2015), the age of mass migration in Latin America, and especially in Brazil, brought immigrants who were more educated than natives, creating a positive human capital shock. In the case of state-sponsored settlements in São Paulo, this pattern was reinforced by a particularly positive selection of immigrants in terms of education and possibly also other unobservable skills. Second, settlements were established before Brazil's industrialization period. Thus, the choices of settlement location and migration decisions in an agrarian economy were not correlated with expected returns from education or entrepreneurship. And third, the areas where state-sponsored settlements were created were not different from other areas, but for the fact that they received different immigrants.

Our main empirical strategy compares literacy rates and other outcomes across municipalities that received a state-sponsored settlement and municipalities that did not receive a settlement but had similar preexisting 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 nineteenth 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 of São Paulo. 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. We focus most of our discussion throughout the analysis on educational outcomes, but recognize that these may be associated with other unobservable differences in individual traits.

We show that this initial shock persisted not only until 1940, but also 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 trajectories was also followed by 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 in the beginning of the century when compared to other municipalities.

We implement a series of robustness tests to make sure the results are not driven by other factors. First, we show that all results are robust to controlling for geographic and presettlement (1872) socioeconomic characteristics. Second, the results remain significant when we use municipality variation and control for municipal fixed effects (when possible). The results are unchanged as well when we control for proxies of culture and social norms associated with nationality (e.g., Italian versus Portuguese) and religion (e.g., Catholic versus Protestant). Finally, we show that other potential immediate effects of the settlements, such as those on property rights or income, do not drive our results. We find no short-term (1920) effect of settlements on land inequality, agricultural productivity, land prices, and wages, and no long-term effect on entrepreneurship.

In sum, we show that the initial shock represented by the skill composition of immigrants—along educational and possibly other unobservable dimensions—persisted through time and was associated, in the long run, with 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 over time, we analyze the evolution in the supply and use of educational inputs and in the structure of employment. We show that in 1920, 1940, and 2000 enrollment rates and the number of teachers per school-aged children were higher in locations associated with settlements. In these same locations, employment shifted away from agriculture and toward 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. We also see evidence that settlement municipalities started attracting more educated domestic immigrants in the second half of the twentieth century.

Though we cannot identify precisely why and how these communities demanded more investments in schooling, one might conjecture that the initially higher set

of skills was correlated with higher intrinsic values attributed to schooling or with higher political participation. Irrespectively, human capital was associated with a higher share of employment in manufacture already in 1920, when it did not translate into higher incomes. The persistently higher human capital, nevertheless, possibly together with some previous specialization in manufacture, allowed these communities to take better advantage of the industrialization process initiated in the following decades. This combination, in turn, led to higher income per capita in the long run and attracted even more skilled domestic immigrants.

Our paper speaks to three sets of literature. First, it relates to a large literature that measures the impact of population composition and great migration episodes on long-run economic development (e.g., Putterman and Weil 2010, Easterly and Levine 2016). Distinct 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. By looking at a single state within Brazil, we immediately hold constant the *de jure* macro institutions that have been the focus of a large and influential literature (e.g., Acemoglu, Gallego, and Robinson 2014). We also provide evidence on the specific channels through which the composition of immigrants affected long-term development. Finally, while there is a large literature on the consequences of the Age of Mass Migration for the United States (e.g., Abramitzky, Boustan, and Eriksson 2014), there is very little evidence for South America. Our paper is one of the first to provide evidence on the long-run effects of more skilled immigrants on economic development.<sup>3</sup>

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, Klačnjaz, and Novta (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 correlation with income per capita.<sup>4</sup>

And third, our work sheds light on the role of skill composition as a determinant of structural change and long-term development. A growing number of papers have 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 seventeenth century

<sup>3</sup> See also the related work of Droller (forthcoming) for the case of Argentina.

<sup>4</sup> For a related literature that focuses on Latin America, see Summerhill (2010), de Carvalho Filho and Colistete (2010), de Carvalho Filho and Monasterio (2012), Valencia Caicedo (2014), and Droller (forthcoming). Our paper differs from this literature by having a clear identification strategy based on the location of state-sponsored settlements and by analyzing the specific channels behind the historical persistence of human capital and its association with long-term development.

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, Hornung, and Woessmann (2011) also look 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 Voigtländer (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 recent 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, Scheinkman, and Shleifer 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 a unified and sequential way, several results that appeared fragmented in the previous literature. We show that the initial heterogeneity in education persisted through time via increased investments in schooling, as in Huillery (2009); higher educational levels—possibly also correlated with other skills useful in manufacturing—were associated, as industrialization progressed over time, with shifts in the structure of employment towards skill-intensive sectors, as in Becker, Hornung, and Woessmann (2011); and 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 I outlines the historical background and discusses the establishment and the likely impacts of state-sponsored settlements in the state of São Paulo. Section II presents the data and descriptive statistics. Section III describes the empirical strategy. Section IV presents the main econometric results. Section V discusses channels and performs robustness exercises. Section VI closes the paper with some concluding remarks.

## I. Historical Background

### A. *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 of the 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 twentieth century. This process was fueled by the nineteenth 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 1959). 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 (Eusébio de Queirós Law) and the abolishment of slavery in 1888 (Lei 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.<sup>5</sup> 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 (Petroni 1982). In the second half of the nineteenth 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 (Petroni 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 disseminated in European cities (Petroni 1982). The plots of land were offered by Brazilian representatives in Europe and at the Immigration Station

<sup>5</sup>The first experiences of state-sponsored international immigration in Brazil date back to the early nineteenth 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 1959, Petroni 1982, Paiva 1993, and Siriani 2005).

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.<sup>6</sup> 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 the 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. Online Appendix Figure A1 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 nineteenth to the twentieth century. Yet only a small fraction 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.<sup>7</sup>

<sup>6</sup>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.

<sup>7</sup>Data on the 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.

The specific sequence of creation of state-sponsored settlements in São Paulo can be outlined as follows. In the 1870s, four 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: four in the economically backward Vale do Paraíba, two in the central area of the state, four in the expanding coffee region of the Northwest, two in the Midwest (where the main agricultural products were cotton and sugarcane), and one in the coastal Vale do Ribeira. In the early 1900s, nine additional colonies were founded in still isolated areas of the Northwest.<sup>8</sup> Finally, following a renewed interest in populating lagging regions, the federal government established two 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).<sup>9</sup>

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.

### B. Characteristics of the State-Sponsored Settlements

Table 1 lists the state-sponsored settlements included in our analysis, presenting basic information on dates of foundation and—when available—emancipation and location (in terms of the original and current political division of municipalities). Online Appendix Table A1 provides further information, when available, on the initial area, 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 the nineteenth century and the remaining 11 in the first years of the twentieth 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.

<sup>8</sup>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.

<sup>9</sup>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.



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
Glória	1877	N/A	São Paulo	São Paulo
Santanna	1877	N/A	São Paulo	São Paulo
São Bernardo	1877	1901	São Bernardo	São Bernardo
São Caetano	1877	N/A	São Bernardo	São Caetano
Canas	1885	1893	Lorena	Canas
Cascalho	1885	1893	Limeira	Cordeirópolis
Pariquera-açu	1887	1901	Iguape	Pariquera-au
Senador Antônio Prado	1887	1893	Ribeirão Preto	Ribeirão Preto
Ribeirão Pires	1887	1893	São Bernardo	Ribeirão Pires
Rodrigo Silva	1887	1893	Porto Feliz	Porto Feliz
Boa Vista	1888	1893	Jacareí	Jacareí
Barão de Jundiá	1887	1893	Jundiá	Jundiá
Sabaúna	1889	1901	Mogi das Cruzes	Mogi das Cruzes
Quiririm	1890	1893	Taubaté	Taubaté
Piagui	1892	1901	Guaratinguetá	Guaratinguetá
Bom Sucesso	1894	1899	Sorocaba	Sorocaba
Campos Sales	1897	1903	Campinas	Cosmópolis
Jorge Tibiriçá	1905	1923–24	Rio Claro	Corumbataí
Nova Odessa	1905	1920–21	Campinas	Nova Odessa
Gavião Peixoto	1907	1923–24	Araraquara	Gavião Peixoto
Nova Paulicéia	1907	1920–21	Araraquara	Gavião Peixoto
Nova Europa	1907	1920–21	Ibitinga e Araraquara	Nova Europa
Bandeirantes	1908	N/A	S. José do Barreiro	S. José do Barreiro
Monção	1910	N/A	S. Bárbara do Rio Pardo	Iaras
Nova Veneza	1910	1919	Campinas	Sumaré
Conde de Parnaíba	1911	1919	Mogi-Mirim	Conchal
Martinho Prado Jr.	1911	1920–21	Mogi-Guaçu	Mogi-Mirim
Visconde de Indaiatuba	1911	1923–24	Mogi-Mirim	Mogi-Mirim

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

The most common nationalities were Italian and German, but Spanish, Portuguese, and Brazilian, among others, were also present (Bassanezi et al. 2008).<sup>10</sup>

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 1910 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). Online Appendix Figure A2 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,

<sup>10</sup> Brazilians were allowed to purchase plots of land within settlements under certain conditions.

indicating some improvement in the pool of immigrants over time (for the interested reader, online Appendix Figure A2 plots these numbers). Overall, the population within settlements had substantially higher literacy rates when compared to the remainder of the state.

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.<sup>11</sup> 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 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

<sup>11</sup> 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.

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 (for the interested reader, online Appendix Figure A3 plots 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. The 1910s Annual Statistical Reports of São Paulo provide data on arrivals and departures for 11 state-sponsored settlements (presented in online Appendix Figure A4). For each settlement, we have data on cumulative arrivals, departures, and net growth from 1912 to 1918 (as shares of the 1912 population).<sup>12</sup> The 11 settlements for which data are available 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.

### *C. Why Did State-Sponsored Settlements Matter?*

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 an 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 we present in the next sections shows that the main distinguishing feature of the settlements was the attraction of a different pool of immigrants, with higher educational levels and possibly also other unobservable traits. The literature

<sup>12</sup>This cumulative number excludes 1917, since the Annual Statistical Report of São Paulo was not available for this year.

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, Scheinkman, and Shleifer 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 this initial compositional shock 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, Maschio 2005). Based on descriptive historical information, this literature shows 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. Maschio documents the specific demands of immigrants and respective government responses, culminating with the creation of community schools, as in the case of some state-sponsored colonies in Paraná. It seems natural to think that similar processes could have operated as well whenever more educated immigrants were concentrated in other state-sponsored settlements. But it is important to keep in mind that, as mentioned in the introduction, it is possible that these persistent differences in education were also associated with other unobservable productive characteristics, such as entrepreneurship or previously acquired manufacturing skills, all of which may contribute to long-term development.

We also consider some additional 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, Johnson, and Robinson 2005; Galor, Moav, and Vollrath 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 far as Weber (1930) and including more recent work such as Landes (1999) and Fernández (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 and Hall 1996, Baldwin and Martin 2004, and Combes et al. 2010). This process could have been further intensified by the development of transportation

infrastructure, as in Donaldson (forthcoming). We try to assess the relevance of each of these hypotheses while conducting our empirical exercise.

## II. Data

We use three sets of historical data. First, we build a dataset 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 dataset 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 analyses 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.<sup>13</sup>

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 preexisting 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-term and long-term impacts of the settlements.

<sup>13</sup> It is worth mentioning that many official settlements became independent municipalities during the first half of the twentieth 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.

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.<sup>14</sup> 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.<sup>15</sup> 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 presettlement characteristics.<sup>16</sup>

The second dataset 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 dataset allows us to estimate our model using municipal fixed effects.

We select a comprehensive set of variables from the 1872 census to characterize preexisting 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 twentieth 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).

<sup>14</sup>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.

<sup>15</sup>A similar procedure has been used in the United States (see Hornbeck 2012).

<sup>16</sup>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 division.

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). 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).<sup>17</sup> In the online Appendix, we provide the definition of each variable included in our analysis.

*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. 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, the major agricultural crop in the late 1800s.

Panels B to E of Table 2 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 received settlements with those of other municipalities in the state. We do this comparison for the 88 municipalities that existed in the 1872 census. 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 of the 16 variables considered (and, even in this case, at the 10 percent level). In the online Appendix (Table A2), we reproduce the comparison from Table 3 in a multivariate setting. Table A2 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. In general,

<sup>17</sup> Data available from [www.estacoesferroviarias.com.br](http://www.estacoesferroviarias.com.br).

TABLE 2—SUMMARY STATISTICS FOR GEOGRAPHIC CHARACTERISTICS AND FOR SOCIOECONOMIC VARIABLES BY CENSUS YEARS

Variables	Mean	SD	Min	Max
<i>Panel A. Geography</i>				
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
<i>Panel B. Vars in 1872</i>				
Share literate (aged 6+)	0.20	0.11	0.05	0.45
Share children attending school	0.14	0.10	0.03	0.76
Teachers/children ( $\times 1,000$ )	5.07	4.38	0.00	25.51
Population (in 1,000)	11.15	7.20	1.57	41.75
Share foreigners	0.01	0.02	0.00	0.08
Share slaves	0.15	0.09	0.04	0.53
Dummy for railway	0.02	0.16	0.00	1.00
Share employment agriculture	0.59	0.10	0.35	0.91
Share employment manufacturing	0.11	0.04	0.02	0.24
<i>Panel C. Vars in 1920</i>				
Share literate (aged 6+)	0.30	0.10	0.10	0.70
Schools/children ( $\times 1,000$ )	0.39	0.38	0.00	1.86
Teachers/children ( $\times 1,000$ )	9.56	6.94	0.83	46.00
Population (in 1,000)	22.6	43.1	2.9	577.6
Share foreigners	0.13	0.10	0.00	0.36
Share literate foreigners	0.44	0.16	0.00	1.00
Share 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
Share employment agriculture	0.78	0.13	0.07	0.96
Share employment manufacturing	0.09	0.07	0.01	0.49
Share employment services and retail	0.12	0.07	0.03	0.55
<i>Panel D. Vars in 1940</i>				
Share literate (aged 5+)	0.43	0.13	0.14	0.84
Share children attending school	0.34	0.12	0.08	1.00
Schools/children ( $\times 1,000$ )	5.63	1.84	1.61	15.42
Population (in 1,000)	37.0	102.2	3.8	1,326.3
Share employment agriculture	0.76	0.18	0.03	0.95
Share employment manufacturing	0.08	0.10	0.01	0.70
Share employment services and retail	0.16	0.11	0.03	0.77
<i>Panel E. Vars in 2000</i>				
Share literate (aged 5+)	0.89	0.03	0.79	0.93
Share children attending school	0.96	0.02	0.86	0.99
Schools/children ( $\times 1,000$ )	6.84	4.20	2.35	29.19
Teachers/children ( $\times 1,000$ )	82.83	21.23	19.67	152.64
Years of schooling (aged 5+)	5.54	0.66	3.53	7.11
Population (in 1,000)	183.3	818.4	2.9	11,100
Income per capita	307.4	80.1	120.4	597.1
Share employment agriculture	0.20	0.13	0.00	0.60
Share employment manufacturing	0.25	0.09	0.10	0.55
Share employment 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 are 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 are originally from [www.estacoesferroviarias.com.br](http://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		Diff.	SE
	Mean	SD	Mean	SD		
log distance to capital (in km)	4.92	0.73	4.87	0.58	0.053	0.159
Latitude	-22.81	0.76	-22.94	0.90	0.126	0.226
Longitude	-47.14	1.49	-46.84	1.20	-0.299	0.328
Altitude (in 100m)	5.74	1.53	5.96	2.57	-0.217	0.619
Latosol (0/1)	0.53	0.38	0.44	0.44	0.091	0.111
Literacy rate	0.19	0.07	0.17	0.10	0.019	0.024
Share children attending school	0.15	0.15	0.16	0.13	-0.015	0.036
Share foreigners	0.02	0.02	0.01	0.01	0.010	0.004
Share slaves	0.20	0.10	0.17	0.10	0.025	0.026
Population density	9.72	7.49	9.47	6.88	0.249	1.817
Railway	0.11	0.32	0.01	0.12	0.091	0.054
Public administration	1.15	1.05	1.01	1.11	0.137	0.285
Legal professionals	0.95	0.57	0.86	0.94	0.087	0.227
Teachers	0.96	0.72	1.03	0.64	-0.07	0.182
Share emp. agriculture	0.61	0.10	0.59	0.12	0.014	0.029
Share emp. manufacturing	0.11	0.04	0.10	0.05	0.012	0.013
Share emp. services and retail	0.28	0.09	0.31	0.11	-0.027	0.027

*Notes:* This table contains information on 88 municipalities based on the map of São Paulo compatible with the 1872 census. There are 19 observations that received a settlement after 1872 (column “Settlements”) and 69 observations that did not receive a settlement after 1872 (column “No Settlements”). All variables are computed according to the 1872 census boundaries. Data for geographic variables are originally from Ipeadata (distance to the capital, latitude, longitude, and elevation) and Embrapa Solos (indicator for presence of latosol). Indicator for railway built is from [www.estacoesferroviarias.com.br](http://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 computed over population aged 6+. Share of children attending school computed over total number of children aged 6–15. Population density is total population per hectare. Public administration, legal profession, and teachers refer to total number of workers in the given occupation relative to total population  $\times 1,000$ . Share of workers in agriculture, manufacturing, services, and retail is computed over total number of occupied workers.

predetermined characteristics have limited, if any, predictive power for settlement location. Once we add all variables simultaneously, 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).

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. Still, since some of these 1872 characteristics could be correlated with future schooling outcomes, we control for all of them in our benchmark specification.

### III. 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:

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

where  $y_i$  is the outcome of interest in municipality  $i$  measured in 1920 for the short-run effects, 1940 for the middle-run effects, and 2000 for the long-run effects;  $S_i$  is a variable indicating whether the municipality received a state-sponsored settlement between 1872 and 1920; and  $\epsilon_i$  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 predetermined characteristics based on this administrative organization, we report standard errors that are clustered at the more aggregated 1872 division. Our main regressions focus on schooling outcomes, but we use identical specifications in complementary exercises to analyze whether there were other dimensions of heterogeneity across municipalities that received and did not receive settlements in 1920, immediately after their initial establishment; long-term effects on income per capita; whether other factors, such as nationality or religion of immigrants, intervened in the relationship between settlements and long-term outcomes; and the evolution of educational inputs, enrollment rates, and structure of employment over time.

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 dataset 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  $\beta$  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

$$(2) \quad y_{it} = \alpha + \beta_t S_i + \gamma_t \mathbf{X}_i + \pi_t \mathbf{W}_i + \eta_i + \mu_t + \epsilon_{it},$$

where  $i$  indexes a municipality and  $t$  indicates the census year. The variables  $\eta_i$  and  $\mu_t$  represent municipality and year fixed effects, and  $\epsilon_{it}$  is an error term. Here,  $\mathbf{X}_i$  is a vector of geographic controls and  $\mathbf{W}_i$  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.

## IV. Results

### A. *The Educational Effects of State-Sponsored Settlements*

We start the analysis by examining the short-term effects of state-sponsored settlements on education and other local characteristics, including the presence of foreigners, population density, land distribution, and agricultural productivity in coffee. The first three columns of Table 4, panel A, report 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; 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.

We next investigate whether the presence of state-sponsored settlements also affected the composition of immigrants or other local characteristics that might have had an effect on long-term development and, indirectly, on educational outcomes. In Table 4, columns 4–5 of panel A, and 1–6 of panel B report 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 as column 3 of panel A, controlling for geographic and pre-1872 socioeconomic characteristics.

Columns 4 and 5 of panel A in Table 4 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 5 shows that the settlements increased the percentage of literate foreigners in 5 percentage points (an increase of 11 percent

TABLE 4—THE SHORT-TERM AND MEDIUM-TERM EFFECTS OF SETTLEMENTS ON LITERACY RATES

<i>Panel A</i>						
	Literacy rate 1920			Foreigners 1920		
	No controls (1)	Control for geography (2)	Control for charact. 1872 (3)	Share foreigners (4)	Share literate (5)	
Settlement	0.104 [0.030]	0.102 [0.028]	0.081 [0.027]	0.025 [0.011]	0.051 [0.028]	
Observations	202	202	202	202	172	
Adj. $R^2$	0.095	0.310	0.382	0.598	0.260	
Geography	No	Yes	Yes	Yes	Yes	
Charact. 1872	No	No	Yes	Yes	Yes	

  

<i>Panel B</i>						
	Other economic and demographic characteristics in 1920					
	Population density (1)	Share small farms (2)	Coffee product. (3)	In land prices (4)	log wages construc. (5)	log wages agric. (6)
Settlement	14.22 [15.06]	0.046 [0.029]	−0.008 [0.007]	0.185 [0.144]	0.040 [0.062]	0.031 [0.056]
Observations	202	202	202	202	148	125
Adj. $R^2$	0.244	0.180	0.202	0.468	0.373	0.553
Geography	Yes	Yes	Yes	Yes	Yes	Yes
Charact. 1872	Yes	Yes	Yes	Yes	Yes	Yes

*Notes:* Robust standard errors are in brackets, clustered at the 1872 census boundaries. All columns report the results from OLS regressions where the dependent variable is the municipality literacy rate in 1920 (share of literate individuals) for columns 1–3, panel A, and the variable listed in the top of the column from columns 4–5, panel A, and 1–6 in panel B. The variable settlement is a dummy that equals one if the municipality received at least one state-sponsored settlement before 1920. Geographic controls include: distance to the capital, latitude, longitude, elevation, and indicators for different types of soil—latosol, argisol, cambisol, and spondosol. Characteristics in 1872 include presence of railway, 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 regressions estimated for 202 municipalities are based on the 1920 census boundaries.

over a base of 44 percent). Thus, settlements brought slightly more foreigners to the respective areas, but particularly more educated foreigners.<sup>18</sup>

In columns 1 to 6 of panel B, we examine other potential effects of the settlements associated with alternative theoretical channels that have been highlighted in the literature, such as land inequality, population agglomeration, and agricultural productivity. We find no evidence that state-sponsored settlements affected population density, the percentage of small farms (proxy for land inequality), or agricultural productivity in 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

<sup>18</sup>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.

TABLE 5—THE MEDIUM-TERM AND LONG-TERM EFFECTS OF SETTLEMENTS ON SCHOOLING AND INCOME

Dep. variable:	Schooling 1940		Schooling 2000			Income 2000
	Literacy (age 5+) (1)	Literacy (age 15–19) (2)	Literacy (age 5+) (3)	Literacy (age 15–19) (4)	Schooling (age 5+) (5)	Income per capita (6)
Settlement	0.107 [0.034]	0.135 [0.040]	0.015 [0.005]	0.001 [0.002]	0.519 [0.138]	0.153 [0.054]
Geography	Yes	Yes	Yes	Yes	Yes	Yes
Charact. 1872	Yes	Yes	Yes	Yes	Yes	Yes
Observations	202	202	202	202	202	202
Adj. $R^2$	0.405	0.406	0.361	0.089	0.365	0.443

*Notes:* Robust standard errors are in brackets, clustered at the 1872 census boundaries. All columns report the results from OLS regressions where the dependent variable is listed on the top of the column. In columns 1–4, the dependent variables refer to literacy rates computed for individuals aged within the range listed respectively in each column. In column 5, the dependent variable is measured in completed years of schooling. All columns include geographic controls (distance to the capital, latitude, longitude, elevation, and indicators for different types of soil—latosol, argisol, cambisol, and spondosol), presence of railway, demographic and economic characteristics 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 are computed according to the 1920 census boundaries.

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 induced a positive selection of foreigners to localities where the higher level of human capital and skills had no immediate economic return, at a moment when coffee production was still the dominant and most profitable economic activity (Dean 1969).

As a robustness test of the previous results, Table A3 in the online Appendix shows the same set of regressions using a continuous variable measuring the intensity of treatment: the share of foreigners in 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 lose some observations under this specification. Despite these limitations, all specifications show that an increase in the relative size of the settlement was associated with an increase in the literacy rate in 1920, while little effect is found in other dimensions.

### B. The Medium-Term and Long-Term Effects of State-Sponsored Settlements

We next document the persistent effects of the establishment of state-sponsored settlements. Table 5 shows results from estimating equation (1) for literacy rates in 1940 (medium-term effects) and literacy rate, years of schooling, and income per capita 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 of Table 5 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 municipal level or intergenerational transmission of human capital.

In columns 3 to 5 of Table 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. Finally, in column 6 of Table 5, we examine whether the establishment of settlements and the persistent increase in human capital also induced a differential development path. Our estimates show that income per capita in 2000 was 15 percent higher in municipalities that received a state-sponsored settlement when compared to other municipalities.

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 6, 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. Different from the sample used in Tables 4 and 5, these regressions use the 1872 map with the initial 88 municipalities. For 1920 and 1940, the municipalities are aggregated into the original 88 municipalities. We use as a dependent variable the literacy rate, which is the only educational outcome that can be compared across censuses. We restrict the analysis to 1872, 1920, and 1940, since by 2000, differences across municipalities were not reflected anymore on literacy but on completed years of schooling (see Table 5).

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 socioeconomic 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,

TABLE 6—MID-TERM AND LONG-TERM EFFECTS OF SETTLEMENTS ON HUMAN CAPITAL: PANEL-DATA SPECIFICATIONS

	Dependent variable		
	Literacy rate		
	(1)	(2)	(3)
Settlement × 1920	0.044 [0.024]	0.043 [0.024]	0.041 [0.024]
Settlement × 1940	0.097 [0.037]	0.086 [0.031]	0.068 [0.031]
Municipality fixed effect	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes
Geography times year effects	No	Yes	Yes
Demographics times year effects	No	No	Yes
Observations	264	264	264
Municipalities	88	88	88
$R^2$	0.748	0.799	0.810

*Notes:* Robust standard errors are in brackets. In all columns, the dependent variable is the literacy rate (share of literate individuals). Dependent and independent variables for 1872, 1920, and 1940 are computed according to the 1872 census boundaries. All specifications include municipality fixed effects and year fixed effects. Column 2 adds interaction terms between year 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, and share of children attending school).

these results are consistent with the cross-sectional estimates presented in Tables 4 and 5 showing that state-sponsored settlements led to a significant increase in literacy rates in the short and medium run.

## V. Mechanisms of Persistence

### A. School Inputs

Tables 4 to 6 show that state-sponsored settlements represented a positive shock to localities and the effects of the shock lasted until 2000. We next document how this initial shock persisted through time. Table 7 presents a series of cross-sectional 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 of Table 7, we show that settlement municipalities 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 of Table 7, however, 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 of Table 7, 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

TABLE 7—THE EFFECTS OF SETTLEMENTS ON SCHOOL INPUTS AND SCHOOL ATTENDANCE

Dependent variable	Schools per school-aged child (1)	Teachers per school-aged child (2)	Share children attending school (3)
<i>Panel A. 1920 census</i>			
Settlement	0.222 [0.108]	5.053 [1.874]	NA
$R^2$	0.090	0.373	
<i>Panel B. 1940 census</i>			
Settlement	-0.251 [0.294]	NA	0.080 [0.025]
$R^2$	0.133		0.327
<i>Panel C. 2000 census</i>			
Settlement	-0.622 [0.721]	11.235 [4.601]	0.006 [0.003]
$R^2$	0.54	0.045	0.354
Geographic controls	Yes	Yes	Yes
Characteristics 1872	Yes	Yes	Yes
Observations	202	202	202

*Notes:* Robust standard errors are in brackets, clustered at the 1872 census boundaries. 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 are computed as the total number of schools (1) or teachers (2) per school-aged children  $\times$  1,000. 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 characteristics 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.

close to universal in the state of São Paulo (data on school enrollment were not available from the 1920 census).

### B. Structural Transformation

We next 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. In Table 8, 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 of Table 8 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 4 indicates, this initial shift in



TABLE 8—THE LONG-RUN EFFECTS OF SETTLEMENTS ON EMPLOYMENT AND STRUCTURAL TRANSFORMATION

Dependent variable	Share emp. agriculture (1)	Share emp. manufacturing (2)	Share emp. retail and services (3)
<i>Panel A. 1920 census</i>			
Settlement	-0.121 [0.039]	0.072 [0.024]	0.048 [0.017]
Adjusted $R^2$	0.457	0.390	0.503
<i>Panel B. 1940 census</i>			
Settlement	-0.165 [0.046]	0.092 [0.027]	0.073 [0.028]
Adjusted $R^2$	0.496	0.472	0.396
<i>Panel C. 2000 census</i>			
Settlement	-0.053 [0.019]	0.006 [0.016]	0.048 [0.017]
Adjusted $R^2$	0.391	0.315	0.280
Observations	202	202	202
Geographic controls	Yes	Yes	Yes
Characteristics 1872	Yes	Yes	Yes

*Notes:* Robust standard errors are in brackets, clustered at the 1872 census boundaries. 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 characteristics 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 are computed according to the 1920 census boundaries.

employment toward manufacturing and services was not reflected on higher urbanization, higher wages, or higher value of land. It was simply a direct result 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).

The process of structural transformation deepened and became 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 of Table 8 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 online Appendix Table A4).

### C. Agglomeration

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.<sup>19</sup>

In Table 9, we examine the effect of state-sponsored settlements on years of schooling by cohort of birth and place of birth using data from the 2000 census. We use data from the census and split the sample between individuals born in the respective municipality and those born in other localities. We then estimate the effects of the state-sponsored settlements across cohorts for both types of individuals. For purposes of comparison, in panel A of Table 9 we show the results for individuals born in the municipality, which corroborate the findings from Tables 3 and 4. In column 1 of Table 9, 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 1920–1929 to 1960–1969 (columns 2 to 6 of Table 9), 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 1920–1929 cohort to 15 percent for the 1960–1969 cohort.

To assess the agglomeration effects induced by the settlements, in panel B of Table 9, we look at how they affected the schooling level of domestic 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 municipalities where they live. Two hypotheses underlie this exercise and our interpretation of its results: first, individuals not born in the settlement municipality where they live are more likely to have been born in municipalities without settlements; and second, older cohorts of individuals not born in the municipality where they live 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 of Table 9, we see that immigrants to settlement municipalities have, on average, 0.54 more year of schooling when

<sup>19</sup>For instance, Ciccone and Hall (1996), Ciccone (2002), and Combes et al. (2010) show that employment density positively affects labor productivity at the regional level in the United States, 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.

TABLE 9—PERSISTENT EFFECTS OF SETTLEMENTS ON YEARS OF SCHOOLING, BY COHORTS OF INDIVIDUALS

Dep. variable	All cohorts	1920–1929	1930–1939	1940–1949	1950–1959	1960–1969
Years of schooling	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A. Individuals born in the municipality</i>						
Settlement	0.883 [0.229]	0.655 [0.236]	0.759 [0.225]	0.941 [0.277]	0.927 [0.303]	1.022 [0.200]
Mean dep. var.	6.63	2.73	3.54	4.64	6.10	7.04
Adj. $R^2$	0.453	0.334	0.415	0.397	0.384	0.451
<i>Panel B. Individuals not born in the municipality</i>						
Settlement	0.544 [0.147]	0.332 [0.174]	0.285 [0.184]	0.534 [0.191]	0.541 [0.198]	0.765 [0.177]
Mean dep. var.	5.79	3.02	3.61	4.56	5.85	6.51
Adj. $R^2$	0.329	0.275	0.272	0.265	0.214	0.202
Observations	202	202	202	202	202	202
Geography	Yes	Yes	Yes	Yes	Yes	Yes
Charact. 1872	Yes	Yes	Yes	Yes	Yes	Yes

*Notes:* Robust standard errors are in brackets, clustered at the 1872 census boundaries. 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 micro data 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 characteristics 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 are computed according to the 1920 census boundaries.

compared to immigrants to other municipalities. Columns 2 to 6 of Table 9 present the effect of the settlements on the characteristics of immigrants 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 panel A of Table 9, 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. In addition, contrary to the results from panel A, 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 of state-sponsored settlements also attracted, over time, more educated individuals from within Brazil. We interpret this evidence to be consistent with agglomeration effects that increase income per capita in the long run.

#### D. Culture and Religion

Up to this point, we have documented that state-sponsored settlements brought positively selected foreigners to some municipalities in the state of São Paulo. These foreigners were more literate, but they might as well have had other traits that were

important for long-run economic development. Specific skills, culture, or social norms associated with these groups could also be related to schooling outcomes and development. Although we have limited ways to measure these characteristics, we have information on nationalities and religion in the 1920 census. In Table 10, we reestimate the regressions presented in Table 5 controlling for the nationality and religion of immigrants. We control for the 1920 percentage of foreigners by major national groups—German, Spaniard, Italian, Portuguese, and Japanese—and for the 1920 percentage of Catholic individuals. The coefficients on state-sponsored settlements remain almost unaffected, despite the fact that the percentage of some nationalities and Catholics appear as positively associated with literacy rates and income per capita. The results suggest that the long-term effects of the presence of state-sponsored settlements were not driven by cultural factors associated with the nationality or religion of immigrants. It remains true, nevertheless, that other unobservable productive attributes correlated with education, such as previously accumulated manufacturing skills, could have been a contributing factor.<sup>20</sup>

## VI. Concluding Remarks

In the end of the nineteenth and beginning of the twentieth century, public authorities established a number of official settlement colonies throughout the state of São Paulo, Brazil. 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 its relationship with long-term development: state-sponsored settlements were not too different from other areas in the state, but for the fact that they received a different pool of immigrants.

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 nineteenth 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, areas that had received settlements were not noticeably different from other areas in terms of density, concentration of land, agricultural productivity, price of land, or wages. 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

<sup>20</sup>It is worth mentioning that we do not find differences in the incidence of entrepreneurship across settlement and non-settlement municipalities when looking at data from the 2000 census (either by cohort or average). So results do not seem to be driven by unobservable traits—cultural or otherwise—associated with entrepreneurship. These results are presented in Table A5 in the online Appendix. Also, if we include an additional control for the political influence of agricultural groups in 1905, results remain unchanged in Table 10, indicating that the initial distribution of political power across agriculture and manufacturing does not seem to be a relevant mechanism behind our results.

TABLE 10—THE EFFECTS OF SETTLEMENTS ON LITERACY RATES, CONDITIONAL ON NATIONAL IDENTITIES AND RELIGION

Dependent variable	Literacy rate in 1920 (1)	Literacy rate in 1940 (2)	Years of schooling 2000 (3)	Income per capita 2000 (4)
Settlement	0.091 [0.026]	0.121 [0.032]	0.521 [0.149]	0.182 [0.054]
Share of Germans	0.206 [0.080]	0.182 [0.100]	0.061 [0.528]	−0.320 [0.240]
Share of Spaniards	0.066 [0.066]	0.113 [0.066]	0.602 [0.356]	0.208 [0.149]
Share of Italians	0.129 [0.055]	0.211 [0.062]	0.540 [0.332]	0.394 [0.155]
Share of Portuguese	0.104 [0.059]	0.242 [0.086]	1.081 [0.378]	0.291 [0.134]
Share of Japanese	0.020 [0.070]	0.034 [0.077]	0.462 [0.411]	0.078 [0.184]
Share of Catholics	0.081 [0.024]	0.064 [0.039]	0.128 [0.307]	−0.013 [0.091]
Geography	Yes	Yes	Yes	Yes
Charact. 1872	Yes	Yes	Yes	Yes
Observations	202	202	202	202
Adj. $R^2$	0.478	0.435	0.368	0.47

*Notes:* Robust standard errors are in brackets, clustered at the 1872 census boundaries. In columns 1 and 2, the dependent variable is literacy rate in 1920 (1940) built on data originally from the 1920 (1940) census. In columns 3 and 4, the dependent variable is, respectively, average years of schooling and income per capita, both 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 1–4 add the shares of individuals from different nationalities computed as the total number of individuals from the given nationality relative to total population in the municipality in 1920, as well as the share of Catholics computed as the number of Catholics relative to total population in 1920. Data on nationalities and number of Catholics are 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 characteristics 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 are computed according to the 1920 census boundaries.

capita 15 percent higher than the remainder of the state. We also show that, throughout the twentieth 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 toward 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. The pattern we uncover is consistent with the idea that state-sponsored settlements brought more skilled foreigners—along educational and possibly other unobservable dimensions—to localities where initially these skills had no immediate economic return. At that point, coffee production—intensive in unskilled agricultural workers—was still the dominant and most

profitable economic activity. But as soon as industrialization took hold, these skills started representing an economic advantage that ended up materializing in increased income per capita in the long run.

Our results document persistence in educational investments and outcomes over a long period of time within the context of homogenous institutions represented by the Brazilian state of São Paulo. Even in this setting, we are able to show that early differences in human capital—possibly also associated with individual heterogeneity along other unobservable dimensions—persisted through time and were correlated with long-run differences in income per capita.

## REFERENCES

- Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson.** 2014. “A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration.” *Journal of Political Economy* 122 (3): 467–506.
- Abreu, Marcelo de P., and Alfonso S. Bevilaqua.** 1996. “Brazil as an export economy, 1880–1930.” Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio) Working Paper 363.
- Acemoglu, Daron, Francisco A. Gallego, and James A. Robinson.** 2014. “Institutions, Human Capital, and Development.” *Annual Review of Economics* 6: 875–912.
- Acemoglu, Daron, Simon Johnson, and James A. Robinson.** 2001. “The Colonial Origins of Comparative Development: An Empirical Investigation.” *American Economic Review* 91 (5): 1369–1401.
- Acemoglu, Daron, Simon Johnson, and James A. Robinson.** 2005. “Institutions as a Fundamental Cause of Long-Run Growth.” In *Handbook of Economic Growth*, Vol. 1A, edited by Philippe Aghion and Steven N. Durlauf, 385–472. Amsterdam: North-Holland.
- Baldwin, Richard E., and Philippe Martin.** 2004. “Agglomeration and Regional Growth.” In *Handbook of Regional and Urban Economics*, Vol. 4, edited by J. Vernon Henderson and Jacques-François Thisse, 2671–2711. Amsterdam: North-Holland.
- Bandiera, Oriana, Myra Mohnen, Imran Rasul, and Martina Viarengo.** 2015. “Nation-Building Through Compulsory Schooling During the Age of Mass Migration.” <http://sticerd.lse.ac.uk/dps/eopp/eopp57.pdf>.
- Barro, Robert J.** 1991. “Economic Growth in a Cross Section of Countries.” *Quarterly Journal of Economics* 106 (2): 407–43.
- Bassanezi, Maria Silva C. Beozzo, Ana Silvia Volpi Scott, Carlos de Almeida Prado Bacellar, and Oswaldo Mário Serra Truzzi.** 2008. *Atlas da Imigração Internacional em São Paulo, 1850–1950*. São Paulo: Editora UNESP.
- Becker, Sascha O., Erik Hornung, and Ludger Woessmann.** 2011. “Education and Catch-Up in the Industrial Revolution.” *American Economic Journal: Macroeconomics* 3 (3): 92–126.
- Ciccone, Antonio.** 2002. “Agglomeration effects in Europe.” *European Economic Review* 46 (2): 213–27.
- Ciccone, Antonio, and Robert E. Hall.** 1996. “Productivity and the Density of Economic Activity.” *American Economic Review* 86 (1): 54–70.
- Ciccone, Antonio, and Elias Papaioannou.** 2009. “Human Capital, the Structure of Production, and Growth.” *Review of Economics and Statistics* 91 (1): 66–82.
- Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux.** 2010. “Estimating Agglomeration Economies with History, Geology, and Worker Effects.” In *Agglomeration Economics*, edited by Edward L. Glaeser, 15–66. Chicago: University of Chicago Press.
- Dean, Warren.** 1969. *The Industrialization of São Paulo, 1880–1945*. Austin: University of Texas Press.
- de Carvalho Filho, Irineu, and Renato P. Colistete.** 2010. “Education Performance: Was It All Determined 100 Years Ago? Evidence From São Paulo, Brazil.” Munich Personal RePEc Archive (MPRA) Paper 24494.
- de Carvalho Filho, Irineu, and Leonardo Monasterio.** 2012. “Immigration and the origins of regional inequality: Government-sponsored European migration to southern Brazil before World War I.” *Regional Science and Urban Economics* 42 (5): 794–807.
- Dell, Melissa.** 2010. “The Persistent Effects of Peru’s Mining Mita.” *Econometrica* 78 (6): 1863–1903.
- Dittmar, Jeremiah, and Ralf R. Meisenzahl.** 2016. “State Capacity and Public Goods: Institutional Change, Human Capital and Growth in Early Modern Germany.” Centre for Economic Performance (CEP) Discussion Paper 1418.

- Donaldson, Dave.** Forthcoming. "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure." *American Economic Review*.
- Droller, Federico.** Forthcoming. "Migration, Population Composition and Long Run Economic Development: Evidence from Settlements in the Pampas." *Economic Journal*.
- Duranton, Gilles, and Diego Puga.** 2004. "Micro-Foundations of Urban Agglomeration Economies." In *Handbook of Regional and Urban Economics*, Vol. 4, edited by J. Vernon Henderson and Jacques-François Thisse, 2063–2117. Amsterdam: North-Holland.
- Easterly, William, and Ross Levine.** 2016. "The European origins of economic development." *Journal of Economic Growth* 21 (3): 225–57.
- Fernández, Raquel.** 2011. "Does Culture Matter?" In *Handbook of Social Economics*, Vol. 1, edited by Jess Benhabib, Alberto Bisin, and Matthew O. Jackson, 481–510. Amsterdam: North-Holland.
- Fourie, Johan, and Dieter von Fintel.** 2014. "Settler skills and colonial development: The Huguenot wine-makers in eighteenth-century Dutch South Africa." *Economic History Review* 67 (4): 932–63.
- Furtado, Celso.** 1959. *Formação Econômica do Brasil*. São Paulo: Companhia Editora Nacional.
- Gadelha, Regina Maria A. F.** 1982. "Os Núcleos Coloniais e o Processo de Acumulação Cafeeira (1850–1920): Contribuições ao Estudo da Colonização de São Paulo." PhD diss. Universidade de São Paulo.
- Galor, Oded, Omer Moav, and Dietrich Vollrath.** 2009. "Inequality in Landownership, the Emergence of Human-Capital Promoting Institutions, and the Great Divergence." *Review of Economic Studies* 76 (1): 143–79.
- Glaeser, Edward L., José A. Scheinkman, and Andrei Shleifer.** 1995. "Economic growth in a cross-section of cities." *Journal of Monetary Economics* 36 (1): 117–43.
- Glaeser, Edward L., Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer.** 2004. "Do Institutions Cause Growth?" *Journal of Economic Growth* 9 (3): 271–303.
- Guiso, Luigi, Paola Sapienza, and Luigi Zingales.** 2016. "Long-Term Persistence." *Journal of the European Economic Association* 14 (6): 1401–36.
- Hornbeck, Richard.** 2010. "Barbed Wire: Property Rights and Agricultural Development." *Quarterly Journal of Economics* 125 (2): 767–810.
- Hornbeck, Richard.** 2012. "The Enduring Impact of the American Dust Bowl: Short- and Long-Run Adjustments to Environmental Catastrophe." *American Economic Review* 102 (4): 1477–1507.
- Hornbeck, Richard, and Suresh Naidu.** 2014. "When the Levee Breaks: Black Migration and Economic Development in the American South." *American Economic Review* 104 (3): 963–90.
- Hornung, Erik.** 2014. "Immigration and the Diffusion of Technology: The Huguenot Diaspora in Prussia." *American Economic Review* 104 (1): 84–122.
- Huillery, Elise.** 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, Luiza Horn.** 2001. *Imigração e Colonização: Legislação de 1747–1915*. Caxias do Sul: Editora da Universidade de Caxias do Sul.
- Iranzo, Susana, and Giovanni Peri.** 2009. "Schooling Externalities, Technology, and Productivity: Theory and Evidence from U.S. States." *Review of Economics and Statistics* 91 (2): 420–31.
- Kreutz, Lúcio.** 2000. "Escolas comunitárias de imigrantes no Brasil: Instâncias de coordenação e estruturas de apoio." *Revista Brasileira de Educação* 15: 159–76.
- Landes, David S.** 1999. *The Wealth and Poverty of Nations: Why Some Are So Rich and Some So Poor*. New York: W. W. Norton and Company.
- Lowes, Sara, Nathan Nunn, James A. Robinson, and Jonathan Weigel.** 2015. "The Evolution of Culture and Institutions: Evidence from the Kuba Kingdom." National Bureau of Economic Research (NBER) Working Paper 21798.
- Lucas, Robert E., Jr.** 1988. "On the mechanics of economic development." *Journal of Monetary Economics* 22 (1): 3–42.
- Martins, José de Souza.** 1973. *A Imigração e a Crise do Brasil Agrário*. São Paulo: Pioneira.
- Maschio, Elaine Cátia Falcade.** 2005. "A Experiência Escolar entre Imigrantes Italianos no Paraná." [http://www.histedbr.fe.unicamp.br/acer\\_histedbr/jornada/jornada6/trabalhos/602/602.pdf](http://www.histedbr.fe.unicamp.br/acer_histedbr/jornada/jornada6/trabalhos/602/602.pdf).
- Moretti, Enrico.** 2004. "Human Capital Externalities in Cities." In *Handbook of Regional and Urban Economics*, Vol. 4, edited by J. Vernon Henderson and Jacques-François Thisse, 2243–91. Amsterdam: North-Holland.
- North, Douglass C.** 2009. *Institutions, Institutional Change and Economic Performance*. Political Economy of Institutions and Decisions. Cambridge: Cambridge University Press.
- Nunn, N.** 2011. "Christians in Colonial Africa." [http://www.econ.yale.edu/~egcenter/Nunn\\_Paper.pdf](http://www.econ.yale.edu/~egcenter/Nunn_Paper.pdf).
- Paiva, Odair da Cruz.** 1993. "Intervenção Governamental e Reordenação Fundiária." Master's diss. Universidade Estadual de Campinas.

- Petrone, Maria Thereza Schorer.** 1982. *O Imigrante e a Pequena Propriedade, 1824–1930*. São Paulo: Brasiliense.
- Putterman, Louis, and David N. Weil.** 2010. “Post-1500 Population Flows and the Long-Run Determinants of Economic Growth and Inequality.” *Quarterly Journal of Economics* 125 (4): 1627–82.
- Rocha, Rudi, Claudio Ferraz, and Rodrigo R. Soares.** 2017. “Human Capital Persistence and Development: Dataset.” *American Economic Journal: Applied Economics*. <https://doi.org/10.1257/app.20150532>.
- Rosenthal, Stuart S., and William C. Strange.** 2004. “Evidence on the Nature and Sources of Agglomeration Economies.” In *Handbook of Regional and Urban Economics*, Vol. 4, edited by J. Vernon Henderson and Jacques-François Thisse, 2119–71. Amsterdam: North-Holland.
- Shapiro, Jesse M.** 2006. “Smart Cities: Quality of Life, Productivity, and the Growth Effects of Human Capital.” *Review of Economics and Statistics* 88 (2): 324–35.
- Siriani, Sílvia Cristina Lambert.** 2005. “Os descaminhos da Imigração alemã para São Paulo no século XIX—aspectos políticos.” *Almanack Braziliense* 1 (2): 91–100.
- Squicciarini, Mara P., and Nico Voigtländer.** 2015. “Human Capital and Industrialization: Evidence from the Age of Enlightenment.” *Quarterly Journal of Economics* 130 (4): 1825–83.
- Summerhill, William R., III.** 2003. *Order against Progress: Government, Foreign Investment, and Railroads in Brazil, 1854–1913*. Stanford: Stanford University Press.
- Summerhill, William R., III.** 2010. “Colonial Institutions, Slavery, Inequality, and Development: Evidence from São Paulo, Brazil.” Munich Personal RePEc Archive (MPRA) Working Paper 22162.
- Valencia Caicedo, Felipe.** 2014. “The Mission: Economic Persistence, Human Capital Transmission and Culture in South America.” [http://sites.bu.edu/neudc/files/2014/10/paper\\_320.pdf](http://sites.bu.edu/neudc/files/2014/10/paper_320.pdf).
- Wantchekon, Leonard, Marko Klašnjaz, and Natalija Novta.** 2015. “Education and Human Capital Externalities: Evidence from Colonial Benin.” *Quarterly Journal of Economics* 130 (2): 703–57.
- Weber, Max.** 1930. *The Protestant Ethic and the Spirit of Capitalism*. London: Allen and Unwin.