

# The Persistence of Human Capital Over Time: Evidence from a Positive Immigration Shock

Anoushka Khatri

The Harker School, San Jose, CA, USA

## ABSTRACT

In this paper, I study the long-run effects of a positive shock to the labor force in Sao Paulo. Using a dataset based on Brazil and an estimation strategy based on regression analysis, I show both a positive direct and indirect effect of settlement on human capital. Being a recipient of highly skilled immigrants in 1872 is found to be positively associated with a rise in literacy rates, as well as higher per capita income and access to running water.

## Introduction

In the last quarter of the 19th century, immigration into Brazil grew quickly. The end of slavery in Brazil brought about a severe need for labor, and, at the same time, emigration from Europe surged. This led to many European immigrants (many with a higher education level than locals) quickly being settled into municipalities to fill the need for workers. As immigration increased, the Brazilian government saw state-sponsored settlement as a way to attract even more immigrants, and began immediately. In this paper, the researchers examine whether this positive supply shock of highly skilled European immigrants in 1872 affected human capital in 2010, seeing if the results of Rocha, Ferraz, and Soares (2017) hold after ten years. To accomplish this, the researchers looked specifically at literacy rates for different age groups, income per capita, and access to running water.

This paper contributes directly to literature analyzing the persistence of human capital over time. Similarly to Acemoglu, Johnson, and Robinson (2001), the researchers looked at the impact of institutions on human capital. However, while Acemoglu, Johnson, and Robinson (2001), look at a variety of European colonies, the researchers focused on Sao Paulo and its municipalities. Rocha, Ferraz, and Soares (2017) also look at the effect of European migration into Brazil, and how they have affected human capital in 2000. In this paper, the researchers examined whether the results of Rocha, Ferraz, and Soares (2017) hold in 2010.

In addition, this paper contributes to literature analyzing the effects of immigration on long-run economic development. Similarly to Putterman and Weil (2010), the researchers looked at the effects of immigration on economic development. However, this paper looks at a single state, and examines municipality-level data.

The remainder of the paper is organized as follows. Section 2 discusses the historical background, explaining the circumstances behind the sudden flow of European immigrants into Sao Paulo, and Section 3 introduces the data used in this paper, and the sources of the data. Section 4 presents the analysis and the main results of the paper, and Section 5 concludes.

## Historical Background

In the early 1820s, soon after Brazil gained independence from Portugal, the Brazilian government wanted to attract more immigrants, but only immigrants of a specific type. There was international pressure on Brazil to end the slave trade, but the Brazilian government needed more workers to support the growing economy, and thus a desire for

immigrants willing to settle in rural areas and contribute to the national economy grew (Lesser 2013). In 1824, Brazil's immigration laws were amended to allow for some religious freedom to attract more immigrants.

In 1850, the international slave trade was abolished in Brazil. With this law came an urgent need for labor in Sao Paulo, as coffee cultivation was expanding rapidly, and therefore needed more workers to sustain the economic boom the region was currently experiencing. At the same time, Europeans immigrated to Brazil due to better work opportunities and difficulties in Europe. Therefore, due to the demand for labor and the ban on the international slave trade, the idea of state-sponsored colonies began to gain traction in the Brazilian government (Lesser 2013).

From 1824 to 1918, the Brazilian government subsidized the settlement of European immigrants in Brazilian municipalities, especially in rural areas (Hall 1969). Management of public land was given to the states, and state governments used the promise of plots of land to attract more immigration.

Sao Paulo specifically became one of the main regions of immigration, with an Immigrant's Hostel (*Hospedaria dos Imigrantes*) built in 1886. Many of the immigrants who came to hostels like these were immediately entered into sharecropping contracts with coffee farmers, but some ended up in state-sponsored settlements.

Eventually, 28 state-sponsored settlements were created in Sao Paulo. This paper examines the effect of these 28 settlements (distributed across 10 percent of Sao Paulo municipalities) on human capital.

## Empirical Strategy

In this paper, the researchers were interested in estimating the effect of settlement in 1872 on literacy rates, per capita income and access to running water. To do this, they considered a cross-section regression at the municipality level, estimating the following specification:

$$Outcome_i = \alpha + \beta_1 Settlement_i + Controls'_i + \gamma + \epsilon_i \quad (1)$$

Where  $Outcome_i$  is one of the three outcomes considered in this paper: literacy rates, per capita income and access to running water.  $Settlement_i$  is a dummy variable taking value of 1 when municipality  $i$  receives the settlement in 1872, and 0 otherwise.  $Controls_i$  is a vector of municipality level controls, evaluated from 1872, including both geographic controls (such as distance to the capital city, latitude, and longitude) and socioeconomics controls (such as literacy, share of foreigners, and coffee production).

$\epsilon_i$  is the error term of the specification: all the factors that remain unobserved and are thus not controlled for, but that might still have an effect on the outcome variables. Lastly,  $\alpha$  is the intercept of the regression, and  $\beta_1$  and  $\gamma$  are the coefficients to be estimated.

## Data

The researchers used data from several sources, and the dataset includes information on socioeconomic characteristics and geographical data.

The census data is pulled from the population censuses that took place in 1872, 1920, 1940, 2000, and 2010. This data was mainly used to find literacy rates, per capita income, and access to running water, along with other socioeconomic characteristics used as controls.

The data on railroads and geography is from Ipeadata (latitude, longitude, elevation, and distance from the capital), Embrapa Solos (soil types), and Estações Ferroviarias do Brasil (railroads).

Ipeadata is a database with data from the Institute of Applied Economic Research (*Instituto de Pesquisa Economica Aplicada*), which is a research organization led by the Brazilian government. The IPEA conducts economics studies to assist the Brazilian government in improving the social, economic, and structural development of the country.

*Embrapa Solos* (Embrapa Soils) is run by Embrapa, which is the Brazilian Agricultural Research Corporation, and is affiliated with the Brazilian Ministry of Agriculture. Embrapa Soils conducts research on soil and soil-environment interactions.

*Estações Ferroviárias do Brasil* (Train Stations of Brazil) is a website that documents the history of railway stations across Brazil. The author of the website, Ralph Mennucci Giesbrecht, is a member of the Brazilian Association of Railway Preservation (*Associação Brasileira de Preservação Ferroviária*) and the Historical and Geographical Institute of São Paulo (*Instituto Histórico e Geográfico de São Paulo*).

For some of the data, the researchers had to complete some manipulations so that they could work with the data. To derive literacy rates, they subtracted each illiteracy rate from 100. Therefore, it is important to note that they made the assumption that everyone who is not illiterate is literate. In addition, the researchers adjusted the per capita income in 2000 by the consumer price index with a base year of 2010.

Table 4 presents the summary statistics of the geographic variables of each municipality, mainly used as control variables, and the socioeconomic variables for 1872 considered in this study.

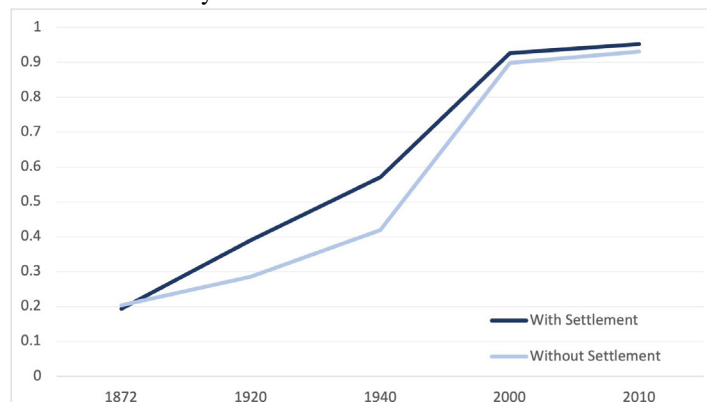
Since the state-sponsored settlement process started in 1872, the researchers did not separate the socioeconomic variables by settlement status, because the average values would be the same.

For the 1872 variables, comparison does not help. This is because state-sponsored settlement started in 1872, so the average values have not yet been affected by immigration into Sao Paulo because no time has passed.

When examining the 2000 variables, there is much more difference seen between the settled and non-settled municipalities. For instance, the average adjusted income per capita is 729 R\$ for settled municipalities and is 569 R\$ for non-settled municipalities, showing a clear difference and indicating an effect of settlement on average adjusted income per capita. In addition, the average share of literacy and average years of schooling are both higher for settled municipalities, showing an effect of settlement on both variables. Also, the share of agricultural employment is lower in municipalities with settlements, indicating a positive effect of settlement.

There is a similar effect seen with several 2010 variables. Average income per capita is higher in municipalities with state-sponsored settlements, and average share of literacy is also higher. In addition, the average share of houses with running water is higher in municipalities with state-sponsored settlements.

**Fig 1.** Growth in Literacy Rates.



## Results

### The Direct Effect of Settlement on Human Capital

The researchers started the analysis by examining the direct effect of state-sponsored settlement on human capital. In this case, the best way to estimate the effect on human capital is to look at the effect on literacy rates, as change in literacy is a good measure of change in human capital.

The first variable the researchers examined was the effect of settlement on the literacy rate of the population aged 15 to 24 years, because Rocha, Ferraz, and Soares (2017) examined the effect on the literacy rate of a similar age group (15 to 19 years of age).

**Table 1.** 2010 Data.

Variables	Municipalities with Settlements			Municipalities without Settlements		
	Mean	Min	Max	Mean	Min	Max
Per capita income (adjusted to 2010 CPI) in R\$	773.550	463.000	1180.000	635.742	303.000	1575.000
Literacy rate (age 15+)	0.953	0.890	0.972	0.931	0.834	0.971
Literacy rates for ages 15-24	0.988	0.970	0.993	0.987	0.971	0.997
Literacy rates for ages 25-39	0.978	0.933	0.989	0.969	0.913	0.990
Literacy rates for ages 40-59	0.952	0.885	0.974	0.930	0.840	0.978
Literacy rates for ages 60+	0.844	0.690	0.920	0.785	0.583	0.910
Share of houses with running water	0.920	0.573	0.991	0.845	0.349	0.992

Notes: This table shows the main variables examined using regression analysis. In addition, per capita income was adjusted using 2010 CPI for ease of comparison.

The researchers examined the effect of settlement using regression analysis, where they controlled for geographic variables (shown in Table 4). When they performed the regression analysis (results seen in Table 1), the estimated coefficient on the variable “settlement in 1872” was statistically insignificant (the p-value is above 0.10) and small in magnitude. In particular, the estimated effect of having a settlement on the literacy rates in 2010 for the young cohort of individuals was 0.1%. Statistically, this number cannot be distinguished from 0, as 0 is part of the confidence interval of this estimate [-0.001;0.003]. Comparing this estimate with the one present in Rocha, Ferraz, and Soares (2017), which is based on literacy rates of individuals aged 15-19 in 2000, highlights that the absence of an effect is still present, although the coefficient can be now more precisely estimated, as there was a lower standard error here (0.001) compared to that in the paper (0.002).

Since the estimated coefficient using the literacy rate of the youngest cohort is statistically insignificant, the researchers decided to use literacy rates for older age groups in the population to examine the effects. When they looked at the literacy rates for the population aged 25 to 39 years, however, the estimated coefficients were statistically insignificant, as the p-values were above 0.10, and were also small in magnitude. Therefore, they then looked at the literacy rate for the population aged 40 to 59 years. When the researchers performed the analysis, the estimated coefficient on the variable “settlement in 1872” was statistically significant at the 5% confidence level because the p-value is below 0.05. The effect of being in a municipality that had a settlement was estimated to be 1.4%. This number is statistically significantly different from 0, as 0 is not part of the confidence interval of this estimate.<sup>1</sup>

This is an improvement compared to the results presented in the paper (Table 5 of Rocha, Ferraz, and Soares 2017), since the researchers also considered a different subset of the population. Not only did they analyze a more recent time period (2010 instead of 2000), but they also considered a subset of the population (40-59 years) that allows for more variation across municipalities. This variation allows for the results to be statistically significant. Rocha, Ferraz, and Soares (2017) instead considered the youngest cohort and, as a result, fails to identify an effect, since there is not much variation in literacy rates across municipalities.

<sup>1</sup> The estimated confidence interval is [0.004; 0.024].

Then, the researchers looked at the literacy rates of the population aged 60 years and over. After completing the analysis, the estimated coefficient on the settlement variable was statistically significant at the 5% confidence level because the p-value was below 0.05, and the effect of being in a municipality that has a settlement was estimated to be 3.2%. This number is statistically significantly different from 0, as 0 is not part of the confidence interval of this estimate.<sup>2</sup>

Similar to the previous analysis for the literacy rates of the population aged 40 to 59 years, this analysis is an improvement compared to the results presented by Rocha, Ferraz, and Soares (2017). The researchers' consideration of a more recent time period and a different, older subset of the population allowed for more variation across municipalities, which leads to significant results.

Therefore, analyzing and comparing the dependent variables of literacy rates for each age group highlighted that the effect of state-sponsored settlement is strongest for the oldest cohorts. This makes sense, because as the age groups become younger, literacy rates in settled and non-settled municipalities become more even due to modernization across all municipalities, causing the effects of settlement to be less prevalent.

**Table 2.** The Effects on Literacy Rates.

Variables	Coefficient	P-value	Lower 90%	Upper 90%
Literacy rate for population aged 15+	0.012**	0.016	0.004	0.020
Literacy rates for ages 15-24	0.001	0.484	-0.001	0.003
Literacy rates for ages 25-39	0.005	0.152	-0.001	0.010
Literacy rates for ages 40-59	0.014**	0.021	0.004	0.024
Literacy rates for ages 60+	0.032**	0.036	0.007	0.056

Notes: The table reports the estimated coefficients and the corresponding p-values. \*\*\*, \*\*, and \* indicate statistical significance at the 1%, 5%, and 10% levels, respectively. All regressions include geographic controls, such as distance from the capital, latitude, longitude, altitude, and soil type. Regressions also include socioeconomic variables, such as whether a municipality was settled in 1872, presence of railways in 1872, share of children in school in 1872, population in 1872, share of foreigners in 1872, share of slaves in 1872, and literacy rates in 1872.

### The Indirect Effect of Settlement on Human Capital

To examine the indirect effect of state-sponsored settlement on human capital, the researchers looked at per capita income and access to running water. Both of these variables reflect how changes to human capital could cause changes to economic factors.

The researchers first considered per capita income in both 2000 and 2010. To ensure that the comparison between the two variables are accurate, they adjusted the per capita income in 2000 according to the 2010 Brazilian consumer price index (CPI). The mean value of per capita income in 2000 was 585 R\$, and the mean value of per capita income in 2010 was 649 R\$. Furthermore, the mean values of per capita incomes for 2000 and 2010 in municipalities that received immigrants were 729 R\$ and 774 R\$, respectively, and the mean values for municipalities that did not receive immigrants were 569 R\$ and 636 R\$, respectively. These data indicate that settlement in 1872 might influence per capita income, so the researchers decided to conduct regression analysis to confirm this.

First, the researchers analyzed the effect of settlement in 1872 on adjusted income per capita in 2000 (results seen in Table 3). After completing this analysis, the estimated coefficient on the settlement variable was statistically significant at the 5% level of significance because the p-value was below 0.05, and the effect of being in a settled

<sup>2</sup> The estimated confidence interval is [0.006; 0.056].

municipality was estimated to be an increase of 100 R\$. This number is statistically significantly different from 0, as 0 is not part of the confidence interval of this estimate.<sup>3</sup> In addition, the average percentage change in per capita income between a settled and a non-settled municipality was 17%.

Next, the researchers analyzed the effect of settlement in 1872 on income per capita in 2010, to see if the effect in 2000 still held. After completing this analysis, the estimated coefficient on the settlement variable was statistically significant at the 5% level of significance because the p-value was below 0.05, and the effect of being in a settled municipality was estimated to be an increase of 86 R\$. This number is statistically significantly different from 0, as 0 is not part of the confidence interval of this estimate.<sup>4</sup> While the estimated effect was lower than that in 2000, this could be accounted for due to the longer timeframe and the modernization of all municipalities, which led the per capita income of each municipality to become closer together, and the effect of settlement to be less prevalent. In addition, the researchers' estimates show that income per capita in 2010 was 13.3% higher in municipalities that received a state-sponsored settlement when compared to other municipalities.

Thus, state-sponsored settlement had a strong effect on per capita income in both 2000 and 2010, although the effect was slightly diminished in 2010 due to the longer timeframe and greater modernization.

The researchers then looked at access to running water in 2010. After completing their analysis, the estimated coefficient on the settlement variable was statistically significant at the 10% level of significance because the p-value was less than 0.1, and the estimated effect of living in a settled municipality on access to running water was estimated to be 5.5%. This number is statistically significantly different from 0, because 0 is not part of the confidence interval of this estimate.<sup>5</sup>

Therefore, state-sponsored settlement had a statistically significant effect on both per capita income and access to running water, which means the indirect effects of settlement on human capital match the direct effects.

**Table 3.** The Effects on Per Capita Income and Access to Running Water.

Variables	Coefficient	P-value	Lower 90%	Upper 90%
Per capita income in 2000	99.806***	0.001	49.903	149.712
Per capita income in 2010	86.387***	0.009	32.343	140.431
Share of houses with running water	5.449*	0.057	0.742	10.156

Notes: See note on Table 2.

## Conclusion

In this paper, the researchers looked at the conclusions of Rocha, Ferraz, and Soares (2017) to see if their results held in 2010. To do this, the researchers used regression analysis to analyze the effect of state-sponsored settlement on human capital, specifically examining the effect on literacy rates, per capita income, and access to running water. They found that state-sponsored settlement had a significant statistical effect on literacy rates in older age groups and has a significant effect on per capita income and access to running water.

Their results showed that state-sponsored settlement in 1872 had a significant effect on human capital in 2010, confirming the conclusions of Rocha, Ferraz, and Soares (2017) and documenting the persistence of human capital through an increase in literacy rates, per capita income, and access to running water.

From a more general perspective, these results mean that an influx of highly educated immigrants into a developing country could have a large effect on human capital.

<sup>3</sup> The estimated confidence interval is [49.903;149.711].

<sup>4</sup> The estimated confidence interval is [32.344;140.431].

<sup>5</sup> The estimated confidence interval is [0.742;10.156].

Several additional avenues for research appear worth exploring: Extending this research to other states in Brazil, could provide valuable insights into the persistence of human capital in Brazil as a whole. In addition, looking at immigration into Brazil in different time periods to examine the different impact each wave of immigration has could be significant. All these questions are left to further research.

## Acknowledgments

This paper has benefitted immensely from conversations with Dr. Andrea Bernini. I would also like to thank Stephen Turban, Dhruva Bhat, Prakriti Sharma, and Sara Ashbaugh for their guidance.

## References

- Acemoglu, D., S. Johnson, and J. A. Robinson (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review* 91(5), 1369–1401. <https://doi.org/10.1257/aer.91.5.1369>
- Lesser, J. (2013). *Immigration, Ethnicity, and National Identity in Brazil, 1808 to the Present*. Cambridge University Press.
- Hall, M. (1969). *The Origins of Mass Immigration in Brazil, 1871-1914*. Columbia University.
- Putterman, L. and D. N. Weil (2010). Post-1500 population flows and the long-run determinants of economic growth and inequality. *The Quarterly Journal of Economics* 125(4), 1627–1682. <https://doi.org/10.1162/qjec.2010.125.4.1627>
- Rocha, R., C. Ferraz, and R. R. Soares (2017). Human capital persistence and development. *American Economic Journal: Applied Economics* 9(4), 105–136. <https://doi.org/10.1257/app.20150532>
- Segal, U. A., D. Elliot, and N. S. Mayadas (2009). *Immigration Worldwide: Policies, Practices, and Trends*. Oxford University Press.