Commodity Booms, Slavery and Illiteracy: Historical Evidence from Brazilian Municipalities

Naercio Menezes-Filho [1] , Leandro Anazawa [2] , Bruno Proença[3],

Bruno Kawaoka Komatsu [4]

Abstract

This paper investigates the effect of commodity booms on illiteracy using panel data from Brazilian municipalities covering a period of more than a century, before and after the abolition of slavery and proclamation of republic. We construct a new time-varying and municipality-specific measure of potential export revenues, composed of a weighted average of each commodity's international price, with weights given by the municipality's suitability to produce each of the commodities. Our results shows that an increase of one standard deviation of potential exports leads to a reduction of 14 percentage points in the share of illiterates in the municipality. With regards to mechanisms, we show that increases in potential exports promote structural transformation and changes the racial composition in the municipality, which explains most of the negative effect of potential exports on illiteracy.

Key words: Illiteracy, commodity booms, economic history

[1] Corresponding author. Ruth Cardoso Professor of Economics at Insper, R. Quatá, 300, Vila Olímpia, São Paulo/SP, Brazil – CEP 04546-042. School of Business, Economics and Accounting of the University of São Paulo, Av. Prof. Luciano Gualberto, 908, Butantã, São Paulo/SP, Brazil – CEP 05508-010. E-mail: naercioamf@insper.edu.br

[2] School of Business, Economics and Accounting of the University of São Paulo, Ribeirão Preto campus, Av. Bandeirantes, 3900, Vila Monte Alegre, Ribeirão Preto/SP, Brazil – CEP 14040-905. E-mail: leandro.anazawa@usp.br

[3] Insper, R. Quatá, 300, Vila Olímpia, São Paulo/SP, Brazil – CEP 04546-042. E-mail: brunokk@insper.edu.br. Tel.: +55(11)4504-2741

[4] Insper, R. Quatá, 300, Vila Olímpia, São Paulo/SP, Brazil – CEP 04546-042. E-mail: brunokk@insper.edu.br. Tel.: +55(11)4504-2741

1 - Introduction

Rising incomes resulting from commodity booms can impact human capital investments through several mechanisms. Export revenues generate tax revenues that can be used to build schools, for example, and richer families can send their children to school and demand public investments in education. But rising profits in agriculture may also increase demand for labor and draw children away from school. The relationship between commodity booms and human capital investments is intermediated by political and economic institutions.[5] [6] This paper investigates how economic shocks resulting from the international trade of commodities impacted the share of illiterates in Brazil between 1872 and 1960. We also highlight a new mechanism explaining this relationship, namely that commodity booms promote structural change and change the demand for skill, and examine whether institutional changes, such as the abolition of slavery and proclamation of republic, changed this relationship.[7]

The relationship between export revenues and investments in human capital can be simultaneously determined, which poses challenges for identification of a causal relationship between them. Municipalities that benefited from improvements in their terms of trade have more resources available to invest in the expansion of their educational system, but agricultural productivity may already be higher in locations that have historically prioritized investments in the human capital of their population. To deal with this simultaneity, we build a variable that captures municipality-specific potential export revenues, based on the interaction between the suitability of geographical conditions to the production of each product in the municipality and the behavior of the international price of this product in 20 years before each census, which we call "Comparative Advantage International Prices" (CAIP henceforth).

[5] Aguëro et al. (2021) find evidences that a redistributive policy of ore exploration revenues in Peru in 2001 led to greater educational investments. Caselli and Michaels (2013) find some evidence of greater investments in education in Brazilian municipalities that presented an increase in tax revenues from oil exploration in 2000.

[6] Feler and Senses (2017) find evidence of deterioration in the quality of public services due to a reduction in tax revenues in localities of the United States that have faced competition from imports of Chinese products. Ferreira and Schady (2009) argue that adverse shocks to local revenues can lead agents to reduce investments in children's education and health, which depends on the agents' ability to mitigate adverse short-term shocks.

[7] Engerman and Sokoloff (1997, 2002) argue that more hierarchical societies rises in locations with favorable conditions for the use of large scale plantations and the use of enslaved labor, which led to a concentration of political and economic powers in certain social groups. Thus, social groups that dominated such powers would be able to decide the destination of public investments and could restrict the expansion of public education (Engerman et al., 2009).

This variable captures the variation in potential export revenues that results from three components of (arguably) exogenous variation: i) the initial (geographic) conditions of the municipality that determines its potential productivity to produce each commodity; ii) the potential for agricultural diversification that also originates from its geography; and iii) the behavior of international prices of the agricultural products. An advantage of the CAIP lies in its capacity to capture an "insurance effect", which represents the municipality capacity to adjust its agricultural production accordingly to the international market fluctuations, based solely on its land suitability for a diverse range of agricultural crops.

Using panel data at the municipal level from 1872 to 1960, our results show that higher potential export revenues reduce the share of illiterates in the population aged 15 or older, even after controlling for the municipalities fixed-effects that capture institutional and geographic permanent differences among them. We find that an increase of one standard deviation in potential revenues leads to a reduction of 14 percentage points (p.p.) in the share of illiterate shares, as compared to an average of 73% illiteracy of the Brazilian population aged 15 years or older in the sample period.

Even though our units of analysis are the municipalities, which are probably "price-takers", the Brazilian export basket was highly concentrated in a few commodities over the 19th and 20th centuries, raising issues about the possible endogeneity of the CAIP, especially with respect to coffee, a product which Brazil had a large chunk of the world exports in the turn of the century.[8] Thus, we carry out several robustness checks changing the commodities basket used to calculate the CAIP, excluding the variation of the price of coffee, for example, and showing that the results remain qualitatively the same. In addition, we experiment with using other time periods to calculate the average price of each commodity and use the median instead of the mean. All the main results are robust to these changes.

We also examine a new mechanism that can mediate the negative relationship between potential export revenues and the share of illiterates over the trade cycle: changes in racial composition brought about by structural change. Several studies show that the

[8] Sugar cane was the main commodity in Brazilian export basket until the early-19th century. Cotton's exports peaked in the 1860s and in the second half of the 1930s. Brazilian coffee dominated the international market in the late 19th century and through the 20th century, representing an average of 60% of the total coffee traded in international markets in this period (see Abreu and Bevilaqua, 2000).

rises in agricultural income cause an expansion of industrial or service activities, a process known as "structural transformation". Our results corroborate this hypothesis by showing that locations with the highest increases in potential export revenues also increase the share of labor in industrial and service sectors to the detriment of the agricultural sector. In addition, we show that commodity booms change the racial composition of the municipality, which affects literacy because of the pre-existent difference in literacy levels between blacks and whites. An increase of one standard deviation in potential revenues reduces the share of blacks by 3.7 pp and the share of mixed-race individuals by 8.4 pp.

We argue that the rising share of industry brought about by the commodity booms increases demand for skilled labor, underrepresented among the black and mixed raced population due to slavery.[9] [10] Our results also show that increases in potential export revenues also lead to a higher share of the young population attending a school, which may be the related to the rising share of whites in the population, as the share of young blacks and mixed-race also decline during commodity booms.

We also investigate the possibility that colonial institutions impact the relationship between potential export revenues and the share of illiterates, analyzing the heterogeneity of the main effects according to the share of enslaved people in each municipality in 1872. Larger shares of enslaved people may indicate locations with the predominance of landowners groups and plantations, which would be associated with the development of less democratic institutions and lower investments in public education.[11] Smaller investments in education would allow these dominant groups to maintain control of the voting group in their localities, since literacy was a pre-condition to vote in Brazil. Our

[9] Mixed race stands for people who claim to be '*pardo*' in the Brazilian society.
[10] Data from the 1872 demographic census, the year near the abolition of the slave labor system in Brazil, indicate that 99.91% of the enslaved population was illiterate
[11] Slavery represented a historical fact in Brazilian society associated with the concentration of economic and political powers in a dominant class, which possibly shaped institutions that have hindered their long-term development trajectory and influenced the persistence of their inequalities to the present day (see Engerman and Sokoloff, 1997, 2002, and Acemoglu et al., 2001, 2002). Bruhn and Gallego (2012) find evidence that the slavery had, in general, negative impacts on long-term economic growth when analyzing seventeen American countries. Specifically in Brazil, the use of slavery relates to lower long-term economic growth (Papadia, 2019) and greater income inequality (Soares et al, 2012). Note that Summerhill (2010) finds that the intensity of slavery do not relate to long-term economic results in the state of São Paulo, possibly due to the arrival of European immigrants in the late 19th and early 20th centuries.

results indicate, however, that the negative relationship between potential export revenues and the share of illiterates is not mediated by the share of population enslaved in 1872.[12]

We also use institutional changes that occurred in Brazil in the late 19th century, the abolition of slavery in 1888 and the proclamation of the Republic in 1889, as natural experiments to investigate whether there was a break or continuity in the effects of potential export revenues on the share of illiterates. The end of slavery allowed for a better distribution of economic and political power and the proclamation of republic allowed state governments to control export revenues and their allocations.[13] Our results indicate that the negative relationship between export revenues and illiteracy was **greater** before these institutional changes.

A possible explanation for this result has to do with the arrival of European immigrants in the late 19th century that were driven to agricultural areas and may have reinforced the positive effect of potential exports on schooling, as immigrants were more educated and had a greater demand for schools.[14] Another explanation is that export revenues were more relevant to family income in the late 19th century than during the 20th century, when a large share of the population was already in the industry and service sectors. Moreover, the late 19th century was a period of greater potential for industrial and services growth, so that the increase in potential export revenues increased the share of labor allocated in the industry and services sectors to a greater extent before institutional changes, which attracted educated whites and expelled blacks and mixed-race individuals. Therefore, structural change, coupled with migration and increase fertility[15] seem to be the main mechanisms driving the relationship between CAIP and illiteracy.

[12] Another possible explanation for the lack of impact would be the endogeneity of the number of enslaved people at a given point in time. This number may be endogenous to the demographic transition process (Lagerlöf, 2009) and to the living conditions of the enslaved people (Soares et al., 2012).

[13] In 1881, changes made to the Brazilian electoral system excluded the possibility of voting for the illiterate population. These changes excluded the majority of the former slave population from the electoral system. This exclusion of the illiterate from the electoral system lasted until 1988, a century after the abolition of the slave labor system.

[14] In the state of São Paulo there was a public policy to attract high-skilled foreigner workers in late 19th century and early 20th century to specific municipalities. Rocha et al. (2017) finds evidence that this policy contributed to a higher income per capita in the long run for these localities. Europeans immigrants were more educated and had a higher demand for schools since they came from countries with more consolidated educational public system relatively to Brazil (see de Carvalho and Colistete, 2010).

[15] Another possibility is that the white population managed to reduce their mortality rates to a greater degree, compared to other groups (see Brea, 2003). This would be possible through better opportunities in the labor market and higher income, which would translate into better living conditions and access to health services.

Finally, CAIP also impacts revenue volatility, as agriculturally more diversified municipalities are less affected by changes in the prices of one specific commodity and are thus less exposed to risks of adverse economic shocks. We investigate the impact of the volatility of potential export revenues on the share of illiterates and the results indicate that the volatility of revenues has no impact on the share of illiterates.

This paper has seven additional sections. The next section places the paper in the context of the literature on commodity booms and human capital investments, inequality, structural change and trade volatility. Section 3 contextualizes the Brazilian economy from 1872 to 1960. Section 4 describes the construction of the potential exports variable, while section 5 describes the rest of the data. Section 6 explains the empirical strategy, and section 7 presents the results. Section 8 concludes.

2 - Literature Review

The main contribution of this paper is to the literature that studies the impacts of commodity export revenues on educational investments. Positive shocks in commodity prices may cause local governments and the families themselves to prioritize the allocation of work in rural properties, to the detriment of educational investments (substitution effect). However, these positive shocks also make larger amounts of resources available to local governments and increase family income, which allows for the possibility of higher levels of educational investments (income effect).[16]

Musacchio et al. (2014) use the proclamation of the Republic and fiscal decentralization in Brazil at the end of the 19th century to analyze the impact of an increase in export tariff revenues on investments in education, finding that States with more democratic institutions invested more in public education in the early 20th century. Kruger (2007) finds that positive shocks in coffee production in the late 20th century encouraged only low-income families to withdraw their children from school and allocate them to agricultural production in the short term.[17] We contribute to this literature by building a new variable of potential export revenues (CAIP) that varies between municipalities and over time, combining the geographical characteristics of the municipalities with the

[16] See Ferreira and Schady (2009) for an extensive review of the literature on the income and substitution effects in the context of the impacts of macroeconomic shocks on educational investments and children's health in several countries.

[17] For analysis of the impacts of shocks in commodity prices on educational or health investments see Cogneau and Jedwab (2012) for the Ivory Coast case, Dammert (2008) for the Peru case, Miller and Urdinola (2010) for the Colombia case, and Maluccio (2005) for Nicaragua case.

temporal variation in commodity prices and proposing a different mechanism to explain the relationship between potential export revenues and illiteracy.

Some studies reveal that increases in local government revenues do not necessarily translate into greater investments in public education, as institutions and the decisions of the local elites are important mediators of this relationship.[18] The theoretical model proposed by Bourguignon and Verdier (2000) considers that those holding political power will adopt strategies to maintain its privileges and will decide not to subsidize public education. In another scenario, the dominant group may adopt redistributive policies when there is a risk of popular revolts, as in Acemoglu and Robinson (2000). In Brazil, literacy was required to vote, therefore, investment decisions in public education had an important political component. We contribute to this literature by examining whether there was a change in the relationship between export revenues and investments in education after the institutional changes that occurred in Brazil at the end of the 19th century, with the end of slavery and proclamation of the republic.

The performance of the export sector may impact other sectors of economic activity and the urbanization process through increases in agricultural productivity, a process known as structural transformation. Bustos et al. (2016) find evidence that technological changes in soy cultivation in Brazil increased the productivity of this crop, reduced the amount of labor required for its cultivation and boosted the growth of the industry in the early of the 21st century.[19] Nunn and Qian (2011) show that the rise in agricultural productivity through the adoption of potato cultivation in Europe contributed to the expansion of European urban centers between 1700 and 1900. However, the spillover of agricultural productivity to other sectors of economic activity can occur only in the short term or simply not occur due to the nature of the technological change. [20] [21] We contribute to this literature by investigating how shocks of potential export revenues

[18] See Engerman et al. (2009) for an analysis of several Latin America countries, Galor et al. (2009) for the United States case, Banerjee and Iyer (2005) for the India case, Musacchio et al. (2014), Naritomi et al. (2012), Wigton-Jones (2020), and Wegenast (2010) for the Brazilian case.

[19] The change was the use of genetically modified soybean seeds that allowed the production of the same amount of soybean using less labor.

[20] Hornbeck and Keskin (2015) find evidence of only short-term overflows to non-rural sectors due to impacts of the adoption of technologies that allowed the exploration of the Ogallala aquifer's waters in the United States and the increase in agricultural productivity in the region.

[21] Bustos et al (2016) analyze a second agricultural crop: corn. The authors note that when using techniques to allow the planting of two corn crops in a given year, there was an increase need for labor, which negatively impact industrial growth in the regions that implemented these two crops.

influenced the allocation of workers among sectors of economic activity in Brazil between 1872 and 1960, the period of the beginning of industrialization in the country.

Latin America had the highest levels of inequality at the end of the 20th century, with Brazil being one of the most unequal.[22] Leff (1972) argues that factor mobility was low until the end of the 19th century in Brazil, which contributed to regional inequalities in the long term. Wagner and Ward (1980) indicate that the internal migrations from rural to urban centers started mainly in the 1930s in Brazil. We add to those studies by showing that regions that suffered persistent negative shocks in early 20th century may have failed to industrialize and became poor in the long-run.

We also contribute to the literature that analyzes the impacts of commodity price volatility on developing countries, as the volatility of economic growth rates[23] and of the prices of commodities that a country exports have been shown to negatively impact its development.[24] Moreover, Bleaney and Greenaway (2001) find evidence that higher levels of trade volatility in sub-Saharan countries reduce investment levels in these locations. We examine how the volatility of potential export revenues impacts the share of illiterate in the municipalities.

3 - Brazilian context in the period 1872-1960

Brazil had agriculture as the main dynamic sector of its economy until the first half of the 20th century and the scenario of international trade in commodities had great repercussions on the national economy. Favorable shocks from international trade and, mainly, the rise of coffee as the main Brazilian export product in the second half of the 19th century contributed to the increase in domestic income through the effect of the internal multiplier and the development of the domestic market.[25] The influx of export

[22] López and Perry (2008) show that the regions of Latin America and Sub-Saharan Africa had the highest levels of income inequality in the late 20th century.
[23] Ramey and Ramey (1995) analyzed 92 OECD countries in the 1960-1985 period and found evidence that countries with high volatility in economic growth rates have lower average economic growth.
[24] Blattman et al. (2007) analyzed 35 developed and developing countries and found that countries with more volatile terms of trade showed less economic development, especially in peripheral countries. This argument is also supported by Jacks et al. (2011) that indicate that poor countries have greater volatility in terms of trade because they specialize in agricultural products and minerals, which have higher price volatility than manufactured products.
[25] Furtado (1959) points out that coffee production was undertaken by groups that worked from agricultural production to commercialization, which allowed the development of agricultural and non-agricultural production to supply the urban centers near coffee farms.

revenues contributed to the expansion of other sectors of society, such as the development of industry until the 1920s.[26]

The Brazilian export basket was predominantly agricultural and extractive throughout the 20th century until the mid-1960s. Still in 1960, the export of basic products represented 85.8% of the export basket.[27] In addition, export tariffs accounted for an important share of total tax revenues of local governments. Foreign trade revenues represented 75% of central government's tax revenues at the end of the 19th century,[28] which shows the primordial importance of the agro-export sector to generate tax revenues in this period. With the proclamation of the Republic in 1889 there was a fiscal decentralization, so that the state governments managed to control the export revenues tariff. In this first republican moment, export revenues represented 40% of the state governments' tax revenues.[29]

The revenues obtained through foreign trade had a strong influence on the possibilities of educational investments in Brazil in the period from 1872 to 1960. State governments were primarily responsible for the decisions of investments in public education in that period. However, until the proclamation of the Republic, the central government maintained control over tax revenues and state governments depended on transfers of resources to make such investments.[30] With fiscal decentralization in 1889, state governments achieved control over export revenues tariffs and greater autonomy in educational investment decisions. Thus, Brazilian exports represented an important potential source[31] of resources for financing public education systems in the late 19th century and throughout the 20th century.

---

[26] Mello (1982) argues that the expansion of the coffee economy allowed the accumulation of resources and the creation of demand for manufactured goods. Thus, at first, periods of good performance in the coffee sector translated into industrial growth. Later, the industrial sector would already be able to absorb shocks from foreign trade, given that there would be a possibility to exploit domestic demand and industrial idle capacity (Mello, 1982, pp. 96-109).

[27] Data from *Anuário Estatístico do Brasil – IBGE/SECEX/DEPLA*.

[28] Villela (2007) indicates that foreign trade tariffs represented, on average, 75% of the total tax revenues of the central government in the period from 1844 to 1889.

[29] Varsano (1996) indicates that the export tariffs represented approximately 40% of the total revenues of the state governments until the 1934 Constitution. After that, the export tariff lost importance to other taxes, such as the sales and consignment tax. Kang (2011) indicates that the export tariff represented less than 1% of the total revenue of state governments in 1958.

[30] Musacchio et al. (2014) indicates if a state wanted to expand its public education system, it should do so with its own resources or depend on transfers from the central government, which retained the power to decide on these investments because of their control over tax revenues in the period of 1824-1889.

[31] Even with the availability of revenues, investments in education were not mandatory. Thus, such investments ultimately depended on the decisions of local governments.

The importance of agriculture for the Brazilian economy created a scenario in which the groups that owned the land (landowners and agro-exporters) also exerted a great influence on the economic and political powers. With export-oriented agricultural production, large farms and the use of enslaved labor predominated.[32] It is important to highlight that the presence of the agro-export sector was not uniform in Brazilian territory or in its composition. The agro-export sector in the northeast region of Brazil focused on the production of sugar cane and cotton and was the main agro-export group until the first half of the 19th century. There was the ascension of the coffee producers in the second half of the 19th century in the southeast region and consequently a transfer of economic dynamism to that region.

The use of enslaved labor until 1888 in Brazil was a possible determinant of a more hierarchical structure in the country's history. In the years before the abolition of its slave labor system and even after the ban on slave trade in the Atlantic, Brazil still had 1,509,339 slaves who represented 15.2% of its total population.[33] Of the people who lived as slaves in Brazil, practically all were illiterate (99.91% of them) and deprived of capital assets.[34] This scenario did not change with the abolition of the slave labor system in 1888, [35] so that the newly freed enslaved people faced risks of expropriation of private property,[36] exclusion of the electoral system[37], and social discrimination.[38]

[32] Engerman and Sokoloff (1997, 2002) argue that the initial distribution of factors and the types of soil and climate present here in Brazil favored the cultivation of plantations aimed at the international market, such as sugar, cotton and coffee. Thus, such agricultural crops had greater economic returns when produced on a large scale through large estates and the use of enslaved labor.
[33] Data from *Recenseamento Geral do Império de 1872*.
[34] Fausto (1994) highlights that the enslaved population had no rights in the legislation. In addition, they did not have the status of a person, but a thing in the legislation. "*The enslaved blacks and mixed race had no rights, thus they were legally considered a thing and not a person*" (Fausto, 1994, p. 54). Translated by the authors.
[35] Fausto (1994) indicates some destinations to former slaves such as became dependent on landowners, turned to subsistence activities, or took on underpaid and unstable occupations. Woodard (2005) argues that it was difficult for the liberated population to participate in both the political system and in prestigious labor positions due to the presence of nepotism in some regions or by the control and administration of a small base of voters in other regions.
[36] Papadia (2019) finds evidence of a lower presence of *quilombos*, communities of enslaved people who fled from farms or freedmen, in regions with higher potential coffee productivity. The author argues that the group of coffee producers would use expropriation of the fertile lands occupied by *quilombos* to expand the frontier of the coffee production.
[37] Voting rules kept this portion of the population without the right to participate through literary restrictions on voting after the period of abolition of slave labor. From 1881 onwards, voters were required to demonstrate that they could read and write.
[38] Bucciferro (2017) finds evidence that black workers received lower wages by up to 50%, compared to white workers, within the same job category in 1920. In addition, employers exhibited a marginal preference for hiring European immigrant workers.

In the first half of the 20th century, Brazilian society was undergoing structural changes, notably the urbanization[39] and industrialization processes. These changes have led to a marginal increase in the demand for schooled labor, which may have increased popular demand for the construction of more schools and attracted people with some schooling to urban centers. Localities that received foreign immigrants (mainly the states of the south and southeast regions) present an additional factor, which is the influx of people with higher education and supposedly with greater demand for educational services.[40] These transformations throughout the 20th century also reduced the importance of export revenues to public revenues, as other urban and domestic trade taxes began to compose the tax revenues of state and municipal governments. This would result in less direct impacts of export revenues on the share of illiterates throughout the 20th century, even though there has been an increase in demand for educational investments.

4 - Comparative Advantage International Prices (CAIP)

Our strategy to deal with the simultaneity problem between export revenues and the share of illiterate in Brazilian municipalities focus on using exogenous variation in the exports revenues in Brazilian municipalities. We follow two steps to construct this variable. First, we obtain the potential agricultural productivity of several agricultural crops that a given municipality inherited at the time of its foundation, considering only factors that are invariant throughout the analyzed period (climatic and soil conditions). Thus, we consider the potential agricultural productivity of the five main export commodities that could in principle be cultivated in each location.

To do that, we use data on the maximum attainable yield of several agricultural crops from the Global Agro-ecological Zones (GAEZ) v3.0 project of the United Nations Food and Agriculture Organization (FAO). These data indicate the maximum productivity of several agricultural crops considering the climatic conditions and restrictions resulting from variations in soil moisture, presence of diseases, pests and weeds, in addition to precipitation and ease in the use of the soil.[41] In addition, these data are calculated according to theoretical models and controlled experiments and measure

[39] Data collected by Wagner and Ward (1980) indicates that 15.3% of the Brazilian population lived in urban centers in 1940. In 1970, this percentage was already 55.8% of the total population.
[40] Rocha et al. (2017) find evidence that localities that received foreign immigrants via programs financed by the government of the state of São Paulo showed a greater provision of educational inputs.
[41] See Fischer et al. (2012) for details on GAEZ v3.0 project.

the potential productivity of all possible agricultural crops in a given region and not just the ones already being produced.

Information on the maximum attainable yield of different commodities considers different types of water availability (for example, rain or irrigation water) and different production technologies (low, medium or high level of technology used). In this paper, we use the potential productivity with rainwater and low technological level only, that represents the production of traditional agricultural crops with the intensive use of labor without the use of chemicals or fertilizers. The data are available in space cells with a dimension of 9.26 km by 9.26 km (measured at the Equator line), encompassing the entire terrestrial globe. The maximum attainable yield data represents the productivity in terms of tons per hectare of a given agricultural crop in a given space cell.

Based on GAEZ/FAO's maximum attainable yields and observed agricultural production value data, we estimate the predicted distribution of the municipalities' potential agricultural production value shares, following the methodology proposed by Fiszbein (2017). We use a fractional multinomial logit model (equation 1) to estimate these potential shares: [42]

$$\widehat{\theta_{ic}} = E[\theta_{ic}|\boldsymbol{\pi_c}] = \frac{e^{\beta_i'\boldsymbol{\pi_c}}}{1 + \sum_{j=1}^{I-1} e^{\beta_j'\boldsymbol{\pi_c}}} \quad (1)$$

where θ_{ic} is the observed production value share for commodity *i* in municipality *c*, and $\boldsymbol{\pi_c}$ is a vector of the averages of maximum attainable yield of the analyzed commodities in the area of municipality *c*. Equation 1 is restricted by $\sum_{i=1}^{I} \theta_{ic} = 1$ e $0 \leq \theta_{ic} \leq 1, \forall i$. The estimates of the production value share of the different crops analyzed represent the exogenous part of these shares that results from their initial conditions (climatic factors, which are considered invariant throughout the analyzed period). From equation 1 and its restrictions we have $\sum_{i=1}^{I} \widehat{\theta_{ic}} = 1$.

The second step in the construction of the exogenous variable consists of interacting the estimate obtained in the first step with the respective component of international prices for each agricultural crop (equation 2), as follows:

$$\omega_{ict} = \widehat{\theta_{ic}} * P_{it} \quad , \qquad i = 1,2, \dots, I \quad (2)$$

[42] See Sivakumar and Bhat (2002) for details of fractional multinomial logit model.

where P_{it} represents the component of international prices of commodity *i* in year *t* and it is the same for all municipalities. We use the average price of commodity *i* in a period of twenty years prior to year *t* as the price component.[43] The sum of the interactions of all commodities, $\sum_{i=1}^{I} \omega_{ict}$ results in the proxy for potential exports, that we call Comparative Advantage International Prices (CAIP):

$$CAIP_{ct} = \sum_{i=1}^{I} \omega_{ict} \quad (3)$$

Therefore, CAIP represents the potential capacity of municipality *c* to generate export revenues from its comparative advantages of the land and the international trade scenario in year *t*, based on the potential productivity of various commodities that can grow in its territory.

In addition to the CAIP level, the present work also analyzes the CAIP volatility. To calculate this volatility, we calculate equation 3 considering the annual price component. Thus, we obtain annual CAIP measures for each municipality *c*. CAIP volatility is measured by the standard deviation of these annual CAIP considering the twenty years prior to the reference date, *t*.

5 - Data

The analysis proposed in this work analyzes 89 years of Brazilian history, considering the period between the years 1872 and 1960.[44] The choice of this period is based on the fact that in 1872 the first Demographic Census of the Brazilian population was carried out and it contains information about the enslaved population at the municipal level. The year of 1960 was chosen for three reasons. The first is because the year 1960 precedes the period of expansion of the Brazilian agricultural frontier through the *cerrado* that involved technological changes in agriculture. The second is because the year 1960 precedes the period of redemocratization in Brazil.[45] The last reason is the availability of information on skin color and race together with literacy information in the 1960 Demographic

[43] We consider other specifications of the price component and the results presented throughout the work. Results remain robust for the tested pricing component specifications.
[44] We perform robustness checks including the years of 1970 and 1980 in the panel data. Main results remained robust to the inclusion of these years. See Appendix 2.
[45] The Brazilian redemocratization occurred with the creation of the 1988 Constitution and resulted in the end of the obligation to know how to read and write in order to participate in the electoral system.

Census.[46] In addition, the interval from 1872 to 1960 represents a period when the agro-export sector was important for the country's economic development, which was predominantly rural until 1960.

Using such a long period of Brazilian history, we face a scenario of changes in the borders of the municipalities, creation and exclusion of municipalities throughout this period. In the first Demographic Census (in 1872), Brazil had 642 municipalities. In 1960, Brazil had 2,766 municipalities. Thus, the concept of *Áreas Minimamente Comparáveis*[47] (AMC henceforth) is used,[48] specifically the compatibilization for the period from 1872 to 2010,[49] to perform the proposed analyzes. This concept uses a methodology for aggregating municipalities, taking into account all changes in borders and the creation or exclusion of municipalities that occurred from 1872 to 2010. The aggregation results in 482 AMC.

For the present work, we analyzed five of Brazil's main export products in the early 20th century. Thus, we guarantee that CAIP captures export earnings from agricultural products that, in fact, had relevance in the Brazilian economic scenario in the early 20th century. We analyze the agricultural crops of coffee, sugar cane, cotton, tobacco and cocoa. Table 1 shows that these commodities represented 72.8% of the Brazilian export basket in the period from 1919 to 1923 and 80.1% of the export basket in the period from 1924 to 1929.

The calculation of the observed production value share of each of these five agricultural crops uses production data in tons from the 1920 *Censo Agrícola*[50] and the average prices in *mil-réis* (per ton) for the year 1912 for these same products obtained in the *Anuário Estatístico do Brasil*[51] of 1908-1912.[52] The choice of agricultural data from

[46] For reasons not explored here, information on skin color and race was exclude from demographic census in Brazil in the early 20th century.
[47] Minimally Comparable Areas.
[48] See Reis et al. (2008) on the AMC construction methodology.
[49] See Ehrl (2017) on the AMC construction methodology, specifically for the period from 1872 to 2010.
[50] Agricultural Census.
[51] Statistical Yearbook of Brazil.
[52] The cocoa crop has no local prices in the *Anuário Estatístico do Brazil* of 1908-1912. Thus, we decided to use real price index from Jacks (2019) and convert its value to *mil-réis.* First, the dollar amount was converted to British pounds using "Lawrence H. Officer," Exchange Rates Between the United States Dollar and Forty-one Currencies, ", MeasuringWorth, 2020" available at http://www.measuringworth.com/exchangeglobal/ and accessed on 22/01/2020. The conversion of pounds sterling to *mil-réis* computed using data from "*Estatísticas históricas do Brasil: séries econômicas, demográficas e sociais de 1550 a 1988. 2. ed. rev. e atual. do v. 3 de Séries estatísticas retrospectivas. Rio de Janeiro: IBGE, 1990*".

1920 stems from the availability of agricultural production data at the municipal level in the early 20th century.

In addition to the observed production value shares, the CAIP calculation uses the maximum attainable yield data obtained from the Global Agro-ecological Zones (GAEZ) v3.0 project of the United Nations Food and Agriculture Organization (FAO), described in the section previous. Figure 1 shows the spatial distribution of the maximum attainable yield (in tons per hectare) of (a) coffee, (b) sugar cane, (c) cotton, (d) tobacco, and (e) cocoa crops, considering climatic conditions. We can see that the coffee crop has lower potential productivity in the *caatinga* biome (northeast region of the map) and in the southern region of Brazil. This pattern is somewhat similar to that presented by the sugar cane culture. The coastal region has high maximum attainable yield for all agricultural crops analyzed, only to a lesser extent for tobacco cultivation. The *caatinga* region (northeast region of the map) has higher levels of potential productivity for cotton and tobacco crops.

With the data of observed agricultural production value (in *mil-réis* per ton) in 1920 and the data of maximum attainable yield (in tons per hectare) of each agricultural crop analyzed and of each AMC, we calculate the estimated shares of agricultural production value (equation 1). Figures 2 to 6 show that the estimated shares maintain a positive correlation with the observed shares of agricultural production value in 1920 for all analyzed commodities, in addition to maintaining the variability of these production values.

Figure 2 shows the spatial distribution of the observed and estimated shares of coffee production value by AMC. We observed that the observed and estimated shares are high in the southeastern region of Brazil. There are regions in the Brazilian inland that present positive estimated shares and null observed shares of this production value. This may be an indication of that AMC was not using all of its agricultural potential for coffee growth or was concentrating on the production of the other agricultural crops analyzed.[53] The correlation index between these two shares is 0.68.

[53] A possible explanation for the null observed shares of production value in 1920 is the fact that many AMC were not integrated into the agricultural production system aimed at exports in that period. It is important to highlight that Brazilian exports' agricultural production was concentrated in regions close to the coast until the 1960s. From 1960 onwards, there are public investments in technological development and in the transport infrastructure that allow the expansion of the agricultural frontier to the interior of the country.

Figure 3 shows the spatial distribution of the observed and estimated shares of sugar cane production value by AMC. The observed shares of this agricultural culture are concentrated in coastal regions and in the interior of Brazil. When we consider the estimated shares, we notice a uniform distribution of these shares throughout Brazil. The correlation index between these two shares is 0.27. Figure 4 shows the spatial distribution of observed and estimated shares of cotton production value by AMC. We note that the observed shares in the northeast region of Brazil are in regions with high estimated shares. The correlation index between these two shares is 0.61.

Figure 5 shows the spatial distribution of the observed and estimated shares of tobacco production value by AMC. The observed shares are concentrated in the southern region, in the interior and in a portion of the coastal region in northeastern Brazil. These regions also have higher levels of estimated shares of tobacco production value, especially in the southern region. The correlation index between these two shares is 0.50. Figure 6 shows the observed and estimated shares of the value of cocoa production value by AMC. We observe that the regions that concentrate the observed shares of this production value are in the north and coastal region of Brazil. These same regions have estimated positive shares. The correlation index between these two shares is 0.49.

The CAIP calculation (equation 3) use these estimated shares of the agricultural production value and the commodity real price index provided by Jacks (2019). The price component used in the present work is the average of the commodity price index in the last twenty years prior to the reference date. We observe in Figure 7 that the spatial distribution of CAIP (standardized to have a mean of 0 and standard deviation of 1) in 1872 presents values above the mean in coastal regions, mainly in the northeast and south regions of Brazil. When we consider the year 1960, the southeastern and southern regions of Brazil concentrate the highest CAIP values.

In Figure 8 we present the CAIP values for all years from 1872 to 1960, considering the distribution of these values among Brazilian AMC. In addition, we incorporate information on variation in the real price index of the analyzed commodities. CAIP showed a downward trend between 1872 and 1910, hitting a plateau between 1910 and 1930. The 1930-1945 period represents an abrupt downward trend of CAIP. The 1945-1960 period represents an upward trend, especially in the 1950-1960 period. These trends presents relationship with the super-cycles of tropical commodities prices from Erten and Ocampo (2012). These authors identifies the first super-cycle's surge from

1895 to 1910 and the fade-out of this super-cycle from 1910 to 1940. The first half of 1940s was especially detriment to the international market of commodities due to World War II. CAIP's upward trend from 1945-1960 matches Erten and Ocampo (2012)'s second super-cycle, which coincides with the period of post-war and reconstruction of Europe.

We can see in Figure 8 that coffee, sugar, and cocoa's price indexes series follows closely with the CAIP value series. In the abrupt downward trend between 1930 and 1940 and its subsequently recovery between 1940 and 1960 we can see all commodities, except tobacco, following these trends. Coffee's price index series resembles the most of Erten and Ocampo (2012)'s super cycle of tropical commodities prices. Tobacco's price index series presents an upward trend from 1880 to 1960, showing detachment from CAIP and from super cycle of tropical commodities prices.[54]

We consider the possibility that some commodities may bias our results since Brazil represented a major exporter of several tropical products through the 19th and 20th centuries. Coffee is a canonical example that Brazilian exports represented at least 60% of the international coffee market in the late 19th century and early 20th century. In Figure 9 we compare six different constructions for CAIP. First in 'All commodities' figure we present standard (see equation 3) CAIP's temporal variation among Brazilian AMC. All the other figures represents alternative CAIP versions (see below in equation 5), which we restrict a commodity's price index at its level of 1872. We can see that restricting sugar's price index generates higher values of CAIP. This pattern is due to the downward trend of sugar price index through 20th century (see Figure 8) and a positive sugar cane's estimated production value share all along Brazil's territory (see Figure 3). Restricting coffee or cotton's price indexes seems to generate only greater dispersion of CAIP values among Brazilian AMC. Restricting tobacco and cocoa's price index does not seems to generate a different pattern from standard CAIP.

Figure 8 already indicates that CAIP captures a type of insurance effect on the agricultural production of Brazilian AMC. This is because an AMC that has a high maximum attainable yield for many agricultural crops would be able to marginally adjust its agricultural production to accommodate adverse shocks in the international prices of

[54] A possible explanation for tobacco's detachment from the other trends lies in its inelastic demand and an increase in its consumption in the first half of the 20th century. See Townsend (1996) for detailed information on the inelastic demand characterization of tobacco.

a given commodity. In Figure 10, it can be seen that the AMC group with high agricultural diversification[55] (the more diversified, the greater the number of potential agricultural crops that a municipality is capable of producing), in fact, has a lower variability of CAIP values throughout the analyzed period. In other words, AMC with high agricultural diversification have greater stability in their potential export earnings, which may indicate the insurance effect of their ability to produce different exportable agricultural crops. On the other hand, municipalities with low agricultural diversification (consequently, less insurance capacity) are more likely to have CAIP levels below the median.

In addition to the variation of CAIP in level, we also analyzed the volatility of CAIP in the period. We can see in Figure 11 that this volatility reached the highest values in the reference dates of 1872, 1950 and in the first half of the 1930s. Coffee prices showed greater volatility between the years 1900 to 1910 and between the years 1950 to 1960, which may have influenced CAIP's greater volatility in these periods. Sugar and cocoa prices show high volatility between 1920 and 1940, with cocoa prices showing high volatility again from 1945. Volatility's peak in the period of 1930-1940 and in the period of 1945-1960 coincides with the end of the first super-cycle and beginning of the second super-cycle of tropical commodities prices, respectively, presented by Erten and Ocampo (2012).

The analysis carried out in the present work use two data panels. The first panel is the most extensive and contains information for the years 1872, 1920, 1940 and 1960. We use this panel to analyze our main question about the relationship between export revenue shocks and the variations in the share of illiterates. We observe in Table 2 that the average share of illiterates among the population aged 15 years or older was 94% among the AMC in the year of 1872. This share reduced to 46.6% by 1960.

Data of 1872 contains information on the enslaved population in Brazil. We see in Table 2 that the average share of enslaved population, in relation to the total population in the AMC of Panel B, was 14.4%. This percentage has a minimum of 0.6% and a maximum of 57.4%, which represents a high variability in this average share. It is important to note that these data were collected at a time when the process of abolishing

[55] Agricultural diversification computed using the Herfindahl index. We compute Herfindahl index as $\left(1 - \sum_{i=1}^{I} \widehat{\theta_{ic}}^{2}\right)$ for each AMC *c* and it assumes values between zero and one. Values close to zero indicate low agricultural diversification and values close to one indicate high agricultural diversification.

the slave labor system was already taking place in Brazilian society (which would take place in 1888) and that the slave labor force was substantially displaced to the coffee producing regions in southeastern Brazil. We also observe in Table 2 that the average share of the black population aged 15 years or older enslaved in 1872 was 52.2%. The situation of the mixed race population aged 15 years or older is that of a smaller share (16.3%) in a situation of slavery in 1872.

Table 2 also shows the descriptive statistics of the potential export variables used in the analysis. We can see in Table 2 that the versions of CAIP that restricts the price index of coffee, sugar, cotton, tobacco or cocoa had smaller reductions in its mean values between 1872 and 1920, relatively to the CAIP version with all agricultural crops. The year of 1940 represents the lowest mean value for all CAIP variables (except for its volatility) and a higher volatility that relates to the international context of war and trade recession. Restricting sugar price index at its level of 1872 generates the highest mean value for CAIP since this commodity presented a downward trend in its price index values through 20th century. CAIP's variability between its alternatives highlights the insurance effect captured by CAIP, which may have a substantial influence on the relationship between shocks from export revenues and the variation in illiteracy. This is because municipalities that have the insurance of having a high agricultural diversification may have had better conditions to invest in their educational systems. On the other hand, there is a possibility that the agricultural production of some product influences international prices, as is the case with coffee, which justifies analyzing the robustness of the results with versions of CAIP restricting a certain agricultural culture.

We use a second panel data that consists of information for the years 1872 and 1960. This panel is used in the analysis of mechanisms of the relationship between CAIP and illiteracy. The reason for restricting the panel to just these two years is the availability of information for population groups by race.[56] Looking at Table 3 it can be seen that the share of illiterates aged 15 years or older maintains mean values close to the unbalanced panel (Table 2), with an average share of 94% in 1872 and 46.8% in 1960.

Considering the distribution of the population among the different sectors of economic activities, we observe (Table 3) a greater allocation of workers in the

[56] The Demographic Census of 1960 was conducted in a sample of municipalities, which explains the reduction from 482 AMC in the full panel to 274 AMC in Table 3.

agricultural sector with an average of 53.4% of the total population aged 15 years or older working in this sector in 1872. The average percentages in industry and services are 5.8% and 40.8% in the same year, respectively. This information shows the importance of agriculture in Brazilian society in the period analyzed, in addition to reflecting its rural characteristics. Even if we consider the year of 1960, 63.3% of the total population aged 15 years or older are working in agriculture. The share of people who migrated from foreign countries to work or live in Brazil has a low average percentage in the period, being 2% of the total population in 1872 and 1.4% in 1960.

The share of young population aged 7 to 14 who attend school is low in 1872, with an average share of 12.7%, and substantially higher in 1960 (average of 50%).[57] The white population was more representative than the black or mixed race populations on average between 1872 and 1960. This pattern is present both in the young population (aged 7 to 14 years of age) and in the population aged 15 years or older. Still in Table 3 we can observe similar mean values for CAIP and CAIP's volatility in 1872 and in 1960.

6 - Empirical strategy

The objective of the present work is to estimate the relationship between shocks in potential export revenues and the variation in the share of illiterates in the AMC of Brazil between 1872 and 1960. We use panel data at the AMC level of the years 1872, 1920, 1940 and 1960. Shocks in potential export revenues are captured by the exogenous variable, CAIP, built in section 3. We use a reduced fixed-effects model at the AMC level to estimate this relationship, as follows:

$$\mathrm{Y}_{ct} = \beta_0 + \beta_1 CAIP_{ct} + \sum_{q=t+1}^{T} \tau_q \lambda_q + \alpha_c + \mu_{ct} \qquad (4)$$

The term Y_{ct} represents the share of illiterates in the population aged 15 years or older in AMC *c* in year *t*. The term $CAIP_{ct}$ represents the potential export revenues in AMC *c* in year *t*. The term λ_q is a binary variable for year *q*, which captures the time fixed-effect from the second year analyzed onwards. The term α_c represents the fixed-effects of AMC *c*, and the term μ_{ct} is an idiosyncratic error. The term of interest in equation 4 is the β_1 coefficient that indicates the average relationship between CAIP and

[57] The share of young population who attended school in 1872 considers that no enslaved young population attended school.

the share of illiterates (aged 15 years or older) in the analyzed AMC and period. Analysis of specific mechanisms and illiteracy by race use the same structure as in equation 4, but use different response variables.

There are three concerns with the construction of CAIP and its exogeneity. First, certain agricultural crops can endogenously determine CAIP, such as coffee. As Brazil maintained its monopoly on the trade of coffee on the international market in the late 19th and early 20th centuries, [58] localities with a high share of Brazilian coffee production could partially affect their international prices. We perform robustness tests by creating alternative versions of CAIP that restrict a particular agricultural crop's price index at its 1872 level. In equation 5, a CAIP is constructed restricting the commodity j, $j \in i$ at its 1872's price index. Thus, it will be possible to verify whether the restriction of any of the analyzed agricultural crops influences the results.

$$CAIP_{ct}^{j} = \widehat{\theta_{jc}} * P_{j,1872} + \sum_{i=1}^{I} \widehat{\theta_{ic}} * P_{it} \quad , \qquad i = 1,2,\dots,I \ , \ i \neq j \tag{5}$$

The second concern is about the sensitivity of the results to the chosen period for the CAIP's price component. Thus, we carry out tests with different price components (average or median), considering different periods (10 or 20 years before year *t*). The last concern is whether the estimation of agricultural production value shares (the first step of building the CAIP) considers only climatic conditions as determinants of these plots. Any other determinant of agricultural crops chosen to be produced in a given AMC must not affect these estimated plots.

The final concern regarding the exogeneity of the CAIP is that it adequately captures potential productivity-related factors. Moreover, we want to test whether there are relevant effects of the increase in commodity prices on education that do not take into account export revenues. We then estimate equation 6, in which we additionally control for interactions between the export crop prices and the geographic features listed below. For commodities $i = 1,2,\dots,I$, P_{it} represents the average price in the twenty years prior to year t and for each AMC c, variable Geo_{cg} represents a geographic trait g. We consider

[58] Abreu and Bevilaqua (2000) present information that the representativeness of Brazilian coffee production in international trade was approximately 60% each year in the period from 1896 to 1942, with this percentage reaching over 70% in half of that period and above 80% in 1906. Even in the 1850s this share was already 50%.

weather and physical factors related to potential productivity (average temperatures and average rainfall in each season, altitude, and latitude), in addition to geographic positions that could affect productivity due to factors related to logistics, infrastructure and proximity to markets (distance from the AMC's centroid to the Atlantic Ocean, and distance from the AMC's centroid to the state's current capital).[59]

$$Y_{ct} = \beta_0 + \beta_1 CAIP_{ct} + \beta_2 \sum_{i=1}^{I} P_{it} * Geo_{cg} + \sum_{q=t+1}^{T} \tau_q \lambda_q + \alpha_c + \mu_{ct} \qquad (6)$$

The interactions in equation 6 represent specific trends of each AMC along the price variation and capture the advantages that AMCs with some characteristics (such as higher temperatures or proximity to the equator line) could have in periods of high international prices. We want to test if CAIP captures potential revenue shocks from export crops and is calculated using the expected proportions based on the potential productivity of those crops. The distance variables to state capitals and to the Atlantic could affect productivity and education due to factors related to logistics, infrastructure, and proximity to markets. [60] Those dimensions would not be captured by the CAIP, which considers climate, soil, and technology conditions. However, as CAIP is exogenous, we expect β_1 should not change much when we include such controls.

Furthermore, we test whether there are relevant effects of the increase in commodity prices on education not related to export revenues. CAIP incorporate the information on physical and weather conditions in the potential productivity indicator and in the estimated proportions, then, conditional on the CAIP, when we use physical and weather conditions as Geo_{cg}, β_2 should be small, unless they represent alternative channels to the effects of the increase in commodity prices on education not related to export revenues.

The estimate to verify the influence of the share of enslaved population present in the year 1872 on the relationship between shocks of potential export revenues and the

[59] Since Brazil's territory is large and presents socioeconomic heterogeneity, we carry out robustness tests to control for different Brazilian states' time-specific fixed effects to rule out possible differences across Brazil's regions. Results are robust for these tests. See Appendix 3.

[60] There is a possibility that Brazilian regions closer to the Atlantic Ocean receives more investment in infrastructure than inland regions for motives other than just agricultural productive. For instance, there could be more investment due to the fact that these regions were the first ones to be settled and concentrated a large population. We carry out robustness tests to control for this heterogeneity and results are robust. See Appendix 3.

share of illiterates include an interaction term between CAIP and the share of enslaved people. In equation 7, $Esc_{c,1872}$ represents the share of total population enslaved in 1872 in AMC *c*. Thus, in equation 7 the parameter β_2 represents the possible change in the average influence of CAIP on the share of illiterates, conditioned on the share of enslaved people present in 1872.

$$Y_{ct} = \beta_0 + \beta_1 CAIP_{ct} + \beta_2 CAIP_{ct} * Esc_{c,1872} + \sum_{q=t+1}^{T} \tau_q \lambda_q + \alpha_c + \mu_{ct} \qquad (7)$$

The analysis that investigate the institutional changes at the end of the 19th century use the same structure of equation 7, but use a binary variable that assumes a value equal to 1 for the year 1872 and 0 otherwise. This binary variable is used in the same way as the share of enslaved people variable in 1872 in equation 7.

7 - Results

7.1 Main results

Our main result indicates that localities (AMC) that have benefited from the increase in the prices of their exportable agricultural products have shown a reduction in their share of illiterates aged 15 years or older. In other words, the results indicate a negative relationship between a positive shock in potential export revenues (CAIP) and the share of illiterates. Table 4 presents the results obtained from the estimation of equation 4, considering the CAIP standardized to have a zero mean and standard deviation of one.[61] Considering the panel with information in the years 1872, 1920, 1940 and 1960, we have that, on average in this period, localities (AMC) with a higher CAIP in one standard deviation (s.d.) reduced their share of illiterates by 14 percentage points (p.p.) (column (1)). This result is statistically significant at 1% and it is highlighted that this estimate already controls by the fixed effects of Brazilian AMC in the period.

Table 5 also presents the results of the estimates with the alternative versions of CAIP that restricts a given commodity's price index (equation 5). The negative relationship remains robust and statistically significant at 1% after restricting each of the commodities considered in the analysis. In column (2) we have the results of the estimation with CAIP restricting the influence of coffee. These results indicate that an

[61] We standardized CAIP variables to present a zero mean and standard deviation of one. The reason for this methodological decision lies in the fact that CAIP does not have a natural scale, so its standardization allows for easier interpretation of results without loss of information.

increase of one s.d. in potential export revenues reduces the share of illiterates by 11.1 p.p. In column (3) we have the results of the estimate with CAIP restricting the influence of sugar cane and the negative relationship is maintained to a lesser extent (the share of illiterates reduces by 2.7 p.p.). CAIP restricting the influence of cotton (column (4)) and of tobacco (column (5)) maintains the negative relationship between shocks of potential export revenues and the share of illiterates. Thus, an increase of one s.d. in potential export revenues reduces the share of illiterates by 6.8 and 13.5 p.p., respectively, for these two alternative version of CAIP. Column (6) indicates that an increase of CAIP by one s.d. reduces the share of illiterates by 13.5 p.p. when we restrict the influence of cocoa crop.

One advantage of CAIP over using separate components of commodities lies in the insurance effect that CAIP captures. Table 5 shows that an estimation using components of commodities (equation 2) instead of CAIP maintain the negative relationship between a positive shock in potential export revenues and the share of illiterates. It is important to highlight the greater variability of the CAIP's individual components relatively to CAIP complete specification, mainly due to CAIP's insurance effect. Thus, we can observe a greater magnitude of CAIP's individual components when compared with results in Table 4 (main results).

Table 6 presents the robustness tests of the main result considering other price components for CAIP. The negative relationship between CAIP and the share of illiterates in the average of the analyzed period remains robust and statistically significant at 1% in all specifications tested in Table 6. In Tables 7.A and 7.B we have the robustness tests of the main result controlling for geographic variables that use the structure of equation 6. All the results obtained in these tables show the robustness of the main result. The negative relationship between CAIP and the share of illiterates remains close to the estimate obtained in column (1) of Table 4 and is statistically significant at 1%.

7.2 Mechanisms

In order to understand a little more about the possible specific mechanisms that determine the negative relationship between CAIP and the share of illiterates, we estimate reduced-form equations. We use the same structure as in equation 4 and only use other response variables that can give us clues about these mechanisms. It is noteworthy that we used the

panel only with information from the years 1872 and 1960 due to the availability of the analyzed variables.[62]

The first set of response variables relates to the distribution of the workforce between the different sectors of economic activity and the share of people who came from foreign countries (immigrants) in the population. We can observe in Table 8.A that an increase of potential export revenues reduces the share of the population (aged 15 years or older) allocated in agriculture (-9.8 p.p.) and increases in industry (+6 p.p.) and in services (+3.8 p.p.). Column (4) presents the estimate of the relationship between CAIP and the share of immigrants in the population, but the result obtained is statistically insignificant. Results in Table 8.A are subsided by the result that the size of the population aged 15 years or more in an AMC is not associated with the level of their CAIP, on average in the analyzed period (column (5)).

These results indicate that AMC with the highest levels of CAIP were those that presented, on average in the period, a greater reallocation of their workforce in the industrial sector. A possible explanation for this result is that the AMC that benefited from a positive shock of export revenues presented spillovers from the gains of the agricultural sector to other sectors of economic activity. Thus, it may be that the development of the industrial sector created a greater need for literate labor and allowed educational investments. The 1872-1960 period represents the rise and consolidation of the industrial sector in Brazil so that an increase in exports revenues great contributed to its funding. Another possible explanation is the urbanization process occurring in this same period in Brazil, which reallocated rural population to the urban centers.

The second set of response variables relates to the population composition of AMC. We can observe in Table 8.B a lower representativeness of the black and mixed race population (aged 15 years or older) relatively to whites in localities with an increase of its potential export revenues. Thus, there is a difference in the relationship between CAIP and population representativeness according to race. One possibility is that a

[62] The main results in section 7.1 indicates that CAIP's impact over the share of illiterates is affected in a larger degree when we restrict sugar or cotton's price index at the year of 1872 level. Robustness results with the CAIP that restricts sugar or cotton are in Appendix 4 and Appendix 5. Results with restriction of sugar's price index reduces the estimates of the impact of CAIP over the share of the young population attending a school and reduces the estimates of the impact of CAIP over the migratory process. This same pattern happens to the results with restriction of cotton's price index. These results may indicate that some localities relied on sugar and cotton potential export revenues variability to expand your school system and therefore the impact of CAIP over the share of illiterates is underestimated when we restrict sugar or cotton's price index at the 1872 level.

portion of the black population had greater incentives to migrate and seek better opportunities after the period of abolition of slavery. Another possibility is that the white population had a higher fertility rate and a lower mortality rate in the analyzed period.

In Table 8.B we can also observe evidence that locations which experienced a positive shock of potential export revenues may have reinvested part of that revenue to expand their school systems. Column (4) of Table 8.B indicates an increase in potential export revenues increases the share of the young population attending a school (+9.4 p.p.). Absence of CAIP's impact over the size of young population (column (5) of Table 8.B) reinforces that this result comes from a school system expansion process.

The expansion of the school system, and consequently the reduction of illiteracy, may have been via an increase in the educational offer in which the government follows an investment agenda in the educational sector as a development policy. In addition, it may have been via an increased demand for educational services by families. We are not able to identify exactly which of these two effects predominated in the analyzed period, but the results of the estimates presented in Table 8 indicate that the reduction in illiteracy was accompanied by a higher school attendance among young people and a reallocation of labor from the agricultural sector to industrial and service sectors.

7.3 Slavery in 1872

Table 9 presents the results of the estimations of equation 7, which analyzes the influence of the share of enslaved people in the population in 1872 on the relationship between CAIP and the share of illiterates aged 15 years or older. One hypothesis considered is that AMC with a higher prevalence of slave labor at the end of the 19th century would have less democratic institutions and, thus, less investment in public education. The result in Table 9 indicates that the share of enslaved people in 1872 does not affect the negative relationship between CAIP and the share of illiterates.

The slave labor system restricted the educational and social opportunities of a considerable portion (15.2% of the total population in 1872) of the Brazilian population. Almost all enslaved people were illiterate (99.91% of the total enslaved in 1872) and this may have represented different starting points for the black, mixed race and white population. Other than the share of enslaved population, maybe institutional changes in the late 19th century in Brazil were more important to determine the educational trajectories for different social groups.

7.4 Institutional changes in the late 19th century

The end of the 19th century represented institutional changes for Brazilian society. The first change was the abolition of slavery in 1888 and the incorporation of the population freed from this system of forced labor into Brazilian society. The second institutional change was the end of the Empire and the proclamation of the Republic in 1889, which included political decentralization among the political changes. Investments in school systems were the responsibility of state governments during the Empire, but revenues from foreign trade tariffs, the main source of government revenue, were under the control of the central government. With the proclamation of the Republic, state governments began to control the revenues from export tariffs and were in a better position to invest in public goods.

Table 10 presents the results of the estimation of equation 4 that were presented in Table 4, but with the addition of the interaction term between CAIP and the binary variable that indicates whether the year is 1872. Thus, we have CAIP estimates conditioned on the periods before and after institutional changes in the late 19th century. Results in Table 10 indicate that a positive shock in potential export revenues reduces the share of illiterates by a greater magnitude in 1872 (-41.2 p.p.), relatively to 1960 (-11 p.p.). These results indicate that localities with higher levels of CAIP reduced their share of illiterates to a greater degree before institutional changes (in 1872) when compared to the result obtained for the subsequent moment.

The results presented in Table 10 can indicate three possible scenarios. First, the potential for industrialization was greater in the late 19th century than throughout the 20th century. Thus, the increase in potential export revenues may have acted to attract rural and foreigner workers, speed up the process of urbanization and increase the demand for education at the end of the 19th century. Concomitantly, the changes generated by the increase in these revenues were expelling the portion of the black population that was liberated by successive abolitionist laws to other less dynamic municipalities.

Another possibility is that export revenues had more importance to the public budget in the late 19th and early 20th centuries. Thus, it would be expected that the state governments' budget would be closely linked to export revenues in that period and, consequently, investments in public education would also be linked to these revenues. Throughout the 20th century tariffs on export revenues lost their importance to the public

budget, reaching less than 1% of state government revenues in 1958. This would explain the smaller magnitude of the relationship between potential export revenues and the share of illiterates after the institutional changes.

The third possible scenario is that educational investments were more efficient in 1872, compared to the later period. This greater efficiency may come from the fact that initially reducing illiteracy rates is possible through the construction of a few schools. Subsequently, reductions in illiteracy rates will increasingly require larger investments in education (building more schools and hiring more teachers and staff, in addition to the existence of a population growth rate and an increase in the cohort of children).

Table 11 shows the results of the mechanisms involving workforce allocations between different sectors of economic activity and the presence of a foreign population, considering the periods before and after the institutional changes. We note in columns (1) and (3) of Table 11 that the impact of potential export revenues on the share of the population aged 15 years or older allocated to the agriculture and services sectors occurs only in the period prior to institutional changes. An increase of CAIP reduces the share of the workforce in agriculture (-36.2 p.p.) and increases in services (+18.1 p.p.) in 1872. In column (2) of Table 11, an increase of potential export revenues increases the share of the workforce allocated in the industry throughout the entire period (+18.1 p.p. in 1872 and +2.8 p.p. in 1960). These results indicate that potential export revenues were closely linked to the reallocation of labor between sectors of economic activity in the period prior to institutional changes. Thus, there was a departure of the agricultural workforce towards industry and services in that period.

The results in Table 11 indicate that the period before the institutional changes of the late 19th century was important for the relationship between an increased potential export revenue and spillovers for industry and services. This can also contribute to the greater magnitudes of CAIP's impacts on reducing the share of illiterates during this period. This is because with the rise of industry and services there may have been a greater demand for schools and a greater budgetary capacity of governments to carry out educational investments.

The localities with the highest CAIP levels in 1872 also seem to have increased the share of their population who came from foreign countries. A positive shock of potential export revenues increases the share of foreigners in the population (+5.3 p.p.) in

1872. These results may be indicating the period of immigrants' arrival that were concentrated in the late 19th century and throughout the first half of the 20th century. Immigrants in Brazil had more schooling and demanded more for the construction of more schools, relatively to the local population. [63] In column (5) of Table 11 we see that an increase in potential export revenues reduces population aged 15 years or older (-16.8%) in average over the period from 1872 to 1960.

Table 12 reinforces the hypothesis that the black population showed migratory movements in the analyzed period. Considering the representativeness of social groups according to their declared race, we have that the black and mixed race population aged 15 years or more has lost representativeness in AMCs with higher levels of CAIP, relatively to whites. Black population lost representativeness before the institutional changes in late 19th century and mixed race population lost representativeness after these changes.

In column (4) of Table 12, we observe that AMC with higher levels of CAIP have higher shares of young people attending school both before and after the changes. The magnitude of the impact is greater in the period before the institutional changes. Results in column (5) of Table 12 indicates a greater young population in localities with higher potential export revenues in 1872, which reinforces the hypothesis of an expansion of the school system. In other words, we have a situation in which AMC with higher levels of CAIP showed an increase in the young population and at the same time that the share of these young people attending a school in 1872 increased. Table 12 (column (5)) also shows that the young population decreased in localities with higher potential export revenues in 1960. This reduction in the youth population may result from the urbanization process, which reduced fertility rates in the urban centers compared to the rural areas.

Results in this section shows evidence that impact of potential export revenues on reducing illiteracy were greater in the period prior to institutional changes in the late 19th century. A major possible explanation for the results presented in this section is the arrival of foreigner population, mainly Europeans, which were more educated and demanded for

[63] Rocha et al. (2017) indicates that settlement policies financed by Brazilian government attract immigrants who had more human capital (in terms of years of schooling) than the local population. A few places like the State of São Paulo adopted specific settlement policies that focused mainly in highly educated and skilled immigrants. De Carvalho Filho and Colistete (2010) indicates that immigrants were from countries with a more consolidated educational public system. When these immigrants settled in Brazil, they organized a more structured demand for public investments in education, mainly the construction of more schools.

more schools in the locations where they settled. Concomitantly, we can observe reallocations of the agricultural workforce to industry and service sectors in the late 19th century, which could incentive investments in education.

There still evidence that corroborates hypothesis of the loss of importance of export earnings to the public budget throughout the 20th century and of expelling of the black and mixed race population from localities with higher potential export revenues. The results indicate that AMC with higher levels of potential export revenues showed a reduction in the representativeness of its black and mixed race population throughout the 20th century. Concomitantly, if we consider the presence of racial discrimination in Brazilian society in the analyzed period, we have that the white population faced better conditions in the labor market and income that may have influenced their higher fertility rates (or lower mortality rates) and their financial capacity to obtain educational services for their children.

7.5 Volatility

The last analysis presented in this paper investigates the relationship between the volatility of CAIP and the share of illiterates in the population aged 15 years or older. There is a possibility that localities that experienced high volatility in their potential export revenues may have had difficulties maintaining positive levels of investment in public goods. However, the results presented in Table 13 are inconclusive in this analysis. This is because in column (1) of Table 13 we see that the volatility of CAIP has no impact on the share of illiterates in the average of the period of 1872, 1920, 1940 and 1960. Column (2) of Table 13 presents the results considering the CAIP restricting coffee's price index and the estimate obtained indicates that an increase of one s.d. in the volatility of CAIP increases the share of illiterates by 7.7 p.p. on the average of the period.

8 - Conclusion

The present work finds that positive and exogenous shocks in export revenues had the potential to reduce the share of illiterates in Brazilian municipalities in the late 19th century and throughout the 20th century. This reduction was stronger in the late 19th century, with the entrance of immigrants and agriculture and foreign trade revenues were paramount for the Brazilian economy and for the public budget of local governments. The results control for the fixed effects of the municipalities that include the institutions and the geographic factors that are constant in the time.

We found two main mechanisms that mediate the negative relationship between potential export revenues and the share of illiterates. The first is the selection of inhabitants in the municipalities. Positive shocks to these revenues attract immigrants in the late 19th century, increase the share of the population working in the industrial and service sectors, and create an expelling movement of the black and mixed race population, especially young people. These changes in the composition of the inhabitants seem to lead to a process of greater demand for educational services and possible expel of the population without (or with little) schooling. Thus, this process contributes to reduce the share of illiterates in these locations.

The second mechanism appears to be via the income and fertility effect of the white population. Municipalities that experienced positive shocks from potential export revenues showed an increase in the representativeness of the white young population in the analyzed period. Considering a context of racial discrimination in the labor market in which the white population could get better jobs and, consequently, a higher income, that population would be able to demand more educational services or even finance their children's education with their own resources.

References

Abreu, M. P.; Bevilaqua, A. S. Brazil as an Export Economy, 1880-1930. In: Cárdenas, E.; Ocampo, J. A.; Thorp, R. (Eds.) An Economic History of Twentieth-Century Latin America, Palgrave Macmillan, Chapter 2, pp. 32-54, 2000.

Acemoglu, D.; Johnson, S.; Robinson, J. A. The Colonial Origins of Comparative Development: An Empirical Investigation. The American Economic Review 91 (5), 1369-1401, 2001.

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

Acemoglu, D.; Robinson, J. A. Why did the West Extend the Franchise? Democracy, Inequality, and Growth in Historical Perspective. The Quarterly Journal of Economics 115, 1167-1199, 2000.

Aguëro, J. M.; Balcázar, C. F.; Maldonado, S. Ñopo, H. The value of redistribution: Natural resources and the formation of human capital under weak institutions. Journal of Development Economics, 148, 2021.

Banerjee, A.; Iyer, L. History, Institutions, and Economic Performance: The legacy of Colonial Land Tenure Systems in India. The American Economic Review, Vol. 95, No. 4, pp. 1190-1213, Sep. 2005.

Blattman, C.; Hwang, J.; Williamson, J. G. Winners and losers in the commodity lottery: The impact of terms of trade growth and volatility in the Periphery 1870-1939. Journal of Development Economics, 82, pp. 156-179, 2007.

Bleaney, M.; Greenaway, D. The impact of terms of trade and real Exchange rate volatility on investment and growth in sub-Saharan Africa. Journal of Development Economics, Vol. 65, pp. 491-500, 2001.

Bourguignon, F.; Verdier, T. Oligarchy, democracy, inequality and growth. Journal of Development Economics, Vol. 62, 285-313, 2000.

Brea, J. A. Population Dynamics in Latin America. Population Bulletin, Vol. 58, No. 1, March 2003.

Bruhn, M.; Gallego, F. A. Good, Bad, and Ugly Colonial Activities: Do They Matter for Economic Development? The Review of Economics and Statistics, 94 (2): 433-461, May 2012.

Bucciferro, J. R. Racial Inequality in Brazil from Independence to the Present. In: Has Latin American Inequality Changed Direction? Looking Over the Long Run. Eds: Bértola, L.; Williamson, J. Springer, pp. 171-194, 2017.

Bustos, P.; Caprettini, B; Ponticelli, J. Agricultural Productivity and Structural Transformation: Evidence from Brazil. American Economic Review, 106 (6): 1320-1365, 2016.

Caselli, F.; Michaels, G. Do Oil Windfalls Improve Living Standards? Evidence from Brazil. American Economic Journal: Applied Economics, Vol. 5, No. 1, January 2013.

Cogneau, D.; Jedwab, R. Commodity Price Shocks and Child Outcomes: The 1990 Cocoa Crisis in Côte d'Ivoire. Economic Development and Cultural Change, Vol. 60, No. 3, pp. 507-534, 2012.

Dammert, A. C. Child labor and schooling response to changes in coca production in rural Peru. Journal of Development Economics, 86, 164-180, 2008.

de Carvalho Filho, I.; Colistete, R. P. Education Performance: Was It All Determined 100 Years Ago? Evidence From São Paulo, Brazil. MPRA Paper No. 24494, Economic History Association – 70th Annual Meeting, 2010.

Ehrl, P. Minimum comparable areas for the period 1872-2010: an aggregation of Brazilian municipalities. Estudos Econômicos, vol. 47, n. 1, pp. 215-229, São Paulo, jan./mar. 2017.

Engerman, S.; Mariscal, E.; Sokoloff, K. The Evolution of Schooling in the Americas, 1800–1925. In: Eltis, D.; Lewis, F.; Sokoloff, K. (Eds.), Human Capital and Institutions: A Long-Run View (pp. 93-142). Cambridge: Cambridge University Press, 2009.

Engerman, S. L.; Sokoloff, K. L. Factor Endowments, Institutions, and Differential Paths of Growth Among New World Economies. A View from Economic Historians of the United States. In: How Latin America Fell Behind. Essays on the Economic Histories of Brazil and Mexico, 1800-1914. Ed: Haber, S. Stanford University Press, Chapter 10, pp. 260-304, 1997.

Engerman, S. L.; Sokoloff, K. L. Factor Endowments, Inequality, and Paths of Development among New World Economies. Economía, Vol. 3, No. 1, pp. 41-109, Fall, 2002.

Erten, B.; Ocampo, J. A. Super-cycles of commodity prices since the mid-nineteenth century.DESA Working Paper No. 110, February 2012.

Fausto, B. História do Brasil. Editora da Universidade de São Paulo, 12 ed., 1 reimpr. (2006), São Paulo, 1994.

Feler, L.; Senses, M. Z. Trade Shocks and the Provision of Local Public Goods. American Economic Journal: Economic Policy, 9(4): 101-143, 2017.

Ferreira, F. H. G.; Schady, N. Aggregate Economic Shocks, Child Schooling, and Child Health. The World Bank Research Observer, vol. 24, no. 2, Aug. 2009.

Fischer, G.; Nachtergaele, F. O.; Prieler, S.; Teixeira, E.; Tóth, G.; van Velthuizen, H.; Verelst, L.; Wiberg, D. GAEZ ver 3.0. Global Agro-ecological Zones, Model

Documentation. International Institute for Applied Systems Analysis. Food and Agriculture Organization of the United Nations, 2012.

Fiszbein, M. Agricultural Diversity, Structural Change and Long-run Development: Evidence from the U.S. NBER Working Paper, No. 23183, 2017.

Furtado, C. Formação Econômica do Brasil. Companhia das Letras, ed. 1 (2007), São Paulo, 1959.

Galor, O.; Moav, O.; Vollrath, D. Inequality in Landownership, the Emergence of Human-Capital Promoting Institutions, and the Great Divergence. Review of Economic Studies, 76, pp. 143-179, 2009.

Hornbeck, R.; Keskin, P. Does Agriculture Generate Local Economic Spillovers? Short-Run and Long-Run Evidence from the Ogallala Aquifer. American Economic Journal: Economic Policy, 7 (2): 192-213, 2015.

Jacks, D. S. From boom to bust: a typology of real commodity prices in the long run. Cliometrica, 13, pp. 201-220, 2019.

Jacks, D. S.; O'Rourke, K. H.; Williamson, J. G. Commodity Price Volatility and World Market Integration Since 1700. The Review of Economics and Statistics, Vol. 93, No. 3, pp. 800-813, August 2011.

Kang, T. H. Descentralização e financiamento da educação brasileira: uma análise comparativa, 1930-1964. Revista Estudos Econômicos, Vol. 41, Nº 3, pp. 573-598, Jul-Set 2011.

Kruger, D. I. Coffee production effects on child labor and schooling in rural Brazil. Journal of Development Economics, 82, 448-463, 2007.

Lagerlöf, N. Slavery and Other Porperty Rights. The Review of Economic Studies, Vol. 76, No. 1, pp. 319-342, Jan. 2009.

Leff, N. H. Economic Development and Regional Inequality: Origins of the Brazilian Case. The Quarterly Journal of Economics, Vol. 86, No. 2, pp. 243-262, May 1972.

López, J. H.; Perry, G. Inequality in Latin America: Determinants and Consequences. Policy Research Working Paper 4504, The World Bank, Feb. 2008.

Maluccio, J. A. Coping with the “Coffee Crisis” in Central America: The Role of the Nicaraguan Red de Protección Social. FCND Discussion Paper 188, Feb. 2005.

Mello, J. M. C. O capitalismo tardio. Brasiliense, 8ª ed. (1991), São Paulo, 1982.

Miller, G.; Urdinola, B. P. Cyclicality, Mortality, and the Value of Time: The Case of Coffee Price Fluctuations and Child Survival in Colombia. Journal of Political Economy, vol. 118, no. 1, 2010.

Musacchio, A.; Fritscher, A. M.; Viarengo, M. Colonial Institutions, Trade Shocks, and the Diffusion of Elementary Education in Brazil, 1889-1930. The Journal of Economic History, Vol. 74, No. 3, Sep. 2014.

Naritomi, J.; Soares, R. R.; Assunção, J. J. Institutional Development and Colonial Heritage within Brazil. The Journal of Economic History, Vol. 72, No. 2, June 2012.

Nunn, N.; Qian, N. The Potato’s Contribution to Population and Urbanization: Evidence from a Historical Experiment. The Quarterly Journal of Economics, 126, 593-650, 2011.

Papadia, A. Slaves, Migrants and Development in Brazil, 1872-1923. European University Institute Working Paper MWP 2019/05, Max Weber Programme, 2019.

Ramey, G.; Ramey, V. A. Cross-Country Evidence on the Link Between Volatility and Growth. The American Economic Review, Vol. 85, No. 5, pp. 1138-1151, Dec. 1995.

Reis, E.; Pimentel, M.; Alvarenga, A. L.; Santos, M. C. H. Áreas mínimas comparáveis para os períodos intercensitários de 1872 a 2000. Rio de Janeiro: Ipea/Dimac, 2008.

Rocha, R.; Ferraz, C.; Soares, R. R. Human Capital Persistence and Development. American Economic Journal: Applied Economics, 9 (4): 105-136, 2017.

Sivakumar, A.; Bhat, C. Fractional Split-Distribution Model for Statewide Commodity-Flow Analysis. Transportation Research Record, 1790(1), pp. 80-88, Jan. 2002.

Soares, R. R.; Assunção, J. J.; Goulart, T. F. A note on slavery and the roots of inequality. Journal of Comparative Economics, 40, pp. 565-580, 2012.

Summerhill, W. Colonial Institutions, Slavery, Inequality, and Development: Evidence from São Paulo, Brazil. MPRA Paper No. 22162, 2010.

Townsend, J. Price and consumption of tobacco. British Medical Bulletin, 52 (No. 1): 132-142, 1996.

Varsano, R. A evolução do sistema tributário brasileiro ao longo do século: anotações e reflexões para futuras reformas. Texto para discussão nº 405, IPEA, 1996.

Villela, A. Distribuição Regional das Receitas e Despesas do Governo Central no II Reinado, 1844-1889. Estudos Econômicos, São Paulo, V. 37, N. 2, pp 247-274, Abr-Jun, 2007.

Wagner, F. E.; Ward, J. O. Urbanization and Migration in Brazil. The American Journal of Economics and Sociology, Vol. 39, No. 3, pp. 249-259, Jul. 1980.

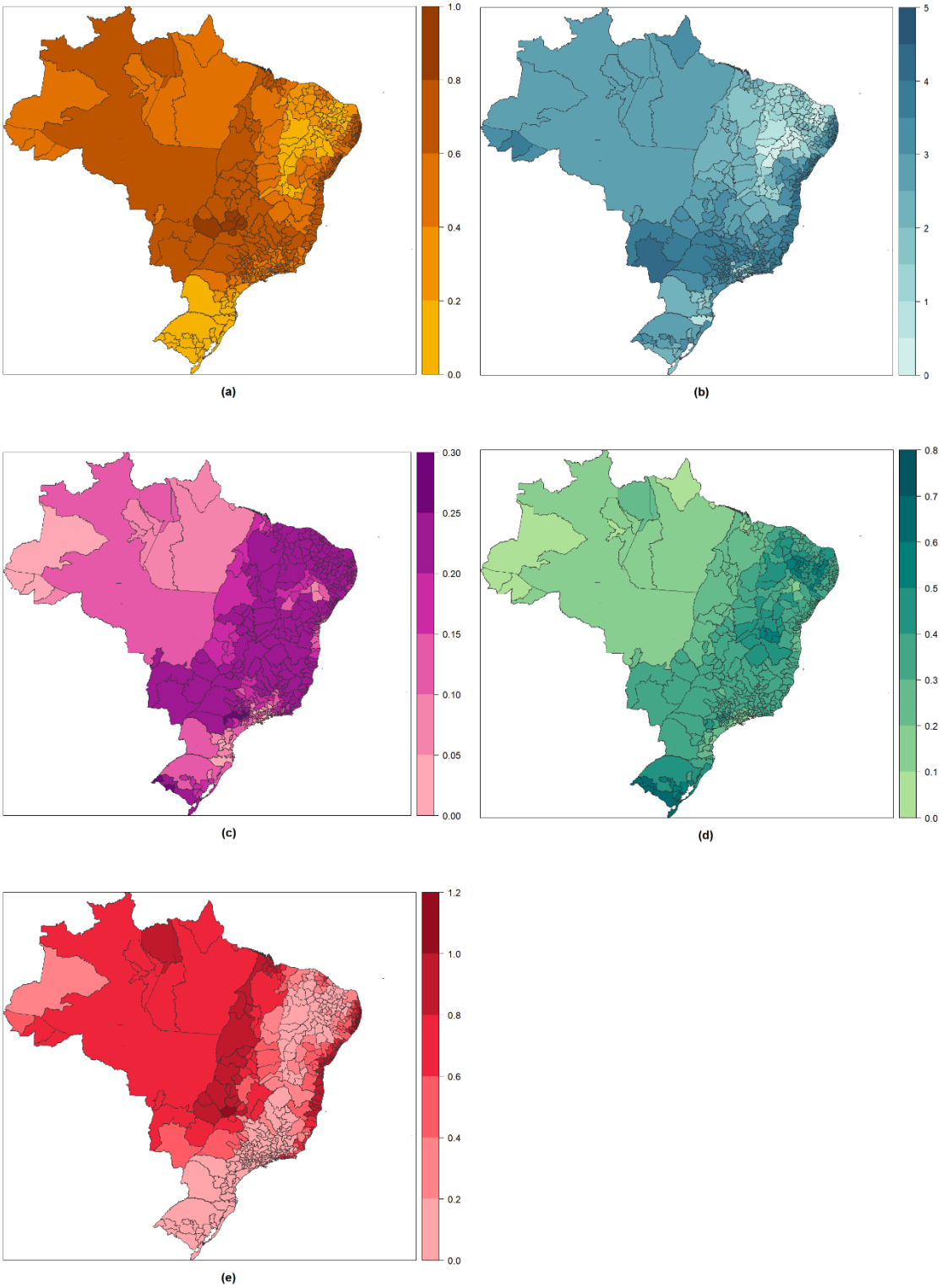
Wegenast, T. Cana, Café, Cacau: Agrarian Structure and Educational Inequalities in Brazil. Revista de Historia Económica, Journal of Iberian and Latin American Economic History, Vol. 28, No. 1: 103-137, 2010.

Wigton-Jones, E. Legacies of inequality: the case of Brazil. Journal of Economic Growth, 25: 455-501, 2020.

Woodard, J. P. Coronelismo in Theory and Practice: Evidence, Analysis, and Argument from São Paulo. Luso-Brazilian Review, Vol. 42, No. 1, pp. 99-117, 2005.

9 Figures

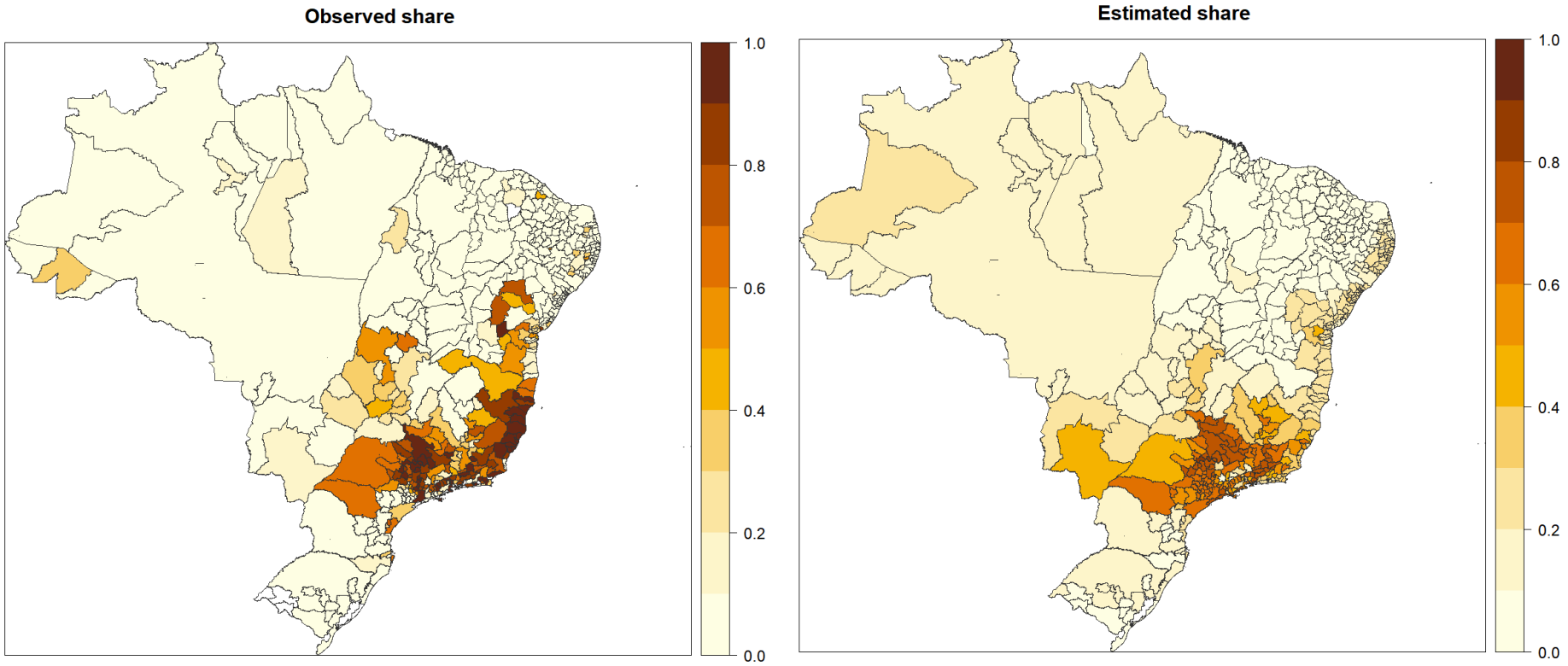
Figure 1. Spatial distribution of maximum attainable yield (in tons per hectare) of (a) coffee, (b) sugar cane, (c) cotton, (d) tobacco, and (e) cocoa crop.



Notes: Values of maximum attainable yield in tons per hectare. Brazilian localities divided through *Áreas Minimamente Comparáveis* (AMC) for the period of 1872-2010.

Sources: See Appendix 1.

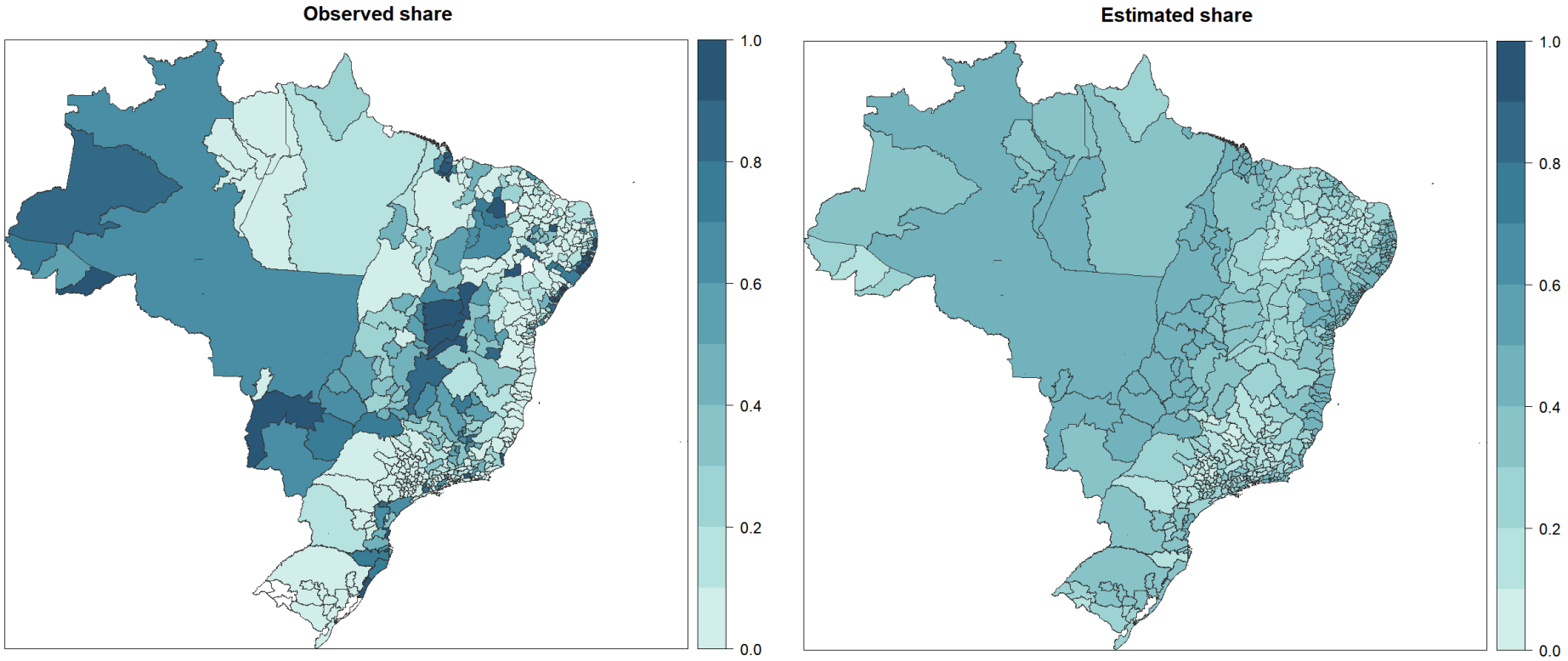
Figure 2. Spatial distribution of the observed and estimated production value share of coffee by AMC.



Notes: Observed production value share computed with observed coffee production (in tons) from 1920 and observed local prices of coffee (in *mil-réis*). Estimated production value share obtained through equation 1. Brazilian localities divided through *Áreas Minimamente Comparáveis* (AMC) for the period of 1872-2010.

Sources: See Appendix 1.

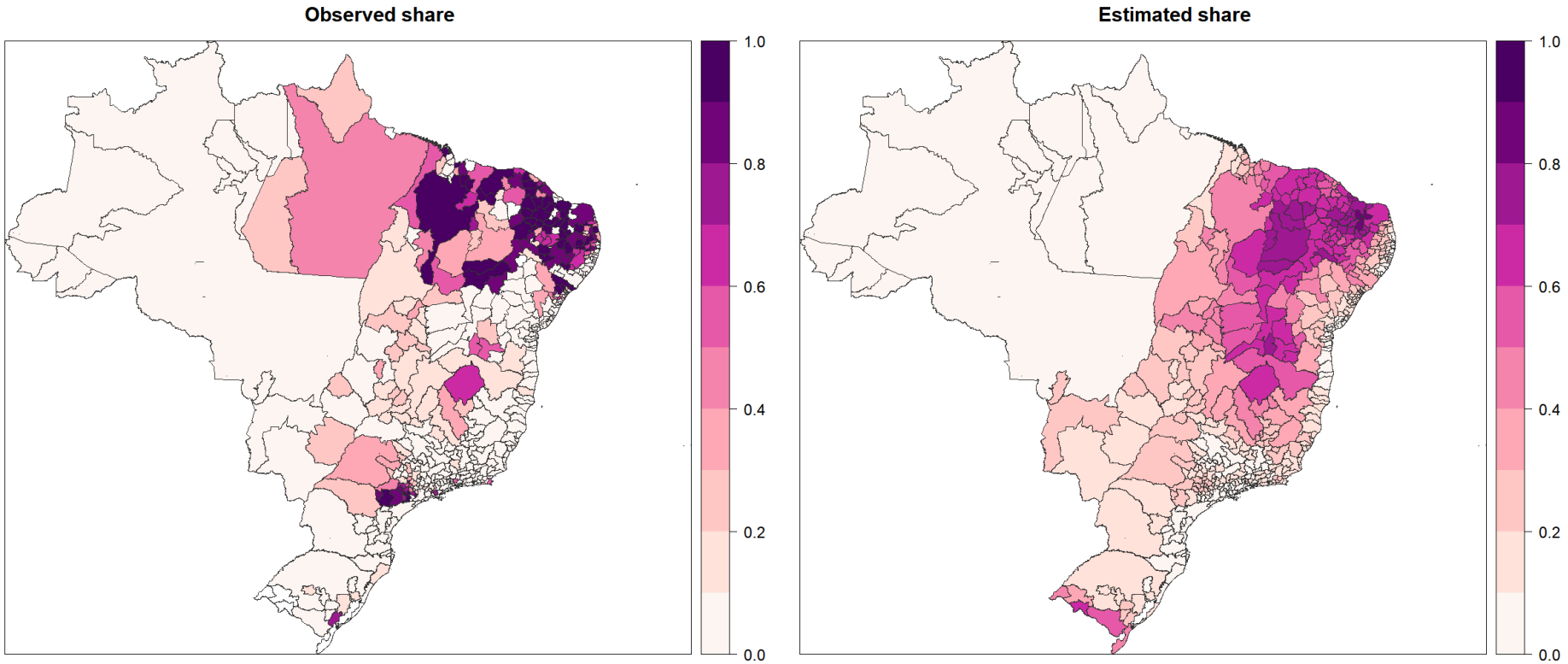
Figure 3. Spatial distribution of the observed and estimated production value share of sugar cane by AMC.



Notes: Observed production value share computed with observed sugar cane production (in tons) from 1920 and observed local prices of sugar (in *mil-réis*). Estimated production value share obtained through equation 1. Brazilian localities divided through *Áreas Minimamente Comparáveis* (AMC) for the period of 1872-2010.

Sources: See Appendix 1.

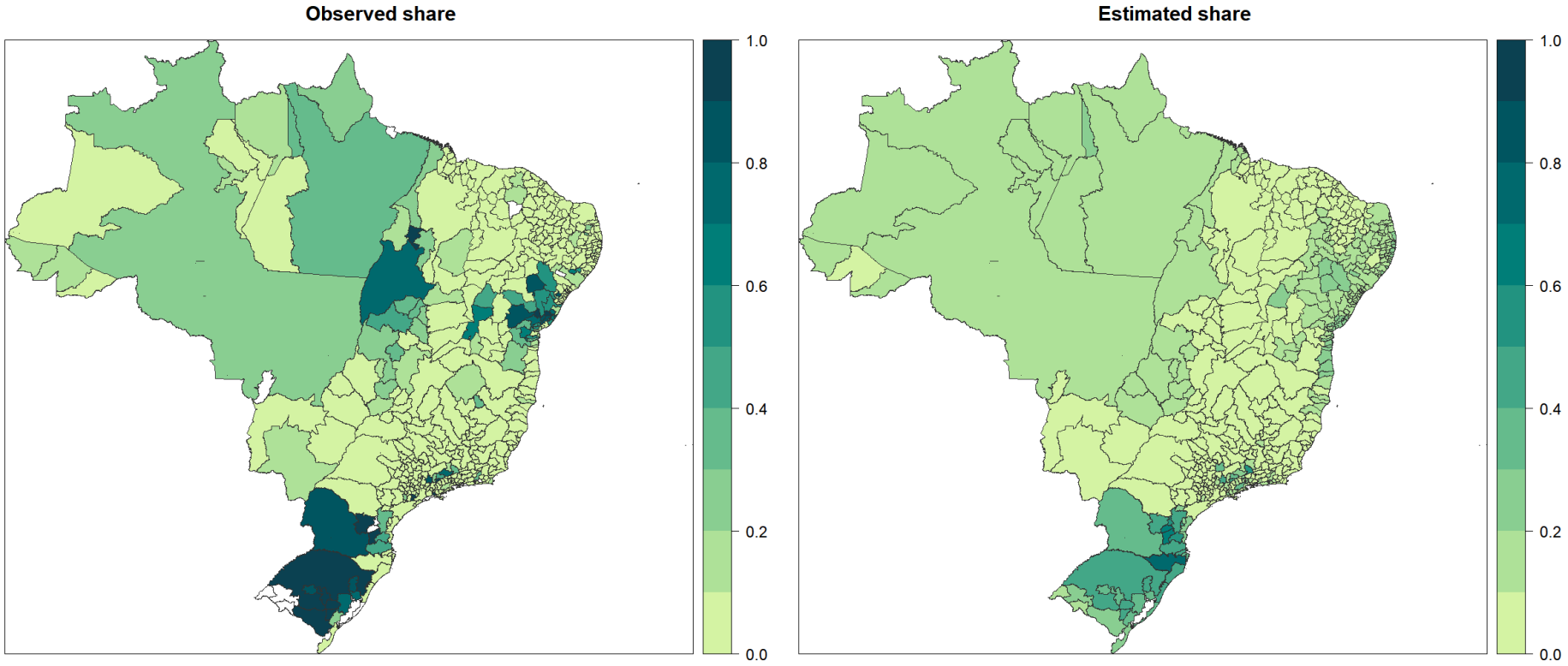
Figure 4. Spatial distribution of the observed and estimated production value share of cotton by AMC.



Notes: Observed production value share computed with observed cotton production (in tons) from 1920 and observed local prices of cotton (in *mil-réis*). Estimated production value share obtained through equation 1. Brazilian localities divided through *Áreas Minimamente Comparáveis* (AMC) for the period of 1872-2010.

Sources: See Appendix 1.

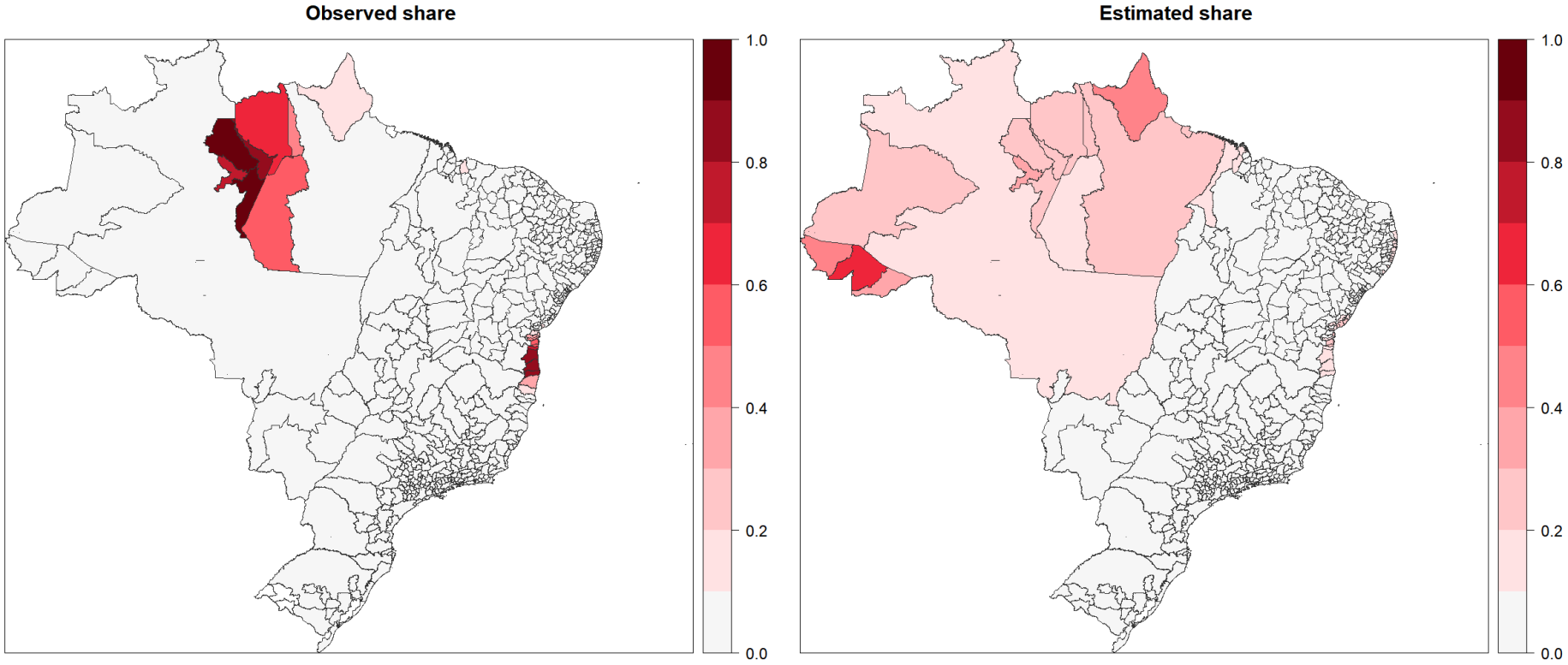
Figure 5. Spatial distribution of the observed and estimated production value share of tobacco by AMC.



Notes: Observed production value share computed with observed tobacco production (in tons) from 1920 and observed local prices of tobacco (in *mil-réis*). Estimated production value share obtained through equation 1. Brazilian localities divided through *Áreas Minimamente Comparáveis* (AMC) for the period of 1872-2010.

Sources: See Appendix 1.

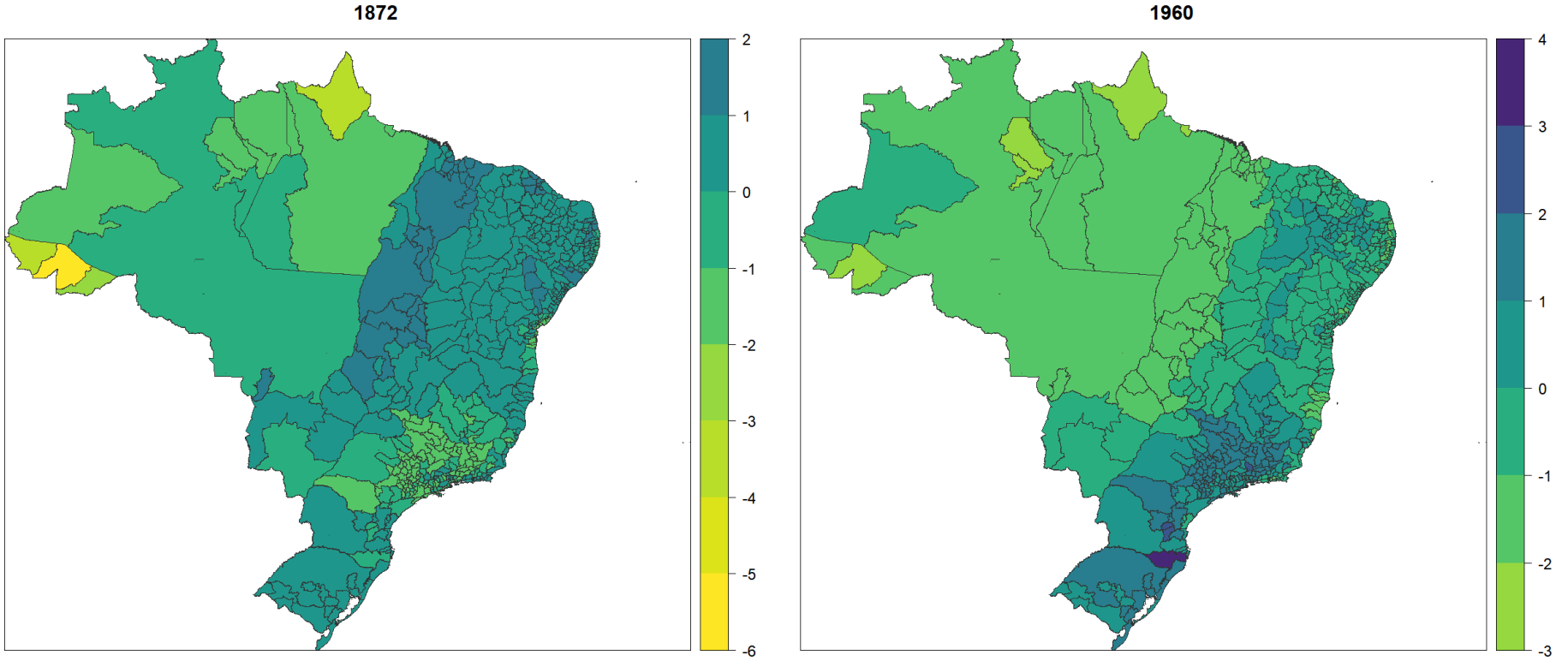
Figure 6. Spatial distribution of the observed and estimated production value share of cocoa by AMC.



Notes: Observed production value share computed with observed cocoa production (in tons) from 1920 and estimated local prices of cocoa (in *mil-réis*). The cocoa crop has no local prices in the *Anuário Estatístico do Brazil* of 1908-1912. Thus, we decided to use the Jacks real price index (2019) and convert its value to *mil-réis*. First, the dollar amount was converted to British pounds using "Lawrence H. Officer," Exchange Rates Between the United States Dollar and Forty-one Currencies, ", MeasuringWorth, 2020" available at http://www.measuringworth.com/exchangeglobal/ and accessed on 22/01/2020. The conversion of pounds sterling to *mil-réis* computed using data from "*Estatísticas históricas do Brasil: séries econômicas, demográficas e sociais de 1550 a 1988. 2. ed. rev. e atual. do v. 3 de Séries estatísticas retrospectivas. Rio de Janeiro: IBGE, 1990*". Estimated production value share obtained through equation 1. Brazilian localities divided through *Áreas Minimamente Comparáveis* (AMC) for the period of 1872-2010.

Sources: See Appendix 1.

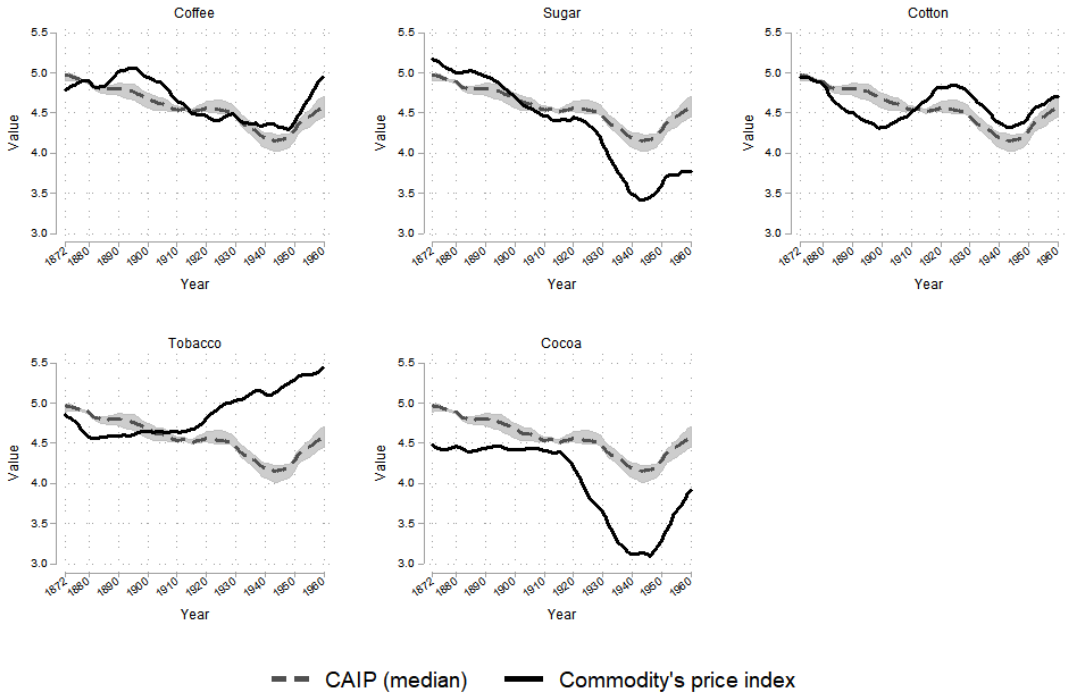
Figure 7. Spatial distribution of CAIP in 1872 and 1960 by AMC.



Notes: CAIP computed as in equation 3. CAIP's price component considers the average commodity price in the twenty years prior to the reference date. Brazilian localities divided through *Áreas Minimamente Comparáveis* (AMC) for the period of 1872-2010.

Sources: See Appendix 1.

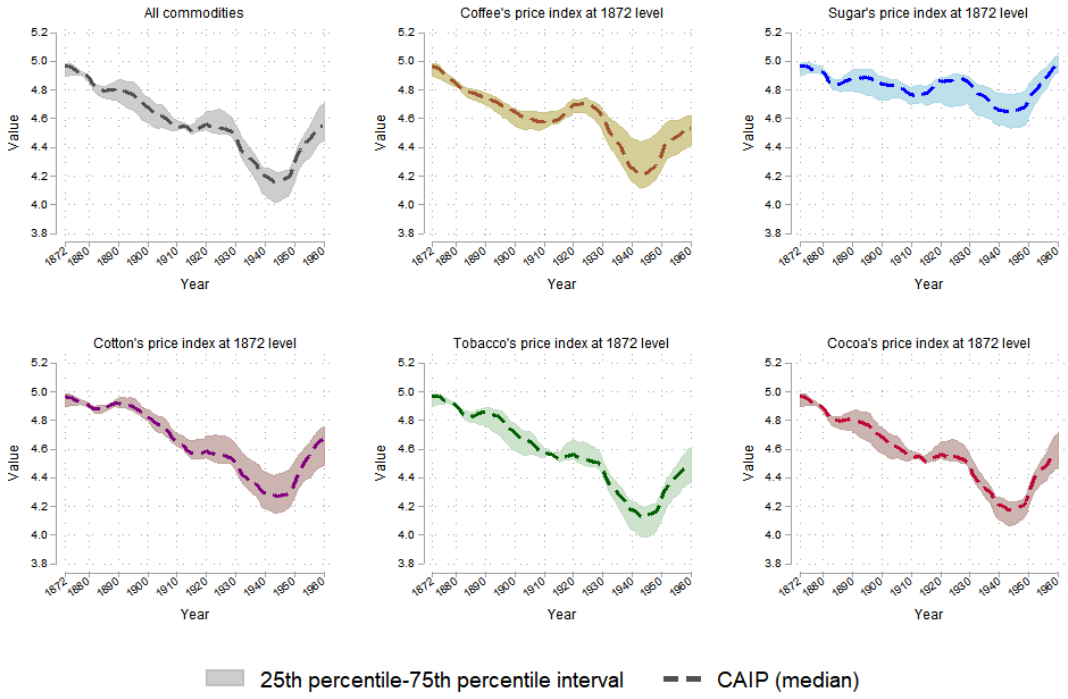
Figure 8. Temporal variation of CAIP and of commodity's price index, considering the CAIP's variability at AMC level.



Notes: Shaded area represents the interval between the 75th percentile (upper bound) and the 25th percentile (lower bound) of the CAIP values for the analyzed AMC. CAIP computed as in equation 3. CAIP computed using the price component considering the average price of the twenty years prior to the reference date. Commodity real price indexes averaged over the twenty years prior to the reference date. Price indexes in natural logarithm.

Sources: See Appendix 1.

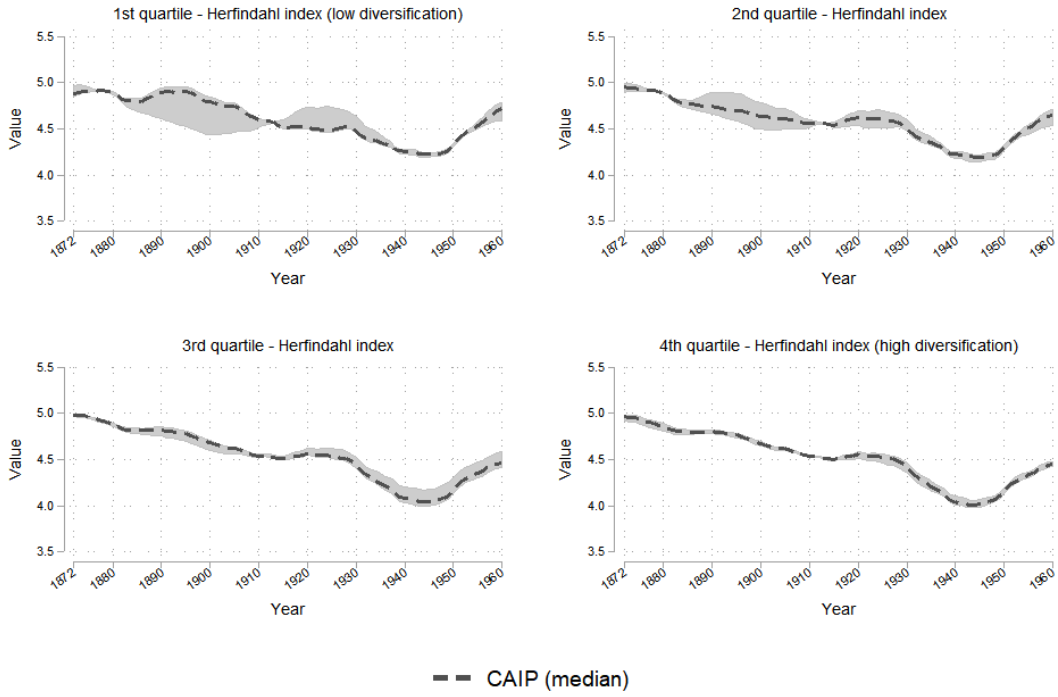
Figure 9. Temporal variation of CAIP and alternative versions of CAIP, considering the CAIP's variability at AMC level.



Notes: Shaded area represents the interval between the 75th percentile (upper bound) and the 25th percentile (lower bound) of the CAIP values for the analyzed AMC. CAIP computed using the price component considering the average price of the twenty years prior to the reference date. Commodity real price indexes averaged over the twenty years prior to the reference date. Price indexes in natural logarithm. 'All commodities' indicate the CAIP calculated in equation 3. 'Commodity's price index at 1872 level' indicates the CAIP calculated in equation 5 that restricts the commodity's price index at its level of 1872.

Sources: See Appendix 1.

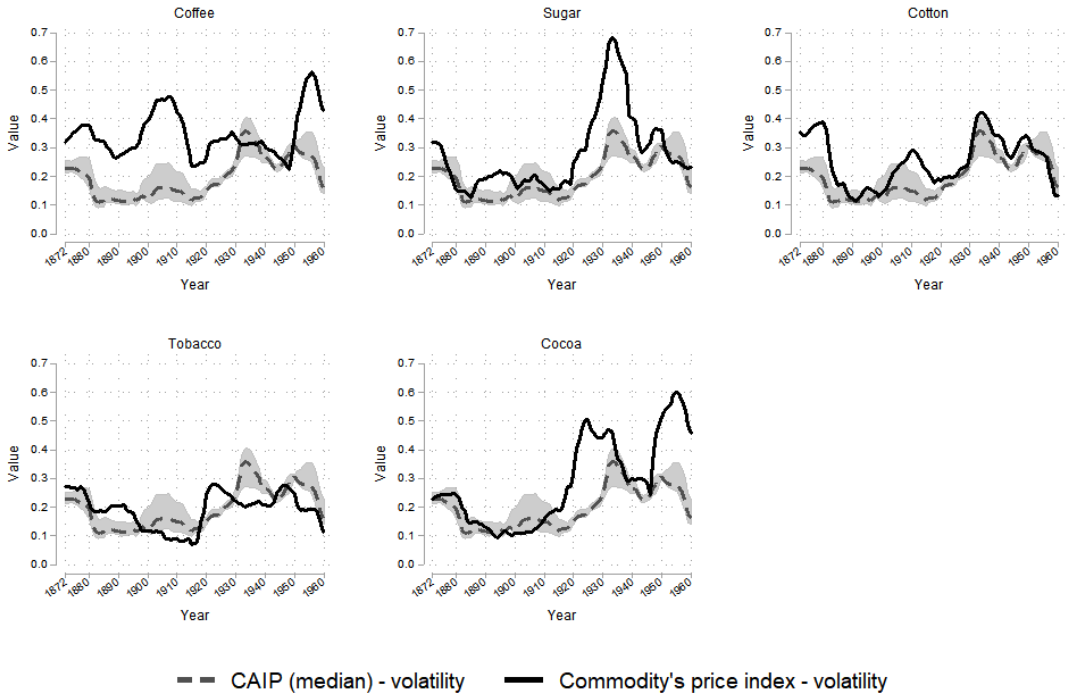
Figure 10. CAIP's dispersion according to the degree of agricultural diversification in AMC from 1872 to 1960.



Notes: 'CAIP (median)' indicates the median value of the CAIP in each group. Shaded area represents the interval between the 75th percentile (upper bound) and the 25th percentile (lower bound) of the CAIP. CAIP computed as in equation 3. CAIP computed using the price component considering the average price of the twenty years prior to the reference date. Agricultural diversification computed using the Herfindahl index. We compute Herfindahl index as $\left(1 - \sum_{i=1}^{I} \widehat{\theta_{ic}}^{2}\right)$ for each AMC c and it assumes values between zero and one. $\widehat{\theta_{ic}}$ represents the estimated production value share (see equation 1) for commodity i and AMC c. Values of the Herfindahl index close to zero indicate low agricultural diversification and values close to one indicate high agricultural diversification. Figure separated by quartiles of the Herfindahl index values.

Sources: See Appendix 1.

Figure 11. Temporal variation of CAIP (volatility) and of commodity's price index, considering the variability of CAIP values (volatility) at AMC level.



Notes: Shaded area represents the interval between the 75th percentile (upper bound) and the 25th percentile (lower bound) of the CAIP values (volatility) of the analyzed AMC. CAIP (volatility) computed using the annual price component and considering the standard deviation of the annual CAIP in the twenty years prior to the reference date. Volatility of commodity real price indexes computed from the standard deviation of these indices in the twenty years prior to the reference date. Price indices in natural logarithm.

Sources: See Appendix 1.

9 Tables

Table 1. Brazil's commodities export basket in 1920s (in percentage of the total of exports value).

Period	Coffee	Sugar	Cotton	Tobacco	Cocoa	Others
1919-1923	58.8	4.7	3.4	2.6	3.3	27.2
1924-1929	72.5	0.4	1.9	2.0	3.3	19.9

Note: Data from early 20th century Brazil, recorded in Villela and Suzigan (1973, p. 70) *Política do Governo e Crescimento da Economia Brasileira, 1889-1945*; reproduced in Fausto (1994, p. 292).

Table 2. Descriptive statistics of variables in the unbalanced panel (years of 1872, 1920, 1940, and 1960).

Variables:	**Mean in 1872**	**Mean in 1920**	**Mean in 1940**	**Mean in 1960**
Share of illiterates	0.940	0.707	0.672	0.466
Share of enslaved population in 1872	0.144			
Share of enslaved black population in 1872	0.522			
Share of enslaved mixed race population in 1872	0.163			
Number of AMC	475	482	479	278[1]
Potential export variables:	**Mean in 1872**	**Mean in 1920**	**Mean in 1940**	**Mean in 1960**
CAIP	4.943	4.588	4.164	4.566
CAIP restricting coffee	4.943	4.684	4.292	4.520
CAIP restricting sugar	4.943	4.804	4.662	4.982
CAIP restricting cotton	4.943	4.620	4.314	4.629
CAIP restricting tobacco	4.943	4.593	4.132	4.496
CAIP restricting cocoa	4.943	4.595	4.198	4.451
CAIP (volatility)	0.232	0.148	0.272	0.438
Number of AMC	482	482	482	482

Notes: Unbalanced panel considers data of 1872, 1920, 1940, and 1960. [1] Data for 1960 is available only for a sample of Brazilian municipalities, which explain the presence of only 278 AMC. Share of illiterates computed over population aged 15+. Mixed race stands for '*pardo*' population. CAIP version restricting a commodity follows equation 5, which restricts commodity's price index at its value of 1872. CAIP considers average of prices in twenty years prior to t for its price component (see equation 3). CAIP (volatility) computed over standard deviation of annual CAIP values (see section 3). Share of population who are enslaved computed over total population.

Sources: See Appendix 1.

Table 3. Descriptive statistics of variables in the balanced panel (years of 1872, and 1960).

Variables:	**Mean in 1872**	**Mean in 1960**
Share of illiterates	0.940	0.468
Share of adult population working in agriculture	0.534	0.633
Share of adult population working in industry	0.058	0.139
Share of adult population working in services	0.408	0.228
Share of immigrants in population	0.020	0.014
Total adult population	9.307	11.383
Share of whites in adult population	0.405	0.560
Share of blacks in adult population	0.195	0.101
Share of mixed race in adult population	0.399	0.339
Total young population	8.218	10.403
Share of young population who attends a school	0.127	0.500
Share of whites in young population	0.413	0.553
Share of blacks in young population	0.175	0.092
Share of mixed race in young population	0.412	0.355
Number of AMC	274	274
Potential export variables:	**Mean in 1872**	**Mean in 1960**
CAIP	4.945	4.571
CAIP (volatility)	0.232	0.438
Number of AMC	274	274

Notes: Data for 1960 is available only for a sample of Brazilian municipalities, which explain the presence of only 274 AMC in balanced panel. Adult population considers population aged 15 years or older. Young population considers population aged 7 to 14 years old. Share of illiterates computed over adult population (aged 15+). Total young population in natural logarithm. Total adult population in natural logarithm. Mixed race stands for '*pardo*' population. CAIP considers average of prices in twenty years prior to t for its price component (see equation 3). CAIP (volatility) computed over standard deviation of annual CAIP values (see section 3). Share of adult population working in agriculture, industry and services calculated to sum to 1.

Sources: See Appendix 1.

Table 4. Relationship between CAIP and share of illiterates aged 15 years or older.

1872, 1920, 1940, and 1960 panel Response variable: share of illiterates aged 15 years or older						
	(1)	(2)	(3)	(4)	(5)	(6)
CAIP (+1 s.d.)	-0.140	-0.111	-0.027	-0.068	-0.135	-0.135
	(0.011)	(0.008)	(0.006)	(0.009)	(0.011)	(0.011)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
CAIP's specification	{All commodities}	{Restrict coffee}	{Restrict sugar}	{Restrict cotton}	{Restrict tobacco}	{Restrict cocoa}
(1,714 observations of 482 AMC)						

Standard errors in parentheses.

P-value in brackets.

Notes: Standard errors (in parentheses) are cluster at the AMC level. Results includes AMC's fixed-effect and binary variables of the years 1920, 1940, and 1960. CAIP variable standardized to have a zero mean and standard deviation of one considering values for all AMC and all panel years. CAIP with all commodities considers all five agricultural crops (coffee, sugar cane, cotton, tobacco, and cocoa) and average of prices in twenty years prior to t for its price component (see equation 3). CAIP version restricting a commodity follows equation 5, which restricts commodity's price index at its value of 1872. Share of illiterates aged 15 years or older computed over adult population aged 15 years or older. Data from unbalanced panel.

Table 5. Relationship between CAIP's components and share of illiterates aged 15 years or older.

1872, 1920, 1940, and 1960 panel Response variable: share of illiterates aged 15 years or older	
Coffee component	-0.316
	(0.058)
	[0.000]
Sugar component	-0.257
	(0.023)
	[0.000]
Cotton component	-0.670
	(0.050)
	[0.000]
Tobacco component	-0.147
	(0.062)
	[0.019]
Cocoa component	-0.094
	(0.021)
	[0.000]
(1,714 observations of 482 AMC)	

Standard errors in parentheses.

P-value in brackets.

Notes: Standard errors (in parentheses) are cluster at the AMC level. Results includes AMC's fixed-effect and binary variables of the years 1920, 1940, and

1960. Component variables standardized to have a zero mean and standard deviation of one considering values for all AMC and all panel years. Commodity component represents interaction between commodity's estimated share of agricultural value production and commodity's price component (see equation 2), which is a component of CAIP. Share of illiterates aged 15 years or older computed over adult population aged 15 years or older. Data from unbalanced panel.

Table 6. Relationship between CAIP and share of illiterates aged 15 years or older, considering alternative CAIP's price component.

1872, 1920, 1940, and 1960 panel Response variable: share of illiterates aged 15 years or older			
	(1)	(2)	(3)
CAIP (+1 s.d.)	-0.117	-0.124	-0.096
	(0.009)	(0.011)	(0.007)
	[0.000]	[0.000]	[0.000]
CAIP's specification	{Average of prices in ten years prior to t}	{Median of prices in twenty years prior to t}	{Median of prices in ten years prior to t}
(1,714 observations of 482 AMC)			

Standard errors in parentheses.

P-value in brackets.

Notes: Standard errors (in parentheses) are cluster at the AMC level. Results includes AMC's fixed-effect and binary variables of the years 1920, 1940, and 1960. CAIP in Table 5 considers average of prices in twenty years prior to t for its price component (see equation 3). Here in Table 7 we test for other specifications for price component that includes 'average of prices in 10 years prior to t', 'median of prices in 20 years prior to t', and 'median of prices in 10 years prior to t'. CAIP variable standardized to have a zero mean and standard deviation of one considering values for all AMC and all panel years. All five commodities considered in CAIP calculation. Share of illiterates aged 15 years or older computed over adult population aged 15 years or older. Data from unbalanced panel.

Table 7. Relationship between CAIP and share of illiterates aged 15 years or older, controlling by geographic variables.

(Table 8.A) - 1872, 1920, 1940, and 1960 panel Response variable: share of illiterates aged 15 years or older						
	(1)	(2)	(3)	(4)	(5)	(6)
CAIP (+1 s.d.)	-0.139	-0.141	-0.137	-0.108	-0.114	-0.123
	(0.011)	(0.011)	(0.012)	(0.012)	(0.013)	(0.012)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Interaction between average prices and {geographic variable}	{Distance to Atlantic Ocean}	{Distance to AMC's capital}	{Altitude}	{Latitude}	{Average temperature in Spring}	{Average temperature in Summer}
(1,714 observations of 482 AMC)						

(Table 8.B) - 1872, 1920, 1940, and 1960 panel Response variable: share of illiterates aged 15 years or older						
	(1)	(2)	(3)	(4)	(5)	(6)
CAIP (+1 s.d.)	-0.112	-0.111	-0.126	-0.127	-0.141	-0.141
	(0.013)	(0.013)	(0.012)	(0.011)	(0.012)	(0.012)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Interaction between average prices and {geographic variable}	{Average temperature in Autumn}	{Average temperature in Winter}	{Average rainfall in Spring}	{Average rainfall in Summer}	{Average rainfall in Autumn}	{Average rainfall in Winter}
(1,714 observations of 482 AMC)						

Standard errors in parentheses.

P-value in brackets.

Notes: Standard errors (in parentheses) are cluster at the AMC level. Results includes AMC's fixed-effect and binary variables of the years 1920, 1940, and 1960. Results based on equation 6. CAIP variable standardized to have a zero mean and standard deviation of one considering values for all AMC and all panel years. All five commodities considered in CAIP calculation. Distance variables consider minimum distance in meters between AMC's centroid and variable's location of interest. AMC's capital stands for a capital city of an AMC in 2020. Altitude in meters computed by mean value of altitude. Latitude computed on AMC's centroid. Average temperature in Celsius computed by mean value between 1960 and 1991. Average rainfall in millimeters per month computed by mean value between 1960 and 1991. Share of illiterates aged 15 years or older computed over adult population aged 15 years or older. Data from unbalanced panel.

Table 8. Relationship between CAIP and specific mechanisms.

(Table 8.A) - 1872 and 1960 panel	Share of adult population working in agriculture	Share of adult population working in industry	Share of adult population working in services	Share of immigrants	Ln(Total adult population)
	(1)	(2)	(3)	(4)	(5)
CAIP (+1 s.d.)	-0.098	0.060	0.038	0.003	-0.070
	(0.022)	(0.010)	(0.016)	(0.004)	(0.064)
	[0.000]	[0.000]	[0.018]	[0.422]	[0.276]
(Table 8.B) - 1872 and 1960 panel	Share of whites in adult population	Share of blacks in adult population	Share of mixed race in adult population	Share of young population who attends a school	Ln(Total young population)
	(1)	(2)	(3)	(4)	(5)
CAIP (+1 s.d.)	0.120	-0.037	-0.084	0.094	-0.041
	(0.013)	(0.007)	(0.013)	(0.014)	(0.066)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.533]
(548 observations of 274 AMC)					

Standard errors in parentheses.

P-value in brackets.

Notes: Results includes AMC's fixed-effect and binary variable of the year 1960. CAIP variable standardized to have a zero mean and standard deviation of one considering values for all AMC and all panel years. All five commodities considered in CAIP calculation. Data for 1960 is available only for a sample of Brazilian municipalities, which explain the presence of only 274 AMC in panel data analyzed. Data from balanced panel. Adult population considers population aged 15 years or older. Young population considers population aged 7 to 14 years old. Share of young population who attends a school computed over population aged 7 to 14 years old. Total young population in natural logarithm. Total adult population in natural logarithm. Mixed race stands for '*pardo*' population.

Table 9. Relationship between CAIP and share of illiterates aged 15 years or older, controlling by the interaction term between CAIP and the share of slaves in the population in 1872.

1920, 1940, and 1960 panel Response variable: share of illiterates aged 15 years or older	
	(1)
CAIP (+1 s.d.)	-0.091
	(0.007)
	[0.000]
CAIP (+1 s.d.) x share of population who are enslaved in 1872	-0.002
	(0.024)
	[0.945]
(1,225 observations of 475 AMC)	

Standard errors in parentheses.

P-value in brackets.

Notes: Standard errors (in parentheses) are cluster at the AMC level. Results includes AMC's fixed-effect and binary variables of the years 1940 and 1960. CAIP variable standardized to have a zero mean and standard deviation of one considering values for all AMC and all panel years. All five commodities considered in CAIP calculation. Seven AMC that do not have information on enslaved population at 1872 excluded from panel data. Share of illiterates aged 15 years or older computed over adult population aged 15 years or older. Share of population who are enslaved computed over total population. Data from unbalanced panel.

Table 10. Relationship between CAIP and share of illiterates aged 15 years or older, controlling by the interaction term between CAIP and the binary variable of the year 1872.

1872, 1920, 1940, and 1960 panel Response variable: share of illiterates aged 15 years or older	(1)
CAIP (+1 s.d.)	-0.110
	(0.011)
	[0.000]
CAIP (+1 s.d.) x 1872 binary	-0.302
	(0.039)
	[0.000]
Marginal effect in 1872	-0.412
	(0.039)
	[0.000]
(1,714 observations of 482 AMC)	

Standard errors in parentheses.

P-value in brackets.

Notes: Standard errors (in parentheses) are cluster at the AMC level. Results includes AMC's fixed-effect and binary variables of the years 1920, 1940 and 1960. CAIP variable standardized to have a zero mean and standard deviation of one considering values for all AMC and all panel years. All five commodities considered in CAIP calculation. Share of illiterates aged 15 years or older computed over adult population aged 15 years or older. Data from unbalanced panel.

Table 11. Relationship between CAIP and specific mechanisms, controlling by the interaction term between CAIP and the binary variable of the year 1872.

1872 and 1960 panel	Share of adult population working in agriculture	Share of adult population working in industry	Share of adult population working in services	Share of immigrants	Ln(Total adult population)
	(1)	(2)	(3)	(4)	(5)
CAIP (+1 s.d.)	-0.028	0.028	0.000	-0.010	-0.168
	(0.033)	(0.015)	(0.024)	(0.007)	(0.098)
	[0.400]	[0.066]	[0.996]	[0.144]	[0.088]
CAIP (+1 s.d.) x 1872 binary	-0.333	0.152	0.181	0.062	0.465
	(0.120)	(0.055)	(0.088)	(0.024)	(0.353)
	[0.006]	[0.006]	[0.040]	[0.009]	[0.189]
Marginal effect in 1872	-0.362	0.181	0.181	0.053	0.297
	(0.097)	(0.045)	(0.071)	(0.019)	(0.286)
	[0.000]	[0.000]	[0.011]	[0.006]	[0.298]
(548 observations of 274 AMC)					

Standard errors in parentheses.

P-value in brackets.

Notes: Results includes AMC's fixed-effect and binary variable of the year 1960. CAIP variable standardized to have a zero mean and standard deviation of one considering values for all AMC and all panel years. All five commodities considered in CAIP calculation. Data for 1960 is available only for a sample of Brazilian municipalities, which explain the presence of only 274 AMC in panel data analyzed. Data from balanced panel. Adult population considers population aged 15 years or older. Total adult population in natural logarithm.

Table 12. Relationship between CAIP and specific mechanisms, controlling by the interaction term between CAIP and the binary variable of the year 1872.

1872 and 1960 panel	Share of whites in adult population	Share of blacks in adult population	Share of mixed race in adult population	Share of young population who attends a school	Ln(Total young population)
	(1)	(2)	(3)	(4)	(5)
CAIP (+1 s.d.)	0.145	-0.012	-0.134	0.051	-0.256
	(0.020)	(0.011)	(0.019)	(0.021)	(0.101)
	[0.000]	[0.294]	[0.000]	[0.016]	[0.011]
CAIP (+1 s.d.) x 1872 binary	-0.118	-0.119	0.237	0.201	1.020
	(0.072)	(0.040)	(0.069)	(0.076)	(0.362)
	[0.102]	[0.003]	[0.001]	[0.009]	[0.005]
Marginal effect in 1872	0.027	-0.131	0.104	0.252	0.763
	(0.058)	(0.032)	(0.056)	(0.061)	(0.293)
	[0.641]	[0.000]	[0.062]	[0.000]	[0.009]
(548 observations of 274 AMC)					

Standard errors in parentheses.

P-value in brackets.

Notes: Results includes AMC's fixed-effect and binary variable of the year 1960. CAIP variable standardized to have a zero mean and standard deviation of one considering values for all AMC and all panel years. All five commodities considered in CAIP calculation. Data for 1960 is available only for a sample of Brazilian municipalities, which explain the presence of only 274 AMC in panel data analyzed. Data from balanced panel. Adult population considers population aged 15 years or older. Young population considers population aged 7 to 14 years old. Total young population in natural logarithm. Mixed race stands for '*pardo*' population.

Table 13. Relationship between CAIP's volatility and share of illiterates in the population aged 15 years or older.

1872, 1920, 1940, and 1960 panel Response variable: share of illiterates aged 15 years or older		
	(1)	(2)
CAIP's volatility (+1 s.d.)	-0.011	0.077
	(0.007)	(0.009)
	[0.129]	[0.000]
CAIP's specification	{All commodities}	{Restrict coffee}
(1,714 observations of 482 AMC)		

Standard errors in parentheses.

P-value in brackets.

Notes: Standard errors (in parentheses) are cluster at the AMC level. Results includes AMC's fixed-effect and binary variable of the years 1920, 1940, and 1960. CAIP variable standardized to have a zero mean and standard deviation of one considering values for all AMC and all panel years. Share of illiterates aged 15 years or older computed over adult population aged 15 years or older. CAIP (volatility) computed using the annual price component and considering the standard deviation of the annual CAIP in the twenty years prior to the reference date. Data from unbalanced panel.

Appendix

Appendix 1. Data sources.

Maximum Attainable yield (in tons per hectare): Global Agro-ecological Zones (GAEZ) v3.0 project of the United Nations Food and Agriculture Organization (FAO).

Áreas Minimamente Comparáveis (AMC): Ehrl (2017) on an adaptation of the methodology proposed by Reis et al. (2008).

Observed agricultural production (in tons): Recenseamento de 1920. Diretoria Geral de Estatística, Volume III, Agricultura, Rio de Janeiro, 1920.

Observed local prices (in mil-réis): Anuário Estatístico do Brasil de 1908-1912. Diretoria Geral de Estatística, Rio de Janeiro, 1917.

Commodities real price indexes: Jacks (2019).

Total population:

1872: Recenseamento Geral do Império.

1920-1960: Demographic Census.

Population by ethnic group:

1872: Recenseamento Geral do Império.

1960: Demographic Census.

Slaves: Recenseamento Geral do Império.

Illiterates:

1872: Recenseamento Geral do Império.

1920-1980: Demographic Census.

School frequency:

1872: Recenseamento Geral do Império.

1960: Demographic Census.

Sector of economic activity:

1872: Recenseamento Geral do Império.

1960: Demographic Census.

Distance variables: Instituto Brasileiro de Geografia e Estatística.

Altitude: Instituto Brasileiro de Geografia e Estatística.

Latitude: Instituto Brasileiro de Geografia e Estatística.

Temperature: CRU CL 2.0 10 of Climate Research Unit of University of East Anglia (CRU-UEA).

Rainfall: CRU CL 2.0 10 of Climate Research Unit of University of East Anglia (CRU-UEA).

Appendix 2. Inclusion of the years of 1970 and 1980 in the panel.

We consider the possibility that our results are influenced by the sample of 1960, which consists of only 270 AMC. However, Table A1 indicates that our results are robust even if we include the years of 1970 and 1980 in the panel.

Table A1. Relationship between CAIP and share of illiterates aged 15 years or older.

Response variable: share of illiterates aged 15 years or older		
	1872, 1920, 1940, 1960, and 1970 panel	1872, 1920, 1940, 1960, 1970, and 1980 panel
CAIP (+1 s.d.)	-0.128	-0.125
	(0.006)	(0.006)
	[0.000]	[0.000]
Number of observations	2,196	2,678
Number of AMC	482	482

Standard errors in parentheses.

P-value in brackets.

Notes: Standard errors (in parentheses) are cluster at the AMC level. Result for 1872-1970 panel includes AMC's fixed-effect and binary variables of the years 1920, 1940, 1960, and 1970. Result for 1872-1980 panel includes AMC's fixed-effect and binary variables of the years 1920, 1940, 1960, 1970, 1980. CAIP variable standardized to have a zero mean and standard deviation of one considering values for all AMC and all panel years. All five commodities considered in CAIP's calculation. Share of illiterates aged 15 years or older computed over adult population aged 15 years or older. Data from unbalanced panel.

Appendix 3. Robustness check for differences between Brazilian states and heterogeneous impact due to distance from Atlantic Ocean.

Potential export revenues can have heterogeneous impacts on the share of illiterates when we consider two aspects: differences between Brazilian states across the analyzed period and distance between an AMC and the Atlantic Ocean. There is a possibility that some Brazilian state had specific educational policies, which may bias our results by allowing between state variations to affect our estimates. We perform a robustness test including an interaction term between year binary variables and state indicator variables in equation 4, resulting in equation A1. Table A2 (column 1) indicates that the negative relationship between CAIP and share of illiterates aged 15 years or older become smaller, but the negative relationship and statistical significance at 1% persists.

$$\mathrm{Y}_{ct} = \beta_0 + \beta_1 CAIP_{ct} + \sum_{q=t+1}^{T} \sum_{s=1}^{S} \vartheta_{qs}(\lambda_q * State\_binary_s) + \sum_{q=t+1}^{T} \tau_q \lambda_q + \alpha_c + \mu_{ct} \qquad (A1)$$

There is also the possibility that localities near the Atlantic Ocean shows different patterns of the relationship between potential export revenues and educational investments. These localities were the first one to be colonized and received a great influx of population, which contributes to the urbanization process and consequently the expansion of industry and service sectors. However, these localities may also lack the suitable land for large plantations of agricultural crops due to your coast-mountain geography, which would positively affect the inland's agricultural production and its subsequently influx of revenues. We perform a robustness test by including an interaction term between CAIP and distance to the Atlantic Ocean from AMC, as equation A2 shows. Table A2 (column 2) indicates that CAIP impact over the share of illiterates aged 15 years or older presents almost the same estimated coefficient from Table 5 and the interaction term is statistically not significant.

$$\mathrm{Y}_{ct} = \beta_0 + \beta_1 CAIP_{ct} + \beta_2 CAIP_{ct} * Dist\_ocean_c + \sum_{q=t+1}^{T} \tau_q \lambda_q + \alpha_c + \mu_{ct} \qquad (A2)$$

Table A2. Relationship between CAIP and share of illiterates aged 15 years or older, controlling for Brazilian state year-specific fixed effects and interaction between CAIP and distance from Atlantic Ocean.

1872, 1920, 1940, and 1960 panel Response variable: share of illiterates aged 15 years or older		
	(1)	(2)
CAIP (+1 s.d.)	-0.056 (0.019) [0.003]	-0.141 (0.011) [0.000]
Control variables	{Interaction between Brazilian states and panel years}	{Interaction between CAIP and distance from Atlantic Ocean}
(1,714 observations of 482 AMC)		

Standard errors in parentheses.

P-value in brackets.

Notes: Standard errors (in parentheses) are cluster at the AMC level. Results includes AMC's fixed-effect and binary variables of the years 1920, 1940, and 1960. CAIP variable standardized to have a zero mean and standard deviation of one considering values for all AMC and all panel years. All five commodities considered in CAIP's calculation. Result in column (1) obtained from equation A1. State of Minas Gerais omitted in estimation of column (1). Result in column (2) obtained from equation A2. Share of illiterates aged 15 years or older computed over adult population aged 15 years or older. Data from unbalanced panel.

Appendix 4. Relationship between CAIP and specific mechanisms with sugar's price index at 1872 level.

(Appendix 4.A) - 1872 and 1960 panel	Share of adult population working in agriculture	Share of adult population working in industry	Share of adult population working in services	Share of immigrants	Ln(Total adult population)
	(1)	(2)	(3)	(4)	(5)
CAIP restricting sugar (+1 s.d.)	-0.083	0.041	0.042	-0.006	0.060
	(0.014)	(0.007)	(0.010)	(0.003)	(0.043)
	[0.000]	[0.000]	[0.000]	[0.026]	[0.158]
(Appendix 4.B) - 1872 and 1960 panel	Share of whites in adult population	Share of blacks in adult population	Share of mixed race in adult population	Share of young population who attends a school	Ln(Total young population)
	(1)	(2)	(3)	(4)	(5)
CAIP restricting sugar (+1 s.d.)	0.060	-0.030	-0.029	0.058	0.063
	(0.009)	(0.005)	(0.009)	(0.009)	(0.044)
	[0.000]	[0.000]	[0.001]	[0.000]	[0.156]
(548 observations of 274 AMC)					

Standard errors in parentheses.

P-value in brackets.

Notes: Results includes AMC's fixed-effect and binary variable of the year 1960. CAIP variable standardized to have a zero mean and standard deviation of one considering values for all AMC and all panel years. CAIP version restricting sugar's price index at its level of 1872. Data for 1960 is available only for a sample of Brazilian municipalities, which explain the presence of only 274 AMC in panel data analyzed. Data from balanced panel. Adult population considers population aged 15 years or older. Young population considers population aged 7 to 14 years old. Share of young population who attends a school computed over population aged 7 to 14 years old. Total young population in natural logarithm. Total adult population in natural logarithm. Mixed race stands for '*pardo*' population.

Appendix 5. Relationship between CAIP and specific mechanisms with cotton's price index at 1872 level.

(Appendix 5.A) - 1872 and 1960 panel	Share of adult population working in agriculture	Share of adult population working in industry	Share of adult population working in services	Share of immigrants	Ln(Total adult population)
	(1)	(2)	(3)	(4)	(5)
CAIP restricting cotton (+1 s.d.)	-0.058	0.040	0.017	0.004	-0.101
	(0.020)	(0.009)	(0.015)	(0.004)	(0.058)
	[0.005]	[0.000]	[0.239]	[0.359]	[0.081]
(Appendix 5.B) - 1872 and 1960 panel	Share of whites in adult population	Share of blacks in adult population	Share of mixed race in adult population	Share of young population who attends a school	Ln(Total young population)
	(1)	(2)	(3)	(4)	(5)
CAIP restricting cotton (+1 s.d.)	0.098	-0.021	-0.077	0.065	-0.090
	(0.012)	(0.007)	(0.011)	(0.013)	(0.060)
	[0.000]	[0.002]	[0.000]	[0.000]	[0.132]
(548 observations of 274 AMC)					

Standard errors in parentheses.

P-value in brackets.

Notes: Results includes AMC's fixed-effect and binary variable of the year 1960. CAIP variable standardized to have a zero mean and standard deviation of one considering values for all AMC and all panel years. CAIP version restricting cotton's price index at its level of 1872. Data for 1960 is available only for a sample of Brazilian municipalities, which explain the presence of only 274 AMC in panel data analyzed. Data from balanced panel. Adult population considers population aged 15 years or older. Young population considers population aged 7 to 14 years old. Share of young population who attends a school computed over population aged 7 to 14 years old. Total young population in natural logarithm. Total adult population in natural logarithm. Mixed race stands for '*pardo*' population.