

# Colonial Institutions, Slavery, Inequality, and Development: Evidence from São Paulo, Brazil (Draft)

William R. Summerhill<sup>1</sup>

[Version 2012]

**JEL Classification:** D63, H7, H27, N16, N36, O12, O13, O18, O54

**JEL Keywords:** Inequality, economic development, wealth, agriculture, regional analysis, growth, slavery, underdevelopment, municipalities, Latin America, public finance

**ABSTRACT:** Brazil is frequently portrayed as exhibiting persistent and structural economic inequality that is rooted in the early colonial experience, and is believed to undermine development in the long run. I construct original measures of agricultural inequality for 1905 in what is today Brazil's largest state, using farm-level micro data for some 50,000 farms. Using these measures of inequality, along with contemporary covariates and other historical variables I assess the impact of colonial institutions, slavery, farm inequality, and political inequality on long-term development in São Paulo. The principal findings are: (1) a potentially coercive colonial institution, the *aldeamento*, is positively correlated with income per capita at the end of the twentieth century; (2) measures of the intensity of slavery have little if any independent impact on income in 2000; (3) farm inequality was not persistent in São Paulo at the county level over the twentieth century; (4) in both OLS and IV estimates, no negative effect can be found for 1905 inequality on long-term development; (5) political inequality in the early twentieth century, measured by the extent of the franchise, is unrelated to contemporary farm inequality, and also unrelated to long-term economic growth; and (6) the provision of local public goods in the early twentieth century, measured by local public education outlays, has a positive impact on long-term development, but was not related to contemporary economic or political inequality. Overall, neither the intensity of slavery nor the pattern of inequality had any discernable negative economic impact in the long run.

---

<sup>1</sup> Department of History, UCLA [summerhill@ucla.edu]. Research for this paper was supported by NSF Award SES-0418350 and the UCLA Center for Economic History. I am particularly grateful to Bianca Giberti, Flávio Santos, Carrie Arboleda, Elena Moroz, and Svetlana Pimkina for their tireless efforts at research and data entry. Molly Ball and Joseph Ryan assisted with historical data, while Leonardo Monastério, Rodrigo Soares, and Eustáquio Reis generously helped with year 2000 data. Much of the work was done while a visiting scholar at the Escola de Pós-Graduação em Economia of the Fundação Getúlio Vargas (EPGE-FGV), and as a visiting professor at École des Hautes Études en Sciences Sociales (EHESS), and I am indebted to both institutions for their generous hospitality. I benefitted from comments received during presentations at the UCLA Von Gremp Workshop at UCLA, and the conference on new frontiers in Latin American Economic History at Columbia University, as well as discussions with Leah Platt Boustan, Renato Perim Colistete, Mark Dincecco, Stanley Engerman, Francisco Gallego, Stephen H. Haber, Herbert S. Klein, and Jean-Laurent Rosenthal. All errors are mine.

Work on growth and development over the last twenty years has increasingly focused on two historical features of economies: institutions, and inequality. While cross-country evidence on institutional characteristics and inequality for recent decades is abundant, the supply of authentically historical evidence beyond anecdotes is far more scarce. This paper examines the impact of colonial institutions, slavery, and inequality on long-term developmental outcomes at the county level for the state of São Paulo, Brazil. The argument of the paper runs in the following terms. São Paulo had colonial deep colonial origins, and was the third largest slave-holding province in Brazil by the late nineteenth century. Moreover, by the start of the twentieth century it exhibited the high levels of farm inequality and political exclusion that one would expect, given its legacy of colonialism, slave-based plantation agriculture, and oligarchic politics. Yet one colonial institution in particular, the *aldeamento*--which presented a mix of extractive (negative) and settler (positive) characteristics-- is positively correlated with income per capita at the end of the twentieth century. Slavery had tremendously negative implications for those who were enslaved, but the consequences of slavery for long-term development are not so clear cut. Measures of the intensity of slavery have little if any independent impact on county-level income in 2000. Farm inequality was high at the start of the twentieth century, yet the pattern of farm inequality was not persistent at the county level over the next 90 years. Moreover, in both OLS and IV estimates it is difficult to locate a negative effect for early twentieth-century inequality on long-term development. Political inequality in the early twentieth century, measured by the extent of the franchise, is also unrelated to long-term economic growth. Local public goods provision in the early twentieth century, measured by local public education outlays, is positively related to long-term development, but not correlated with contemporary inequality. Overall, neither the intensity of slavery nor the pattern of farm inequality in São Paulo had any discernable negative impact in the long run.

The 1990s saw an outpouring of studies on the relationship between inequality and economic growth (key overviews of the this work include Aghion, Caroli, and Garcia-Penalosa, 1999; Benabou, 1996). The upending of the old view that inequality was a necessary condition of successful modern economic growth (Kuznets, 1955; Lewis, 1954) has spawned a vast literature on inequality and development. Theoretical and empirical studies alike have shown how inequality can damage the prospects for growth, by facilitating demands for efficiency--undermining redistribution, reducing the provision of public goods, suppressing the formation of human capital, or fostering political institutions and policies that support rent-seeking by a narrow elite. Several studies showed how inequality led to redistributive policies that worked against growth (Alesina and Rodrik, 1994; Persson and Tabellini, 1994). Inequality in the presence of imperfect credit markets can lower growth by reducing human capital formation (Galor and Zeira, 1993; Perotti, 1996; Galor, Moav, and Vollrath; 2009). Rich elites may repress democracy to secure their rents and privileges (see for example, Sokoloff and Engerman, 2000; Bourguignon and Verdier, 2000; Acemoglu, 2003).

Whether the purported relationship between inequality and underdevelopment holds empirically has been another area of research. Initial results linking inequality to poor economic growth (Alesina and Rodrik, 1994; Persson and Tabellini, 1994; Clarke, 1995; Alesina and Perotti, 1996) have been called into question by the findings of others (Forbes, 2000; Barro, 2000; Banerjee and Duflo, 2003). In particular, Forbes (2000) discovered a positive relationship between inequality and growth, much in the same vein as the early development literature. Yet

more recently, Easterly's investigation (2007) found a strong and credibly causal negative relationship between inequality and growth, reaffirming on a different basis the findings of the early 1990s.

The debate over the role of inequality intersects the larger body of work on the determinants of growth. Both literatures have taken a decidedly historical turn (though historical work on growth has longstanding precedents). One prominent strand views growth as historically constrained by tropical latitudes and geographical conditions (Sachs and Warner, 1997; Bloom and Sachs, 1998). Historical institutions in the form of legal origins have been used to explain both broad patterns of growth, and more specifically the development of finance (LaPorta, et al., 1997; LaPorta, et al, 1998; Levine, 1999; Levine, 2005; Levine, Loayza, and Beck, 2000). Acemoglu, Johnson, and Robinson (2001, 2002, 2005) focused on an institutional mechanisms rooted in the early colonial era in explaining long-term differences in development. They distinguish "settler" institutions from "extractive" institutions, showing that where the former prevailed the trajectory of development was higher. Focusing specifically on the Americas, Engerman and Sokoloff (1997) proposed that differing factor endowments in the 1500s led to different growth paths over five centuries, and that these explained current patterns of underdevelopment in the Americas. Easterly and Levine focused on more recent history, but also found a key role for endowments (Easterly and Levine, 2003). In contrast to the original version of the Engerman and Sokoloff view (1997), North, Summerhill, and Weingast (2000) noted that over an interval of several centuries, factor endowments were effectively endogenous, and argued instead for an independent role for political institutions in the late colonial era and the nineteenth century in determining long-term economic outcomes. In response Engerman, Sokoloff and co-authors refocused attention on institutions (namely slavery, the franchise, and education) that impacted both inequality and growth, while adhering to the argument that early factor endowments were the key determinant of long-term outcomes (Sokoloff and Engerman, 2000; Engerman, Mariscal, and Sokoloff, 2002; Engerman and Sokoloff, 2005; Engerman and Sokoloff, 2008). Institutions, whether rooted in factor endowments or not, are not universally viewed as relevant (Glaeser, et al., 2004). More recently Coatsworth (2008) and Williamson (2009) have made arguments about inequality that, like the North, Summerhill, and Weingast view of long-term development, reject an early "colonial origins" explanation for patterns of inequality in Latin America.

Importantly, a growing body of work has moved beyond international comparisons to look at variations within countries, especially with respect to institutions (Banerjee and Iyer, 2005; Banerjee, Iyer and Somanathan, 2005; Bonet and Meisel, 2006; Bruhn and Gallego, 2008). Huillery, 2009; Naritomi, Soares, and Assunção, 2009). Studies that look at within-country variation of inequality arrive at a variety of findings. Ramchuran (2009) found that farm inequality in the late nineteenth-century U.S. was an important determinant of redistributive transfers, including expenditures on education, as did Galor, Moav, and Vollrath (2009), and Acemoglu, et al. (2008). Nunn (2008) found that the share of the population comprised by slaves in the United State in 1860 was correlated with income in 2000, but that land inequality in 1860 was not. Acemoglu, et al. (2008) determined that there was little role for nineteenth-century land inequality in Colombia in explaining long-term economic outcomes, but that political inequality was a distinct cause of lower levels of development. Most recently, Wegenast (2010), compares educational outcomes across Brazilian states, arguing for a central

role for endowments, crop choices and agricultural inequality in explaining lower levels of education.

## **The Setting and Hypotheses**

São Paulo is the largest and richest state in Brazil. If it were independent it would have nearly the third largest population of any country in Latin America, behind only the rest of Brazil, Mexico, and almost as large as Colombia. Most of the state's territory is comfortably inside the tropics, with a smaller portion in the "temperate" region south of the tropic of Capricorn. Like the rest of Brazil, São Paulo was a colonial possession of Portugal from the start of the sixteenth century. The territory of what was then a donatary captaincy was rich in land, but (from the perspective of the Portuguese) relatively poor in labor, and more importantly for the colonization project, it was poor in natural resources compared to Brazil's northeast. The early colonizers in São Paulo imposed slavery, first on the indigenous population, then using Africans (Monteiro, 1994). Research in the "new historical development" literature has heavily emphasized the importance of the colonial origins of development (Sokoloff and Engerman, 2000; Acemoglu, Johnson, and Robinson, 2001, 2002; Nunn, 2008). Acemoglu, Johnson, and Robinson (2001, 2002) find that early European mortality risk influenced whether the colony established the settler institutions that promoted favorable long-term outcomes, or instead ended up with extractive institutions. For Sokoloff and Engerman (2000) the factor endowments encountered by the Europeans determined what types of institutions were created in the colonial era, along with long-term patterns of inequality and development. On the other side of the Atlantic, Nunn found that the African regions that were most intensively involved early on in the export of slaves suffered worse developmental outcomes hundreds of years after the fact (Nunn, 2008). In the spirit of this literature, and employing a rich data set for Brazil, Naritomi, Soares, and Assunção (2009) find that extractive resource "booms" (sugar and gold) in the colonial era created de facto institutions that undermined development centuries later. Across these studies the sources of colonial institutions differed, but all are seen as having early origins, and as accounting for big differences in development.

### *Colonial Institutions in São Paulo: the Aldeamento*

Although São Paulo was one of the few relatively successful colonial captaincies, it lay well beyond the regions most affected by the first big sugar cycle in the sixteenth century (see Naritomi, Soares, and Assunção, 2009, Figures 5 and 6). It was largely removed from the orbit of the late seventeenth-century gold boom as well. With the mining zones as distant as a two-months' journey for horses and human porters (Monteiro, 1994), the gold boom from 1680 to 1750 was unlikely to have had a long-term institutional impact in São Paulo. São Paulo had always produced sugar, but its own late eighteenth-century sugar "boom" came more than a century and a half after the original sugar cycle involving Brazil's northeast (Klein and Luna, 2003). Only with the expansion of coffee cultivation on a large scale, exploiting fertile land and transport improvements, in the late nineteenth century did São Paulo undergo a resource-based export boom.

If it was largely removed from the colonial extractive booms, São Paulo's colonial origins may have mattered in a different way. A subset of São Paulo's earliest settlements, dating from the mid-sixteenth century, were established as *aldeamentos*. *Aldeamentos* were settlements that the Portuguese created to fix the semi-nomadic indigenous population in place, convert them to Christianity, provide with ongoing religious instruction, and defend against less-cooperative raiding Indians (Petroni, 1995; Perrone-Moises, 1992). *Aldeamentos* were sanctioned by the crown, administered by the Jesuits, and populated with Indians. Characterizing actual colonial institutions (as opposed to de facto institutions inferred ex post from regression results) is not as simple as much of the factor endowments-colonial origins literature suggests, and *aldeamentos* are no exception.<sup>2</sup> What is frequently obscured in the colonial factor endowments view is that the pattern and intensity of colonization and settlement followed not only the profit-seeking motives of the colonizers, but also the fiscal goals of the crown, and a religious and spiritual logic emanating from Rome and running through Spain and Portugal. In creating permanent sites of settlement and religious instruction for a heterogeneous mix of nomadic and semi-nomadic indigenous peoples, the church and crown created an institution with myriad attributes. While designed to protect indigenous peoples from slavery, *aldeamentos* reduced their mobility, concentrating them in a few sites, and may have made them more readily exploitable by Europeans as a labor force. In this sense *aldeamentos* were functionally part of an overall "extractive" effort. Yet, much like settler institutions (in terms suggested by Acemoglu, et al.), *aldeamentos* were established with a goal of permanence by their organizers, rather than short-term gains. Importantly, *aldeamentos* were not enclaves. European and ethnically-mixed settlement sprung up around *aldeamentos* because of the economic activities undertaken under the watchful eyes of the priests.

One cluster of *aldeamentos* was located on the highland plateau, in and around what today is the megapolis of São Paulo. This clustering might hint at geographic and climatic influences on the location of *aldeamentos*, yet the settlements were also found on the near-tropical coast, and well inland, close to the Paraíba valley. There is no information on the radius of influence of an *aldeamento*. In one respect it could be large, in that indigenous peoples were brought to some *aldeamentos* from considerable distances. In other respects the area of influence might be limited to the actual boundaries of the *aldeamento*, and the boundaries of direct Jesuit control. The more expansive measure of influence could encompass the much larger districts of which each *aldeamento* was a part at the peak of its population in the mid seventeenth century.<sup>3</sup> A far

---

<sup>2</sup> The Engerman and Sokoloff thesis, as it pertained to Latin America, is very close to Lockhart and Schwartz's pioneering interpretation of the pattern of colonization. Though not shrouded in an economic apparatus or jargon, the Lockhart and Schwartz view strongly emphasized the characteristics of indigenous social organization, as well the availability of readily exploitable natural resources, in determining the paths of colonization and colonial development. Lockhart and Schwartz's interpretation, once recast in economic language, is by and large the factor endowments story.

<sup>3</sup> The indigenous populations within the *aldeamentos* declined after 1640, when the Jesuits were forcibly expelled by colonists from São Paulo because of their stance against indigenous slavery. The decline was exacerbated by the gold boom in Minas Gerais, which attracted laborers among

more restrictive measure would be the boundaries of the year 2000 county in which the *aldeamento* was located. Here a somewhat conservative middle approach is taken: each *aldeamento*'s area of influence is taken to be the county as it was defined in 1872 that had been created on the basis of the *aldeamento*. The hypothesis to be tested is that, if *aldeamentos* indeed had predominantly "extractive" characteristics, they should be negatively associated with long-term economic outcomes.

### *Colonial and Modern Institutions: African Slavery*

Generations of historians have long believed that slavery and the legacy it left were inimical to development. Sokoloff and Engerman (2000), and more recently Nunn (2008) present evidence that slavery had a long-term negative impact on economic growth in the Americas. Brazil was the single largest importer of African slaves in the western hemisphere over a period of nearly three and a half centuries. Around 1580 the Portuguese finally grasped that the indigenous population of Brazil was never going to provide the relatively abundant source of servile labor that the Spanish had encountered in meso-America and the Andes. Few of the slaves imported to Brazil in the colonial era went to São Paulo, going instead to export-boom regions with high labor demand during the sugar, gold, and early coffee cycles. Several centuries of efforts to enslave the indigenous population in São Paulo gave way by the early 1800s to a system based on the enslavement of Africans and their descendents. African slavery always existed in São Paulo under the Portuguese, but was far more prominent in the decades before abolition in 1888 than at any point before. The demand for slaves in São Paulo increased sharply in the second half of the nineteenth century. The construction of railroads in São Paulo beginning in the late 1860s opened the central and western parts of the state to commercial agriculture. The spread of coffee cultivation beyond the Paraíba valley greatly increased the demand for farm labor. Because the slave trade from Africa had ended in 1850s, this demand for labor was initially met by imports of slaves from elsewhere in Brazil. By 1872 ten percent of Brazil's slave population was concentrated in São Paulo, and only three provinces had more slaves. Between 1872 and the abolition of slavery in 1888 São Paulo had the highest rate of net internal importation of slaves of any province in Brazil (Slenes, 1976). When the available supply of slaves proved insufficient, coffee growers in São Paulo began to draw large numbers of European immigrants (Holloway 1980; Dean, 1976). The Brazilian government began to gradually abolish slavery in 1871, and in 1888 slavery ended. The 1872 census reported the numbers of slaves in each county (Brazil, 1876). This was likely near the peak of slavery in São Paulo. Testing the hypothesis that slavery was inimical to development in São Paulo draws on county-level measures of the intensity of slavery 1872.

### *Agricultural Inequality*

The inequality component of the Engerman-Sokoloff thesis resonated especially strongly with other investigators, since it was applicable beyond the peculiarities of slave societies. Here the recent historical turn in the literature has been a sharp one. The findings, however, are mixed. For the late 20<sup>th</sup> century Easterly (2007) found a strong and negative relationship between

---

the indigenous population (Petroni, 1995; Monteiro, 1994). The effect of the creation of the *aldeamentos* was permanent, however, as the communities that built up around them persisted.

income inequality and income around the world. Frankema (2009) argues for a mix of institutional and factor endowments in accounting for patterns of inequality across countries. Ramcharan (2009); Galor, Moav, and Vollrath (2009); and Acemoglu, et al. (2008) all found a negative relationship between nineteenth-century land inequality and human capital formation (or educational expenditures) within the United States. Nunn (2008), by way of contrast, found no significant relationship between early land inequality and long-term economic performance in the US at either the state or county level. Acemoglu, et al. (2008) similarly did not find a negative relationship between nineteenth-century land inequality, and either income or education, in Cundinamarca province in Colombia. If inequality in São Paulo undermined development, it should be apparent in its effects on later income and related development indicators. This is the hypothesis tested below using an original set of inequality measures based on farm-level data for the fifty thousand farms in 1905.

## Data Overview

Testing the hypotheses that colonial institutions, slavery, and inequality negatively impacted long-term development requires data on several key variables, and a variety of potential controls. Since farm inequality can be calculated at the county level in 1905, I focus on county-level political and economic indicators. Assessments at the county level for São Paulo are complicated by the fact that in most cases the year 2000 counties do not correspond exactly with counties from 1905 and earlier periods. The extent of this problem is revealed by Figure 1.A and Figure 1.B, which show the relentless division of counties over the twentieth century. To render year 2000 municipalities comparable with those of 1905 required a painstaking recombination of nearly 650 municípios in year 2000 into 170 municípios in 1905.<sup>4</sup> The matching of newer counties to their 1905 predecessors relied on a variety of sources detailing the administrative subdivision of counties over time. Of the 171 counties in 1905, only one was permanently absorbed by another at a later point. Maintaining comparability between the 1905 counties and year 2000 counties required that Santo Amaro and the capital of São Paulo be treated as a single composite 1905 county. Appendix II provides detail on the matching process, along with checks for consistency.

The 1905 inequality measures are based on farm-level census results encompassing nearly 50,000 farms in 1904-1905. Census results from 1904-1905 have not previously exploited at the micro level for the state as a whole. While most of the census data is straightforward, the data for a several counties presented considerable challenges to measuring inequality. The construction of the gini coefficients is detailed in the following section. Additional detail on manipulations and adjustments to the farm-level data is provided in Appendix III.

---

<sup>4</sup> Eustáquio Reis has pioneered in reconstructing minimal comparable areas back to 1872 for Brazil as a whole, thereby making it possible for investigators to use historical variables at levels below that of the state, despite the widespread fragmentation of counties over the last century and a half (Reis, Pimentel, and Alvarenga, 2010). The approach here to developing comparable territorial units (CTUs) is similar, but differs in that it attempts a direct and complete matching of 2000 and 1905 counties, and is limited to the state of São Paulo.

Information on the *aldeamentos* and their locations comes from Petrone (1995). Other variables and controls from the first decade of the twentieth century, and before, come mainly from contemporary official publications. Variables for the late twentieth century come from standard official sources. Definitions and sources for each variable are detailed in Appendix I.

## Agricultural Inequality in 1905

One of the principal obstacles to studying the impact of past inequality has been the lack of comprehensive and reliable measures. The farm census of 1905 provides farm-level details on land, wealth, owners, and cultivation, among other fields. The results of this census have long been consulted by historians (see Camargo, 1981; Milliet, 1939; and Holloway, 1980, among others). Previous work has not used it to create measures of farm inequality. Of 171 counties that made up the state in 1905, the census reports farm-level details on more than 50,000 farms in 165 counties. For the other six counties no detailed farm-level data have yet been located. Farm-level information was collected for these counties, since the distribution of farms by size category was reported in the annual report of the state secretary of agriculture, along with farm values, land-use data, and information on proprietors.<sup>5</sup> The distribution of farms across size categories for these six counties provides a crude measure of inequality that is then used to generate estimates of the other measures.

To indicate the degree of farm inequality I construct Gini coefficients for each county in 1905. While a variety of inequality measures are available, Ginis enjoy ease of interpretation, and are directly comparable with late-twentieth century measures of farm inequality. The Gini's are constructed using the standard formula:

$$G = \frac{2}{n^2 \bar{a}} \sum_{i=1}^n i(a_i - \bar{a})$$

Where the  $a$ 's are either the farm size or its value, the  $a$ 's are arranged in ascending order,  $i$  indexes the farm in municipality  $n$ ,  $a$  is the total area (or value) of farmland owned by proprietor

$i$ , and  $\bar{a}_m = \frac{1}{n_m} \sum_{i=1}^n a_i$  is the average farm size in municipality  $n$ .

The five gini coefficients are the *crude farm gini*, *farm land gini*, the *overall farm land gini*, the *farm wealth gini*, and the *overall farm wealth gini*. For these six counties that did not report at the farm level, an approximation to the gini coefficient was constructed by using the midpoints of the each size category as the mean farm size. This is the *crude farm gini coefficient*.

The farm land gini and farm wealth gini consider the distribution of farm land and farm wealth over proprietors. Figure 3 shows the relationship between the two definitions of inequality in practice. Differences between these two measures have two obvious sources. The first is the unit value of land, which varied considerably across farms within counties, and across counties.

---

<sup>5</sup> These were the 1905 counties of Apiahy, Iquape, Mogy das Cruzes, Santa Izabel, Tieté, and Ytú.

The second is the value of non-land farm assets (implements, improvements, structures, and the like). This is especially salient in the case of coffee growing. Coffee cultivation was an investment problem; plantings required a three to five year wait to bear fruit, and then under normal circumstances yielded a crop for decades. Farms with productive coffee groves would be worth substantially more than similar farms planted in annual crops.

The farm land gini and farm wealth gini are the most common measures of agricultural inequality. But they address distribution only across owners, and do not take into account those without property. They will mask high inequality if many people do not own land. In a hypothetical county of 10 equal-sized farms distributed across 10 families, and 90 more families who do not own, the land and value gini coefficient will indicate perfect equality *among owners*, when 90 percent of the population is in fact landless. The *overall* measures are a step toward remedying this problem, taking into account that farms were distributed across not just farm owners, but the entire population (here taken as potential farm-owning *families*, rather than individuals). The number of “landless” families in each county in 1905 is not known. Under the assumption that the average family size of immigrants was roughly typical of the rural population, the number of landless families can be calculated from the county population and the number of farm owners.<sup>6</sup>

For the six counties presenting data only by size categories, imputing the farm land gini coefficients, and the farm wealth gini coefficient, was achieved by regressing each of these variables separately on the crude gini coefficient.<sup>7</sup> The crude farm gini alone was a less effective predictor of the overall ginis. Since the principal distinguishing feature of the overall ginis is their incorporation of the landless population, the crude farm gini and the estimated share of landless families in the county were used together as predictors.<sup>8</sup>

Table 1.A presents summary statistics for the five gini coefficients. As expected, the overall ginis are considerably higher than the ginis considering only farm proprietors, given the large number of people who were not owners. The inequality measures nonetheless show considerable dispersion, with some counties being more than fifty percent less equal than others in the distribution of wealth. Correlations between these measures are presented in Table 1.B. The adjustments made to the farm land gini and farm value ginis to construct the overall measures of inequality are not simply positive monotonic transformations. The share of the population that was landless varied sufficiently across counties in 1905 that the correlations between different ginis declines considerably between the crude farm gini, and the overall farm wealth gini.

The first result on long-term inequality appears in Table 1.C. Farm area gini coefficients are available for almost all counties in 1995. I combine these 1995 county-level ginis to create farm area ginis for the CTUs in 1995, weighted by the share of farmland accounted for by each of the 1995 counties within the CTU. The result is a set of comparable farm land gini coefficients for

---

<sup>6</sup> Holloway’s (1980) detailed work on immigrants shows that the average immigrant family size in the first two decades of the twentieth century was 4.7 persons. I employ a family size of five.

<sup>7</sup> The overall fit of these simple regressions was excellent; details available on request.

<sup>8</sup> Adding the share of landless families improved the fit dramatically over the simpler specification using the crude farm gini coefficient alone.

168 of the 170 CTUs. A comparison of the farm area ginis from 1995 and 1905 reveal that the average level of farm land inequality in 1995 is strikingly similar to its level in 1905, with gini coefficients of 0.688 and 0.677 respectively. The fact that the degree of farm land inequality remains largely unchanged for nearly a century in São Paulo might be taken as support for the prevailing view that inequality in Brazil is both structural and persistent (Easterly, 2007). The average levels of inequality in 1905 and 1995 obscure a key detail. Agricultural inequality in São Paulo is high. But the state-wide average masks both considerable diversity, and change over time. The correlation between inequality across counties in 1905 and 1995 is strikingly low. Less than ten percent of the county variation in inequality in 1995 is associated with the county variation in inequality in 1905. The low correlation shows that farm land inequality in São Paulo was not persistent at the county level. Such a shift in the spatial distribution of inequality across counties strongly suggests that farm inequality cannot be “structural,” in that it was predetermined by political and economic institutions established at some previous point in time. As the results below will demonstrate, agricultural inequality may in fact be irrelevant in explaining many outcomes of interest, whether in 1905 or in 2000.

## Colonialism, Slavery, Inequality and Development: Reduced-Form Estimates

The outcome of interest is income per capita in year 2000 in each comparable territorial unit corresponding to a 1905 county.<sup>9</sup> The use of comparable territorial units permits ready exploitation of the 1905 inequality measures, as well as over contemporary covariates. Several other factors are examined as well, including indicators of colonial heritage, and slavery. I use OLS to estimate reduced-form equations of the following type:

$$\ln y_i = \alpha + \beta I_i + \gamma C_i + \varepsilon_i$$

where  $\ln y$  is the natural logarithm of per capita income in 2000 for the CTU in 2000 that corresponds to the 1905 county;  $I$  is the variable of interest, and  $C$  is a vector of controls. Table 2A provides summary statistics of the variables used below, and Table 2B reports the correlations between key variables. Table 3 presents regression results that reveal the extent of the correlation between per capita income and variables for colonial origins and slavery. The simplest representation is found in the first specification. It looks at two factors. The first is a dichotomous variable that measures whether the first population center to be formally recognized in each county was founded before independence (1822). This is a direct and blunt measure of whether local institutions were tautologically “colonial.” The second factor is whether the municipality was created specifically from a colonial *aldeamento*. Both parameter estimates are positive, but only colonial *aldeamento* is significant (this specification, and all that follow, use a robust estimator of the variance-covariance matrix to accommodate heteroskedasticity in the errors). Since all of the *aldeamentos* also date from the colonial period, this simple specification

---

<sup>9</sup> The mean of the level of year 2000 per capita income is close to the median, yet considerably skewed to the right, and exhibits thick tails. Taking the natural logarithm eliminated the skewness and excessive kurtosis.

necessarily suffers from multicollinearity. The second specification adds the distance from the coast as a control, and may capture several countervailing effects. If distance from port proxies the high cost of overland transport in the colonial era, and such isolation negatively impacted development, the coefficient should be negative. Alternatively, if population centers founded earlier in the colonial era were also systematically closer to the coast, where transport costs were lower, founding year and distance to coast could be picking up a similar effect. As a further alternative, if up-country lands were more fertile on average, and higher fertility had a durable impact, distance from the coast would be positively correlated with income per capita in 2000. In this specification the estimate on distance is positive and statistically significant, as is colonial founding.

The third specification adds a variable for the intensity of slavery in 1872. Each 1905 county is assigned the ratio of slaves-to-total population for the 1872 county of which it was part. In effect this measures the impact of an aggregated variable on a more “micro” variable (there were more counties in 1905 than in 1872), and requires cluster robust standard errors to guard against overstating the statistical significance of the coefficients (Moulton 1990). The effect of slavery on income per capita in 2000 is positive and statistically significant. A larger share of the population enslaved in 1872 is associated with higher levels of development in 2000. Finally, São Paulo experienced considerable variation in the timing and type of settlement, which may have influenced long-term outcomes in ways not captured by the dichotomous variable for colonial founding. Specification 4 takes this into account by dropping the dummy variable on “colonial” and adding instead a continuous variable for the founding year of the first settlement in the county. In this specification the signs of the coefficients on the other variables remain unchanged, and all estimates are significant. The result is negative, with an earlier founding associated with higher levels of income in 2000. This result, too, is surprising, given that most specialists associate the emergence of modern economic growth in São Paulo with areas of more recent settlement. Based on this specification, Figure 4 illustrates the partial correlation between colonial *aldeamento* and income per capita in 2000.

Causality cannot be established from regressions in which several variables may be endogenous. The locations of *aldeamentos*, even those based on existing indigenous *aldeias*, could have been heavily influenced by other conditions. Local climate, geography, and forest resources are all factors that could have mattered both for the location of *aldeamentos*, and for income today. The timing of the founding of settlements (*aldeamentos* or not) could be similarly affected by other preconditions. Even distance to the coast, which is indisputably exogenous, may capture several distinct effects that run in opposing directions. The OLS estimates are nonetheless striking: colonial founding and (and, alternately an earlier founding year) is associated with *higher* levels of year 2000 income. If the county’s origins were in an *aldeamento*—the principal distinguishing institutional feature among colonial settlements in São Paulo—its year 2000 level of income per capita was higher still. The intensity of slavery during the nineteenth-century coffee boom also gives the appearance of a strong and positive effect on income in 2000.

What role did inequality play in long-term outcomes, once colonial conditions and slavery are taken into account? The regressions in Table 4 incorporate the measures of farm inequality and overall inequality constructed in the preceding section. The results are again striking. In no specification does 1905 inequality have a negative impact on income in year 2000. In the first

three specifications, using the crude gini, the farm land gini, and the farm wealth gini, the coefficients are all positive but with standard errors that are large, making them statistically indistinguishable from zero. The specifications using the overall gini measures, however, result in a significant coefficient on inequality, but one that is positive, rather than negative. Figure 5 shows the relationship between overall farm wealth inequality and income per capita in 2000, while Figure shows the same for overall farm wealth. Using OLS, the effect of inequality is the opposite of the predicted one: higher inequality in 1905 is associated with better development outcomes in 2000.

To help ensure that 1905 inequality is not merely capturing other contemporary factors that are positively correlated with income per capita in 2000, the specifications in Table 5 adds controls to a set of specifications using the overall farm wealth gini, since it is the most encompassing of the inequality measures. Specification 2 adds a measure of population density in 1907. Density has been found to be a remarkably reliable indicator of relative incomes, and captures several key economic characteristics of place (Acemoglu, et al., 2002). The effect of density in 1907 on income in 2000 is strongly positive—counties that were already prosperous in the early twentieth century remained relatively so for the rest of the century. Its other effect is to reduce the parameter estimate on slavery. The third specification adds controls for transport conditions, all of which are related to greater exposure to distant markets, and—in agricultural economies—higher rents to land and higher farm incomes. The coefficients on all three controls—Port, River Navigation, and Railroad—are positive, but River Navigation is not significant. Their inclusion further reduces the parameter on the slavery variable, and strips it of its last remaining statistical significance. The railroad effect in particular is strong in the OLS specifications, a result consistent with finding in other frameworks (Summerhill, 2003; 2005). The final specification removes the density variable and adds an economically more specific measure of the average productivity of land (gross farm product divided by the land in farms), which takes on the expected sign and is significant.

The loss of statistical significance on the slavery variable in specifications that include population density and transportation facilities warrants elaboration. Slavery in Brazil was a property-rights institution dating from the colonial era. But instead of proxying for *de facto* institutions, the intensity of slavery within São Paulo in 1872 was an indicator of local production possibilities—i.e., endowments. Slavery's *intrinsic* long-term consequences are poorly captured by the distribution of the slave population in São Paulo in 1872 (or anywhere else in Brazil, for that matter). Table 7 presents estimates of the relationship between income in 2000 and the intensity of slavery at a couple of points in history for all of Brazil. The 1872 slavery variable is positive and significant by itself, and when late colonial population density and the intensity of slavery at the end of the colonial era are taken into account. But the lion's share of the variance is explained by regional inequality. Figure 7 displays the added-variable plots. Once region is controlled for, late colonial population density (an indicator of the economy) takes on the expected sign, as does the intensity of slavery in 1819. The intensity of slavery in 1819 has a negative impact on income per capita nearly two hundred years later.<sup>10</sup> The intensity of slavery in 1872 has no statistically significant impact.

---

<sup>10</sup> A cautionary note is in order: the 1872 figures, and even the late eighteenth-century population data, come from primary sources that are familiar to historians. The 1819 slavery

The economic changes of the nineteenth century constituted a reversal of fortune across Brazil's regions. The increasing integration of domestic markets, shifting demand abroad for goods, altered regional (and local) prospects for profit. Mobile factors of production, including (or especially) slave labor, followed those prospects. From the perspective of owners, slaves were nothing more than capitalized streams of labor services. Slave owners moved slaves to the regions where the return on the investment in the slave was highest. Those were the regions with the best productive possibilities based on growing conditions and distance to market. Slave owners and slaves agglomerated in areas with superior agronomic endowments, which were also frequently along the planned or likely routes of railroads. Those same regions of São Paulo also had agglomerations of free persons in the early 20<sup>th</sup> century, and the later population density was higher.

In summary, São Paulo was wholly unexceptional within Brazil in terms of the relationship between slavery in 1872, and economic outcomes in 2000. Negative consequences of the local intensity of slavery are hard to identify. The intensity of slavery itself was an effect, an outcome of the ensemble of local and regional conditions conducive to future prosperity.

## Robustness

To assess robustness I first inspect the sensitivity of the results to the exclusion of counties with unusual characteristics. I then conduct specification tests with respect to the functional form, and check for outliers. Finally I consider another approach to estimation.

One source of potential concern is the three extraordinarily large frontier counties in the west in 1905. These alone divided into hundreds of counties in year 2000. There is considerable variation among the present-day counties that descended from these three, which is impossible to assess with historical variables, given that the lowest level at which we have most early twentieth-century variables is the year 2000 CTU matched to the "old" 1905 county. None of these suggested any problems in tests of outliers and leverage. Nonetheless, to assess the impact of including these large western frontier counties on the estimates, every specification reported here was run again, excluding these three cases. The parameter estimates and test statistics proved remarkably insensitive to the exclusion of the three big western 1905 counties. In general, their exclusion improved the overall fit very slightly, with nearly imperceptible changes in the parameter estimates.

Moreover, the specification of the model underpinning the OLS results proves adequate, in light of the results from Ramsey's omitted-variable specification error test. Additionally, to gauge whether some observations exercised undue influence on the estimates, I computed outlier statistics and measures of leverage. Using the DFITS statistic revealed 15 outlier cases overall (less than 10 percent of the cases here), with the number of negative outliers exceeding positive outliers 2 to 1. However, the value of the two largest positive outliers (the counties of Lorena,

---

figure, by way of contrast, is a contemporary estimate, the basis of which is not clear. See Summerhill, *Origins of Economic Backwardness*, ch. 4 (in progress).

and Sao Paulo/Santo Amaro) exceed the absolute value of the largest negative outlier by 92% and 64% respectively. As in the approach taken to the three large western counties, observations for Lorena and Sao Paulo/Santo Amaro were dropped, and the regressions run again without the large positive outliers. The key change in the results was that the estimates on farm area gini, and farm wealth gini, became statistically significant, yet remained strongly positive. As a last test of the sensitivity to outliers, all 15 outliers were dropped, and the estimates run again (again, with cluster robust standard errors). This further strengthened the results, both with respect to the statistical significance of the coefficients, and in terms of the overall fit and the F statistic. But outliers are in no way driving the result by which inequality fails to negatively correlate with long-term income. Removing the outliers did not alter the general thrust of the OLS estimates--both higher inequality, and the presence of a colonial aldeamento, were associated with higher levels of per capita income in 2000.

Finally, one concern is that the effects of inequality and other covariates could differ across the (conditional) distribution of the log of per capita income in 2000. If the parameter estimate on inequality is not the same across quantiles, the OLS estimates can obscure a potentially important relationship between inequality and development for a subset of the counties. Quantile regression is more robust to outliers than OLS, and is semi-parametric, making it especially useful in the presence of heteroskedasticity. I performed a quantile regression analysis using bootstrap standard errors, with parameters estimated at the 25<sup>th</sup>, 50<sup>th</sup>, and 75<sup>th</sup> percentiles. The coefficients on inequality in fact differ across quantiles. Yet the results of quantile estimation leave the qualitative findings from the OLS results unchanged. In particular, the sign on farm wealth inequality remains positive.

In the following subsections I return to the OLS setting, and employ additional economic controls to better ensure that the true effect of inequality on long-term income is not negative. I further incorporate a new set of variables that indicate at least partially the degree of political inequality. I finally consider alternative outcome variables, to check if the effects identified here are simply a result of using per capita income in 2000 as the dependent variable.

### *Other economic controls: early twentieth-century agriculture*

Crop mix is often associated with long-term effects, particularly if there is some sort of lock-in based on crop choice. The weakness of crop-mix variables (and output-mix variables) is that they are endogenous not only to “endowments” but also innovations, such as transport improvements. Crop choices thus reflect a number of features, including climate, geography, and access to market. The advantage of Easterly’s (2007) instrumental variable on the suitability of land for wheat vs. sugar becomes clear, since it moves away from actual plantings to address potentials, better isolating endowments. Here, three approaches were taken to testing for the effect of crop mix (the results are suppressed in the interest of space). The first tack included the share of cultivated land in coffee in 1905, along with the shares of land in other major crops in the regression on year 2000 income. The parameter estimate on the share of cultivated land in coffee was positive and significant. This suggests that a durable effect from coffee cultivation that may have locked in patterns of labor relations for several decades, even if coffee was no longer the principal activity of municípios in 2000. The second approach uses the share of crop

output in gross farm product (including extractive and livestock). Areas reliant on livestock and forestry have lower outputs per unit of land, and may exhibit poorer long term growth. Here, however, there was no effect. The third approach goes beyond the distinction between coffee and non-coffee. Output figures by county and crop are available for 1905 and 1906.<sup>11</sup> They must be used with caution, since they fail in some cases to report information on locally idiosyncratic crops. This is less of a problem for inland counties. But to illustrate the nature of the measurement error that can arise, the farms in the county of São Vicente produced mainly bananas, oranges, and manioc. The last three products are not included in the state-level crop production figures for 1905, and do not enter in the measures for either the value of agricultural output or gross farm product for any year. Similarly, the county-wide aggregates for Conceição de Itanhaem show mainly some small amounts of rice and beans in the 1906 crop as the county's food production. Yet the farm-level census from 1905 indicates every farm in the county engaged principally in manioc cultivation, a non-reported category in crop output in 1906. Some measurement bias is inevitable, but the cases like these counties are few overall. Importantly, the share of coffee in total crop output in 1905 was found to figure positively and strongly in year 2000 per capita income, while the shares of rice and beans in particular are *negatively* correlated with year 2000 income. These effects come in addition to the market-exposure effect of having a railroad, and likely reflect specific features of the agronomic endowment, a feature addressed in more detail in the IV estimates below.

*Other economic controls: industry, banking, immigration, and local public goods*

Table 6 presents further results from alternative specifications that incorporate other economic variables. These controls are of particular interest because they either capture key features of the São Paulo economy in this period, or have otherwise been singled out for their potential to foster growth.

Most of São Paulo's counties were overwhelmingly agricultural in 1905. Brazilian manufacturing was still concentrated in the city and state of Rio de Janeiro. Nonetheless, pockets of consumer industry had already appeared across São Paulo, which at the beginning of the twentieth century was on its way to becoming Brazil's principal industrial center (Dean, 1969; Cano, 1997). It is well known that manufacturing can exhibit agglomeration economies that prove highly durable over long periods of time. To test the hypothesis that early location of industry in the state of São Paulo mattered for year 2000 income, and might offset the inequality variable (especially in counties with more manufacturing), Specification 1 of Table 6 adds a

---

<sup>11</sup> Every county except Amparo reported its crop output in 1906. For Amparo, physical outputs from 1905 are used instead, and are valued and aggregated using the product prices for 1906 from its largest neighboring county.

variable for industrial capital per capita in 1907.<sup>12</sup> The parameter estimate has the "correct" sign, but lacks both statistical and economic significance.

Beyond industry, banking may have played a role both in the level of economic activity in the first decade of the twentieth century, and in the long run. Banking in the state was concentrated heavily in the city of São Paulo, but there were local banks based in other counties as well. Proximity to a bank may have made access to credit easier for farmers, either because it lowered monitoring costs for the lender, or even because insider connections were important in gaining access to capital. The second specification adds banking assets per capita. The coefficient has the right sign, and proves to be both statistically (and economically) significant.

The third leg of the tripod underpinning São Paulo's rapid growth, beyond railroads and coffee, was immigration. Brazil was second only to the United States and Argentina in the number of immigrants it received in this period. Within Brazil São Paulo was the single most important destination of immigrants. Holloway (1980) in particular highlighted the possibilities for upward mobility that São Paulo agriculture offered to immigrants. The county's share of Brazilian-born farmers in 1905 had strongly negative effect on income in 2000. This effect was robust to variety of other controls not reported here, including measures of the partial productivity of land and coffee trees. Counties with more immigrant farm owners had systematically better long-term outcomes.

Most importantly, the coefficient on inequality (Overall Farm Wealth Gini) remained positive and significant irrespective of the controls. Inequality in 1905 had no detectable deleterious effects on income in 2000.

## **Inequality and Other Outcomes**

### *Long-term outcomes*

For some investigators income per capita remains the best overall long-term indicator of human well being. A growing number of scholars, however, view income alone as far too narrow of a measure. Lest there be suspicion that the results thus far are some fluke of using income per capita as the outcome variable, Table 9 presents OLS results on a variety of other measures of development. The first three specifications drop income per capita as the dependent variable, and use instead Human Development Indices (HDI) for income, education, and longevity. Despite the appealing reference to "development" in the name, HDIs are little more than gussied-up indices of gross domestic product per capita, literacy and schooling, and average life expectancy (and the overall HDI, not used here, is nothing more than a simple average of the indices for the three component HDIs). The fourth specification in the table uses a related measure, the IDFM, a composite index of employment, health, and education at the municipal level. Specifications 5 and 6 are proxies for human capital, and together are related to the HDI for Education in Specification 2. Specification 5 uses the average years of schooling completed

---

<sup>12</sup> Contemporaries and later observers noted that the 1907 manufacturing census fell short of a full count. Nonetheless, a careful comparison of the textile firms enumerated in 1907 with those of the textile census in 1905 shows that the 1907 survey was far more complete in São Paulo.

by adults, and Specification 6 employs the literacy rate for people 15 years of age and above. Finally, Specification 7 uses an index of Physicians per capita in year 2000. As with income per capita, these variables are combined into 1905 CTUs using population shares of year 2000 counties.

Of the seven specifications with different outcome variables, in no instance are higher levels of inequality in 1905 associated with lower levels of income in 2000. In most cases 1905 farm inequality is both positive and statistically significant. The results are sensitive to the use of other definitions of the inequality variable; for most specifications the parameter estimate on inequality becomes statistically insignificant using other ginis. In some instances it actually turns negative, taking on the “expected” sign, but never at levels of significance that are conventionally acceptable. Using alternative indicators of development, and differing measures of inequality, not a single result showed farm inequality in 1905 to be significantly and negatively related to outcomes in 2000.

### *Contemporary outcomes*

The reduced form relationship between inequality and contemporary outcomes can be readily checked. Table 10 looks at inequality and three variables taken as indicative of the level of development in 1905: infant mortality, farm output per capita, and local public expenditures per capita on education. It is noteworthy that the impact of colonial *aldeamentos* turns negative for two of these outcomes. To the extent that the *aldeamento* variable by 1905 proxied higher levels of urbanization, it could easily be correlated with lower *agricultural* output per capita (as in Specification 2). Its correlation lower *local* expenditures on education (Specification 3) quite likely stems from the fact that in places with higher populations the state government, rather than local governments, was the primary provider of primary schooling.

A more extensive treatment of the determinants of infant mortality is in preparation, but Specification 1 reports correlations in a simple setting. It shows the expected positive relationship between infant mortality rates in the first decade of the twentieth century and inequality, but without statistical significance. Stripping out the covariates (not reported here) does not change the relationship, and the direct correlation between infant mortality and all the gini coefficients is very low. Infant mortality in the first decade of the twentieth century is better explained by land productivity (with higher productivity lowering mortality), the time since settlement, and the non-immigrant share of farm owners.

In the second specification higher inequality is associated with lower farm output per capita, but in light of the standard errors the inequality effect is once again effectively nil. Per capita agricultural productivity in 1905 is, unsurprisingly, correlated with conventional features: the higher fertility of more recently settled land (founding year, and output per unit of land), and access to more distant markets (railroad).

Specification 3 illuminates a political economy factor that is often believed to be important, and paves the way to the next subsection. Where inequality is high, and the franchise restricted as it was in Brazil, local elites may successfully limit public education, both to economize on redistributive transfers, and to reduce the mobility of workers (Sokoloff and Engerman, 2000).

The local provision of a public good in a place with high inequality would be expected to be low if elites had captured the local political process. But in specification 3 there is no effect from inequality (nor is there an effect using any of the other gini measures, not reported). Nor does higher levels of farm product per capita appear to matter. The key proximal determinant of local public goods provision in São Paulo was the share of immigrant land owners (Summerhill, in preparation). Where there were higher proportions of immigrant owners, local elites decided to spend more on education. Given that the immigrants were not enfranchised, this was most likely a “supply-side” phenomenon. By offering higher levels of local public education, counties could recruit more immigrant workers, some of whom became farm owners.

## **Political Inequality, Farm Inequality, and Development**

Acemoglu et al (2008) found that farm inequality in late nineteenth-century Colombia was *positively* related to development. They stressed, however, that it was political inequality, rather than farm inequality, that mattered there for poorer developmental outcomes. Ramcharan (2009) found that farm inequality and political inequality in the early twentieth-century US were related, and both were negatively associated with outlays on public education. These results confirm the view that enfranchised elites would work to block redistribution as well as institutional changes that could dilute their control. This political-capture-by-elites phenomenon figures prominently in various strands of the colonial origins literature, as well as more recent works (Acemoglu and Robinson 2000a, 2000b).

By 1900 Brazil was a caricature of an oligarchic Latin American polity based on political exclusion, patronage, and clientelism. The 1881 electoral reform eliminated the strikingly broad electoral franchise that Brazil had for most of the nineteenth century, along with the system of two-stage elections. The overthrow of the constitutional monarchy in 1889 reduced the franchise even further. The Federal Constitution of 1891 guaranteed the right to vote to literate Brazilian males 21 years age and older, with only a few exceptions (Love, 1970). The *de facto* franchise was far more exclusive. Literacy had to be proven by writing name, age, residence, occupation, marital status, and parents’ names. Local electoral boards, consisting of a judge and leading taxpayers, screened aspiring voters for literacy and other requirements. Those who met the legal requirements still had hurdles to jump. To register to vote required proof of two-months’ residency (conveniently waived if three property-owning citizens vouched in writing for the voter). The opportunities for disenfranchisement, or highly selective enfranchisement by a partisan board, were rich. Turnout in national elections was much lower than one would expect, even based on the *de jure* restrictions.

As the story goes, state-level Republican parties held a monopoly on electoral outcomes within each state. In São Paulo, between 1892 and 1930 the Republican party won all state-level offices with nearly 100 percent of the votes (Love, 1980). The system was similarly rigged at the national level, where the Republican party machines of the richest (and most powerful) states preordained the president, often returning nearly unanimous, and highly manipulated, electoral results. The consensus view is that local control of politics exercised by a wealthy, powerful and often armed class (*coronelismo*, or “colonel-ism”), which served to further enrich the wealthy, with little to no regard for the immiseration of the poor.

New evidence, however, suggest there were limits to how closely São Paulo’s politics hewed to this stereotype of oligarchic, rent-seeking control. Political cleavages within the state’s elite were apparent during the “Old” Republic period, and grew over time (Font, 1990; Perissinotto, 2003). Woodard (2009) in particular shows that local politics were more varied and complex than the longstanding characterizations suggest. There is no disputing that the franchise was tightly restricted in law and practice. But it also varied by county, which is of particular interest given the hypotheses here.

Four variables for county-level political inequality in the first decade of the twentieth century are constructed. The first two are simple ratios: (1) the overall share of the population registered to vote; (2) the portion of registered voters made up by men whose principal occupation was agriculture. The third variable measures the breadth of the effective franchise in terms of farm owners:

$$B = \frac{R}{A}$$

where R is the number of registered voters in the county who occupation is agriculture, and A is number of Brazilian (citizen) farm owners in the census of 1905. A measure of  $B < 1$  indicates a county where even Brazilian citizen farm owners had trouble getting on the electoral rolls. A measure of  $B > 1$  is a county where “agriculturalist” voters exceeded the number of Brazilian farm owners. Counties for which  $B < 1$  are taken as having an especially restrictive franchise, while those for which  $B > 1$  had a more expansive franchise. Where  $B > 1$ —in effect, more democratic—one might expect better long-term outcomes.

The fourth variable is an annual average of the number of newspapers in the county containing political content during the first decade of the twentieth century. Where a local political boss held a firm grip on the voters, there would be little reason to have a political newspaper. An example would be Xiririca, where there was no political newspaper, and which was seen a reliably “rotten borough.” A single political newspaper in a county might be seen as a mouthpiece for the local political boss, and perhaps a vehicle to discourage political entry by rivals. Alternatively, a single political newspaper might indicate a political upstart taking on an established machine. In either case the level of political competition would be low. Two or more political newspapers, however, would be unnecessary if a single *coronel* or local machine had firm control, and would be indicative of greater political competition.

The descriptive statistics of Table 2 show that the franchise varied widely, from less than 1/3 of one percent of the population, to fully 10 percent (a figure consistent with the old, broader franchise of the imperial era). The other variables also indicate a lack of uniformity in the rates of political inclusion and opportunities across the counties. Franchise breadth

The first insight from the data on political and economic inequality is that the correlations among them are low. The principal exception is the franchise variable, which is *positively* correlated with economic inequality. Table 8 presents the results from several different specifications that include political inequality as a control in the regressions of farm inequality on income. Of the first three political variables, only the simple franchise ratio takes on statistical significance, but

has the “wrong” sign—the greater the effective franchise in 1905, the worse the development outcomes in 2000. Neither the agricultural intensity of the electorate, nor the breadth of the franchise, is significant. Higher numbers of political newspapers are strongly associated with higher income. However, inclusion of the Density variable (not reported) completely strips away the effect on political newspapers. More prosperous areas were no doubt more literate, more urbanized, consumers had higher purchasing power, and those counties supported more newspapers in general, including newspapers with political content.

Overall, political inequality at the county level in the early twentieth century is no more negatively associated with long-term outcomes than is economic inequality. The non-negative impact of economic inequality remains robust to the inclusion of controls for political inequality.

## **Colonial Institutions, Slavery, Inequality and Income: IV Results**

OLS estimates, like those above, suffer from a well-known and potentially serious weakness. Violations of the zero-conditional mean assumption will render the estimates inconsistent, such that the coefficients on inequality could be exaggerated, or even take on the wrong sign. Two potential violations are of particular concern here. The first is measurement error in the regressors. Dichotomous variables of whether a county had a railroad or not, or a port, do not convey the quality of the port facilities, nor how cheap the unit cost of shipment was by rail. Census-based measures of farm activity also clearly suffer from survey error of various sources (ranging here from the overstatement of farm sizes, to the assignment of a single monetary value to a large group of farms). A second potential violation of the zero-conditional mean assumption that is of concern is omitted variables. If one or more variables that explain the outcome of interest are omitted, the OLS estimates will be both inconsistent and biased.

To obtain estimates that are unbiased and consistent requires an instrumental variable. Properly instrumenting the endogenous regressor when there is omitted variable bias may even change the sign of the coefficient. To insure that the positive coefficients on inequality produced by the OLS specifications are not the result of bias, several different instrumental variables are employed. The instrument must be both uncorrelated with the error term, yet sufficiently correlated with the endogenous regressor.

Inequality is widely seen as an outcome of other conditions. Easterly (2009), building on historical insights from Engerman and Sokoloff, has stressed that growing conditions (topography, climate, and soil) favorable to the production of cash crops contribute to higher inequality. To the extent that the share of land planted in coffee directly reveals intrinsic conditions that could cause farm wealth to concentrate, it would be useful as an instrument. Exploratory analysis reveals that share of cultivated land in coffee in 1905 is strongly and positively associated with all five measures of inequality. But it is also endogenous, and likely violates the exclusion condition, ruling out its use.<sup>13</sup> Several plausible instruments, including latitude and altitude, were rejected on empirical grounds.

---

<sup>13</sup> Consider two parcels of land, of identical soil and growing conditions, equidistant from port. Let one be autarkic, and the other exposed to the market by virtue of a transport cost-reducing

Growing conditions are an obvious candidate for an instrument. The farm census of 1904-1905 meticulously characterized the predominant “type” of soil on each farm. For two reasons this information is not exploited. The vast variety of soil types reported, and the colloquial labels used for them, make it extremely difficult to combine the soil conditions across farms into a county-wide indicator. Second, contemporary understanding of morphology and pedogenesis, even when based on geological observation and soil sampling, was much less accurate than in recent decades. The soil best suited for coffee, and generating the highest yields, was the “terra roxa” (reddish-purple earth). Terra Roxa originated in diabase deposits, and presented high porosity and high iron content. In the 1905 census terra roxa appears frequently in one guise or another. Yet terra roxa was also widely misidentified by contemporaries, with “almost any soil with a reddish color being confidently and optimistically classified as terra roxa.” (James, 1932). The first recognizably modern and comprehensive soil surveys of São Paulo were completed only at midcentury (Setzer, 1949; Neto, et al., 1951). Given that the quality and precision of information on soil conditions have only increased over the intervening six decades, the most modern soil data is the most informative, even for historical work.

All of the 1905 inequality measures are highly correlated with the share of surface area consisting of two of the five major soil types present in São Paulo at the end of the twentieth century. Moreover, there is a negative relationship between farm inequality in 1905 and the share of land classified as least suitable for agriculture (taking into account soil, declivity, topography, and drainage). There is also a strong negative relationship between inequality and the area made up of cambisols. There is further a strong positive relationship between inequality and the area made up of latosols. Experiments with over-identification tests in the IV framework used each of these as instruments (along with distance to coast as a second instrument). The overall suitability of land for cultivation was a weak instrument. The presence of cambisols exerted a direct impact on year 2000 income per capita, thus violating the exclusion restriction. The area made up of latosols, however, impacted income per capita solely through their effect on farm wealth inequality. Latosols are generally viewed as relatively poor for cultivation, and thus frequently associated with livestock raising. However, they present the iron content (and associated reddish-purple color) typical of terra roxa, and are commonly (but not exclusively) used for coffee planting.

In all of the IV specifications I use limited information maximum likelihood (LIML) estimation rather than the generalized method of moments (2SLS) estimator. With small samples, and possibly weak instruments, the bias of LIML is less than 2SLS. The IV regressions in Table 11 use the share of each CTU made up of latosols as an instrument for farm wealth inequality. For each specification in the table presents the parameter estimate on the farm wealth gini, along with robust standard errors. It also presents the F-statistic for the (joint) significance of the instrument(s) excluded from the structural model (note that this is not the same as first-stage F-statistic for the joint significance of all of the variables, which can be much larger when covariates are included). It further presents the Stock-Yogo critical value for the LIML

---

technology. The latter will have a crop mix that includes more “commercial” crops, even though the two farms are otherwise identical.

estimator when the actual size of the 5% Wald test exceeds 15% (Stock, Wright, and Yogo, 2002). An F-statistic less than this critical value suggests that the estimation may suffer from weak instruments, which would bias the parameter estimate on the inequality variable in the direction of the OLS estimate.<sup>14</sup>

The next rows present the Wald confidence intervals, and most importantly for instances where the instrument may be weak, the size-corrected confidence sets based on Moreira's conditional approach (Moreira, 2003). Assessing instruments for potential weakness using rules-of-thumb for the first-stage F-statistics (Staiger and Stock, 1997), and even the more precise Stock-Yogo criteria (2002), can lead to rejection of an instrument as weak when in fact it is valid and sufficiently "strong." At the same time, F-statistics that suggest the instrument is not weak, when in practice it is, can be misleading (Cruz and Moreira, 2005). The advantage of Moreira's conditional approach is that the confidence sets from the conditional likelihood ratio test (along with critical values and p-values) are of the correct size, irrespective of the weakness of the instruments. Moreover, if the instrument is in fact weak, it is visible in an appreciable difference between the "incorrect" Wald confidence interval and the correct conditional confidence set.

The bottom panel presents the corresponding OLS estimates. The first insight gleaned from these results is that for two of the specifications the IV estimates turn negative, whereas in the OLS estimates they were positive. Yet while in four of the five OLS estimates the coefficient on farm inequality was significant, *none* of the IV estimates are significant, and both the Wald confidence intervals, along with the conditional confidence sets, span zero. One could "blame" the covariates for having too high a correlation with inequality, thereby raising the standard error on its coefficient. Yet in Specification 1 where there is no covariate the coefficient on inequality is also insignificant. Omitted variables clearly biased the OLS estimates on inequality upward. Most importantly, instrumenting the inequality variable did not reveal a statistically significant negative relationship between inequality in 1905 and income per capita in 2000.

Table 12 presents results that add an additional instrument, the distance from the municipality to the coast. The statistics reported are as in the previous table, but can now include tests of over-identifying restrictions, which help assess the validity of the instrumental variables strategy using soil type. Instruments must be statistically independent of the disturbance process, and the exclusion restriction states that the instruments for inequality only affect income in 2000 indirectly through their impact on inequality, rather than having any direct effect on income. The over-identification tests assume that one instrument is truly exogenous, and then test for the exogeneity of the other instrument. Failure to reject the null hypothesis that the instrument is uncorrelated with the disturbance term means that the over-identifying restriction is valid. The p-values on the two over identification tests in the table that exceed 0.05 fail to reject the null, validating the restriction. In addition to the over-identification tests reported in the table, specifications were run with the soil instrument included as a regressor in the second-stage

---

<sup>14</sup> I use the F-statistic for the joint significance of the excluded instruments; if the minimum eigenvalue statistic were used instead, it would make it even easier to reject the null hypothesis of weak instruments, since it is larger than the F-statistic. It is an open question as to which statistic to use, since the tests for weak instruments are predicated on homoskedastic errors, and the estimates here use the robust estimator to correct for heteroskedasticity.

regression. If the exclusion restriction was not valid, the parameter estimate on the instrument in the second stage would be significant. In every case, the area in latosolos met the exclusion restriction, by both the over-identification tests, and by the inclusion of the latosolos variable on the right-hand side in the second stage.

Just as in the first set of IV estimates using latosols as the single instrument, most of the results on farm wealth inequality based on two instruments lose all statistical significance compared to the OLS results. Two of them actually turn negative. Among the positive results, one remains statistically significant. Adding *aldeamento* as a control causes the coefficient on inequality to become significant at the lowest conventional level. All three of the positive IV coefficients exceed their OLS counterparts, including the one for Specification 1. This suggests that in the OLS results the negative attenuation bias from measurement error may have exceeded the positive bias from omitted variables. The covariates all remain positive and significant. The IV estimates in these over-identified specifications differ from those in the just-identified versions. But in no case is inequality negatively and significantly related to year 2000 income per capita.

To further test the sensitivity of the IV results to the choice of instrument for inequality, Table 13 presents results that instrument inequality with the natural logarithm of winter rainfall. Elsewhere, farm inequality has been associated with weather conditions, and in particular their variability (Ramcharan, 2009). Data on long-term weather patterns are available for São Paulo for the latter decades of the twentieth century at the county level.<sup>15</sup> Despite the importance assigned to climate, neither the average temperature over a thirty-year period, nor the standard deviation of temperature, are correlated with the measures of inequality from 1905. Nor were the annual average levels of rainfall, or the standard deviation of rainfall over several decades. While overall rainfall levels were not correlated with inequality, seasonal patterns could work quite differently. James (1932) noted that winter (dry season) rainfall levels were negatively related to coffee yields. This is a climate feature that could concentrate farm land and wealth, via its effects on crop mix and the minimum efficient scale of production. The natural logarithm of average winter (dry-season) rainfall levels from 1960 to 1990 is correlated with the share of cultivated land in coffee in 1905, and with all of the inequality measures. Specification 1 with no covariates gives a result that positive but insignificant. It is also smaller than the OLS version, suggesting the positive bias that is typical of OLS. The addition of covariates risks creating a weak instruments problem, since the F-statistic falls below the Stock-Yogo critical value. This makes the conditional confidence sets particularly important. Overall the results using this instrument are much like those obtained when the area in latosols is the instrument. Of the few negative parameter estimates on the farm wealth variable, none have confidence intervals that are exclusively to the left of zero. There is no evidence that farm wealth inequality in 1905 was negatively related to income in 2000.

Finally, I test the sensitivity of the results to the choice of the inequality indicator. The OLS results of the reduced-form sections above relied principally on the overall farm wealth gini as

---

<sup>15</sup> Establishing long-term weather patterns conventionally requires at least 30 years of data. The available weather data for the late nineteenth and early twentieth centuries offer coverage of less than 25 percent of the state's counties, so I use the more comprehensive data from the later period.

the dependent variable. Table 14 looks at the relationship between the overall wealth gini and income per capita in 2000 with IV regression. For overall farm wealth, the only available instrument that performs adequately is distance to the coast. In three of the five IV specifications (estimated again with LIML) the coefficient on inequality is positive, and on two of these with covariates it is strongly positive and statistically significant, even more so than in the OLS results. Only in Specifications 3 and 5 are the coefficients on inequality negative, and neither is statistically significant. The conditional confidence sets give no reason to believe that the coefficients on inequality in these two specifications are reliably negative.

The upshot of the IV regressions for income per capita in 2000, using two different inequality variables, and three distinct instruments for inequality, is that they reveal not a single statistically-significant negative coefficient on inequality. There is no evidence that 1905 land inequality had a persistent or negative impact on long term development.

## Conclusions

This paper tests the hypotheses that colonial institutions, slavery, and inequality reduced income and undermined development. The analysis uses actual historical measures of these variables, rather than proxies, including an original set of measures of farm inequality in São Paulo in 1905.

Institutions created in the early colonial era did have an impact on long run development. The agenda-setting papers by Engerman and Sokoloff, and Acemoglu, Johnson, and Robinson, have stressed the importance of colonial conditions for development. The findings here show that in São Paulo a previously-unexplored local colonial institution mattered. The insight is different in flavor from the previous literature. The Jesuit-run, indigenously-populated *aldeamentos* of the sixteenth and seventeenth centuries had both “extractive” and “settler” elements. Many actual colonial institutions (rather than those inferred ex post) quite likely exhibited a similarly ambiguous character. The net effect of the *aldeamentos* on long-term development was positive. Counties that had begun as an *aldeamento* in the 1500s had higher levels of per capita income in 2000.

The intensity of slavery, measured near its likely peak in 1872 during the upsurge of export agriculture in São Paulo, failed to account for variations in long-term income across the states. In reduced-form estimates the effect of slavery on income was actually positive. The addition of controls that proxy for the developmental prospects of each area eliminated the positive impact of slavery on development. The absence of a negative effect of “peak” slavery on income runs counter to the Engerman-Sokoloff thesis on the impact of slavery on development. Evidence for Brazil as a whole, however, suggests that the intensity of late *colonial* slavery is associated with lower levels of development. This possibility cannot be assessed yet within the state of São Paulo in the absence of reliable, county-level estimates of the slave population.

Most strikingly, farm inequality has no negative impact on long-term income. An original set of inequality indicators based on some fifty-thousand farms is constructed to test the hypothesis that inequality undermined development. OLS estimates range from no effect of inequality on

growth, to a positive effect of inequality on growth. Inequality is unlikely to be exogenous. The IV approach, using instruments that are both conceptually and statistically appropriate, did not alter key finding obtained from the OLS regressions: there is no negative relationship between inequality in 1905 and income per capita in 2000. The estimated relationship between the two varies, from strongly positive, to insignificantly different from zero. Adding a second well-performing instrument to soil type as an instrument made it possible to conduct over-identification tests, confirming that the effect of soil type ran solely through inequality, with no direct impact on income per capita in 2000.

Inequality, when it is not persistent, does not necessarily indicate dire circumstances that favor later underdevelopment. Nor is it the case that unfavorable outcomes can be pinned mechanically on earlier inequality. The within-state result here using an original set of measures of historical inequality runs counter to an array of other findings, whether across countries (e.g., Easterly, 2007), or within Brazil (e.g., Wegenast, 2010). The principal implication of this very different result is that inequality in São Paulo was not only not persistent, but also unlikely “structural.” The high mobility of factors, especially people, both around the globe and within the state, helped insure that inequality would not persist, despite a legacy of slavery, colonialism, and political oligarchy. Early twentieth-century farm inequality reflected an equilibrium level of market-based inequality that was part of the overall process of economic change, one that in no way worked against long-term growth or development.

## Appendix I: Variable Definitions and Sources

### *Early Twentieth-Century Variables*

Note: most variables are averages of the available annual figures from the first decade of the twentieth century. The cases where data were reported annually (infant mortality, by way of example) were few. For some variables there was only one year in which the information was available. All monetary variables were adjusted to mid-decade real prices using the Catão wholesale price index.

**Inequality Variables (crude farm gini, farm area gini, farm wealth gini, overall farm land gini, overall farm wealth gini):** as detailed in the text.

**Colonial Aldeamento:** as detailed in the text.

**Founding Year/Colonial Founding:** the earliest year that any population center of the 1905 county (not necessarily the county itself) was assigned administrative “status” by a higher level of government. Settlement in almost all instances clearly preceded this date. As such, in most instances it is a lagging indicator that people had begun to concentrate in that location. Sources: Forjaz, 1941; São Paulo, *Divisão Administrativa*, 1908.

**Railroad:** dichotomous variable coded 1 if the county had a railroad in its boundaries in 1908. Sources: Brazil, *Questionários*; Cunha, *Estudo Descritivo*; Silva, *Reforma das Tarifas*; São Paulo, Secretaria de Agricultura, *Relatório*, 1899.

**Port:** dichotomous variable coded 1 if the county had a coastal port in 1908. Source: Brazil, *Questionários*.

**River Navigation:** Nearly every county reported several rivers, but most of these were not suitable for navigation over economically significant distances. The variable is coded 1 if the county was reported as having significant commercial river navigation in 1908. Sources: Brazil, *Questionários*; Companhia Paulista, *Relatório*, 1900-1910.

**Total Farm Area:** area reported held in farms in 1905. Sources: São Paulo, *Estatística Agrícola*; São Paulo. Secretaria dos Negócios da Agricultura, *Relatório*, 1906.

**Population:** total population in 1907. This is the basis for all per capita variables from the early twentieth century. Source: *Anuário Estatístico do Brasil, 1908-1912*. The farm census of 1905 was not a population census. The state *Anuário Estatístico* reported, in conspicuously round numbers, admittedly conjectural estimates of population by county in 1905. For several counties the 1907 the figures from the national *Anuário Estatístico* differed greatly from those estimated by the state for 1905. The state *Anuário Estatístico* in 1909 again reported round-numbered figures for each county, which in several instances were quite different still. One way to try and resolve the conflicting figures is to interpolate 1900 and 1920 census figures. This provides yet a different set of estimates for many counties. Linear (or any other type of) interpolation is probably a poor approximation to actual population figures in many counties, given the very rapid immigration underway between 1900 and 1920. The 1907 estimates were part of a nationwide effort by contemporaries to establish inter-census county-level population figures, and are used here.

**Total Cultivated Area:** area under cultivation in 1905. Sources: São Paulo, *Estatística Agrícola*; São Paulo. Secretaria dos Negócios da Agricultura, *Relatório*, 1906.

**Cultivated Share:** Total Cultivated Area divided by Total Farm Area in 1905.

**Land use by crop:** area planted in principal crops in 1905. Sources: São Paulo, *Estatística Agrícola*; São Paulo. Secretaria dos Negócios da Agricultura, *Relatório, 1906*.

**Coffee Share of Cultivated Area:** Coffee Area divided by Total Cultivated Area in 1905

**Coffee Trees:** number of coffee trees reported in 1905. Sources: São Paulo, *Estatística Agrícola*; São Paulo. Secretaria dos Negócios da Agricultura, *Relatório, 1906*.

**Coffee Output:** quantity of coffee produced in 1905. Sources: São Paulo, *Estatística Agrícola*; São Paulo. Secretaria dos Negócios da Agricultura, *Relatório, 1906*.

**Tree Yield:** Coffee crop divided by number of coffee trees in 1905

**Land Yield:** Coffee crop divided by area planted in 1905

**Farm Value:** Total value of farms, in milréis. Sources: São Paulo, *Estatística Agrícola*; São Paulo. Secretaria dos Negócios da Agricultura, *Relatório, 1906*.

**Share of Owners who were Brazilian:** number of Brazilian farm owners in 1905 divided by total farm owners. Sources: São Paulo, *Estatística Agrícola*; São Paulo. Secretaria dos Negócios da Agricultura, *Relatório, 1906*.

**Share of Brazilian Owners:** number of Brazilian farm owners divided by total farm owners. Sources: São Paulo, *Estatística Agrícola*; São Paulo. Secretaria dos Negócios da Agricultura, *Relatório, 1906*.

**Local Public Education Expenditures:** annual average expenditures on public education, 1901-1910. Sources: São Paulo, *Anuário Estatístico do Estado de São Paulo*, various years.

**Total Municipal Expenditures:** annual average municipal expenditures net of debt service, 1901-1910. Sources: São Paulo, *Anuário Estatístico do Estado de São Paulo*, various years.

**Total Municipal Revenues:** real annual average municipal revenues, net of loans and previous year's budget surplus, 1901-1910. Sources: São Paulo, *Anuário Estatístico do Estado de São Paulo*, various years.

**Coffee Tax Share:** real annual average revenues from tax on coffee cultivation divided by Total Municipal Revenues, 1901-1910. Sources: São Paulo, *Anuário Estatístico do Estado de São Paulo*, various years.

**Industrial Capital:** capital of manufacturing firms in 1907. Source: Centro Industrial do Brasil, *O Brasil*, Vol. 3, 1909.

**Political Newspapers:** annual average number of newspapers carrying political content. Sources: São Paulo, *Anuário Estatístico do Estado de São Paulo*, various years.

**State Voters:** voters enrolled for state elections, 1903 and 1904. Sources: São Paulo, *Anuário Estatístico do Estado de São Paulo*, various years.

**Share Agricultural Voters:** share of total electors listed as agriculturalists on state electoral rolls, 1903 and 1904. Sources: São Paulo, *Anuário Estatístico do Estado de São Paulo*, various years.

**Breadth of Franchise:** Agricultural electors in 1903 and 1904 divided by Brazilian farm owners in 1905.

**Output Mix Variables:** shares of the value of various farm crops the total value of agricultural output in 1905 and 1906; Sources: São Paulo, *Anuário Estatístico do Estado de São Paulo*, various years.

**Land share Variables:** share of cultivated land in each major crop in 1905.

**Gross Farm Product:** sum of agricultural output, livestock output, and extractive output. Sources: São Paulo, *Anuário Estatístico do Estado de São Paulo*, various years.

**Agriculture Share:** value of crop output divided by Gross Farm Product. Sources: São Paulo, *Anuário Estatístico do Estado de São Paulo*, various years.

**Slave Share:** share of the population in 1872 consisting of slaves.

**Density:** population in 1907 divided by total area of the CTU

**Banking Assets:** total assets of banks based in the county. Foreign banks, which were all located in the municipio of São Paulo, are included. Sources: São Paulo, *Anuário Estatístico do Estado de São Paulo*, various years.

### *Year 2000 Variables*

Note: county-level data from 1990s and year 2000 are aggregated up to the CTU based on 1905 counties. Values for each county are aggregated in an intuitive way. Income per capita, HDIs, average years of schooling, and literacy rates are population-share weighted. Farm land gini coefficients are farm area-share weighted across counties in the CTU. Distance to coast, rainfall, temperature, altitude, and latitude are unweighted averages.

**Farm Gini Coefficients, 1995:** data assembled by EMBRAPA

**Soil type:** data assembled by EMBRAPA

**Seasonal and Annual rainfall and temperature:** estimates from the Climate Research Unit, University of East Anglia, from 1960 to 1990, provided by Eustáquio Reis of IPEA.

For all other late twentieth-century and early twenty-first century variables the sources are:

[www.ibge.gov.br](http://www.ibge.gov.br)

[www.seade.gov.br](http://www.seade.gov.br)

[www.ipeadata.gov.br](http://www.ipeadata.gov.br)

## Appendix II: the Construction of Comparable Territorial Units (CTU), 1905-2000

Before the long-run developmental implications of early twentieth-century farm inequality could be assessed, aided by the rich set of contemporaneous controls, São Paulo's counties (*municípios*) from 2000 had to be matched to their counterparts in 1905, and recombined into comparable territorial units to correspond as closely as possible to their original boundaries. The nature and extent of the task is revealed by the maps of São Paulo's counties in 1900 and 1997 in Figure 2. From 171 counties in 1905, the state subdivided into nearly 650 counties. The creation of new counties by the fragmentation of old ones was especially pronounced in the large, sparsely-settled, distant western regions of the state at the start of the century.

The construction of the year 2000 CTUs that match the 1905 counties was done manually. Each year-2000 county was traced back to its 1905 origins using sources that detailed legislation and decrees dictating the administrative subdivision counties over the century. The main source was Forjaz's running compilations of the progressive "dismemberment" of counties beginning in 1929 (São Paulo, 1941). The most recent version of the equivalent source was current through 1993 (São Paulo, 1995). Only a few counties were created between 1993 and 2000, and these could be traced at the state government's statistical agency website for historical demographic statistics ([www.seade.gov.br](http://www.seade.gov.br)). At points it was useful to consult the national statistical agency's on-line reports of each county's origins ([www.ibge.gov.br](http://www.ibge.gov.br)), although not all counties were found. None of the sources was found to be without error, as revealed by frequent cross checks, with the most reliable being Forjaz

There were two main complications in matching counties from 2000 to counties in 1905. The first was when the name of the county changed (which was usually unrelated to subdivision). By way of example, and independent of any later fragmentation, the county of Patrocínio de Santa Izabel was abruptly renamed Igaratá; Similarly, Remédios do Tieté became Anhembi. Name changes made matching even more complicated when a new, and entirely different county, took up the retired name of an old one (with the county of Bocaina being the most bedeviling case). The second complicating feature was the subordination of old counties to new ones, followed at some later date by their re-emancipation. The best indicator that an "old" county which had been subsumed was being recreated was the reappearance of the old county name, though this was not always true, as the Bocaina example illustrates.

Tracing the counties by name and legislation solved the matching problem even in the most difficult cases. Nonetheless, the area of the year 2000 comparable units and 1905 counties might not be the same, even if all of the subdivided counties are properly matched to their 1905 forebears. The following consistency check was applied to the CTUs: if the matching and recombination were done properly, total farm area in the 1905 county should be no larger than the overall area of the comparable territorial unit in 2000. This was checked using the modern figures on county size, and the 1905 farm census figures on farm area. The check revealed that twenty-five of the 170 comparable territorial units in 2000 were *smaller* than the total farmland in their 1905 counties, violating the upper consistency bound.

There are several likely sources of these violations. It is important to note that while counties comprising the CTUs may be “matched” correctly, the areas of those counties may not be perfectly aligned. In 1905, boundaries and areas of many counties remained uncertain or even unknown. The three far western counties of Baurú, São José do Rio Preto, and Campos Novos de Paranapanema (Paulista) subdivided into more than 200 counties total by 2000. The eastern-most reaches of these counties comprised the western frontier in 1905. Only in the early twentieth century did survey-explorers sort out many of the basic geographical features of the *paulista* west, and even then were attacked by Indians while conducting their fieldwork.

Given the nearly complete absence of land surveying in 1905, the possibility of shifting county boundaries (as opposed to merely subdivision), and census survey error, I ignored all violations of the upper consistency bound in which the difference between 1905 county total farm area and the year 2000 CTU area was twenty percent or less. This left ten CTUs for which the size discrepancy was greater than twenty percent. The territorial and administrative history of every constituent county of these ten CTUs was scrutinized in an effort to locate the source of the discrepancy. In a couple of cases the most likely source was indeed the problem of having been subsumed into another county, and then later separated again. In these instances the re-separated counties likely had less territory than they possessed before they had been absorbed. The 1905 Villa Bella, which is 49% larger than its year 2000 CTU, is a case in point. In 1934 it was reduced to the status of a district within the neighboring São Sebastião. Later the same year it was separated anew as a county, but of unknown size. The likely result is that its CTU in 2000 is too small (and that of São Sebastião is now too “large”). The counties of Yporanga and Espírito Santo do Turvo also fall into this category. Without knowing precisely which share of each 1905 area should be reassigned to a different county, no further adjustment can be made.

A few cases simply lack any explanation altogether on the basis of the sources. For example, Bananal is 93 percent over size, and Silveiras some 150 percent larger in 1905 than its CTU. Such a difference, however, cannot be a result of failing to properly match the newer counties to their 1905 counterparts. The records do not reveal even a *single* administrative division between 1905 and 2000 for these counties. This leaves two alternatives for explaining the difference in area between the 1905 county and the 2000 CTU. Either the 1905 assessments of farm sizes were grossly off, or the county lost territory between 1905 and 2000 that is not noted in official sources. In either case, this becomes one source of unavoidable measurement error.

## Appendix III: Measuring Farm Inequality

Gini coefficients for the distribution of farm land and farm wealth in 1905 were calculated by:

$$G = \frac{2}{n^2 \bar{a}} \sum_{i=1}^n i(a_i - \bar{a})$$

Five different gini coefficients were calculated, four of which are based on the farm-level census data for 165 of 171 counties: the *farm land gini*, the *overall farm land gini*, the *farm wealth gini*, and the *overall farm wealth gini*. The fifth measure—the *crude farm gini*-- used the distribution of farms across discrete size categories reported by the state secretary of agriculture. The number of farms for each size category was reported for all 171 counties, and the gini was calculated under the assumption that the mid-point of each category was the mean size of the farms in the category (for farms greater than 1000 alqueires in size, the mean was taken to be 1000 alqueires). The *overall* ginis added to left tail of the Lorenz curve the estimated number of landless families in the municipality, as detailed in the text. This increased the gini coefficients considerably.

The farm census reported the size and value of farms in a straightforward way for most of the 165 counties that were detailed. For a subset of counties, however, omissions and arbitrary aggregations complicated the construction of gini coefficients. The purpose of this appendix is to highlight some of these, and describe the measures taken to arrive at indicators of inequality.

Since an owner might have several farms, care is taken in all of the counties to combine multiple farms under the same owner into a single unit. Names that occur frequently complicate a simplified reliance on owner name. To combine farms, the census had to provide some information indicating that more than one farm belonged to the same person. The most typical cases arose when an owner and farm were listed, and then immediately following for the next farm the census noted the owner as “o mesmo” (“the same”). Sometimes farms were listed under the owner’s name, followed by “herança” (estate), or “herdeiros” (heirs). This indicated the farm was in probate, or had otherwise not been divided among heirs. Such a farm was treated as one unit without any effort to allocate shares of the farms across the potential heirs.

Some anomalies were idiosyncratic. For two counties (Caraguatatuba, and São Sebastião), farm size was reported in the census, but not the farm values. The soil type was listed uniformly as “good” across all of the farms, and the census reported a constant unit value of land for all farms in each county. For these two counties the inequality statistics for farm area, and farm wealth, are necessarily identical. For the county of Belém do Descalvado many of the farm values were missing. However, most farms reported a unit value of land, as well as the farm size. The product of these provided estimates of the value of each farm. The county of Conceição de Itanhaem was yet another anomaly. It had fewer than 100 farms, every one of which produced only manioc. For most farms the value of the property was reported, but for many farms no area was reported, and the farm land ginis and farm wealth ginis diverge accordingly.

Beyond these three idiosyncratic cases, two types of anomalies stand out in the census: (1) for some counties the census delegates were particularly inconsistent in their approach to reporting

farm data, (2) the census compilers in turn were not consistent in how they dealt with such discrepancies across counties.

One of the most significant challenges arose from counties in the number of farm owners what different from the number of farms. In some instances this was authentic in-family ownership. For example, in the county of Avaré three men with the same surname, and three minors, owned a single farm. But in several counties there were numerous instances of an area comprising a single farm, usually sizeable, listed with multiple proprietors of different surnames. It is not made explicit whether these were tenants (something higher than a worker), or whether the census taker had simply lumped together the areas of several farms into one unit because they could not secure sufficient detail about the holdings. The presumption here is that listing in the census as a proprietor indicates ownership.

The problem presented by farms listed with numerous owners was complicated by the inconsistent treatment of such cases by the census' statisticians across counties. For example, in Avaré, 387 separate proprietors were enumerated by the census. A number of them were grouped together as holding a single unit of land, such that only 228 separate areas were enumerated as farms, in the hands of the 387 proprietors. Yet in the Agriculture Secretary's summary by size categories, Avaré is shown having 225 farms. The census officials clearly did not allocate the area or values of these farms across the individual proprietors.

The census' compilers took a very different approach, however, to Santo Antonio da Boa Vista. There, 325 proprietors were enumerated in the census, but with only 71 farms. Nonetheless, in the Agriculture Secretary's summary, the county had 325 farms (equal to the number of proprietors in the census). In order to have arrived at such a result census officials clearly did assign the areas of the 71 farms in the census across the 325 farm owners.

Whether or not the contemporary analysts assigned land areas to proprietors correctly in their statistical manipulations is one concern. I assume that they did so, based on the best information, both statistical and qualitative, that available to them. A different concern is whether the staffers working on the census were consistent in their treatment, county to county. The examples of Avaré and Santo Antonio da Boa Vista show that they were not, and this is a potential source of error that makes comparability of the gini coefficients across these counties difficult.

Little purpose is served by detailing the features for every county that exhibited this grouping problem.<sup>16</sup> The approach taken here for counties in which there were "grouped" farm owners is to allocate the grouped farms in equal shares to all the listed proprietors, before calculating the gini coefficients. These were then checked against the crude farm gini coefficients calculated from the size categories of the Agriculture Secretary's report for 1906 for consistency.

---

<sup>16</sup> In addition to Avaré and Santo Antônio da Boa Vista, this problem was especially salient in the cases of Baurú, São Pedro do Turvo, and São Miguel Archanjo.

## References

- Acemoglu, Daron. "The Form of Property Rights: Oligarchic Vs. Democratic Societies." Working paper: MIT, 2003.
- Acemoglu, Daron, Maria Angelica Bautista, Pablo Querubin, and James A. Robinson. "Economic and Political Inequality in Development: The Case of Cundinamarca, Colombia." Chap. in Elhanan Hellman, ed., *Institutions and Economic Performance*. Cambridge, MA, 2008.
- Acemoglu, Daron, Simon Johnson, and James A. Robinson. "The Colonial Origins of Comparative Development: An Empirical Investigation." *The American Economic Review* 91, no. 5 (2001): 1369-401.
- Acemoglu, Daron, Simon Johnson, and James A. Robinson. "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution." *Quarterly Journal of Economics* 117 (2002): 1231-94.
- Acemoglu, Daron and James A. Robinson. "Political Losers as a Barrier to Economic Development." *American Economic Review* 90 (2000a): 126-30.
- . "Why Did the West Extend the Franchise? Democracy, Inequality and Growth in Historical Perspective." *Quarterly Journal of Economics* 115 (2000b): 1167-99.
- Aghion, Philippe, Eva Caroli, and Cecilia Garcia-Penalosa. "Inequality and Economic Growth: The Perspective of the New Growth Theories." *Journal of Economic Literature* 37 (1999): 1615-60.
- Alesina, Alberto, and Danni Rodrik. "Distributive Politics and Economic Growth." *Quarterly Journal of Economics* 109 (1994): 465-90.
- Alesina, Alberto and Roberto Perotti. "Income Distribution, Political Instability and Investment." *European Economic Review* 40 (1996): 1203-25.
- Banerjee, Abhijit, and Esther Duflo. "Inequality and Growth: What Can the Data Say?" *Journal of Economic Growth* 8 (2003): 267-98.
- Banerjee, Abhijit, and Lakshmi Iyer. "History of Institutions and Economic Performance: The Legacy of Colonial Land Tenure Systems in India." *American Economic Review* 95, no. 5 (2005): 1190-213.
- Banerjee, Abhijit V., Lakshmi Iyer, and Rohini Somanathan. "History, Social Divisions, and Public Goods in Rural India." *Journal of European Economic Association* 3, no. 2-3 (2005): 639-47.
- Barro, R.J. "Inequality and Growth in a Panel of Countries." *Journal of Economic Growth* 5 (2000): 5-32.
- Benabou, Roland. "Inequality and Growth." In *NBER Macroeconomics Annual*, 11-74. Cambridge: MIT Press, 1996.
- . "Unequal Societies: Income Distribution and the Social Contract." *American Economic Review* 90 (2000): 96-129.
- Bloom, David E., and Jeffrey D. Sachs. "Geography, Demography, and Economic Growth in Africa." *Brookings Papers on Economic Activity* 2. Washington D.C., 1998.
- Bonet, Jaime, and Adolfo Meisel. "El Legado Colonial Como Determinante Del Ingreso Per Capita Departamental En Colombia." edited by Banco de la Republica. Cartagena: Documentos de Trabajo Sobre Economia Regional, 2006.
- Bourguignon, Francois, and Thierry Verdier. "Oligarchy, Democracy, Inequality and Growth."

- Journal of Development Economics* 62 (2000): 285-313.
- Brasil, Centro Industrial do. *O Brasil: Suas Riquezas Naturaes, Suas Industrias*. Vol. 3: Industria de Transportes, Industria Fabril. Rio de Janeiro, 1909.
- Brazil. Ministerio da Agricultura, Industria e Commercio. *Anuário Estatístico Do Brasil: 1908-1912*. Rio de Janeiro, 1917.
- . *Recenseamento Do Brasil, 1920*. Rio de Janeiro, 1923.
- Bruhn, Miriam, Gallego, Francisco A. “Good, Bad, and Ugly Colonial Activities: Studying Development across the America.” In *Policy Research Working Paper: The World Bank*, 2008.
- Camargo, José Francisco de. *Crescimento Da População No Estado De São Paulo E Seus Aspectos Econômicos, Ensaio Econômicos*. São Paulo, 1981.
- Cano, Wilson. *Raizes Da Concentração Industrial Em São Paulo*. São Paulo, 1997.
- Catão, Luis A. “A New Wholesale Price Index for Brazil During the Period 1870-1913.” *Revista Brasileira de Economia* 46 (1992).
- Clarke, George “More Evidence on Income Distribution and Growth.” *Journal of Development Economics* 47 (1995): 403-27.
- Coatsworth, John H. “Inequality, Institutions and Economic Growth in Latin America.” *Journal of Latin American Studies* 40, no. 03 (2008): 545-69.
- Carvalho Filho, Irineu de and Renato Perim Colistete. “Education Performance: Was it All Determined 100 Years Ago? Evidence from Sao Paulo, Brazil” MPRA Working Paper 24494. Munich, 2010.
- Cruz, Luiz M. , and Marcelo J. Moreira. “On the Validity of Econometric Techniques with Weak Instruments: Inference on Returns to Education Using Compulsory School Attendance Laws.” *The Journal of Human Resources* XL, no. 2 (2005).
- Dean, Warren. *The Industrialization of São Paulo, 1880-1945, Latin American Monographs*,. Austin: Published for the Institute of Latin American Studies by the University of Texas Press, 1969.
- Easterly, William. “Inequality Does Cause Underdevelopment: Insights from a New Instrument.” *Journal of Development Economics* 84, no. 2 (2007): 755-76.
- Easterly, William, and Ross Levine. “Tropics, Germs, and Crops: How Endowments Influence Economic Development.” *Journal of Monetary Economics* 50, no. 1 (2003): 3-39.
- Engerman, Stanley L., Elisa V. Mariscal, and Kenneth L. Sokoloff. “The Evolution of Schooling Institutions in the Americas, 1800-1925.” Working paper: UCLA, 2002.
- Engerman, Stanley L., and Kenneth L. Sokoloff. “The Evolution of Suffrage Institutions in the New World.” *Journal of Economic History* 65, no. 4 (2005): 891-921.
- Engerman, Stanley L., and Kenneth L. Sokoloff. “Factor Endowments, Institutions, and Differential Paths of Growth among New World Economies.” In *How Latin America Fell Behind: Essays on the Economic Histories of Brazil and Mexico, 1800 -1914*, edited by Stephen H. Haber. Stanford, 1997.
- . “Once Upon a Time in the Americas: Land and Immigration Policies in the New World.” In *Paper presented at the Harvard Economic History Workshop*, 2008.
- Engerman, Stanley L. and Kenneth L. Sokoloff. “Factor Endowments, Inequality, and Paths of Development among New World Economies.” *Economia* 3-1, Fall (2002): 41-109.
- Font, Mauricio. *Coffee, Contention, and Change in the Making of Modern Brazil*. Cambridge, MA, 1990.
- Forbes, Kristin. “A Reassessment of the Relationship between Inequality and Growth.”

- American Economic Review* 90, no. 4 (2000): 869-87.
- Frankema, Ewout. "The Colonial Roots of Land Inequality: Geography, Factor Endowments, or Institutions?" *The Economic History Review* (2009).
- Galor, Oded, Orner Moav, and Dietrich Vollrath. "Inequality in Land Ownership, the Emergence of Human Capital Promoting Institutions and the Great Divergence." *Review of Economic Studies* 76, no. 1 (2009): 143-79.
- Galor, Oded, and Joseph Zeira. "Income Distribution and Macroeconomics." *Review of Economic Studies* 60 (1993): 35-52.
- Glaeser, Edward, La Porta, Rafael, Lopes-de-Silanes, Florencio, Shleifer, Andrei. "Do Institutions Cause Growth?" *Journal of Economic Growth* 9, no. 3 (2004): 271-303.
- Holloway, Thomas H. *Immigrants on the Land : Coffee and Society in São Paulo, 1886-1934*. Chapel Hill, 1980.
- Huillery, Elise. "History Matters: The Long-Term Impact of Colonial Public Investment in French West Africa " *American Economic Journal: Applied Economics* 1, no. 2 (2009).
- James, Preston. "The Coffee Lands of Southeastern Brazil." *Geographical Review* 22, no. 2 (1932): 225-44.
- Kuznets, Simon. "Economic Growth and Income Inequality." *The American Economic Review* 45, no. 1 (1955): 1-28.
- La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert W. Vishny. "Law and Finance." *The Journal of Political Economy* 106, no. 6 (1998): 1113-55.
- Levine, Ross. "Law, Endowments and Property Rights." *The Journal of Economic Perspectives* 19, no. 3 (2005): 61-88.
- Levine, Ross, Norman Loayza, and Thorsten Beck. "Financial Intermediation and Growth: Causality and Causes." *Journal of Monetary Economics* 46, no. 1 (2000): 31-77.
- Lewis, W. Arthur. "Economic Development with Unlimited Supplies of Labor." *Manchester School of Economic and Social Studies* 22 (1955): 139-91.
- Love, Joseph L. "Political Participation in Brazil, 1881-1969." *Luso-Brazilian Review* 7, no. 2 (1970): 3-24.
- Love, Joseph LeRoy. *São Paulo in the Brazilian Federation, 1889-1937*. Stanford, Calif.: Stanford University Press, 1980.
- Luna, Francisco Vidal, and Herbert S. Klein. *Slavery and the Economy of São Paulo, 1750-1850, Social Science History*. Stanford, Calif.: Stanford University Press, 2003.
- Milliet, Sergio. *Roteiro Do Café E Outros Ensaio: Contribuição Para O Estudo Da História Economica E Social Do Brasil*. São Paulo, 1939.
- Monteiro, John M. *Negros Da Terra : Índios E Bandeirantes Nas Origens De São Paulo*. São Paulo, 1994.
- Moreira, Marcelo J. "A Conditional Likelihood Ratio Test for Structural Models." *Econometrica* 71, no. 4 (2003): 1027-48.
- Moulton, Brent. "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units." *Review of Economics and Statistics* (1990): 334-38.
- Naritomi, Joana, Rodrigo R. Soares, and Juliano J. Assunção. "Institutional Development and Colonial Heritage within Brazil." PUC-Rio Texto para Discussão, 2009.
- Neto, J.E. de Paiva, R.A. Catani, A. Kupper, H. Penna Medina, F.C. Verdade, M. Gutmans, and A.C. Nascimento. "Observações Gerais Sobre Os Grandes Tipos De Solo Do Estado De São Paulo." *Bragantia* 11, no. 7-9 (1951): 227-53.
- Nunn, Nathan. "Slavery, Inequality, and Economic Development in the Americas: An

- Examination of the Engerman-Sokoloff Hypothesis.” In *Institutions and Economic Performance*, edited by E. Helpman. Harvard, 2008.
- North, Douglass, William Summerhill, and Barry Weingast. “Order, and Disorder, and Economic Change, Latin America vs. North America,” chap. In *Governing for Prosperity*, Bruce Bueno de Mesquita and Hilton L. Root, eds. (New Haven, 2000).
- Perissinotto, Renato Monseff. “State and Coffee Capital in São Paulo's Export Economy (Brazil 1889-1930).” *Journal of Latin American Studies* 35, no. 1 (2003): 1-23.
- Perrone-Moisés, Beatriz. “Índios Livres E Índios Escravos: Os Princípios Da Legislação Indigenista Do Período Colonial (Séculos Xvi a Xviii).” In *História Dos Índios No Brasil*, edited by Manuela Carneiro da Cunha and Francisco M. Salzano, 611 p. São Paulo, 1992.
- Perrotti, Roberto. “Growth, Income Distribution, and Democracy: What Does the Data Say.” *Journal of Economic Growth* 1 (1996): 149-88.
- Persson, Torsten, and Guido Tabellini. “Is Inequality Harmful for Growth?” *American Economic Review* 84 (1994): 600-21.
- Petrone, Pasquale. *Aldeamentos Paulistas*. São Paulo: Edusp, 1995.
- Ramcharan, Rodney. “Inequality and Redistribution: Evidence from U.S. Counties and States, 1890-1930.” *Review of Economics and Statistics* forthcoming (2009).
- Reis, Eustáquio, Márcia Pimentel, and Ana I. Alvarenga. “Áreas Mínimas Comparáveis Para Os Períodos Intercensitários De 1872 a 2000.” Rio de Janeiro: IPEA, 2010.
- Sachs, Jeffrey, and Andrew Warner. “Fundamental Sources of Long-Run Growth.” *American Economic Review Papers and Proceedings* 87, no. 2 (1997): 184-88.
- São Paulo, Departamento Estadual de Estatística. *Ensaio De Um Quadro Demonstrativo Do Desmembramento Dos Municípios*. Edited by Djalma Forjaz. 3 ed. São Paulo, 1941.
- São Paulo, Instituto Geográfico e Cartográfico. *Quadro Do Desmembramento Territorial-Administrativo Dos Municípios Paulistas*. São Paulo, 1995.
- São Paulo, Repartição de Estatística e Arquivo do Estado. *Anuario Estatístico De São Paulo (Various Years)*. São Paulo: , 1900-1910.
- . *Divisão Administrativa E Divisas Municipaes Do Estado De S. Paulo*. São Paulo, 1908.
- São Paulo, Secretaria dos Negócios da Agricultura, Comércio e Obras Públicas. *Estatística Agrícola E Zootécnica Do Estado De São Paulo, 1904-1905*. São Paulo.
- Setzer, José. *Os Solos Do Estado De São Paulo: Relatório Técnico Com Considerações Práticas*. Rio de Janeiro, 1949.
- Shleifer, Andrei, Florencio Lopez-De-Silanes, Rafael La Porta, and Robert W. Vishny. “Legal Determinants of External Finance.” *Journal of Finance* 52, no. 3 (1997).
- Slenes, Robert Wayne. “The Demography and Economics of Brazilian Slavery, 1850-1888.” Doctoral Thesis, Stanford University, 1975.
- Sokoloff, Kenneth L. , and Stanley L. Engerman. “Institutions, Factor Endowments, and Paths of Development in the New World.” *Journal of Economic Perspectives* 14, no. 3 (2000): 217-32.
- Staiger, D., and J.H. Stock. “Instrumental Variables Regression with Weak Instruments.” *Econometrica* 65 (1997): 557-86.
- Summerhill, William R. “Big Social Savings in a Small Laggard Economy: Railroad-Led Growth in Brazil.” *The Journal of Economic History* 65, no. 01 (2005): 72-102.
- . *Order Against Progress: Government, Foreign Investment, and Railroads in Brazil, 1854-1913*. Stanford, 2003.

- Wegenast, Tim. "Cana, Café, Cacao: Agrarian Structure and Educational Inequalities in Brazil." *Revista de Historia Economica, Journal of Iberian and Latin American Economic History* (2010): 1-35.
- Williamson, Jeffrey. "History without Evidence: Latin American Inequality since 1491." conference paper, 2009.
- Woodard, James. *A Place in Politics: São Paulo, Brazil, from Seigneurial Republicanism to Regionalist Revolt*. Durham, 2009.

**Figure 1. State of São Paulo**



**Figure 2. Evolution of municipalities in São Paulo, 1900-2000**

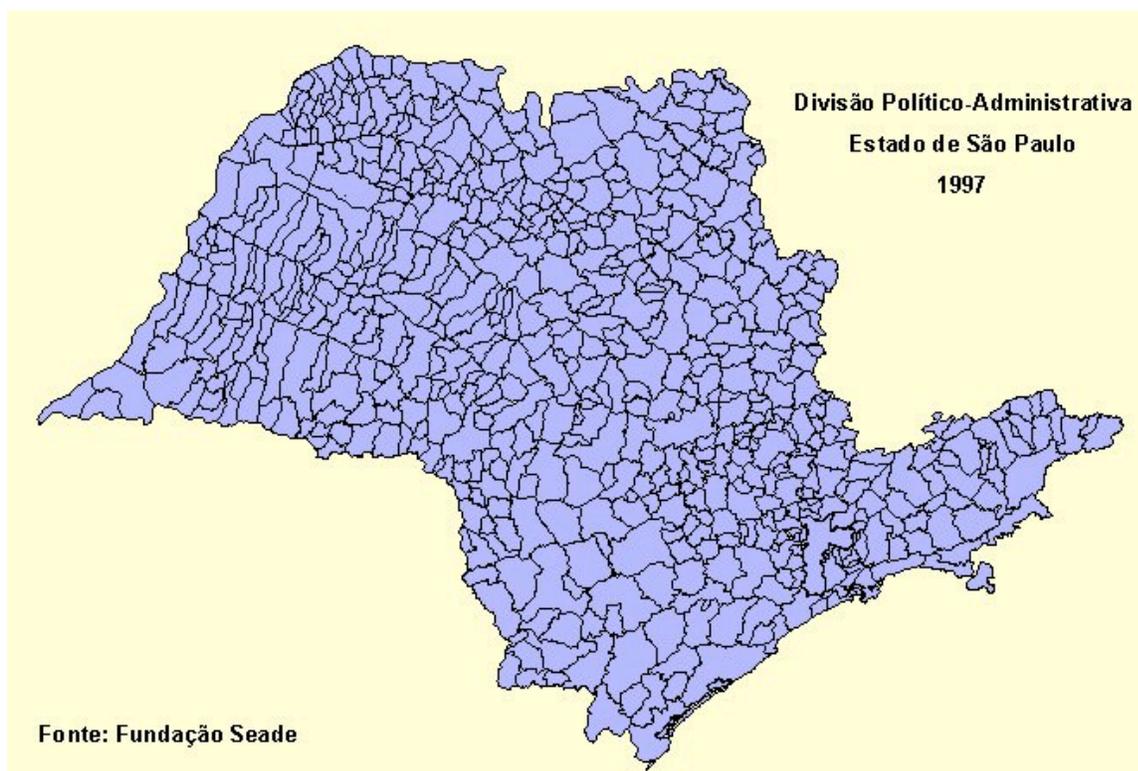
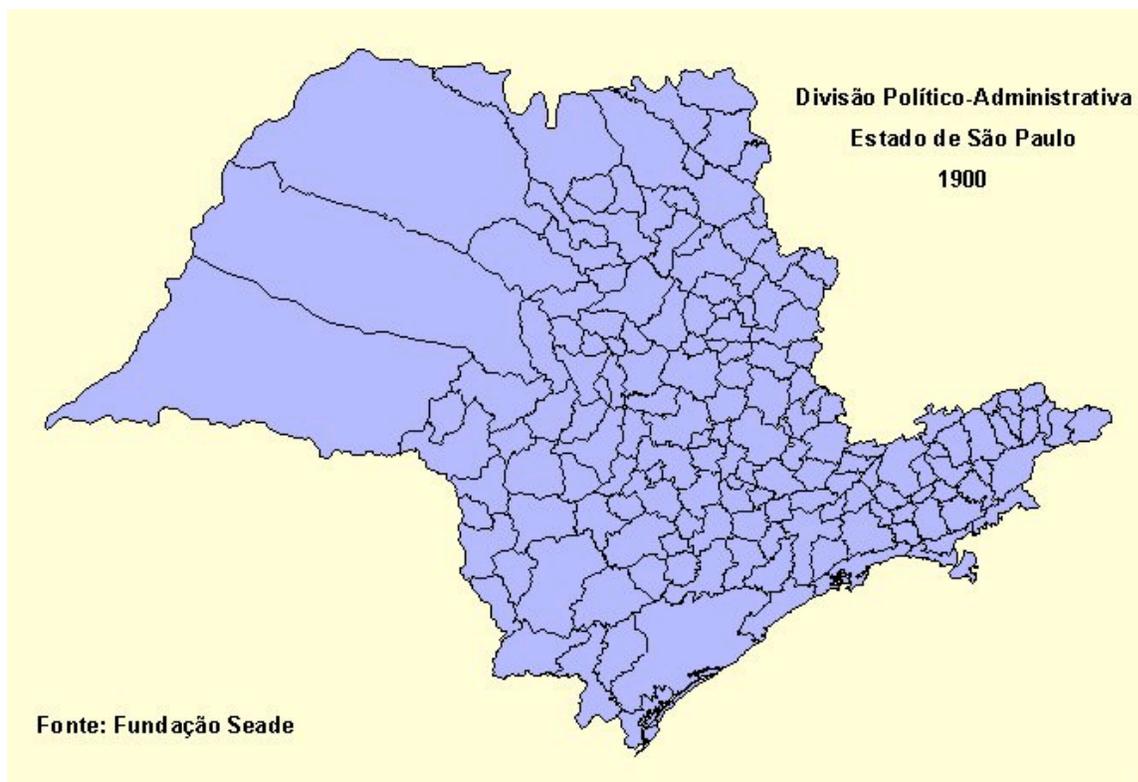


Figure 3. Inequality in 1905: Farm Land and Farm Wealth Gini Coefficients

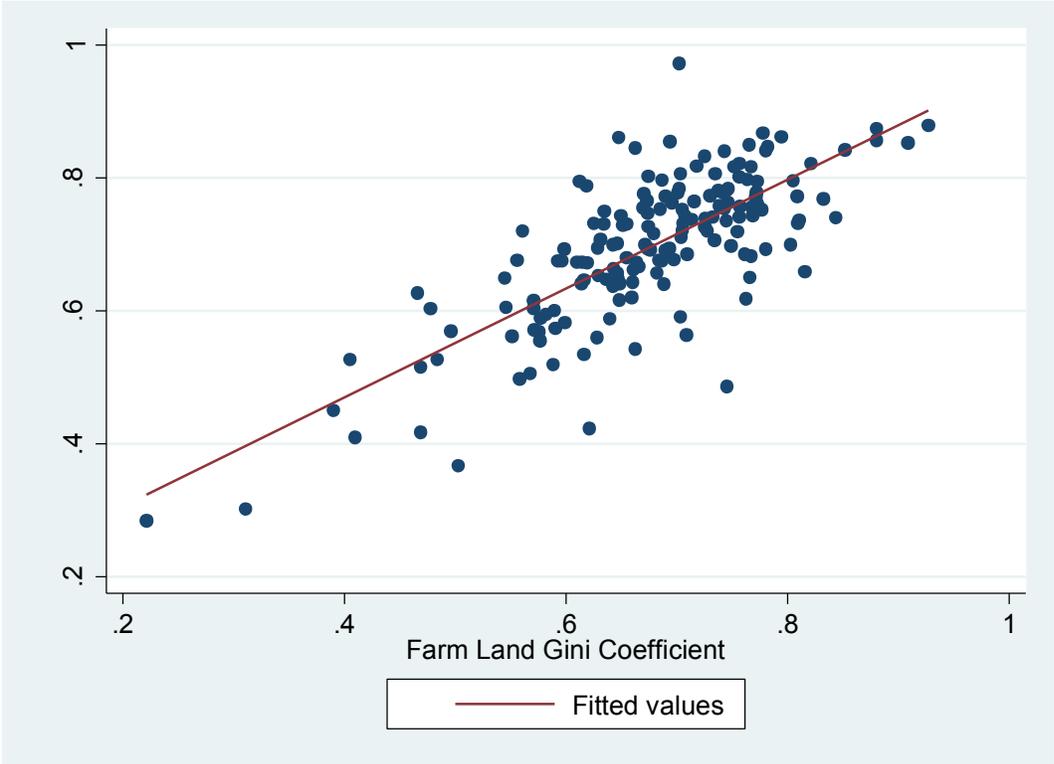
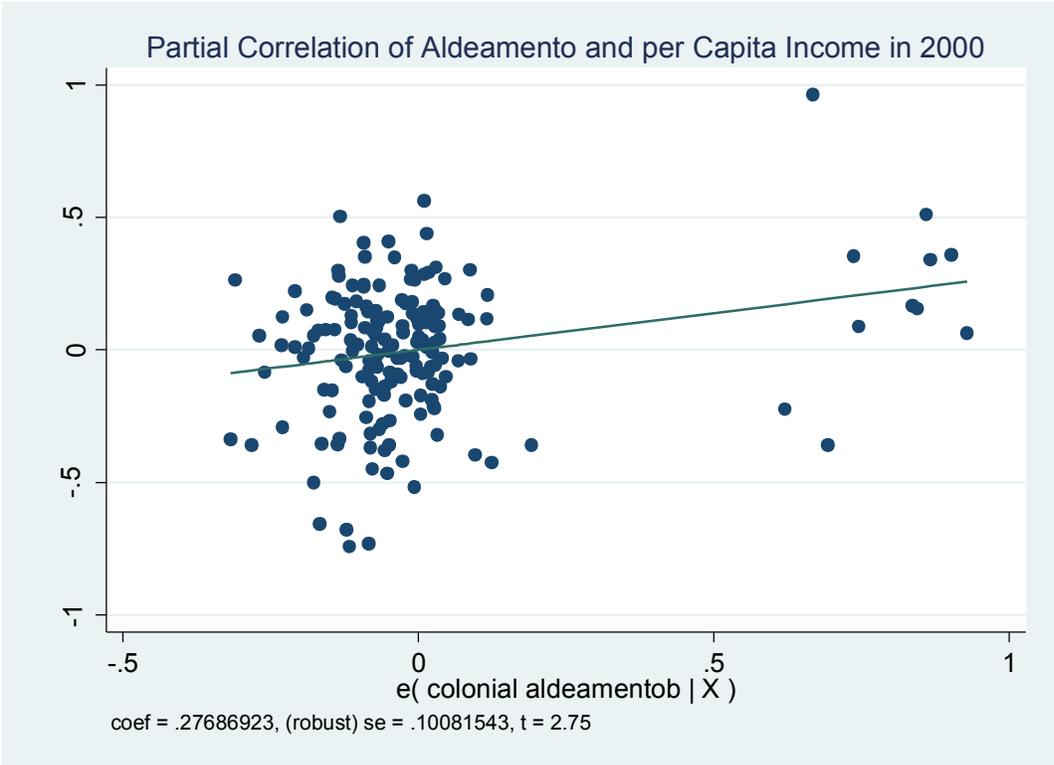
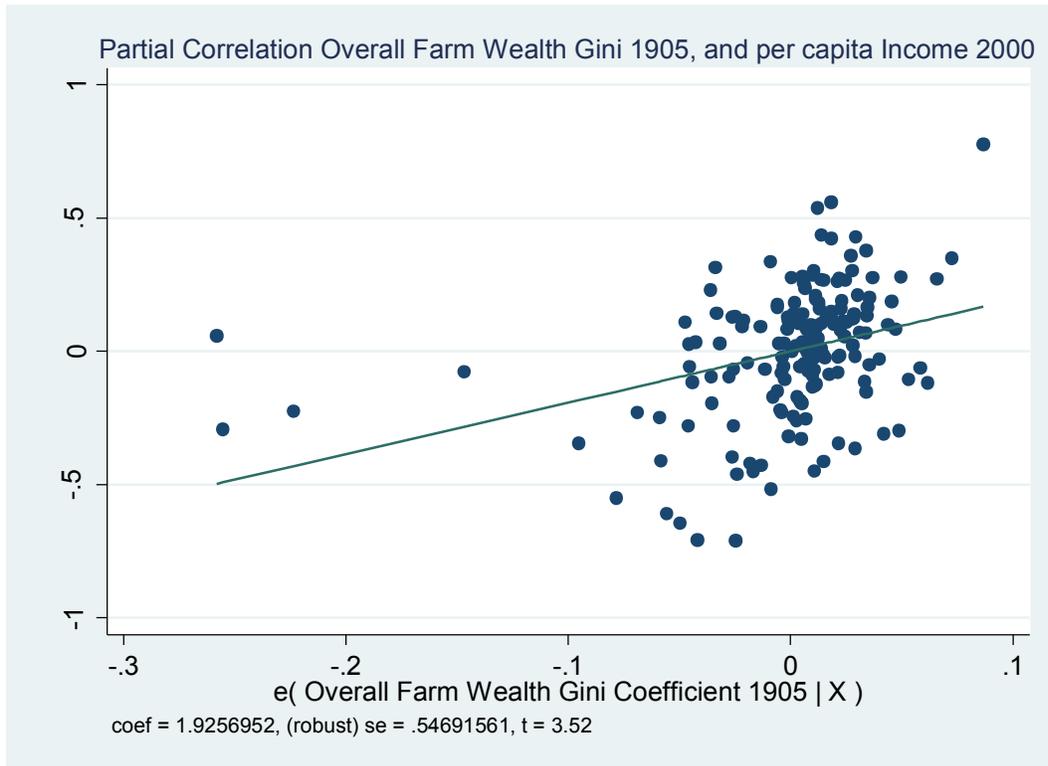


Figure 4. Colonial *Aldeamentos* and per capita Income in 2000



**Figure 5. Overall Farm Wealth Inequality in 1905 and per capita Income in 2000**



**Figure 6. Farm Wealth Inequality in 1905 and per capita Income 2000**

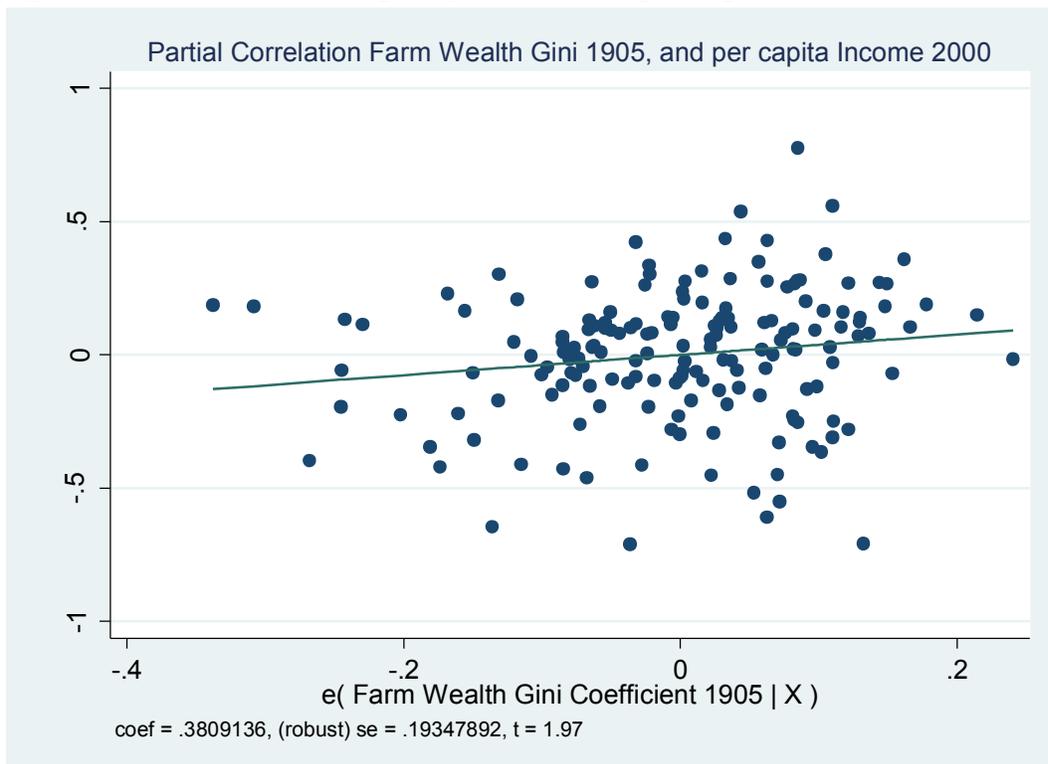
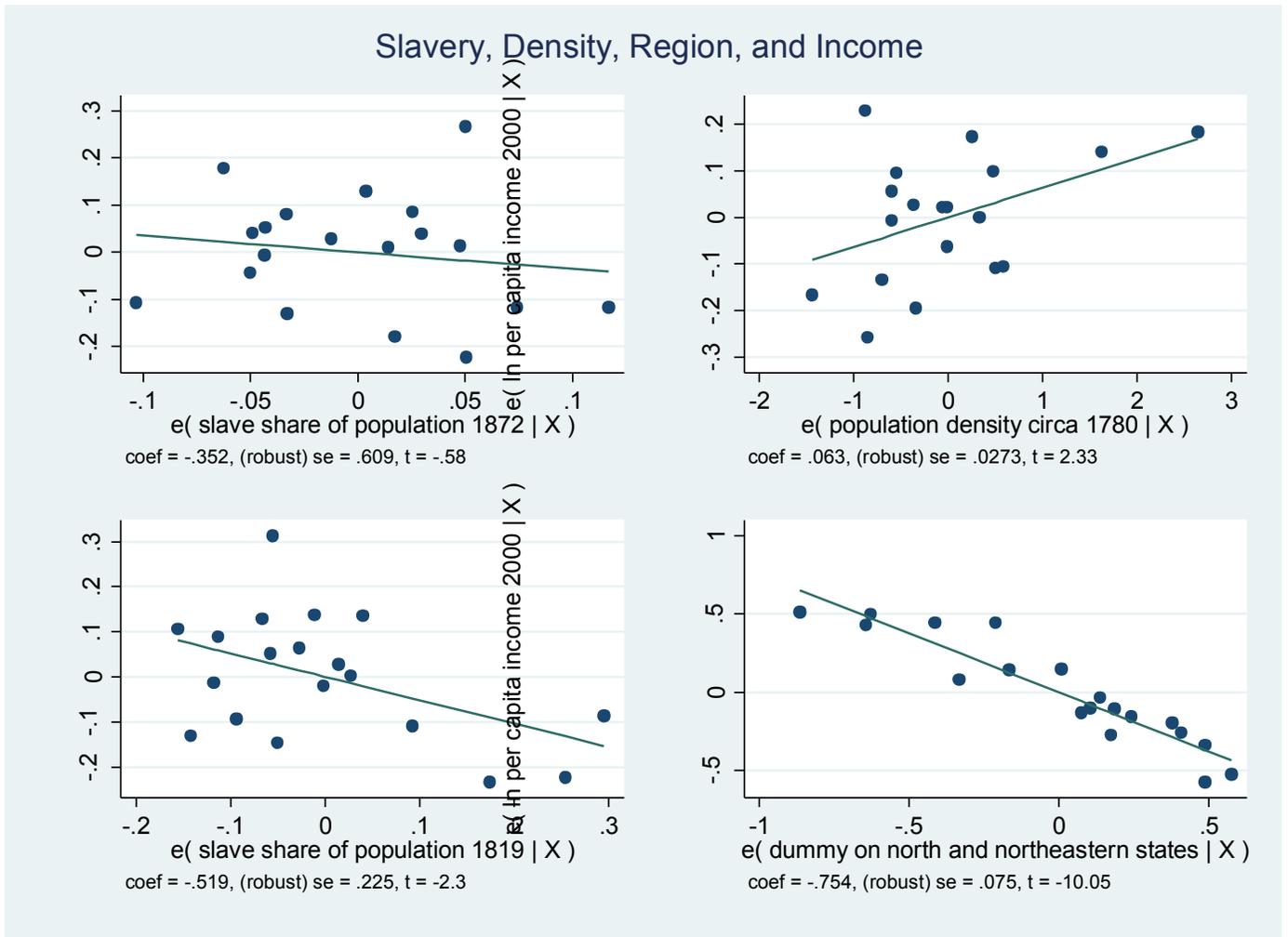


Figure 7. Slavery and Growth: All Brazil, 1780-2000



**Table 1. Farm Gini Coefficients, Descriptive Statistics and Correlations**

**Panel A: Descriptive Statistics of 1905 Inequality Measures**

<b>Variable</b>	<b>Obs</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Min</b>	<b>Max</b>
Crude Farm Gini	170	0.630918	0.10317	0.168971	0.795514
Farm Land Gini	164	0.677151	0.104658	0.22112	0.92738
Farm Wealth Gini	164	0.696343	0.113105	0.28406	0.97292
Overall Farm Area Gini	164	0.958974	0.050434	0.600467	0.998215
Overall Farm Wealth Gini	164	0.962191	0.046284	0.681941	0.999467

**Panel B: 1905 Inequality Correlations**

	<b>Crude Farm Gini</b>	<b>Farm Area Gini</b>	<b>Farm Wealth Gini</b>	<b>Overall Farm Area Gini</b>	<b>Overall Farm Wealth Gini</b>
<b>Crude Farm Gini</b>	1				
<b>Farm Land Gini</b>	0.7923	1			
<b>Farm Wealth Gini</b>	0.6411	0.7562	1		
<b>Overall Farm Area Gini</b>	0.3173	0.2782	0.2223	1	
<b>Overall Farm Wealth Gini</b>	0.273	0.2765	0.3322	0.9235	1

**Panel C: Correlation Between Farm Land Ginis, 1905 and 1995**

	<b>Farm Gini 1905</b>	<b>Farm Gini 1995</b>
<b>Farm Gini 1905</b>	1	
<b>Farm Gini 1995</b>	0.0895	1

**Table 2A. Summary Statistics**

<b>Variable</b>	<b>Obs</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Min</b>	<b>Max</b>
Log per capita Income (2000)	170	5.698	0.284	4.791	6.908
Colonial Founding	170	0.306	0.462	0.000	1.000
Aldeamento	170	0.065	0.247	0.000	1.000
Distance to Coast	170	198.238	150.592	0.027	1137.948
Slave Share 1872	170	0.160	0.086	0.039	0.531
Founding Year	170	1822.735	79.662	1545.000	1897.000
Density 1907	170	23.416	20.851	0.365	189.559
Port	170	0.041	0.199	0.000	1.000
River Navigation	170	0.129	0.337	0.000	1.000
Railroad	170	0.665	0.473	0.000	1.000
Gross Farm Product per capita	170	111.497	81.125	1.423	425.599
Gross Farm Land Productivity	170	98.279	102.771	3.599	535.952
Industrial Capital per capita	170	19.288	89.788	0.000	894.010
Bank Assets per capita	170	5.801	60.587	0.000	784.656
Outlays on Public Education per capita	156	0.165	0.197	0.000	0.851
Brazilian Share of Farm Owners	170	0.845	0.147	0.386	1.000
Voters per capita	170	0.036	0.017	0.003	0.101
Agriculturalist Share of Voters	170	0.665	0.166	0.047	0.974
Franchise Breadth	170	1.821	1.401	0.148	8.125
Political Newspapers	170	0.847	2.078	0.000	25.000
HDI Income (2000)	170	0.722	0.045	0.573	0.839
HDI Education (2000)	170	0.865	0.032	0.775	0.933
HDI Life Expectancy (2000)	170	0.776	0.039	0.600	0.853
Municipal Development Index (2000)	170	0.696	0.057	0.542	0.856
Average Years of Schooling for Adults (2000)	170	5.650	0.864	3.300	7.747
Literacy Rate, 15 Years and Older (2000)	170	90.414	2.990	79.354	95.417
Physicians per thousand (2000)	170	0.728	0.709	0.000	4.566
Infant Mortality Rate	170	0.274	0.139	0.067	0.911
Share of county in Latosols	170	0.534	0.412	0.000	1.000
log Winter Rain	170	4.566	0.233	4.302	5.589