The Unpromised Land: Elites and Human Capital Development in 20th-Century Peru*

Mauricio Espinoza†
Wageningen University & Research

Héctor Paredes‡
Paris School of Economics

Draft Version: January 14, 2026

Abstract

Can landed elites prevent mass education? If so, under what conditions? This article examines the long-term human capital impacts of the Spanish hacienda, the predominant system of agrarian estates with coerced labor granted to local elites during the colonial era in Latin America. Using original district-level data on the placement of haciendas in Peru since 1876 and over a century of population census records, we exploit variation in hacienda crop-suitable areas and age-cohort groups to show that districts with higher hacienda prevalence persistently exhibited lower literacy rates throughout the 20th century. Consistent with a supply-side mechanism, hacienda-dominated districts received lower investments in schools and teachers—without comparable reductions in other public goods—and observed negligible impacts on demand-side proxies such as local economic activity, agricultural specialization, or broader development outcomes. We further show that haciendas were associated with lower indigenous political participation, and that adverse educational effects were strongest in areas where local elites had greater capacity to act—specifically, where a few landowners controlled a large share of the local workforce. Although industrialization, urbanization, and top-down land redistribution later emerged, they failed to reverse the human capital deficits entrenched by the hacienda system.

Keywords: education, land concentration, labor coercion

JEL Classification: I24; I25; Q15.

*We would like to thank Ruerd Ruben, Oliver Vanden Eynde, Claudio Ferraz, Felipe Valencia, Ricardo Fort, Michael Albertus, Karen Macours, Eduardo Montero, Ekaterina Zhuravskaya, Thomas Piketty, Sylvie Lambert, Denis Cogneau, Sara Lowes, Marc Goñi, Elias Papaioannou, Erick Hornung, James Fenske, Ellen Munroe, Manuel Glave, Alex Rivadeneira, Guillermo Woo-Mora, Marcelo Gantier, Andrés Irarrázaval and seminar participants at PSE, LSE, UBC, Lund University, Tilburg University, Maastricht University, the Group for the Analysis of Development, the Pearson Institute at the University of Chicago, the UDEP Seminar on Applied Microeconomics, and LACEA-RIDGE for helpful comments and suggestions. All remaining mistakes are exclusively ours.

†mauricio.espinozahermoza@wur.nl

‡hector.paredes@psemail.eu

"(...) many different ethnic groups under Inca sway regarded the earth as sacred, so sacred that it could not be possessed in perpetuity by anyone."
— Ramírez (2016), p.34.

1. Introduction

During early stages of state formation, the provision of public goods often reflects the outcome of bargaining among local elites and other social groups (Acemoglu and Robinson, 2000; Persson et al., 2000). In agrarian societies, where land access was the primary source of wealth and mobility, the institutions that governed land distribution and labor relations played a central role in shaping this bargaining process. Among the most consequential of these were landed-elite institutions, which concentrated property rights and political influence in the hands of a few. A central tension in the literature concerns the conditions under which these elites chose to obstruct public education to preserve their rents, rather than support it to enhance productivity and capture its gains.

Although a substantial body of research has explored the relationship between historical land inequality and human capital development, empirical findings remain mixed and the mechanisms are poorly understood. On the one hand, it has been argued that the historical formation of large estates fostered economic development by creating local economic interlinkages, stabilizing land tenure systems, and promoting public goods provision (Acemoglu et al., 2008; Dell, 2010; Arias and Flores-Peregrina, 2024). On the other, the political power derived from such estates has been linked to efforts by landed elites to suppress mass education, especially when schooling threatened to increase labor mobility or expand political participation under literacy-based voting rules (Engerman and Sokoloff, 1997; Galor and Moav, 2006; Galor et al., 2009). Yet, micro-level evidence on how these incentives operated—and under what conditions elites had the capacity to act on them—remains scarce.

This paper examines the long-term human capital impacts of the Spanish hacienda, the predominant system of estates under coerced labor force granted to local elites during the colonial period across Latin America. We leverage original district-level archives that detail the complete roll-out of haciendas in Peru, the center of the Spanish Empire in the Americas circa 1500, and one of the last countries under Spanish ruling in the continent. We compile over a century's worth of population census data to link the share of a district's population living in haciendas since 1876 with literacy rates and other measures of human capital development across seven population censuses (1876-2017). To the best of our knowledge, this is the first-time census district-level data on

hacienda exposure is systematically compiled for an entire country, tracking changes in district boundaries and incorporating demographic and economic outcomes spanning more than a century. Additional original data on early 20th century public goods, local economic activity and electoral registration allow us to explore the interactions of supply and demand mechanisms, as well as the incentives and capacities of landed elites at shaping this interaction.

The hacienda system emerged out of the gradual dismantling of the colonial *encomienda* —a Crown-granted regime of indigenous tribute and forced labor—which proved unsustainable due to demographic collapse, legal constraints, and fears of autonomous elite power (Lockhart, 1969; Yeager, 1995; Abad and Maurer, 2025). In its place, haciendas consolidated as large landholdings controlled by Spanish and criollo elites, operating under alternative forms of labor coercion. A process that paved the way for the consolidation of a local elite with significant political power (Davies, 1984; Ramirez, 1994; Hurley, 2011). The digitization of the complete set of republican censuses provides, for the first time, a district-level measure of hacienda presence in Peru since 1876 (Figure 1). Consistent with historical accounts, two main regional clusters emerge: export-oriented zones along the Northern Coast, specializing in sugar and cotton, and more geographically isolated areas in the Southern Highlands, where production remained centered on traditional crops such as maize and potatoes.

While the Catholic Church dominated education under colonial rule—offering religious instruction to indigenous populations and formal schooling only to criollo elites —the republican state struggled to assert educational authority well into the 20th century (Contreras and Cueto, 2013; Walker, 1999). In this vacuum, landed elites in hacienda regions exercised *de facto* control over rural education, often resisting its expansion. The incentives in place were clear: under regimes where voting rights were conditioned on literacy, schooling threatened to undermine their political and labor control (Handelman, 1975; McClintock, 1981; Albertus, 2015). This scenario is consistent with literacy trends throughout 20th century, reflecting stark regional and social disparities. By 1925, only 12% of students enrolled in a school were indigenous and only 33% of adults on the same group were registered as voters. Educational inequalities persisted well into the latter half of the 20th century despite major structural shifts including industrialization, the near-universal coverage of primary education, and land reform.

Given that the initial placement of haciendas was hardly random—potentially following, for instance, preexisting population settlements—, we exploit plausibly exogenous variation in soil suitability to identify the causal effect of early hacienda exposure on long-term human capital development. Specifically, we instrument the share of the

population living and working in haciendas within a district with a composite index that combines: (1) soil suitability for sugar and cotton—the dominant coastal crops of the 19th century—and (2) potential agricultural output (measured in calories) based on staple crops prevalent in the highlands after 1500 CE (Galor and Özak, 2015). The logic behind the instrument is that agro-climatic soil quality, by determining crop suitability, influenced historical land use patterns and the emergence of economies of scale in agriculture. This, in turn, shaped the geographic distribution of haciendas in ways that are plausibly exogenous to later human capital outcomes (Galor and Özak, 2016; Easterly, 2007; Frankema, 2009a; Cinnirella and Hornung, 2016).

The instrument strongly predicts the share of resident population in haciendas, even after controlling for a wide set of baseline covariates—geographic, demographic, and economic—as well as region fixed effects. While the exclusion restriction cannot be directly tested, we show that the instrument's predictive power for educational outcomes only operates in areas where haciendas consolidated—precisely where a mechanism of reduced educational supply would be expected to operate. Furthermore, beyond directly testing a wide set of alternative mechanisms, we argue that any remaining plausible violations, such as better soils raising school demand or improving market access, would likely bias our estimates upward, making the negative effects identified a plausible lower bound of the true impact.

Our results show that a higher share of resident population in haciendas in 1876 persistently lowered literacy rates throughout the 20th century. A one standard deviation increase in hacienda prevalence is associated with a 4 to 23 percentage point reduction in literacy rates before 1961, with the negative effect gradually diminishing over time but remaining significant nearly a century after initial exposure. The findings are robust to a wide set of controls and regional fixed effects, and are not driven by a selected set of compliers, specific levels of hacienda exposure, or an arbitrary subset of covariates. Notably, a complementary age-cohort IV design exploiting individual-level data from modern censuses rules out the possibility that our estimates are primarily driven by selective migration.

Consistent with the central role of landed elites over the local supply of education, we show that hacienda-dominated districts had lower public investments in education—without comparable reductions in other public goods—and negligible impacts on demand-side proxies, such as local economic activity, agricultural specialization, or broader development outcomes. Crucially, we document how the local power of landed elites—measured through proxies of hacienda consolidation—combined with their political incentives to limit the literacy of indigenous workers, were pivotal at ex-

plaining the deliberate underprovision of schooling within hacienda zones. Moreover, while potential mitigating factors such as the rise of competing non-agrarian elites (industrialization), structural shifts in labor markets and top-down land redistribution later emerged, we show that they were ultimately insufficient to counterbalance the long-term adverse impacts of hacienda elites on human capital development.

The paper contributes to several strands of literature. First, the discussion speaks directly to studies on the human capital implications of early land inequality and labor coercion (Acemoglu et al., 2008; Galor et al., 2009; Acemoglu and Wolitzky, 2011; Cinnirella and Hornung, 2016; Ashraf et al., 2024). We extend this literature by documenting the persistence of adverse effects on educational outcomes and showing that these effects are conditional on the political incentives and economic consolidation of landed elites. In this way, our results align with Acemoglu et al. (2008)'s framework, suggesting that when economic and political inequality go hand in hand, unlike their study case in Cundinamarca (Colombia), historical land inequality can be unambiguously detrimental for long-term development. Using a newly compiled nationwide measure of hacienda exposure, we also uncover heterogeneous impacts tied to geographic isolation and market integration, helping to contextualize theoretical predictions and reconcile mixed empirical findings from studies on specific forms of labor coercion or focused on regional clusters sharing specific production characteristics (Dell, 2010; Arias and Flores-Peregrina, 2024; Rivadeneira, 2023; Tabares-Pozos et al., 2025).

Second, our findings speak to the pivotal role of elites for economic development (Acemoglu et al., 2006, 2014, 2021; Bisin and Verdier, 2024). More precisely, by understanding how the responses of elites during early stages of state formation can shape the long-term impacts of mass schooling. From the Church's exclusionary control during the colonial period, to the dominance of hacendados well into the 20th century, our results underscore the enduring influence of local power structures over the course of development. Our findings challenge the view that land concentration necessarily fostered stable land-tenure systems conducive to public goods provision, as proposed on other settings and conjectured by Dell (2010) for Peru's Southern Highlands. Instead, we show that under specific economic and political conditions, the strategic value of public goods for elites can decline, leading to underinvestment in education, while leaving other public goods unaffected.

Third, the paper relates with the growing but relatively scarce literature on competing mechanisms of persistence (Nunn, 2009; Feir et al., 2024; Ashraf et al., 2024; Fenske et al., 2025). This by emphasizing how the legacy of historical institutions—whether beneficial or detrimental—interacts with subsequent economic and political trajecto-

ries, which can either amplify or mitigate their long-term effects. As such, we provide evidence on the relative predominance of a supply-side mechanism, but also assess directly potentially competing forces over the course of a century, including industrialization, sectoral shifts in labor markets and top-down land redistribution. In particular, our findings on mitigating factors connect to recent work that explicitly incorporates such dynamics when assessing the enduring impacts of historical land institutions linked to coerced labor (Goñi, 2023; Abad and Maurer, 2025), their inter-generational transmission (Rivadeneira, 2023; Jaramillo-Echeverri and Alvarez, 2023), or their interaction with later processes of labor emancipation (Ashraf et al., 2024; Markevich and Zhuravskaya, 2018).

The remainder of the paper is organized as follows. Section 2 traces the origins and expansion of the hacienda system in Peru. Section 3 examines how haciendas shaped rural educational provision, focusing on the economic and political mechanisms that enabled landed elites to restrict schooling. Section 4 describes the data. Section 5 presents the empirical strategy, outlines key identification challenges, and reports the main results along with robustness checks. Section 6 explores the role of supply- and demand-side mechanisms, with particular attention to elite incentives and local political power. Section 7 investigates whether major economic and social shifts—such as industrialization and land reform—mitigated the long-run impacts of haciendas on education. Section 8 concludes.

2. The Hacienda System: origins, expansion and consolidation

The Spanish Empire's expansion and consolidation in the Americas—circa 1500—was confronted with a critical problem: how to maximize rent extraction while simultaneously incentivize Spaniards to settle in and effectively govern the newly conquered territories (Arteaga, 2016; Abad and Maurer, 2025). In a context defined by abundant land and natural resources but constrained by legal restrictions on slavery, the scarcity of native labor in labor-intensive sectors such as mining emerged as a central concern (Yeager, 1995; O'Toole, 2012). Furthermore, the composition of settlers—largely private expeditions rather than professional armies, with minimal Crown oversight— exacerbated agency problems in pacifying, controlling, and developing the new settlements (Abad and Maurer, 2025). Consequently, alternative forms of coerced labor, often superimposed on Pre-Columbian or Iberian institutions, became predominant.

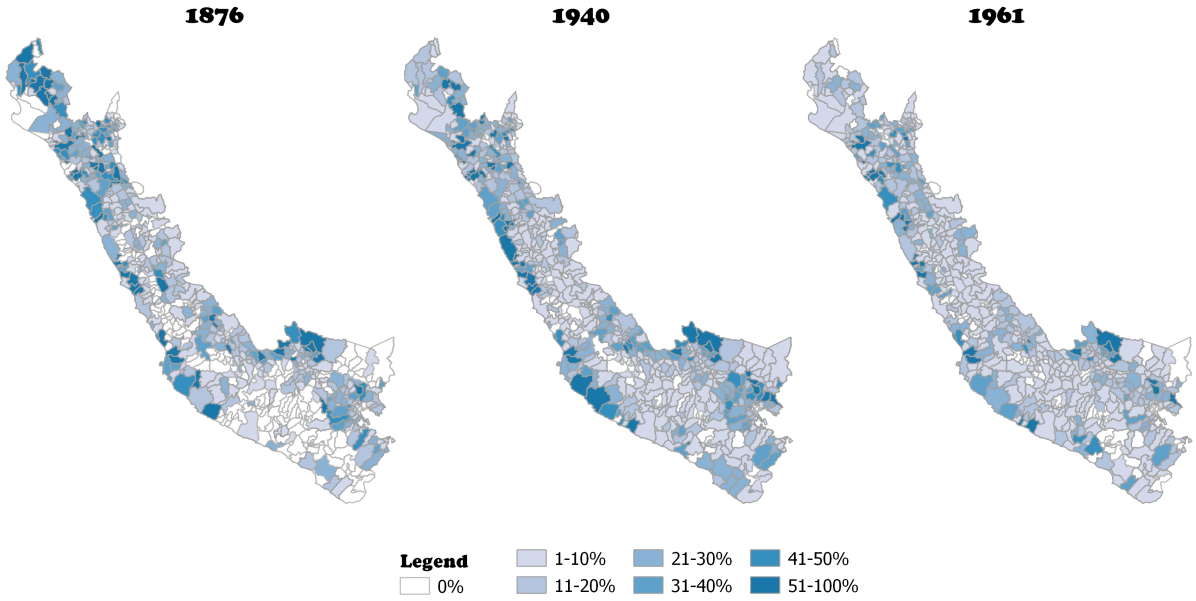
One of the earliest and widespread of these institutions was the *encomienda*: Crown granted rights to extract resources from indigenous populations, as in-kind tribute and forced labor. Originated during the Christian Reconquista in Spain, the model was later exported during the colonization of the Americas and granted to conquerors (local elites) to reward their military services in exchange of providing education in the Catholic faith and "protection" of native populations (Lockhart, 1969). The *encomienda* system facilitated a rapid expansion at minimal fiscal cost, subjugating approximately 20% of the indigenous population at its peak (Abad and Maurer, 2025). However, over time, several factors undermined its sustainability. These included inheritance restrictions that incentivized the accelerated depletion of the indigenous labor force, mobility constraints that limited higher-revenue opportunities and hindered economies of scale, and the Crown's fear of a powerful local elite emerging (Yeager, 1995).

The transition from the encomienda system to the emergence of haciendas has motivated different takes among historians. Although without a unified definition, the term *hacienda* in the Americas generally refers to landed estates -or land rights- of significant size, owned almost exclusively by Spaniards and criollos, and operating under alternative forms of labor coercion[1]. On the transition, authors have emphasized the decline in indigenous population and the growing demand from expanding Spanish cities as main drivers of the emergence of commercial agricultural producers in the form of haciendas (Gill, 2018). Further discussions have described the emergence of haciendas

[1]Throughout the paper, we will conceptualize local exposure to haciendas as a bundled treatment, wherein landownership concentration and coerced labor structures interact to reinforce a persistent domination dynamic between owners (*hacendados*) and workers. Notably, while this dynamic of domination is rooted in economic relations, it extends beyond them, shaping a rigid social hierarchy across multiple dimensions (cultural, political, etc.) (Matos Mar and Mejía, 1980; Favre, 1976).

as a replacement of encomiendas as the main mediator between Spanish cities and indigenous settlements (Lockhart, 1969; Keith, 1976; Frankema, 2009b). Overall, while the transition from *encomiendas* to *haciendas* took diverse forms across regions and over time, one consistent outcome stands out: the persistence of local elites, maintained either through direct lineage or via extended networks of elite families.

Figure 1: Hacienda concentration in 1876, 1940, and 1961



Notes: The figure shows the percentage of the district population residing in haciendas. The sample consists of 665 districts from the coastal and highland regions of Peru (65 districts from the Amazon region were excluded). The average hacienda population share was 0.15 in 1876, 0.16 in 1940, and 0.12 in 1961. The correlation between these measures was 0.64 ($p<0.001$) for 1876 and 1940, and 0.82 ($p<0.001$) for 1940 and 1961, indicating a high persistence of hacienda elites up to the 1960s. In 1969, land reform expropriated virtually all haciendas, so no data is available afterward.

Thus far, research on haciendas in Peru has relied primarily on case study evidence, owing to the absence of systematic data on their historical location and characteristics at the national level. To address this gap, we digitized over a century of population censuses and compile district-level figures of the share of resident population in haciendas since the late 19th century. Figure 1 shows its regional distribution between 1876 and 1961. As observed, the concentration of haciendas resident population remained stable throughout this period, accounting for roughly 15% of Peru's total population and reaching about one million people—most of them Indigenous—by 1940. Accordingly, as late as 1961, nearly 80 percent of all arable land remained concentrated in the hands of less than 1% landowners in the country (Handelman, 1975)[2]. The rest of the coun-

[2]Landholding inequality in Peru was exceptionally high by international standards. According to Frankema (2009a), Peru ranked among the top four countries with the highest landholding inequality circa 1960, based on Gini coefficients across a sample of nearly 120 countries.

tryside consisted of dispersed indigenous communities and smallholders outside the direct control of haciendas.

Consistent with historical accounts, Figure 1 reveals a segmented regional pattern in the distribution of the population residing in haciendas. On the one hand, haciendas were concentrated in areas more suitable for export-oriented agriculture—such as the Northern Coast, where crops like sugar and cotton thrived amid rising international prices in the 1890s (Gonzales, 1993; Thorp and Bertram, 1978; Bértola and Ocampo, 2012). On the other hand, in the Central and Southern Highlands, hacienda elites established control over the most productive lands for livestock and staple crops, often through coercive means, including the dispossession of land from peasant and indigenous communities (Handelman, 1975; Favre, 1976).

The structure of the agricultural sector in the early 20th century led to several scholars to characterize Peru as an "enclave economy", in which productive units operated with limited integration into the national economy and weak oversight from the central state (Thorp and Bertram, 1978). This was not solely the result of institutional choices but also a consequence of Peru's fragmented geography and precarious transport infrastructure, which allowed haciendas to function as semi-autonomous domains under the effective control of local elites. As the next section will discuss, this isolation had far-reaching implications beyond economic organization, particularly for education.

3. From Colonial Exclusion to Local Control: Haciendas and Educational Provision in Rural Peru

During the colonial period and well into the 19th century, the Catholic Church had near-total control over education in Peru. It managed elite schools, seminaries, and universities, while indigenous majorities received religious indoctrination but little academic instruction (Contreras and Cueto, 2013). For the vast indigenous population, educational access was limited almost entirely to catechism and preparation for sacraments —delivered in Spanish only to the extent necessary to recite prayers— rather than on literacy or numeracy (Walker, 1999; Contreras and Oliart, 2014). This exclusion reflected a broader vision of the colonial order in which indigenous peoples were to be governed and moralized, not integrated into the literate public sphere.

After independence, the new Peruvian state began to wrest educational responsibilities from the Church and sought to extend schooling beyond the colonial elite. Yet weak state capacity and high connectivity costs in a geographically fragmented country resulted in unstable jurisdictional authority over education (Arroyo Abad, 2016).

Since the creation of the Ministry of Public Instruction in 1837, the provision of primary schooling oscillated between national and municipal control, with successive legal reforms between 1870 and 1920 aiming to expand coverage. However, implementation lagged significantly behind legal mandates, resulting in persistently limited school coverage and stark regional disparities. By 1940, 49% of coastal children aged 6–14 received instruction, compared to just 24% in the Highlands—barely above the 21% recorded in 1902 (Contreras and Oliart, 2014). In practice, the limited reach of public institutions created a significant vacuum in educational provision across rural areas, creating conditions for education to be *de facto* provided by local dominant actors, resulting in incentives prone to underprovision or clientelistic use.

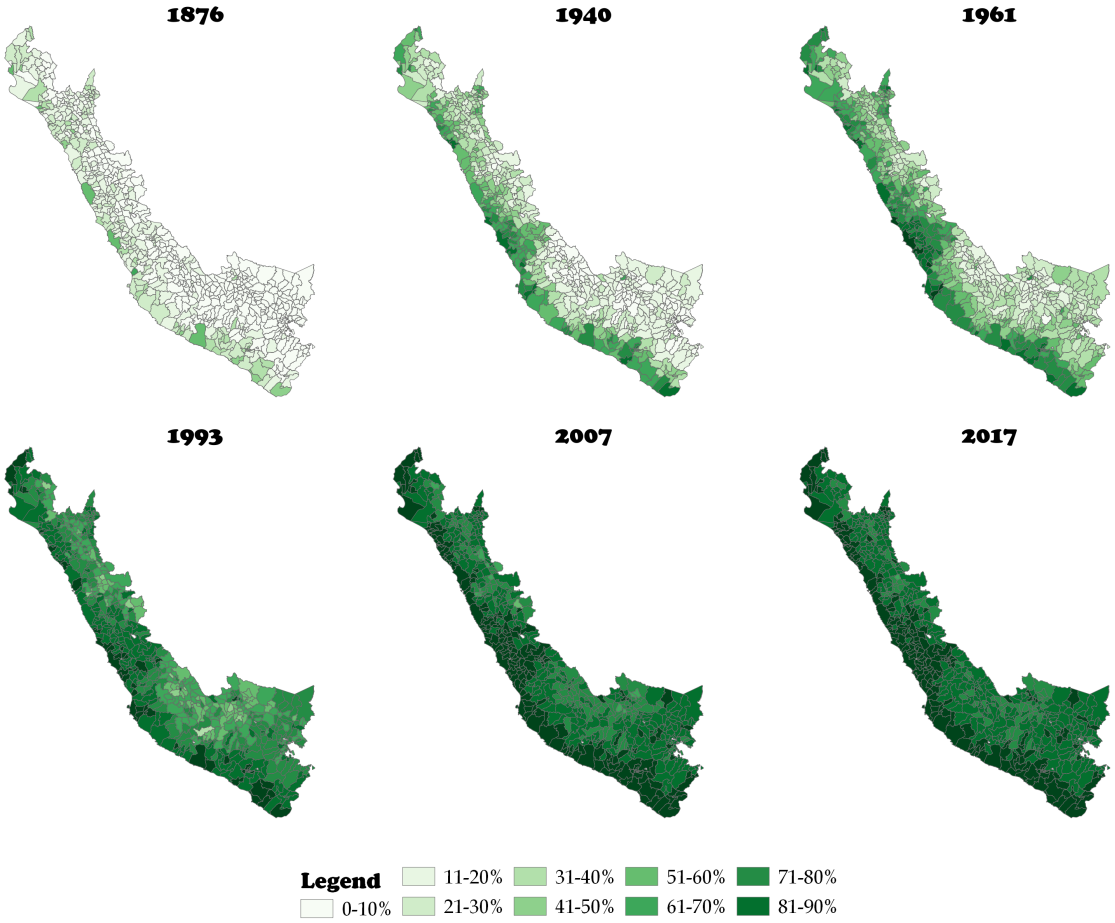
In a context of severe land inequality, hacienda zones epitomized the educational neglect of rural areas at the discretion of local elites. Their influence can be initially understood through the economic relations that structured labor. As agricultural enterprises, haciendas relied on coercive labor systems exerted over a vast majority of indigenous population. Two predominant mechanisms were debt-peonage -—where workers were formally free but remained perpetually indebted to the estate through wages, rents, or advances—- and sharecropping, which allowed peasants to cultivate estate land in exchange for surrendering a substantial portion of their harvest. Both reinforcing a system where workers were legally free yet economically bound to the estate, often across generations (Handelman, 1975; Alberti, 1972, 1976).

Beyond economic dependence, however, haciendas also operated as localized systems of social and political control, reinforced by geographical isolation, limited labor mobility, and weak market integration (Matos Mar and Mejía, 1980; Favre, 1976). Historical accounts suggest hacendados exercised *de facto* control over land access, basic services, justice, credit, and even religion to extend clientelistic power beyond labor relations (Handelman, 1975; Favre, 1976; Matos Mar and Mejía, 1980). Furthermore, their political dominance was sustained through multiple channels, including direct management or capture of key municipal positions (Cotler, 1976; Favre, 1976; McClintock, 1981), instrument peasant union leadership (Albertus, 2015), or by influencing votes through misinformation, intimidation, or outright violence (Stein, 1980; Keith, 1976).

Crucially, because voting rights were conditioned on Spanish literacy until 1980, restricting access to education became a deliberate strategy of political control (Handelman, 1975; McClintock, 1981; Albertus, 2015). As accounted by historians, among landed elites, literacy was considered a threat to the political and economic order by enabling mobility, organizing capacity, and the right to vote (e.g. Handelman (1975)). As

a consequence, hacendados had strong incentives to obstruct access to education. Case studies in mid-20th-century haciendas suggest that most landowners actively avoided providing schools, as education was associated with increased worker autonomy and potential political contestation. In addition, even when schools existed, they often prevented children from attending, or imposed labor obligations incompatible with study (Craig, 1967; Handelman, 1975; Cotler, 1976). Furthermore, in some cases, efforts to establish schools by organized peasants met violent repression (Paige, 1975). Despite widespread exclusion, some highland indigenous communities—operating outside hacienda control—established rudimentary, semi-formal schools that, in response to state absence, persisted well into the 1970s. (Favre, 1976; Puente, 2023)

Figure 2: Literacy rate across census years (1876-2017)



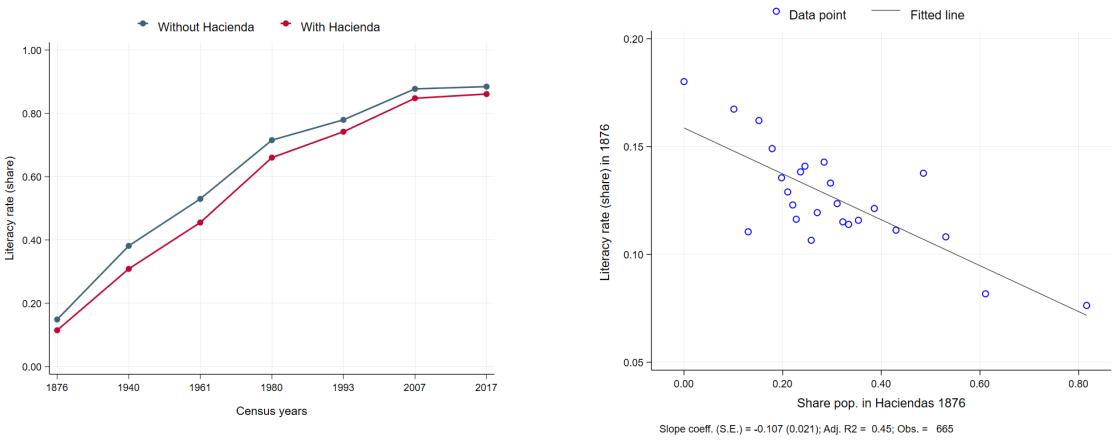
Notes: The figure shows the mean literacy rate at the district level, defined as the proportion of individuals aged 5 or older who could read, write, or both. The sample consists of 665 historical districts from the coastal and highland regions. The mean literacy rate was 0.13 in 1876, 0.33 in 1940, 0.48 in 1961, 0.76 in 1993, 0.86 in 2007, and 0.87 in 2007.

The incentives and capacities of landed elites at preventing access to education are consistent with literacy trends throughout 20th century Peru, as shown by Figure 2.

The evolution of literacy during this period reflect stark regional and social disparities masked by nationwide trends. While literacy rose from about 13% of the population in 1876 to roughly 35% by 1940—and the number of primary schools increased from around 1,400 in 1876 to nearly 5,000 by 1940 (Contreras and Oliart, 2014), most of this growth occurred almost exclusively in urbanized/coastal districts. In the southern and central Sierra (highlands), where the majority of Peru's Indigenous peoples lived, school attendance lagged far behind. As late as 1940, these regions averaged under 20% literacy, compared to over 50% in Lima and coastal areas.

The negative association between historical hacienda presence and literacy rates is both persistent and statistically robust (Figure 3). Panel A illustrates that, despite overall increases in literacy rates over the 20th century, districts with historical haciendas consistently exhibit lower literacy levels. Panel B strengthens this pattern by showing a negative relationship between the share of the population residing in haciendas and local literacy rates. An association that holds even after controlling for geographic characteristics and regional fixed effects.

Figure 3: Literacy and Hacienda Exposure



(a) Panel A

(b) Panel B

Notes: Panel A depicts the districts' mean literacy rate across census years, categorized by historical hacienda establishment. The sample consists of 436 districts (66%) with at least one hacienda and 229 districts (34%) without any haciendas in 1876. Literacy rate is defined as the proportion of individuals aged 5 or older who could read, write, or both. Panel B displays a binned scatter plot showing the relationship between literacy and hacienda exposure across districts in 1876, controlling for longitude, latitude, mean elevation, and department fixed effects. The hacienda variable represents the percentage of the district population residing in haciendas in 1876. The binned scatter plot groups the x-axis variable into 25 equal-sized bins, computes the mean of the x-axis and y-axis variables within each bin, and then creates a scatter plot of these data points.

Educational inequalities persisted well into the latter half of the 20th century despite major structural shifts. These included the dismantling of the hacienda system by top-

down land redistribution under military ruling (1969-1975) and its contested legacy across economic (Albertus and Popescu, 2020; Espinoza and Escobal, 2025), social (Albertus, 2019) and political dimensions (Paredes, 2023); the growth of the industrial sector -albeit delayed and uneven- (Thorp and Bertram, 1978); and the near-universal coverage of primary education, which failed to address persistent deficits in quality and equity (Contreras and Oliart, 2014).

In the following sections, we will provide causal evidence on the hampering educational impacts of haciendas in districts exposed to the system throughout the 20th century. Crucially, an impact driven by the incentives and capacities of landed elites at restricting the local supply of education (schools, teachers, etc.). While potential mitigating factors later emerged -e.g. industrialisation-, we show they were ultimately insufficient to counterbalance the long-term adverse impacts of haciendas on human capital development.

4. Data

To evaluate the effects of hacienda exposure on education in Peru, we analyze district-level data on historical hacienda concentration alongside district- and age-cohort-level census data on literacy, education, and other relevant characteristics. We further supplement this information with data from various sources on early 20th-century public goods, historic and contemporaneous measures of economic dynamism and agricultural modernization, as well as geographic and topographic characteristics.

4.1. Data sources and district harmonization procedure

Our primary data sources include seven national population censuses spanning from the late 19th century to the present (1876, 1940, 1961, 1980, 1993, 2007, and 2017). These censuses provide consistent district-level information on literacy, hacienda concentration, population, ethnicity, and occupations (See Table A1 in the Online Appendix for descriptive statistics on all variables used in this paper). We digitized archival data from censuses conducted in 1961 and earlier using records from the General Directorate of Statistics of Peru (for 1876, 1940, and 1961). The more recent censuses from 1981 onward were accessed in electronic format through the National Institute of Statistics and Informatics of Peru (INEI 1981, 1993, 2007, and 2017).

The number of districts in Peru grew significantly over the study period, increasing from 730 in 1876 to 1,491 by 1961 and to 1,875 in 2017. This expansion primarily reflects district subdivisions due to population growth and political factors. To account for

changes across multiple censuses, we harmonize the universe of districts for each census period following the administrative boundaries in effect at the baseline year—the earliest available (1876)[3]. This harmonization procedure involved reviewing district creation laws and tracing historical administrative division changes from the present back to 1876.

Our analysis thus focuses on the 730 districts in place in 1876. However, we excluded 65 districts from the Amazon region, as they were sparsely populated (with less than 5% of the national population in late 19th century) and lacked haciendas. This results in a final sample of 665 "historical" districts from 1876, covering the main populated areas of Peru.

4.2. Outcome variables

Our main dependent variable is the district-level literacy rate, defined as the proportion of individuals aged 5 or older who could read, write, or both. The average literacy rate increased from 13% in 1876 to 48% in 1961 and further to 87% in 2017[4]. This data thus allow us to trace the complete trajectory of literacy expansion, covering nearly 150 years of progress from a low baseline to near-universal literacy. In addition, we also examine alternative educational outcomes available in more recent censuses, including average years of schooling and school enrollment.

4.3. Hacienda concentration

Data on hacienda concentration is compiled from archives of the 1876, 1940, and 1961 population censuses. They provide a complete list of haciendas in the country, including each hacienda's name, district location, and resident population size. We digitized this data and aggregated it at the district level, creating alternative measures of hacienda concentration. Our preferred measure is the share of the district population residing in haciendas, though additional measures were also explored [5].

Unfortunately, the census data do not provide information on hacienda's land sizes. While this would be a useful measure of hacienda exposure, the use of the share of the resident population in haciendas have several advantages. First, it captures the degree of social penetration and control of haciendas more directly than land area. Second,

[3] In the Appendix, we implement an alternative approach using the districts in effect at each census period, while clustering the standard errors at the 1876 district level.

[4] An inspection of the data reveals strong autocorrelation in district-level literacy rates over time, indicating consistency across population censuses. The correlation between literacy rates in 1876 and 1940 is 0.67 ($p<0.001$), between 1940 and 1993 is 0.80 ($p<0.001$), and between 1993 and 2017 is 0.91 ($p<0.001$).

[5] See Panel M in Table A1 in the Online Appendix for a list of alternative measures

it deals with a relatively lower extent of potential measurement error—for instance, due to heterogeneity in land productivity or use or incomplete/biased cadastral data. The resulting measure displays strong autocorrelation over time, indicating substantial persistence of hacienda elites into the 1960s. Following the 1969 land reform, nearly all haciendas were expropriated, and by the 1980s, most had been subdivided into small farms.

4.4. Control variables

We constructed two types of district-level control variables: one set related to demographic and economic characteristics, and another to geographic and topographic attributes. For the first group, we include census-based measures such as logged population, population density, urban population share, ethnicity composition (white, indigenous, and mestizo), and workforce sectoral shares (agriculture, mining, and manufacturing). These variables are drawn from census data across the seven available periods.

The geographic control variables include district mean elevation, slope, ruggedness, surface area, and centroid coordinates (longitude and latitude). Average elevation, slope, and ruggedness were calculated by the authors in QGIS using the NASA SRTM's Elevation Model from Google Earth Engine Data, while district centroid coordinates and surface area were derived from administrative boundaries in effect in 1876.

5. The effect of hacienda concentration on human capital development

This section analyses the relationship between hacienda concentration and literacy rates. We begin with OLS results from seven separate cross-sections, followed by an instrumental variable approach based on soil characteristics to address endogeneity concerns. Next, we present analyses by age-cohort groups, examine effects on other educational outcomes, and address potential inferential threats through a series of robustness checks.

5.1. OLS results

We start by assembling a district-level panel dataset consisting of 665 historical districts observed at seven points in time (1876, 1940, 1961, 1981, 1993, 2007, and 2017). We examine the relationship between hacienda concentration and literacy rates by estimating seven separate cross-sectional OLS models, one for each census period. This baseline

specification can be expressed as follows:

$$y_{i,t} = \alpha + \beta_1 \text{Hac}_{i,1876} + G_i'\phi_1 + X_{i,t}'\gamma_1 + \delta_d + \varepsilon_{i,t} \quad (1)$$
$$\forall t \in \{1876, 1940, 1961, 1981, 1993, 2007, 2017\}$$

Here $y_{i,t}$ is the literacy rate for district i at census year t; $\text{Hac}_{i,1876}$ is the share of population residing in haciendas in district i as of the 1876 census, standardized to have zero mean and unit standard deviation; the vector G_i includes time-invariant geographic and topographic controls (district area, elevation, slope, ruggedness, latitude and longitude); and $X_{i,t}$ represents a vector of time-variant demographic and economic controls (log of population size, population density, urban share, ethnicity composition, and shares of employment in agriculture, mining, and services) in census years; γ_1 denotes region fixed effects, and $\varepsilon_{i,t}$ the error term. The coefficient of interest, β_1, captures the association between hacienda concentration and literacy rates in each census period.

The OLS estimates, shown in Panel A of Table 1, reveal a significant negative association between historical hacienda concentration and literacy rates across all census periods. This negative relationship remains strong from 1876 to 1961, gradually weakening over time but remaining significant up to 2017. The magnitude and significance of point estimates are remarkably stable across models, even after accounting for a full set of geographical and demographic controls, regional or provincial fixed effects, alternative standard error structures—including robust, wild cluster bootstrap, and Conley standard errors[6], and accounting for an overall positive trend in literacy rates throughout the study period (see Panel A in Figure 3). These findings suggest that hacienda-exposed districts persistently lag behind those not exposed in terms of literacy rates.

5.2. IV design

Unobserved factors correlated with both hacienda concentration and literacy rates, or hacienda estates initially established in areas with lower education levels—e.g., where a larger indigenous population lived—might confound the observed negative association captured by OLS estimates. To address these concerns, we implement an instrumental variable design that leverages plausibly exogenous variation in soil suitability for a subset of pre-19th-century crops, driving differential levels of hacienda spatial concentration.

[6]Tables A5, A6 and Figure A5 in the Online Appendix present these tests for the seven cross-sections.

Historians of the early colonial period in Peru have long noted that the concentration and location of hacienda estates by the late 19th century followed a segmented pattern of production possibilities: areas suited for export-oriented crops, such as sugarcane and cotton along the Coast, contrasted with Highland regions dedicated to traditional farming and livestock, including staple crops like maize and potatoes. Hacienda elites strategically established in areas with higher economic potential, targeting either export markets or local urban demand (Matos Mar and Mejía, 1980; Thorp and Bertram, 1978; Vasquez, 1961). Their ability to secure the most desirable lands—often through coercive means—reflects the considerable economic and political power they wielded, unlike other country cases in Europe and the Americas where elites established estates in areas with lower land demand.[7]

We build an instrument that leverages this segmented pattern of production potential, exploiting geographic differences in crop suitability for a subset of pre 19th century predominant crops, while controlling for a comprehensive set of baseline factors—including proxies for labor availability, another key driver of hacienda placement. The rationale of this approach is that agro-climatic soil quality, by shaping crop suitability, influenced land use patterns through economies of scale, leading to differential patterns of land concentration and hacienda formation, independently of other factors linked to human intervention.[8]

Specifically, we instrument the share of the population living and working in haciendas within a district using a composite measure that captures (1) soil suitability for sugarcane and cotton—the predominant coastal crops of the 19th century—based on irrigated land, low-input adjusted World Agroecological Zones data from FAO,[9] and (2) potential agricultural output (measured in calories per hectare), accounting for staple crops dominant in the highlands post-1500CE (Galor and Özak, 2016). We averaged these indexes at the district level according to 1876 district boundaries, standardized

[7]For instance, in 19th-century Prussia, Cinnirella and Hornung (2016) show that regions that exhibit relative poorer soil quality (i.e., high sand content), and therefore a lower marginal value of land, experienced a lower demand for land, were settled relatively late, and were thus characterized by higher landownership concentration. Other example along these lines is Arias and Flores-Peregrina (2024) for the case of Mexico.

[8]The method draws on a large body of prior studies that exploit natural land characteristics as instruments to address inequality and land concentration. From Easterly (2007)'s seminal work using the abundance of land suitable for growing wheat relative to sugarcane as an instrument for inequality, all the way to more recent examples such as Cinnirella and Hornung (2016) using the share of loamy soils as an instrument for land ownership concentration in Prussia, Ashraf et al. (2024) who employ terrain slope to instrument for the initial abundance of elite-owned physical capital, Rivadeneira (2023) who uses the potential yield of maize relative to potatoes as an instrument for *concertaje*, a colonial-era forced labor institution in Ecuador, or Tabares-Pozos et al. (2025), using suitability for growing potatoes to study the effects of land inequality on human capital formation in Colombia.

[9]Due to the strong correlation of the measures for sugarcane and cotton (0.96), the distribution of the instrument treating both crops jointly or separately generate similar results.

the measures to have a mean of zero and a standard deviation of one, and then selected the maximum value between both measures.[10]

The distribution of the composite instrument, which we label the Soil Crop Suitability Index (CSI), is shown in Panel (a) of Figure 4. As expected, the spatial distribution of the CSI visually reflects the two main regional clusters of hacienda concentration in the 19th century: the Northern Coast and the Central and Southern highlands of the country. Panel (b) shows our measure of historical hacienda concentration, while Panel (c) shows the relationship between the two, visualized in a binned scatter plot. The result reveals a notable spatial overlap between the CSI and the concentration of haciendas, confirmed by a statistically significant positive correlation. Notably, standard coefficient stability tests report that the relative selection of unobservables would need to be at least two times as important with respect to the set of already included controls to make the instrument irrelevant to explain exposure to haciendas (Oster, 2019; Diegert et al., 2022).[11]

Our approach rests on the assumption that soil suitability affects literacy rates solely through its influence on historical hacienda placement. Any direct effect of soil suitability on education access would violate the exclusion restriction, undermining the validity of our estimates. Although it is not possible to test this directly, we argue that such a direct effect, if present, would likely go in the opposite direction of our estimates—better soils should be expected to increase household income and educational demand and investments, improving human capital. This would mean that our estimates underestimate the true negative impact of hacienda concentration on literacy. However, to mitigate these concerns, we provide evidence against direct effects of our instrument in proxies of educational demand and investments, as well as the robustness of our results to selective migration by implementing an age-cohort group IV design that assigns exposure to haciendas at the district of birth of each individual.

We further asses the validity of our instrument through a falsification test. We examined whether soil crop suitability correlates negatively with literacy rates in areas with high hacienda concentration, where an indirect influence via landed elites restricting education is expected, compared to areas with minimal or no hacienda presence, where such an effect is not anticipated (no hacienda elites). Our results confirm that our instrument only exhibits explanatory power in districts with significant hacienda concentration.[12]

[10] We also explored alternative methods for combining both measures and the results were similar. Figure A1 in the Online Appendix maps their distribution and provides further details on the data.

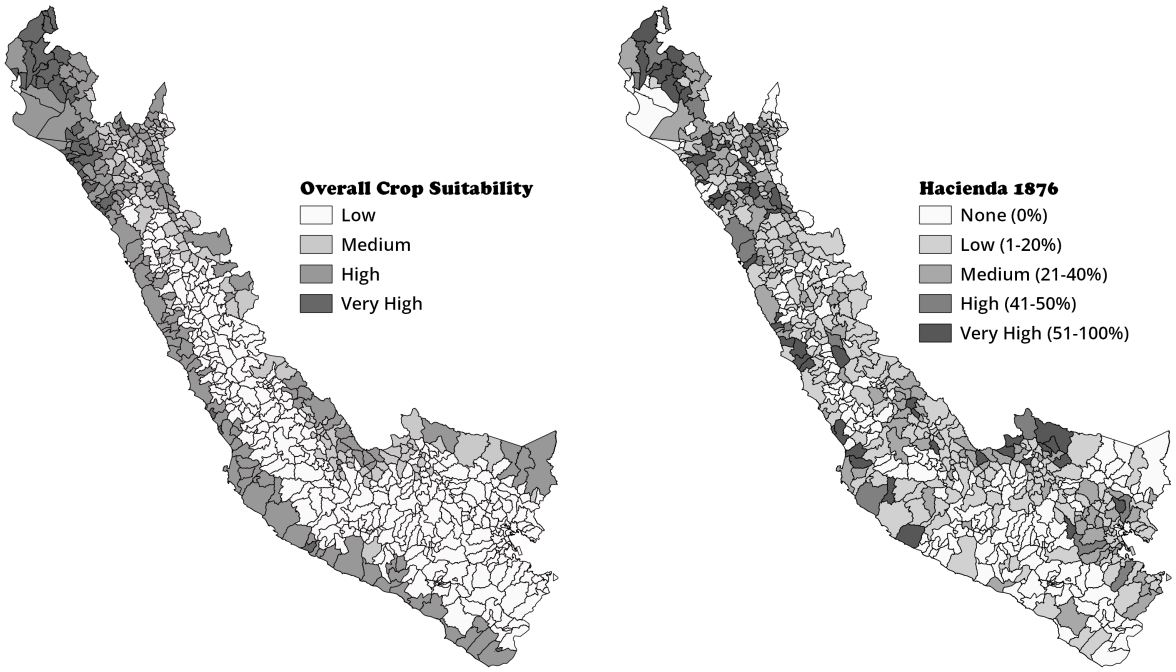
[11] Figure A5 in the Online Appendix

[12] These findings are detailed in Figure A2 in the Online Appendix.

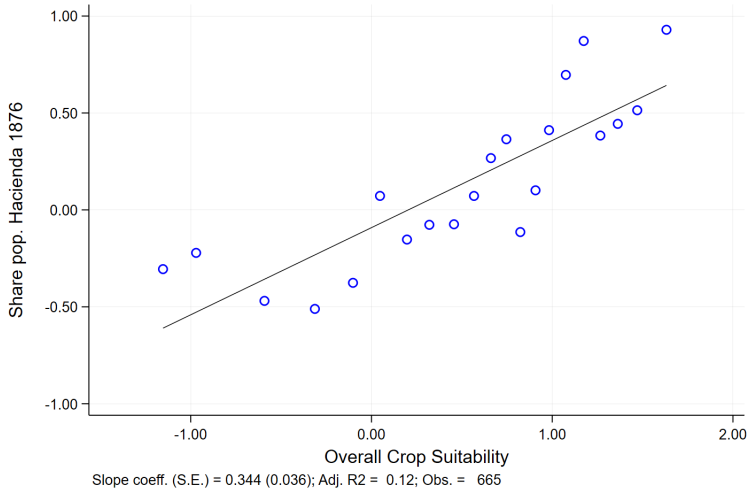
Figure 4: Soil Crop Suitability (instrument) and Hacienda Concentration

(a) Crop suitability

(b) Hacienda exposure



(c) First-stage



Notes: Panel A maps the average soil crop suitability by district, which measures the maximum standardized soil suitability between cotton (GAEZ) and other caloric-intensive crops that were available for cultivation (e.g., potato, maize) (Galor and Özak, 2016). Panel B maps the percentage of the district population residing in haciendas in 1876. Panel C displays a binned scatter plot showing the simple correlation between the two previous measures, with both variables standardized to have zero mean and unit standard deviation. The sample for all figures consists of 665 historical districts from the coastal and highland regions.

5.3. IV results

In this section, we present and discuss the results from the instrumental variables estimation. First-stage estimates, shown in Panel B of Table 1, reveal a strong and significant association between our instrument and historical hacienda concentration, even after accounting for a full set of geographical and demographic controls, as well as regional fixed effects. In addition, the Kleibergen-Paap F-statistics exceed the standard threshold level, mitigating the possibility of weak instrument bias (Stock and Yogo, 2005; Staiger and Stock, 1997).[13] As initial placement drivers at baseline include the location of former population settlements, labor availability (rural population, mestizo/indigenous) and proximity to mining centers, we include all of these as controls in all our regressions.

Table 1: OLS and IV Estimates for Literacy and Hacienda Concentration

Census Years:	1876	1940	1961	1981	1993	2007	2017
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: OLS — Literacy and Hacienda Concentration							
Dep. var.: literacy rate (share) at census year							
Share pop. *Hacienda* 1876	-0.019***	-0.026***	-0.020***	-0.011***	-0.007***	-0.004**	-0.002*
	(0.005)	(0.004)	(0.005)	(0.003)	(0.002)	(0.002)	(0.001)
Mean DV	0.126	0.334	0.481	0.679	0.755	0.858	0.869
Observations	665	665	665	613	665	665	665
R-squared	0.548	0.870	0.824	0.794	0.795	0.789	0.784
Panel B: IV First Stage — Hacienda Concentration and Crop Suitability							
Dep. var.: population in Haciendas in 1876 (share)							
Soil Crop Suitability	0.302***	0.274***	0.235***	0.202***	0.207***	0.215***	0.203***
	(0.062)	(0.065)	(0.064)	(0.069)	(0.066)	(0.064)	(0.064)
Kleibergen-Paap F statistic	23.57	17.89	13.69	8.48	9.81	11.31	10.16
Observations	665	665	665	613	665	665	665
R-squared	0.358	0.312	0.307	0.294	0.279	0.279	0.276
Panel C: IV Second Stage — Literacy and Hacienda Concentration							
Dep. var.: literacy rate (share) at census year							
Share pop. *Hacienda* 1876	-0.043**	-0.131***	-0.232***	-0.137***	-0.102***	-0.051***	-0.033**
	(0.020)	(0.035)	(0.065)	(0.052)	(0.034)	(0.017)	(0.013)
Upper A-R 95% CI	[-0.003	[-0.077	[-0.143	[-0.067	[-0.055	[-0.025	[-0.013
Lower A-R 95% CI	-0.089]	[-0.241]	-0.476]	-0.396]	-0.255]	-0.118]	-0.082]
Observations	665	665	665	613	665	665	665
R-squared	0.522	0.706	0.654	0.265	0.270	0.455	0.536

Notes: The table shows district-level OLS and IV-estimates for seven separate cross-sections. The dependent variable is the literacy rate at each census year, defined as the proportion of individuals aged 5 or older who could read, write, or both. Historical hacienda concentration is captured by the percentage of the district population residing in haciendas in 1876, standardized to have a mean of zero and a standard deviation of one. Hacienda concentration is instrumented using the soil crop suitability index. All regressions control for geographic covariates (elevation, slope, ruggedness, area, longitude, and latitude), contemporaneous demographic covariates (population, population density, urban share, ethnic diversity, arable share, mining, and working share), and department fixed effects. Robust standard errors are reported in parentheses and Anderson-Rubin confidence intervals in brackets. Coefficients significantly different from zero are denoted by: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[13] Complete first-stage covariate coefficients are detailed in Table A2 in the Online Appendix, along with correlations between the instrument and baseline control variables from 1876.

Second-stage estimates are presented in Panel C of Table 1. A one standard deviation increase in historical hacienda concentration is associated with a reduction in literacy rates of 13 percentage points in 1940, 23 percentage points in 1961, 10 percentage points in 1993, and only 4.4 percentage points in 2017. Given that literacy rates rose from 13% in 1876 to nearly 90% by 2017, these reductions, when evaluated at mean values, correspond to a decrease in literacy of 34–48% from 1876 to 1961, diminishing to 20% in 1981 and 4% in 2017. Consistent with our OLS estimates, this relationship intensifies through the mid-20th century before gradually diminishing since 1981. Notably, estimates are robust to an extended set of controls and fixed effects, treatment specifications, alternative clustering structures (e.g. 1876 district boundaries) and remain significant under Anderson-Rubin confidence intervals (Keane and Neal, 2024).

The local nature of IV estimates reflect an estimation over a subset of districts where the link between soil suitability and hacienda formation holds (our *compliers*). In this sense, the IV-OLS coefficient gap can be suggestive of the degree of selection among those compliers' characteristics or treatment levels. In our case, for instance, the IV coefficient for hacienda concentration is 2.3 times that of the OLS estimate in 1876, and 5 times larger in 1940.

A novel method to formally evaluate the components of variation in the IV-OLS gap is developed by Ishimaru (2024). The authors decompose the IV-OLS coefficient gap into components related to how much OLS is biased by uncontrolled confounders and how much the true effect varies across units (IV weights). By taking the same weights on covariates and treatment levels, the residual OLS-IV difference would support the IV correcting for an initial OLS endogeneity bias, rather than other forms of selection. The results from Ishimaru's decomposition procedure, detailed in Table A3 of the Online Appendix, indicate that our estimates for all census years are not driven by a selected set of compliers, specific levels of hacienda exposure, or an arbitrary subset of covariates. Instead, they support the presence of an initial downward bias in the OLS estimates.

5.4. Age-cohort results

A natural concern related to the previous estimation relates to the role of selective migration. In particular, the possibility that our estimates are mostly capturing more educated individuals leaving hacienda zones, regardless of changes in local characteristics induced by haciendas. To tackle this concern directly, individual-level microdata from modern censuses (1993, 2007, and 2017) enable us to exploit age-cohort groups based on their decade of birth to examine how hacienda concentration impacted literacy rates

across 11 cohort groups born between 1901–1910 and 2001–2010. Since these data provide information on each individual's district of birth, rather than just their district of residence, we can assign hacienda exposure based on the former.[14]

The analysis is based on eleven district-of-birth level regressions, estimated separately for each age-cohort group. Estimates for the older eight age-cohort groups (1901-1910 to 1971-1980) are compiled from the 1993 census, while estimates for 1981-1990 age-cohort come from the 2007 census, and those for the 1991-1999 and 2000-2010 cohorts are based on the 2017 census. The OLS specification can be expressed as follows:

$$y_{i,c} = \alpha + \beta_1 \text{Hac}_{i,1876} + G_i'\phi_1 + C_i'\vartheta_1 + \delta_d + \varepsilon_{i,c} \quad (2)$$
$$\forall c \in \{1901-1910, 1911-1920, \ldots, 2001-2010\}$$

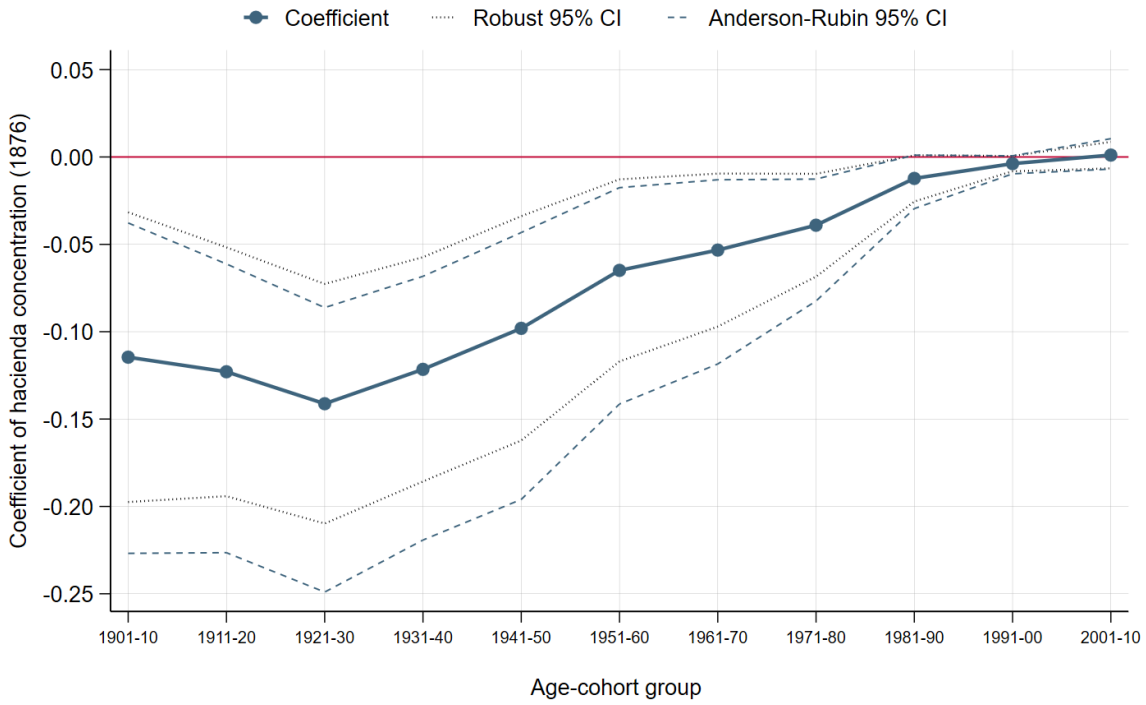
The dependent variable $y_{i,c}$ represents the literacy rate of age-cohort group c born in district i. Explanatory variables include $\text{Hac}_{i,1876}$, the district's hacienda concentration in 1876; district-level geographic and topographic controls G_i; and cohort characteristics (C_i), such as age-cohort sample size, mother tongue, and gender. Department fixed effects (δ_d) are also included. The model is estimated for each age-cohort group using both OLS and IV estimators, with soil crop suitability serving as an instrument for hacienda concentration in 1876[15].

The coefficients of interest, β_1, capturing the effect of historical hacienda concentration on literacy for each age-cohort group, are depicted in Figure 5. These correspond to the IV estimates, shown with both robust and Anderson-Rubin 95% confidence intervals. The results show that, even after accounting for the possibility of selective migration, the hampering effects of hacienda exposure on literacy holds. Stronger effects are found among older cohorts born before 1950, which gradually diminish for cohorts born between 1951 and 1980 and become non-significant for cohorts born after 1980. This pattern aligns with the near-universal literacy achieved among younger cohorts, with literacy rates reaching 98% for those born in the 1980s and later. In terms of magnitude, a one standard deviation increase in hacienda concentration in 1876 is associated with a 10-15% decrease in literacy rates for cohorts born in the first half of the 20th century, a 4-7% decrease for those born between 1951 and 1980, and an effect

[14] We assume that the district of birth serves as a reasonable proxy for the district of education. According to the 1993 census, 85 percent of children aged 8–10 still resided in their district of birth. Assuming that age 8–10 roughly marks the typical age for beginning school and acquiring literacy, this percentage provides a conservative estimate for the proportion of people whose district of birth corresponds to their district of education. It is likely that this proportion was even higher in the early to mid-20th century, a period characterized by low migration rates (Dell (2010), p. 1876). The assumption that district of birth is a good proxy for the district of education aligns with the approach used by Duflo (2001).

[15] Detailed results are available in Table A10 in the Online Appendix.

Figure 5: Age-Cohort Estimates of the Effect of Historical Hacienda Concentration on Literacy



Notes: The figure displays the IV estimates of the effect of district-of-birth historical hacienda concentration on literacy across age-cohort groups surveyed in the 1993, 2007 and 2017 population censuses. The coefficients were estimated in separate district of birth-level regressions, each using one of the eleven different age-cohort samples. Estimates for age-cohort groups 1901-1910 to 1971-1980 were derived from the 1993 census, estimates for 1981-1990 age-cohort from the 2007 census, and estimates for age-cohort groups 1991-1999 and 2000-2010 from the 2017 census. The dependent variable is the literacy rate of each age-cohort group. The hacienda variable represents the percentage of the district of birth population residing in haciendas in 1876, standardized to have a mean of zero and a standard deviation of one. Hacienda concentration is instrumented using the soil crop suitability index. All regressions control for district of birth geographic covariates (elevation, slope, ruggedness, area, longitude, and latitude), age-cohort group characteristics (age-cohort sample size, mother tongue, and gender), and department fixed effects. Robust 95% confidence intervals are depicted with black dotted lines, while Anderson-Rubin 95% confidence intervals are shown with blue short-dashed lines. The Kleibergen-Paap F-statistics range from 13.7 to 18.1 across all age-cohort group regressions.

approaching zero for cohorts born after 1980.

5.5. Broader effects on education beyond literacy

Thus far, our analysis has focused on the long-term negative effects of historical hacienda concentration on literacy. While literacy provides a consistent and comparable indicator over time, especially during the early phases of educational expansion, its informative power diminishes as literacy rates approach universal levels. To better understand the broader legacy of hacienda systems, we now extend our analysis to additional educational outcomes: (i) school enrollment among children aged 6 to 12,

and (ii) average years of schooling among young adults aged 18 to 30. These outcomes capture both early-life access to education and cumulative educational attainment, offering a broader picture of the long-term human capital implications of haciendas.

Table 2: Long-term effect of historical hacienda concentration on other educational outcomes

Dependent variable:	**Share of households with children not attending school**			**Mean years of education, adults 18-30**		
	1993	2007	2017	1993	2007	2017
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: OLS — Other Educational Outcomes and Hacienda Concentration						
Share pop. *Hacienda* 1876	0.006***	0.003***	0.001	-0.045**	-0.074**	-0.078***
	(0.002)	(0.001)	(0.001)	(0.035)	(0.033)	(0.026)
Observations	665	665	665	665	665	665
R-squared	0.551	0.492	0.343	0.803	0.818	0.801
Mean DV	0.14	0.05	0.02	6.7	9.0	10.4
Panel B: IV Second Stage — Other Educational Outcomes and Hacienda Concentration						
Share pop. *Hacienda* 1876	0.061**	0.016*	-0.003	-1.011**	-0.820**	-0.299
	(0.026)	(0.009)	(0.005)	(0.446)	(0.361)	(0.255)
Upper A-R 95% CI	[0.165]	[0.046]	[0.006]	[-0.342]	[-0.256]	[0.201]
Lower A-R 95% CI	[0.021]	[0.001]	[-0.018]	[-2.830]	[-2.131]	[-1.079]
Kleibergen-Paap F statistic	9.81	11.31	10.16	9.81	11.31	10.16
Observations	665	665	665	665	665	665
R-squared (2nd stage)	0.103	0.339	0.263	0.550	0.678	0.778

Notes: The dependent variable in column 1 to 3 is the share of households with children not attending school at census years 1993, 2007, and 2017, respectively. The dependent variable in columns 4 to 6 is the mean years of education of adults aged 18 to 30 at census years 1993, 2007, and 2017, respectively. Historical hacienda concentration is captured by the percentage of the district population residing in haciendas in 1876, standardized to have a mean of zero and a standard deviation of one. Hacienda concentration is instrumented using the soil crop suitability index. All regressions control for geographic covariates (elevation, slope, ruggedness, area, longitude, and latitude), contemporaneous demographic covariates (population, population density, urban share, ethnicity share, agricultural, mining, and manufacturing share), and department fixed effects. Robust standard errors are reported in parentheses and Anderson-Rubin confidence intervals in brackets. Coefficients significantly different from zero are denoted by: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2 presents the results from both OLS and IV specifications using data from the 1993, 2007, and 2017 population censuses. The evidence indicates that the negative effects of hacienda concentration extended beyond literacy, affecting school enrollment and educational attainment well into the 21st century. IV estimates in columns (1) to (3) show that one standard deviation increase in the share of resident population in

haciendas in 1876 is associated with a 6.1% increase in the share of households with children not attending school in 1993 and a 1.6% increase in 2007. These represent substantial increases over the mean values (44% and 32%, respectively). The effect fades by 2017, suggesting some degree of convergence in basic educational access. Similarly, the impact on average years of schooling among young adults shows a decline of 1 year in 1993 and 0.8 years in 2007, with no significant difference remaining by 2017.

5.6. Robustness checks

This section addresses potential threats to identification of the negative effect of historical hacienda exposure on education. Previous sections presented evidence supporting the instrument's relevance, provided a discussion on the rationale of the exclusion restriction, and showed the robustness of our results under alternative specifications, fixed effects and covariates. Notably, evidence against selective migration was also provided exploiting age-cohort estimates (i.e. assigning historical hacienda exposure at a cohort-level based on the district of birth).

Clustering at 1876 (baseline) district boundaries

An additional consideration we have not yet addressed involves changes in district administrative boundaries over time. It is possible that the creation of new districts was partially endogenous to the hacienda economy, potentially influencing our measure of hacienda concentration given our unit of analysis. The results presented in Table 1 follow a conservative approach with respect to this issue by aggregating variables to a common set of districts as defined at the start of the study period. Alternatively, one could use the current district boundaries at the time of each census period while clustering standard errors at the 1876 district level, as our primary independent variable—historical hacienda concentration—is based on these boundaries. As shown in Table A8 in the Online Appendix, this alternative approach yields results that are consistent and closely aligned with those in Table 1. Moreover, the larger number of districts in 1940 and subsequent periods results in even higher F-statistics, further reinforcing the robustness of our findings.

Measurement error in 1876 exposure to haciendas

Another concern involves potential measurement error in our historical hacienda concentration variable, as the 1876 census did not utilize modern data collection techniques. To address this, we conduct two additional tests to assess the reliability of this measure. First, we examine whether our findings hold when using more recent

measures of hacienda presence, available for 1940 and 1961. Figure A3 in the Online Appendix demonstrates that hacienda concentration in both 1940 and 1961 is also significantly associated with lower literacy rates in the short and long term, with magnitudes remarkably similar to those obtained using the 1876 hacienda measure. These results suggest that changes in hacienda concentration over time played a minimal role in shaping human capital outcomes; rather, it was the early establishment and persistence of longstanding hacienda elites that were most influential.

In addition, we test the robustness of our findings to alternative definitions of hacienda concentration. Table A9 in the Online Appendix explores five alternative approaches for measuring hacienda concentration: (1) the share of a district's rural population residing in haciendas in 1876; (2) a dummy variable indicating whether a district had at least one hacienda in 1876; (3) the log of the population residing in haciendas; (4) the log of the number of haciendas in the district; and (5) the number of haciendas per 100 people in 1876. The results remain consistent across these alternative measures, showing similar negative impacts, though with weaker statistical significance due to lower explanatory power. Our chosen measure, based on the share of a district's total population residing in haciendas, yields the highest F-statistic and R-squared, underscoring its appropriateness and explanatory strength.

6. Mechanisms: local supply and demand for education in haciendas

What accounts for the persistent negative relationship between hacienda concentration and educational outcomes in Peru? A large body of theoretical and empirical work has linked land concentration and labor coercion with low educational development, primarily through two channels. On the supply side, it has been argued that elite landowners—faced with low complementarity between land and human capital—had strong incentives to block mass education, which would raise rural wages and enable labor mobility (Galor et al., 2009; Acemoglu and Wolitzky, 2011). This through direct investments—or lack thereof— in their areas of influence (e.g. school construction, maintenance, teacher recruitment) or by influencing taxation, education laws or public spending.

On the demand-side, its possible that underinvestments in human capital may have also stemmed from structural disincentives and constraints faced by rural households living in haciendas. Specific mechanisms include direct labor coercion, which limited the time and autonomy needed for schooling (Cinnirella and Hornung, 2016); restricted outside options, which made the returns to education appear limited (Ashraf

et al., 2024); and broader patterns of agricultural specialization, which could reinforce dependence on low-skill labor.

Using a set of proxy variables for both supply and demand factors, we investigate these mechanisms and provide evidence consistent with a supply-side explanation. We show that hacienda-dominated districts had lower public investments in education—without comparable reductions in other public goods—and negligible impacts on demand-side proxies, such as local economic activity, agricultural specialization, or broader development outcomes. Crucially, we document how the local power of landed elites—measured through proxies of hacienda consolidation—combined with their political incentives to limit the literacy of indigenous workers, were pivotal at explaining the deliberate underprovision of schooling within hacienda zones.

6.1. Supply-side factors

6.1.1. Educational investments: If hacienda owners had both the incentives and the local capacity to restrict educational opportunities, a first order effect of this control would be to restrict local investments in schooling within their areas of influence. This is precisely what we observe when analyzing the impacts of hacienda concentration on the number of schools and teachers (standardized by population) since 1918[16]. As shown in Table 3, both OLS and IV estimates reveal large and significant negative impacts on the supply of schooling—fewer schools and lower teacher density—in districts with a higher share of resident population in haciendas at least before 1950.

Until 1930, the magnitudes of the IV coefficients are substantial: a one standard deviation increase in hacienda exposure reduced the number of schools per capita by roughly eight schools, representing over a 50% decline relative to the mean. Similar patterns emerge for teacher density. By 1950, this impact starts fading, likely reflecting the onset of increased rural education spending and literacy campaigns by the central government starting that period.

[16]We employ province-level data from 1918, 1930, and 1950, sourced from the Statistics Department of Peru, as district-level data was not available. To account for this, we conducted the regressions at the district level and clustered the standard errors at the province level. There were 100 provinces in 1918, 101 provinces in 1930, and 114 provinces in 1950.

Table 3: Long-term effect of hacienda concentration on the supply of schooling

Dependent variable:	**Number of schools per population**			**Number of teachers per population**		
	1918	1930	1950	1918	1930	1950
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: OLS — Schools and Hacienda Concentration						
Share pop. *Hacienda 1876*	-1.365***	-1.569***	-1.576***	-1.423***	-1.402**	-2.252***
	(0.413)	(0.521)	(0.464)	(0.484)	(0.564)	(0.722)
Observations	665	665	665	665	665	665
R-squared	0.629	0.472	0.557	0.607	0.641	0.558
Mean DV	13.03	14.67	18.83	15.21	20.96	33.58
Panel B: IV Second Stage — Schools and Hacienda Concentration						
Share pop. *Hacienda 1876*	-7.895**	-8.372**	0.043	-8.337*	-9.525*	-2.529
	(3.570)	(3.787)	(2.678)	(4.331)	(5.290)	(5.197)
Upper A-R 95% CI	[-1.672	[-1.877	[6.104	[-0.602	[-0.206	[8.761
Lower A-R 95% CI	-18.049]	-19.326]	-5.889]	-20.364]	-24.422]	-14.524]
Kleibergen-Paap F statistic	16.09	16.07	15.52	16.09	16.07	15.52
Clusters (provinces)	100	101	114	100	101	114
Observations	665	665	665	665	665	665
R-squared	0.313	0.265	0.539	0.335	0.42	0.558

Notes: The dependent variable in columns 1 to 3 is the number of schools in the province per 10,000 inhabitants, and in columns 4 to 6 the number of teachers in the province per 10,000 inhabitants. The data sources are the 1918, 1930, and 1950 Statistical Abstract of Peru (Statistics Department). Historical hacienda concentration is captured by the percentage of the district population residing in haciendas in 1876, standardized to have a mean of zero and a standard deviation of one. Hacienda concentration is instrumented using the soil crop suitability index. All regressions control for geographic covariates (elevation, slope, ruggedness, area, longitude, and latitude), 1940 demographic covariates (population, population density, urban share, ethnicity share, agricultural, mining, and manufacturing share), and department fixed effects. Cluster standard errors at the province level are reported in parentheses and Anderson-Rubin confidence intervals in brackets. Coefficients significantly different from zero are denoted by: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

6.1.2. Public goods and property rights: Were these constrained local investments unique to education? It has been argued that early land inequality has the potential to encourage public goods, secure property rights and promote market integration (e.g. Acemoglu et al. (2008); Dell (2010)). These outcomes may arise when investments that enhance productivity and secure existing rents align with the interests of landed elites. Such investments can have broader implications for long-run development trajectories. As documented earlier, the case of the local provision of education within haciendas goes in the opposite direction. However, even in this scenario, it remains possible that

elites prioritized other types of public goods that were more closely aligned with their economic interests. To explore this possibility, we examine the long-term impact of hacienda concentration on a broader set of public goods since 1950, including road infrastructure, land titling, and irrigation.[17]

Table 4: Historical hacienda concentration and the provision of other public goods

Dependent variable:	**Roads** *Log distance to the national road network*			**Property rights** *Registered land titled 1994*		**Irrigation** *sh. farmland w/ irrigation*
	1950	1975	2002	%Farmers	%Farmland	1961
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: OLS — Other Public Goods and Hacienda Concentration						
Share pop. *Hacienda* 1876	-0.029	-0.159	-0.038	0.002	0.019*	0.002
	(0.112)	(0.121)	(0.149)	(0.007)	(0.011)	(0.005)
Observations	665	665	665	665	665	665
R-squared	0.691	0.681	0.47	0.363	0.187	0.617
Mean DV	10.062	8.133	5.727	0.25	0.46	0.119
Panel B: IV Second Stage — Other Public Goods and Hacienda Concentration						
Share pop. *Hacienda* 1876	0.837	1.097	1.080	0.087	0.102	-0.091
	(0.891)	(1.014)	(1.464)	(0.083)	(0.121)	(0.100)
Upper A-R 95% CI	[2.963	[3.951	[5.27	[0.352	[0.466	[0.111
Lower A-R 95% CI	-1.119]	-0.793]	-1.825]	-0.066]	-0.136]	-0.360]
Kleibergen-Paap F statistic	13.69	13.69	15.52	10.41	10.41	12.57
Observations	665	665	665	665	665	664
R-squared	0.654	0.631	0.431	0.224	0.112	0.392

Notes: The dependent variable in columns 1 to 3 is the logged distance, in meters, from the district to the nearest paved road in the national road network. In columns 4 and 5, the dependent variables are the share of farmers and farmland with registered land titles (Ag. Census 1994). The dependent variable in column 6 is the percentage of agricultural land in the province with some form of irrigation (Ag. Census 1961). Historical hacienda concentration is captured by the percentage of the district population residing in haciendas in 1876, standardized to have a mean of zero and a standard deviation of one. Hacienda concentration is instrumented using the soil crop suitability index. All regressions control for geographic covariates (elevation, slope, ruggedness, area, longitude, and latitude), contemporaneous demographic covariates (population, population density, urban share, ethnicity share, agricultural, mining, and manufacturing share), and department fixed effects. Robust standard errors are reported in parentheses and Anderson-Rubin confidence intervals in brackets. Coefficients significantly different from zero are denoted by: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4 presents the OLS and IV results for these outcomes. As shown, districts with a higher historical share of resident population in haciendas exhibit no statistically significant differences in the provision of any of the public goods analyzed –including connectivity to the national road network,[18] the coverage of registered land titles, and

[17]For roads, we digitized Peru's national road network maps from three periods in the latter half of the 20th century and calculated the logged distance (in meters) from each district's centroid to the nearest paved road in the national network for 1950, 1975, and 2002. These maps were sourced from the International Petroleum Co., Ltd. (Lima, 1950), Touring y Automóvil Club del Perú (Lima, 1975), and the Ministry of Transport (Lima, 2002). For property rights, we examined the share of farmers and farmland with a registered land title recorded in the 1994 Agricultural Census. Finally, for irrigation, we analyse the share of agricultural land equipped with some form of irrigation, based on data from the 1961 First National Agricultural Census (DNEC, 1961).

[18]Additionally, we tested road network density measures (i.e., total paved road length in each district

the availability of irrigation infrastructure. A result that holds even after accounting for a rich set of geographic, demographic and economic controls, as well as regional fixed effects.

How can we account for the lack of significant differences in the accumulation of public goods in former hacienda zones? A plausible explanation lies in the more isolated and enclave-like nature of the Peruvian hacienda economy, particularly in the highlands. For instance, unlike better integrated regions such as Cundinamarca (Colombia), at the heart of the Spanish colonial system, where landed elites invested in public services (Acemoglu et al., 2008), or in central Mexico where hacienda elites operated as centers linking rural economic activity to the main colonial economic activities (Arias and Flores-Peregrina, 2024), with the exemption of the Northern Coast, Peruvian haciendas often operated with limited integration into broader commercial or state networks, reducing the strategic value of public goods investment beyond the immediate needs of estate production.

Taken together, these results suggest that the long-term development impacts of haciendas seem to be dependent on local conditions shaping the incentives and capacities different elites had at exerting specific types of investments. This emphasizes the importance of regional conditions in explaining the plausibly heterogeneous legacies of haciendas.

a) A regional comparison case: the Southern Highlands

As noted earlier, research on haciendas in Peru has relied primarily on descriptive case studies, with two prominent regional clusters receiving attention: the Northern Coast and the Southern Highlands. While not the central focus of her study, an important exception is Dell (2010), who analyzes the Southern Highlands to examine the long-term development impacts of the *Mita* system—a colonial forced labor institution imposed by the Spanish Crown around 1600. Using a geographic regression discontinuity design across districts straddling the historical Mita boundary, the author finds persistent negative impacts on a range of outcomes, including household consumption, child stunting, literacy, educational attainment, road infrastructure, agricultural specialization, and market integration. Although the role of haciendas is not directly tested, the author offers an alternative interpretation of their historical significance: areas subjected to the Mita would have failed to develop land concentration in the form of haciendas, which in turn inhibited the establishment of stable land-tenure systems that could encourage public-goods provision.

divided by the district's surface area in square kilometers) and alternative road types (e.g., non-paved, dirty roads) and found a similarly negligible impacts.

With this in mind as a comparison case, we replicate our IV design on the subset of Southern Highlands districts studied by Dell (2010) and obtain two key findings. First, although districts subjected to the Mita exhibited lower hacienda population by 1876[19]—as expected given the different suitability of agricultural versus mining areas—we find no statistically significant direct effect of hacienda concentration on contemporaneous literacy rates. These null effects do not appear to stem from limited statistical power, as shown in Table A19 and Figure A7. Second, within the same sample, hacienda concentration has null or negative direct impacts on investment in public goods such as schools and roads (Table A20). These results support a conditional version of Dell (2010)'s conjecture on the mediating role of the formation of haciendas over broader development outcomes: even in mita-treated districts that observed a later formation of haciendas, landed elites incentives appear to have undermined development-enhancing investments in education and infrastructure.[20]

6.2. Demand-side factors

Even if hacienda elites had not directly restricted the long-run supply of education, persistently lower educational outcomes in hacienda-dominated districts could still reflect underinvestment in human capital by households——particularly if local returns to schooling were perceived as low. To test this demand-side channel, we compile an original set of district-level measures of economic conditions dating back to 1961 (the earliest available at this level of disaggregation) and test them under our IV design.

The results are presented in Figure 6, with two panels grouping proxies of economic well-being (A) and structural change (B). Across all outcomes, no significant effects are found as the result of a higher resident population in haciendas. As Panel A shows, our OLS-IV estimates reveal null effects on household well-being indicators—income per capita, asset endowments, and child labor use. Likewise, Panel B shows no significant shifts in sectoral employment (agriculture, manufacturing, services) or urbanization rates well into the second half of the 20th century, mitigating differential patterns of structural transformation in districts historically exposed to haciendas[21]. We further corroborate these null effects using more recent and granular indicators of economic

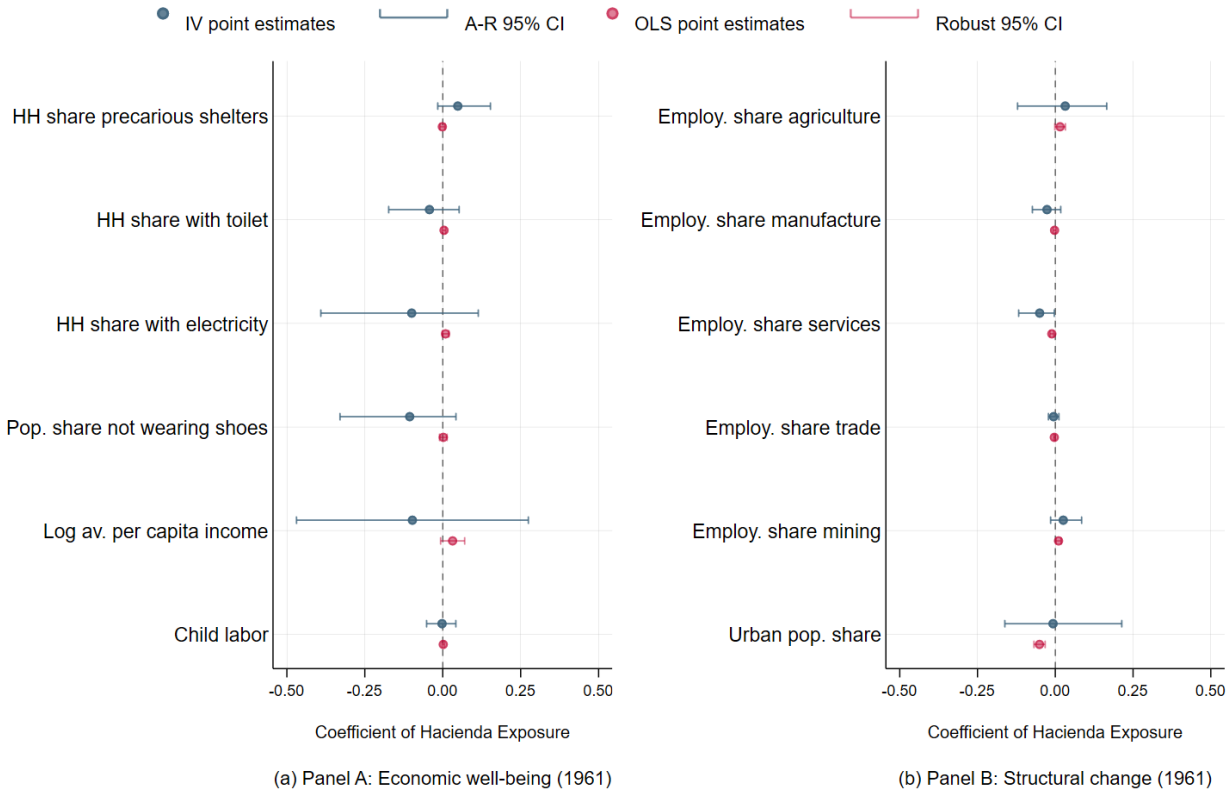
[19] As shown in the maps of Figure A6 and the conditional negative correlation of Table A18.

[20] Moreover, as recently discussed by Abad and Maurer (2025), further replications of Dell's study using the same RDD with alternative samples, measures of exposure and outcomes measured at different points in time suggest that the long-term legacies of the mita system would be historically contingent, spatially heterogeneous, and could be underplaying the importance of republican-era processes linked to land concentration after the mita was abolished.

[21] These findings align with Galor et al. (2009)'s theoretical model, which predicts that land inequality delays both the emergence of human-capital–promoting institutions (such as public schools) and the structural transition from agriculture to industry.

well-being such as nightlights, poverty rates, or early childhood development proxies (Table A13). Again, none show consistent evidence of enhanced or suppressed local economic demand that could rationalize persistent educational deficits in hacienda regions.[22]

Figure 6: OLS-IV estimates - Economic well-being and structural change



(a) Panel A: Economic well-being (1961)

(b) Panel B: Structural change (1961)

Notes: The figure presents OLS and IV estimates of the association between historical hacienda concentration and indicators of economic well-being (Panel A) and structural transformation (Panel B) measured in 1961 (or 1981 for income and child labor). Each outcome point estimate and corresponding confidence interval was estimated in a separate regression. Panel A includes the following outcomes at the province level in 1961: the share of households living in precarious shelters (chozas), the share with access to a toilet, the share with electricity, and the share of individuals not wearing shoes. It also includes, at the district level in 1981, the logged average per capita income and the share of individuals aged 6 to 17 engaged in paid work. Panel B reports results for the district-level employment share in agriculture, manufacturing, services, trade, and mining in 1961, as well as the urban population share. The main explanatory variable—hacienda concentration—measures the standardized share of the district population living in haciendas in 1876. This variable is instrumented using the soil crop suitability index. All regressions control for geographic characteristics, demographic and economic variables from 1961 or 1981, and department fixed effects. Robust standard errors are depicted with red lines and Anderson–Rubin confidence intervals with blue lines.

Perhaps more surprisingly, historical exposure to haciendas does not appear to have

[22]Complementary evidence consistent with a supply-side underprovision of education in hacienda-dominated areas comes from the analysis of gender-disaggregated literacy impacts. If the observed effects were driven primarily by household demand for education, stronger negative impacts would be expected among males, who were more economically valuable in agricultural labor. Results in Table **??** and Figure A4 show the opposite: hacienda exposure reduced literacy rates for both men and women to a similar degree, with no statistically significant differences from 1940 to 2017. This pattern further supports the interpretation that the observed deficits stemmed primarily from constraints on educational supply rather than differential demand.

translated into lasting specialization or productivity gains in agriculture. Table A14 presents results using data from the 1994 Agricultural Census. We find no significant effects of hacienda concentration on average farm sizes, land inequality, or the adoption of agricultural technologies linked to electricity, fertilizer or insecticide use. If anything, consistent with our previous result, the only robust negative effect appears in the years of education attained by farmers. Taken together, while regional heterogeneity may partly account for these patterns, there is little evidence to support the notion that former hacienda zones became more dynamic or specialized agricultural regions—let alone ones in which low schooling investments reflected a lower demand for education as a rational adaptation to labor market needs.

6.3. Mechanisms running through the local power of landed elites?

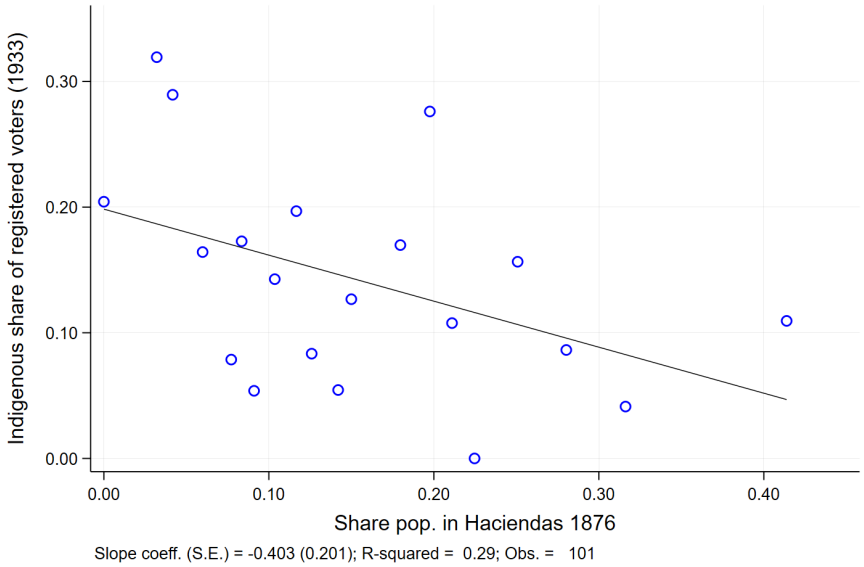
Throughout the text, we have implicitly justified a deliberate restriction of the supply of education as a key mechanism at play based on the local power of landed elites enabling them to do so. While testing this directly is hard, we will provide evidence consistent with high levels of political capture in hacienda-dominated districts, as well as proportionally stronger impacts in hacienda zones with better consolidated and less fragmented landed elites.

6.3.1. Political incentives: As discussed in Section 3, the restriction of voting rights to literate citizens—specifically those literate in Spanish—until 1980 created strong incentives for local elites to limit educational access as a means of political control. Historical evidence suggests that in hacienda-dominated regions, this exclusionary strategy was not incidental but deliberate (e.g. Cotler (1976); McClintock (1981)). To test this hypothesis we compile province-level shares of registered voters by ethnicity from the 1933 Electoral Census, the earliest available at this level of disaggregation. Both as a conditional correlation (Figure 7) and under our IV design (Table A15), we show that provinces with a higher share of resident population in haciendas exhibit a significantly lower proportion of indigenous population registered as voters in 1933 — precisely those mostly illiterate in Spanish and predominant in the labor force of haciendas.

The result reinforce our broader argument: hacienda-dominated areas not only received a restricted supply of education, but this restriction translated very effectively in the political exclusion of indigenous workers. This result is consistent with theoretical models of local political capture linked to coerced labor and non secret voting (e.g. Baland and Robinson (2008)), but also with evidence for the Peruvian case showing that only once the 1969 land reform *de facto* dismantled the hacienda system, did both

political participation and competition in rural areas significantly improve (Paredes, 2023).

Figure 7: Hacienda exposure and indigenous registered voters



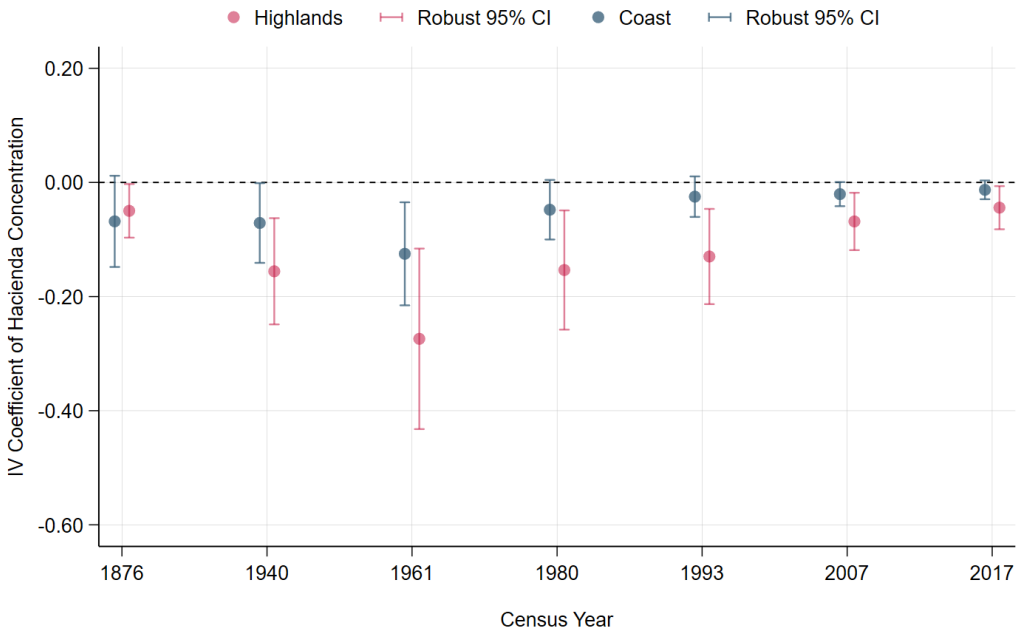
Notes: The figure displays a binned scatter plot illustrating the relationship between the share of indigenous registered voters in 1933 and the share of the resident population living in haciendas in 1876 at the province level. The analysis controls for geographic covariates (slope, ruggedness, and area) and 1940 demographic characteristics (logged population, indigenous share, and agricultural share). The x-axis variable is divided into 20 equal-sized bins; within each bin, the mean values of both the x-axis and y-axis variables are computed and plotted to visualize the underlying relationship.

6.3.2. Economic consolidation: We have argued that both the incentives and the capacity of landed elites to restrict educational access were key enabling factors behind our main findings. If this interpretation is correct, then the negative impact of hacienda concentration on educational outcomes should intensify with the economic strength of landed elites—that is, their ability to exert effective local control. The relationship is not straightforward. As shown by Ashraf et al. (2024), for example, the consolidation of landed estates and the accumulation of elite-owned physical capital can, under certain conditions, increase the profitability of labor emancipation rather than coercion. This shift occurs as the costs of monitoring rise in the context of industrialization, thereby weakening the incentive to suppress education over time.

To test this hypothesis, we rely on two main set of results under our IV design. First, we interact our measure of hacienda concentration with regional dummies for Coastal and Highland districts across all available census years (Figure 8). The estimates show that, particularly prior to the dismantling of the hacienda system in the

1970s, the negative effect of hacienda concentration on educational outcomes is significantly more pronounced in the Highlands. As noted earlier, three main factors help explain this regional disparity. First, greater geographic isolation and weaker integration into commercial networks likely led highland haciendas to rely more heavily on coercive labor regimes. Second, a disproportionally large share of indigenous population, which increased the strategic value of limiting access to education—especially under the literacy-based voting requirements. Third, lower exposure to international markets reduced modernization pressures, such as the shift toward wage labor or incentives to invest in public goods.

Figure 8: Estimates of the Effect of Historical Hacienda Concentration on Literacy by Natural Region



Notes: The figure displays OLS and IV estimates for seven cross-sectional regressions on the association between hacienda concentration and literacy rates by natural region (i.e., Coast and Highlands). The dependent variable is the literacy rate at each census period, and the main explanatory variables are interaction terms between our measure of hacienda concentration and dummy variables indicating whether the district is located in the Coastal or Highland region. Hacienda concentration represents the percentage of the district population residing in haciendas in 1876, standardized to have a mean of zero and a standard deviation of one. Hacienda concentration is instrumented using the soil crop suitability index and its interaction with the natural region binary indicator. All regressions control for geographic covariates (elevation, slope, ruggedness, area, longitude, and latitude), contemporaneous demographic covariates (population, population density, urban share, ethnicity share, agricultural, mining, and manufacturing share), and department fixed effects. Robust 95% confidence intervals are illustrated with blue and red lines corresponding to coastal and highlands districts, respectively.

Next, we examine whether the degree of hacienda consolidation within a district shaped the severity of long-run educational impacts. To do so, Table A17 presents OLS and IV estimates interacting our measure of historical hacienda exposure with an in-

dex of hacienda fragmentation—constructed as the inverse of a Herfindahl-Hirschman Index (HHI), where the shares correspond to the proportion of the district population residing in each hacienda. Higher values of this index indicate greater fragmentation, plausibly reflecting a more dispersed distribution of economic power among hacienda owners. Consistent with the idea that intra-elite competition mitigated their capacity to exert local control, we find that more fragmented hacienda structures significantly attenuate the negative effect of hacienda concentration on literacy. These results suggest that in districts where no single elite actor held dominant control, the ability to suppress educational investment or limit worker mobility was weakened—potentially due to reduced economies of scale in coercion or increased competition for labor.

7. Mitigating factors on the persistence of hacienda exposure?

The results presented thus far emphasize the hindering educational legacies of haciendas, driven by the incentives and capacities of landed elites to restrict local investments in schooling. Yet, institutional legacies rarely persist in isolation. Over the course of the 20th century, major structural transformations—such as urbanization, industrialization, and the expansion of the modern state—interacted with preexisting patterns of elite dominance shaped their long-term impact on human capital formation.[23] In particular, shifts from agriculture to industry, the development of urban labor markets, and state-led policy reforms may have progressively eroded the bargaining power of traditional landed elites by increasing labor mobility and expanding workers' outside options (Acemoglu and Wolitzky, 2011; Ashraf et al., 2024; Mares, 2015).

In this section, we examine three key mitigating factors that may have attenuated the detrimental legacy of historical hacienda concentration on literacy. First, the rise of competing non-agrarian elites—particularly those in urban, mining, or textile sectors—who often had stronger incentives to support educational expansion due to their reliance on skilled labor (Ashraf et al., 2024; Goñi, 2023). Second, structural shifts in labor markets, including growing labor scarcity, declining mobility costs, and expanding non-agricultural employment opportunities, which likely undermined the effectiveness of coercive labor regimes and may have compelled hacienda elites to ease re-

[23] A growing body of work in the persistence literature suggests that the long-term effects of coercive colonial institutions are contingent on subsequent structural changes, elite adaptation, and shifts in the broader opportunity structure (Abad and Maurer, 2025; Bobonis and Morrow, 2014; Frankema, 2009b). In Latin America, this interplay was especially complex: while entrenched inequality and limited institutional reform constrained inclusive growth (Coatsworth, 2008; Mariscal and Sokoloff, 2000), the regional wave of redistributive land reforms, the decline of coercive labor systems, and the emergence of competing economic interests sometimes eroded elite control or created incentives for broader educational investment.

strictions on education to retain or attract workers (Abad and Maurer, 2025; Méndez and Van Patten, 2022; Mares, 2015). Third, the 1970s land reform—a top-down policy implemented under military rule—which dismantled large estates and transferred land to organized peasant communities[24] (Albertus et al., 2020).

For the first two set of factors, we conducted a supplementary analysis exploring its interaction with our measure of hacienda concentration. We constructed five dummy variables indicating the presence of urban, manufacturing, and mining elites, districts with high labor scarcity, and those with availability of outside options for workers.[25] We estimated five OLS models, one for each mitigating factor, by modifying equation (1) to include an interaction term between the hacienda measure and the dummy variable for each mitigating factor:

$$y_{i,1940} = \alpha + \beta_1 \text{Hac}_{i,1876} + \pi_1 \big(\text{Hac}_{i,1876} \times MF_{i,m}\big) + G_i'\phi_1 + X_{i,1940}'\gamma_1 + \delta_r + \varepsilon_i \qquad (3)$$
$$\forall m \in \{\text{urban elite, textile elite, mining elite, labor scarcity, outside option}\}$$

The dependent variable $y_{i,1940}$ is the literacy rate in each district i in 1940, allowing us to examine these relationships during the highest point of hacienda dominance. The term $\text{Hac}_{i,1876}$ represents the historical concentration of hacienda in 1876, and $\text{Hac}_{i,1876} \times MF_{i,m}$ captures the interaction between hacienda exposure and a dummy variable for each mitigating factor m. The remaining terms include geographic and demographic control variables, department fixed effects, and a random error term. In this framework, the coefficients of interest, β_1, captures the baseline impact of hacienda concentration on literacy, while π_1 measures how this effect is moderated by each mitigating factor. To address endogeneity concerns, we extended our analysis using the IV framework. In this setup, both the hacienda measure and its interaction with the mitigating factor are treated as endogenous, and instrumented using the crop suitability index and its interaction with the mitigating factor.

Table 5 presents the results from both the OLS (Panel A) and IV (Panel B) specifications. Columns (1) to (3) explore whether the presence of competing elites mitigated the negative hacienda effect. These elites represent key economic sectors in early 20th-

[24]Peru's land reform was a top-down, collectivist redistribution implemented under military rule between 1969 and 1980. It was one of the most far-reaching reforms in Latin America, affecting roughly half of all agricultural land (i.e., 10 million of ha.) and eliminating nearly every hacienda by 1985.

[25]Districts associated with urban elites are defined as those located within 50 km of the nearest city with a population exceeding 20,000 in 1940. Manufacture industry districts are identified where the labor force share in manufacturing was above the 75th percentile in 1940, and mining industry districts are those where the mining labor force share was above the 75th percentile in 1940. High labor scarcity districts are defined as areas with a worker density below two workers per square km in 1940. Finally, districts offering outside opportunities are identified as those situated within 10 km of the nearest paved road in the national road network as of 1950.

Table 5: OLS and IV Estimates for Literacy and Historical Hacienda Concentration

Mitigating factors:	**Competing elites**			**Labor supply**	**Outside options**
	Urban elite (cities)	Manufacture industry elite	Mining industry elite	High labor scarcity	Close to national road netw.
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS — Literacy and Historical Hacienda Concentration					
Dep. var.: literacy rate (share) in 1940					
Share pop. *Hacienda* 1876	-0.027***	-0.029***	-0.026***	-0.021***	-0.024***
	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)
Hacienda x mitigating factor	0.007	0.007	0.001	-0.011	-0.008
	(0.009)	(0.007)	(0.011)	(0.007)	(0.008)
Observations	665	665	665	665	665
R-squared	0.87	0.87	0.87	0.87	0.87
Mean mitigating factor	0.13	0.29	0.26	0.39	0.17
Panel B: IV Second Stage — Literacy and Historical Hacienda Concentration					
Dep. var.: literacy rate (share) in 1940					
Share pop. *Hacienda* 1876	-0.133***	-0.147***	-0.122***	-0.154***	-0.153***
	(0.036)	(0.051)	(0.030)	(0.040)	(0.044)
Hacienda x mitigating factor	0.006	0.038	-0.024	0.080**	0.095**
	(0.040)	(0.077)	(0.075)	(0.040)	(0.045)
Kleibergen-Paap F statistic	8.72	8.04	9.77	9.82	7.58
Observations	665	665	665	665	665
R-squared	0.702	0.705	0.719	0.713	0.691

Notes: The dependent variable is the district literacy rate in 1940 (mean = 0.334). Each column reports OLS (Panel A) and IV (Panel B) estimates of the association between historical hacienda concentration (the share of population living in haciendas in 1876, standardized) and literacy, including an interaction with a "potential mitigating factor" dummy: proximity (50 km) to a city of >20,000 inhabitants in 1940 (col 1), manufacture labor-force share above the 75th percentile in 1940 (col 2), mining labor-force share above the 75th percentile in 1940 (col 3), worker density below 2 workers/km² in 1940 (col 4), and proximity (10 km) to a paved road in 1950 (col 5). Hacienda concentration and its interaction term are instrumented by the soil crop suitability index and its interaction with the mitigating-factor dummy. All regressions include geographic controls (elevation, slope, ruggedness, area, longitude, latitude), 1940 demographic controls (population, population density, urban share, ethnic composition), and department fixed effects. Robust standard errors are in parentheses, Anderson–Rubin confidence intervals in brackets. Coefficients significantly different from zero are denoted by: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

century Peru, including urban services, manufacture (particularly textiles), and mining (Thorp and Bertram, 1978). Across both OLS and IV models, the interaction terms remain small and statistically insignificant, suggesting that alternative elite groups did not meaningfully challenge the dominance of landed elites in shaping educational outcomes. This pattern aligns with historical accounts emphasizing Peru's weak industrial development and continued reliance on primary exports during this period. The lack of a strong industrial base limited the emergence of countervailing elite coalitions, leaving landowners with few incentives to diversify their rents or invest in broader public goods such as education (Thorp and Bertram, 1978; Bértola and Ocampo, 2012).

Columns (4) and (5) shift focus to labor scarcity and the availability of outside options. Here, the IV estimates reveal statistically significant and positive interaction terms, indicating that these conditions attenuated the negative effect of haciendas on

literacy. For example, in districts facing high labor scarcity, the estimated effect of hacienda concentration is reduced from –15.4 to –5.9 percentage points, though the negative impact is not fully offset. These findings suggest that when workers had stronger fallback options or greater bargaining power, the influence of hacienda elites on local institutions weakened—highlighting the role of broader structural shifts in mediating institutional persistence.

For the third mitigating factor—the 1970s land reform—we rely on district-level data on reform exposure from Albertus et al. (2020). The authors find that greater exposure to the reform reduced educational attainment, as measured by years of schooling. This decline stemmed largely from demand-side constraints: binding cooperative obligations and scarce off-farm opportunities trapped beneficiaries in stagnant rural areas, reduced household income over time, and increased reliance on child labor—dynamics that limited families' ability to keep children in school.

Due to selection concerns in the roll-out of the reform, which cannot be addressed by our IV strategy (designed to capture quasi-exogenous variation in hacienda formation rather than reform exposure), we cannot interact land reform exposure with historical hacienda concentration in the same manner as other mitigating factors analyzed in Table 5.[26] Instead, we show our results are robust to the inclusion of the share of expropriated agricultural land per district by the reform (Table A7). Aligned with Albertus et al. (2020), these results suggest that, while the reform weakened the political and economic power of landed elites, limitations stemmed from both implementation shortcomings and unintended incentives that encouraged beneficiaries to remain in subsistence agriculture, limited mobility and continued access to educational opportunities.

8. Conclusion

This paper examines how landed elites shaped the trajectory of human capital development in Peru over the past 150 years. To this end, we compiled original district-level data on the rollout of haciendas in Peru since 1876 and over a century of population census records. For causal identification, we implemented an instrumental variable design that leverages plausibly exogenous variation in soil suitability for a subset of pre 19th-century crops, driving differential levels of hacienda spatial concentration, as well

[26]Districts that were eventually targeted by the land reform already exhibited significantly higher literacy levels prior to the reform. This is evident from a positive and statistically significant correlation between reform exposure and literacy rates in both pre-reform years (1940 and 1961) and post-reform years (1981 and 1993).

as variation across age-cohort groups. Our findings show that historical hacienda concentration significantly hindered literacy expansion in affected districts throughout the 20th century, with negative effects also observed in other educational outcomes, albeit diminishing in recent decades.

The evidence highlights supply-side constraints as the primary mechanism behind this impact: hacienda-dominated districts received lower public investments in education, while funding for other public goods remained unaffected. Additionally, these districts showed minimal changes in local economic activity, agricultural modernization, or broader development outcomes. Although factors like the rise of competing non-agrarian elites (industrialization), structural shifts in labor markets and top-down land redistribution later emerged, they ultimately proved insufficient to offset the enduring adverse effects of hacienda elites on human capital formation. These findings highlight the active role landed elites' incentives and capacities could play in constraining development trajectories—challenging the view that such outcomes are merely passive consequences of long-run land inequality.

Even after the formal abolition of haciendas, many of the underlying structures that had enabled elite control—weak local governance, a poorly educated population, and deficient policy-making—remained intact. These self-reinforcing mechanisms may have helped entrench the legacy of extractive institutions, allowing inequalities rooted in colonial landholding patterns to shape educational outcomes decades after the collapse of the agrarian order. Our findings thus resonate with a growing body of research highlighting how early land concentration can delay structural transformation and mass schooling (Galor et al., 2009; Cinnirella and Hornung, 2016; Goñi, 2023).

In this context, the Peruvian experience stands in contrast to cases where elite adaptation fostered more inclusive development. As documented in studies from Colombia and central Mexico (Acemoglu et al., 2008; Faguet et al., 2022; Arias and Flores-Peregrina, 2024), some coercive landholding systems eventually gave rise to local governance institutions, trade linkages, or broader fiscal capacity. In Peru, however, the narrow agricultural base of elite rents, combined with a weak and fragmented industrial sector (Thorp and Bertram, 1978), seem to have left few incentives for such transformations. Instead, the country's landowning elites retained concentrated power over labor and land, actively resisting educational investments and reinforcing exclusionary development trajectories well into the 20th century.

References

Abad, L. and N. Maurer (2025) "The long shadow of history? The impact of colonial labor institutions on economic development in Peru," *Journal of Economic Growth*, 10.1007/s10887-024-09249-9.

Acemoglu, D., M. A. Bautista, P. Querubin, and J. A. Robinson (2008) *Economic and political inequality in development: the case of Cundinamarca*, 181–248, Working paper.

Acemoglu, Daron, Georgy Egorov, and Konstantin Sonin (2021) "Institutional Change and Institutional Persistence," in Bisin, Alberto and Giovanni Federico eds. *Handbook of Historical Economics*, 365–389, Amsterdam: Elsevier, 10.1016/B978-0-12-815874-6.00015-1.

Acemoglu, Daron, Simon Johnson, and James A. Robinson (2006) "Institutions as the Fundamental Cause of Long-Run Growth," in Aghion, Philippe and Steven N. Durlauf eds. *Handbook of Economic Growth*, 1A, 385–472, Amsterdam: Elsevier, 10.1016/S1574-0684(05)01006-3.

Acemoglu, Daron, Tristan Reed, and James A. Robinson (2014) "Chiefs: Economic Development and Elite Control of Civil Society in Sierra Leone," *Journal of Political Economy*, 122 (2), 319–368, 10.1086/674988.

Acemoglu, Daron and James A. Robinson (2000) "Why Did the West Extend the Franchise? Democracy, Inequality, and Growth in Historical Perspective," *Quarterly Journal of Economics*, 115 (4), 1167–1199.

Acemoglu, Daron and Alexander Wolitzky (2011) "The Economics of Labor Coercion," *Econometrica*, 79 (2), 555–600, 10.3982/ECTA8976.

Alberti, G. (1972) "Educación y movilización colectiva," in Alberti, Giorgio and Julio Cotler eds. *Aspectos sociales de la educación rural en el Perú*, 97–124, Lima: Instituto de Estudios Peruanos.

——— (1976) "Ruptura de la Estructura de Poder Urbano Provincial y Surgimiento de Movimientos Campesinos," in *Hacienda, Comunidad y Campesinado en el Peru*, 343–367: IEP Ediciones.

Albertus, Michael (2015) *Autocracy and Redistribution: The Politics of Land Reform*: Cambridge University Press, 1-353, https://doi.org/10.1017/CBO9781316227107.

——— (2019) "Land Reform and Civil Conflict: Theory and Evidence from Peru," *American Journal of Political Science*, 63 (2), 250–266, 10.1111/ajps.12409.

Albertus, Michael, Mauricio Espinoza, and Ricardo Fort (2020) "Land reform and human capital development: Evidence from Peru," *Journal of Development Economics*, 147, 102540, https://doi.org/10.1016/j.jdeveco.2020.102540.

Albertus, Michael and Bogdan G. Popescu (2020) "Does Equalizing Assets Spur Development? Evidence From Large-Scale Land Reform in Peru," *Quarterly Journal of Political Science*, 15 (2), 255–295, 10.1561/100.00019033.

Arias, L. M. and D. Flores-Peregrina (2024) "Colonial agricultural estates and rural development in twentieth-century Mexico," *Economic History of Developing Regions*, 39 (2), 105–144.

Arroyo Abad, Leticia (2016) "THE LIMITS OF THE ESTADO DOCENTE: EDUCATION AND POLITICAL PARTICIPATION IN PERU, 1876-1940," *Revista de Historia Económica / Journal of Iberian and Latin American Economic History*, 34 (1), 81–109, 10.1017/S0212610915000233.

Arteaga, F. (2016) "Fragmentation of an Empire: The Political Economy of Spain and Its Colonies in the Late Bourbon Period," SSRN Working Paper.

Ashraf, Q. H., F. Cinnirella, O. Galor, B. Gershman, and E. Hornung (2024) "Structural Change, Elite Capitalism, and the Emergence of Labour Emancipation," *Review of Economic Studies*, 10.1093/restud/rdae043.

Baland, J. M. and J. A. Robinson (2008) "Land and power: Theory and evidence from Chile," *American Economic Review*, 98 (5), 1737–1765, 10.1257/aer.98.5.1737.

Bértola, L. and A. Ocampo (2012) *The Economic Development of Latin America Since Independence*: OUP Oxford.

Bisin, Alberto and Thierry Verdier (2024) "On the Joint Evolution of Culture and Political Institutions: Elites and Civil Society," *Journal of Political Economy*, 132 (5), 1485–1564, 10.1086/727711.

Bobonis, Gustavo J. and Peter M. Morrow (2014) "Labor coercion and the accumulation of human capital," *Journal of Development Economics*, 108, 32–53, https://doi.org/10.1016/j.jdeveco.2014.01.004.

Cinnirella, F. and E. Hornung (2016) "Landownership concentration and the expansion of education," *Journal of Development Economics*, 121, 135–152.

Coatsworth, John H. (2008) "Inequality, Institutions and Economic Growth in Latin America," *Journal of Latin American Studies*, 40 (3), 545–569, http://www.jstor.org/stable/40056706.

Contreras, Carlos and Marcos Cueto (2013) *Historia del Perú contemporáneo: Desde las luchas por la independencia hasta el presente*, Lima: Instituto de Estudios Peruanos, 6th edition.

Contreras, Carlos and Patricia Oliart (2014) *Modernidad y educación en el Perú*, Diversidad Cultural, Serie 8, Lima: Ministerio de Cultura del Perú.

Cotler, J. (1976) "Haciendas y Comunidades Tradicionales en un Contexto de Mobilizacion Politica," in *Hacienda, Comunidad y Campesinado en el Peru*, 311–342: IEP Ediciones.

Craig, W. W. (1967) *From Hacienda to Community: An Analysis of Solidarity and Social Change in Peru* in , Cornell University Latin American Program Dissertation Series (6), Ithaca: Cornell University.

Davies, K. A. (1984) *Landowners in Colonial Peru*: University of Texas Press.

Dell, M. (2010) "The Persistent Effects of Peru's Mining Mita," *Econometrica*, 78 (6), 1863–1903, 10.3982/ECTA8121.

Diegert, P., M. A. Masten, and A. Poirier (2022) "Assessing Omitted Variable Bias When the Controls Are Endogenous," *arXiv preprint arXiv:2206.02303*.

Duflo, E. (2001) "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," *American Economic Review*, 91 (4), 795–813.

Easterly, W. (2007) "Inequality does cause underdevelopment: Insights from a new instrument," *Journal of Development Economics*, 84 (2), 755–776, 10.1016/j.jdeveco.2006.11.002.

Engerman, S. and K. Sokoloff (1997) *Factor Endowments, Institutions, and Differential Paths of Growth Among New World Economies*: Stanford University Press.

Espinoza, Mauricio and Javier Escobal (2025) "The Impact of Peru's Land Reform on National Agricultural Productivity: A Synthetic Control Study," *Land Use Policy*, 157, 107619, 10.1016/j.landusepol.2025.107619.

Faguet, J.-P., C. Matajira, and F. Sánchez (2022) "Constructive extraction? Encomienda, the colonial state, and development in Colombia," Tech. Rep. No. 12.

Favre, H. (1976) "Evolucion y Situation de la Hacienda Tradicional de la Region Huancavelica," in *Hacienda, Comunidad y Campesinado en el Peru*, 105–138: IEP Ediciones.

Feir, Donn L, Maggie EC Jones, and Angela Redish (2024) "American Indian Wealth in the Early Twentieth Century," in *AEA Papers and Proceedings*, 114, 210–214, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Fenske, James, Bishnupriya Gupta, and Anwesh Mukhopadhyay (2025) "Colonial Persistence," The Warwick Economics Research Paper Series (TWERPS) 1557, University of Warwick, Department of Economics, None.

Frankema, E. (2009a) "The colonial roots of land inequality: Geography, factor endowments, or institutions?" *Economic History Review*, 63 (2), 418–451, 10.1111/j.1468-0289.2009.00479.x.

——— (2009b) *Has Latin America Always Been Unequal? A Comparative Study of Asset and Income Inequality in the Long Twentieth Century*, 3: Brill.

Galor, O. and O. Moav (2006) "Das Human-Kapital: A Theory of the Demise of the Class Structure," *The Review of Economic Studies*, 73 (1), 85–117.

Galor, O., O. Moav, and D. Vollrath (2009) "Inequality in landownership, the emergence of human-capital promoting institutions, and the great divergence," *Review of Economic Studies*, 76 (1), 143–179.

Galor, O. and Ö. Özak (2016) "The Agricultural Origins of Time Preference," *American Economic Review*, 106 (10), 3064–3103.

Galor, Oded and Ömer Özak (2015) "Land Productivity and Economic Development: Caloric Suitability vs. Agricultural Suitability," Brown University Working Paper.

Gill, N. (2018) *Between Two Worlds: Andean Haciendas in Colonial History*: Southern Affairs.

Gonzales, Michael J. (1993) *Azúcar y trabajo: La transformación de las haciendas en el norte del Perú*, Lima, Peru: Instituto de Estudios Peruanos.

Goñi, M. (2023) "Landed elites and education provision in England: Evidence from school boards, 1871-99," *Journal of Economic Growth*, 28, 125–171.

Handelman, H. (1975) *Struggle in the Antes: Peasant Political Mobilization in Peru*, Texas, USA: Institute of Latin American Studies, University of Texas Press, Austin.

Hurley, D. (2011) "Spanish Colonial Economies: An Overview of the Economy of the Viceroyalty of Peru, 1542-1600."

Ishimaru, S. (2024) "Empirical Decomposition of the IV-OLS Gap with Heterogeneous and Nonlinear Effects," *The Review of Economics and Statistics*, 106 (2), 505–520.

Jaramillo-Echeverri, Juliana and AndrÃ©s Alvarez (2023) "The persistence of segregation in education: Evidence from historical elites and ethnic surnames in Colombia," Cuadernos de Historia Económica 58, Banco de la Republica de Colombia, 10.32468/chee.58.

Keane, M. and T. Neal (2024) "A Practical Guide to Weak Instruments," *Annual Review of Economics*, 16, 185–212.

Keith, R. (1976) *Conquest and Agrarian Change: The Emergence of the Hacienda System on the Peruvian Coast*: Harvard Historical Studies 93, Harvard University Press.

Lockhart, J. (1969) "Encomienda and Hacienda: The Evolution of the Great Estate in the Spanish Indies," *The Hispanic American Historical Review*, 49 (3), 411–429.

Mares, I. (2015) *From Open Secrets to Secret Voting: Democratic Electoral Reforms and Voter Autonomy*: Cambridge University Press.

Mariscal, Elisa and Kenneth Sokoloff (2000) "Schooling, Suffrage and the Persistence of Inequality in the Americas," in Haber, Stephen ed. *Political Institutions and Economic Growth in Latin America: Essays in Policy, History and Political Economy*, 159–218, Stanford: Hoover Institution Press.

Markevich, Andrei and Ekaterina Zhuravskaya (2018) "The Economic Effects of the Abolition of Serfdom: Evidence from the Russian Empire," *American Economic Review*, 108 (4-5), 1074–1117, 10.1257/aer.20160144.

Masten, M. A. and A. Poirier (2022) "The Effect of Omitted Variables on the Sign of Regression Coefficients," *arXiv preprint arXiv:2208.00552*.

Matos Mar, J. and J. M. Mejía (1980) *La Reforma Agraria en el Perú*: Editorial Perú Problema.

McClintock, C. (1981) *Peasant Cooperatives and Political Change in Peru*, Princeton, NJ: Princeton University Press.

Méndez, E. and D. Van Patten (2022) "Multinationals, Monopsony, and Local Development: Evidence from the United Fruit Company," *Econometrica*, 90 (6), 2685–2721, https://doi.org/10.3982/ECTA19514.

Nunn, Nathan (2009) "The Importance of History for Economic Development," *Annual Review of Economics*, 1 (Volume 1, 2009), 65–92, https://doi.org/10.1146/annurev.economics.050708.143336.

Oster, E. (2019) "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business & Economic Statistics*, 37 (2), 187–204.

O'Toole, R. S. (2012) *Bound Lives: Africans, Indians, and the Making of Race in Colonial Peru*: Pittsburgh University Press.

Paige, J.M. (1975) *Agrarian Revolution: Social Movements and Export Agriculture in the Underdeveloped World*: Free Press, https://books.google.fr/books?id=_O6FAAAAIAAJ.

Paredes, Héctor (2023) "Land Without Masters: Local Political Competition Since the Peruvian Land Reform (1969–1980),"Technical report, Hal Dumas Working Paper, https://dumas.ccsd.cnrs.fr/dumas-04188002, Working paper.

Persson, Torsten, Gerard Roland, and Guido Tabellini (2000) "Comparative Politics and Public Finance," *Journal of Political Economy*, 108 (6), 1121–1161.

Puente, Javier (2023) *Rural State: Making Community and Claiming Power in Twentieth-Century Peru*, Austin: University of Texas Press, https://utpress.utexas.edu/9781477326305/.

Ramirez, S. E. (1994) "Instability at the top: A social history of the landed elite in colonial Peru," *Colonial Latin American Historical Review*, 3 (3), 327.

Ramírez, S. E. (2016) "Land and Tenure in Early Colonial Peru: Individualizing the Sapci, "That Which is Common to All"," *The Medieval Globe*, 2 (2), 33–70, 10.17302/TMG.2-2.3.

Rivadeneira, A. (2023) "The Legacy of Concertaje in Ecuador," in *Roots of Underdevelopment: A New Economic and Political History of Latin America and the Caribbean*, 127–162: Springer International Publishing.

Staiger, D. and J. Stock (1997) "Instrumental variables regression with weak instruments," *Econometrica*, 65 (3), 557–586.

Stein, S. (1980) *Populism in Peru: The Emergence of the Masses and the Politics of Social Control*, Madison: University of Wisconsin Press.

Stock, J. and M. Yogo (2005) "Testing for Weak Instruments in Linear IV Regression," in *Identification and Inference for Econometric Models*, 80–108: Cambridge University Press.

Tabares-Pozos, Germán, Irina España-Eljaiek, and Juan Carlos Muñoz-Mora (2025) "Education and land inequality: Evidence on Colombia land tenure and human capital formation," *Economic History of Developing Regions*, 40 (1), 49–79, 10.1080/20780389.2025.2459909.

Thorp, R. and G. Bertram (1978) *Peru 1890-1977: Growth and Policy in an Open Economy*: Columbia University Press.

Vasquez, M. (1961) *Hacienda, Peonaje y Servidumbre en los Andes Peruanos*, Lima, Peru: Editorial Estudios Andinos.

Walker, Charles (1999) *Smoldering Ashes: Cuzco and the Creation of Republican Peru, 1780–1840*, Durham: Duke University Press.

Yeager, T. J. (1995) "Encomienda or Slavery? The Spanish Crown's Choice of Labor Organization in Sixteenth-Century Spanish America," *The Journal of Economic History*, 55 (4), 842–859.

Online Appendix for:

The Unpromised Land: Elites and Human Capital Development in 20th-Century Peru

Authors:

Mauricio Espinoza
Wageningen University & Research

Héctor Paredes
Paris School of Economics

January 14, 2026

Table of Contents

Appendix A — Descriptive Statistics **2**
Table A1 Descriptive Statistics 2

Appendix B — Instrumental Variable: Descriptives and Tests
Table B1 Hacienda Concentration and Instrument Correlates at Baseline (1876) . . 7
Table B2 IV–OLS Gap Decomposition in Hacienda Concentration Estimates 8
Table B3 First-Stage IV Estimates using alternative Soil Suitability Measures 9
Figure B1 Soil Suitability Indexes 26
Figure B2 Soil Crop Suitability and Literacy by Hacienda Concentration 27

Appendix C — Robustness Checks
Table C1a Specification Tests (1876-1981) 12
Table C1b Specification Tests (1993-2017) 13
Table C2 Excluding Subsamples and Controlling for Land Reform Exposure 14
Table C3 Historical Hacienda Concentration on Literacy (by Census Period) 10
Table C4 Robustness to Alternative Hacienda Concentration Definitions 11
Figure C1 Regression Sensitivity Analysis 30
Figure C2 Persistence of the Hacienda Effect on Literacy 28

Appendix D — Additional Results
Table D1 Hacienda Concentration and Literacy by Age–Cohort 15
Table D2 OLS and IV: Hacienda Concentration and Economic Well-being 16
Table D3 OLS and IV: Hacienda Concentration and Structural Transformation . . . 17
Table D4 Long-Term Economic and Health Outcomes 18
Table D5 Farming Conditions (1994 Ag. Census) 19
Table D6 Hacienda Concentration and Indigenous Political Participation 20
Table D7 Hacienda Concentration and Literacy by Natural Region 21
Table D7 OLS and IV Estimates Linking Hacienda Concentration and Literacy: The Mediating Role of Hacienda Fragmentation 22
Figure D1 Gendered Effects on Literacy by Census Period (A–R) 29

Appendix E — Dell (2010) Comparing Study
Table E1 Mita and Hacienda Concentration — Dell (2010) Subsamples 23
Table E2 Hacienda Concentration and Literacy — Dell (2010) Subsamples 24
Table E3 Hacienda Concentration, Literacy and Public Goods 25
Figure E1 Dell (2010) Study Area and 1876 Haciendas around Mita Boundary . . 31
Figure E2 Hacienda Concentration and Literacy by Dell (2010) Subsamples 32

Tables

Table A1: Descriptive Statistics

Variable	Mean	SD	Min	Max
Panel A: Literacy Rate				
Literacy rate (share) - 1876	0.125	0.112	0.002	0.642
Literacy rate (share) - 1940	0.334	0.217	0.035	0.894
Literacy rate (share) - 1961	0.481	0.218	0.040	0.940
Literacy rate (share) - 1981	0.679	0.149	0.265	0.956
Literacy rate (share) - 1993	0.755	0.113	0.372	0.962
Literacy rate (share) - 2007	0.858	0.070	0.584	0.983
Literacy rate (share) - 2017	0.869	0.058	0.660	0.981
Panel B: Hacienda Concentration				
Share pop. Hacienda 1876	0.148	0.193	0.000	0.943
Share pop. Hacienda 1940	0.161	0.185	0.000	0.930
Share pop. Hacienda 1961	0.122	0.161	0.000	0.935
Panel C: Geographic Controls				
Elevation (1000 meters above the sea level)	3.008	1.323	0.007	4.770
Mean slope	8.415	4.195	0.070	18.827
Mean Ruggedness	4.363	2.188	0.034	9.949
Surface area (squared km.)	814.98	876.46	4.52	8,368.82
District centroid longitude	-75.016	3.221	-81.092	-68.980
District centroid latitude	-11.678	3.581	-18.083	-3.552
Panel D: Demographic and Econ. Controls - 1876				
Population (logged)	7.843	0.900	0.000	11.245
Population density (10 people per sq. km.)	1.110	4.128	0.000	94.774
Share urban population	0.290	0.306	0.000	1.000
Share indigenous ethnicity	0.639	0.281	0.001	1.000
Share white ethnicity	0.114	0.152	0.000	0.942
Share foreign population	0.008	0.032	0.000	0.355
Work force share in agriculture	0.679	0.244	0.011	1.000
Work force share in mining	0.006	0.030	0.000	0.429
Work force share in manufacturing	0.199	0.213	0.000	0.925
Panel E: Demographic and Econ. Controls - 1940				
Population (logged)	8.556	0.956	5.438	13.377

Variable	Mean	SD	Min	Max
Population density (10 people per sq. km.)	2.091	5.754	0.032	111.950
Share urban population	0.275	0.224	0.005	1.000
Share indigenous ethnicity	0.520	0.325	0.000	0.989
Work force share in agriculture	0.653	0.126	0.023	0.790
Work force share in mining	0.021	0.043	0.000	0.389
Work force share in manufacturing	0.159	0.056	0.064	0.352
Panel F: Demographic and Econ. Controls - 1961				
Population (logged)	8.818	1.009	5.687	14.428
Population density (10 people per sq. km.)	3.058	10.265	0.026	194.927
Share urban population	0.301	0.240	0.000	1.000
Spanish mother tongue	0.398	0.416	0.000	1.000
Work force share in agriculture	0.738	0.193	0.046	0.972
Work force share in mining	0.023	0.079	0.000	0.701
Work force share in manufacturing	0.052	0.056	0.000	0.440
Panel G: Demographic and Econ. Controls - 1981				
Population (logged)	9.006	1.171	5.366	15.335
Population density (10 people per sq. km.)	5.101	19.324	0.050	292.446
Share urban population	0.381	0.279	0.000	1.000
Indigenous mother tongue	0.450	0.406	0.000	0.998
Work force share in agriculture	0.666	0.214	0.023	1.000
Work force share in mining	0.031	0.088	0.000	0.636
Work force share in manufacturing	0.055	0.066	0.000	0.570
Panel H: Demographic and Econ. Controls - 1993				
Population (logged)	9.090	1.223	5.892	15.663
Population density (10 people per sq. km.)	6.240	25.304	0.084	384.653
Share urban population	0.400	0.278	0.013	1.000
Indigenous mother tongue	0.406	0.382	0.000	0.995
Work force share in agriculture	0.677	0.223	0.019	0.960
Work force share in mining	0.022	0.077	0.000	0.766
Work force share in manufacturing	0.068	0.057	0.000	0.366
Panel I: Demographic and Econ. Controls - 2007				
Population (logged)	9.188	1.275	5.631	15.954
Population density (10 people per sq. km.)	7.621	32.197	0.062	506.629
Share urban population	0.463	0.284	0.016	1.000
Indigenous mother tongue	0.368	0.369	0.000	0.987

Variable	Mean	SD	Min	Max
Work force share in agriculture	0.574	0.218	0.012	0.930
Work force share in mining	0.036	0.095	0.000	0.741
Work force share in manufacturing	0.038	0.039	0.000	0.393
Panel J: Demographic and Econ. Controls - 2017				
Population (logged)	9.067	1.372	5.583	16.074
Population density (10 people per sq. km.)	8.423	36.813	0.038	562.560
Share urban population	0.534	0.282	0.026	1.000
Indigenous mother tongue	0.395	0.371	0.000	0.993
Work force share in agriculture	0.558	0.234	0.010	0.950
Work force share in mining	0.031	0.084	0.000	0.667
Work force share in manufacturing	0.033	0.032	0.000	0.236
Panel K: Soil Suitability				
Suitability cotton (GAEZ)	-5.115	3.187	-6.908	2.063
Caloric Suitability Index (Galor and Özak)	1.606	1.736	0.000	4.902
Crop Suitability Index	0.578	0.989	-0.563	2.252
Panel L: Literacy by Gender				
Male Literacy rate (share) - 1876	0.175	0.137	0.002	0.930
Male Literacy rate (share) - 1940	0.446	0.230	0.065	0.923
Male Literacy rate (share) - 1961	0.786	0.173	0.137	1.000
Male Literacy rate (share) - 1981	0.777	0.112	0.391	0.968
Male Literacy rate (share) - 1993	0.836	0.081	0.510	0.965
Male Literacy rate (share) - 2007	0.915	0.046	0.661	0.988
Male Literacy rate (share) - 2017	0.921	0.034	0.778	0.980
Female Literacy rate (share) - 1876	0.076	0.107	0.000	0.879
Female Literacy rate (share) - 1940	0.231	0.218	0.000	0.866
Female Literacy rate (share) - 1961	0.545	0.282	0.000	1.000
Female Literacy rate (share) - 1981	0.584	0.186	0.106	0.952
Female Literacy rate (share) - 1993	0.676	0.146	0.249	0.960
Female Literacy rate (share) - 2007	0.801	0.098	0.452	0.983
Female Literacy rate (share) - 2017	0.818	0.084	0.551	0.982
Panel M: Alternative Hacienda Measures				
Share of rural pop. living in haciendas- 1876	0.287	0.335	0.000	1.000
At least one hacienda dummy - 1876	0.675	0.469	0.000	1.000
Log population living in haciendas - 1876	3.949	3.029	0.000	9.295
Log number of haciendas - 1876	1.024	1.123	0.000	5.147

Variable	Mean	SD	Min	Max
Number of haciendas per 100 people - 1876	.287	0.335	0.000	1.000
Panel N: Other Educational Outcomes				
Share of hh. w/ children not att. school - 1993	0.139	0.070	0.015	0.450
Share of hh. w/ children not att. school - 2007	0.046	0.029	0.000	0.193
Share of hh. w/ children not att. school - 2017	0.022	0.013	0.000	0.093
Mean years of educ., adults 18-30 - 1993	6.691	1.640	2.171	10.225
Mean years of educ., adults 18-30 - 2007	9.041	1.702	3.854	13.494
Mean years of educ., adults 18-30 - 2017	10.419	1.236	6.501	13.970
Panel O: Supply of Education				
Density of schools in the province - 1918	13.026	9.736	2.678	57.651
Density of schools in the province - 1930	14.668	12.513	0.870	110.582
Density of schools in the province - 1950	18.833	10.028	2.785	54.343
Density of teachers in the province - 1918	15.208	11.109	2.695	68.764
Density of teachers in the province - 1930	20.960	14.453	1.740	77.690
Density of teachers in the province - 1950	33.584	16.545	9.590	114.521
Panel P: Other Public Goods				
Distance to the nat. road network - 1950	10.062	3.789	0.000	12.844
Distance to the nat. road network - 1975	8.133	4.709	0.000	12.653
Distance to the nat. road network - 2002	5.727	4.853	0.000	11.436
Sh. farmers with registered land title - 1994	0.249	0.189	0.000	1.000
Sh. farmland with registered land title - 1994	0.458	0.251	0.000	1.000
Share of irrigated farmland - 1961	0.119	0.166	0.000	0.778
Panel Q: Economic Well-Being				
Sh. living in precarious shelters - 1961	0.153	0.084	0.026	0.478
Sh. with access to a toilet - 1961	0.078	0.106	0.001	0.876
Sh. with access to electricity - 1961	0.126	0.157	0.001	0.763
Sh. of individuals not wearing shoes - 1961	0.624	0.208	0.293	0.997
Log of average per capita income - 1981	1.837	0.636	0.030	3.749
Sh. aged 6–17 engaged in paid work - 1981	0.079	0.050	0.000	0.332
Panel R: Structural Change				
Employment share in agriculture - 1961	0.738	0.193	0.046	0.972
Employment share in manufacturing - 1961	0.052	0.056	0.000	0.440
Employment share in services - 1961	0.064	0.076	0.000	0.507
Employment share in trade - 1961	0.033	0.027	0.000	0.161
Employment share in mining - 1961	0.023	0.079	0.000	0.701

Variable	Mean	SD	Min	Max
Urban population share - 1961	0.301	0.240	0.000	1.000
Panel S: Long-Term Economic and Health Outcomes				
Log Nighttime lights (2016)	3.090	0.450	1.099	5.541
Stunted growth, Children 6–9 (1993)	0.562	0.169	0.082	1.000
Child mortality - 2007	0.248	0.089	0.095	0.539
Poverty rate (%) - 2007	0.605	0.220	0.039	0.997
Log Per capita expenditure (PEN) - 2009	5.860	0.320	5.102	7.241
Panel T: Potential Mitigating Factors				
Urban elite (cities) - 1940	0.134	0.341	0.000	1.000
Manufacture industry elite - 1940	0.290	0.454	0.000	1.000
Mining industry elite - 1940	0.256	0.437	0.000	1.000
High labor scarcity - 1940	0.389	0.488	0.000	1.000
Close to paved road - 1950	0.170	0.376	0.000	1.000
Panel U: Farming (1994 Agricultural Census)				
Average farm size (ha. of ag, land) - 1994	2.649	1.776	0.000	15.217
Share of farms with less than one hectare - 1994	0.414	0.211	0.031	1.000
The agricultural land Gini coefficient - 1994	0.552	0.109	0.000	0.949
Farmers' Ave. yrs of education - 1994	4.706	1.360	1.683	9.990
Share of farmers who use electric power- 1994	0.006	0.015	0.000	0.285
Share of farmers who use fertilizer- 1994	0.393	0.294	0.000	0.983
Share of farmers who use insecticides- 1994	0.439	0.283	0.000	0.988
Panel V: Natural Regions				
District sample share located in the Coast	0.168	0.375	0.000	1.000
District sample share located in the Highlands	0.832	0.375	0.000	1.000
Panel W: Indigenous Political Participation (1933 Electoral Census)				
Indigenous share of registered voters - 1933	0.334	0.246	0.000	0.930
Panel X: Hacienda Fragmentation				
Inverse HHI of hacienda population in 1876	0.323	0.476	0.000	5.342

Table A2: Hacienda Concentration and Instrument correlates at baseline (1876)

Dependent variable:	**Historical hacienda concentration (1876)**		**Soil Crop Suitability (instrument)**	
	Coef.	**S.E.**	**Coef.**	**S.E.**
Soil Crop Suitability Index	0.302***	(0.062)		
Elevation (mean, thousands meters)	0.070	(0.060)	-0.670***	(0.027)
Slope (mean, degrees)	-0.165*	(0.091)	-0.056	(0.061)
Ruggedness (mean)	0.300*	(0.178)	0.216*	(0.116)
Area (squared km.)	0.000	(0.000)	-0.000	(0.000)
Longitude	-0.120	(0.082)	0.209***	(0.049)
Latitude	-0.104	(0.074)	0.321***	(0.035)
Log population 1876	0.098**	(0.043)	0.087***	(0.032)
Population density 1876	-0.006	(0.007)	-0.012***	(0.003)
Share urban 1876	-0.917***	(0.116)	0.063	(0.074)
Share indigenous ethnicity 1876	0.350	(0.254)	0.111	(0.105)
Share mestizo ethnicity 1876	1.153***	(0.309)	0.005	(0.172)
Share foreigners 1876	2.004	(1.409)	0.528	(0.622)
Share Agriculture 1876	0.302	(0.268)	-0.204	(0.174)
Share Mining 1876	2.722***	(0.899)	-1.124***	(0.431)
Share Manufacture 1876	-0.150	(0.333)	-0.229	(0.195)
Constant	-11.407*	(6.830)	20.367***	(4.003)
Observations	665		665	
R-squared	0.34		0.804	
Adjusted R-squared	0.31		0.79	
Department FE	Yes		Yes	

Notes: The units of observation are 665 historical districts from 1876. In column 1, the dependent variable is the share of the population residing in *haciendas* in 1876 (standardized with zero mean and unit standard deviation), while in column 2 is our instrument, the soil crop suitability index. Robust standard errors are reported in parentheses. Coefficients significantly different from zero are denoted by: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A3: IV-OLS Gap decomposition in Hacienda Concentration Estimates

Census year	**Coefficients**			**Decomposition**		
	OLS	**IV**	**Gap**	Δ_{CW}	Δ_{TW}	Δ_{ME}
1876	-0.019***	-0.043**	-0.024	0.011	0.000	-0.035*
	(0.005)	(0.020)	(0.019)	(0.008)	(0.000)	(0.020)
1940	-0.026***	-0.131***	-0.105**	0.014	0.000	-0.119***
	(0.004)	(0.035)	(0.035)	(0.011)	(0.000)	(0.039)
1961	-0.020***	-0.232***	-0.212***	-0.005	0.000	-0.208***
	(0.005)	(0.065)	(0.065)	(0.015)	(0.000)	(0.064)
1981	-0.011***	-0.137***	-0.126***	0.007	0.000	-0.133**
	(0.003)	(0.052)	(0.052)	(0.013)	(0.000)	(0.054)
1993	-0.007***	-0.102***	-0.096***	0.004	0.000	-0.100***
	(0.002)	(0.034)	(0.035)	(0.011)	(0.000)	(0.036)
2007	-0.004**	-0.051***	0.048***	-0.001	0.000	-0.046**
	(0.002)	(0.017)	(0.018)	(0.005)	(0.000)	(0.018)
2017	-0.002*	-0.044**	-0.031***	0.003	0.000	-0.034***
	(0.001)	(0.017)	(0.013)	(0.005)	(0.000)	(0.013)

Notes: The units of observation are 665 historical districts from 1876. The first three columns report the OLS estimates, the IV estimates, and their gaps, concerning the effect of historical hacienda concentration on literacy rates. The next three columns report the estimates of the covariate weight difference (Δ_{CW}), the treatment-level weight difference (Δ_{TW}), and the endogeneity bias or marginal effect difference (Δ_{ME}) components. Coefficients significantly different from zero are denoted by: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A4: First Stage IV Estimates of *Hacienda* Concentration and Alternative Soil Suitability Measures

Alternative Instruments:	**Cotton Suitability**	**Sugarcane Suitability**	**Caloric Suit. Index**	**Soil Crop Suit. [Max.(1)-(3)]**
Panel A: IV fist stage—*hacienda* concentration and alternative measures of soil suitability				
Dep. var.: hacienda concentration 1876				
Instrument (as in the headings)	0.288***	0.224***	0.136**	0.302***
	(0.068)	(0.069)	(0.060)	(0.062)
Kleibergen–Paap F statistic	17.74	10.44	5.07	23.57
Observations	665	665	665	665
R-squared	0.360	0.352	0.346	0.358

Notes: The table presents first-stage IV estimates of the relationship between alternative soil suitability measures and *hacienda* concentration in 1876 at the district level. The dependent variable is the standardized share of the district population living in *haciendas* in 1876. Columns 1 to 4 use different instruments: cotton suitability (GAEZ), sugarcane suitability (GAEZ), caloric suitability index (Galor & Özak), and our composite soil crop suitability index (maximum value between standardized cotton suitability and standardized caloric suitability). All regressions include geographic controls (i.e., elevation, slope, ruggedness, area, latitude, and longitude), 1876 demographic and economic controls (i.e., population, density, urban share, ethnicity, and sectoral labor shares), and department fixed effects. Robust standard errors are reported in parentheses. Coefficients significantly different from zero are denoted by: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A8: The Effect of Historical Hacienda Concentration on Literacy using District Boundaries in Effect at Each Census Period

	1876	1940	1961	1981	1993	2007	2017
Panel A: OLS — literacy and hacienda concentration							
Dep. var.: literacy rate (share) at census year							
Share pop. *Hacienda* 1876	-0.019***	-0.027***	-0.022***	-0.010***	-0.007***	-0.004**	-0.004***
	(0.005)	(0.004)	(0.005)	(0.004)	(0.002)	(0.002)	(0.001)
Mean DV	0.126	0.359	0.504	0.693	0.764	0.864	0.872
Observations	665	955	1299	1312	1559	1594	1621
R-squared (OLS)	0.548	0.856	0.798	0.759	0.764	0.774	0.752
Panel B: IV first stage — hacienda concentration and crop suitability							
Dep. var.: population in Haciendas in 1876 (share)							
Soil Crop Suitability	0.302***	0.281***	0.263***	0.214***	0.231***	0.232***	0.229***
	(0.062)	(0.064)	(0.066)	(0.060)	(0.061)	(0.061)	(0.061)
Kleibergen-Paap F statistic	23.57	19.22	17.79	10.42	14.68	14.53	13.87
Observations	665	955	1299	1312	1559	1594	1621
R-squared (1st stage)	0.358	0.290	0.308	0.286	0.298	0.294	0.292
Panel C: IV second stage — literacy and hacienda concentration							
Dep. var.: literacy rate (share) at census year							
Share pop. *Hacienda* 1876	-0.043**	-0.135***	-0.212***	-0.118***	-0.086***	-0.039***	-0.037***
	(0.020)	(0.034)	(0.050)	(0.039)	(0.024)	(0.013)	(0.012)
Upper A-R 95% CI	[-0.003	[-0.082	[-0.137	[-0.061	[-0.048	[-0.017	[-0.017
Lower A-R 95% CI	[-0.089]	[-0.236]	[-0.374]	[-0.278]	[-0.166]	[-0.077]	[-0.074]
Observations	665	955	1299	1312	1559	1594	1621
R-squared (2nd stage)	0.522	0.706	0.360	0.421	0.506	0.633	0.576

Notes: The table show district-level OLS and IV-estimates for seven separate cross-sections. The dependent variable is the literacy rate at each census year, defined as the proportion of individuals aged 5 or older who could read, write, or both. Historical hacienda concentration is captured by the percentage of the district population residing in haciendas in 1876, standardized to have a mean of zero and a standard deviation of one. Hacienda concentration is instrumented using the soil crop suitability index. All regressions control for geographic covariates (elevation, slope, ruggedness, area, longitude, and latitude), contemporaneous demographic covariates (population, population density, urban share, ethnicity share, agricultural, mining, and manufacturing share), and department fixed effects. Standard errors clustered at the 1876 district-level are reported in parentheses and Anderson-Rubin confidence intervals in brackets. Coefficients significantly different from zero are denoted by: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A9: Robustness to Different Hacienda Concentration Definitions

Hacienda concentration variable: (1876)	**Share rural pop. in hacienda (1)**	**At least one hacienda in the district (2)**	**Log hacienda population (3)**	**Log number of haciendas (4)**	**Number of haciendas per 100 pop. (5)**
Panel A: OLS — literacy and hacienda concentration					
Dep. var.: literacy rate 1940 (share)					
Hacienda measure (as headings)	-0.024***	-0.013***	-0.023***	-0.019***	-0.011***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Observations	665	665	665	665	665
R-squared (OLS)	0.868	0.863	0.866	0.864	0.861
Mean of hacienda measure	0.21	0.68	3.95	1.02	0.18
Panel B: IV first stage — hacienda concentration and crop suitability					
Dep. var.: Hacienda measure (as headings)					
Soil Crop Suitability	0.285***	0.195***	0.224***	0.237***	0.199***
	(0.070)	(0.080)	(0.071)	(0.069)	(0.069)
Observations	665	665	665	665	665
R-squared (1st stage)	0.312	0.333	0.405	0.378	0.263
Panel C: IV second stage — literacy and hacienda concentration					
Dep. var.: literacy rate 1940 (share)					
Hacienda measure (as headings)	-0.126***	-0.182**	-0.160***	-0.151***	0.180***
	(0.025)	(0.078)	(0.054)	(0.047)	(0.065)
Upper A-R 95% CI	[-0.073]	[-0.089]	[-0.088]	[-0.086]	[-0.096]
Lower A-R 95% CI	[-0.238]	[-0.838]	[-0.401]	[-0.336]	[-0.518]
Kleibergen-Paap F statistic	16.77	5.94	9.93	11.83	8.44
Observations	665	665	665	665	665
R-squared (2nd stage)	0.712	0.455	0.623	0.627	0.405
Geographic controls	Yes	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes	Yes
Department FE	Yes	Yes	Yes	Yes	Yes

Notes: The table presents district-level OLS and IV estimates of the relationship between four measures of historical hacienda concentration and literacy in 1940. The dependent variable is the district-level literacy rate in 1940, defined as the proportion of individuals aged 5 or older who could read, write, or both. Hacienda concentration in 1876 is measured as follows: column (1) uses the percentage of the district rural population residing in haciendas; column (2) includes a dummy variable indicating whether the district had at least one hacienda; column (3) uses the log of the population residing in haciendas; column (4) uses the log of the number of haciendas in the district; and column (5) uses the number of haciendas per 100 people in the district. All measures are standardized to have a mean of zero and a standard deviation of one. Hacienda concentration measures are instrumented using the soil crop suitability index. All regressions control for geographic covariates (elevation, slope, ruggedness, area, longitude, latitude), 1940 demographic covariates (population, population density, urban share, ethnicity share, and shares of agricultural, mining, and manufacturing activities), and department fixed effects. Robust standard errors are shown in parentheses, and Anderson-Rubin confidence intervals are reported in brackets. Coefficients significantly different from zero are denoted as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A5: Specification tests

	(1)	(2)	(3)	(4)
DV: Literacy rate in 1876				
Haciendas 1876	-0.013	-0.021	-0.019	-0.016
Robust S.E.	(0.004)***	(0.005)***	(0.005)***	(0.007)**
Conley S.E.	(0.006)**	(0.005)**	(0.005)**	(0.007)**
Wild cluster bootstrap (p-value)	0.046**	0.002***	0.004***	0.016**
R-squared	0.321	0.472	0.548	0.637
DV: Literacy rate in 1940				
Haciendas 1876	-0.022	-0.035	-0.026	-0.020
Robust S.E.	(0.005)***	(0.005)***	(0.004)***	(0.004)***
Conley S.E.	(0.007)***	(0.006)***	(0.004)***	(0.004)***
Wild cluster bootstrap (p-value)	0.024**	0.002***	0.004***	0.000***
R-squared	0.590	0.773	0.870	0.914
DV: Literacy rate in 1961				
Haciendas 1876	-0.022	-0.032	-0.020	-0.010
Robust S.E.	(0.006)***	(0.005)***	(0.005)***	(0.004)**
Conley S.E.	(0.007)***	(0.007)***	(0.004)**	(0.003)**
Wild cluster bootstrap (p-value)	0.018**	0.004***	0.006***	0.020**
R-squared	0.575	0.733	0.824	0.895
DV: Literacy rate in 1981				
Haciendas 1876	-0.013	-0.019	-0.011	-0.007
Robust S.E.	(0.005)***	(0.004)***	(0.003)***	(0.003)**
Conley S.E.	(0.006)**	(0.005)**	(0.004)**	(0.003)**
Wild cluster bootstrap (p-value)	0.056*	0.002***	0.014**	0.072*
R-squared	0.438	0.618	0.794	0.859
Observations	665	665	665	665
Department FE (n=18)	Yes	Yes	Yes	No
Province FE (n=88)	No	No	No	Yes
Geographic controls	No	Yes	Yes	Yes
Demographic controls	No	No	Yes	Yes

Notes: The dependent variable is the district-level literacy rate at each census year, defined as the proportion of individuals aged 5 or older who could read, write, or both. The hacienda variable represents the percentage of the district population residing in haciendas in 1876, standardized to have zero mean and unit standard deviation. Column (1) includes department FE, column (2) adds geographic and topographic control variables (district's elevation, mean slope, ruggedness, surface area, and centroid longitude and latitude), column (3) adds demographic and economic controls at each census year (population (log), population density (log), urban share, ethnicity share (white, indigenous, and mestizo ethnicity), and workers sectoral share (agricultural, mining, and manufacturing)). Robust SE are shown in parentheses, and Conley SE (assuming a cutoff distance of 200km) are in brackets. Wild bootstrap p-values are clustered at the department level in columns (1) to (3) and at the province level in column (4). Coefficients significantly different from zero are denoted by: $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A6: Specification tests (Part 2)

	(1)	(2)	(3)	(4)
DV: Literacy rate in 1993				
Haciendas 1876	-0.010	-0.015	-0.007	-0.004
Robust S.E.	(0.003)***	(0.003)***	(0.002)***	(0.002)*
Conley S.E.	(0.005)**	(0.004)***	(0.002)***	(0.002)**
Wild cluster bootstrap (p-value)	0.102	0.006***	0.058*	0.110
R-squared	0.440	0.599	0.795	0.865
DV: Literacy rate in 2007				
Haciendas 1876	-0.006	-0.010	-0.004	-0.003
Robust S.E.	(0.002)***	(0.002)***	(0.002)**	(0.001)**
Conley S.E.	(0.002)***	(0.002)***	(0.001)***	(0.002)*
Wild cluster bootstrap (p-value)	0.042**	0.006***	0.050**	0.038**
R-squared	0.372	0.531	0.789	0.869
DV: Literacy rate in 2017				
Haciendas 1876	-0.004	-0.007	-0.002	-0.000
Robust S.E.	(0.002)**	(0.002)***	(0.001)*	(0.001)
Conley S.E.	(0.002)**	(0.002)***	(0.001)***	(0.001)
Wild cluster bootstrap (p-value)	0.092*	0.004***	0.130	0.778
R-squared	0.368	0.516	0.784	0.845
Observations	665	665	665	665
Department FE (n=18)	Yes	Yes	Yes	No
Province FE (n=88)	No	No	No	Yes
Geographic controls	No	Yes	Yes	Yes
Demographic controls	No	No	Yes	Yes

Notes: The dependent variable is the district-level literacy rate at each census year, defined as the proportion of individuals aged 5 or older who could read, write, or both. The hacienda variable represents the percentage of the district population residing in haciendas in 1876, standardized to have zero mean and unit standard deviation. Column (1) includes department FE, column (2) adds geographic and topographic control variables (district's elevation, mean slope, ruggedness, surface area, and centroid longitude and latitude), column (3) adds demographic and economic controls at each census year (population (log), population density (log), urban share, ethnicity share (white, indigenous, and mestizo ethnicity), and workers sectoral share (agricultural, mining, and manufacturing)). Robust SE are shown in parentheses, and Conley SE (assuming a cutoff distance of 200km) are in brackets. Wild bootstrap p-values are clustered at the department level in columns (1) to (3) and at the province level in column (4). Coefficients significantly different from zero are denoted by: $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A7: Robustness to Excluding Different Subsamples and Controlling by Land Reform Exposure

Dep. var.: literacy rate (share) at each census period	1876	1940	1961	1981	1993	2007	2017
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: IV second stage (excluding 212 districts without any *hacienda* in 1876)							
Share pop. *Hacienda* 1876	-0.034*	-0.101***	-0.186***	-0.116**	-0.085***	-0.054***	-0.030**
	(0.013)	(0.035)	(0.062)	(0.047)	(0.032)	(0.020)	(0.014)
Kleibergen–Paap F statistic	25.68	13.19	10.97	6.93	7.88	8.71	8.35
Mean literacy rate	0.11	0.31	0.45	0.66	0.74	0.85	0.86
Observations	452	452	452	421	452	452	452
Panel B: IV second stage (excluding 175 districts with urban share above 0.5 in 1876)							
Share pop. *Hacienda* 1876	-0.050**	-0.116***	-0.228***	-0.157***	-0.107***	-0.056***	-0.037***
	(0.020)	(0.033)	(0.066)	(0.055)	(0.035)	(0.019)	(0.013)
Kleibergen–Paap F statistic	16.64	16.20	12.02	8.42	10.38	12.10	11.52
Mean literacy rate	0.11	0.30	0.45	0.66	0.74	0.85	0.86
Observations	489	489	489	446	489	489	489
Panel C: IV second stage (excluding 197 districts with workforce share in agriculture below 0.5 in 1876)							
Share pop. *Hacienda* 1876	-0.031**	-0.112***	-0.223***	-0.136***	-0.094***	-0.041***	-0.033***
	(0.016)	(0.024)	(0.050)	(0.040)	(0.027)	(0.013)	(0.011)
Kleibergen–Paap F statistic	31.44	24.31	19.71	12.82	13.63	18.15	15.56
Mean literacy rate	0.12	0.33	0.48	0.68	0.76	0.86	0.87
Observations	468	468	468	423	468	468	468
Panel D: IV second stage (excluding 238 districts with low crop soil suitability)							
Share pop. *Hacienda* 1876	-0.067**	-0.123**	-0.179*	-0.081	-0.079*	-0.047**	-0.031*
	(0.031)	(0.056)	(0.075)	(0.066)	(0.047)	(0.023)	(0.018)
Kleibergen–Paap F statistic	12.49	5.56	6.11	3.11	4.68	6.03	6.38
Mean literacy rate	0.14	0.36	0.49	0.68	0.75	0.85	0.86
Observations	427	427	427	396	427	427	427
Panel E: IV second stage (controlling by Land Reform exposure)							
Share pop. *Hacienda* 1876				-0.195**	-0.135**	-0.051***	-0.033***
				(0.090)	(0.056)	(0.017)	(0.013)
Kleibergen–Paap F statistic				4.67	5.98	11.31	10.16
Mean literacy rate				0.68	0.75	0.86	0.87
R-squared				-0.214	-0.042	0.455	0.575
Observations				611	665	665	665

Notes: The table reports second-stage IV estimates linking historical *hacienda* concentration to literacy rates across seven census years, using the same specification as in Table 1 but restricting the sample or adding controls in each panel. Panel A excludes 212 districts that had no *haciendas* in 1876. Panel B excludes 175 districts with an urban population share above 0.5 in 1876. Panel C excludes 197 districts with a workforce share in agriculture below 0.5 in 1876. Panel D excludes 238 districts with low soil crop suitability. Panel E uses the full sample of districts but includes a control for land reform exposure, measured as the share of privately owned agricultural land in a district expropriated during the land reform between 1969 and 1985. The dependent variable is the literacy rate at each census year. Historical *hacienda* concentration is captured by the percentage of the district population residing in *haciendas* in 1876. *Hacienda* concentration is instrumented using the soil crop suitability index. All regressions include geographic controls (e.g., elevation, slope, ruggedness, area, latitude, and longitude), demographic controls (e.g., population, population density, urban share, ethnicity share, and sectoral employment shares), and department fixed effects. Robust standard errors are reported in parentheses. Coefficients significantly different from zero are indicated as follows: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A10: The Effect of Historical Hacienda Concentration on Literacy by Age-Cohort Groups

	1901-1910	**1911-1920**	**1921-1930**	**1931-1940**	**1941-1950**	**1951-1960**	**1961-1970**	**1971-1980**	**1981-1990**	**1991-2000**	**2000-2010**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: OLS — literacy and historical hacienda concentration											
Dep. var.: literacy rate (share) by age-cohort group in 1993, 2007, or 2017											
Share pop. *Hacienda* 1876	-0.023***	-0.028***	-0.027***	-0.024***	-0.019***	-0.013***	-0.010***	-0.007***	-0.003***	-0.001***	0.000
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.002)	(0.001)	(0.000)	(0.000)
Observations	665	665	665	665	665	665	665	665	665	665	665
R-squared	0.764	0.805	0.807	0.775	0.713	0.619	0.54	0.456	0.442	0.463	0.204
Mean DV	0.464	0.522	0.590	0.665	0.759	0.861	0.911	0.946	0.976	0.988	0.971
Panel B: IV second stage — literacy and historical hacienda concentration											
Dep. var.: literacy rate (share) by age-cohort group in 1993, 2007, or 2017											
Share pop. *Hacienda* 1876	-0.115***	-0.123***	-0.141***	-0.122***	-0.098***	-0.065**	-0.053**	-0.039***	-0.012*	-0.004*	0.001
	(0.042)	(0.036)	(0.035)	(0.033)	(0.033)	(0.027)	(0.022)	(0.015)	(0.007)	(0.002)	(0.004)
Upper A-R 95% CI	[-0.227	[-0.226	[-0.249	[-0.219	[-0.196	[-0.141	[-0.118	[-0.082	[-0.030	[-0.010	[-0.007
Lower A-R 95% CI	-0.038]	-0.061]	-0.086]	-0.068]	-0.043]	-0.018]	-0.013]	-0.013]	0.001]	0.001]	0.010]
Kleibergen-Paap F statistic	18.07	17.81	17.73	17.45	15.79	14.64	13.65	15.06	15.23	13.82	15.56
Observations	665	665	665	665	665	665	665	665	665	665	665
R-squared	0.641	0.658	0.579	0.579	0.533	0.461	0.337	0.193	0.351	0.399	0.201

Notes: The Table displays the OLS and IV estimates of the effect of district-of-birth historical hacienda concentration on literacy across age-cohort groups surveyed in the 1993, 2007, and 2017 population census. The coefficients were estimated in separate district of birth-level regressions, each using one of the eleven different age-cohort samples. Estimates for age-cohort groups 1901-1910 to 1971-1980 were derived from the 1993 census, estimates for 1981-1990 are from the 2007 census, and estimates for age-cohort groups 1991-1999 and 2000-2010 from the 2017 census. The dependent variable is the literacy rate of each age-cohort group. The hacienda variable represents the percentage of the district population residing in haciendas in 1876, standardized to have a mean of zero and a standard deviation of one. Hacienda concentration is instrumented using the soil crop suitability index. All regressions control for geographic covariates (elevation, slope, ruggedness, area, longitude, and latitude), age-cohort group characteristics (age-cohort sample size, squared age-cohort sample size, mother tongue, and gender), and department fixed effects. Robust standard errors are reported in parentheses, and Anderson-Rubin confidence intervals in brackets. Coefficients significantly different from zero are denoted as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A11: OLS and IV Estimates Linking *Hacienda* Concentration to Economic Well-being (1961, 1981)

Dependent Variable:	HH share precarious shelters	Share with toilet	Share with electricity	Share w/o shoes	Log average per capita Income	Child Labor
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: OLS—Economic Well-being (1961, 1981) and *Hacienda* concentration						
Share pop. *Hacienda* 1876	-0.001	0.004	0.009*	0.002	0.032	0.002
	(0.004)	(0.003)	(0.005)	(0.006)	(0.020)	(0.002)
Mean DV	0.15	0.08	0.13	0.62	1.84	0.13
Observations	665	665	665	665	613	613
R-squared	0.551	0.703	0.596	0.729	0.584	0.377
Panel B: IV second stage—Economic Well-being (1961, 1981) and *Hacienda* concentration						
Share pop. *Hacienda* 1876	0.048	-0.043	-0.099	-0.106	-0.097	-0.002
	(0.035)	(0.048)	(0.107)	(0.078)	(0.190)	(0.018)
Upper A–R 95% CI	[0.153;	[0.053;	[0.114;	[0.042;	[0.316;	[0.024;
Lower A–R 95% CI	-0.016]	-0.173]	-0.391]	-0.330]	-0.685]	-0.052]
Kleibergen–Paap F statistic	12.52	12.52	12.52	12.52	8.48	8.48
Observations	665	665	665	665	613	613

Notes: The table presents OLS and IV estimates of the relationship between *historical hacienda* concentration and indicators of economic well-being, measured in 1961 and 1981. Panel A shows OLS estimates of the association between the standardized share of the district population living in *haciendas* in 1876 and six outcomes: the share of households living in precarious shelters (*chozas*), the share with access to a toilet, the share with access to electricity, and the share of individuals not wearing shoes in 1961—all measured at the province level; as well as the log of average per capita income and the share of individuals aged 6–17 engaged in paid work in 1981, measured at the district level. Panel B reports the corresponding second-stage IV estimates using a soil crop suitability index as the instrument for *hacienda* concentration. All regressions control for geographic covariates (elevation, slope, ruggedness, area, latitude, and longitude), demographic and economic covariates from 1961 or 1981 (population, population density, urban share, ethnic composition, and sectoral employment shares), and department fixed effects. Robust standard errors (columns 1–4) and cluster standard errors at the province level (columns 5–6) are reported in parentheses, and Anderson–Rubin 95% confidence intervals are shown in brackets. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A12: OLS and IV Estimates Linking *Hacienda* Concentration to Structural Change (1961)

Dependent Variable:	Employment share in:					Urban
	Agriculture	Manufacture	Services	Trade	Mining	Population share
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: OLS—Structural Change (1961) and *hacienda* concentration						
Share pop. *Hacienda* 1876	0.015*	-0.002	-0.011***	-0.003***	0.011**	-0.051***
	(0.009)	(0.002)	(0.003)	(0.001)	(0.004)	(0.009)
Mean DV	0.74	0.05	0.06	0.03	0.02	0.30
Observations	665	665	665	665	665	665
R-squared	0.311	0.299	0.357	0.290	0.190	0.350
Panel B: IV second stage—Structural Change (1961) and *hacienda* concentration						
Share pop. *Hacienda* 1876	0.032	-0.027	-0.050**	-0.005	0.026	-0.007
	(0.063)	(0.020)	(0.025)	(0.007)	(0.021)	(0.081)
Upper A–R 95% CI	[0.165;	[0.017;	[-0.003;	[0.012;	[0.085;	[0.214;
Lower A–R 95% CI	-0.121]	-0.074]	-0.118]	-0.022]	-0.015]	-0.162]
Kleibergen–Paap F statistic	14.10	14.10	14.10	14.10	14.10	14.10
Observations	665	665	665	665	665	665

Notes: The table presents OLS and IV estimates of the association between historical *hacienda* concentration and structural change, measured in 1961. Panel A reports OLS estimates of the relationship between the standardized share of the district population living in *haciendas* in 1876 and the district-level employment share in five sectors—agriculture, manufacturing, services, trade, and mining—as well as the urban population share. Panel B presents the corresponding second-stage IV estimates, instrumenting *hacienda* concentration using a soil crop suitability index. All regressions include controls for geographic characteristics (elevation, slope, ruggedness, area, latitude, and longitude), 1961 demographic and economic variables (population, population density, urban share, ethnic composition, and sectoral employment shares), and department fixed effects. Robust standard errors are reported in parentheses, and Anderson–Rubin 95% confidence intervals are shown in brackets. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A13: Historical hacienda concentration and economic and health outcomes

Dependent variable:	**Log nighttime lights (2016)**	**Poverty rate (%) (2007)**	**Per capita expenditure (PEN) (2009)**	**Stunted growth kids 6–9 (1993)**	**Child mortality (2007)**
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS — Economic and Health Outcomes and Hacienda Concentration					
Share pop. *Hacienda* 1876	-0.013	-0.068	-0.004	0.001	-0.738***
	(0.009)	(0.451)	(0.009)	(0.065)	(0.239)
Observations	665	665	665	599	665
R-squared	0.783	0.828	0.723	0.464	0.611
Mean DV	3.10	60.48	5.86	0.56	24.82
Panel B: IV Second Stage — Economic and Health Outcomes and Hacienda Concentration					
Share pop. *Hacienda* 1876	-0.073	1.972	-0.030	-0.099	-3.443
	(0.110)	(4.010)	(0.069)	(0.065)	(2.536)
Upper A-R 95% CI	[0.148;	[11.676;	[0.127;	[0.004;	[1.079;
Lower A-R 95% CI	-0.401]	-7.434]	-0.21]	-0.338]	-11.518]
Kleibergen-Paap F statistic	10.16	11.31	11.31	10.59	11.31
Observations	665	665	665	599	665
R-squared	0.769	0.821	0.719	0.201	0.543

Notes: The dependent variable in column 1 is the logged value of the average remote-sensed nighttime light emissions for each district in 2016, sourced from NASA's Black Marble VIIRS nighttime light data (accessed on 02/07/2024 at https://earthobservatory.nasa.gov/). In column 2, the dependent variable is the district's prevalence of stunted growth in children aged 6 to 9 in 1993, defined as children whose heights are more than 2 standard deviations below the age-specific median (National School Height Census 1993, Ministry of Education). Column 3's dependent variable is the child mortality rate in 2007, while column 4 uses the monetary poverty rate (both variables are sourced from the 2007 Poverty Map, INEI). In column 5, the dependent variable is the logged value of the average per capita expenditure in each district in 2009, sourced from the 2009 Poverty Map (INEI, 2009). Historical hacienda concentration is measured by the percentage of the district population residing in haciendas in 1876, standardized to have a mean of zero and a standard deviation of one. Hacienda concentration is instrumented using the soil crop suitability index. All regressions control for geographic covariates (elevation, slope, ruggedness, area, longitude, and latitude), contemporaneous demographic covariates (population, population density, urban ethnic composition, and shares of agriculture, mining, and manufacturing), and department fixed effects. Robust standard errors are reported in parentheses and Anderson-Rubin confidence intervals in brackets. Coefficients significantly different from zero are denoted by: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A14: Long-term effect of historical Hacienda concentration on Farming Conditions

Dependent variable:	Farm size		Inequality	Education	Technology and Inputs		
	(1) Average farm size	(2) Share <1ha	(3) Gini (ag. land)	(4) Farmers' years of educ.	(5) Electrification	(6) Fertilizer use	(7) Insecticide use
Panel A: OLS — Agriculture and Hacienda concentration							
Share pop. Hacienda 1876	0.310***	0.000	0.014**	-0.104***	-0.000	0.009	0.002
	(0.088)	(0.008)	(0.006)	(0.032)	(0.001)	(0.009)	(0.010)
Observations	665	665	662	663	665	665	665
R-squared	0.417	0.473	0.220	0.732	0.328	0.566	0.527
Mean DV	2.649	0.414	0.552	4.706	0.006	0.393	0.430
Panel B: IV Second Stage — Agriculture and Hacienda concentration							
Share pop. Hacienda 1876	-0.120	-0.043	-0.005	-1.830**	0.009	-0.118	-0.143
	(0.680)	(0.071)	(0.047)	(0.735)	(0.010)	(0.098)	(0.095)
Upper A-R 95% CI	1.327	0.118	0.112	-0.840	0.036	0.068	0.023
Lower A-R 95% CI	-2.014	-0.231	-0.137	-5.630	-0.011	-0.426	-0.469
Kleibergen-Paap F-statistic	10.41	10.41	8.18	8.17	10.41	10.41	10.41
Observations	665	665	662	663	665	665	665
R-squared	0.376	0.443	0.200	-0.372	0.067	0.437	0.345

Notes: The table presents OLS and IV estimates of the association between historical hacienda concentration and district-level farming outcomes from the 1994 Agricultural Census. Panel A reports OLS estimates, while Panel B reports second-stage IV estimates using a soil crop suitability index as an instrument for hacienda concentration, measured as the standardized share of the district population residing in haciendas in 1876. The dependent variables are: (1) average farm size (hectares of agricultural land); (2) share of farms with less than one hectare; (3) agricultural land Gini coefficient (INEQDECO module, Jenkins 1999); (4) farmers' mean years of education; (5–7) share of farmers using electric power, fertilizers, and insecticides. All regressions control for geographic covariates (elevation, slope, ruggedness, area, latitude, longitude), 1993 demographic characteristics (population, density, urban share, indigenous language), and 1994 farm characteristics (female-headed farms, logged number of farms). Department fixed effects are included. Standard errors are in parentheses. Anderson–Rubin confidence intervals in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A15: OLS and IV Estimates of Hacienda Concentration and Indigenous Voting registration (1933)

	(1)	(2)	(3)
Dep. Var.: Indigenous share of registered voters (1933)			
Panel A: OLS results			
Share pop. Hacienda 1876	-0.022*	-0.017	-0.000
	(0.011)	(0.014)	(0.010)
R-squared	0.503	0.187	0.611
Mean DV	0.33	0.33	0.33
Panel B: IV Second Stage results			
Share pop. Hacienda 1876	-0.282***	-0.510*	-0.168*
	(0.087)	(0.267)	(0.091)
Upper A-R 95% CI	-0.141	-0.123	-0.005
Lower A-R 95% CI	-0.543	-1.736	-0.418
Kleibergen-Paap F statistic	17.54	8.88	16.37
Department FE	Yes	No	Yes
Geographic controls	No	Yes	Yes
Demographic controls	No	No	Yes
Clustered SE	Yes	Yes	Yes
Clusters (provinces)	101	101	101
Observations	665	665	665

Notes: The table presents OLS and IV estimates of the relationship between historical hacienda concentration and indigenous political participation, measured by the share of registered voters identified as indigenous in the 1933 electoral registry. Panel A reports OLS estimates, while Panel B presents second-stage IV estimates using a soil crop suitability index as an instrument for hacienda concentration, defined as the standardized share of the district population residing in haciendas in 1876. Column (1) includes department fixed effects only; column (2) adds geographic and demographic controls but omits fixed effects; and column (3) includes all controls. Geographic covariates include elevation, slope, ruggedness, area, latitude, and longitude. Demographic controls include 1940 population, population density, urban share, ethnic composition, and sectoral employment shares. Standard errors clustered at the province level are reported in parentheses, and Anderson–Rubin confidence intervals are presented in brackets. Coefficients significantly different from zero are denoted by: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A16: OLS and IV Estimates Linking *Hacienda* Concentration and Literacy by Natural Region

Dep. var.: literacy rate (share)	1876	1940	1961	1981	1993	2007	2017
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: OLS—literacy and *hacienda* concentration							
Share pop. Hac. 1876 * Coast	-0.033***	-0.020***	-0.001	0.004	0.005*	-0.001	-0.001
	(0.009)	(0.006)	(0.009)	(0.004)	(0.003)	(0.002)	(0.001)
Share pop. Hac. 1876 * Highlands	-0.011***	-0.028***	-0.029***	-0.020***	-0.014***	-0.005**	-0.003*
	(0.004)	(0.005)	(0.005)	(0.005)	(0.003)	(0.002)	(0.002)
Observations	665	665	665	613	665	665	665
R-squared	0.556	0.871	0.828	0.800	0.800	0.790	0.785
Mean DV	0.126	0.334	0.481	0.679	0.755	0.858	0.869
Panel B: IV second stage—literacy and *hacienda* concentration							
Share pop. Hac. 1876 * Coast	-0.068*	-0.071**	-0.125***	-0.048*	-0.025	-0.020*	-0.013
	(0.041)	(0.036)	(0.046)	(0.027)	(0.018)	(0.011)	(0.008)
Share pop. Hac. 1876 * Highlands	-0.050**	-0.156***	-0.274***	-0.153***	-0.130***	-0.068**	-0.044**
	(0.024)	(0.047)	(0.081)	(0.053)	(0.043)	(0.026)	(0.019)
Kleibergen–Paap F statistic	10.36	7.31	6.12	5.01	4.65	4.89	3.96
Observations	665	665	665	613	665	665	665

Notes: The table reports OLS and IV estimates of the relationship between historical *hacienda* concentration and literacy rates, disaggregated by natural region. The dependent variable is the literacy rate (share) in census years between 1876 and 2017. Panel A reports OLS estimates from regressions that interact *hacienda* concentration in 1876 with a dummy for whether the district belongs to the coast or the highlands. Panel B presents second-stage IV estimates using the same specification, where *hacienda* concentration is instrumented with a soil crop suitability index. *Hacienda* concentration is defined as the standardized share of the district population living in *haciendas* in 1876. All regressions include geographic controls (elevation, slope, ruggedness, area, latitude, and longitude), demographic and economic controls from the corresponding census year (population, population density, urban share, ethnic composition, and employment shares by sector), and department fixed effects. Robust standard errors are reported in parentheses. Coefficients significantly different from zero are denoted by: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A17: OLS and IV Estimates Linking Hacienda Concentration and Literacy- The Mitigating Role of Hacienda Fragmentation

	Hacienda Fragmentation Index				
Interaction term:	>70th	>75th	>80th	>85th	>90th
Interaction term:	percentile	percentile	percentile	percentile	percentile
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS — Literacy and Historical Hacienda Concentration					
Dep. var.: Literacy rate (share) in 1940					
Share pop. Hacienda 1876	-0.030***	-0.029***	-0.029***	-0.029***	-0.030***
	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)
Interaction term	0.008	0.005	0.008	0.010	0.016*
	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)
Observations	665	665	665	665	665
R-squared	0.870	0.870	0.870	0.870	0.871
Mean mitigating factor	0.30	0.25	0.20	0.15	0.10
Panel B: IV Second Stage — Literacy and Historical Hacienda Concentration					
Dep. var.: Literacy rate (share) in 1940					
Share pop. Hacienda 1876	-0.253***	-0.216***	-0.168***	-0.156***	-0.152***
	(0.094)	(0.077)	(0.052)	(0.043)	(0.040)
Interaction term	0.215**	0.166**	0.105**	0.101***	0.123***
	(0.096)	(0.077)	(0.047)	(0.036)	(0.036)
Kleibergen-Paap F statistic	4.38	5.12	6.89	8.50	8.48
Observations	665	665	665	665	665
R-squared	0.524	0.595	0.681	0.692	0.694

Notes: The table presents OLS and IV estimates of the effect of historical hacienda concentration on literacy rates in 1940, interacted with the degree of hacienda fragmentation. Hacienda concentration is measured as the standardized share of the district population residing in haciendas in 1876. Fragmentation is measured as the inverse of the Herfindahl-Hirschman Index (HHI), where each hacienda's share corresponds to its proportion of the total resident population in the district. The interaction term captures whether districts above each fragmentation percentile threshold experience differential effects. Robust standard errors are reported in parentheses. Statistical significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A18: Mita and Hacienda Concentration – Dell(2010)'s Subsamples

	Dell's Sample		
	< 100 km of Bound. (1)	< 75 km of Bound. (2)	< 50 km of Bound. (3)
Dep. var.: Population in Haciendas 1876 (share)			
Inside Mita Catchment (dummy)	-0.424*	-0.525**	-0.444*
	(0.252)	(0.244)	(0.260)
Observations	200	151	116
R-squared	0.278	0.329	0.334

Notes: The dependent variable is the share of the resident population living in haciendas in 1876. Column 1 restricts the sample to districts located within 100 km of the mita boundary, as defined in Dell (2010); columns 2 and 3 further restrict the sample to within 75 km and 50 km, respectively. The main explanatory variable is a binary indicator for whether the district is located inside the mita catchment area. Additional control variables include altitude, slope, logged 1876 population, urban share, and share of indigenous and mestizo population. Robust standard errors are reported in parentheses. Coefficients significantly different from zero are denoted by: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A19: Hacienda Concentration and Literacy - Dell(2010)'s Subsamples

	Dell(2010)'s Sample			**Random Sample**		
	< 100 km	< 75 km	< 50 km	< 100 km	< 75 km	< 50 km
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: OLS — literacy and hacienda concentration						
Dep. var.: literacy rate 1940 (share)						
Share pop. Hacienda 1876	-0.005	0.002	0.007	-0.029***	-0.028***	-0.028**
	(0.006)	(0.006)	(0.006)	(0.006)	(0.009)	(0.011)
Observations	200	151	116	201	152	117
R-squared	0.906	0.850	0.805	0.868	0.885	0.893
Mean DV	0.22	0.18	0.17	0.38	0.39	0.40
Panel B: IV second stage — literacy and hacienda concentration						
Dep. var.: literacy rate 1940 (share)						
Share pop. Hacienda 1876	-0.101	-0.106	-0.094	-0.125**	-0.109*	-0.183
	(0.067)	(0.116)	(0.114)	(0.061)	(0.063)	(0.151)
Kleibergen-Paap F statistic	3.48	1.12	0.99	5.35	3.21	1.15
Observations	200	151	116	201	152	117
R-squared	0.780	0.546	0.407	0.713	0.798	0.641
Mean DV	0.22	0.18	0.17	0.38	0.39	0.40

Notes: The dependent variable is the literacy rate (share) in 1940. Columns 1 to 3 restrict the sample to districts located within 100 km, 75 km, and 50 km of the *mita* boundary, respectively, as defined in Dell (2010). Columns 4 to 6 use random samples of districts matched in size to those in columns 1 to 3. The main explanatory variable—hacienda concentration—measures the standardized share of the district population living in haciendas in 1876. This variable is instrumented using a soil crop suitability index. All regressions control for geographic covariates (elevation, slope, ruggedness, area, longitude, and latitude), 1940 demographic and economic covariates (population, population density, urban share, ethnicity share, agricultural, mining, and manufacturing share), and department fixed effects. Robust standard errors are reported in parentheses. Coefficients significantly different from zero are denoted by: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

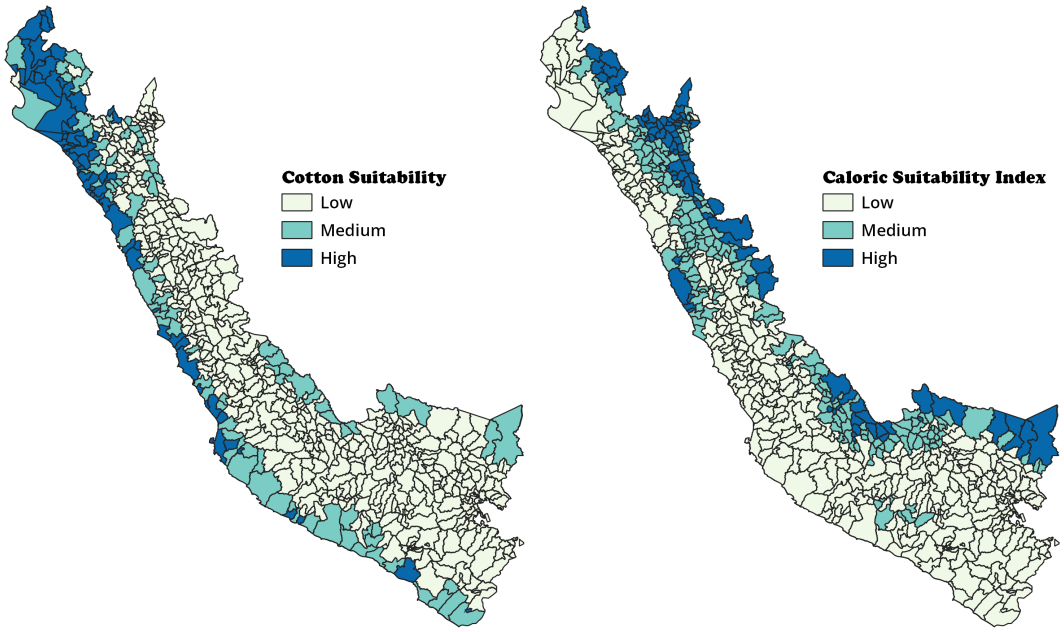
Table A20: Hacienda Concentration, Literacy and Public Goods — Dell(2010)'s Subsamples

Dependent Variable as heading:	Literacy Rate (1876) (1)	Literacy Rate (1940) (2)	Schools per 10K (1918) (3)	Teachers per 10K (1918) (4)	Log Dist. Nat. Road (1950) (5)	Log Dist. Nat. Road (1975) (6)	Log Dist. Nat. Road (2000) (7)
Sample <100 km. of Bound.							
Share pop. Hacienda 1876	-0.001	-0.005	-0.961***	-0.817**	0.163*	-0.221	0.362
	(0.004)	(0.006)	(0.318)	(0.319)	(0.088)	(0.146)	(0.367)
Observations	200	200	200	200	200	200	200
R-squared	0.751	0.906	0.820	0.855	0.758	0.767	0.501
Mean DV	0.08	0.22	10.92	12.38	11.58	10.35	6.43
Sample <75 km. of Bound.							
Share pop. Hacienda 1876	-0.002	0.002	-0.761*	-0.584	0.081	0.006	0.664*
	(0.005)	(0.006)	(0.394)	(0.405)	(0.103)	(0.103)	(0.373)
Observations	151	151	151	151	151	151	151
R-squared	0.614	0.850	0.803	0.819	0.776	0.746	0.492
Mean DV	0.06	0.18	10.61	11.70	11.78	11.45	7.34
Sample <50 km. of Bound.							
Share pop. Hacienda 1876	-0.006	0.007	-0.719*	-0.535	-0.012	-0.090**	0.575
	(0.005)	(0.006)	(0.408)	(0.422)	(0.021)	(0.035)	(0.370)
Observations	116	116	116	116	116	116	116
R-squared	0.459	0.805	0.763	0.798	0.925	0.913	0.534
Mean DV	0.05	0.17	10.56	11.76	12.04	11.78	7.84

Notes: The table presents OLS estimates of the relationship between historical hacienda concentration and various long-term outcomes across different district subsamples based on proximity to the mita boundary, following the approach in Dell (2010). The main explanatory variable is the standardized share of the district population living in haciendas in 1876. Columns report estimates for the following dependent variables: stunted growth among children aged 6–9 in 1993 (col. 1), literacy rate in 1940 (col. 2), number of schools per 10,000 people in 1918 (col. 3), number of teachers per 10,000 people in 1918 (col. 4), log distance to the national road network in 1950 (col. 5), and log distance to the national road network in 1975 (col. 6). Rows correspond to three spatial subsamples: districts within 100 km, 75 km, and 50 km of the mita boundary. All regressions control for geographic covariates (elevation, slope, ruggedness, area, longitude, and latitude), contemporaneous demographic and economic covariates (population, population density, urban share, ethnicity share, agricultural, mining, and manufacturing share), and department fixed effects. Robust standard errors are reported in parentheses. Coefficients significantly different from zero are denoted as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

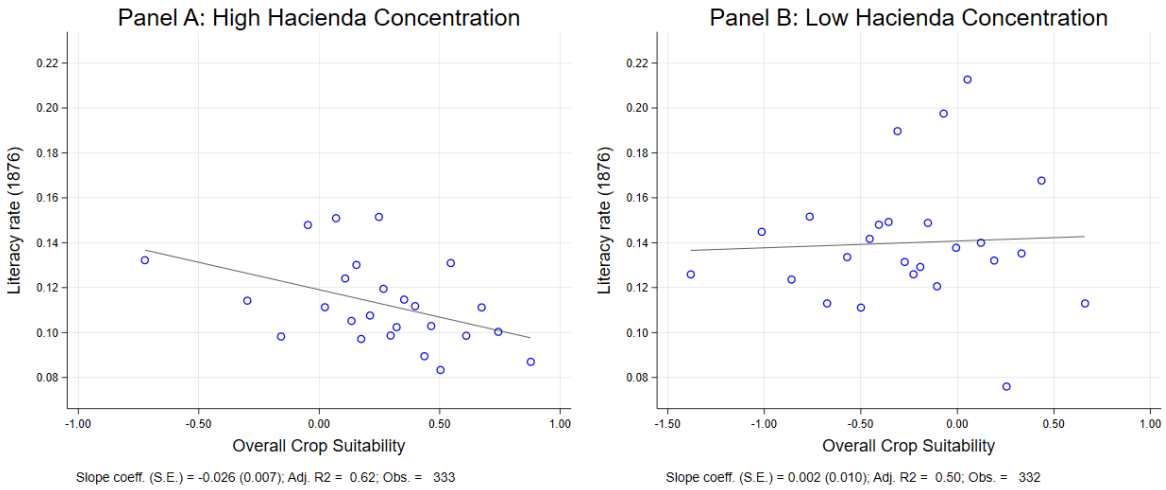
Figures

Figure A1: Soil Suitability Indexes



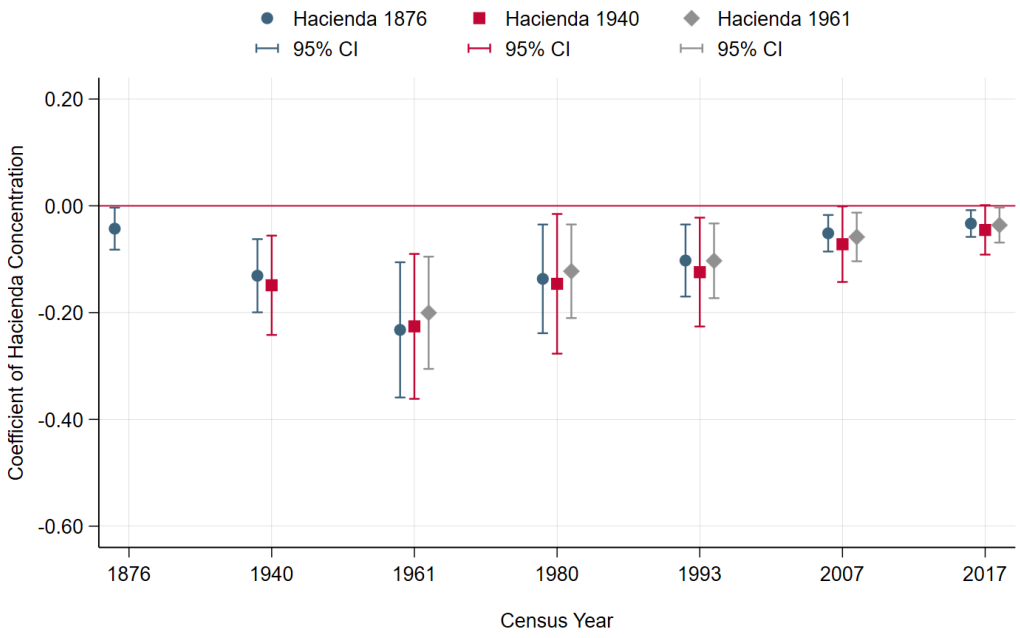
Notes: The data source for Cotton Suitability Index is the Global Agro-Ecological Zones (GAEZ) data from FAO. It captures the district-level average suitability of cropland for Cotton under irrigation conditions and high input level and with CO2 fertilization using climate data source CRUTS32. This data is available at 5 arc-minutes (about 9 x 9 km) and was averaged at the district-level according to the boundaries in place in 1876. The data source for the Caloric Suitability Index is Galor and Özak (2016). This measure captures the potential agricultural output (measured in calories) based on crops that were available for cultivation. Both measures exhibit a weak negative correlation equal to -0.07 (p-value=0.069).

Figure A2: Soil Crop Suitability and Literacy by Hacienda Concentration



Notes: The figures display binned scatter plots showing the relationship between districts' average soil crop suitability and literacy rates in 1876, controlling for geographic covariates (elevation, slope, ruggedness, area, longitude, and latitude), contemporaneous demographic covariates (population, population density, urban share, ethnicity share, agricultural, mining, and manufacturing share), and department fixed effects. Panel A shows this relationship for the sample of 333 districts with high hacienda concentration, defined as districts where the share of the population living in haciendas is above the median value (ranging from 7% to 94%, with a mean value of 29%). Panel B shows this relationship for the 332 districts with low hacienda concentration (where the hacienda share ranges from 0% to 7%, with a mean of 1%). Literacy is defined as the proportion of individuals aged 5 or older who could read, write, or both. The soil crop suitability index measures the maximum standardized cropland suitability between cotton (GAEZ) and other caloric-intensive crops available for cultivation (e.g., potato, maize) (Galor and Özak, 2016). The binned scatter plot divides the x-axis variable into 25 equal-sized bins, calculates the mean values of the x-axis and y-axis variables within each bin, and then plots these data points.

Figure A3: The Persistence of Hacienda Effect on Literacy



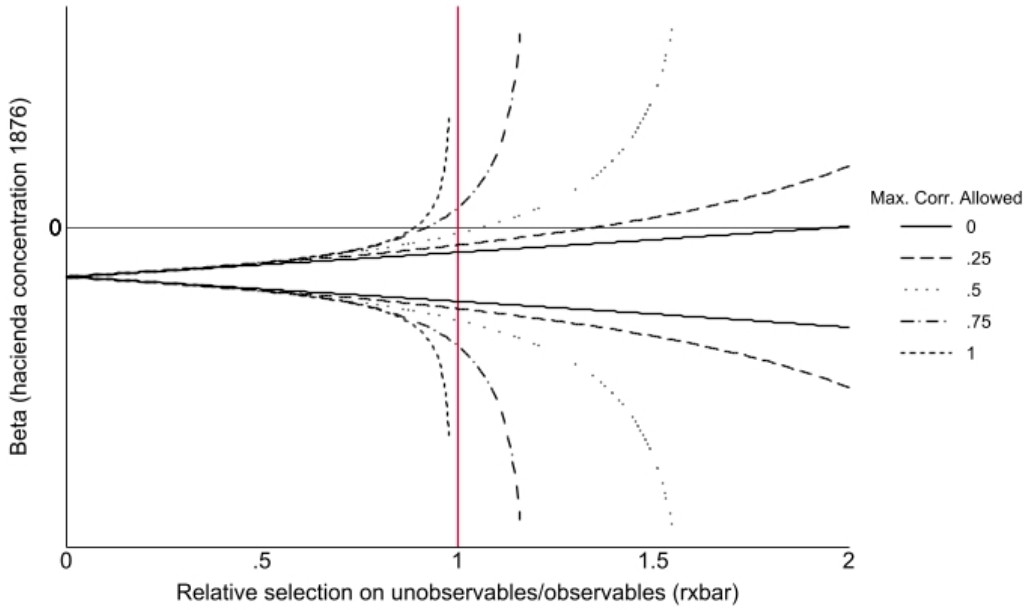
Notes: The figure displays district-level IV estimates of the effect of hacienda concentration in 1876, 1940, and 1961 on literacy rates across seven population censuses. The coefficients were estimated from separate district-level cross-sectional regressions, where the dependent variable is the literacy rate for each census year. The effect of hacienda concentration in 1876 on contemporary and subsequent literacy rates is represented by blue circles, the effect of hacienda concentration in 1940 by red squares, and the effect of hacienda concentration in 1961 by grey diamonds. These measures reflect the percentage of the district population residing in haciendas during each period, standardized to have a mean of zero and a standard deviation of one. The hacienda concentration measures are instrumented using the soil crop suitability index. All regressions control for geographic covariates (elevation, slope, ruggedness, area, longitude, and latitude), contemporaneous demographic covariates (population, population density, urban share, ethnic composition, and shares of agriculture, mining, and manufacturing), as well as department fixed effects. Robust 95% confidence intervals are depicted by blue, red, and grey lines corresponding to hacienda concentration in 1876, 1940, and 1961, respectively.

Figure A4: Gendered Effects of Hacienda Concentration on Literacy



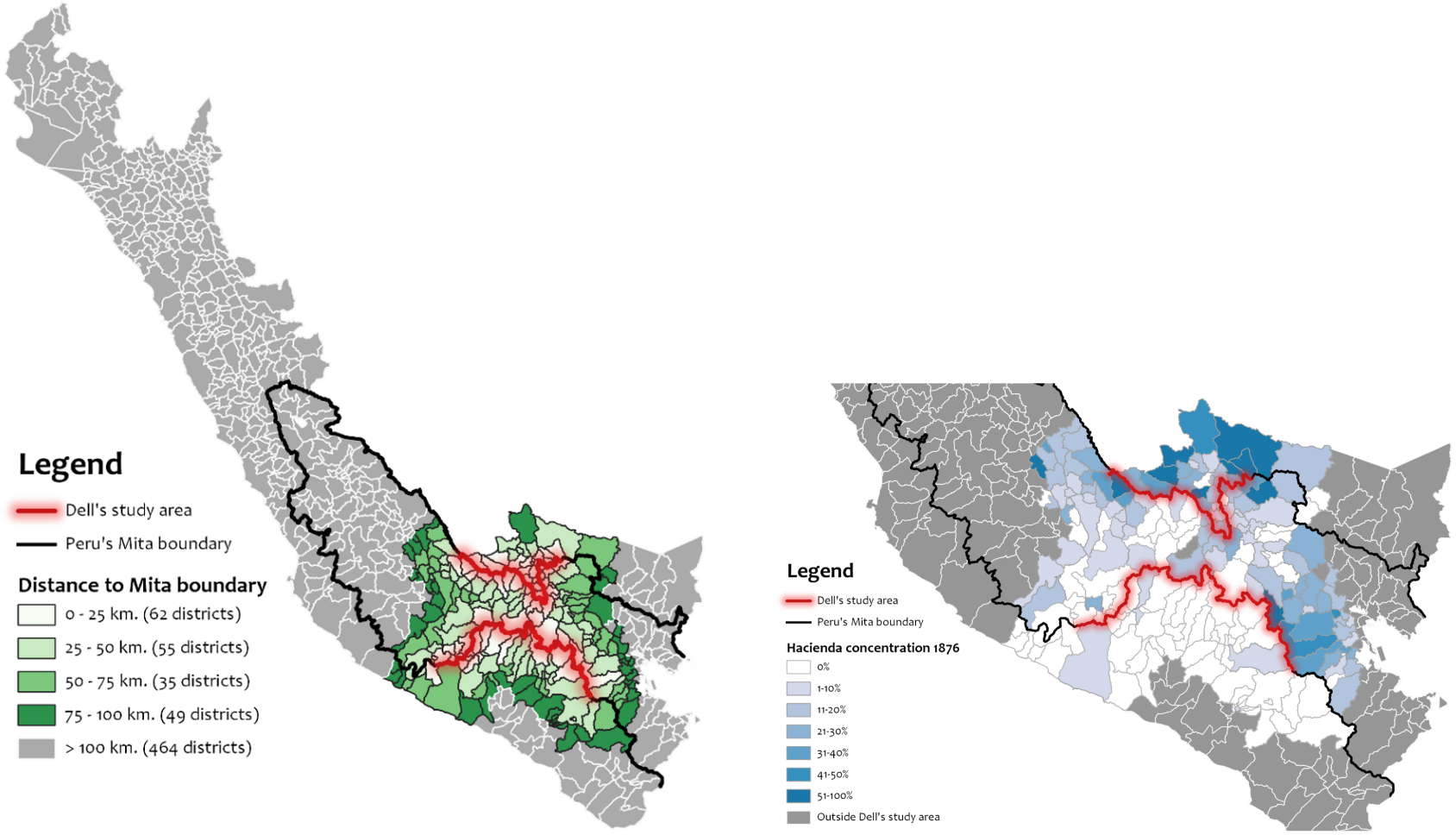
Notes: The figure presents IV estimates of the effect of historical hacienda concentration on male and female literacy rates across seven population censuses. The coefficients were estimated from separate cross-sectional regressions at the district level, following the boundaries in effect in 1876. The effect on female literacy is shown by red circles, while the effect on male literacy is represented by blue circles. Historical hacienda concentration is measured by the percentage of the district population residing in haciendas in 1876, standardized to a mean of zero and a standard deviation of one. Hacienda concentration is instrumented using the soil crop suitability index. All regressions control for geographic covariates (elevation, slope, ruggedness, area, longitude, and latitude), demographic covariates at each census period (population, population density, urban share, ethnic composition, and shares of agriculture, mining, and manufacturing), and department fixed effects. Anderson-Rubin 95% confidence intervals are illustrated with red and blue lines corresponding to female and male literacy rates, respectively. The Kleibergen-Paap F-statistics range from 9.8 to 21.8 across all cross-sectional regressions.

Figure A5: Regression Sensitivity Analysis (DMP 2022)



Notes: following Oster (2019), Diegert et al. (2022) and Masten and Poirier (2022), the graph plots estimated bounds on regression coefficients by relaxing the assumption that the omitted variables are uncorrelated with the included variable of interest. Altitude control is used to calibrate deviations from identifying assumptions. Bounds are calculated for a range of alternative assumptions about the allowed correlation between observable and omitted variables. In our main specification, taking Oster (2019) bounds (max. corr=0), the relative selection on unobservables/observables would need to have twice the explanatory power of the already included covariates to make the estimated effect of hacienda concentration on literacy zero.

Figure A6: Dell(2010)'s study area and 1876 haciendas around Mita boundary

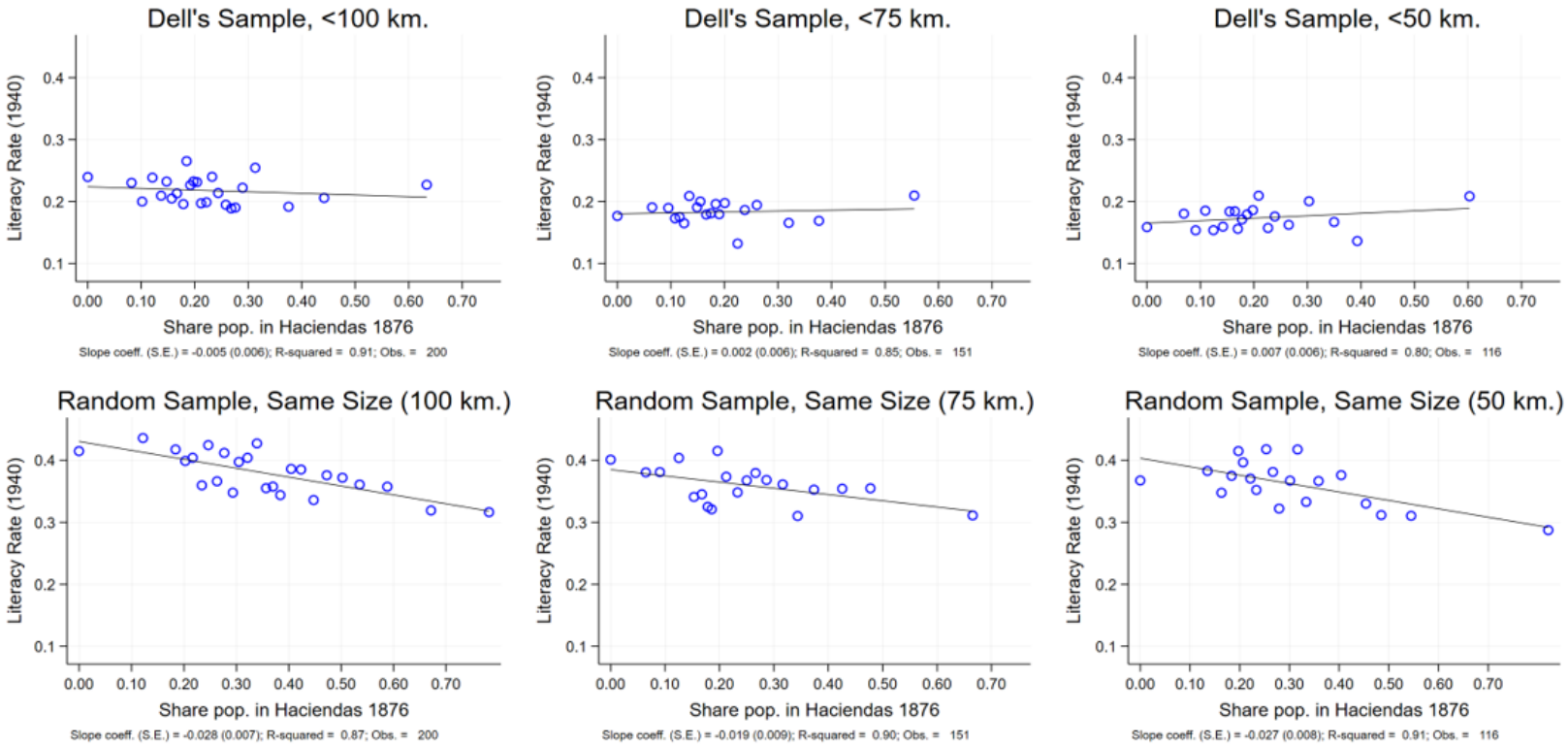


(a) Dell(2010)'s study area

(b) Hacienda concentration 1876

Notes: Panel A displays Dell (2010)'s study area in southern Peru and classifies districts based on their distance to the mita boundary. Panel B overlays the same study area with the distribution of hacienda concentration in 1876, measured as the share of the district population living in haciendas. The mita boundary is shown in black, while the boundary of Dell's study is outlined in red. Districts outside the study area are shaded in gray.

Figure A7: Hacienda Concentration and Literacy according to Dell(2010)'s sumsamples



Notes: The figure presents binned scatter plots illustrating the relationship between hacienda concentration in 1876 and literacy rates in 1940 across Dell (2010)'s subsamples. The top row shows results for districts located within 100 km, 75 km, and 50 km of the mita boundary, respectively. The bottom row shows corresponding plots for randomly selected district samples matched in size to the top-row subsamples. The x-axis measures the share of the district population living in haciendas in 1876, and the y-axis shows the literacy rate in 1940. Each plot includes a fitted linear regression line with slope coefficients and robust standard errors reported below.