



# Landownership concentration and human capital accumulation in post-unification Italy

Vitantonio Mariella<sup>1,2</sup> 

Received: 21 December 2020 / Accepted: 3 May 2022  
© The Author(s) 2022

## Abstract

This paper contributes to the debate on institutions and economic development by assessing the relationship between landownership concentration and education. Using historical data at both the district and province levels in post-unification Italy from 1871 to 1921, I find evidence of an adverse effect of land inequality on literacy rates. Instrumental variable estimates using malaria pervasiveness as a source of exogenous variation rule out concerns regarding potential endogeneity. Exploration of the panel dimension of the data reveals that several shocks during this period affected the relationship between land inequality and literacy rates. In addition, this paper provides insights into the mechanism behind this relationship by analyzing the impact on intermediate outputs, such as enrollment rates in primary school, child-teacher ratio, school density, child labor, and municipality expenditures. Land inequality may have adversely affected literacy rates not only by influencing schooling supply through the political process but also through the private demand for education.

**Keywords** Land inequality · Human capital · Political power · Post-unification Italy

**JEL classification:** I25 · J24 · N54 · O43

---

*Responsible editor:* Oded Galor

---

✉ Vitantonio Mariella  
vitantonio.mariella@irpps.cnr.it

<sup>1</sup> IRPPS, Italian National Research Council, Via Palestro, 32-00185 Rome, Italy

<sup>2</sup> Department of Economics and Law, Sapienza University, Via del Castro Laurenziano, 9-00185 Rome, Italy

# 1 Introduction

Countries characterized by extractive institutions and the concentration of political power in the hands of small elites tend to display a pattern of limited long-term economic growth (Acemoglu et al. 2001, 2005). Recent studies have also shown that human capital is essential in explaining the link between institutions and economic development (Gennaioli et al. 2012; Glaeser et al. 2004). Under the assumption that political power is linked to landownership concentration in pre-industrial economies, assessing the role of land inequality becomes fundamental for understanding the path of development of countries and regions.

This paper aims to analyze the relationship between land inequality and literacy rates in the late nineteenth century and the early twentieth century in Italy.<sup>1</sup> I build a new dataset at both the district and province levels in Italy from 1871 to 1921, which allows me to exploit both the temporal and cross-sectional dimensions of the relationship between landownership concentration and education. I estimate several regressions at each point during the considered period, showing how this relationship seems to be fairly stable over time. However, when I explore the panel dimension of the dataset, the relationship vanishes over time. This was likely due to different institutional shocks such as population enfranchisement and polity regime changes to implement public policies (i.e., decentralization vs centralization). The most important shock was the introduction of the *Daneo-Credaro* law in 1911, which transferred the decisions to invest in public schooling from municipalities to the central government, thus limiting the political power of local notables. By interacting the initial level of landownership concentration with time shocks, I employ a stepwise difference-in-differences strategy to account for the differential effect of those institutional changes on areas with a different distribution of political power.

The dataset's cross-sectional and panel structure creates several advantages in terms of identification strategy. The panel dimension allows me to control for time-invariant unobserved heterogeneity, whereas the cross-sectional dimension permits me to exploit the advantages of an instrumental variable approach that helps identify a causal relationship between landownership concentration and education levels. More specifically, I use malaria pervasiveness as a source of exogenous variation in land inequality. Extensive research documents the link between *latifundia* and malaria. This thesis was put forth by Celli and Fraentzel (1933), who linked malaria endemicity with the creation of large estates, a pattern observed for centuries in the Pontine marshes, a plain in the surroundings of Rome. At the same time, extensive literature documents the greater capacity of large farmers to cope with the risks of malaria-infested areas (see Snowden (2008)). They were forced to manage adverse shocks and adopt appropriate incentive-compatible strategies, while rural workers were discouraged from living close to the lands, thus influencing their settlement pattern.

<sup>1</sup> For a more comprehensive view of the idea of scarcity of human capital as a drag to modern economic growth in Italy, see Di Martino and Vasta (2017).

I find the following results. First, separate cross-section estimates at different points in time reveal a negative impact of the measure of land inequality on literacy rates. Second, IV estimates, using malaria as an instrument, rule out concerns regarding potential endogeneity. Third, the validity of the exclusion restriction hypothesis is addressed in Section 5.3. Although malaria may affect education outcomes through health channels, robustness checks demonstrate that the main findings are hardly changed under second-order assumptions. Finally, once I explore the panel dimension of the dataset, further evidence to support the negative association between land inequality and education levels is provided, albeit it seems to vanish over time, in line with the “passive modernization process” Italy was facing at that time (Felice 2013).

The main contribution is twofold. First, although much literature at a global level has already documented the impact of extractive institutions and inequality on growth factors, I find that the epidemiological environment is a strong predictor of historical land inequality. Specifically, this paper demonstrates that malaria virulence was a major driver of landownership concentration in Italy. Second, I explore the mechanism of the effect transmission behind the observed relationship to distinguish between the impact of supply and demand factors. I argue that landownership concentration may have adversely affected literacy rates, not only influencing the supply of schooling through the political process but also affecting the private demand for education of landless peasants.

A vast amount of literature has primarily focused on the role of landowning elites in preventing the poor majority from gaining education and power. Assuming that large landowners rule local municipalities and acquire decision autonomy to invest in human capital, as in Italy in the post-unification decades, their interest lies in the vote against the expansion of public spending on mass schooling. I gather information on the number of public schools and the corresponding appointed teachers and municipal expenditures and restrict the sample to rural areas with low urbanization rates to test the relationship between supply factors in education and land inequality. I find partial support for the hypothesis that the transmission mechanism worked through political voice. Consequently, while local notables, on the one hand, might have low returns to invest in public schooling, on the other hand, they could have burgeoning stakes in the industrial sector.<sup>2</sup> However, since the industrialization process presumably made for a greater demand for education (see Galor and Moav (2006)), I assess whether demand factors played a role too. Using data on enrollment rates and child labor, I find that the poor majority, which faces tightening budget constraints and limited opportunities for upward mobility and offering their labor outside agriculture, had fewer incentives to invest in human capital. Hence, land inequality affects human capital accumulation in diverse ways. This influence not only operates through the role of extractive institutions blocking the supply of public schooling infrastructures but also via demand effects.

I proceed as follows. In Section 2, I survey the corresponding literature review. In Section 3, I briefly provide a historical background of post-unification Italy.

<sup>2</sup> Henceforth, large landowners, landowning elites, and local notables will assume the same meaning.

Section 4 describes data, while Section 5 outlines the econometric methodology and reports the main empirical results, focusing on the role of malaria as a source of exogenous variation of *latifundia*. Moreover, I conduct several robustness checks, enabling the identification of a plausibly causal relationship between land inequality and education. Particular attention is also devoted to panel estimates, with a step-wise diff-in-diffs approach. Section 6 focuses on the different possible channels of transmission of the effect. Finally, Section 7 provides concluding remarks.

## 2 Literature review

The literature on the role of inequality and its long-run effect on human capital is vast and presents broad empirical support. Galor et al. (2009) build a theoretical model where capitalists, on the one hand, push for policies aimed to promote the education of the masses to have a skilled labor force, whereas landowners, on the other hand, support policies that deprive the masses of education. The interest of landowners is to reduce labor mobility to keep wages low and have an available labor force in the countryside. That is especially true in later stages of development, and industrialization has already emerged, increasing the return to education and the incentives to invest in human capital (Galor 2011). Besides, Galor et al. (2009) assess the inverse relationship between land inequality and education, drawing from the theoretical model an empirically testable prediction tested using US data for the 1900–1940 period, finding that changes in landownership concentration negatively affect education expenditures. Moreover, to demonstrate the causality of the effect, they exploit as an instrument the interaction between changes in the relative price of crops and climatic characteristics across states.

Ramcharan (2010) investigates the effect of land inequality on redistributive policies. By looking at US census data for the 1890–1930 period, he finds a negative association between inequality and expenditures in education, identifying a causal relationship using geographic variables as a source of exogenous variation in land inequality.<sup>3</sup> Similarly, Vollrath (2013) shows a negative correlation between landownership concentration and taxes for funding public schooling in US counties in 1890. The same results can be found in Chaudhary (2009), who links low public spending on primary schooling with greater castes' interests in India.<sup>4</sup>

At the same time, Engerman and Sokoloff (1997); Sokoloff and Engerman (2000) assert that Latin American elites blocked the expansion of mass schooling and suffrage enlargement to retain their political power. Go and Lindert (2010) aim to explain differences in enrollment rates between the North and the South of the USA by looking at local autonomy and more widespread access to political voice for the masses in the North. Also, they analyze the effect of enfranchisement on tax levels for public schooling and on enrollment rates. Along the same line of research,

<sup>3</sup> More specifically, the author uses data for crop choice, rainfall intensity, and surface elevation.

<sup>4</sup> The same results have been found for Korea and UK, respectively, by Go and Park (2012) and Goni (2015).

focusing on estimating the correlation between schooling and education levels with inequality in political power, Mariscal and Sokoloff (2000) find that inequality in Latin America is negatively associated with enrollment and literacy rates. Besides, they argue that the extension of the franchise promoted mass schooling, a result also found by Acemoglu and Robinson (2000) and Gallego (2010). The latter states that democratization enhances primary education while decentralizing political power impacts only secondary and higher education levels. At odds with these latter findings, Aghion et al. (2019) argue that the role of democratization is irrelevant to explaining higher education levels, and they propose military rivalry as the real incentive for countries to invest in mass schooling.

Recently, another body of literature has called this view into question, as incentives to invest in human capital entering individuals' utility functions and concerns regarding their limited budget constraints may be considered non-mutually exclusive channels of transmission of the effect. Indeed, a more equally distributed land ownership may have contributed to unleashing resources to invest in acquiring literacy and numeracy skills (Cinnirella and Hornung 2016; Beltrán Tapia and Martínez-Galarraga 2018), leaving the supply-side mechanism linked to the political process aside.<sup>5</sup> Overall, less empirical research has been carried out on the demand-based viewpoint concerning the historical link between land inequality and human capital. The first attempt dates back to Galor and Zeira (1993), which highlights that in the presence of frictions in the credit market, unequal distribution of resources can have a long-lasting effect on the lower class' ability to invest in human capital. Individual underinvestment in human capital resulting from capital market imperfections is also found by Deininger and Squire (1998). At the same time, Reis (2005) provides a somewhat similar explanation of disparities in education levels across European countries in pre-industrial times. He emphasizes the role of education opportunity cost for the vast majority of the population rather than for the powerful landed nobility acting as a constraint for the supply of schooling. Cinnirella and Hornung (2016) find that emancipated peasants had more incentives and resources to invest in education after serfdom abolition in nineteenth-century Prussia. Thus, the negative relationship between land inequality and literacy was mitigated in Prussian counties with a larger number of land and labor redemption cases, indicating the achievement of complete independence from the feudal landlord. Similarly, Beltrán Tapia and Martínez-Galarraga (2018), under the assumption that girls were not fully engaged in agricultural tasks, indicate that while farm laborers were negatively associated with male and female enrollment rates, they had a negative effect only on the supply of male school teachers. In the same direction, they also argue that land access inequality affects boys' schooling enrollment more than the provision of male teachers. Lastly, the results seem unaltered when the analysis is restricted to a rural sample, where landowners' power would plausibly be more extensive. In another work, relying on a newly assembled dataset on individuals living in Madrid in 1880 and 1905, the authors find that despite the supply of mass schooling improved access

<sup>5</sup> Andersson and Berger (2019) document that educational expenditure was *higher* where the distribution of political power was more skewed in nineteenth-century Sweden.

to education for children with disadvantaged backgrounds, the public effort was insufficient to offset the constraints these families were to face (Beltrán Tapia and de Miguel Salanova 2019).

I contribute to this literature by providing new insights on the role of land inequality on human capital accumulation in post-unification Italy, a country characterized by severe inequality levels and a North-South gap. Before describing data and the empirical strategy to test the proposed hypotheses, I briefly outline a historical background of the Italian economy in the liberal age, focusing on its educational and agricultural structure.

### 3 Historical context

#### 3.1 Italy's education system, 1859–1821

During the unification process, the Casati law (1859), firstly operating only in the Kingdom of Sardinia, was extended to the annexed regions that later became part of the Kingdom. The subsequent education system consisted of first-grade primary schools free of charge, delivered by municipalities according to their fiscal capacity. The first two years of primary school were mandatory for boys and girls, but households were not to pay sanctions in case of no compliance with the rule. Municipalities with more than 4000 inhabitants were responsible for setting up second-grade schools, characterized by two additional years of primary schooling. Each municipal council was then authorized to build schoolhouses, hire teachers, pay their salaries, and enforce attendance.

In 1877, the Coppino reform amended the electoral system established by the Casati Law, whose main contribution consisted in raising the compulsory years of schooling from two to three. Also, it introduced the possibility for municipal councils to receive subsidies for building schoolhouses and furnishing didactic material and established sanctions for households not complying with the rule of law. Generally, it enhanced the power of local councils to enforce compulsory attendance, especially in rural and less developed areas of the country. The Nasi and Orlando laws followed this reform, approved respectively in 1903 and 1904, which further raised the age of mandatory education and increased teachers' salaries, blocked since 1886 (see Cives (1990)). Finally, in 1906, the Special Law for the South of Italy was enacted, including reforms aiming to bridge the socioeconomic gap in schooling between northern and southern regions. However, although these reforms improved the education system, none decisively boosted human capital.

A further crucial step was taken with the introduction of the *Daneo-Credaro* Law in 1911, following the Corradini Inquiry published one year before. The reform significantly changed the education system in Italy, centralizing the funding of schooling to a large extent. The main characteristic consisted of creating the provincial school boards, the so-called *Consiglio Scolastico Provinciale* (CSP), which worked as a link between municipalities and the central government. On the one hand, they limited the power of municipal councils, whereas, on the other hand, they allowed a certain degree of administrative decentralization in the management of public

resources. While municipalities could apply for free loans to fund the construction of schoolhouses, the state was responsible for paying teachers' salaries. Moreover, municipalities included in the CSP system were obliged to transfer an amount of money equal to the previous years' budget to the Treasury. The state would, in turn, re-transfer these funds to the CSPs, which were charged with the duty of financing public education (see Cappelli (2015) for further details).

### 3.2 Agrarian structure of the Italian economy

Italy, soon after unification, was a predominantly rural economy. In 1911, according to the Population Census, the share of agricultural employment over the total labor force was about 60%, and it remained relatively unchanged until the Second World War (see Felice (2018)). Since the first years after unification, land inequality has always been considered one of the main issues at the core of political projects aimed at improving the conditions of the southern economy. The debate around *latifundia* involved prominent scholars such as Gramsci, Nitti, Sereni, and Croce, at least as late as the implementation of the land reform after the Second World War.

First of all, land inequality is a measure of wealth dispersion per se, whose observation contributes to shedding light on the level of concentration of resources among the population. Second, land constitutes the main productive input in a prevalently agrarian economy. Thereby, it may create oligopsonistic local labor markets if highly concentrated in the hands of few landowners. As a result, many rural laborers lose bargaining power and are left with the only alternative to migrate, looking for better conditions outside agriculture. Naturally, large landowners' interests will consist of keeping wages low, trying to impede the migration process, and maintaining a large mass labor force at disposal.

Traditional historical literature considered *latifundia* negatively, attributing them as one of the profound causes of the backwardness of the South. Accordingly, large landowners were deemed absentee and not rational economic agents, more concerned with rent extraction than undertaking productive investments on their land (Sereni 1971).

A more recent branch of literature revised the traditionalist view. Specifically, the hypothesis of southern large landowners' irrational behavior is called into question (Petrusewicz 1989; Placanica 1990). Instead, their decisions are attributed to the existence of specific non-removable constraints they were facing, different from the ones landowners were to address in the Po valley, such as drought in the summer, lack of rain, and the presence of malaria (Galassi and Cohen 1994; Bevilacqua 2005; Lupo 1990).

Notwithstanding an entrenched controversy and long-term debate on the role of land inequality in the North-South divide in Italy, there have been no attempts to integrate the discussion with quantitative evidence for the post-unification period. I try to fill this void to assess its correlation with human capital accumulation. Then, I focus on one specific component of the development process, leaving aside other possible channels through which land inequality affects long-run economic growth, such as efficiency in agriculture (Deininger and Squire 1998), excessive taxation (Persson and Tabellini 1994), trade policy that protects rent extraction (Adamopoulos 2008), extractive institutions (Acemoglu et al. 2002), and market power (Martinelli 2014).



The following section describes the main variables employed in the empirical analysis.

## 4 Data

I rely on a newly assembled dataset that includes land inequality and human capital information in post-unification Italy. I digitized data from various sources and referred to six years (1871, 1881, 1891, 1901, 1911, 1921) for the 69 provinces and to only three points in time (1881, 1901, and 1921) for 205 districts to test whether the concentration of landownership is associated with the expansion of education.<sup>6</sup> Data coming from Population Censuses contain several available information, including literacy rates, all agrarian contracts employed in agriculture, their percentage out of the total labor force, surface extent, child labor, and population. I integrated this information with further data coming from alternative sources. Specifically, data on municipal expenditures, literacy rates for the population aged between 15 and 19, the number of children and teachers, enrollment rates in primary school, and the number of schools divided by gender come from Bozzano et al. (2019) and Cappelli and Quiroga Valle (2020) and are kindly provided by the authors.

Since Population Censuses only provide information on contractual agreements employed in agriculture, without including information on the size of landholdings and their distribution among individuals, I cannot build up a proper measure of land inequality. Vollrath (2013) uses a Gini index as the first measure of land inequality, despite several issues to be addressed. It is computed across all operationalizing farms divided by bin size and adjusted to account for males aged 21 or more. Similarly, Ramcharan (2010) studies the effect of land inequality on redistributive policy in the USA, employing the Gini coefficient at the county level for 1890–1930. Another accurate measure of land inequality alternative to the Gini index would have been the size of landholdings (Ramcharan 2010; Cinnirella and Hornung 2016; Vollrath 2013). However, it has been used to approximate the extent of serfdom or the tax base in rural areas. Therefore, I follow Beltrán Tapia and Martínez-Galarraga (2018) and Oto-Peralías and Romero-Ávila (2016), which measure land inequality, respectively, with the percentage of landless peasants over the total agricultural active workforce and the fraction of farm laborers over the total agricultural population.<sup>7</sup> In this context, the adequate measure of the variable of interest is the fraction

<sup>6</sup> The Population Census was not conducted for the year 1891 because of financial issues. Therefore, I have estimated data for this year through linear interpolation.

<sup>7</sup> Advocating the use of such a measure, Oto-Peralías and Romero-Ávila (2016) affirm that: “We consider this as a clear-cut proxy for historical structural inequality, which is referred to as a type of inequality that is historical in the sense that has exhibited high persistence over centuries, and structural in the sense that it is a class-based inequality that measures the relative size of the landless workers class relative to landowners and tenants. In a primary-sector economy where land is a major factor of production, if landownership is highly concentrated, broad segments of the population have to work for landlords, earning low wages and living in miserable conditions. That was indeed the situation for a broad mass of the population in large estate regions.”



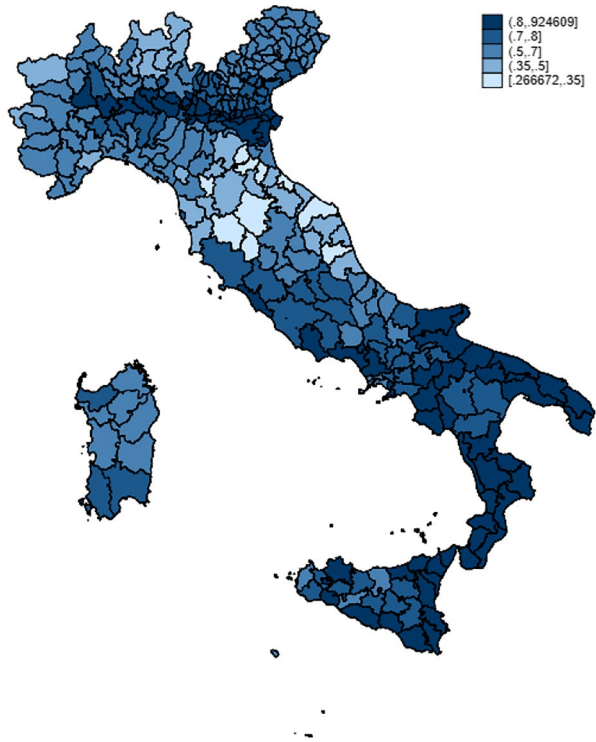
of farm laborers divided by the total labor force engaged in agriculture. In particular, the numerator includes the number of daily laborers, the so-called *braccianti*, and wage workers, generally hired on an annual basis. The two agrarian figures constitute the mass of landless peasants, counterposed to large farm owners. Furthermore, the Censuses explicitly state that if a person had declared to exercise simultaneously or, alternately, two or more professions, only the profession declared as the primary one had been taken into account. Besides, the standard computation applies throughout all the Italian territory homogeneously. Hence, double-counting issues and heterogeneous misrepresentation of the occupational structure in circumscribed areas are simultaneously avoided. Besides, the share of farm workers is more appropriate than the Gini index to measure large landowners' de facto political power in the first phase of the industrialization process. Indeed, I am primarily interested in the class conflict *between* landowners and landless peasants rather than among landowners. Thus, farm workers' share is better suited to capture the extensive margin of political power distribution rather than the intensive distribution of landownership. I expect the fraction of farm workers to be negatively associated with literacy rates.

In Fig. 1, I report the share of farm laborers in 1881 (one of the years included in my sample). Although Fig. 1 reflects what I expected to find about land inequality in Italy, it might be possible that those registered as landless indeed owned some small plots of land, yet not sufficient to ensure they reached the subsistence level. Nevertheless, this type of rural worker was more prevalent where *braccianti* were less widespread, balancing the disparities in land inequality among regions. Moreover, farm workers had to rely on wage work and child labor to meet their needs, further incentivizing them not to invest in education because of its opportunity cost. Generally, incentives to acquire education are associated with occupations attaching economic value to literacy skills, most widespread in urban areas. Indeed, towns and cities offered more accessible admittance to schools and provided inhabitants with more opportunities to bump into market exchanges and frequent encounters with the law and the authorities. Therefore, it was indispensable to read and write bills of exchange and other trade documents in a similar context. In contrast to urban areas, the countryside was mainly characterized by a less dynamic environment. Thus, it is more likely that only local notables had access to education and the opportunity to acquire literacy and numeracy skills.

Figure 2 plots literacy rates for Italian districts for the same reference year in the liberal age, 1881. They were much higher in the northwestern part of the country, coinciding with the so-called industrial triangle around Turin, Genova, and Milan, Italy's three most industrialized cities. Although some income disparities already existed between the North-West and the rest of the country at the unification and in the first decades after 1861, industrialization was only taking its first steps, and we do not observe noticeable disparities yet.<sup>8</sup> Differences in the industrialization index would have enlarged from that period, bringing about an even larger gap in income,

<sup>8</sup> Federico et al. (2019) reveal enlargement in the Northwest-South gap during the liberal age, with human capital as its primary determinant.

**Fig. 1** Farm workers (share over total labor force in agriculture) in 1881. Author's elaborations on data coming from 1881 Population Census



particularly accentuated in the interwar period.<sup>9</sup> Remarkably, the North of Italy was not urban, except in Turin, Milan, and Genova, but was mainly constituted by sparsely settled small towns, with a high number of small municipalities. The area was marked by the settlement pattern of the *cascina irrigua*, the typical Po Valley agrarian regime, which favored the settlement of peasants close to their land. The highest urbanization rates were registered in some southern areas, mostly coinciding with the so-called *agro-towns*, composed of a huge mass of agrarian laborers settled on top of hills and far from cultivated lands, often infested by malaria (Malanima 2005).

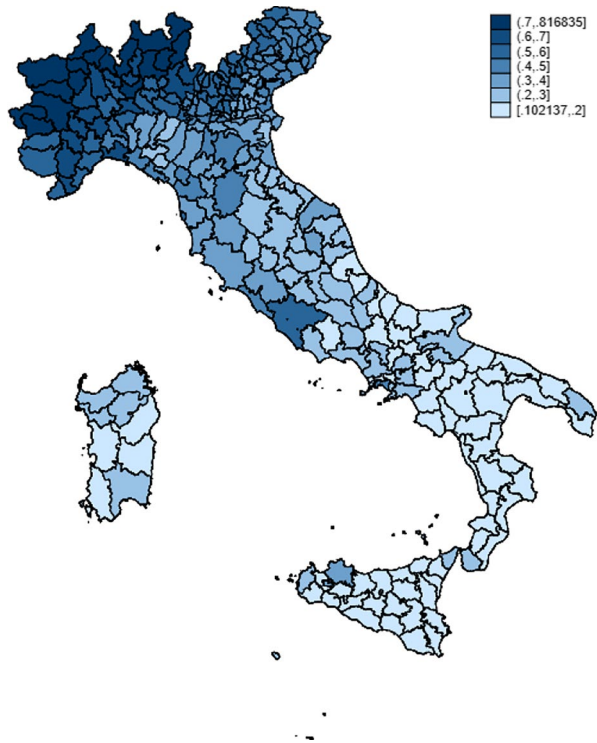
Figure 3 displays the variation of literacy rates over time for the four main Italian macroregions. Whereas the graph shows a clear North-South gap, it seems that literacy rates grew steadily at the same pace throughout the period under analysis. Neither divergence nor convergence seems to occur (at best modest convergence rates) until the implementation of the *Daneo-Credaro* law. From that moment on, the growth rates' speed seemed enhanced for all Italian regions, yet slightly more in southern regions.<sup>10</sup>

I test the hypothesis that the degree of landownership concentration is negatively associated with the level of human capital. A preliminary overview confirms the

<sup>9</sup> It is still debated whether industrialization created incentives to invest in human capital or it was the other way around, i.e., higher education levels paved the way for the industrial revolution in the circumscribed area of northwestern Italy. For further details, see A'Hearn (1998) and Felice (2018).

<sup>10</sup> In the Appendix, Fig. 9, I report the temporal trend for the share of farm workers and literacy divided by gender.

**Fig. 2** Literacy rate (1881).  
Author's elaborations on 1881  
Population Census

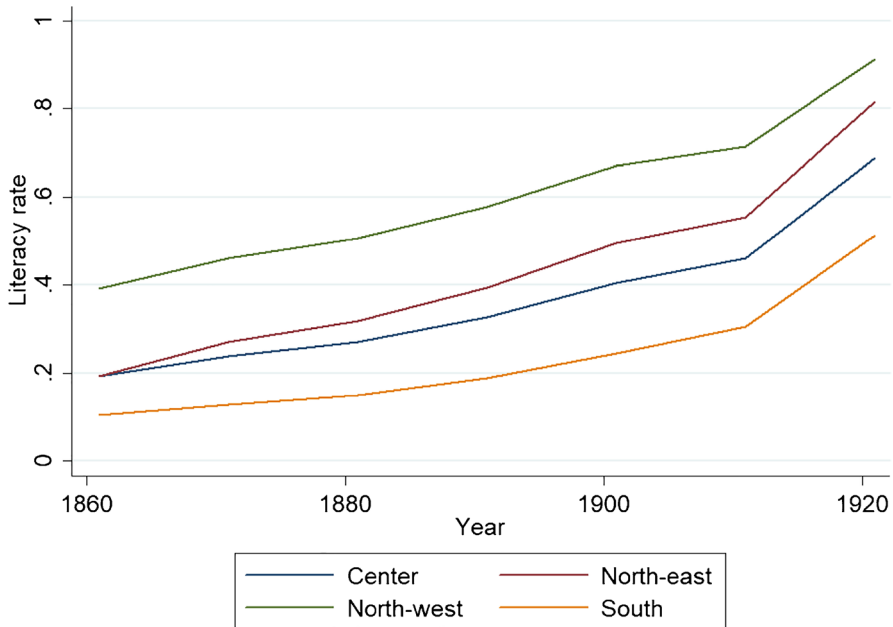


conjecture, but the results may be driven by some omitted or unobserved factors related to both variables. Hence, in the next section, I attempt to clarify similar concerns by including additional controls and implementing an identification strategy based on an instrumental variable approach. Furthermore, I exploit the panel structure of data by employing a stepwise diff-in-diffs strategy, accounting for the differential effect of structural reforms on literacy between areas with low and high initial land inequality. Among them, the centralization of public investment decisions in mass schooling with the introduction of the *Daneo-Credaro* law in 1911 could have played a decisive role to prompt education, acting as a “substitute for prerequisites” (Allen 2011; Gerschenkron 1962; Cappelli and Vasta 2020).

Descriptive statistics referred to variables employed in the empirical analysis are reported in Tables 14 and 15 in the Appendix, both for districts and provinces and each year. For a thorough definition of the variables and their source, see Table 16 in the Appendix.

## 5 Empirical strategy

This section evaluates the contemporaneous and lagged relationship between landownership concentration and literacy rates. First, I present OLS results from separate cross-sections at different points in time, both for districts and provinces in



**Fig. 3** Literacy rate by macroregion (1861–1921)

post-unification Italy. Then, I introduce an instrumental variable approach based on the presence of malaria, which aims to address potential endogeneity. Next, I conduct a robustness check analysis by including further controls and validating the instrument with a reduced form regression and several falsification tests. Lastly, I explore the temporal dimension of the dataset presenting panel estimates, which allow accounting for time-invariant unobserved heterogeneity.

## 5.1 OLS estimates

I estimate a standard OLS model where literacy rates are a function of the share of farm laborers in each district or province  $i$ :

$$Lit_{it} = \alpha_1 + \beta_1 LandIneq_{it} + \gamma_1 X' + \epsilon_{it}; \quad (1)$$

$\forall t \in 1871, 1881, 1891, 1901, 1911, 1921$ , where  $Lit$  is the literacy rate,  $LandIneq$  is the share of farm workers, and  $\beta_1$  the coefficient of interest. One may contend that literacy of the overall population is determined much earlier than farm laborers are measured. Hence,  $LandIneq$  is measured both at time  $t$  and  $t - 1$ , whenever possible.  $\epsilon$  is the standard error term (henceforth, SE), clustered at the district- and provincial levels. For robustness, I also compute Conley SE (Conley 1999) to account for the possibility that errors are cross-correlated among neighboring districts and provinces, and wild cluster bootstrap SE, following

Cameron et al. (2008).<sup>11</sup>  $X$  is a vector of socio-economic and geographical covariates, which permits controlling for other factors related to education and land inequality that may drive the results. These variables, indeed, capture other aspects of demand and supply of schooling potentially associated with landownership concentration. I include population (in logs), urbanization rate, sharecroppers out of total agricultural workforce, and the fraction of workers engaged in agriculture. In the absence of disaggregated data for GDP and GDP per capita for the period under analysis, the inclusion of these variables permits to account for opportunities outside agriculture that create an incentive to invest in human capital. Indeed, urban areas are typically more prosperous, as more productive agriculture can sustain other activities where literacy skills are required. Following the same argument, I include the share of the agricultural labor force to distinguish between rural and urban geographic areas. Then, I expect it to be negatively associated with literacy rates. The proportion of sharecroppers allows controlling for areas where it is likely to register high land inequality levels not accounted for, given how the variable is constructed.<sup>12</sup> Furthermore, I control for a set of geographic factors: a dummy that takes on a value equal to 1 whether the district or the province is landlocked and 0 otherwise; the latitude and four dummies for the Italian macroregions. The first variable allows capturing the openness of the local environment to trade and cultural flows, commonly due to better access to the sea. In contrast, the latitude controls for a diverse climate and temperature throughout the peninsula, potentially driving different agrarian regimes. Finally, four departmental dummies allow capturing most of the unobserved heterogeneity at the macroregional level.

OLS results are presented in Tables 1 and 2, respectively, at the district- and province-level. The coefficient of interest is negative and statistically significant for all years under analysis, both for the district and the province estimates.<sup>13</sup> Furthermore, concerning estimates based on a more detailed territorial level (district-level), the

<sup>11</sup> I calculate Conley SE assuming a cutoff distance of 200km, using the “acreg” command for Stata implemented by Colella et al. (2019) and imposing a distance linear decay between observations within the cutoff in the correlation structure. In unreported regressions, I also assume a cutoff distance of 100km for the district-level dataset, and the results remain robust. I report the more conservative specification with 200km in the district-level dataset because of consistency with provincial estimates. The choice of the cutoff is based on the simple consideration that the literature reports a threshold of 300km for regional estimates. Then, choosing a cutoff of 200km is consistent with more granular observation units such as provinces and districts. Besides, I report  $p$ -values associated with the wild cluster bootstrap SEs, clustering at the regional level. To compute them, I use the “boottest” package uploaded by Roodman et al. (2019). Instead, the stars relate to SE clustered at the region-level before imputing the bootstrap correction method.

<sup>12</sup> I find a negative correlation between the share of farm workers and sharecroppers for each year, meaning that they represent almost mutually exclusive contractual agreements. At the same time, Martinelli (2014) highlights the presence of a high level of landownership concentration even in the case of the adoption of sharecropping agreements. To avoid any doubt, I opt for including sharecroppers among baseline controls, given its ambiguous relationship with the variable of interest.

<sup>13</sup> The low level of significance for 1871 may be due to measurement error, given a different construction of the variable of interest. Indeed, the classification of tenancy contracts in the 1871 Population Census is based on alternative categories, and they are collected differently from other Censuses. Thus, they may overlap with each other, producing different results.

**Table 1** Literacy and farm workers, OLS, district level

	1881	1901	1901 lagged effect	1921	1921 lagged effect
	(1)	(2)	(3)	(4)	(5)
Farm workers	-0.286*** (0.055)	-0.201*** (0.038)	-0.340*** (0.052)	-0.134*** (0.033)	-0.154*** (0.032)
Agrarian labor force	-0.277*** (0.053)	-0.482*** (0.048)	-0.298*** (0.056)	-0.306*** (0.046)	-0.360*** (0.050)
Population (log)	0.027*** (0.010)	0.016** (0.008)	0.023** (0.010)	0.009 (0.008)	0.007 (0.007)
Urbanization rate	0.022 (0.020)	-0.005 (0.018)	0.042* (0.023)	-0.014 (0.021)	-0.001 (0.021)
Sharecroppers	-0.311*** (0.050)	-0.239*** (0.031)	-0.392*** (0.047)	-0.144*** (0.041)	-0.165*** (0.031)
Landlocked	0.020** (0.010)	0.022** (0.009)	0.020** (0.010)	0.008 (0.009)	0.006 (0.009)
Latitude	0.017*** (0.003)	0.021*** (0.003)	0.020*** (0.003)	0.020*** (0.004)	0.021*** (0.004)
Conley S.E. 200km	[0.080]***	[0.033]***	[0.078]***	[0.020]***	[0.039]***
<i>p-value farm workers</i>					
Wild bootstrap	{0.025}***	{0.014}***	{0.004}***	{0.001}***	{0.006}***
<i>N</i>	205	205	205	205	205

All regressions include macroregional dummies. Heteroskedastic-robust standard errors are reported in parentheses for separate cross-sections. *P*-values of wild cluster bootstrap SE of farm workers clustered at the regional level are reported in curly brackets (stars relate to SE clustered at the regional level); Conley SE (200 km) in square brackets. Columns 3 and 5 report estimates with lagged explanatory variables.

\*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

coefficient magnitude appears to decrease over time, meaning that ongoing changes associated with landownership concentration had a role in reducing its impact on education levels. Therefore, even though OLS results are not conclusive, they suggest that the relationship between land inequality and literacy rates was likely to be more accentuated in the first stages of development. The share of the agrarian labor force, distinguishing between rural vis-à-vis urban areas, shows a negative and statistically significant coefficient, as expected. Even the share of sharecroppers over the total labor force in the primary sector is negatively associated with literacy, in line with the fact that only northwestern regions had higher levels of education. Latitude is positively associated with literacy, while other controls seem less relevant, except the dummy for being landlocked, positively associated with the provincial estimates and the district-level outcomes in 1881 and 1901. Moreover, in columns 3 and 5 of Table 1 and columns 3, 5, 7, 9, and 11 of Table 2, I report estimates with lagged levels of farm workers ( $LandIneq_{it-1}$ ), presenting an even greater magnitude of the parameter of interest.<sup>14</sup> Finally, to account for spatially correlated standard

<sup>14</sup> I also employed lagged levels of the covariates.

errors, I report wild-bootstrap  $p$ -values (1,000 repetitions) for SE of farm workers clustered at the regional level and Conley SE adjusted considering a radius of 200km at the bottom of the table. Both alternative ways of computing SE produce similar results about the significant impact of farm workers on literacy. Overall, while OLS estimates highlight a robust association between land inequality and human capital in post-unification Italy, they still present several sources of endogeneity to be addressed, and the estimated coefficients cannot be interpreted as causal relations.

## 5.2 IV estimates: the role of malaria

While empirical evidence of OLS regressions appears robust, the findings do not reproduce a causal effect as unobserved omitted factors correlated with land inequality and literacy might bias the results. On the one hand, more educated peasants may be more able to rent plots of land they can redeem and obtain better tenant conditions. This could be particularly true in a fixed-rent regime, ultimately resulting in a more equally distributed land ownership. Conversely, more educated peasants may decide to sell their lands to seize better rents in occupations outside agriculture, with a higher payoff for high-skilled workers, potentially leading to increased land inequality. While in this latter case, I eventually find a downwardly biased coefficient, in the first, the estimates may become upwardly biased toward zero. Thus, I address endogeneity by adopting an instrumental variable approach. Specifically, I exploit a new measure related to malaria pervasiveness in the liberal age as a source of exogenous variation in land inequality. Before moving to the empirical analysis, I briefly describe the peculiar relationship between malaria and land inequality and their role in Italian historiography.

### 5.2.1 Malaria and *latifundia* in the Italian countryside

Malaria is the basis of all social life. It determines relations of production and the distribution of wealth. Malaria lies at the root of the most important demographic and economic facts. The distribution of property, the prevailing crop systems, and patterns of settlement are under the influence of this one powerful cause.

Francesco Saverio Nitti

Italy registered a long record of malaria exposure throughout the centuries. Archaeological evidence dates back to the eighth century BC the presence of the disease on the Italian territory, attesting to abandoned settlements beside the southern Italian and Sicilian coasts attributable to malaria epidemics (Corbellini and Merzagora 1998, pp. 22–23). First, the ancient Romans met one of their greatest enemies in malaria, as the disease rendered the Pontine swamplands around Rome completely uninhabitable (Nájera 2001). Then, around 400–100 BC, it spread along Tuscan coasts, either coming from the South of Italy or directly from North Africa (Salares et al. 2004). During the last decades of the nineteenth century and the first half of the twentieth century, the disease was still considered one of the main political and socio-economic issues because of its association with rural poverty and



**Table 2** Literacy and farm workers, OLS, province level

	1871	1881	1881 lag	1891	1891 lag	1901	1901 lag	1911	1911 lag	1921	1921 lag
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Farm workers	-0.153* (0.077)	-0.324*** (0.102)	-0.174** (0.073)	-0.331*** (0.075)	-0.383*** (0.093)	-0.261*** (0.050)	-0.380*** (0.069)	-0.247*** (0.050)	-0.263*** (0.045)	-0.147*** (0.045)	-0.172*** (0.045)
Agrarian labor force	-0.278** (0.117)	-0.309*** (0.102)	-0.281** (0.111)	-0.416*** (0.076)	-0.321*** (0.095)	-0.419*** (0.069)	-0.462*** (0.071)	-0.435*** (0.071)	-0.396*** (0.067)	-0.320*** (0.075)	-0.347*** (0.078)
Population (log)	0.013 (0.018)	0.015 (0.017)	0.013 (0.017)	0.006 (0.013)	0.016 (0.016)	0.002 (0.012)	0.007 (0.012)	0.000 (0.012)	0.004 (0.010)	0.005 (0.010)	0.006 (0.010)
Urbanization rate	0.038 (0.029)	0.024 (0.036)	0.038 (0.029)	0.021 (0.028)	0.032 (0.036)	0.040 (0.028)	0.031 (0.029)	0.031 (0.034)	0.060** (0.029)	-0.004 (0.043)	0.002 (0.040)
Sharecroppers	-0.167*** (0.058)	-0.340*** (0.079)	-0.204*** (0.055)	-0.321*** (0.050)	-0.411*** (0.073)	-0.275*** (0.036)	-0.379*** (0.049)	-0.334*** (0.044)	-0.278*** (0.034)	-0.195*** (0.046)	-0.253*** (0.043)
Landlocked	0.028 (0.017)	0.034** (0.015)	0.028* (0.017)	0.037** (0.015)	0.036** (0.015)	0.040*** (0.014)	0.040*** (0.014)	0.044*** (0.013)	0.043*** (0.013)	0.020 (0.016)	0.024 (0.015)
Latitude	0.017*** (0.006)	0.014*** (0.005)	0.018*** (0.006)	0.014*** (0.005)	0.015*** (0.005)	0.016*** (0.005)	0.015*** (0.005)	0.016*** (0.005)	0.014*** (0.005)	0.020*** (0.007)	0.021*** (0.006)
Conley S.E. 200km	[0.069]**	[0.090]***	[0.064]***	[0.056]***	[0.081]***	[0.035]***	[0.053]***	[0.035]***	[0.029]***	[0.040]***	[0.038]***
<i>p-value farm workers</i>											
Wild bootstrap	{0.520}	{0.059}**	{0.300}	{0.000}***	{0.011}***	{0.001}***	{0.001}***	{0.001}***	{0.001}***	{0.017}***	{0.003}***
<i>N</i>	69	69	69	69	69	69	69	69	69	69	69

All regressions include macroregional dummies. Heteroskedastic-robust standard errors are reported in parentheses for separate cross-sections. *P*-values of wild cluster bootstrap SE of farm workers clustered at the regional level are reported in curly brackets (stars relate to SE clustered at the regional level); Conley SE (200 km) in square brackets. Columns 3, 5, 7, 9, and 11 report estimates with lagged explanatory variables. \*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

underpopulation (Amorosa Jr et al. 2005). Malarial zones in Sardinia, in the Pontine marshes (*Agro Romano*), and along the coasts of *Mezzogiorno* were all characterized by low agricultural productivity and population density (Celli and Fraentzel 1933; Missiroli 1949).

While the interrelationship between high land inequality and malaria has a long track record, environmental factors contributed to entrenching high inequalities once institutional factors became less relevant after the subversion of feudalism. Feudal residues and fiscal privileges were removed between the Napoleonic wars and the unification. Inheritance laws such as *maggiorasco* and *fideicommeso*, which had maintained high levels of wealth inequality, were eventually abolished.<sup>15</sup> Nevertheless, the lands of the fiefs, which remained deprived of the feudal lords' possession, returned to the availability of pre-unification governments, which reconfigured their ownership through assignments, taking the "useful possession" and, in its absence, direct assignment as the main guideline criteria. Therefore, it was easy for local notables to exhibit all the necessary titles regarding their "useful possession" of lands<sup>16</sup>. Besides, many workers who received a share of land could not pay land taxes for the scarcity of resources and means at their disposal. They were compelled to sell it to the wealthiest owners, too large were the rights still recognized for the barons that the individuals could redeem only by resorting to the courts, exposing themselves to intolerable economic costs. Consequently, after the subversive laws of feudalism, local notables maintained property on most of the feudal lands de facto (Villani 1965; Anatra 1977).

Thus, although this process of institutional change did not profoundly modify the agrarian structure of the economy, geographical non-removable constraints played a major role in exacerbating high wealth and land inequality well after the unification. Malaria-infested regions, impeding rural households from living close to their lands, contributed to further concentrating land ownership and maintaining *latifundism*. A similar agrarian structure of society was perpetrated as late as the interwar period. Arrigo Serpieri, former Minister of Agriculture during fascism, attested that a different geographical and social environment was necessary to switch from an extensive to an intensive agrarian regime. He also added that "(...)intensified cultivation systems are not possible where good and tolerable hygienic conditions are not determined before or in parallel with their introduction. Many large estates are now infested with malaria, in front of which the current system has the great merit of not requiring the permanence in the countryside of several people, especially during the period of the most serious infection" (my translation from Boccini and Piccialuti (2003)).

The hypothesis that malaria was a driver of *latifundia* creation was originally suggested by Celli and Fraentzel (1933). More recently, it was tested by Chaves (2013), who uses Markov chain models to study the dynamics of *latifundia* creation in 1930

<sup>15</sup> By *maggiorasco* is meant elder brother's inheritance. In contrast, goods, usually real estate, that were the subject of a *fideicommissum* could not be sold, donated, or diminished in any case, except in very exceptional circumstances, and consequently were transmitted unaltered from generation to generation.

<sup>16</sup> Such a mechanism of the reestablishment of lords' land ownership after the subversion of feudalism is documented in detail by Villani (1965), above all for the Bourbon royal power.

Spain. Celli observes a high correlation between malaria endemicity and the creation of large estates in the Pontine marshes, a plain in the surroundings of Rome.<sup>17</sup> More specifically, the historical archives of the Roman Pontine marshes document the debilitating effects of malaria on farmers, which do not allow them to harvest crops and properly cultivate lands. Thus, malaria, forcing rural households to abandon their lands, brought about that only fewer wealthy landowners could afford them. Then, large landowners would exploit the land as *latifundia*, based on extensive and land-intensive agriculture, requiring a landless peasants' workforce. During harvest time, farm workers were forced to settle in towns located on hills and far from infested areas. Idle lands reinforced the presence of *latifundia*, ultimately resulting in a vicious circle that maintained such a feudal agrarian regime as late as the Second World War.

Moreover, extensive literature documents the greater capacity of large farmers to cope with climatic and environmental risks (Binswanger and Rosenzweig 1993; Eastwood et al. 2010 and Ramcharan (2010)). Among these factors, malaria played a major role, posing the infested areas with a constant uncertain and risky condition. It forced large landowners to manage adverse shocks and adopt appropriate incentive-compatible strategies. Snowden (2008) documents the greater capacity of larger farmers to promptly replace sick workers during the harvest season, further complicated by their location far away from cultivated lands. Extensive and labor-saving agriculture, farming practices based on specific cultivars, and measures to cut transaction costs in hiring external laborers prevailed in malaria-infested regions. At the local level, land-use regimes and tenancy systems based on short-term leases involving an annual circulation of tenants dominated. Extensive mono-cultivar cereal farming demanded work only at certain times of the year. Hence, residence on the land was discouraged, and peasants preferred living in the "agro-towns," located on top of hills (King and Strachan 1978).<sup>18</sup> The settlement pattern of the population was therefore highly influenced by the presence of malaria.

To exploit information on malaria at a very granular territorial level, I digitized the map "*Carta Della Malaria dell'Italia*," illustrated by Luigi Torelli in 1882 (see Fig. 4), obtaining a measure of its diffusion at provincial and district levels. The map was obtained as a result of the governmental investigation entitled "Reclamation of malaria regions along the Italian railways,"<sup>19</sup> led by Senator Luigi Torelli in 1882, and contained within the documentation of bill 17 of the 15th legislature. Malaria

<sup>17</sup> The same pattern was observed in Spain in the 1930s by Beauchamp (1988).

<sup>18</sup> Snowden (2008) wrote: "At harvest and threshing time, latifundia set in motion extensive migrations of nomadic workers. During the winter off-season, these peasants were recruited in distant hill towns to join workgangs that journeyed to the region's most fertile and lethal fields. Working based on oral agreements and sometimes for meals alone, the migrants worked from sunrise to sunset, and slept on the estates, where they sheltered in stables or straw huts or bivouacked in the open fields."

<sup>19</sup> My translation of "*Bonificazione delle regioni di malaria lungo le ferrovie d'Italia*."

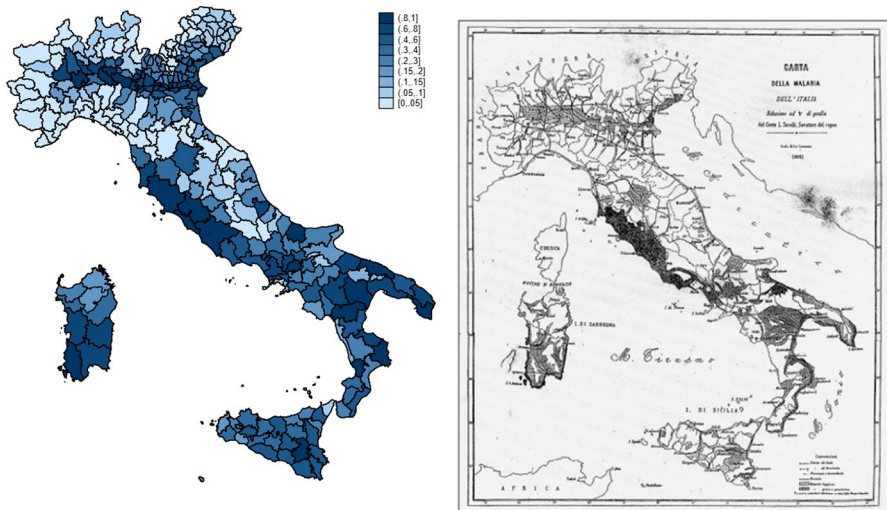


Fig. 4 “Carta della malaria dell’Italia,” Torelli map (1880)

prevalence is classified into three degrees of intensity: weak, moderate, and severe.<sup>20</sup> To my knowledge, there exists no other similar historical data on malaria virulence at such a high spatial resolution for pre-industrial Italy.

Figure 4 shows more than one similarity with the pattern observed for farm workers in the same year. As well as in the South and the two islands, I find malaria widespread in the areas of Maremma and Pontine marshes, Tuscany and Latium, and some provinces of the Po Valley. I claim that this measure is representative of the diffusion of malaria throughout Italy in all the liberal age, even though it is observed at only one point in time, and I am not able to exploit its panel dimension. Indeed, according to Snowden (2008), no remarkable changes occurred in the South until the Second World War, despite the introduction of quinine at the turn of the twentieth century, which somewhat reduced mortality due to malaria, especially in the North.<sup>21</sup> Nevertheless, mortality rates from malaria represent a less proper index for the extension of the phenomenon. Indeed, like other parasitic diseases adapted to their hosts, it is generally associated with a low case-mortality ratio. In 1882, the Italian government conducted a survey attesting that around 11 million people

<sup>20</sup> A similar index was proposed by Buonanno et al. (2020) on the grounds of Torelli report. Wherever a municipality contained at least a share of land with malaria, the authors considered the municipality affected by malaria in 1882 and attributed them the corresponding degree of severity. The digitalization of the map allowed obtaining an even more exact measure computed as the share of geographic territory covered by the disease. First, the territory was divided into grid cells; then, the variable was reaggregated first at municipal and then at district and provincial levels.

<sup>21</sup> Malaria was finally eradicated with the introduction of DDT by American troops after 1944. A malaria endemicity pattern similar to 1881 was registered before its application, between 1945 and 1950, despite being more moderate in severity.

were at constant risk of infection, and another 2 million contracted the disease every year (see Snowden (2008)).<sup>22</sup> Hence, while the disease can contaminate a substantial percentage of the population, the number of deaths directly attributable to malaria endemicity accounts for only a low fraction of the overall annual mortality rate.

I use the variable as an excluded instrument in all cross-sectional IV regressions at different points in time, both for districts and provinces. The following equation expresses the first-stage regression:

$$LandIneq_{it} = \alpha_2 + \beta_2 Malaria_{it} + X'_{it} \gamma_2 + u_{it} \quad (2)$$

where *LandIneq* is the proxy for landownership concentration, *Malaria* is the share of geographic territory covered by malaria, and *X* is the same vector of covariates included in the second-stage equation. The identification strategy hinges on the assumption that malaria endemicity affects literacy rates only through land inequality. Nonetheless, if malaria can influence agricultural productivity and crop mixes requiring different skill intensities, this would violate the exclusion restriction. Following the same argument, malaria could affect households' fertility decisions and variations of child labor in agriculture, keeping children out of school. This phenomenon might even be reinforced by adopting crops requiring high skills and monitoring levels, inducing households not to invest in their children's education. In light of these considerations, the estimates should be interpreted with caution, as a violation of the exclusion restriction may still occur.

### 5.2.2 Malaria and *latifundia*: an empirical analysis

First-stage results are reported in Tables 17 and 18 in the Appendix, showing the relationship between malaria and farm workers for each year at the district and provincial levels. The share of the surface covered by malaria is significantly and positively correlated with the fraction of farm workers. The first-stage F-statistic is constantly high, confirming the instrument's validity and power, significantly related to the land ownership structure. Indeed, values exceed the threshold of ten as a rule of thumb, ranging from 11.18 to 26.49 for provincial estimates and from 36.32 to 76.86 for the district dataset.<sup>23</sup> Columns gradually add controls to evaluate the robustness of the association to different model specifications. Only the inclusion of the share of sharecroppers seems to substantially reduce the coefficient size in the first-stage regressions, while estimates remain stable conditioning for the remaining variables. In particular, residual baseline covariates and time-invariant geographic controls do not alter the association between malaria and land inequality. Besides, the F-test of excluded instruments slowly decreases as long as gradually conditioning for baseline and geographic controls, yet never plummeting below ten.

To provide further evidence of malaria as a fundamental driver of landownership concentration, I gather data on the share of the population scattered in the

<sup>22</sup> For instance, in Sardinia before 1920, about 80% of a community suffered from malaria annually, whereas only at most 7% of deaths resulted from the disease (Brown 1983).

<sup>23</sup> Stock and Yoko and Sanderson-Windmeijer tests confirm that the model is not weakly identified.

**Table 3** Scattered dwellings and malaria, OLS, district level

	1881	1901	1921
	(1)	(2)	(3)
Malaria 1880	-0.092** (0.036)	-0.073** (0.033)	-0.067* (0.035)
<i>N</i>	205	205	205

Regressions include geographic controls and macroregional dummies. Heteroskedastic-robust standard errors are reported in parentheses for separate cross-sections. \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$

countryside and not located in towns (over the total population) for the district-level dataset at each point in time. Then, in another specification, I regress this measure on malaria and the set of geographic controls. Living close to cultivated lands was dangerous in mosquitos-infested regions. Thus, farmer families implemented a risk-coping strategy based on their settlement in towns located on top of hills, far away from malarious lands dispersed in the plains. The results reported in Table 3 confirm my conjecture. I find fewer people located in scattered dwellings where malaria was pervasive but more concentrated in small centers and villages. A 10 percentage points increase in the share of territory covered by malaria induces a reduction of 0.9 percentage points in the population scattered in the countryside. This indicates that a similar settlement pattern, while it seemed inevitable to offer protection from contagion, moved the agricultural workers away from the fields. The increased distance to the workplace favored the adoption of labor-saving agrarian regimes and extensive agriculture based on large estates concentrated in the hands of absentee landowners.

Second-stage results are reported in Tables 4 and 5.<sup>24</sup> The two-stage least-squares (2SLS) regressions confirm the significant negative relationship between landownership concentration and human capital at the end of the nineteenth century and Italy's first years of the twentieth century. In terms of magnitude, if I increased by one percentage point the share of farm workers for the Italian districts, I would find a literacy rate 0.41 percentage points lower in 1881, 0.24 percentage points lower in 1901, and 0.221 percentage points lower in 1921. In particular, consider the difference between the farm workers' share in the districts of Bari and Lugo in 1881. In Lugo, *LandIneq* = 0.556 (which is the 25th percentile of the distribution of *LandIneq* across districts in Italy), and in Bari, *LandIneq* = 0.8032 (which is the 75th percentile). Using the estimates in the first column of Table 4 implies that Bari's literacy rates in 1881 would have been almost 10% higher if it had a farm workers' share as small as Lugo's. The coefficient magnitude is also high for provincial regressions, ranging from 0.381 percentage points in 1881 to about 0.27 in 1921, even though inferior to the specification considering districts, plausibly due to unobserved factors not accounted for that make parameters downwardly biased.<sup>25</sup> To account for the

<sup>24</sup> See Figs. 10 and 11 in the Appendix to have a visual representation of residual variation between land inequality and literacy rates and between malaria and land inequality, conditional on control variables employed in the empirical analysis.

<sup>25</sup> The coefficient referred to 1871 is almost not statistically significant. See note 13 for a possible explanation.

**Table 4** IV estimates, district level

	1881	1901	1901 lagged effect	1921	1921 lagged effect
	(1)	(2)	(3)	(4)	(5)
Farm workers	−0.413*** (0.140)	−0.243*** (0.055)	−0.511*** (0.143)	−0.221*** (0.058)	−0.227*** (0.053)
Agrarian labor force	−0.291*** (0.052)	−0.496*** (0.047)	−0.317*** (0.056)	−0.322*** (0.046)	−0.386*** (0.051)
Population (log)	0.029*** (0.010)	0.016** (0.008)	0.026** (0.010)	0.010 (0.007)	0.007 (0.007)
Urbanization rate	0.025 (0.021)	−0.001 (0.017)	0.046** (0.022)	−0.008 (0.022)	0.005 (0.020)
Sharecroppers	−0.407*** (0.104)	−0.261*** (0.034)	−0.522*** (0.104)	−0.186*** (0.046)	−0.205*** (0.036)
Landlocked	0.021** (0.010)	0.022** (0.009)	0.022** (0.010)	0.007 (0.009)	0.006 (0.009)
Latitude	0.015*** (0.004)	0.020*** (0.003)	0.017*** (0.004)	0.019*** (0.004)	0.019*** (0.004)
Conley S.E. 200km	[0.222]*	[0.071]***	[0.212]**	[0.085]***	[0.083]***
<i>p-value farm workers</i>					
Wild bootstrap	{0.095}**	{0.013}***	{0.044}***	{0.012}***	{0.006}***
<i>N</i>	205	205	205	205	205

All regressions include macroregional dummies. Heteroskedastic-robust standard errors are reported in parentheses for separate cross-sections. *P*-values of wild cluster bootstrap SE of farm workers clustered at the regional level are reported in curly brackets (stars relate to SE clustered at the regional level); Conley SE (200 km) in square brackets. Columns 3 and 5 report estimates with lagged explanatory variables. \*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

small number of regions at which SEs are clustered, I additionally report wild bootstrap *p*-values and, alternatively, Conley SE adjusted for spatial correlation (cutoff of 200km) at the bottom of the table. The statistical inference produces similar findings under both approaches.

Overall, the IV coefficients are higher than the OLS coefficients at all points in time. The magnitude of the bias is in line with the results obtained by Cinnirella and Hornung (2016) and Ramcharan (2010). The authors find that IV results are more than three times larger than OLS estimates. The bias toward zero referred to OLS results might arise for several reasons, ranging from measurement error to the presence of unobservables only partially captured by the model and the identification strategy. However, the size of the bias varies according to the period. The  $\beta$  coefficient of the district level IV regression is almost 1.5 times higher than the corresponding OLS coefficient of 1881, while about 1.2 and 1.3 times higher in 1901 and 1921. This further suggests that the bias might be due to measurement error of farm workers in the first years, presumably because of a less refined grade of data collection in cadastres and censuses. Finally, similar to OLS results, we observe a decreasing effect over time by comparing



Table 5 IV estimates, provincial level

	1871 (1)	1881 (2)	1881 lag (3)	1891 (4)	1891 lag (5)	1901 (6)	1901 lag (7)	1911 (8)	1911 lag (9)	1921 (10)	1921 lag (11)
Farm workers	-0.271* (0.157)	-0.381** (0.173)	-0.327** (0.154)	-0.326*** (0.108)	-0.456*** (0.166)	-0.285*** (0.076)	-0.386*** (0.106)	-0.350*** (0.097)	-0.309*** (0.075)	-0.270*** (0.095)	-0.355*** (0.118)
Agrarian labor force	-0.317*** (0.112)	-0.324*** (0.105)	-0.332*** (0.105)	-0.414*** (0.078)	-0.340*** (0.097)	-0.425*** (0.068)	-0.464*** (0.071)	-0.443*** (0.070)	-0.409*** (0.067)	-0.3197*** (0.074)	-0.361*** (0.079)
Population (log)	0.013 (0.017)	0.016 (0.017)	0.014 (0.017)	0.006 (0.013)	0.018 (0.016)	0.001 (0.012)	0.007 (0.012)	0.001 (0.011)	0.003 (0.010)	0.009 (0.011)	0.008 (0.010)
Urbanization rate	0.035 (0.033)	0.023 (0.037)	0.035 (0.033)	0.021 (0.029)	0.030 (0.037)	0.042 (0.028)	0.031 (0.029)	0.038 (0.034)	0.064** (0.029)	0.006 (0.043)	0.015 (0.041)
Sharecroppers	-0.245** (0.115)	-0.380*** (0.131)	-0.306*** (0.113)	-0.318*** (0.066)	-0.463*** (0.126)	-0.286*** (0.044)	-0.383*** (0.066)	-0.397*** (0.069)	-0.300*** (0.044)	-0.245*** (0.063)	-0.365*** (0.083)
Landlocked	0.028 (0.019)	0.035** (0.016)	0.028 (0.019)	0.037** (0.015)	0.037** (0.016)	0.040*** (0.014)	0.040*** (0.015)	0.044*** (0.014)	0.043*** (0.014)	0.016 (0.017)	0.024 (0.016)
Latitude	0.015** (0.007)	0.013** (0.006)	0.015** (0.007)	0.014*** (0.005)	0.014** (0.006)	0.015*** (0.005)	0.015*** (0.005)	0.013** (0.005)	0.012** (0.006)	0.017*** (0.006)	0.015** (0.007)
Conley S.E. 200km	[0.133]**	[0.156]**	[0.136]**	[0.096]***	[0.151]***	[0.067]***	[0.093]***	[0.080]***	[0.058]***	[0.091]***	[0.126]***
<i>p-value farm workers</i>											
Wild bootstrap	{0.1932}*	{0.144}**	{0.138}**	{0.080}***	{0.096}**	{0.029}***	{0.034}***	{0.002}***	{0.004}***	{0.029}***	{0.026}***
N	69	69	69	69	69	69	69	69	69	69	69

All regressions include macroregional dummies. Heteroskedastic-robust standard errors are reported in parentheses for separate cross-sections. *P*-values of wild cluster bootstrap SE of farm workers clustered at the regional level are reported in curly brackets (stars relate to SE clustered at the regional level); Conley SE (200 km) in square brackets. Columns 3, 5, 7, 9, and 11 report estimates with lagged explanatory variables. \*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

coefficients across the three separate cross-sections for the district-level dataset. The same holds for the provincial estimates, except for an increase in size in 1911.

As to the other control variables, the share of the total agricultural labor force, capturing opportunities outside agriculture and distinguishing between rural and urban areas, shows a negative and statistically significant sign, as expected. However, excluding this variable does not affect the results. Interestingly, the negative relationship between sharecropping activities and human capital remains robust, although slightly more modest than the variable of interest, unlike in the OLS framework. The macroregional coefficients (not reported for brevity) are alternate in significance, with northern areas positively associated with literacy rates and latitude. The remaining control variables are not related to literacy. Specifically, urbanization rates are not associated with more education, likely because of high levels registered in the South due to the *agro-towns*' presence, as already mentioned in Section 4. Then, the dummy for the landlocked area is slightly positively associated with literacy rates in the first stages of the period under analysis. Finally, odd columns show results for lagged values of farm workers. Interestingly, the size of the associated parameter is more than twice the size of the contemporary effect in the district level regression in 1901, while it seems in line with the respective coeval regressions in all remaining years. Then, findings seem consistent with the conjecture that the structure of the agricultural population and the share of farm workers are determined earlier than observed literacy rates.

In Section 5.3, I also report the model specification with the inclusion of the industrialization index, mortality rates due to malaria, agricultural productivity, and the role of domestic market access. Again, changes in the results are not significant, confirming the goodness of the model specification and the robustness of the identification strategy to factors potentially invalidating the exclusion restriction hypothesis. Moreover, I go even further by undertaking three falsification tests: the first consisting of a reduced-form regression of literacy rates on the indicator for malaria; the second allowing for a deviation from the perfect exogeneity assumption; and the third consisting in making assumptions on the degree and sign of the variable of interest's endogeneity. Next, in Section 5.4, I exploit the panel dimension of the dataset.

### 5.3 Robustness checks

Malaria-endemic countries exhibit lower economic growth rates. The channels through which malaria prevents a country from undertaking a process of development are multiple, including effects on fertility, poverty, population growth, saving and investment, worker productivity, mortality rates, and medical costs (Sachs and Malaney 2002). Malaria is also seen to reduce the labor force and impede the adoption of new technologies and practices in agriculture (see Asenso-Okyere et al. (2011) for a review). All the more, the effects of malaria on school attendance and performance may directly affect human capital accumulation. Bleakley (2010) finds a positive effect of eradication campaigns in terms of income and an increase in years of schooling, whereas Cutler et al. (2010) find support for a mixed but

indicative effect only for female education, with no impact on male income and education values. Similarly, Lucas (2010) finds that malaria eradication increased years of educational attainment and literacy. In contrast, high dropout rates are generated by school absenteeism, resulting from the diffusion of the disease.<sup>26</sup>

Furthermore, a new and increasing branch of research also points toward ways in which malaria can permanently affect the cognitive performance of children (see Al Serouri et al. (2000); Holding and Snow (2001)). According to these studies, parasitemic children, for instance, score lower grades on specific tests than non-parasitemic children. Also, the long-term cognitive performance of a child may be inhibited by its previous in-utero experience during the pregnancy of the malaria-infected mother (McCormick et al. 1992).

### 5.3.1 Additional controls

In this section, I address the problem of potential time-varying confounding factors that can bias previous results. By doing so, I carry out some robustness checks of the IV baseline specification, including additional controls that can affect literacy through their interaction with landownership concentration. Results are reported in Table 6 at the provincial level.<sup>27</sup> For simplicity, I report regressions only for 1901, as all variables are at hand for this year, and comparison among different specifications is possible.

In the first column, I add mortality rates due to malaria in 1900, the first year this information is available Ministero di Agricoltura 1900. According to Malaney et al. (2004), in areas characterized by high levels of malaria endemicity, its mortality burden generally falls most heavily on children less than 5 years of age. Nonetheless, including mortality rates due to malaria does not affect the results, revealing that differences in mortality caused by malaria morbidity were not relevant in explaining literacy rates.

Column 2 includes a regional measure of labor productivity in agriculture to control for heterogeneity in agricultural productivity.<sup>28</sup> Indeed, malaria can undermine land suitability, ultimately affecting crop mixes and the adopted agrarian regimes. Nevertheless, the magnitude and significance of the coefficients are very similar to the estimates for 1901 in the IV baseline specification.<sup>29</sup>

Moreover, a shift from an agricultural to an industrial economy, creating a suitable environment for commerce and trade, might cause heterogeneity in the incentives

<sup>26</sup> A further attempt to evaluate the direct role of malaria on human capital is the work by Kere et al. (1993), who estimate the loss of educational investment expenditures as a result of lost school days in the Solomon Islands. Regarding the consequences of malaria eradication on education outcomes in Italy, Percoco (2013) provides further positive evidence at the regional level in post-war Italy.

<sup>27</sup> Unfortunately, I do not have at disposal these additional controls at the district level for each year.

<sup>28</sup> The regional measure of agricultural productivity is borrowed by Federico et al. (2007). They compute it by dividing gross saleable production over the total labor force engaged in agriculture. Unfortunately, values of agricultural productivity are not available at the provincial or district levels.

<sup>29</sup> In an alternative specification (not reported for the sake of brevity), I also gathered data for provincial per-hectare yields for wheat and corn in 1871. Including crop yields to account for differences in agricultural productivity produces similar results.

**Table 6** Robustness checks, 1901

	(1)	(2)	(3)	(4)
Farm workers	−0.297*** (0.081)	−0.284*** (0.075)	−0.280*** (0.078)	−0.286*** (0.078)
Agrarian labor force	−0.416*** (0.069)	−0.426*** (0.068)	−0.374*** (0.093)	−0.419*** (0.071)
Population (log)	0.003 (0.012)	0.002 (0.012)	−0.001 (0.012)	0.002 (0.012)
Sharecroppers	−0.286*** (0.044)	−0.284*** (0.044)	−0.289*** (0.045)	−0.286*** (0.044)
Urbanization rate	0.049 (0.030)	0.027 (0.031)	0.039 (0.027)	0.041 (0.028)
Landlocked	0.042*** (0.014)	0.038*** (0.014)	0.039*** (0.014)	0.040*** (0.015)
Latitude	0.016*** (0.005)	0.018*** (0.005)	0.013** (0.005)	0.015*** (0.005)
Mortality rate due to malaria	0.012 (0.010)			
Agricultural productivity (Federico)		0.000 (0.000)		
Industrialization index (Ciccarelli & Fenoaltea)			0.026 (0.024)	
Domestic market potential (Missiaia)				0.000 (0.000)
Conley S.E. 200km	[0.070]***	[0.063]***	[0.070]***	[0.066]***
<i>p-value farm workers</i>				
Wild bootstrap	{0.042}***	{0.032}***	{0.036}***	{0.032}***
<i>N</i>	69	69	69	69

All regressions include macroregional dummies. Heteroskedastic-robust standard errors are reported in parentheses for separate cross-sections. *P*-values of wild cluster bootstrap SE of farm workers clustered at the regional level are reported in curly brackets (stars relate to SE clustered at the regional level); Conley SE (200 km) in square brackets. \*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

to invest in education. Then, I include a provincial industrialization index in column 3, coming from the work by Ciccarelli and Fenoaltea (2010). High industrialized areas, coinciding with the provinces of the so-called industrial triangle, are plausibly negatively associated with high land inequality levels, capturing its effect on literacy rates. Nevertheless, once again, the coefficient of the variable of interest retains its significance and remains almost unaltered.

The same holds once I add a regional measure of market access, coming from Missiaia (2016), to account for heterogeneity in opportunities to trade. Indeed, market access may have favored education expansion and the acquisition of literacy skills thanks to the closeness with markets and more flourishing trade. In her work, the author argues that only domestic market potential, representing the home market, shows a traditional north-south divide. In contrast, the South does not show

remarkable disparities with the rest of the country whenever international markets are introduced. Therefore, I only use such a measure of domestic market potential, and I add it within the set of covariates. The estimates presented in column 4 with its inclusion reveal that the magnitude and significance of the coefficient of interest are virtually unchanged.

### 5.3.2 Validating the instrument: falsification tests

As is common in an IV framework, although it is easy to assess whether an instrument is relevant, less obvious is the satisfaction of the exclusion restriction hypothesis. As mentioned before, a problem with the choice of malaria as an instrument is that it may impact schooling through health channels, thus violating the exclusion restriction hypothesis. To rule out this channel, or at least to show that it is not the dominant one, I provide the reader with some insight into the instrument's validity by performing specific falsification tests.

*Reduced-form* The first test consists of running a reduced form regression of literacy on the instrument, i.e., the indicator of malaria prevalence in a territory.<sup>30</sup> Results are reported in Table 7, in the row “Reduced form (full sample).” Not surprisingly, the presence of malaria has a negative and statistically significant correlation with literacy rates in 1881. To verify that the effect is not spurious, I replicate the exercise by running two regressions dividing the sample into subsamples with low and high values of the farm workers' share (respectively, rows “Direct effect (remaining sample)” and “Direct effect (zero-first-stage group)”). Specifically, I consider the median of the variable of interest as the threshold value, being 0.6952 in 1881, 0.4196 in 1901, and 0.4155 in 1921. In the former case, the correlation between malaria and literacy rates is negative but not significantly different from zero (positive for 1901). By contrast, the regression conducted over the subsample with high values of farm workers shows a negative and statistically significant correlation between the indicator of malaria and literacy rates (except in 1921). Moreover, the magnitude of the effect almost reaches the one shown for the whole sample (i.e.,  $-0.04$  against  $-0.05$  in 1881), meaning that the significant correlation displayed in the baseline case is mainly driven by districts exhibiting a substantial landownership concentration. In other words, when the conditions for *latifundia* creation are not met, the presence of malaria is not a sufficient condition to have lower human capital levels.

*Plausibly (and beyond) exogenous test* The second and most reliable falsification test comes from the relaxation of the assumption that the instrument is strictly exogenous. As widely argued, several concerns can be raised about the instrumental variable's validity and strictly exogenous nature. Hence, I relax the hypothesis that malaria is not correlated with other unobservables influencing literacy; that is, I argue that malaria is *plausibly exogenous*. Such a definition comes from Conley

<sup>30</sup> As malaria pervasiveness is only observed for 1881, it is possible to evaluate the contemporaneous correlation between malaria and literacy rates only in one year. Cross-sectional estimates in 1901 and 1921 might be slightly biased because of measurement error due to omitted factors not controlled for that could potentially alter its intensity throughout the territory.

**Table 7** Falsification tests:  
summary of regression results

	1881 (1)	1901 (2)	1921 (3)
<i>Effect of farm workers on literacy</i>			
OLS	−0.286*** (0.055)	−0.201*** (0.038)	−0.134*** (0.033)
IV	−0.413*** (0.140)	−0.243*** (0.055)	−0.221*** (0.058)
Plausibly exogenous	−0.341 (0.409)	−0.284* (0.150)	−0.069 (0.145)
Plausibly exogenous (with uncertainty)	−0.368** (0.149)	−0.315*** (0.058)	−0.103* (0.061)
<i>Effect of malaria on literacy</i>			
Reduced form (full sample)	−0.054*** (0.019)	−0.067*** (0.016)	−0.058*** (0.014)
Direct effect (zero-first-stage-group)	−0.006 (0.049)	0.020 (0.037)	−0.031 (0.034)
Direct effect (remaining sample)	−0.044** (0.017)	−0.037* (0.019)	−0.032 (0.023)
<i>Effect of malaria on farm workers</i>			
First stage (full sample)	0.131*** (0.022)	0.278*** (0.032)	0.263*** (0.037)
First stage (zero-first-stage-group)	0.104*** (0.039)	0.187*** (0.043)	0.138*** (0.040)
First stage (remaining sample)	0.003 (0.021)	0.048 (0.034)	0.058* (0.033)

Heteroskedastic-robust standard errors are reported in parentheses. \*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$ . The row “Plausibly exogenous” assumes  $\omega_\gamma = S_0^2$  and “(with uncertainty)” uses  $\omega_\gamma = (0.125 \sqrt{S_0^2 + S_{-0}^2})$ ; “remaining sample” indicates the full sample bar the zero-first-stage group

et al. (2012), who present a tractable and straightforward test of conducting inference consistent with the hypothesis that the exclusion restriction does not hold exactly. Expressly, the authors assume that the instrument enters the second-stage regression linearly as a further control and directly affects the outcome variable. Hence, the associated coefficient, which is the parameter on which they infer, is allowed to have a non-zero value, in contradiction with the exclusion restriction hypothesis. Based on two distinct approaches, the union-of-confidence intervals (henceforth, UCI) and the local-to-zero approach (henceforth, LTZ), they make prior assumptions on the true value of such a parameter that I will call  $\gamma$ . In the first case, the support for  $\gamma$  is restricted to be an interval between two values, whereas, in the second case, further hypotheses are advanced concerning its distribution function.<sup>31</sup>

<sup>31</sup> For a broader and more detailed discussion, see Conley et al. (2012).

Therefore, the plausibly exogenous method is advantageous if we have prior information on the sign and size of the violation of the exclusion restriction. However, no guidance is provided on how to obtain a plausible and credible prior. Of course, priors over  $\gamma$  centered at 0 may be inappropriate in this case, as the related literature suggests a negative effect of malaria on education (Bleakley 2010; Lucas 2010). A recent contribution by van Kippersluis and Rietveld (2018) proposed a two-step procedure to provide a novel sensitivity analysis for IV estimation. First, this approach estimates the direct effect of the IV on the outcome in a subsample group for which the first stage (that is, the effect of the IV on the treatment variable) is zero. The intuition is that in the same subsample, the reduced form (the effect of the IV on the outcome) should also be zero if the exclusion restriction is satisfied. After, this estimate of the direct effect is used as input for the plausibly exogenous method developed by Conley et al. (2012).<sup>32</sup>

Although we do not know the subsample for which the first-stage effect is exactly zero, we might rely on the reduced form specification for low levels of the treatment variable. In such a case, we expect malaria pervasiveness to have, at the very least, a lower association with land inequality. Hence, the estimates of the reduced-form relationship between malaria and literacy where the percentage of farm workers is low provide consistent estimates of  $\gamma$ , as we expect malaria to be less likely to produce an effect on literacy.<sup>33</sup>

Table 7 summarizes regression results and reports estimated beta coefficients of the main variables for each year.<sup>34</sup> While malaria retains a statistically significant adverse effect in the zero-first-stage group (despite being smaller in magnitude), the respective direct effect is not significantly different from zero for each year.<sup>35</sup> This is reassuring, as a significant association between the instrument and the outcome in the subgroup for which we expect to have a zero-first-stage effect would have us believe in a substantial violation of the exclusion restriction. The corresponding results displayed in rows “plausibly exogenous” and “plausibly exogenous (with uncertainty)” are mixed and, to some extent, inconclusive. While the “plausibly exogenous” row shows that the effect of farm workers on literacy completely disappears when correcting for the estimated direct effect of the IV on the outcome in the reference sample,

<sup>32</sup> In terms of assumptions, one substitutes the exclusion restriction hypothesis with a weaker restriction assuming heterogeneous first-stage effects, yet homogeneous direct effects  $\gamma$  across the zero-first-stage group and the full sample. Hence, one can incorporate uncertainty around  $\gamma$  through specifying non-zero elements in the variance-covariance matrix  $\omega$  by borrowing Imbens and Rubin (2015)’s rule of thumb. That is, to fix the variance such that the normalized difference in the direct effects between the zero-first-stage group and the full sample does not exceed one-quarter in 95% of the cases. In such a case, one sets  $\omega_\gamma = (0.125 \sqrt{S_0^2 + S_{-0}^2})$ , where  $S_0$  is the standard error of  $\hat{\gamma}$  in the zero-first-stage group and  $S_{-0}$  is the standard error of  $\hat{\gamma}$  in the remaining sample (see van Kippersluis and Rietveld (2018) for further details).

<sup>33</sup> Generally, the reduced form captures both the effect of  $\beta$  and  $\gamma$ . However, if we are confident that  $\beta$  is zero, it captures only the effect of  $\gamma$ . In this context, reliable values for  $\gamma$  are provided in the reduced form with low levels of land inequality, that is, when  $\beta$  is believed to be almost zero.

<sup>34</sup> For ease of exposition, I also report OLS and IV estimates.

<sup>35</sup> While a more restricted sample taking only observations up to the 25th percentile of the endogenous variable reports a non-significant first-stage effect, the corresponding reduced-form effect is shown to be positive. Hence, I do not take this subsample into account in the analysis.



the subsequent row incorporating uncertainty around the direct effect shows that its size does not seem sufficiently large to alter the main conclusions substantively.

Figure 5 shows that while in some cases, there is still a significant effect of land-ownership concentration on literacy rates, even allowing for plausible amounts of imperfect exogeneity, in some others, more caution is required in the interpretation of results.<sup>36</sup> For instance, in the UCI approach (left-side) for 1881 regression, if I allow  $\gamma$  to be negative and ranging from  $-0.05$  to  $0$ , the corresponding confidence set for  $\beta$  is approximately  $[-0.6879; 0.2453]$ , while for 1921 is  $[-0.3346; 0.0703]$ . By contrast, in 1901, it remains negative at almost every value in the support of  $\gamma$ . As for the LTZ approach, we can straightforwardly verify that the bounds of the confidence intervals for  $\beta$  go across  $0$  as soon as I allow the direct effect to be greater than, respectively,  $-0.018$ ,  $-0.02$ , and  $-0.026$  approximately.<sup>37</sup> This means that priors consistent with beliefs above this threshold value are sufficient to lose confidence in finding a significantly negative effect of land inequality on education. By contrast, below this prior belief for  $\gamma$ , I find evidence that  $\beta$  still represents an economically relevant impact.

To conclude, applying Conley et al. (2012), we can show that the negative estimate of *gamma* reported in Table 7 for the zero-first-stage group is not above (in absolute value) the value of *gamma* necessary to lose confidence in the finding of a negative impact of land inequality on literacy in 1881 and 1901. For instance, in 1881, for the 90% confidence interval for *beta* to include zero, *gamma* must be larger than  $-0.018$  circa. Thus, it is more than three times the estimate of  $-0.006$  in Table 7. Therefore, even allowing for reasonable amounts of imperfect exogeneity, we can still confirm the presence of an adverse effect of land inequality on education. However, the opposite holds for 1921, as  $-0.031$  is larger than  $-0.026$ . This implies that we cannot reject a zero effect of land access inequality on literacy rates. The significant OLS and IV results might be the consequence of the selection of comparatively more malaria-pervasive districts into those with higher farm worker proportions.<sup>38</sup>

*Degree of endogeneity* An alternative identification strategy that does not rely on the exclusion restriction hypothesis comes from Kiviet (2020). Instead of constructing conservative confidence sets that allow for a mild violation of the exclusion restrictions, assuming a plausible range of direct effects for the instruments, the author imposes assumptions on the degree of regressor endogeneity, which is left unrestricted in an IV world. In other words, this approach identifies the confidence bounds of regression coefficients by allowing the admissible correlation of the regressors with the error term to be within plausible bounds.<sup>39</sup> The kinky least

<sup>36</sup> I followed Clarke and Matta (2018) to implement the “plausibly exogenous” test and for the setting-up of the graphing options.

<sup>37</sup> Estimates for 1881 and 1921 are in line with the estimated value of  $-0.006$  and  $-0.031$  found in the reduced form regression with low levels of land inequality, whereas the estimated value for 1901 (0.02) is positive and thus not reliable.

<sup>38</sup> Remind that malaria pervasiveness is only observed for 1881, and cross-sectional estimates in 1901 and 1921 might be potentially slightly biased.

<sup>39</sup> Akin to the plausibly exogenous test, in terms of assumptions, this test replaces one set of possibly strong though speculative assumptions (exclusion restrictions) with another set of plausibly less restrictive assumptions (degrees of endogeneity).

squares approach (henceforth, KLS) developed by Kiviet (2020) corrects the OLS estimator for all values on a grid of endogeneity correlations to provide a set of consistent coefficient estimates under the postulated endogeneity range.<sup>40</sup> Although inference can thus be conducted without instruments, the KLS approach enables testing the validity of any potential exclusion restrictions by adding them into the KLS regression and (set)-identifying their direct effect.<sup>41</sup>

Several hypotheses can be made on the farm workers' plausible degree of endogeneity. On the one hand, omitted variables proxying for economic performance and prosperity can be conducive to a negative correlation between the error term and the endogenous variable. Similarly, omitted variables capturing channels through which malaria might impact literacy, such as cognitive development of children, childhood exposure to malaria, childhood mortality due to malaria, and school absenteeism, might induce the endogenous regressor to further negatively correlate with the disturbances. On the other hand, allowing for a mild measurement error of farm workers, an imperfect measure of landownership concentration and political power, induces a positive correlation with the error term by construction.<sup>42</sup>

KLS results are presented in Fig. 6 in conjunction with 2SLS estimates.<sup>43</sup> While KLS confidence intervals are narrower than IV's, the actual correlation rate is unknown, and one should consider the union of confidence intervals over the whole range of correlations, or at least a reasonable one. For high levels (in absolute value) of correlations, the union of KLS confidence sets is around three times larger than the 2SLS confidence interval. Notably, however, KLS and 2SLS confidence intervals overlap for median endogeneity correlations, and the 2SLS point estimates are inside the KLS intervals over a sub-range of slightly positive values. This is at the very least reassuring about the appropriateness of the chosen instrumental variable. Nevertheless, without prior information on the reasonable range of endogeneity, I cannot sharpen the KLS inference.

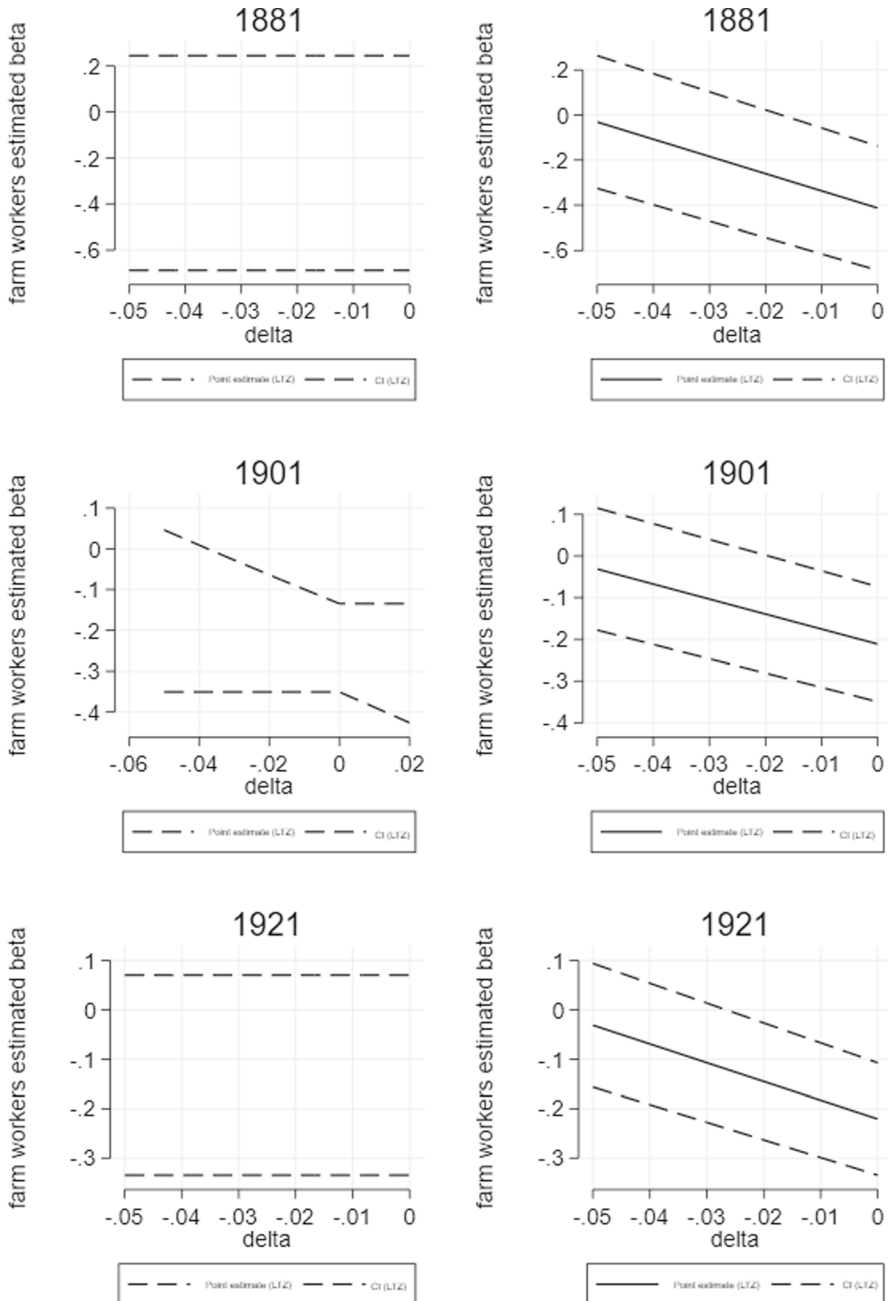
Generally, every year we observe that for a positive range of postulated endogeneity correlations, the KLS confidence intervals never intersect 0, always resulting significantly negative. On the contrary, assuming a negative range of endogeneity, we always notice a cutoff value (vertical red lines) above which we lose confidence in farm workers' significantly negative economic impact on literacy. The opposite holds with priors consistent with a postulated endogeneity range below the same threshold value. Similarly, we find that we do not reject the null hypothesis that the

<sup>40</sup> For a reasonably narrow range of endogeneity correlations, KLS confidence sets are generally narrower than those from IV estimations. Asymptotically conservative confidence intervals are obtained by taking the union of confidence intervals over the considered grid. For further details, see Kiviet (2020).

<sup>41</sup> Similar approaches with the aim to bound a causal impact of a single endogenous regressor in the absence of strong and valid instruments are not new in this literature. Altonji et al. (2005); Oster (2019) impose restrictions on the size of the correlation between the endogenous regressor and the error term relative to the correlation of the endogenous regressor with the other control variables. Virtually bounding this sensitivity parameter may be less intuitive than directly imposing restrictions on the endogeneity correlation itself.

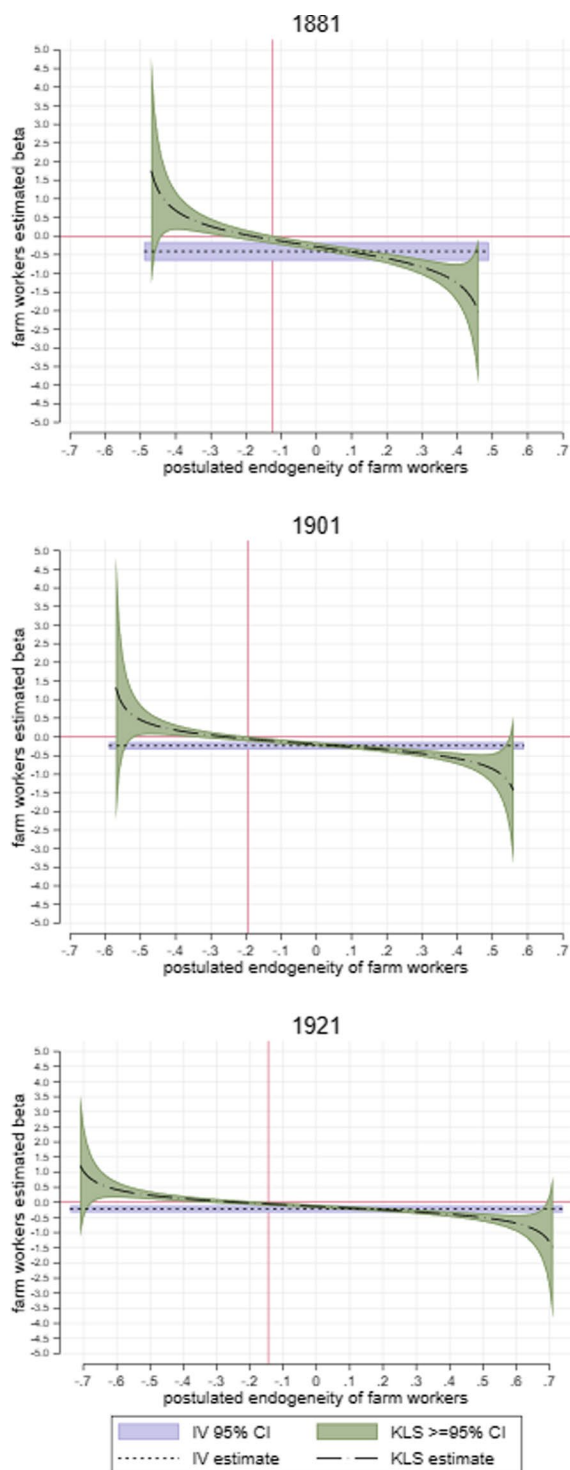
<sup>42</sup> See Section 5.2.2 for a plausible explanation of measurement error of land inequality, especially in the first period of the analysis.

<sup>43</sup> KLS estimates with correlation equal to 0 correspond to OLS estimates.



**Fig. 5** Plausibly exogenous test: 95% interval estimates. *Notes:* The figure presents 95% confidence intervals for the effect of farm workers (share over total labor force in agriculture) on literacy rates in 1881 at the district level. The definition of  $\gamma$  differs between the support-only intervals (UCI, left side) and the interval that uses the full prior (LTZ, right side). In the first approach, I let  $\gamma$  assume a range of values between  $-0.05$  and  $0$ , while in the LTZ approach, I impose a normal distribution with a mean equal to  $\hat{\gamma}$  in the zero-first-stage group and variance equal to  $S_0^2$

**Fig. 6** KLS and 2SLS coefficient estimates and confidence intervals of farm workers for each year. *Notes:* Vertical red lines indicate the point at which KLS coefficient estimates become insignificant



instrument is validly excluded from the model for values ranging from slightly negative to small positive correlation levels.<sup>44</sup> Hence, I add malaria among the set of covariates in the following specification, as the KLS approach has the advantage that we can obtain valid inference without any instruments.<sup>45</sup>

I report graphs of interest in Fig. 7. Directly implied by the exclusion restrictions tests, the coefficient of malaria seems to take on negative values when allowing for negative endogeneity correlations to occur, although the process takes place at a slower speed than with the coefficient of farm workers. This suggests that the coefficient associated with malaria is more stable in comparison and steadily remains at zero for a broader range of endogeneity correlations.

Nevertheless, I can further test the exclusion restriction by employing the level of malaria coefficient in the previous KLS specification as a prior in the plausibly exogenous test, applying Conley et al. (2012). As seen, shaping a prior belief about a plausible range of direct effects could generally be equally complex or even more challenging than agreeing on a reasonable range of endogeneity correlations. Hence, I take the value corresponding to the KLS specification with a negative correlation level equal to the threshold above which the coefficient of farm workers loses significance.<sup>46</sup> Then, I assume that the direct effect is halved and standing at an equal distance between 0 and such a value. Thus, I can finally apply Conley et al. (2012) to obtain the corresponding union of confidence intervals.

Instead of reporting the graph from the plausibly exogenous test, I add the resulting confidence bands,  $[-0.6437; 0.0073]$  in 1881,  $[-0.3333; -0.0142]$  in 1901, and  $[-0.3164; -0.0188]$  in 1921 to the graphical output of the KLS regression.<sup>47</sup> The results are displayed in Fig. 8. Notice that imposing zero as the lower limit for the direct effect of malaria pervasiveness, for it would be a validly excluded instrument, the lower bounds of the plausibly exogenous test coincide with the respective 2SLS bounds. The principal result of this empirical exercise is that we cannot reject the hypothesis that malaria is a valid instrument for the range of endogeneity correlations that do not overstep the previous cutoff values. Only the widened PE confidence bounds for 1881 boiling down from allowing malaria to have a (small) nonpositive direct effect make it more difficult to infer meaningful conclusions. In contrast, the KLS inference remains informative for the 1901 and 1921 regressions as long as I restrict the attention to a subset of postulated endogeneity correlations.

Overall, maintaining the assumption that the endogeneity of malaria is due to measurement error (thus positive), the KLS estimate of the farm workers is still significantly negative. In contrast, if we are sure that omitted variables play a crucial role and the correlation between the endogenous variable and the error term is negative, two possible outcomes may arise. On the one hand, when land inequality is mildly

<sup>44</sup> Results are not reported for brevity. If malaria is not validly excluded from the model, KLS estimates will suffer from omitted-variables bias if any of the included covariates be correlated with the instrument.

<sup>45</sup> For the sake of brevity, the outcomes of the regression are omitted.

<sup>46</sup> This value is equal to  $-0.125$  for 1881,  $-0.194$  for 1901, and  $-0.145$  for 1921.

<sup>47</sup> I followed Kripfganz et al. (2020) for the implementation of KLS regressions and the setting-up of the graphing options.

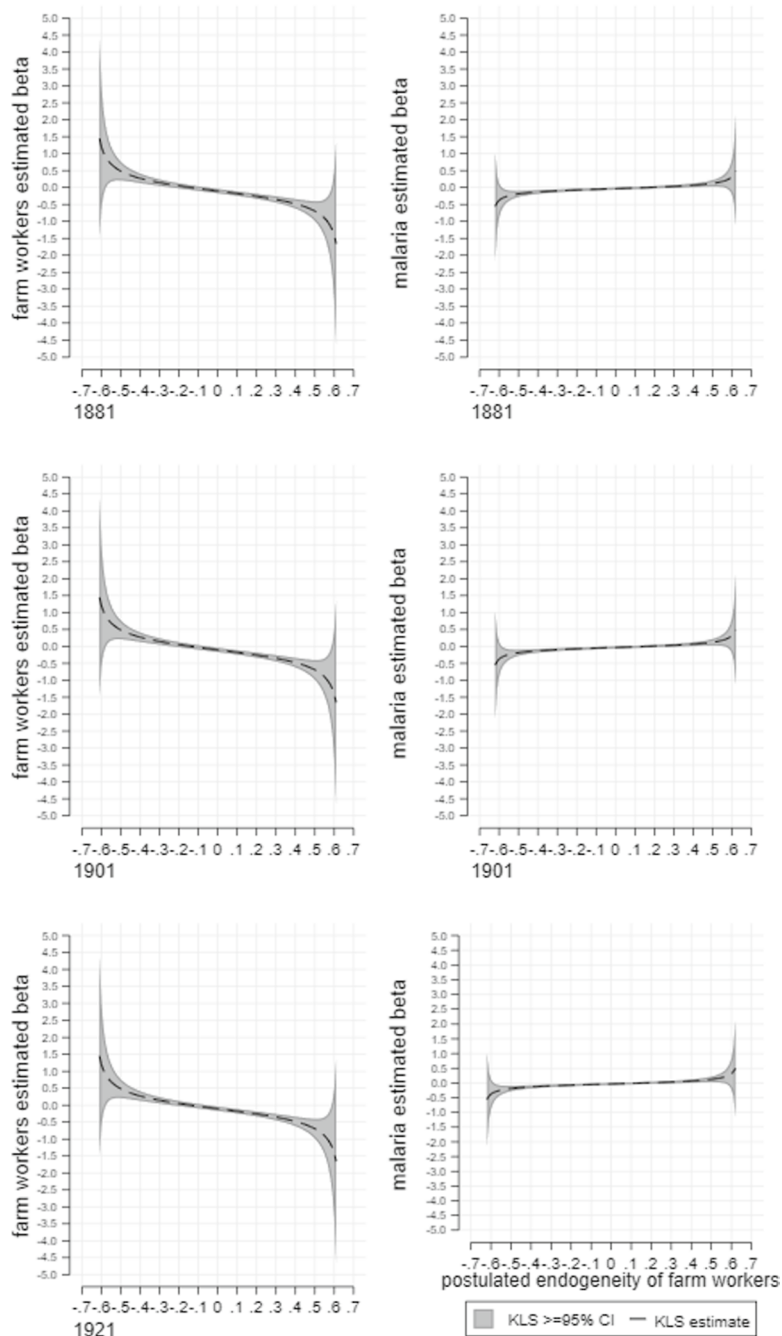
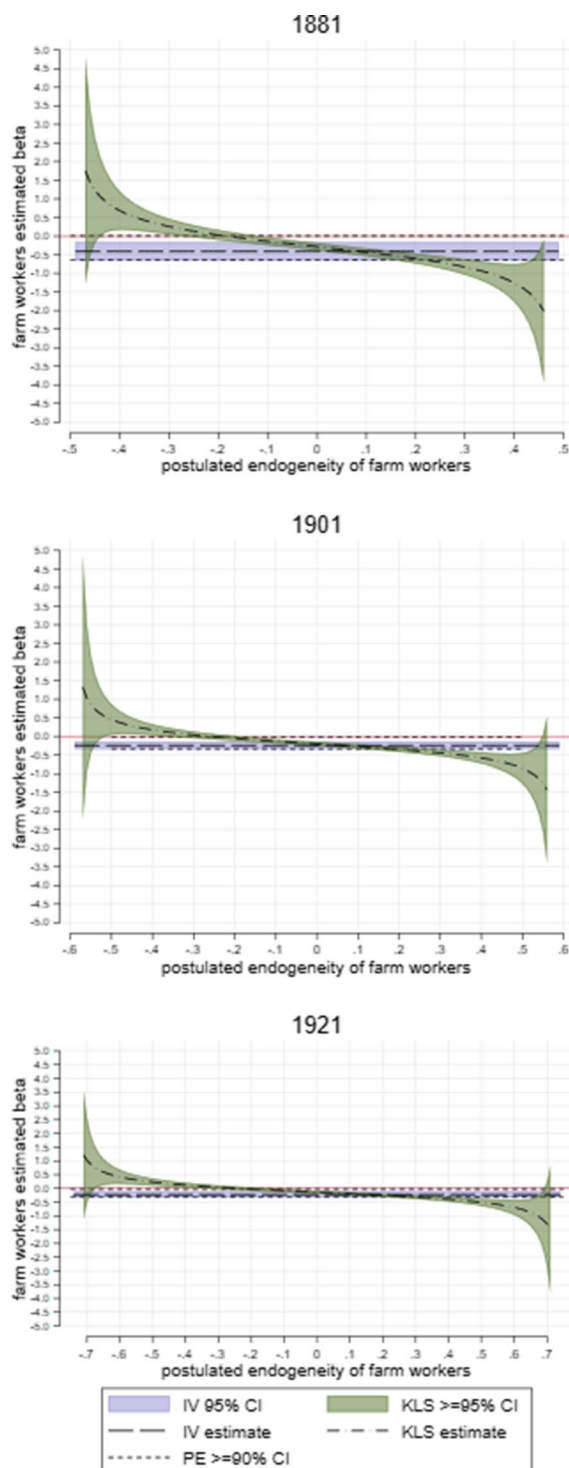


Fig. 7 KLS coefficient estimates with malaria among regressors for each year

**Fig. 8** KLS, 2SLS, and PE coefficient estimates and confidence intervals in the specification with malaria pervasiveness among covariates





endogenous, the associated coefficient remains statistically significantly negative (but smaller), and estimates' bounds encompass the respective 2SLS point estimates. This indicates that malaria pervasiveness might indeed be a valid and relevant instrument. On the other hand, if we accept that the omitted-variables bias is so large that we cannot maintain that our prior belief about the endogeneity correlation ranges between 0 and a small negative value, we are forced to acknowledge that malaria becomes an invalid instrument. However, well-established literature points to measurement error when evaluating the effect of land inequality on education, justifying the upwardly biased OLS coefficient and the difference in magnitude with IV estimates (Cinnirella and Hornung 2016; Easterly 2007; Ramcharan 2010; Beltrán Tapia and Martínez-Galarraga 2018). Therefore, it is unrealistic to assume a largely negative true endogeneity correlation, thus providing support for an identified causal adverse effect of landownership concentration on education during this juncture of Italian history.

#### 5.4 Panel results

Exploration of the panel dimension of the dataset to estimate the within-province and within-district relationship between changes in land inequality and literacy produces similar results. The advantage I have is twofold: consider all the unobserved time-invariant heterogeneity potentially correlated with the land ownership structure (fixed effects) and time shocks common to all units of observation, potentially producing changes in the variable of interest then affecting education levels (time dummies). Furthermore, shocks could have determined a change in the political power of landowning elites. Hence, I interact land inequality with time dummies to account for institutional changes common to all units in a stepwise diff-in-diffs approach. The corresponding equation assumes the following form:

$$Edu_{it} = \alpha_i + \delta_t + \sum_i \beta Land_{i,1881} \delta_t + \sum_i \gamma X'_{it} \delta_t + v_{it} \quad (3)$$

where  $Edu$  are literacy rates (overall and by gender),  $\alpha_i$  and  $\delta_t$  are, respectively, fixed effects and time shocks, while  $X'_{it} \delta_t$  represent the interaction between baseline controls and time dummies. The variable of interest is represented by the interaction term  $\sum_i \beta Land_{i,1881} \delta_t$ ; finally,  $v_{it}$  is the error term. Land inequality is fixed in 1881, and the coefficients for the interaction terms are to be interpreted with respect to this initial value. By doing so, we avoid that changes in farm workers are endogenously driven by those in education levels. I am primarily interested in  $\beta$ 's.

Results are reported in Tables 8 and 9, both for districts and provinces. As mentioned above, I opt for a model specification in which all regressors are interacted with time dummies to account for their marginal effect to change over time (see Cinnirella and Hornung (2016) and Cappelli (2016)). This is due to the different nature of the variables: literacy rates follow a constantly increasing trend, while the fraction of farm workers appears to be slowly changing, only showing remarkable geographical differences *between* regions.<sup>48</sup> The coefficient sign referred to the initial

<sup>48</sup> For a visual representation of the trends in the main variables, see Fig. 9 in the Appendix.

level of land inequality is negative and statistically significant in 1901 and positive but not significantly different from zero in 1921. In the provincial case, with a longer time series at hand, it becomes significant in 1891, reaching a peak in 1901 and then decreasing subsequently, eventually losing significance in 1921. When I consider literacy by gender, the evidence is more mixed. In the district-level panel dataset, while the impact on female literacy rates seems to follow the same path, the effect on male literacy rates is surprisingly positive and significant in 1921. Similarly, in the provincial case, I find a negative and statistically significant effect throughout the considered period on female literacy rates, with the coefficient magnitude becoming smaller and insignificant over time, while the impact on male literacy vanishes even before, eventually turning positive in 1921 (though not significant). In columns 1, 4, and 7, the control variables are allowed to change over time. However, changes in the covariates might be endogenous to those in the distribution of political power, thus biasing our estimates. Similar to farm workers' share, I fix all time-varying control variables in 1881 and interact them with time shocks (columns 2, 5, and 8). The coefficients for interacted landownership concentration become even higher in size, reinforcing the effect on literacy. Interestingly, in the district panel dataset, the impact on literacy for all people and males becomes significantly positive in 1921. The pattern of results for the provincial estimates is hardly affected. In columns 3, 6, and 9, I move one step forward and predict farm laborers in 1881 with malaria in 1881 and use the predicted values interacted with time dummies as the main regressor. Again, while the size of coefficients increases, their time pattern remains unaffected, with a negative impact vanishing before turning positive in 1921 (though not significant). Finally, spatial analysis accounting for spatially correlated standard errors does not alter the main conclusions. Wild bootstrap *p*-values (1,000 repetitions) substantially confirm the main results, although slightly hampering coefficients significance in columns 1, 2, and 5 of Table 8, and columns 1, 2, 4, and 5 of Table 9. Conley SEs adjusted considering a threshold of 200km corroborate previous findings, only reducing the significance of the coefficient interacted with the 1891 dummy.

Overall, a comparison of the coefficients for the interaction terms indicates that districts and provinces exhibit a change over time in the relationship between land ownership concentration and literacy, with the former becoming less and less influential due to a weakening of the political power of local notables. That is, districts and provinces with an initially stronger landownership concentration, while first exhibiting lower literacy rates relative to 1881, see their education levels increase at the end of the period. Take Table 8, column 2 as an example: while in 1901 literacy rates are 1.3 percentage points lower relatively to 1881 in districts with 10 percentage points higher initial share of farm workers, those same districts exhibit in 1921 literacy 1.2 percentage points higher with respect to 1881. Remarkably, the differential seems to increase after 1911, suggesting a greater scope for convergence for more unequal areas.

Similar results conform with the passive modernization process Italy was experiencing at that time, with the imposition of many top-down public policies that, along with minor changes in the land ownership structure, may have reduced its impact on

**Table 8** Panel results: district level

Dep. var.:	Literacy	Literacy	Literacy	Literacy M	Literacy M	Literacy F	Literacy F	Literacy F
Model specification:	Main	Controls	Malaria	Main	Controls	Main	Controls	Malaria
	(1)	(2)	(3)	(4)	(5)	(7)	(8)	(9)
Farm workers x 1901	-0.126*** (0.032)	-0.139*** (0.029)	-0.337*** (0.085)	-0.039 (0.029)	-0.053** (0.025)	-0.194*** (0.038)	-0.210*** (0.036)	-0.530*** (0.115)
Conley SE (200 km)	[0.058]**	[0.055]**	[0.085]***	[0.051]	[0.046]	[0.067]***	[0.066]***	[0.115]***
<i>p-value farm workers</i>								
Wild bootstrap	{0.155}	{0.121}*	{0.000}***	{0.599}	{0.440}	{0.080}**	{0.084}**	{0.000}***
Farm workers x 1921	0.048 (0.053)	0.123** (0.055)	0.178 (0.144)	0.135*** (0.048)	0.232*** (0.049)	-0.003 (0.066)	0.053 (0.066)	0.064 (0.173)
Conley SE (200 km)	[0.071]	[0.069]*	[0.144]	[0.065]**	[0.061]***	[0.080]	[0.078]	[0.173]
<i>p-value farm workers</i>								
Wild bootstrap	{0.662}	{0.522}	{0.481}	{0.472}	{0.163}**	{0.976}	{0.624}	{0.768}
<i>N</i>	615	615	615	615	615	615	615	615
<i>p-value: <math>\beta_{1901} = \beta_{1921}</math></i>	0.000	0.000	t-stat: [-3.06]***	0.000	0.000	0.000	0.000	t-stat: [-3.35]***

Fully flexible district-level panel estimates. The share of farm workers is interacted with time dummies for the years 1881, 1901, and 1921. The omitted reference year is 1881. Baseline controls are interacted with time dummies. All columns use the initial level of farm workers; columns 2, 3, 5, 6, 8, and 9 fix all control variables in 1881; columns 3, 6, and 9 use predicted values of farm workers using malaria as instrument. All regressions include time shocks and fixed effects. Heteroskedastic-robust standard errors are reported in parentheses for separate cross-sections. *P*-values of wild cluster bootstrap SE of farm workers clustered at the regional level are reported in curly brackets (stars relate to SE clustered at the regional level); Conley SE (200 km) in square brackets. Robust standard errors are bootstrapped to account for uncertainty stemming from using a predicted variable in the second-stage estimates in columns 3, 6, and 9. \*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

**Table 9** Panel results: province level

Dep. var.: Model specification:	Literacy Main	Literacy Controls	Literacy Malaria	Literacy M Main	Literacy M Controls	Literacy M Malaria	Literacy F Main	Literacy F Controls	Literacy F Malaria
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Farm workers x 1891	-0.100*** (0.034)	-0.112*** (0.030)	-0.200** (0.082)	-0.066** (0.027)	-0.069** (0.026)	-0.108 (0.066)	-0.133*** (0.044)	-0.154*** (0.037)	-0.289*** (0.106)
Conley SE (200 km)	[0.069]	[0.072]	[0.082]**	[0.056]	[0.056]	[0.066]	[0.085]	[0.090]*	[0.106]***
<i>p-value farm workers</i>									
Wild bootstrap	{0.082}*	{0.055}**	{0.002}**	{0.190}	{0.114}*	{0.042}	{0.081}**	{0.022}**	{0.000}***
Farm workers x 1901	-0.199*** (0.065)	-0.241*** (0.065)	-0.434*** (0.156)	-0.105*** (0.051)	-0.135*** (0.052)	-0.209* (0.116)	-0.286*** (0.085)	-0.338*** (0.083)	-0.642*** (0.206)
Conley SE (200 km)	[0.066]***	[0.069]***	[0.156]***	[0.053]**	[0.053]**	[0.116]*	[0.081]***	[0.086]***	[0.206]***
<i>p-value farm workers</i>									
Wild bootstrap	{0.082}**	{0.053}**	{0.001}***	{0.239}	{0.148}*	{0.044}*	{0.036}**	{0.027}**	{0.000}***
Farm workers x 1911	-0.163** (0.075)	-0.202*** (0.072)	-0.370** (0.153)	-0.048 (0.060)	-0.055 (0.057)	-0.085 (0.114)	-0.268*** (0.097)	-0.336*** (0.093)	-0.627*** (0.215)
Conley SE (200 km)	[0.065]**	[0.070]***	[0.153]**	[0.055]	[0.055]	[0.114]	[0.087]***	[0.088]***	[0.215]***
<i>p-value farm workers</i>									
Wild bootstrap	{0.181}	{0.113}*	{0.007}**	{0.572}	{0.531}	{0.435}	{0.064}**	{0.041}**	{0.000}***
Farm workers x 1921	-0.088 (0.102)	-0.081 (0.104)	-0.343 (0.263)	0.005 (0.085)	0.051 (0.088)	-0.093 (0.221)	-0.178 (0.124)	-0.207 (0.124)	-0.569* (0.318)

**Table 9** (continued)

Dep. var.:	Literacy Main	Literacy Controls	Literacy Malaria	Literacy M Main	Literacy M Controls	Literacy M Malaria	Literacy F Main	Literacy F Controls	Literacy F Malaria
Model specification:									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Conley SE (200 km)	[0.086]	[0.092]	[0.263]	[0.072]	[0.077]	[0.221]	[0.104]*	[0.110]*	[0.318]*
<i>p-value farm workers</i>									
Wild bootstrap	{0.583}	{0.600}*	{0.133}	{0.969}	{0.688}	{0.641}	{0.349}	{0.277}	{0.041}*
<i>N</i>	345	345	345	345	345	345	345	345	345
<i>p-value: <math>\beta_{t_1,t_2} = \beta_{t_3,t_4}</math></i>	0.598	0.439	t-stat: [0.29]	0.110	0.009	t-stat: [-0.62]	0.808	0.645	t-stat: [0.80]

Fully flexible province-level panel estimates. The share of farm workers is interacted with time dummies for the years 1881, 1891, 1901, 1911, and 1921. The omitted reference year is 1881. Baseline controls are interacted with time dummies. All columns use the initial level of farm workers; columns 2, 3, 5, 6, 8, and 9 fix all control variables in 1881; columns 3, 6, and 9 use predicted values of farm workers using malaria as instrument. All regressions include time shocks and fixed effects. Heteroskedastic-robust standard errors are reported in parentheses for separate cross-sections. *P*-values of wild cluster bootstrap SE of farm workers clustered at the regional level are reported in curly brackets (stars relate to SE clustered at the regional level); Conley SE (200 km) in square brackets. Robust standard errors are bootstrapped to account for uncertainty stemming from using a predicted variable in the second-stage estimates in columns 3, 6, and 9. \*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

education through time. If the conjecture is correct, reforms of the end of the nineteenth century and the beginning of the twentieth century somewhat changed the relationship between the landed nobility and the peasantry and their incentives to invest in education. Indeed, democratization and the achievement of political equality in Western European countries during the late nineteenth and early twentieth century were generally followed by rapid public policy changes. Lindert (2004) defined it as “the 1880-1930 laboratory,” documenting the historical proximity between franchise extension and public provision of education, increased spending in social transfers, labor market reforms, and creation of income tax systems. Post-unitary Italy is part of this laboratory, as local governments were in charge of providing many pro-growth public goods, including education and healthcare. Their decisions were likely to have determined long-lasting regional variation in income per capita. The *Daneo-Credaro* bill, enacted in 1911, is only one example of similar top-down policies. Another fundamental structural break is the piecemeal enfranchisement process experienced in Italy at different stages, with the stepwise enlargement of the electoral franchise for male people as late as 1919. All these institutional shocks weakened the political power of landowning elites, thus limiting their ability to intervene in public choice in later stages of development. In this case, it is reasonable to expect such a variable to have had lesser explanatory power regarding human capital accumulation.

In the next section, I explore the mechanism of transmission of the effect by focusing on the effect of land inequality on some intermediate measures of supply and demand of schooling.

## 6 Mechanism of transmission: supply vs demand

Literacy rates represent a so-called equilibrium measure of human capital. Despite finding a negative and statistically significant relationship with the share of farm workers, approximating land inequality, I still do not know whether this impact runs through either a demand or a supply channel. On the one hand, an unequal land ownership structure implies a more concentrated economic and political power in the hands of a few large landowners. This is particularly relevant for the purpose of this study since local notables, who took over municipal councils, could influence the decisions to invest in mass schooling (certainly until 1911) and prevent the majority of the population from gaining decisional power. On the other hand, a high fraction of annual and daily workers in agriculture, a social class owing few economic resources and facing a binding budget constraint, had lower incentives to invest in human capital, given its low payoff and uncertain long-term returns.

As discussed in Section 1, Galor et al. (2009) assert that human capital is complementary to industry-related tasks, most widespread in urban and industrialized areas. In contrast, the rural labor force is not required to possess literacy skills. Thus, large landowners' stake was to reduce and possibly oppose human capital-enhancing institutions, limit migration, and maintain low agricultural wages. By contrast, capitalists were in favor of supporting institutions promoting education. To explore how land inequality affects human capital accumulation, I re-estimate the IV model

restricting the analysis to rural districts and comparing results to those obtained using the entire sample.<sup>49</sup> We expect the coefficient associated with land inequality to be even larger where landowning elites were more able not to favor the expansion of the provision of mass schooling in rural areas.

Results are reported in Table 10. I employ malaria pervasiveness as the exogenous instrument in all specifications and include the complete set of controls.<sup>50</sup> Once comparing the estimated coefficients with those reported in Table 4, we observe that the size ranges from 1.75 to 1.3 times higher in the subsample containing districts with urbanization below the 33rd percentile. In contrast, it ranges from 1.4 to 1.16 in the subsample containing districts registering urbanization rates below the median. The supply-side mechanisms stressed by Galor et al. (2009) seem to be confirmed by this evidence.

The landed nobility retaining political power in rural areas was liable for constructing schooling infrastructures, school teachers' appointments, and municipal fiscal resources management. Thus, I employ these intermediate outputs approximating schooling supply to test further whether landowning elites could obstruct education provision. Specifically, I use child-teacher, school density, and municipal expenditures per child aggregated at the provincial level. Besides, I gather information on enrollment rates and child labor to assess the role of rural households' demand for education. Unfortunately, I can barely conduct IV regressions without the possibility to exploit the panel dimension of the dataset because this information is only observed for 1881. Moreover, data at a more granular district level is only available for child labor, while enrollment rates are only obtainable for provinces.<sup>51</sup> They are constructed by dividing people enrolled in primary schools over those aged between six and fifteen. Moreover, it is possible to disentangle the variable by gender.

The baseline specification reports IV estimates, employing malaria as an excluded instrument and including the set of baseline controls.<sup>52</sup> Notwithstanding that these variables aim to proxy supply and demand factors, they are somehow endogenous. Indeed, a larger number of teachers appointed to school and a higher level of school density may induce children to attend school. Conversely, enrollment rates, capturing children's demand to go to school, may boost the supply of schooling. The same argument applies to municipal expenditures per capita, which may be greater where education demand is high. Nonetheless, assessing the impact of inequality on these schooling outputs remains highly informative per se.

Results are reported in Table 11. I find a positive and statistically significant association between farm workers and the ratio between children and teachers. This finding suggests that where landownership was highly concentrated, the number

<sup>49</sup> Restricted rural sample excludes those districts with urbanization rates, alternatively, higher than 33rd percentile and the median value of the cross-sectional sample for each year. Eventually, the two remaining subsamples will count 67 out of 205 districts in the first and 102 out of 205 districts in the second.

<sup>50</sup> It corresponds to the specification used in Table 4. Note that the direct effect of urbanization was already accounted for in the main specification of the same table.

<sup>51</sup> For further details on the definition and measurement of the variables referred to supply and demand factors of schooling, see Table 16, in the Appendix.

<sup>52</sup> It corresponds to the specification in Table 4.

**Table 10** Restricted sample: IV with rural districts

	IV 1881 (urb. rates <33rd perc.) (1)	IV 1901 (urb. rates <33rd perc.) (2)	IV 1921 (urb. rates <33rd perc.) (3)
Farm workers	−0.720*** (0.193)	−0.326*** (0.067)	−0.290*** (0.108)
<i>N</i>	67	67	67
<i>F</i>	97.22	237.2	115.7
	IV 1881 (urb. rates <median)	IV 1901 (urb. rates <median)	IV 1921 (urb. rates <median)
Farm workers	−0.590*** (0.155)	−0.288*** (0.062)	−0.285*** (0.080)
<i>N</i>	103	102	103
<i>F</i>	144.2	418.8	175.6

2SLS with districts having urbanization rates smaller than the 33rd percentile and the median. Heteroskedastic-robust standard errors are reported in parentheses for separate cross-sections. \*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

of teachers appointed to schools was lower, likely resulting from a minor effort to expand education. The significance seems to vanish over time, suggesting this supply channel worked more vigorously in the first stages of the period under analysis and lost strength over time with changes influencing power balance. When it comes to school density, I find a negative association, though not statistically significant.<sup>53</sup> This means that disparities in the construction of schoolhouses between areas with greater and lower levels of landownership concentration had been almost non-existent as late as the rise of the Fascist party. The effort spent to build them up appears to be the same across Italy's municipalities.<sup>54</sup> Similarly, except for the first period under analysis, I find no evidence of a significant systematic correlation when investigating the relationship between farm workers and the value of municipal expenditures per child.<sup>55</sup>

The sufficient cross-sectional and temporal variation allows the implementation of panel estimates for all three outcomes. I do not consider the presence of malaria as an exogenous instrument due to its lack of time variation. However, as in Section 5.4, it is possible to account for unobserved provincial heterogeneity (fixed effects) and the presence of temporal shocks common to all provinces (time dummies). Furthermore, I employ the exact model specification as in Table 9, with the share of farm workers fixed at the beginning of the period and interacting with time shocks. Even covariates interact with time dummies in the vein of the already employed stepwise diff-in-diffs estimation. Columns 2, 5, and 8 fix all control

<sup>53</sup> I do not have available data on school density in 1871 and 1921.

<sup>54</sup> Nevertheless, further concerns may be raised if one takes into consideration the role of road infrastructures. Indeed, following this argument, in the most disadvantaged and mountainous areas, they may have prevented a vast majority of the population from easily reaching schoolhouses.

<sup>55</sup> I do not have data referred to municipal expenditures available for 1921.



**Table 11** Mechanism of transmission: supply

	1871	1881	1891	1901	1911	1921
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: IV						
<i>Dep. var.:</i>	<i>Child-teacher ratio</i>					
Farm workers (share)	0.491 (0.302)	0.488* (0.253)	0.317** (0.151)	0.236** (0.115)	0.286** (0.135)	0.154** (0.073)
Conley SE (200 km)	[0.215]**	[0.200]**	[0.112]***	[0.086]***	[0.106]***	[0.061]**
<i>p-value farm workers</i>						
Wild bootstrap	{0.031]**	{0.082]**	{0.028}***	{0.030}***	{0.006}***	{0.022]**
Panel B: IV						
<i>Dep. var.:</i>	<i>School density</i>					
Farm workers (share)		−0.008 (0.008)	−0.005 (0.006)	−0.003 (0.004)	−0.002 (0.005)	
Conley SE (200 km)		[0.006]	[0.005]	[0.004]	[0.004]	
<i>p-value farm workers</i>						
Wild bootstrap		{0.158}	{0.337}	{0.407}	{0.593}	
Panel C: IV						
<i>Dep. var.:</i>	<i>Municipal expenditures (per child)</i>					
Farm workers (share)	−25.817** (11.606)	−22.761** (11.370)	−17.522 (13.574)	−15.414 (13.822)	−34.882 (28.707)	
Conley SE (200 km)	[11.115]**	[9.259]**	[12.546]	[12.813]	[21.764]	
<i>p-value farm workers</i>						
Wild bootstrap	{0.053]**	{0.066]**	{0.183}*	{0.411}	{0.322}	

2SLS province-level estimates. Heteroskedastic-robust standard errors are reported in parentheses for separate cross-sections. *P*-values of wild cluster bootstrap SE of farm workers clustered at the regional level are reported in curly brackets (stars relate to SE clustered at the regional level); Conley SE (200 km) in square brackets. \*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

variables in 1881, and columns 3, 6, and 9 use predicted values of farm workers in 1881 using the malaria index as an instrument.

Results are reported in Table 12. In the first three columns, I find a negative and statistically significant effect of farm workers on the child-teacher ratio. Also, the impact displays an increasing trend in magnitude. Take column two as an example: the parameter ranges from −0.1 for 1891 to a value of −0.2 for 1921. It means that geographic areas that exhibit high levels of land inequality at the beginning of the period register the highest decreases in child-teacher ratios through time. Besides, the effect magnitude is even reinforced moving forward in time, apparently suggesting an increasing role of the supply process in expanding public schooling.

Nevertheless, while this finding might seem in contrast with cross-sectional results and appear surprising and somehow counterintuitive, the stepwise diff-in-diffs estimates must be interpreted differently. They suggest that the channel referred to the appointment of teachers into schoolhouses becomes relevant as long as the distribution of political power diverts away from landholders. This is evidenced

**Table 12** Mechanism of transmission: panel estimates

Dep. var.:	Child- teacher	Child-teacher	Child- teacher	Mun. exp.	Mun. exp.	Mun. exp.	School density	School density	School density
Model specification:	Main	Controls	Malaria	Main	Controls	Malaria	Main	Controls	Malaria
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Farm workers x 1891	-0.100*** (0.027)	-0.104*** (0.027)	-0.184** (0.075)	-5.92 (5.973)	-5.746 (5.572)	-12.791 (15.118)	-0.000 (0.001)	-0.000 (0.001)	0.000 (0.003)
Conley SE (200 km)	[0.029]***	[0.028]***	[0.075]**	[7.200]	[7.316]	[15.118]	[0.001]	[0.001]	[0.003]
<i>p-value farm workers</i>									
Wild bootstrap	{0.020}**	{0.005}**	{0.048}**	{0.312}	{0.301}	{0.044}	{0.891}	{0.892}	{0.971}
Farm workers x 1901	-0.088** (0.034)	-0.081** (0.031)	-0.133 (0.081)	-15.184 (9.365)	-14.511* (8.438)	-28.686 (17.951)	-0.000 (0.001)	-0.001 (0.001)	-0.001 (0.002)
Conley SE (200 km)	[0.029]***	[0.027]***	[0.081]	[8.235]*	[7.848]*	[17.951]	[0.001]	[0.001]	[0.002]
<i>p-value farm workers</i>									
Wild bootstrap	{0.061}**	{0.053}*	{0.228}	{0.207}	{0.193}	{0.045}	{0.872}	{0.678}	{0.661}
Farm workers x 1911	-0.110** (0.043)	-0.124*** (0.041)	-0.198** (0.098)	-32.921* (18.911)	-46.791** (19.612)	-64.811 (47.625)	0.001 (0.001)	-0.000 (0.001)	-0.000 (0.002)
Conley SE (200 km)	[0.031]***	[0.027]***	[0.098]**	[12.493]***	[11.319]***	[47.625]	[0.001]	[0.001]	[0.002]
<i>p-value farm workers</i>									
Wild bootstrap	{0.097}*	{0.029}**	{0.085}**	{0.174}*	{0.040}**	{0.101}	{0.683}	{0.983}	{0.838}
Farm workers x 1921	-0.150*** (0.052)	-0.203*** (0.048)	-0.379*** (0.145)						

Table 12 (continued)

Dep. var.:	Child-teacher	Child-teacher	Child-teacher	Mun. exp.	Mun. exp.	Mun. exp.	School density	School density
Model specification:	Main	Controls	Malaria	Main	Controls	Malaria	Main	Controls
	1881	1881	predicted	1881	1881	predicted	1881	1881
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Conley SE (200 km)	[0.044]***	[0.034]***	[0.145]***					
<i>p-value farm workers</i>								
Wild bootstrap	{0.088}**	{0.020}***	{0.013}***					
<i>N</i>	345	345	345	276	276	276	276	276
<i>p-value: <math>\beta_{t_1,t_2} = \beta_{t_3,t_4}</math></i>	0.144	0.004	t-stat: [1.73]*	0.403	0.052	t-stat: [0.49]	0.344	0.558
								t-stat: [-0.19]

Fully flexible province-level panel estimates. The share of farm workers is interacted with time dummies for the years 1881, 1891, 1901, 1911, and 1921. The omitted reference year is 1881. Baseline controls are interacted with time dummies. All columns use the initial level of farm workers; columns 2, 3, 5, 6, 8, and 9 fix all control variables in 1881; columns 3, 6, and 9 use predicted values of farm workers using malaria as instrument. All regressions include time shocks and fixed effects. Heteroskedastic-robust standard errors are reported in parentheses for separate cross-sections. *P*-values of wild cluster bootstrap SE clustered at the regional level are reported in curly brackets (stars relate to SE clustered at the regional level); Conley SE (200 km) in square brackets. Robust standard errors are bootstrapped to account for uncertainty stemming from the use of a predicted variable in the second-stage estimates in columns 3, 6, and 9. \*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

by the coefficient magnitude associated with the interaction term in 1921, which is almost twice the previous ones. After registering steady negative values in the first three periods, we observe the coefficient plummeting sharply, in conjunction with the introduction of the *Daneo-Credaro* law in 1911. Thus, higher investments in teachers are to be attributed to the central government.

The following columns display findings using municipal expenditures per child and school density as outcome variables. While we observe a modest role of political power in affecting public expenditure, school density is never influenced by the interaction between the initial level of farm workers and time shocks. Finally, spatial analysis clustering standard errors with the Conley method and wild bootstrap estimation confirm the main findings in both tables. Therefore, while findings corroborate the results of the IV model specifications, further research is needed to find out more about which supply factors mattered most to transmit the effect, if any.

As for the demand side of the story, I test the effect on enrollment rates and child labor. Although there is a high correlation with literacy rates, enrollment rates should accurately capture decisions to attend school and then the demand for education. This is because literacy rates are referred to the whole population aged more than six, thereby including people who could have acquired such a skill years before. In this case, the contemporaneous correlation with land inequality would say little about the presence of literate within a district or a province. The literacy endowment of a specific territory may thus represent the result of private agents' decisions undertaken years or even decades earlier.<sup>56</sup> By contrast, people enrolled in primary schools are prevalently children, thus only representing a population cohort. Hence, they should better reflect the decisions of the families to invest in their children's education.

Results are reported in Table 13, panel A. Unlike estimates with school density and municipal expenses as outcomes, land inequality negatively affects enrollment rates in every specification (overall and by gender). Moreover, the magnitude of the effect is stronger than the one on the schooling supply outcomes. For instance, relative to the regression with the child-teacher ratio as the outcome variable for the same year (Table 11, column 2), the coefficient is more than 1.5 times higher. Unfortunately, contrary to the case of the child-teacher ratio, here we cannot investigate the robustness of the correlation between the two phenomena by evaluating whether the effect vanishes over time.

Nevertheless, most often, the number of children enrolled at school does not coincide with the proportion of children attending school. Absenteeism rates were elevated in depressed areas, and the law was not equally enforced throughout the territory. Since attendance rates are not observed for the period under analysis, I rely on district-level data on child labor in 1881 by gender to approximate the decision of families to send their children to work.<sup>57</sup> Indeed, child labor is the alternative to education.

<sup>56</sup> The concern due to lagged gap between the decision to invest in human capital and the resulting outcome has been already addressed considering lagged values of the variable of interest in Section 5.

<sup>57</sup> Population Census 1881 reports data on children performing all work duties. Nevertheless, I collect data only on children engaged in agricultural tasks because total child labor coincides with the entire population aged between 9 and 15. Hence, we do not distinguish between children enrolled in primary school and children involved in work duties outside agriculture.

**Table 13** Mechanism of transmission: demand

	1881 (1)	1881 (2)	1881 (3)
Panel A (IV, province level)			
<i>Dep. var.:</i>	<i>Enrollment rate</i>	<i>Enrollment rate F</i>	<i>Enrollment rate M</i>
Farm workers	−0.743** (0.302)	−0.628** (0.288)	−0.858** (0.343)
Conley SE (200 km)	[0.357]**	[0.310]**	[0.426]**
<i>p-value farm workers</i>			
Wild bootstrap	{0.047}**	{0.100}*	{0.036}**
Obs.	69	69	69
Panel B (IV, district level)			
<i>Dep. var.:</i>	<i>Agrarian child labor</i>	<i>Agrarian child labor F</i>	<i>Agrarian child labor M</i>
Farm workers	0.154* (0.092)	−0.054 (0.106)	0.339*** (0.130)
Conley SE (200 km)	[0.120]	[0.081]	[0.200]*
<i>p-value farm workers</i>			
Wild bootstrap	{0.368}	{0.681}	{0.161}
Obs.	205	205	205

2SLS estimates. Heteroskedastic-robust standard errors are reported in parentheses for separate cross-sections. *P*-values of wild cluster bootstrap SE of farm workers clustered at the regional level are reported in curly brackets (stars relate to SE clustered at the regional level); Conley SE (200 km) in square brackets. \*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

Doepke and Zilibotti (2005) propose a theoretical approach in which unskilled workers compete with children in the labor market, facing a tradeoff. On the one hand, unskilled workers support child labor regulations to raise their own wages. On the other hand, they oppose it when their working children provide plenty of family revenue. Furthermore, historical evidence observes a preference for child labor regulations in the first phases of industrialization because of an increasing return to education. Hence, I expect a higher proportion of child labor to be associated with greater landownership concentration.

Panel B of Table 13 reports results using agricultural child labor as the outcome variable, overall and divided by gender. When I do not distinguish by gender, regression outputs only show a modest significant result. Nevertheless, a positive and statistically significant impact is found once I account only for male child labor. This aligns with the argument that only males were devoted entirely to working the land, albeit sometimes female involvement in agricultural tasks was likely to occur.<sup>58</sup> Both methods to spatially cluster SE produce consistent results.

<sup>58</sup> “Given the gender division of labor in the South, where women were rarely employed in the most infection-exposed occupations, malaria in many provinces was also largely a disease of men and boys. In the interior of Sicily, for example, the Faina report found that agricultural work “is unthinkable for any women except the most impoverished - wives abandoned by their husbands, widows, and young girls who have experienced some disaster” (Snowden 1999). See also Beltrán Tapia and Martínez-Galarra (2018) for a similar explanation for interwar Spain.

Finally, I move one step forward by considering only people aged between 15 and 19 who have acquired literacy skills. In this case, the pattern of literate should represent the coeval result of the decisions of households to invest in education, the alternative to their decision to send their children to work. Following a similar argument, I expect farm workers to be negatively correlated with literacy computed for people aged between 15 and 19, in contrast to coefficient signs for agricultural child labor. Hence, I run regressions employing this measure at the provincial level as the outcome, finding a negative and significant effect of land inequality. Results are robust to different model specifications and estimators (OLS and IV) and different years.<sup>59</sup> The results are less striking in panel regressions, employing the usual step-wise diff-in-diffs approach. Only when controls are held fixed at the beginning of the period do we find provinces with a larger share of farm workers registering a significant increase in literacy for people aged between 15 and 19 in 1911. This finding is in line with the vanishing size of the coefficients in the cross-sectional estimates. De facto political power and landownership concentration play a constantly decreasing role in the ongoing modernization process of Italy.

Therefore, although the findings do not sweep away the role of mass schooling supply and decisions of landowning elites in promoting it, they evidence the existence of alternative factors explaining the negative relationship between land inequality and education. Similar demand-based incentives may be related to opportunities outside agriculture, to which payoffs of literacy skills are attached, and binding budget constraints of the rural population during this stage of development. Further research would enhance knowledge on the identification of the proper mechanism.

## 7 Conclusions

Relying on an adequate indicator of land inequality in post-unification Italy, a uniquely assembled database both at the district and provincial levels offers the possibility to investigate whether and how landownership concentration affected literacy rates. I find that areas that exhibit higher levels of landownership concentration show systematically lower literacy rates throughout the period under analysis. Exploration of the panel dimension of the dataset shows that the magnitude and significance of the effect decrease over time, in line with the “passive modernization process” hypothesis. An instrumental variable approach based on malaria virulence addresses omitted variables and reverse causality concerns. Under the assumption that malaria is exogenously determined, IV regressions corroborate OLS results, confirming a negative relationship between landownership concentration and literacy. Moreover, the findings seem particularly robust to a different size of the sample (districts and provinces) and the inclusion of additional controls, whose omission might have violated the exclusion restriction hypothesis. Furthermore, several falsification tests to validate the identification strategy confirm the hypothesis of impact causality.

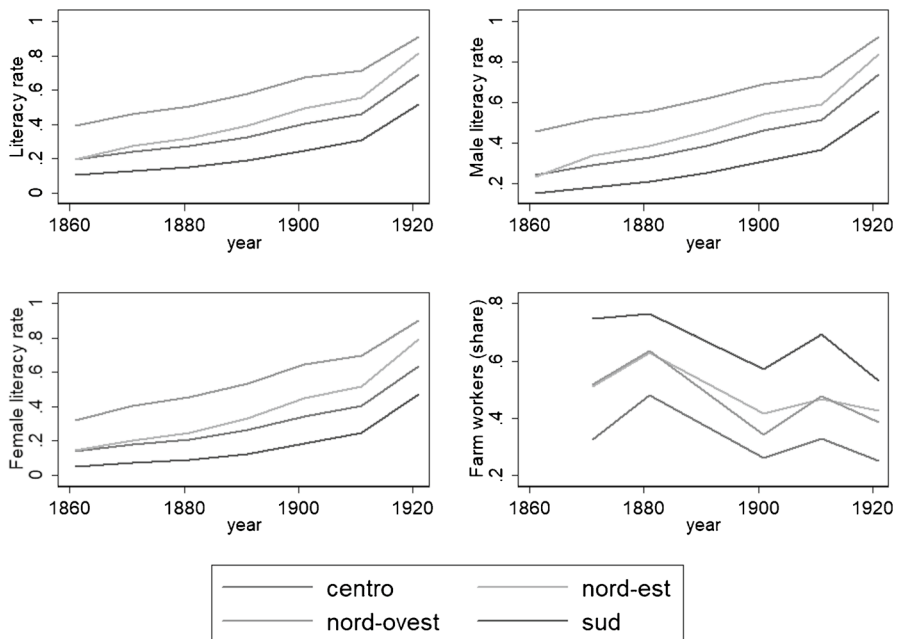
Another remarkable result is obtained when analyzing the effect’s transmission mechanism. Notwithstanding the difficulty of distinguishing between supply and

<sup>59</sup> The results are reported in the Appendix, Tables 19 and 20.

demand factors, not only land inequality may have adversely affected literacy rates influencing schooling supply through the political process, but also through private education demand. Results employing several measures of intermediate outputs of schooling supply and demand confirm this conjecture, showing that the latter played an even major role in determining disparities in human capital accumulation. Indeed, once I employ enrollment rates, child labor, and literacy rates for people aged between 15 and 19 as outcomes, results show a robust association with landownership concentration, indicating that rural households preferred utilizing their children's workforce to perform work duties rather than sending them to school. Thus, where land ownership was highly concentrated, the economic condition of the vast majority of the rural population was astonishingly precarious, and incentives to invest in education were beyond the binding budget constraints of large segments of society. Likewise, the opportunities to acquire literacy skills were limited because the possibilities of upward mobility were almost non-existent.

Therefore, although alternative explanations are not to discard, the findings of this work shed new light on the mechanism through which land inequality shapes different human capital accumulation paths, highlighting the role of malaria in creating and maintaining *latifundism* in the first place.

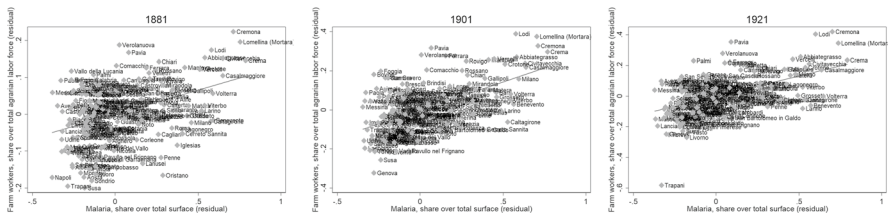
## Appendix



**Fig. 9** Literacy rate by gender and for each macroregion, and share of farm workers over total labor force in agriculture (1861–1921)



**Fig. 10** Each dot indicates a district in the sample. The y-axis in the three panels plots the residual variation in literacy rates in 1881 (left panel), 1901 (middle panel), and 1921 (right panel) after controlling for the baseline controls and the excluded instrument employed in the empirical analysis (see Tables 5 and 6). The x-axis plots the residual variation in land inequality, proxied by the fraction of farm workers over the total agrarian labor force, conditional on the baseline controls and the excluded instrument



**Fig. 11** Each dot indicates a district in the sample. The y-axis in the three panels plots the residual variation in the fraction of farm workers over the total agrarian labor force in 1881 (left panel), 1901 (middle panel), and 1921 (right panel) after controlling for the baseline controls employed in the empirical analysis (see Tables 5 and 6). The x-axis plots the residual variation in the presence of malaria, conditional on the baseline controls



**Table 14** Summary of the variables: district level

	Mean	Median	Std. dev.	10th perc	90th perc
1881					
Literacy	0.35	0.28	0.20	0.15	0.65
Literacy (male)	0.42	0.35	0.19	0.23	0.73
Literacy (female)	0.28	0.21	0.20	0.07	0.60
Farm workers (share over total labor force in agriculture)	0.66	0.70	0.16	0.40	0.86
Population (log)	11.64	11.62	0.62	10.92	12.48
Urbanization rate	0.28	0.22	0.24	0.00	0.64
Latitude	47.14	47.77	2.91	42.23	50.49
Landlocked	0.56	1.00	0.50	0.00	1.00
Sharecroppers (share over total labor force in agriculture)	0.13	0.05	0.18	0.00	0.43
Agrarian labor force (share over total population)	0.39	0.38	0.11	0.25	0.53
Malaria 1880	0.30	0.21	0.29	0.00	0.77
Scattered dwellings	0.27	0.21	0.21	0.03	0.59
Agrarian child labor (male)	0.24	0.23	0.08	0.13	0.34
Agrarian child labor (female)	0.16	0.15	0.10	0.03	0.30
Agrarian child labor	0.20	0.20	0.08	0.09	0.30
1901					
Literacy	0.48	0.41	0.22	0.23	0.82
Literacy (male)	0.54	0.49	0.20	0.31	0.84
Literacy (female)	0.42	0.34	0.24	0.16	0.79
Farm workers (share over total labor force in agriculture)	0.44	0.42	0.21	0.19	0.73
Population (log)	11.78	11.75	0.63	10.99	12.63
Urbanization rate	0.32	0.25	0.26	0.00	0.73
Latitude	47.14	47.77	2.91	42.23	50.49
Landlocked	0.56	1.00	0.50	0.00	1.00
Sharecroppers (share over total labor force in agriculture)	0.21	0.13	0.21	0.03	0.59
Agrarian labor force (share over total population)	0.40	0.41	0.12	0.25	0.54
Scattered dwellings	0.28	0.22	0.20	0.05	0.60
1921					
Literacy	0.68	0.66	0.18	0.45	0.92
Literacy (male)	0.71	0.70	0.17	0.50	0.94
Literacy (female)	0.65	0.63	0.20	0.42	0.92
Farm workers (share over total labor force in agriculture)	0.43	0.42	0.18	0.20	0.67
Population (log)	11.85	11.80	0.68	11.00	12.72
Urbanization rate	0.36	0.30	0.27	0.00	0.79

**Table 14** (continued)

	Mean	Median	Std. dev.	10th perc	90th perc
Latitude	47.14	47.77	2.91	42.23	50.49
Landlocked	0.56	1.00	0.50	0.00	1.00
Sharecroppers (share over total labor force in agriculture)	0.16	0.08	0.19	0.01	0.52
Agrarian labor force (share over total population)	0.37	0.37	0.12	0.21	0.53
Scattered dwellings	0.27	0.22	0.20	0.05	0.59
<i>N</i>	205				

Descriptive statistics at the district level for each year. Data for child labor and malaria are only available for 1881

**Table 15** Summary of the variables: provincial level

	Mean	Median	Std. dev.	10th perc	90th perc
1871					
Literacy	0.25	0.22	0.13	0.11	0.49
Literacy (male)	0.31	0.27	0.14	0.16	0.53
Literacy (female)	0.19	0.18	0.13	0.06	0.45
Farm workers (share over total labor force in agriculture)	0.56	0.59	0.22	0.21	0.83
Population (log)	12.74	12.74	0.51	12.21	13.46
Urbanization rate	0.30	0.22	0.21	0.08	0.62
Latitude	42.69	43.43	2.62	38.06	45.33
Landlocked	0.41	0.00	0.49	0.00	1.00
Shareroppers (share over total labor force in agriculture)	0.19	0.09	0.22	0.01	0.64
Agrarian labor force (share over total population)	0.32	0.33	0.10	0.19	0.43
Malaria 1880	0.30	0.24	0.24	0.02	0.69
Child-teacher ratio	0.21	0.17	0.15	0.07	0.39
Municipal exp. (p.c.)	1.04	0.99	0.40	0.54	1.60
Domestic market potential	702.06	747.00	131.05	509.00	908.00
Industrialization index	0.93	0.87	0.29	0.60	1.37
1881					
Literacy	0.29	0.27	0.14	0.13	0.52
Literacy (male)	0.35	0.31	0.14	0.19	0.57
Literacy (female)	0.22	0.20	0.15	0.07	0.49
Farm workers (share over total labor force in agriculture)	0.65	0.68	0.17	0.39	0.85
Population (log)	12.80	12.78	0.52	12.24	13.50
Urbanization rate	0.32	0.22	0.21	0.10	0.68
Latitude	42.69	43.43	2.62	38.06	45.33
Landlocked	0.41	0.00	0.49	0.00	1.00
Sharecroppers (share over total labor force in agriculture)	0.15	0.06	0.19	0.01	0.52
Agrarian labor force (share over total population)	0.37	0.37	0.10	0.25	0.49
Malaria 1880	0.30	0.24	0.24	0.02	0.69
Child-teacher ratio	0.15	0.13	0.09	0.06	0.29
School density	0.002	0.002	0.003	0.001	0.003
Enrollment rate	0.35	0.31	0.16	0.15	0.60
Enrollment rate (male)	0.38	0.32	0.17	0.17	0.63
Enrollment rate (female)	0.32	0.29	0.15	0.14	0.56
Municipal exp. (p.c.)	1.07	1.01	0.37	0.65	1.56
Domestic market potential	852.46	889.00	118.38	642.00	1033.00
Industrialization index	0.93	0.79	0.31	0.62	1.43

**Table 15** (continued)

	Mean	Median	Std. dev.	10th perc	90th perc
1901					
Literacy	0.42	0.41	0.17	0.21	0.67
Literacy (male)	0.47	0.45	0.16	0.28	0.69
Literacy (female)	0.37	0.35	0.19	0.14	0.66
Farm workers (share over total labor force in agriculture)	0.43	0.40	0.20	0.19	0.71
Population (log)	12.93	12.95	0.54	12.31	13.64
Urbanization rate	0.36	0.29	0.24	0.12	0.74
Latitude	42.69	43.43	2.62	38.06	45.33
Landlocked	0.41	0.00	0.49	0.00	1.00
Sharecroppers (share over total labor force in agriculture)	0.23	0.16	0.21	0.04	0.63
Agrarian labor force (share over total population)	0.39	0.41	0.11	0.26	0.51
Child-teacher ratio	0.13	0.11	0.07	0.06	0.26
School density	0.003	0.002	0.003	0.001	0.004
Municipal exp. (p.c.)	1.89	1.89	0.59	1.20	2.69
Domestic market potential	1447.59	1414.00	214.34	1182.00	1779.00
Industrialization index	0.90	0.78	0.38	0.55	1.54
Malaria mortality rate	4.74	1.60	9.20	0.20	12.60
Agricultural productivity	87.40	78.50	28.42	59.90	145.50
1911					
Literacy	0.48	0.46	0.17	0.27	0.72
Literacy (male)	0.52	0.52	0.15	0.33	0.74
Literacy (female)	0.43	0.42	0.19	0.20	0.70
Farm workers (share over total labor force in agriculture)	0.52	0.48	0.20	0.25	0.79
Population (log)	12.97	13.01	0.55	12.39	13.70
Urbanization rate	0.39	0.31	0.24	0.14	0.77
Latitude	42.69	43.43	2.62	38.06	45.33
Landlocked	0.41	0.00	0.49	0.00	1.00
Sharecroppers (share over total labor force in agriculture)	0.21	0.11	0.21	0.03	0.62
Agrarian labor force (share over total population)	0.36	0.37	0.11	0.23	0.47
Child-teacher ratio	0.12	0.10	0.07	0.06	0.25
School density	0.003	0.002	0.003	0.001	0.004
Municipal exp. (p.c.)	3.33	3.02	1.60	2.10	4.69
Domestic market potential	2506.01	2550.00	373.96	2063.00	3094.00
industrialization index	0.88	0.75	0.40	0.51	1.63
1921					
Literacy	0.70	0.73	0.17	0.48	0.92

**Table 15** (continued)

	Mean	Median	Std. dev.	10th perc	90th perc
Literacy (male)	0.74	0.76	0.16	0.52	0.93
Literacy (female)	0.67	0.67	0.18	0.42	0.91
Farm workers (share over total labor force in agriculture)	0.42	0.39	0.16	0.20	0.66
Population (log)	13.04	13.04	0.56	12.42	13.77
Urbanization rate	0.41	0.34	0.24	0.16	0.78
Latitude	42.69	43.43	2.62	38.06	45.33
Landlocked	0.41	0.00	0.49	0.00	1.00
Sharecroppers (share over total labor force in agriculture)	0.19	0.09	0.20	0.02	0.57
Agrarian labor force (share over total population)	0.36	0.37	0.11	0.22	0.49
Child-teacher ratio	0.10	0.09	0.05	0.05	0.15
<i>N</i>	69				

Descriptive statistics at the provincial level for each year. Data for enrollment rates and malaria are only available for 1881, the mortality rate due to malaria only for 1901, and agricultural productivity at the regional level only for 1891 (here shown in 1901)

**Table 16** Description of the variables

Variables	Description	Source
Literacy rate	Share of people > 6 able to read and write	Population Census
Farm workers (share over the total labor force in agriculture)	Sum of daily workers ( <i>braccianti</i> ) and annual workers ( <i>salariati</i> ) over total labor force engaged in agriculture. For 1871 tenancy contracts overlap each other and are not perfectly comparable with Censuses of the subsequent years.	
Population (log)	Fraction of population settled in cities > 10000 inhabitants.	
Urbanization rate	Fraction of sharecroppers over total labor force engaged in agriculture	
Sharecroppers (share over the total labor force in agriculture)	Share of labor force engaged in agriculture over total population	
Agrarian labor force (share over total population)	Share of people aged between 6 and 15 engaged in agriculture	
Child labor (agrarian)	over the total number of people aged between 6 and 15	<a href="https://www.databasecomuni.it/">https://www.databasecomuni.it/</a>
Latitude		Author's elaboration.
Landlocked		Digitalization of the map "Carta della Malaria dell'Italia" by Senator Luigi Torelli.
Malaria 1880	Share of total territory covered by malaria.	Population Census
Scattered dwellings	Share of population scattered in the countryside and not settled within the town over total population	
Child-teacher ratio	Ratio between school-age children (males and females) and primary school teachers. (divided by 1000)	Data are kindly provided by the authors and come from Bozzano, Monica and Cappelli, Gabriele, "The legacy of history or the outcome of reforms? Primary education and literacy in Liberal Italy (1871 - 1911)," Working paper (2019) and Cappelli, Gabriele and Quiroga Valle,

**Table 16** (continued)

Variables	Description	Source
School density	Sum of public, private and mixed schools divided by the territory (surface in square km) of each province.	Gloria. "Schooling and literacy in the provinces of Italy and Spain, 1860 - 1920: new harmonised data," Mimeo (2019).
Enrollment rate	Number of children enrolled in public schools divided by the total number of school-age children.	
Municipal expenditures (per child)	Sum of municipal expenditures in education divided by the number of children.	
Mortality rate due to malaria	Number of deaths caused by malaria fever over total deaths.	Ministero dell'Agricoltura, Industria e Commercio, "Statistica delle cause di morte," 1900. Roma, 1902.
Agricultural productivity	Gross saleable production divided by the number of agrarian employees (regional level).	Federico, Giovanni et al., "Ma l'agricoltura meridionale era davvero arretrata?",
Industrialization index	Ratio between the share of added value, except construction sector	Rivista di politica economica 97, 3/4 (2007), pp. 317. Ciccarelli, C and S Fenoaltea (2010). "Attraverso la lente d'ingrandimento: aspetti provinciali della crescita industriale nell'Italia post-unitaria." In: Quaderni di storia economica 4.
Domestic market potential	Market potential of region A: sum of the GDP of all the adjacent regions each weighted by their distance from region A	Missiati, Anna (2016). "Where do we go from here? Market access and regional development in Italy (1871-1911)." In: European Review of Economic History 20.2, pp. 215-241.

**Table 17** First-stage regression for each year, district level

Outcome: farm workers	1881	1881	1881	1881	1881	1881	1881
Malaria 1880	0.288*** (0.028)	0.276*** (0.029)	0.276*** (0.029)	0.274*** (0.029)	0.159*** (0.020)	0.157*** (0.021)	0.131*** (0.022)
Agrarian labor force		-0.182** (0.076)	-0.178** (0.078)	-0.211** (0.100)	-0.153** (0.069)	-0.144** (0.071)	-0.091 (0.067)
Population (log)			0.004 (0.015)	0.007 (0.015)	-0.009 (0.008)	-0.009 (0.008)	0.005 (0.008)
Urbanization rate				-0.029 (0.056)	0.088*** (0.029)	0.087*** (0.029)	0.044 (0.031)
Sharecroppers					-0.703*** (0.026)	-0.701*** (0.026)	-0.705*** (0.029)
Landlocked						-0.006 (0.013)	0.017 (0.012)
Latitude							-0.011** (0.004)
Departmental dummies	No	No	No	No	No	No	Yes
Constant	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	205	205	205	205	205	205	205
First-stage F-stat	103.61	89.41	89.37	87.79	62.56	57.89	36.32
Outcome: farm workers	1901	1901	1901	1901	1901	1901	1901
Malaria 1880	0.441*** (0.035)	0.422*** (0.034)	0.427*** (0.034)	0.430*** (0.032)	0.327*** (0.030)	0.317*** (0.031)	0.278*** (0.032)
Agrarian labor force		-0.563*** (0.092)	-0.634*** (0.087)	-0.445*** (0.098)	-0.362*** (0.089)	-0.336*** (0.089)	-0.345*** (0.075)
Population (log)			-0.044*** (0.016)	-0.056*** (0.016)	-0.049*** (0.013)	-0.048*** (0.013)	-0.026** (0.012)
Urbanization rate				0.163*** (0.054)	0.232*** (0.037)	0.222*** (0.039)	0.129*** (0.043)
Sharecroppers					-0.479*** (0.037)	-0.473*** (0.037)	-0.431*** (0.042)
Landlocked						-0.030* (0.018)	0.010 (0.018)
Latitude							-0.021*** (0.007)
Departmental dummies	No	No	No	No	No	No	Yes



**Table 17** (continued)

Outcome: farm workers	1881	1881	1881	1881	1881	1881	1881
Constant	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<i>N</i>	205	205	205	205	205	205	205
First-stage F-stat	158.46	156.04	160.51	176.24	117.21	105.25	76.86
Outcome: farm workers	1921	1921	1921	1921	1921	1921	1921
Malaria 1880	0.367*** (0.033)	0.369*** (0.032)	0.370*** (0.032)	0.366*** (0.033)	0.294*** (0.033)	0.283*** (0.034)	0.263*** (0.037)
Agrarian labor force		-0.299*** (0.080)	-0.310*** (0.087)	-0.252*** (0.086)	-0.160* (0.084)	-0.144* (0.084)	-0.249*** (0.082)
Population (log)			-0.005 (0.013)	-0.012 (0.014)	-0.013 (0.012)	-0.012 (0.012)	-0.005 (0.012)
Urbanization rate				0.069 (0.042)	0.151*** (0.039)	0.136*** (0.039)	0.076* (0.041)
Sharecroppers					-0.374*** (0.048)	-0.367*** (0.048)	-0.335*** (0.058)
Landlocked						-0.029 (0.018)	0.001 (0.018)
Latitude							-0.004 (0.008)
Departmental dummies	No	No	No	No	No	No	Yes
Constant	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<i>N</i>	205	205	205	205	205	205	205
First-stage F-stat	125.61	132.06	129.83	125.47	80.52	69.59	50.53

OLS estimates. Standard errors in parentheses are clustered at the district level for separate cross-sections. \*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

**Table 18** First-stage regression for each year, provincial level

Outcome: farm workers	1871	1871	1871	1871	1871	1871	1871
Malaria 1880	0.522*** (0.081)	0.447*** (0.082)	0.443*** (0.081)	0.399*** (0.082)	0.251*** (0.057)	0.234*** (0.059)	0.210*** (0.060)
Agrarian labor force		-0.681*** (0.210)	-0.641*** (0.197)	-1.010*** (0.216)	-0.406* (0.225)	-0.389* (0.218)	-0.236 (0.194)
Population (log)			0.056 (0.041)	0.048 (0.041)	0.006 (0.030)	0.012 (0.030)	0.002 (0.029)
Urbanization rate				-0.277** (0.119)	0.043 (0.095)	0.012 (0.101)	0.034 (0.095)
Sharecroppers					-0.678*** (0.067)	-0.670*** (0.067)	-0.588*** (0.064)
Landlocked						-0.037 (0.026)	0.009 (0.032)
Latitude							-0.015 (0.010)
Departmental dummies	No	No	No	No	No	No	Yes
Constant	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<i>N</i>	69	69	69	69	69	69	69
First-stage F-stat	41.99	29.98	30.01	23.85	19.63	15.81	12.18
Outcome: farm workers	1881	1881	1881	1881	1881	1881	1881
Malaria 1880	0.364*** (0.051)	0.335*** (0.053)	0.330*** (0.054)	0.310*** (0.055)	0.178*** (0.035)	0.177*** (0.035)	0.156*** (0.039)
Agrarian labor force		-0.348*** (0.120)	-0.293** (0.122)	-0.526** (0.202)	-0.263* (0.138)	-0.262* (0.136)	-0.196 (0.124)
Population (log)			0.059** (0.027)	0.059** (0.027)	0.009 (0.017)	0.009 (0.017)	0.015 (0.016)
Urbanization rate				-0.165 (0.107)	0.042 (0.056)	0.041 (0.064)	0.022 (0.068)
Sharecroppers					-0.663*** (0.040)	-0.662*** (0.041)	-0.657*** (0.042)
Landlocked						-0.001 (0.017)	0.022 (0.017)
Latitude							-0.012* (0.007)
Departmental dummies	No	No	No	No	No	No	Yes
Constant	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<i>N</i>	69	69	69	69	69	69	69
First-stage F-stat	50.04	40.22	36.54	32.09	26.11	24.9	15.92
Outcome: farm workers	1891	1891	1891	1891	1891	1891	1891
Malaria 1880	0.429***	0.390***	0.389***	0.385***	0.264***	0.261***	0.229***

**Table 18** (continued)

Outcome: farm workers	1871	1871	1871	1871	1871	1871	1871
	(0.056)	(0.056)	(0.057)	(0.058)	(0.044)	(0.045)	(0.046)
Agrarian labor force		−0.497***	−0.453***	−0.497***	−0.270	−0.269	−0.254**
		(0.133)	(0.124)	(0.175)	(0.162)	(0.162)	(0.126)
Population (log)			0.031	0.031	−0.006	−0.004	−0.004
			(0.029)	(0.030)	(0.025)	(0.026)	(0.020)
Urbanization rate				−0.030	0.140***	0.134**	0.083
				(0.096)	(0.051)	(0.061)	(0.070)
Sharecroppers					−0.566***	−0.564***	−0.496***
					(0.047)	(0.048)	(0.050)
Landlocked						−0.007	0.025
						(0.022)	(0.022)
Latitude							−0.021**
							(0.009)
Departmental dummies	No	No	No	No	No	No	Yes
Constant	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<i>N</i>	69	69	69	69	69	69	69
First-stage F-stat	58.51	47.9	46.42	43.39	36.3	34.2	24.39
Malaria 1880	0.493***	0.460***	0.459***	0.471***	0.350***	0.345***	0.304***
	(0.069)	(0.068)	(0.069)	(0.068)	(0.057)	(0.058)	(0.059)
Agrarian labor force		−0.471***	−0.450**	−0.267	−0.136	−0.135	−0.183
		(0.174)	(0.189)	(0.228)	(0.201)	(0.203)	(0.160)
Population (log)			0.013	0.015	−0.015	−0.014	−0.018
			(0.040)	(0.039)	(0.035)	(0.036)	(0.028)
Urbanization rate				0.146	0.275***	0.267***	0.180**
				(0.100)	(0.059)	(0.072)	(0.085)
Sharecroppers					−0.510***	−0.509***	−0.385***
					(0.056)	(0.057)	(0.068)
Landlocked						−0.008	0.031
						(0.030)	(0.030)
Latitude							−0.031**
							(0.012)
Departmental dummies	No	No	No	No	No	No	Yes
Constant	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<i>N</i>	69	69	69	69	69	69	69
First-stage F-stat	51.04	45.08	44.14	47.2	36.96	35.23	26.49
Outcome: farm workers	1911	1911	1911	1911	1911	1911	1911
Malaria 1880	0.484***	0.479***	0.471***	0.472***	0.319***	0.301***	0.250***
	(0.067)	(0.068)	(0.067)	(0.067)	(0.050)	(0.051)	(0.053)
Agrarian labor force		−0.193	−0.094	−0.011	0.013	0.014	−0.044
		(0.164)	(0.167)	(0.244)	(0.181)	(0.175)	(0.153)

**Table 18** (continued)

Outcome: farm workers	1871	1871	1871	1871	1871	1871	1871
Population (log)			0.052 (0.036)	0.053 (0.036)	0.001 (0.025)	0.008 (0.025)	-0.002 (0.020)
Urbanization rate				0.056 (0.116)	0.224*** (0.059)	0.188*** (0.068)	0.117 (0.077)
Sharecroppers					-0.637*** (0.050)	-0.627*** (0.050)	-0.517*** (0.057)
Landlocked						-0.036 (0.028)	0.015 (0.028)
Latitude							-0.025** (0.010)
Departmental dummies	No	No	No	No	No	No	Yes
Constant	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<i>N</i>	69	69	69	69	69	69	69
First-stage F-stat	51.63	48.92	48.79	49.16	41.32	34.21	22.45
Outcome: farm workers	1921	1921	1921	1921	1921	1921	1921
Malaria 1880	0.419*** (0.062)	0.420*** (0.063)	0.407*** (0.064)	0.406*** (0.065)	0.307*** (0.059)	0.295*** (0.061)	0.301*** (0.060)
Agrarian labor force		-0.242 (0.162)	-0.125 (0.180)	-0.098 (0.201)	0.058 (0.188)	0.049 (0.188)	-0.053 (0.173)
Population (log)			0.056 (0.038)	0.056 (0.038)	0.019 (0.032)	0.024 (0.032)	0.020 (0.027)
Urbanization rate				0.021 (0.077)	0.183*** (0.060)	0.154** (0.069)	0.094 (0.084)
Sharecroppers					-0.429*** (0.058)	-0.419*** (0.060)	-0.253*** (0.077)
Landlocked						-0.026 (0.028)	-0.022 (0.031)
Latitude							-0.009 (0.013)
Departmental dummies	No	No	No	No	No	No	Yes
Constant	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<i>N</i>	69	69	69	69	69	69	69
First-stage F-stat	45.16	44.68	39.77	38.87	27.17	23.22	25.38

OLS estimates. Standard errors in parentheses are clustered at the district level for separate cross-sections. \*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

**Table 19** Literacy 15–19 and farm workers, provincial level

	1871 (1)	1881 (2)	1891 (3)	1901 (4)	1911 (5)
Panel A:					
	<i>OLS</i>				
Farm workers	−0.266*** (0.092)	−0.561*** (0.123)	−0.477*** (0.086)	−0.327*** (0.058)	−0.282*** (0.064)
Conley SE (200 km)	[0.081]***	[0.114]***	[0.067]***	[0.042]***	[0.052]***
<i>p-value farm workers</i>					
Wild bootstrap	{0.101}**	{0.004}***	{0.001}***	{0.000}***	{0.007}***
Panel B:					
	<i>IV</i>				
Farm workers	−0.503** (0.202)	−0.802*** (0.238)	−0.569*** (0.148)	−0.426*** (0.110)	−0.489*** (0.165)
Conley SE (200 km)	[0.162]***	[0.203]***	[0.126]***	[0.097]***	[0.160]***
<i>p-value farm workers</i>					
Wild bootstrap	{0.021}***	{0.001}***	{0.001}***	{0.003}***	{0.020}***
<i>N</i>	69	69	69	69	69

All regressions include baseline controls and macroregional dummies. Heteroskedastic-robust standard errors are reported in parentheses for separate cross-sections. *P*-values of wild cluster bootstrap SE of farm workers clustered at the regional level are reported in curly brackets (stars relate to SE clustered at the regional level); Conley SE (200 km) in square brackets. \*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

**Table 20** Literacy 15–19 and farm workers, panel estimates, provincial level

Dep. var.:	Literacy rate 15–19	Literacy rate 15–19	Literacy rate 15–19
Model specification:	Main	Controls	Malaria
		1881	predicted
	(1)	(2)	(3)
Farm workers x 1891 dummy	0.010 (0.042)	0.013 (0.040)	–0.015 (0.103)
Conley SE (200 km)	[0.072]	[0.071]	[0.103]
<i>p-value farm workers</i>			
Wild bootstrap	{0.853}	{0.855}	{0.869}
Farm workers x 1901 dummy	0.038 (0.084)	0.022 (0.083)	–0.050 (0.197)
Conley SE (200 km)	[0.073]	[0.076]	[0.197]
<i>p-value farm workers</i>			
Wild bootstrap	{0.761}	{0.856}	{0.756}
Farm workers x 1911 dummy	0.169 (0.119)	0.268** (0.121)	0.399 (0.304)
Conley SE (200 km)	[0.090]*	[0.094]***	[0.304]
<i>p-value farm workers</i>			
Wild bootstrap	{0.357}	{0.228}*	{0.233}
<i>N</i>	276	276	276
<i>p-value: <math>\beta_{1891,1901} = \beta_{1921}</math></i>	0.045	0.000	t-stat: [-1.72]*

Fully flexible province-level panel estimates. The share of farm workers is interacted with time dummies for the years 1881, 1891, 1901, and 1921. The omitted reference year is 1881. Baseline controls are interacted with time dummies. All columns use the initial level of farm workers; columns 2 and 3 fix all control variables in 1881; column 3 uses predicted values of farm workers using malaria as instrument. All regressions include time shocks and fixed effects. Heteroskedastic-robust standard errors are reported in parentheses for separate cross-sections. *P*-values of wild cluster bootstrap SE of farm workers clustered at the regional level are reported in curly brackets (stars relate to SE clustered at the regional level); Conley SE (200 km) in square brackets. Robust standard errors are bootstrapped to account for uncertainty stemming from the use of a predicted variable in the second-stage estimates in column 3. \*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

**Acknowledgements** I am grateful to Prof. Mauro Rota for his comments and suggestions that significantly improved the quality of the paper. I would also like to thank the Editor, Oded Galor, three anonymous referees for several valuable comments that have remarkably improved the quality of research, and Madeline Zavodny for proposing additional suggestions that improved the article's clarity. I have also benefited from comments and feedback from Sascha O. Becker, Alessandro Belmonte, David Chilosi, Daniel Gallardo-Albarran, Pablo Martinelli, Gabriele Cappelli, Giovanni Federico, Emanuele Felice, Jacob L. Weisdorf, Marco Ventura, Francesco Cinnirella, Alessandro Nuvolari, Carlo Ciccarelli, and Francesco Bloise. Finally, I am also indebted to the participants at the EHES Conference in Paris, the seminar participants at Groningen University, and the internal seminars at Sapienza University in Rome.

**Funding** Financial support is provided by the “Innovation and Global Challenges” PRIN (Projects of National Interests), Prot. 20177J2LS9, funded by Ministero dell’Istruzione, dell’Università e della Ricerca.

## Declarations

**Competing interests** The author declares no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Acemoglu D, Robinson JA (2000) Why did the west extend the franchise? Democracy, inequality, and growth in historical perspective. *Q J Econ* 115(4):1167–1199
- Acemoglu D, Johnson S, Robinson JA (2001) The colonial origins of comparative development: an empirical investigation. *Am Econ Rev* 91(5):1369–1401
- Acemoglu D, Johnson S, Robinson JA (2002) Reversal of fortune: geography and institutions in the making of the modern world income distribution. *Q J Econ* 117(4):1231–1294
- Acemoglu D, Johnson S, Robinson JA (2005) Institutions as a fundamental cause of long-run growth. *Handb Econ Growth* 1:385–472
- Adamopoulos T (2008) Land inequality and the transition to modern growth. *Rev Econ Dyn* 11(2):257–282
- Aghion P, Jaravel X, Persson T, Rouzet D (2019) Education and military rivalry. *J Eur Econ Assoc* 17(2):376–412
- A'Hearn B (1998) Institutions, externalities, and economic growth in southern Italy: evidence from the cotton textile industry, 1861–1914. *Econ Hist Rev* 51(4):734–762
- Al Serouri AW, Grantham-McGregor SM, Greenwood B, Costello A (2000) Impact of asymptomatic malaria parasitaemia on cognitive function and school achievement of schoolchildren in the Yemen Republic. *Parasitology* 121(4):337–345
- Altonji JG, Elder TE, Taber CR (2005) Selection on observed and unobserved variables: assessing the effectiveness of Catholic schools. *J Polit Econ* 113(1):151–184
- Amorosa LF Jr, Corbellini G, Coluzzi M (2005) Lessons learned from malaria: Italy's past and sub-Saharan's future. *Health Place* 11(1):67–73
- Anatra B (1977) Abolizione del feudalesimo e rivoluzione agricola. *Stud Stor* 18(3):113–126
- Andersson J, Berger T (2019) Elites and the expansion of education in nineteenth-century Sweden. *Econ Hist Rev* 72(3):897–924
- Asenso-Okyere K, Asante FA, Tarekegn J, Andam KS (2011) A review of the economic impact of malaria in agricultural development. *Agric Econ* 42(3):293–304
- Beltrán Tapia F, Martínez-Galarra J (2018) Inequality and education in pre-industrial economies: evidence from Spain. *Explor Econ Hist* 69:81–101
- Binswanger H, Rosenzweig M (1993) Wealth, weather risk and the composition and profitability of agricultural investments. *Econ J* 103(416):56–78
- Bleakley H (2010) Malaria eradication in the Americas: a retrospective analysis of childhood exposure. *Am Econ J Appl Econ* 2(2):1–45
- Brown PJ (1983) Demographic and socioeconomic effects of disease control: the case of malaria eradication in Sardinia. *Med Anthropol* 7(2):63–87
- Cameron AC, Gelbach JB, Miller DL (2008) Bootstrap-based improvements for inference with clustered errors. *Rev Econ Stat* 90(3):414–427
- Cappelli G (2015) Escaping from a human capital trap? Italy's regions and the move to centralized primary schooling, 1861–1936. *Eur Rev Econ Hist* 20(1):46–65

- Cappelli G (2016) One size that didn't fit all? Electoral franchise, fiscal capacity and the rise of mass schooling across Italy's provinces, 1870–1911. *Cliometrica* 10(3):311–343
- Cappelli G, Quiroga Valle G (2020) Literacy and schooling in Italy and Spain (1860–1921): a comparative analysis. *Riv Storia Econ* 26(1):87–123
- Cappelli G, Vasta M (2020) Can school centralization foster human capital accumulation? A quasi-experiment from early twentieth-century Italy. *Econ Hist Rev* 73(1):159–184
- Chaudhary L (2009) Determinants of primary schooling in British India. *J Econ Hist* 69(1):269–302
- Chaves LF (2013) The dynamics of latifundia formation. *PLoS ONE* 8(12):e82863
- Cinnirella F, Hornung E (2016) Landownership concentration and the expansion of education. *J Dev Econ* 121:135–152
- Clarke D, Matta B (2018) Practical considerations for questionable IVs. *Stat J* 18(3):663–691
- Conley TG (1999) GMM estimation with cross sectional dependence. *J Econ* 92(1):1–45
- Conley TG, Hansen CB, Rossi PE (2012) Plausibly exogenous. *Rev Econ Stat* 94(1):260–272
- Corbellini G, Merzagora L (1998) La malaria: una lezione italiana. *KOS (Milano)* 148–49:46–51
- Cutler D, Fung W, Kremer M, Singhal M, Vogl T (2010) Early-life malaria exposure and adult outcomes: evidence from malaria eradication in India. *Am Econ J Appl Econ* 2(2):72–94
- Deininger K, Squire L (1998) New ways of looking at old issues: inequality and growth. *J Dev Econ* 57(2):259–287
- Doepke M, Zilibotti F (2005) The macroeconomics of child labor regulation. *Am Econ Rev* 95(5):1492–1524
- Easterly W (2007) Inequality does cause underdevelopment: insights from a new instrument. *J Dev Econ* 84(2):755–776
- Eastwood R, Lipton M, Newell A (2010) Farm size. *Handb Agricult Econ* 4:3323–3397
- Federico G et al (2007) Ma l'agricoltura meridionale era davvero arretrata? *Riv Polit Econ* 97(3/4):317
- Federico G, Nuvolari A, Vasta M (2019) The origins of the Italian regional divide: evidence from real wages, 1861–1913. *J Econ Hist* 79(1):63–98
- Felice E (2018) The socio-institutional divide: explaining Italy's long-term regional differences. *J Interdiscip Hist* 49(01):43–70
- Galassi FL, Cohen JS (1994) The economics of tenancy in early twentieth-century southern Italy 1. *Econ Hist Rev* 47(3):585–600
- Gallego FA (2010) Historical origins of schooling: the role of democracy and political decentralization. *Rev Econ Stat* 92(2):228–243
- Galar O, Moav O (2006) Das human-kapital: a theory of the demise of the class structure. *Rev Econ Stud* 73(1):85–117
- Galar O, Zeira J (1993) Income distribution and macroeconomics. *Rev Econ Stud* 60(1):35–52
- Galar O, Moav O, Vollrath D (2009) Inequality in landownership, the emergence of human-capital promoting institutions, and the great divergence. *Rev Econ Stud* 76(1):143–179
- Gennaioli N, La Porta R, Lopez-de Silanes F, Shleifer A (2012) Human capital and regional development. *Q J Econ* 128(1):105–164
- Glaeser EL, La Porta R, Lopez-de Silanes F, Shleifer A (2004) Do institutions cause growth? *J Econ Growth* 9(3):271–303
- Go S, Lindert P (2010) The uneven rise of American public schools to 1850. *J Econ Hist* 70(1):1–26
- Holding PA, Snow RW (2001) Impact of Plasmodium falciparum malaria on performance and learning: review of the evidence. *Am J Trop Med Hyg* 64(1-suppl):68–75
- Kere N, Keni J, Kere J, Bobogare A, Webber R, Southgate B (1993) The economic impact of Plasmodium falciparum malaria on education investment: a Pacific Island case study. *Southeast Asian J Trop Med Public Health* 24(4):659–663
- Kiviet JF (2020) Testing the impossible: identifying exclusion restrictions. *J Econ* 218(2):294–316
- Lupo S (1990) I proprietari terrieri nel mezzogiorno. *Stor Agric Ital Contempor: Uomini Classi* 2:105–50
- Martinelli P (2014) Latifundia revisited: market power, land inequality and agricultural efficiency. Evidence from interwar Italian agriculture. *Explor Econ Hist* 54:79–106
- McCormick MC, Brooks-Gunn J, Workman-Daniels K, Turner J, Peckham GJ (1992) The health and developmental status of very low-birth-weight children at school age. *Jama* 267(16):2204–2208
- Missiaia A (2016) Where do we go from here? Market access and regional development in Italy (1871–1911). *Eur Rev Econ Hist* 20(2):215–241
- Vollrath D (2013) Inequality and school funding in the rural United States, 1890. *Explor Econ Hist* 50(2):267–284
- Allen RC (2011) Global economic history: a very short introduction, vol 282. Oxford University Press



- Beauchamp C (1988). Fièvres d'hier, paludisme d'aujourd'hui, vie et mort d'une maladie. In: *Annales. Histoire, Sciences Sociales*, vol 43. Cambridge University Press, pp. 249–275
- Beltrán Tapia F, de Miguel Salanova S (2019) Class, education and social mobility: Madrid, 1880–1905. Technical report, European Historical Economics Society (EHES)
- Bevilacqua P (2005) Breve storia dell'Italia meridionale: dall'Ottocento a oggi. Donzelli editore
- Bocchini F, Piccialuti M (2003) Fonti per la storia della malaria in Italia. Ministero per i beni e le attività culturali, Direzione generale per gli archivi
- Bozzano M, Cappelli G et al (2019) The legacy of history or the outcome of reforms?: Primary education and literacy in Liberal Italy (1871–1911). Università di Siena
- Buonanno P, Esposito E, Gulino G (2020) The epidemiological roots of inequality: historical evidence from malaria in Italy. Working Papers 1220, Barcelona Graduate School of Economics
- Celli A, Fraentzel A (1933) The history of malaria in the Roman Campagna from ancient times. London, pages pp.viii + 226 pp. ref.39 pp
- Ciccarelli C, Fenoaltea S (2010) Attraverso la lente d'ingrandimento: aspetti provinciali della crescita industriale nell'Italia post-unitaria. *Quad Di Storia Econ* , 4
- Cives G (1990) La scuola italiana dall'Unità ai nostri giorni, vol 446. La Nuova Italia
- Colella F, Lalive R, Sakalli SO, Thoenig M (2019) Inference with arbitrary clustering. IZADiscussion Paper No. 12584
- Di Martino P and Vasta M (2017) Ricchi per caso: la parabola dello sviluppo economico italiano. Il mulino
- Engerman SL, Sokoloff KL (1997) Factor endowments, institutions, and differential paths of growth among new world economies. *How Latin America Fell Behind*, pp 260–304
- Felice E (2013) Perché il Sud è rimasto indietro. il Mulino Bologna
- Galor O (2011) Inequality, human capital formation, and the process of development. In: *Handbook of the economics of education*, vol 4. Elsevier, pp 441–493
- Gerschenkron A (1962) Economic backwardness in historical perspective: a book of essays. Technical report, Belknap Press of Harvard University Press, Cambridge
- Go S, Park K-J (2012) The elite-biased growth of elementary schooling in colonial Korea
- Goni M (2015) Landed elites and education provision in England and Wales. evidence from school boards, 1870–99. Manuscript, University of Vienna
- Imbens GW, Rubin DB (2015) Causal inference in statistics, social, and biomedical sciences. Cambridge University Press
- King R, Strachan A (1978) Sicilian agro-towns (sizilianische agro-städte). *Erdkunde*, pp 110–123
- Kripfganz S, Kiviet JF et al (2020) *kinkyreg: instrument-free inference for linear regression models with endogenous regressors*. London Stata Conference 2020 15, Stata Users Group
- Lindert PH (2004) Growing public: volume 1, the story: social spending and economic growth since the eighteenth century, vol 1. Cambridge University Press, Cambridge
- Lucas AM (2010) Malaria eradication and educational attainment: evidence from Paraguay and Sri Lanka. *Am Econ J Appl Econ* 2(2):46–71
- Malaney P, Spielman A, Sachs J (2004) The malaria gap. *Am J Trop Med Hyg* 71(2\_suppl):141–146
- Malanima P (2005) Urbanisation and the Italian economy during the last millennium. *Eur Rev Econ Hist* 9(1):97–122
- Mariscal E, Sokoloff K (2000) Schooling, suffrage and the persistence of inequality in the Americas. Political institutions and economic growth in Latin America: essays in policy, history and political economy. Hoover Institution Press, Stanford, pp 159–218
- Ministero di Agricoltura IECM (1900) Statistica delle Cause di Morte. Tip. nazionale di G. Bertero, Roma
- Missiroli A (1949) L'influenza della scomparsa della malaria sulle condizioni sociali ed economiche dell'Italia del sud. *Riv Parassitol* 10:271–272
- Nájera JA (2001) Malaria control: achievements, problems and strategies. *Parassitologia* 43(1–2):1–89
- Oster E (2019) Unobservable selection and coefficient stability: theory and evidence. *J Bus Econ Stat* 37(2):187–204
- Oto-Peralías D, Romero-Avila D (2016) The economic consequences of the Spanish Reconquest: the long-term effects of medieval conquest and colonization. *J Econ Growth* 21(4):409–464
- Percoco M (2013) The fight against disease: malaria and economic development in Italian regions. *Econ Geogr* 89(2):105–125

- Persson T, Tabellini G (1994) Is inequality harmful for growth? *The American Economic Review*, pages 600–621
- Petrusewicz M (1989). *Latifondo: economia morale e vita materiale in una periferia dell'Ottocento* Marsilio Venice
- Placanica A (1990) Il mondo agricolo meridionale: usure, caparre, contratti. Bevilacqua, P. (a cura di). *Storia dell'agricoltura* 2:261–324
- Ramcharan R (2010) Inequality and redistribution: evidence from US counties and states, 1890–1930. *Rev Econ Stat* 92(4):729–744
- Reis J (2005) Economic growth, human capital formation and consumption in western Europe before 1800. *Living Standards in the Past*. Oxford University Press, Oxford, pp 195–225
- Roodman D, Nielsen MØ, MacKinnon JG, Webb MD (2019) Fast and wild: bootstrap inference in Stata using boottest. *Stata J* 19(1):4–60
- Sachs J, Malaney P (2002) The economic and social burden of malaria. *Nature* 415(6872):680
- Sallares R, Bouwman A, Anderung C (2004) The spread of malaria to southern Europe in antiquity: new approaches to old problems. *Med Hist* 48(3):311–328
- Sereni E (1971) *Il capitalismo nelle campagne: 1860–1900, con un nuovo saggio introduttivo*, vol 112. Einaudi
- Snowden F (2008) *The conquest of malaria: Italy, 1900–1962*. Yale University Press
- Snowden FM (1999) 'fields of death': malaria in Italy, 1861–1962. *Mod Italy* 4(1):25–57
- Sokoloff KL, Engerman SL (2000) Institutions, factor endowments, and paths of development in the new world. *J Econ Perspect* 14(3):217–232
- van Kippersluis H, Rietveld CA (2018) Beyond plausibly exogenous. *Econ J* 21(3):316–331
- Villani P (1965) La feudalità dalle riforme all'eversione. *Clio*, pages 600–622