ORIGINAL PAPER

Geography, landownership inequality, and literacy: historical evidence from Greek regions

Nikos Benos[1] · Stelios Karagiannis[2] · Sofia Tsitou[3]

Abstract

Our work sheds light on the joint role of human capital and geography during the early stages of the transition from stagnation to growth in early twentieth century Greece. We uncover a robust association between geography and literacy. We also show that geography is correlated with land inequality and thus establish that land distribution is a channel through which geography influences literacy. Finally, the impact of geography on human capital formation weakens with industrialization. Our work contributes to the literature on geography and human capital in the transition from stagnation to growth since Greece was at the early stages of the industrial era during the study period.

Responsible editor: Oded Galor

✉ Nikos Benos
  nbenos@uoi.gr

  Stelios Karagiannis
  stelioskarayanis@hotmail.com

  Sofia Tsitou
  sofitsitou@gmail.com

[1] Department of Economics, University of Ioannina, 45100 Ioannina Campus, Ioannina, Greece

[2] European Training Foundation, Turin, Italy

[3] Hellenic Open University, Patras, Greece

1 Introduction

The role of human capital formation, geography, institutions, and ethnic fractional-ization has been the center of a discussion about the origin of the differential tim-ing of the transition from stagnation to growth. These factors have led to important changes in cross-country income distribution since the onset of the Industrial Rev-olution. In particular, geography has been shown to have persistent effects on past and current economic outcomes. A large literature emphasizes the critical role of variations in climatic and geographic conditions for comparative economic devel-opment (e.g., Galor et al. 2009; Ashraf and Galor 2013; Easterly 2007; Baten and Hippe 2018; Michalopoulos and Papaioannou 2017; Adamopoulos and Restuccia 2022; Gershman et al. 2022). In this framework, inequality in land distribution is a mediator through which geography has affected human capital formation and the transition from an agricultural to an industrial economy, as a precondition for development (e.g., Galor et al. 2009; Gershman et al. 2022). Therefore, land ine-quality can be an obstacle to human capital accumulation and a decelerating factor of industrialization as well as economic growth.

In line with this literature, our research sheds light on the role of geography in human capital formation during the early stages of the transition from stagnation to growth. We study the case of Greece in early 1900s, an ideal setting for the investi-gation of the geography-literacy nexus since large topographic disparities are asso-ciated with substantial literacy differentials. In this context, large properties were dominant in lowland fertile areas, possibly affecting human capital accumulation. In this period, industrialization is already present but important only in certain areas; overall, the dominant sector is still agriculture. In particular, our work uniquely inte-grates four related questions: Has geography affected literacy? Has geography influ-enced land inequality? Has land inequality been a channel through which geography exerts an impact on literacy? Does the relationship between geography and liter-acy vary depending on industrialization? In order to examine empirically the above issues, we have built an entirely novel province-level data set on geography, literacy, landownership, and industrialization. We find that geography is associated with both literacy and land inequality. The dominance of large properties has a sizable detri-mental effect on literacy accounting for a variety of geographic factors. Finally, the impact of geography on human capital accumulation is weaker in more industrial-ized provinces compared to less industrialized ones.

A rich literature has identified geography as a deep-rooted factor behind devel-opment differences both across and within countries. For instance, Diamond (1997) presents a compelling case for the impact of geography and environment on shaping the contemporary world. He posits that societies which gained an early advantage in food production transcended beyond the hunter-gatherer stage and subsequently developed religious beliefs, alongside the emergence of harmful diseases and power-ful weapons. Sachs and Warner (1999) show that booms based on natural resources can stimulate industrialization but can also forestall or even reverse it.

A seminal paper by Galor et al. (2009; henceforth GMV) provides the theo-retical foundation behind the prediction that inequality in landownership had an

adverse effect on human capital accumulation, industrialization and growth. During the transition from the agricultural to the industrial economy, a major conflict arose between agricultural landholders and capitalists. Landholders would benefit less from an increase in human capital of their workers than capitalists: human capital raises productivity of workers more in industry than in agriculture because land and human capital are less complementary than physical and human capital. Return on land declines as wages of workers rise due to higher education than individuals obtain, while educated workers have stronger incentives to migrate to industrial areas than less educated ones. Thus, landownership is an obstacle to human capital formation, thereby slowing down industrialization and economic growth. More recent studies focus on literacy rates and the channels through which landownership and geographic factors influence them.

The value-added of our work is that, first, we study the geography-human capital link in Greece, which was at the early industrialization phase during the early twentieth century, while much literature focuses on mature industrial economies. Specifically, our paper is the first among empirical studies which utilizes Greek historical data at the province-level around 1900. Second, using a regional sample from the early twentieth century, we are able to isolate the potential impact of sweeping educational reforms, which took place in Greece after that period, on literacy rates from the effect of geography through land inequality. Third, we use more disaggregated (province-level) data than related literature, which better capture the spatial variation of the phenomena analyzed (e.g., GMV; Easterly 2007; Baten and Hippe 2018). Fourth, our unit of analysis is much more homogenous in terms of institutions, human capital, culture, and geography than cross-country samples. Here, we should emphasize that the administrative division of Greece analyzed has existed since 1833, well before the appearance of private large land holdings and the onset of industrialization and thus orthogonal to the issues analyzed.[1] A study closely associated with our work is by Baten and Hippe (2018), who find that the distribution of landownership is a mechanism behind the correlation of human capital and geographic factors, using a European multi-country regional level analysis.

A conclusive exploration of the impact of geography on literacy ought to overcome significant empirical hurdles. Our investigation takes into account numerous challenges. First, a large number of confounding factors affecting our dependent variables are included in our specifications, e.g., demographics, ethno-religious composition, and structural economic characteristics. Second, we utilize fixed effects and initial values of explanatory variables to deal with endogeneity. Third, we explore omitted variable bias by applying Oster's strategy. Fourth, we estimate clustered standard errors at the prefecture level to allow for spatial error dependence. Finally, we apply alternative fractional estimators to account for the nature of our dependent variables. We empirically demonstrate that geography is associated with literacy; geography is linked with land inequality; inequality in landownership distribution is a mediator in the geography-literacy nexus having a negative effect on human capital formation; and the impact of geography on literacy weakens with industrialization.

---

[1] Royal Decree 3, 15th April 1833.

The rest of the paper is structured as follows. In Section 2, we present the theoretical framework and empirical literature on geography, literacy, and landownership. In Section 3, we outline the historical context of landownership and education in Greece. In Section 4, we describe the data and empirical specification. In Section 5, we discuss our empirical results. In Section 6, we apply Oster analysis to check for omitted variable bias. Section 7 concludes. We perform additional robustness checks in the Appendix.

2 Literature review

Recent work investigates literacy rates and the channels through which geography influences human capital. A strand of literature emphasizes geographic factors as explanations for long-run development. A classic paper is by Ashraf and Galor (2013) who argue that the distance to the cradle of mankind—east Africa—matters for human genetic diversity. Genetic variety has both advantages and disadvantages. If there are many genetic differences among the population, ideas and potential innovation are high, but this comes at a cost. This is so because there are more conflicts, less trust, and less social capital, which lead to a vicious circle of civil war and violent conflict. On the other hand, homogeneous populations enjoy high social capital and trust, but are characterized by less ideas and low innovation potential. Therefore, there is hump-shaped relationship between genetic diversity, which is directly related to geography, and comparative development. Another influential paper on geographic factors and economic development is by Galor and Ozak (2016), who argue that time preferences and long-term orientation are affected by geography. They use the expansion of suitable crops for cultivation during the Columbian Exchange to establish that pre-industrial agro-climatic characteristics implying higher return to agricultural investment and have triggered selection, adaptation, and learning processes which have generated a persistent positive effect on the prevalence of long-term orientation in the modern period. Here, we argue that geography matters for Greek regional development, partly because it affects land inequality.

As an element of our inquiry, we use the framework of GMV, who demonstrate that land inequality is a channel through which geography affects human capital formation. They show that inequality in land distribution has negatively affected educational institutions (e.g., public schooling and child labor regulations) in US states in 1900–1940. Along the same lines, Gersbach and Siemers (2010) demonstrate theoretically that land redistribution allows the transition of a society from an agriculture-based to a human capital–based developed economy. Deininger and Squire (1998) use cross-country and panel data for 103 countries in 1960–1990 and find a negative relationship between initial land inequality and long-run growth. One possible mechanism behind this relationship is the individual underinvestment in education as a consequence of capital market imperfections. Large landownership is negatively correlated with education, while high levels of education lead to more investment. Alternatively, land inequality can be itself a direct source of inefficiency with negative effects on long-run development (Martinelli 2014). The mechanism

linking land inequality and inefficiency is market power, since a high degree of landownership concentration endows landowners with market power in poorly integrated rural labor markets. Sokoloff and Engerman (2000) show that a higher degree of inequality in landownership distribution among Latin American countries was reflected in lower human capital investment. Erickson (2004) demonstrates that landownership does not impinge on the relationship between geographic endowments and income inequality, but high landownership concentration leads to low education levels. Easterly (2007) uses cross-country data for developing countries and concludes that high inequality in landownership constitutes the most important drawback to welfare, high-quality institutions, and high schooling. Wegenast (2009) suggests that the agricultural production system in Asia and Latin America constitutes a major factor behind educational outcomes. Countries with higher agricultural plantations provide less broadly based educational policies in contrast to countries organized around family farming. He concludes that exports of plantation crops—a proxy for the political strength of the agrarian elite—reduce secondary education attainment and related government investments, but they are associated with higher tertiary education levels. Thus, educational and political inequalities, especially in Latin America, are due to the historical influence of landlords on schooling. Baten and Juif (2014) show the detrimental causal influence of early land inequality on math and science skills using a global sample. Baten and Hippe (2018) examine the relationship between geographic factors and numeracy in more than 300 European regions in nineteenth century. They find that human capital is negatively correlated with landownership inequality, which is also related to geographic factors. Their underlying rationale is that capitalists benefit more from an increase in human capital of their workers than landholders, and thus, landowners do not affiliate education-promoting policies.

Another strand of literature concentrates on single-country analysis. For instance, Tapia and Martinez-Galarraga (2018) establish a negative relationship between the fraction of farm laborers and male literacy rates in mid-nineteenth century Spain, before industrialization. Cinnirella and Hornung (2016) establish a negative effect of landownership concentration on education in nineteenth century Prussia and show that it is causal and weakens over time. Gershman et al. (2022) find that Prussian counties in which peasant emancipation in large landholdings occurred at a higher degree by mid-nineteenth century exhibit higher contemporaneous share of skilled manufacturing workers and a faster subsequent rate of general human capital accumulation.

Other historical studies yield insignificant and sometimes contradictory results. Andersson and Berger (2019) conclude that elites were historically not always a barrier to the diffusion of elementary education. They find that Swedish regions governed by local elites have higher educational expenditure relative to regions where the distribution of political power was more equal. Goñi (2022) concludes that high concentration of landownership has a significant negative effect on education during 1871–1899 in English counties, but this effect is significant only for changes that began after the Industrial Revolution. High landownership concentration reduces the ratio of state to private schools, the number and salaries of teachers, and the facilities per pupil in areas where landowners promote land elites. Overall, some studies

support the significant impact of land ownership, geography, and other factors on human capital formation whereas others are inconclusive.

3 Historical context: landownership and education in Greece

3.1 Landownership in historical perspective

During the Ottoman period in Greece, the legal framework prohibited private ownership, and the land of the Empire belonged to the Sultan. The first major reform was initiated by the decree of 1839 (Gülhane), which guaranteed the right to life and property for all subjects, including non-Muslims. In 1858, the Land Code secured the local administrators (known as beys and aghas) the right to full ownership of estates, if they had been cultivating them for at least 10 years. By the end of nineteenth century, the land tenure system divided the land into two categories: the chiftliks (large estates) and the head villages, i.e., Christian villages under the supervision of the central government (Vergopoulos 1975; McGrew 1985; Petmezas 2003, 2012; Kostis and Petmezas 2006; Kontogiorgi 2006).

Chiftliks, populated by 20 to 30 families of sharecroppers, were usually located in the fertile lowlands close to main roads and were conferred as the estates of a landlord. The sharecroppers were obliged to deliver part of the grain and other plow production, e.g., cotton, to the landlords based on the number of plows and grain provided by them. The composition of production was oriented towards grain, characterized by low but relatively stable prices, and much less towards potentially high price-high volatility crops, e.g., tobacco, grapes etc. Landowners also provided sharecroppers with a small piece of land, which could be used to cultivate vegetables, grapevines, tobacco, and other commercial crops along with housing without the obligation to pay rents. It was illegal to expel the sharecroppers from their land, and the rights of usufruct on the land were bestowed to the farmers' descendants. The landlords had the moral obligation to secure the sharecroppers low interest loans necessary for cultivation expenses.

A new historical phase was initiated after the annexation of new territories to Greece. According to the Berlin Treaty (1878) Ottoman, large estate owners would maintain their property rights on land in Thessaly region. However, most of them sold their properties to wealthy Greeks from abroad (Vergopoulos 1975; Petmezas 2003, 2012). During the first decade of Greek administration of Thessaly, where the majority of land was occupied by chiftliks, the position of the sharecroppers worsened, because government recognized full ownership of large estates by the landlords. As a result, the farmers became agricultural workers with no continuous rights on the land they cultivated, which created tensions between sharecroppers and landlords leading to violent protests (Petmezas 2012).

In light of these, a fund was created to lend farmers and help them buy large properties in 1907–1914 in Thessaly, while very strict terms were set regarding the possibility of selling these pieces of land. Also, the new Liberal government implemented a series of reforms: (a) changed the constitution in 1911 so that mandatory expropriation of private property was possible for the common

benefit; (b) passed laws in 1911–1915 which made the expulsion of sharecroppers illegal; (c) passed a law in 1913 which prohibited any action modifying the land property rights in the New Lands (Macedonia, Epirus, Eastern and Northern Aegean and Crete). Thus, it became impossible for rich Greeks to buy large land holdings of foreigners (Petmezas 2012). The final stage of the agricultural reforms was initiated with the 1917 law for the mandatory expropriation of large landholdings in the entire Greece. A second more comprehensive law was voted in 1919, which simplified the procedure for the expropriation of large landholdings with strict terms regarding their future status.

After the arrival of Greek refugees and the exchange of populations in 1922–1924, the Liberal governments passed a series of laws during 1922–1926, according to which large and middle landholdings were dismantled, distributed to their cultivators, or landless peasants (natives and refugees) (Petmezas 2012). Most expropriations took place between 1923 and 1928, but continued until after the Second World War. Specifically, 2259 large properties existed before 1917, but only 76 of these were alienated before 1923. Overall, around 1730 properties were expropriated until 1938 (Alivizatos 1938). These developments marked the end of large landownership in Greece and established a nation of small land property holders (Map 1).

3.2 The evolution of the Greek education system

The first Governor of Greece (Kapodistrias) in 1828 organized primary, secondary, and professional education in order to reduce illiteracy and provide agriculture as well as light industry with trained labor force. Also, he aimed at providing the public sector with well-trained teachers, priests, public servants, and the armed forces with trained personnel. However, after his murder, most policies were discontinued, and the majority of schools founded stopped functioning.

During the reign of King Otto (1832–1862), the institutionalization of primary, secondary, and tertiary education took place, with the introduction of 7-year compulsory education. Primary education would be funded by local communities, while secondary and tertiary education would be funded by the state. However, local authorities were deprived of their fiscal responsibilities forcing them to close many schools. Overall, the system was very centralized, every education level was meant to prepare students for the next level and not the labor market, the language used was close to ancient Greek, and the education content became very theoretical and away from the focus on professional education instituted by Kapodistrias. No attention was paid to the education and living standards of teachers and the improvement of infrastructure. The operation of private schools was allowed. The first major reform after King Otto's reign took place in 1880, when the interdependent teaching method was replaced by the co-educational method supported by new school books. Also, the first public school to educate teachers was founded in 1878. Moreover, ancient Greek was replaced gradually in primary education by modern language, and a curriculum was compiled for the first time. University was accessible to all secondary education graduates without entrance examinations. These characteristics dominated the Greek educational system until 1928 (Chantzopoulos 1998).

Map 1 Distributed land in Greece in 1936. The map shows the allocation of land into distributed (in red), non-distributed (in yellow) and heath (check green pattern). Source: Hellenic Ministry of Agriculture; available in Alivizatos (1938)

The landslide victory of the Liberal party in 1928 marked the most significant turning point in the history of modern Greek education (Stefanidis 2006). According to its leader Eleftherios Venizelos, education was one of the principal tiers of this reform package. He had stressed its importance by stating: "You should know that I consider our educational reform to be the greatest title of glory of my premiership and my greatest service to the motherland" (Dimaras 2006). The aim of the reforms was to provide education for all and not only for the elites, in order to narrow social divide. The main reforms of the program included (i) 6-year compulsory education, which had been instituted by King Otto, but not implemented; (ii) establishment of more professional and technical schools at the expense of the classical secondary schools whose number was reduced; (iii) introduction of national entrance examinations for higher education; (iv) flexible curricula, according to the specific needs of each school and its students; and (v) an enriched and updated list of textbooks and school libraries. Also, the Education Ministry engaged in an extensive construction effort, which resulted in 3167 new school buildings—twice as many as constructed during a century of independent statehood. These reforms constituted a breakthrough, and they were developed continuously and consistently through a succession of legislative adjustments, over a 3-year period. Following their abolition by the next government, many of these reforms still remain unrealized goals in the early twenty-first century (Stefanidis 2006).

4 Data and specification

The empirical analysis proceeds as follows. First, we investigate if regional differences in geography influence human capital in Greece. Second, we examine whether geography also affects land inequality. The underlying rationale is that regional variation in geography had generated a heterogeneous demand for land over space and contributed to the formation of large properties in Greece in certain areas (see discussion in Section 3.1; Easterly 2007; Cinnirella and Hornung 2016). Third, we explore if land inequality is one channel through which geography influences human capital accumulation; in other words, how much of the geography-human capital relationship is due to inequality in landholdings. Fourth, we check whether this relation depends on the level of industrialization, i.e., if the role of geography in human capital accumulation declines as the economy moves from the agricultural to the industrial stage of development.

To this end, we build an entirely novel data set with Greek historical data at the provincial level, starting with the first Census of 1861. Our sample consists of 142 provinces. We mainly use data from the 1928 Census of the Hellenic Statistical Authority because it is the first period, which includes extended coverage at the provincial level for Greece after the annexation of the Northern provinces, Eastern and Northern Aegean Islands as well as Crete in 1913–1920, after which Greece almost doubled in size.[2]

---

[2] Definitions of all variables are available in Table A.1, in the Appendix. The Supplementary Appendix provides additional maps.
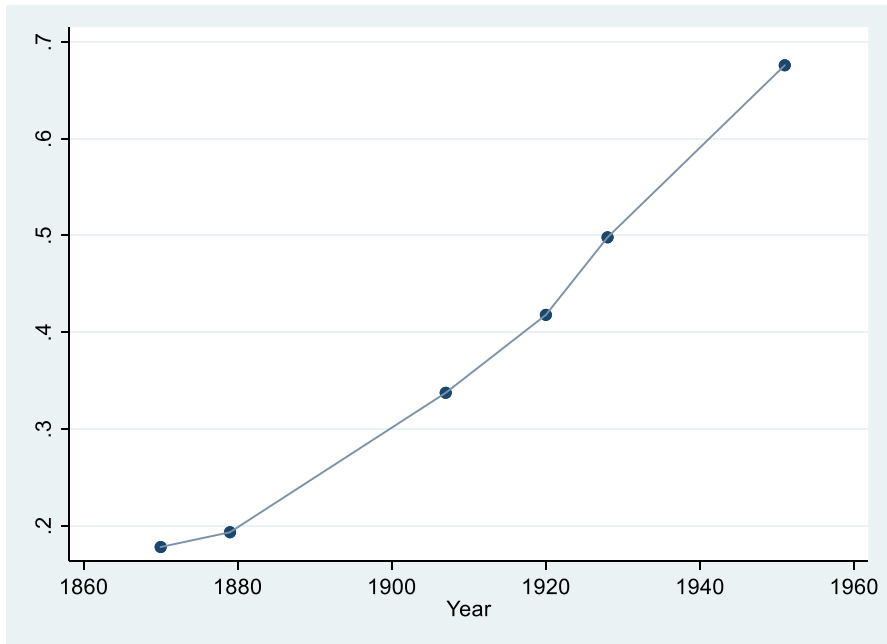
Overall, literacy starting from below 20% in 1861 gradually increased over time (Fig. 1). We collect literacy rates for both natives and refugees. Our estimations are based on the literacy of the natives, as it is the one acquired through education in Greece. At a later stage, we include refugee literacy as a control variable, to account for possible human capital spillovers (see discussion in Appendix Section A.2).[3]

Map 2 illustrates the geographical distribution of native literacy rates in Greek provinces in 1928. The Attica province, which includes the capital city of Athens, together with the major urban centers of Thessaloniki, Kavala (in the north), Volos, Larissa (central), Patras (south-west), and Ermoupolis (in Cyclades islands) exhibit the highest literacy rates. The lowest literacy rates are present in Thrace, parts of Central and western Macedonia, western Thessaly, western Epirus, and western Central Greece.

The geographic factors we use are soil suitability (wheat, rice, grain, cereal, pulse, oil, sugar, cotton), temperature, ruggedness, and precipitation. Ruggedness is the standard deviation of altitude, which is collected from the 1951 Census of the Hellenic Statistical Authority. The 10 remaining variables are obtained using GIS techniques. Specifically, we have downloaded maps, which contain suitability data from the Food and Agriculture Organization (FAO) and climatic data from the IPUMS Terra (Integrated Population and Environmental Data) databases. We have taken the coordinates of Greek municipalities from Geodata, a Greek geospatial

---

[3] Note that the language of instruction in all schools was Greek.
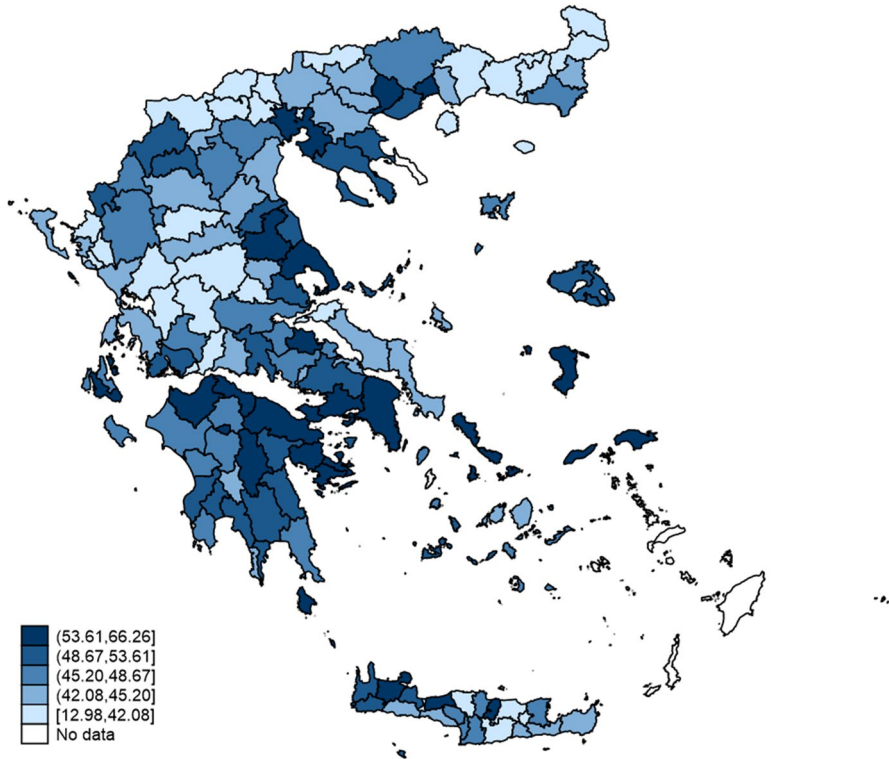
Map 2 Native literacy rates in 1928. Source: Hellenic Statistical Authority, 1928 Census

database, and combined maps and coordinates using the QGIS software in order to extract geographic, suitability, and climatic data at provincial level (Maps S.2-S.11 in the Supplementary Appendix). We also compute climate (temperature and precipitation) variability as a proxy for trust from World Bank historical weather data (Appendix Section A.3).

The main non-geographic variable of interest concerns land inequality and demonstrates if large land holdings are dominant in a province or not. Direct official quantitative data regarding early twentieth century exist for very few Greek provinces. In light of this, we build four land inequality variables: (i) a qualitative one based on historical sources; (ii) a quantitative one extracted from an official map; (iii) two indicators based on the 1929 Agricultural Census.

First, we construct a qualitative variable for all provinces based on economic history sources on land tenure (Vergopoulos 1975; McGrew 1985; Petmezas 2003, 2012; Kostis and Petmezas 2006; Kontogiorgi 2006). We use alternative sources in order to ensure that the qualitative information we obtain is credible and to achieve the desired geographic coverage at the required level of aggregation. The information on land tenure does not indeed differ between these alternative sources. In other words, these present the same picture with regard to the areas where large properties

dominated vs. the areas with mostly small landholdings. Only the geographic coverage and level of regional aggregation vary between relevant studies, and this is the second reason for which we use multiple sources of information to construct the land inequality dummy variable. According to these sources, large landholdings in Greece were concentrated in areas where chiftliks existed (Section 3.1). Thus, land inequality is measured by a dummy variable taking values 0 and 1, which correspond to regions with weak and strong presence of large land holdings until 1922, respectively. For instance, we set this dummy equal to 1 for provinces of Attica, Thivon, Lokridas, Fthiotidas, and Xirochoriou and provinces belonging to the former Ottoman kazas of Kilkis, Edessis, Langada, Katerinis, Sidirokastrou, Dramas, Elassonas, and Servion, for which there is strong evidence of dominance of large properties (Petmezas 2003, 2012). We set the dummy equal to 0 in provinces for which it is mentioned that small properties prevailed, e.g., Karatzova, Enotias, and Florinas in Macedonia (Petmezas 2012). In rare cases, specific numbers are given, e.g., in Trikala prefecture, where there were 213 large properties, and only 48 settlements were not chiftliks (Petmezas 2012). For very few regions, there are tables with detailed data on the area of large properties and the respective share of agricultural population at the province level (e.g., Table 1.20, 1.29 in Petmezas 2012 and in p. 176 in Vergopoulos 1975). Regional differences in landownership are very stable over time until 1922, i.e., the occurrence of mass refugee inflows, because mass expropriations of large landholdings started soon afterwards in order to distribute this land mainly to refugees (Vergopoulos 1975).

We construct a map for the geographical distribution of landownership in Greek provinces referring to the period before the expropriation of large properties by the state, where dark grey corresponds to dominance of large properties and light grey corresponds to weak presence of large properties (Map 3). Provinces with strong prevalence of large land holdings are found mainly in Central Greece/Euboea, Thessaly, Epirus, and Macedonia. Provinces characterized by weak occurrence of large properties are mainly located in Peloponnese, Crete, Ionian Islands, Cyclades, Aegean Islands, and Thrace.

Second, we generate another variable measuring land inequality, from the digitization of the only available official (governmental) map which presents the composition of land in Greece in 1936 (Map 1 in Section 3.1).[4] This map distinguishes between distributed and not distributed land to farmers, as well as heath land. Specifically, we proxy large land holdings by the ratio of distributed land to total land for each province. This is because distributed land comprised mostly former large

---

[4] The basemap has been obtained in the form of raster data by scanning the paper map in 600 dpi, stored in a lossless compression image data format (TIFF). In order to assign spatial reference information and metric properties to the scanned map, a georeferencing process has been applied. The Georeferencer module of the QGIS was the software utilized. Due to the lack of coordinate grid on the raster map, well-defined objects within the image and the country vector layer, such as ports, capes, and other land features, have been identified as ground control points. A total number of 100 control points, spread over the entire raster, have been collected. A second-order polynomial transformation has been applied, resulting in a RMS error of approximately 2 km. The target raster map has been resampled to 250 m spatial resolution, projected to the Hellenic Geodetic Reference System 1987 (HGRS87, EPSG:2100) to derive the required metrics.
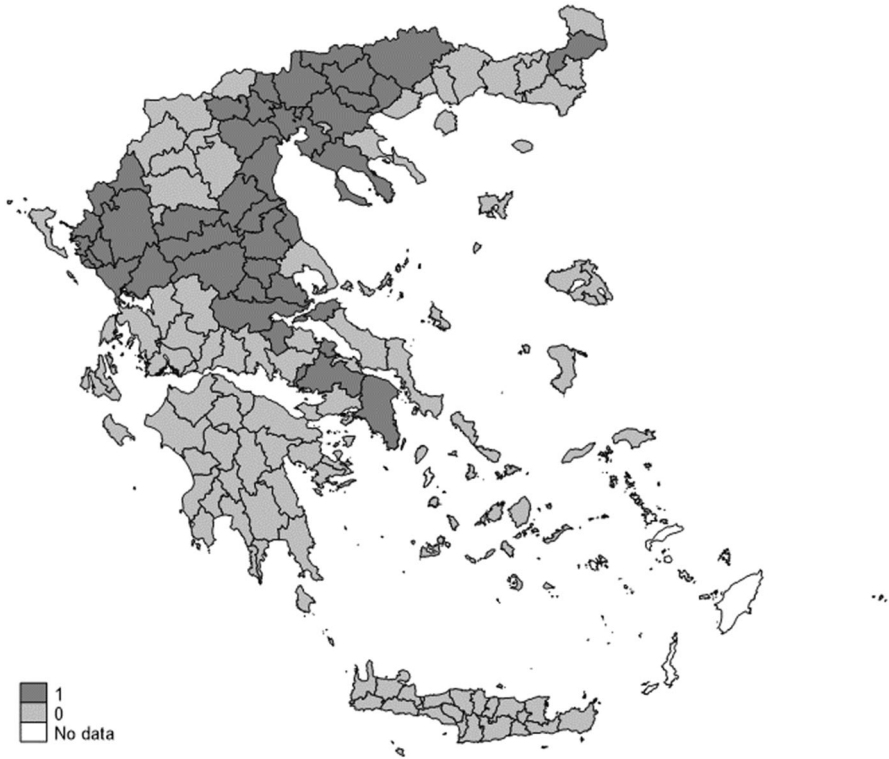
Map 3 Land inequality in early twentieth century. Number "0" corresponds to provinces with dominance of small land holdings and number "1" corresponds to provinces with dominance of large land holdings. Source: Various historical sources (see Section 4 in text for details)

land properties, which were expropriated and allocated to refugees as well as landless native farmers after 1922, since the vast majority of the expropriations of large landholdings took place between 1923 and 1928 (see Section 3.1 for details). So, the newly constructed variable provides a good proxy of the prevalence of large land holdings before 1922, although it also includes marshes and lakes that were desiccated along with small properties of Bulgarians and Turks that left the country in early 1920s (Map A.1 in the Appendix). Hence, our land inequality variable concerns the period before 1928, when literacy is observed.

Finally, we use the earliest Agricultural Census of 1929 covering Greece in 1928, which contains data at the prefecture level for the size of land properties in 27 bins to derive two alternative measures of land inequality. First, we define large properties as those of 101–300,000 stemmata and calculate their respective shares in arable land in each prefecture.[5, 6] We define prefectures with high land inequality

---

[5] Stremma (or stremmata in plural) was the land metric unit used at the time and is equal to 1000 m^2.

[6] The cutoffs of 101 and 501 stremmata (see also below) were commonly used to define large properties in Greece at the time (Seferiades 1999).

those where the above share is higher than the 75th percentile of the distribution of this variable (52.5%) and create a binary dummy equal to one in prefectures where the share is higher than 52.5% and zero otherwise. We transform these data into province-level data by giving the same dummy value to all provinces in the same prefecture. In very few prefectures characterized by strong heterogeneity in terms of land inequality between their provinces according to historical evidence, we change the value of the dummy (see Section 3.1, 4). This is intended to account for two factors: (a) aggregation bias, due to the fact that in some prefectures, the land distribution of their largest provinces dominates the distribution of small provinces, hiding significant disparities within prefectures; (b) expropriation of large land holdings initiated prior to 1929 and completed for many of them before the Census took place, so that the latter did not reflect land distribution around 1900 (Map A.2 in the Appendix). Second, as a supplement to the qualitative variable described above, we use the share of very large properties (those between 501 and 300,000 stremmata) in arable land in each prefecture without any adjustment in order to check the robustness of our findings with a quantitative variable derived directly from the Census (Map A.3 in the Appendix). The disadvantages are of course aggregation bias, not taking into account the expropriation of large properties during 1923–1929 (see above) and low variation due to prefecture-level data. Overall, we have used all available data sources on land tenure during our sample period for the construction of land inequality measures, which makes us confident about our findings.

5 Empirical results

5.1 Has geography affected human capital formation?

First, we examine whether geographic factors are related to literacy. We employ OLS and regional (prefecture) fixed effects because the regressions are likely to suffer from endogeneity in the form of omitted variable bias. We also allow for clustered standard errors at the prefecture level to account for spatial error dependence (Table 1).

The effects of the geographic variables as a whole on literacy are significant at the 1% level as shown by the F statistic at the bottom of the table. This is in line with the high prevalence of large properties in specific areas of Greece, mainly the eastern and northern part. These findings are consistent with a large number of papers, which find a significant association between land inequality and geography measures (see for example GMV; Easterly 2007; Cinnirella and Hornung 2016; Baten and Hippe 2018).

Table 1 Literacy and geography

| | Literacy rate | | |
	(1)	(2)	(3)
Wheat suitability	0.036 (0.051)		0.023 (0.059)
Rice suitability	−0.049 (0.036)		−0.048 (0.042)
Grain suitability	0.054 (0.037)		0.056 (0.046)
Cereal suitability	−0.047 (0.047)		−0.037 (0.056)
Pulse suitability	0.025 (0.038)		0.021 (0.042)
Oil suitability	0.016 (0.014)		0.020 (0.015)
Sugar suitability	−0.031 (0.033)		−0.028 (0.032)
Cotton suitability	−0.006 (0.025)		−0.007 (0.028)
Ruggedness		0.000 (0.000)	0.000 (0.000)
Precipitation		−0.001 (0.001)	−0.001 (0.001)
Temperature		0.001 (0.001)	0.001 (0.001)
Constant	0.368*** (0.014)	0.333*** (0.100)	0.344*** (0.109)
R^2	0.62	0.60	0.63
F-stat	56.30***	2.57*	12.05***
Observations	140	138	138

Regressions include prefecture fixed effects. Standard errors are clustered at the prefecture level. Standard errors are presented in parentheses for the OLS regressions. F-statistic concerns the joint significance of geographic variables

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

Next, we provide evidence on the relationship between land inequality and geography in order to check if land distribution is a potential channel through which geography affects literacy. To this end, we regress land inequality on the same geographic variables used in the literacy specification presented above using the estimation methodology used before (Table 2). The geographic indicators are very important as a whole in explaining land distribution regardless of the specific measure of land inequality used.

Overall, geography is associated with both literacy and land inequality. Hence, land distribution is shown to be a mechanism behind the effect of geography on literacy in line with GMV theory. In light of this, in the following section, we estimate literacy as a function of geography and land inequality. We incorporate all geographic controls in the estimated specifications in order to check whether and how they affect the land inequality-literacy relationship.

Table 2 Land inequality and geography

| | Land inequality | | | |
| | Historical dummy | Census dummy | Census continuous | Map-based |
	(1)	(2)	(3)	(4)
Wheat suitability	0.034 (0.376)	− 0.168 (0.178)	− 0.158** (0.071)	0.187 (0.123)
Rice suitability	− 0.680*** (0.182)	− 0.330** (0.150)	− 0.124** (0.047)	− 0.234*** (0.083)
Grain suitability	0.701*** (0.199)	0.244 (0.186)	0.085 (0.051)	0.299*** (0.102)
Cereal suitability	− 0.151 (0.341)	0.045 (0.129)	0.134** (0.057)	− 0.262** (0.114)
Pulse suitability	− 0.236 (0.181)	0.046 (0.175)	− 0.023 (0.046)	− 0.201** (0.098)
Oil suitability	− 0.062 (0.074)	− 0.048 (0.089)	0.046* (0.024)	− 0.001 (0.040)
Sugar suitability	0.110 (0.095)	0.097 (0.110)	− 0.011 (0.036)	0.108 (0.071)
Cotton suitability	0.287* (0.149)	0.114 (0.152)	0.055* (0.031)	0.116 (0.072)
Ruggedness	0.001 (0.001)	0.001 (0.001)	0.000 (0.000)	0.000 (0.000)
Precipitation	− 0.003* (0.002)	0.001 (0.001)	− 0.000 (0.001)	− 0.003*** (0.001)
Temperature	− 0.001 (0.002)	0.001 (0.001)	− 0.000 (0.001)	0.001 (0.001)
Constant	0.521 (0.377)	− 0.159 (0.303)	0.341*** (0.121)	0.321 (0.228)
R^2	0.31	0.15	0.14	0.43
F-stat	147.36***	465.31***	3.67***	105.14***
Observations	139	138	138	137

Standard errors are clustered at the prefecture level. Standard errors are presented in parentheses for the OLS regressions. F-statistic concerns the joint significance of geographic variables

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

5.2 Has land inequality affected human capital?

We provide evidence on the implication of GMV theory concerning the presence of a negative relationship between land inequality and literacy in Greek regions, found in most studies for other countries (Easterly 2007; Wegenast 2009; Cinnirella and Hornung 2016; Baten and Hippe 2018; Gershman et al. 2022). We estimate a regression model of literacy on land inequality and geographic controls. We use four different variables for land inequality. First, we employ a qualitative variable for all provinces based on economic history sources on land tenure. Next, we replicate the initial estimations using two land inequality variables derived from the 1929 Agricultural Census (Table 3). Finally, we use an alternative measure of large land holdings extracted from the digitization of the only available map which presents the composition of land in Greece in 1936 (Section 4). We also employ 11 geographic factors and regional (prefecture) fixed effects. Table 3 provides the *OLS* estimates.

The effect of land inequality on literacy is sizeable and negative for all inequality measures, and it is significant when the dummy based on historical sources is used. Note that in the fractional estimations, all inequality variables are significant except the continuous Census one (see Table A.4 in the Appendix). For instance, in the OLS estimations, the dominance of large land holdings reduces literacy by approximately 5.4% (or 11.6% lower than the average of 46.5%; see col. 1, Table 3). This underlines the importance of land distribution as a mediator in the geography-literacy

Table 3 Literacy and land inequality-baseline estimations

Land inequality	Literacy rate			
	Historical dummy	Census dummy	Census continuous	Map-based
	(1)	(2)	(3)	(4)
Land inequality	− 0.054*** (0.016)	− 0.035 (0.030)	− 0.021 (0.078)	− 0.024 (0.042)
Wheat suitability	0.014 (0.042)	0.009 (0.055)	− 0.046 (0.050)	− 0.038 (0.050)
Rice suitability	− 0.054 (0.045)	− 0.050 (0.042)	0.047 (0.042)	0.044 (0.042)
Grain suitability	0.068 (0.048)	0.061 (0.046)	− 0.074* (0.039)	− 0.069 (0.043)
Cereal suitability	− 0.036 (0.039)	− 0.029 (0.050)	0.042 (0.050)	0.032 (0.051)
Pulse suitability	0.020 (0.045)	0.023 (0.043)	0.067** (0.029)	0.063* (0.031)
Oil suitability	0.021 (0.016)	0.021 (0.015)	0.007 (0.019)	0.005 (0.019)
Sugar suitability	− 0.029 (0.034)	− 0.028 (0.033)	− 0.012 (0.028)	− 0.010 (0.029)
Cotton suitability	− 0.004 (0.028)	− 0.007 (0.026)	− 0.035 (0.028)	− 0.033 (0.028)
Ruggedness	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Precipitation	− 0.001 (0.001)	− 0.001 (0.001)	− 0.000 (0.000)	− 0.000 (0.000)
Temperature	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.001* (0.001)
Constant	0.329*** (0.102)	0.330*** (0.097)	0.319** (0.136)	0.319** (0.132)
R^2	0.65	0.64	0.18	0.19
F-stat	83.35***	15.83***	8.20***	7.97***
Oster ratio	0.89	1.79	0.08	0.11
Observations	138	138	138	136

Columns 1–2 include fixed effects at the prefecture level and cluster standard errors at the prefecture level. Standard errors are presented in parentheses for the OLS regressions. *F*-statistic concerns the joint significance of geographic variables. Oster ratio exhibits the degree of selection of unobservables relative to observables such that the effect of land inequality vanishes, accounting for variations in R^2 *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$

relationship. Regarding geography, the corresponding variables are jointly significant for literacy at the 1% level in all specifications as is the case in Table 1. These findings are consistent with a large number of papers, which find a negative association between land inequality and human capital measures (see for example GMV; Easterly 2007; Cinnirella and Hornung 2016; Baten and Hippe 2018).

5.3 Threshold estimations

Threshold regression extends linear regression allowing coefficients to vary across regions. Those regions are determined by a threshold variable according to whether it is higher or lower than a specific threshold value. We use this methodology to investigate the possibility of varying relationship between geography and literacy depending on the level of industrialization. If GMV theory holds for Greek provinces, the impact of geography would be weaker in more industrialized provinces compared to less industrialized ones.

Table 4 Threshold estimations-baseline estimations

Literacy rate

Land inequality	Historical dummy	Census dummy	Census continuous	Map-based	Historical dummy	Census dummy	Census continuous	Map-based
	Region 1	Region 1	Region 1	Region 1	Region 2	Region 2	Region 2	Region 2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Land inequality	0.012 (0.026)	−0.016 (0.021)	−0.023 (0.065)	−0.001 (0.030)	−0.047 (0.029)	−0.020 (0.027)	−0.014 (0.089)	−0.044 (0.040)
Wheat suitability	−0.129*** (0.041)	−0.133*** (0.038)	−0.131*** (0.038)	−0.131*** (0.038)	0.027 (0.042)	0.045 (0.047)	0.046 (0.050)	0.063 (0.058)
Rice suitability	0.116** (0.046)	0.102** (0.041)	0.105*** (0.038)	0.107*** (0.039)	−0.008 (0.032)	0.017 (0.026)	0.023 (0.032)	0.010 (0.022)
Grain suitability	−0.053 (0.036)	−0.043 (0.034)	−0.042 (0.035)	−0.044 (0.034)	−0.015 (0.031)	−0.041 (0.028)	−0.047 (0.031)	−0.028 (0.025)
Cereal suitability	0.118*** (0.030)	0.117*** (0.026)	0.116*** (0.027)	0.117*** (0.027)	−0.017 (0.027)	−0.028 (0.028)	−0.024 (0.033)	−0.045 (0.040)
Pulse suitability	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.103*** (0.023)	0.115*** (0.024)	0.104*** (0.021)	0.108*** (0.021)
Oil suitability	0.027* (0.015)	0.024 (0.015)	0.026* (0.015)	0.027* (0.015)	−0.035* (0.021)	−0.036* (0.022)	−0.037 (0.023)	−0.043* (0.022)
Sugar suitability	0.034 (0.023)	0.040* (0.024)	0.035* (0.021)	0.036 (0.022)	−0.083*** (0.025)	−0.090*** (0.028)	−0.084*** (0.028)	−0.089*** (0.027)
Cotton suitability	−0.110*** (0.039)	−0.104*** (0.035)	−0.106*** (0.034)	−0.108*** (0.034)	0.019 (0.018)	0.008 (0.016)	0.009 (0.019)	0.014 (0.019)
Ruggedness	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)
Precipitation	0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Temperature	0.001*** (0.000)	0.002*** (0.000)	0.001*** (0.000)	0.001*** (0.000)	−0.000 (0.000)	−0.000 (0.000)	0.000 (0.000)	−0.000 (0.000)
Constant	0.190*** (0.094)	0.187** (0.094)	0.203** (0.093)	0.201** (0.093)	0.548*** (0.079)	0.518*** (0.070)	0.507*** (0.102)	0.561*** (0.077)
F-stat	73.53***	192.11***	218.19***	76.10***	77.93***	76.27***	94.99***	123.66***
Observations	138	138	138	136	138	138	138	136

Robust standard errors. Standard errors are presented in parentheses. F-statistic concerns the joint significance of geographic variables

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

Formally, consider a threshold specification with two regions defined by a threshold γ as follows:

$$y_i = x_i\beta + z_i\delta_1 + \varepsilon_i \text{ if } -\infty < w_\iota \leq \gamma$$
$$y_i = x_i\beta + z_i\delta_2 + \varepsilon_i \text{ if } \gamma < w_\iota < \infty$$

where y_i is literacy, x_i is land inequality, β is a $k \times 1$ vector of region-invariant parameters, z_i is the vector of geographic variables with region-specific coefficient vectors δ_1 and δ_2, w_i is a threshold variable (industrial employment), and ε_i is an IID error term with mean 0 and variance σ^2. The parameters of interest are δ_1, δ_2. Region 1 is the subset of observations in which w_i is less than the threshold, while region 2 is the subset of observations in which the value of w_i is higher than the threshold (Hansen 1997, 2000). Threshold estimation uses conditional least squares to estimate the parameters of the model. The estimated threshold $(\hat{\gamma})$ is one of the values of the threshold variable w_i and is estimated by minimizing the SSR obtained for all tentative thresholds.

We use industrial employment as threshold variable and allow geographic variables to vary over the two regions separated by the estimated threshold value, since we want to capture the possibility of varying effects of geography on literacy depending on the level of industrialization. According to our results, there is indeed a threshold at 10.91% of manufacturing employment. Geography plays a more important role when provinces exhibit industrialization below the threshold than above it. Specifically, six geographic variables are individually significant in less industrialized provinces vs four variables in more industrialized provinces (Table 4, cols 1 and 5 respectively). A test of their joint significance shows that geography as a whole is a very important determinant of literacy in both low- and high-industrialization regions. Overall, these findings are in line with the implications of GMV theory for the declining role of agriculture as industrialization gains pace.

6 Omitted variables

We apply Oster analysis, to check for omitted variable bias (Oster 2019). This is based on the idea that both the stability and importance of the coefficients matter for the detection of this bias. So, when we scale coefficient movements according to changes in R^2, the bias is proportional to the resulting changes in coefficients. Thus, we calculate the degree of selection of unobservables relative to observables such that the effect of land inequality vanishes, accounting for variations in R^2.[7]

In all tables, we show the relative degree of selection on unobservables such that the impact of land inequality on literacy is statistically zero, accounting for R^2 movements. The degree of relative selection on unobservables is 0.89 and 1.79 in the baseline specifications with the historical dummy and the Census dummy respectively suggesting that the unobservables must be up to almost twice as important as

---

[7] Here, we should emphasize that Oster ratios cannot be applied in fractional estimations.

the observables in order to eliminate the effect of land inequality. There is a potential issue of bias only in the two continuous land inequality baseline specifications. However, in the equations with the most comprehensive set of controls, the selection bias parameter ranges from 0.83 to 2.16 for all four land inequality measures (please see Table A.5 in the Appendix). Therefore, our estimates are unlikely to suffer from omitted variable bias.

7 Concluding remarks

The role of geography has been the center of a discussion about the origin of differences in the timing of transition from stagnation to growth. Theory suggests that land inequality is a plausible mechanism, being influenced by geographic characteristics. Much literature on the relationship between geography, human capital, and landownership focuses on mature industrial economies, while our work examines Greece at the early stage of the transition to the industrial era.

In this study, we investigate how geography affects literacy, thus the evolution of human capital, using a newly constructed historical province-level dataset in Greek regions around 1900. We demonstrate a strong association between geography and literacy. We confirm that land distribution is a mediator through which geography influences literacy, as predicted by GMV and Gershman et al. (2022). Overall, provinces characterized by large landholdings were falling behind in terms of human capital accumulation, which impeded rapid transformation from agriculture to industry. Our research underlines the importance of human capital in the development process, reflecting its complementarity with physical capital and technology. The evidence suggests that a more equal distribution of land in early twentieth century would help to foster educational attainment and economic growth in Greece.

Data Availability The datasets analyzed in this paper are available from the corresponding author on reasonable request.

Declarations

Disclaimer The views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Training Foundation or the European Union institutions.

Conflict of interest The authors declare no competing interests.

References

Adamopoulos T, Restuccia D (2022) Geography and agricultural productivity: cross-country evidence from micro plot-level data. Rev Econ Stud 89(4):1629–1653

Alivizatos B (1938) State and Agricultural Policy National Printing House, (in Greek)

Andersson J, Berger T (2019) Elites and the expansion of education in nineteenth-century Sweden. Econ Hist Rev 72(3):897–924

Ashraf QH, Galor O (2013) The 'Out of Africa' hypothesis, human genetic diversity, and comparative economic development. Am Econ Rev 103(1):1–46

Baten J, Hippe R (2018) Geography, land inequality and regional numeracy in Europe in historical perspective. J Econ Growth 23:79–109

Baten J, Juif D (2014) A story of large landowners and math skills: Inequality and human capital formation in long-run development, 1820–2000. J Comp Econ 42(2):375–401

Chantzopoulos K (1998) The education in Greece (1821–1985). *Lecture notes Alexandroupoli*, (in Greek)

Cinnirella F, Hornung E (2016) Landownership concentration and the expansion of education. J Dev Econ 121:135–152

Deininger K, Squire L (1998) New ways of looking at old issues: inequality and growth. J Dev Econ 57(2):259–287

Diamond J (1997) Guns, germs, and steel. New York: A short history of everybody for the last 13,000 years. W.W. Norton & Company

Dimaras A (2006) Modernization and reaction in Greek education during the Venizelist era. *Eleftherios Venizelos: The Trials of Statesmanship*, 319–345

Easterly W (2007) Inequality does cause underdevelopment: insights from a new instrument. J Dev Econ 84(2):755–776

Erickson LA (2004) Dimensions of land inequality and economic development. *IMF working paper No. 04/158*

Galor O, Ozak O (2016) The agricultural origins of time preference. Am Econ Rev 106(10):3064–3103

Galor 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

Gersbach H, Siemers LHR (2010) Land reforms and economic development. Macroecon Dyn 14(4):527–547

Gershman B, Ashraf QH, Cinnirella F, Galor O, Hornung E (2022) Structural change, elite capitalism, and the emergence of labor emancipation. *Working Paper 2022–07*, American University, Department of Economics

Goñi M (2022) Landed elites and education provision in England: evidence from school boards, 1871–99. J Econ Growth 1–47

Hansen BE (1997) Approximate asymptotic p values for structural-change tests. J Bus Econ Stat 15(1):60–67

Hansen BE (2000) Sample splitting and threshold estimation. Econometrica 68(3):575–603

Kontogiorgi E (2006) Population exchange in Greek Macedonia: the rural settlement of refugees 1922–1930. Oxford University Press on Demand

Kostis K, Petmezas, S (2006) The development of Greek Economy in 19th century. Alexandria Editions, Athens, (in Greek)

Martinelli P (2014) Latifundia revisited: market power, land inequality and agricultural efficiency. Evidence from interwar Italian agriculture. Explor Econ Hist 54:79–106

McGrew WW (1985) Land and revolution in modern Greece, 1800–1881: the transition in the tenure and exploitation of land from Ottoman rule to independence. Kent State University Press

Michalopoulos S, Papaioannou E (2017) The long economic and political shadow of history: Europe and the Americas. CEPR Working Paper

Oster E (2019) Unobservable selection and coefficient stability: theory and evidence. J Bus Econ Stat 37(2):187–204

Petmezas S (2003) The Greek agricultural economy in the 19th century. University of Crete Editions, Crete, (in Greek)

Petmezas S (2012) Introduction in the history of Greek agricultural economy in the interwar period. Alexandreia Editions, Athens, (in Greek)

Sachs JD, Warner AM (1999) The big push, natural resource booms and growth. J Dev Econ 59(1):43–76

Seferiades S (1999) Small rural ownership, subsistence agriculture, and peasant protest in interwar Greece: the agrarian question recast. J Mod Greek Stud 17(2):277–323

Sokoloff KL, Engerman SL (2000) Institutions, factor endowments, and paths of development in the new world. J Econ Perspect 14(3):217–232

Stefanidis ID (2006) Reconstructing Greece as a European state: Venizelos' last premiership 1928–32. The Trials of Statesmanship, Edinburgh University Press, Eleftherios Venizelos, pp 193–233

Tapia FJB, Martinez-Galarraga J (2018) Inequality and education in pre-industrial economies: evidence from Spain. Explor Econ Hist 69:81–101

Vergopoulos K (1975) The agricultural issue in Greece-the problem of social integration of agriculture. Exantas editions, Athens, (in Greek)

Wegenast T (2009) The legacy of landlords: educational distribution and development in a comparative perspective. Zeitschrift Für Vergleichende Politikwissenschaft 3(1):81–107