Department of Economics

School of Economics and Management Science

University of Ioannina

ESSAYS ON HUMAN CAPITAL AND REGIONAL ECONOMICS

Sofia Tsitou

A dissertation submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN ECONOMICS

Ioannina

December 2021

ESSAYS ON HUMAN CAPITAL AND REGIONAL ECONOMICS

Benos Nikos, Associate Professor
Department of Economics
University of Ioannina

Karagiannis Stelios, Research Executive Agency
European Commission*, Belgium*
Centre of Planning and Economic Research*, Greece*

Karkalakos Sotiris, Professor
Department of Economics
University of Piraeus

Cappelli Gabriele, Assistant Professor
Department of Economics and Statistics
University of Siena

Litina Anastasia, Assistant Professor
Department of Economics
University of Macedonia

Mocan H. Naci, Professor
Department of Economics
Louisiana State University

Panagiotidis Theodore, Professor
Department of Economics
University of Macedonia

Examination Date: December 03, 2021

This research is co-financed by Greece and the European Union (European Social Fund-ESF) through the Operational Programme «Human Resources Development, Education and Lifelong Learning» in the context of the project “Strengthening Human Resources Research Potential via Doctorate Research” (MIS-5000432), implemented by the State Scholarships Foundation (IKY).

Acknowledgments

Throughout the writing of my doctoral dissertation I have received great help and support.

First, I would like to thank and express my appreciation to my supervisor, Associate Professor Nikos Benos, for his guidance, ideas, patience, support and valuable advice during this process. I could not have imagined finishing this dissertation without his assistance. I will always appreciate him for the time he devoted, for his wonderful ideas and for everything he has done whenever it was needed. In addition, I would like to thank him for the courage and advice he gave me at difficult times.

I am also so grateful to Professor Sotiris Karkalakos, who offered me his continuous support, motivation and immense knowledge throughout my PhD study. I truly appreciate him for providing the initial research idea of the spatial analysis and his assistance and support. The third chapter of my dissertation would not have been possible without his help. Specifically, his research input and feedback have greatly contributed to the completion of the third part and the whole present thesis.

Of course, I would like to express my sincere gratitude to Stelios Karagiannis for his very supportive comments and advices through my entire PhD thesis. His advice and assistance from the first moment has been an inspiration and a driving force throughout the course of this dissertation, which would not have been finished without his directions and help.

I am grateful to Assistant Professor Anastasia Litina for extremely insightful comments and suggestions which greatly contributed to the improvement of this thesis. Her continuous encouragement has been crucial throughout the whole development of this dissertation. I appreciate her for offering the original idea for the second chapter and her overall subsequent help for this chapter.

Special thanks to Professors Theodore Panagiotidis, H. Naci Mocan and Gabriele Cappelli who accepted to be members of my defense committee. I appreciate them very much and it is an honor for me to be members of my dissertation. They are honorable professors with a lot of important work in the field of economic research. I have also greatly benefited from Assistant Professors, Achilleas Vassilopoulos and Pantelis Kazakis for detailed remarks and recommendations. I am also grateful to the participants of the conferences and seminars that I have presented my work for their valuable comments and suggestions.

Last but not least, many sincere and special thanks to my husband Dimitris and my kids Nikos and Giorgos for all their love, patience and encouragement during this challenging period. I would also like to thank my parents Vangelis and Georgia, as well as my sister Eleftheria, for supporting me during all these years and in my life in general.

Table of Contents

Introduction.....9

Human Capital.....9

Regional Economics.....10

Empirical Literature Review.....12

Chapter 1: Geography, Land Ownership and Literacy: Historical Evidence from Greek Regions.....18

1.1 Introduction.....18

1.2 Literature Review.....21

1.3 Data and Methodology.....26

1.3.1 Data.....26

1.3.2 Methodology.....37

1.4 Empirical Results.....40

1.5 Long-Run Development.....49

1.6 Concluding Remarks.....51

Appendix of Chapter 1 54

Maps.....54

Tables.....67

Figures.....93

Chapter 2: The Implications of the Population Exchange between Greece and Turkey on Literacy and Employment Rates of the Native Population.....99

2.1 Introduction.....99

2.2 Related Literature.....101

2.3 Historical Background……111

2.3.1 The Turkish-Greek War of 1920-1922……111

2.3.2 Refugees: Conditions on Arrival and Allocation over Greece……112

2.3.3 Composition of the Refugee Population……114

2.3.4 Education of Refugees and Natives……115

2.3.5 Labor Force Participation and Sectoral Allocation for Refugee and Native Population……117

2.4 The Data……118

2.4.1 Descriptives……123

2.5 Empirical Specification (Diff-in-Diff)……134

2.5.1 Forced Displacement as a Natural Experiment and Identification Concerns……135

2.5.2 Treatment……136

2.5.3 Specification……137

2.5.4 Parallel Trend Assumption……138

2.6 Results……139

2.6.1 Literacy……139

2.6.2 Employment Allocation……142

2.7 Mechanisms……146

2.8 Conclusions……149

Appendix of Chapter 2 ……150

Definitions of Variables at the Provincial Level……150

Chapter 3: A European metropolitan analysis of enterprise births……153

3.1 Introduction……153

3.2 Literature Review……158

3.3 Data and Methodology……174

3.3.1 Data……174

3.3.2 Methodology……178

3.4 Empirical Results……181

3.4.1 Baseline Results……181

3.4.2 GMM Results……183

3.4.3 Spatial Results……184

3.5 Concluding Remarks……187

Appendix of Chapter 3……189

Tables……189

Maps……201

Conclusions……203

References ……205

Introduction

Human capital

Human capital determinants and the impact of human capital on various aspects of economic outcomes have been long studied in the literature. The role of human capital formation has been the center of a discussion about the origin of the differential timing of the transition from stagnation to growth in an economy. The country's ability to develop its human resources, specifically in providing skilled workforces in various fields, becomes the key to success to economic policies in every country. Becker (1962) defines human capital as an aggregation of knowledge, skills, competencies and abilities acquired by human beings over the course of their lives, developed through participation in various forms of formal and informal education and training that enhance individual productivity. Economists refer to expenditures on education and training as investments in human capital. Human capital contributes to economic growth and development of a country in a big way. Over the years, growth theories have evolved to include human capital as a significant determinant of economic growth. Many theories explicitly connect investment in human capital development to education. The role of human capital in economic development, productivity growth, and innovation has frequently been cited as a justification for government subsidies for education and job skills training. As the world accumulated more and more physical capital, the opportunity cost of going to school declined, thus education became an increasingly important component of the workforce.

Human capital development has contributed to higher levels of economic productivity and profitability, thus it is a major factor in many growth models. From a theoretical aspect, there are two basic categories of growth models; exogenous and endogenous growth models. Exogenous growth models include the Solow-Swan model and its extensions (neoclassical growth models) and focus on the role of physical capital. On the other hand, endogenous growth models broaden the concept of capital to include human capital. Because of this, endogenous growth models can be further distinguished into three subcategories. The first category is growth models with human capital accumulation. Human capital is considered as a major factor of production, and its accumulation influences the growth process of the economy. The second category is growth models promoting human capital and technological progress. Economic growth depends on the existing stock of human capital, which can be useful for generating and adopting new technologies. The last category is growth models with human capital and threshold effects. These models makes the assumption that there could be a threshold level of human capital below which poorer economies do not catch up with the richer ones.

Regional Economics

Regional economics is a relatively young branch of economics. Specifically, regional economics are a sub-discipline of economics and is often regarded as one of the fields of the social sciences. Their late start depicts the regrettable tendency of formal professional disciplines to lose contact with one another and to neglect some important problem areas that require a mixture of approaches. Regional economics are all economic activities taking place in a specific geographically defined area, help

to determine where different types of economic activity will prosper. Regional economics analyze how space and location influence the economic behavior of agents (consumers, producers, government, and others), and how these behaviors affect their locational configuration that could shape the spatial patterns of economic development.

Regional economics is the study of regions focuses on the consideration of space, proximity, transportation cost, and location in production and consumption decisions, increasing returns to scale and externalities. Particularly, this field of economics highlights the spatial arrangements of households, firms, and capital in specific areas, externalities which arise from the proximity of households and land uses and the public policy issues which arise from the interplay of these economic forces.

In general, regional or "spatial" economics might be summed up in the question "What is where, and why—and so what?" (Hoover and Giarratani, 2020). The first what refers to every type of economic activity: not only production establishments in the narrow sense of factories, farms, and mines, but also other kinds of businesses, households, private and public institutions. Where refers to location in relation to other economic activity; it involves questions of vicinity, congregation, diffusion, and similarity or disparity of spatial patterns, and it can be argued either in broad terms, such as among regions, or micro geographically, in terms of zones, neighborhoods and sites. The why and the so what refer to interpretations within the somewhat elastic limits of the economist's faculty and venture.

Until fairly recently, traditional economists ignored the where question altogether, finding plenty of problems to occupy them without giving any spatial dimension to their analysis. Traditional geographers, though directly concerned with what is where, lacked any real technique of explanation in terms of human behavior and institutions to supply the why, and resorted to mere description and mapping. As time has moved forward, regional economies have surpassed national and state level strategies in terms of importance and impact. These public-private partnerships are playing a vital role in improving the product of their regions through advocacy, strategic planning and economic development projects.

Empirical literature review

Empirical studies on human capital and regional economics attempt to empirically test the determinants and the impacts of human capital on economic outcomes in specific regions. A long time ago, several studies tried to explain the factors that influence the formation of human capital. Forced migration, landownership, religion, culture, urbanization, climatic factors and geographic factors have significant effects on human capital formation, while in turn human capital formation affects economic growth of specific regions. The present thesis is related to the analysis of human capital and regional economics, as I examine in which way aspects of forced migration, landownership, urbanization, culture and geography determine the quantity and quality of human capital formation in a specific geographically defined area. In addition, I investigate how human capital formation affects economic development in terms of starting new businesses. Overall, this dissertation is related to the literature that examines the historical

determinants of human capital and provides evidence on the causal relationship between human capital and economic development.

This thesis can be divided in three district parts, each of which focuses on human capital and regional economics from a different perspective. In the first and second part, I shed light on the role of human capital during the early stages of the transition from stagnation to growth in Greece around 1900. I investigate how historical determinants across Greek regions shaped the disparities in terms of human capital. On the other hand, the third part aims to highlight the effects of human resources on the creation of new firms using recent European data at the metropolitan level.

The first chapter investigates historical determinants of human capital formation. In order to examine empirically these determinants, I have built an entirely novel data set with Greek historical data on literacy, landownership, geography, fertility and demography at the provincial and municipal levels based on the 1907 and 1928 Census. The theoretical foundation of this chapter is the seminal paper of Galor et al. (2009, GMV). Specifically, I examine the validity of GMV theory for the first time in a regional Greek sample of provinces and municipalities around 1900. Galor et al. (2009) provide that inequality in land ownership may be viewed as an obstacle to human capital accumulation and thus as a factor decelerating the process of industrialization and economic growth via US state-level evidence for the first decades of the twentieth century. The rising importance of human capital in production has made capitalists to support the expansion of public education in favor of workers (Galor and Moav, 2006). Landholders would benefit much 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. Erickson and Vollrath (2004) argue that high concentration of landownership across agricultural populations leads to low levels of education. Moreover, Easterly (2007) concludes that high inequality in landownership constitutes the most important drawback to welfare, high quality institutions and high schooling. A panel data analysis of Wegenast (2009) suggests that educational and political inequalities, especially in Latin America, are due to the historical influence of landlords on schooling. Another strand of this literature focus on human capital and the channels through which landownership and geographic factors influence it. For instance, Baten and Hippe (2017) 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. I combine the above strands of the empirical literature and pursue empirical analysis taking into account regional (department and prefecture) fixed effects because the regressions are likely to suffer from endogeneity in the form of omitted variable bias. Also, I use a treatment effect analysis (regression adjustment, nearest-neighbor matching and propensity-score matching) to address endogeneity. In line with several previous studies I find that unequal distribution of land has a substantial negative effect on human capital formation in Greek Regions. In addition, I provide that geographic factors, fertility and demographic factors have significant impacts on literacy in Greek provinces and municipalities.

The second part of this dissertation contributes to the empirical historical literature of human capital and regional economics, discussing about how Greece evolved after the Greco-Turkish war of 1919–1922, in the short run. This historical forced migration led to substantial territorial adjustments and large-scale population movements in the wider region. My study aims to uncover the role of a historical event, critical for the Greek history in many different ways and uncover its effects on human capital and employment. I build a unique Greek historical dataset, with data derived from the 1929 Census of the Hellenic Statistical Authority at the provincial level for the years 1920 and 1928. A huge part of the previous literature supports that forced migration consists a determinant of human capital formation and employment. After the mass flight and expulsion of millions of Germans during and after World War II, migrant children acquired more education than their native peers (Bauer et al., 2013). Bharadwaj et al. (2015) suggest that partition related flows in India resulted in an increase in literacy rates of natives. Similarly, Kreibaum (2016) supports that Ugandans living near refugee settlements are more likely to have access to primary schools run by NGOs or other private organizations. There is also evidence that descendants of displaced Poles invested more in education as a result of a shift in preferences towards human capital (Becker et al., 2020). Benos et al. (2021) indicate that refugees living in Greek settlements close to native farming settlements accumulate more education and as a result enter manufacturing more than the nearby natives. For the effect of forced migration on employment, Braun and Mahmoud (2014) find that an increase in the share of German expellees in the labor force decreased the employment rate of natives and Braun and Kvasnicka (2014) suggest that this inflow of German expellees have reduced low-productivity

agricultural employment and led to higher employment rates of the nonagricultural sector. In addition, Murard and Sakalli (2019) find that regions with a higher share of Greek refugees in 1923 display higher level of prosperity and industrialization because of the provision of new agricultural know-how and the transfer of technological knowledge in textile. Regarding empirical analysis, I exploit a quasi-natural experiment and focuses on the impact of refugee inflows on the total natives but also on the basis of gender. I use a diff-in-diff approach with the share of refugees in the total population as the treatment term. My results provide that refugee inflows, as a historical determinant of human capital and employment, increase native literacy population for both men and women. Concerning employment, I find that upon the arrival of refugees, the male native population turned away from agriculture to the industry, transportation, stock farming and trade sectors.

The last part of this thesis conducts a systematic examination of how human capital and other major factors integrates for higher regional enterprise births in 230 European metropolitan regions for the period 2008-2013. This study is the first among empirical studies which focuses on European metropolitan regions, while most literature focuses on specific countries and especially in the U.S. Metropolitan areas with larger, growing, denser markets as well as more foreign-born and higher labour turnover have higher enterprise births in all sectors (Behrens et al., 2014; Motoyama and Malizia, 2017). Recently literature focuses on enterprise births, its determinants, and its economic impacts (Decker et al., 2016; Gourio et al., 2016; OECD, 2017; Bijnens and Konings, 2018; Pugsley and Sahin, 2019). Regions with

higher levels of entrepreneurship and employment are characterized by better local governance, spend more on intellectual property rights and have a more educated workforce (OECD 2017). Patent applications and human resources have a substantial and long-lasting impact on enterprise births (Farre-Mensa et al, 2020). In addition, knowledge is a key source for enterprise births, particularly in innovative industries. As a result innovative enterprise births can be observed as manifestations of knowledge spillovers extant knowledge sources (Fritsch and Wyrwich, 2018; Farre-Mensa et al., 2020). Motoyama and Malizia (2017) conclude that tertiary education and high-tech industry concentrations are statistically related to high-tech business births. Eriksson and Rataj (2019) show that human capital is a significant and positive factor explaining the regional variation of new firms. I pursue empirical analysis taking into account country fixed effects because the regressions are likely to suffer from endogeneity. Besides cross-section and panel *OLS* fixed effects estimations I employ *GMM* estimates. To further strengthen my results, I also provide spatial estimations because it is widely recognized that sample data collected from geographically close entities are not independent but spatially correlated. I empirically demonstrate that total patent applications in combination with human capital factors (e.g. human resources in science and technology) stimulate enterprise births in all sectors.

Chapter 1: Geography, Land Ownership and Literacy: Historical Evidence from Greek Regions

1.1 Introduction

The role of geographical factors, institutional factors and human capital formation has been the center of a discussion about the origin of the differential timing of the transition from stagnation to growth and the important changes in income distribution which have taken place since the onset of the industrial revolution. There is a huge literature analyzing geography as a very important factor for comparative economic development (Baten and Hippe, 2017; Easterly, 2017; Adamopoulos and Restuccia, 2018). Also a number of papers support that inequality in the distribution of land has affected human capital formation and the transition from an agricultural to an industrial economy. . Galor et al. (2009, GMV) provide the theoretical foundation behind the fact that inequality in land ownership has adverse effect on human capital accumulation. During the transition from agricultural to industrial economy, major conflict arose between agricultural landholders and capitalists. Landholders benefit much 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. Return on land declines as wages of workers rise due to higher education that individuals obtain and educated workers have stronger incentives to migrate to industrial areas than less educated workers. Departure of workers from land is contrary to the interests of landholders. Landowners inhibit educational policies aimed to augment general education. Consequently, land ownership is

obstacle to human capital formation, thus factor slowing down industrialization and economic growth. Land ownership is a driving force behind divergence in income per capita with long-run implications even today. So, variations in the distribution of land ownership across countries generated variations in the industrial composition of the economies. This is because the transition to an industrial economy raises the importance of human capital in the production process, reflecting its complementarity with physical capital and technology. GMV confirm the hypothesis that inequality in land distribution has a negative correlation with human capital formation via US state-level evidence for the first decades of the twentieth century. Countries with high land inequality are overtaken in the procedure of industrialization. Education improves the productivity of industrial labour more than that of agricultural labour. In contrast, economies with low land inequality promote public education, benefit from industrialization and leach into a cursory process of development. In order to address causality they instrument landownership inequality through the interaction between global changes in the relative price of agricultural crops that are associated with economies of scale and variation in climatic characteristics across states. More recent studies focus on literacy rates and the channels through which land ownership, geographic factors and other determinants influence them. For instance, Baten and Hippe (2017) find that the distribution of land ownership is a mechanism behind the correlation of human capital and geographic factors using a European multi-country regional level.

This research aims to shed light on land ownership as driving force of the divergence in educational attainment and economic development across regions in

modern Greece. In line with GMV theoretical predictions, I show that inequality in land ownership has a negative correlation with human capital formation. Specifically, this paper examines the validity of GMV theory for the first time in a regional Greek sample of provinces and municipalities around 1900. This differs from most literature, which focuses on industrial economies because Greece was at the early stages of the transition to the industrial era during early 20th century. In order to examine empirically the above theoretical hypothesis I build an entirely novel data set with Greek historical data, regarding literacy rates(male, female, total, refugees) at the provincial and municipal levels based on the 1907 and 1928 Census, which are the first ones to include an extended educational attainment coverage at the provincial and municipal levels. Generally, I use more spatially disaggregated data (provincial and municipal levels) than the related literature. My dependent variables are literacy rates and the rest of the explanatory variables are similar to e.g. Baten and Hippe (2017). According to historical sources large landholdings in Greece were concentrated in areas where chiftliks (large manors farmed for the market) existed, while these sources focus on the different forms of land tenure in various regions of Greece. My first set of regressions uses land ownership as dependent variable and geographic variables as explanatory variables. In a second step, I study the correlation of literacy rates with geographic factors without including land ownership and finally I estimate regressions with literacy rates as dependent variables and land ownership, geographic factors and other controls as independent. To further strengthen my results (i.e. with respect to endogeneity), I also provide treatment effect estimations to address endogeneity. Moreover, I construct some graphs in order to sketch the idea that these historical data affect long-run

development. My research suggests that inequality in the distribution of land ownership adversely affected human capital formation, and thus the process of the transition from an agricultural to an industrial economy.

The rest of the chapter is structured as follows: in Section 2 I revisit the theoretical framework on literacy, land ownership and geography. In Section 3 I describe the data and the methodology used in this study. In Section 4 I present my empirical results. In Section 5 I present the effect of my data on the long-run development and Section 6 concludes.

1.2 Literature Review

There are quite a few studies which focus on human capital formation and its determinants. Recent work focuses on literacy rates and the channels through which land ownership, geographic factors and fertility influence them. As mentioned above, Galor et al. (2009) argues that land inequality has negatively affected human capital promoting educational institutions (e.g. public schooling, child labour regulations) in the United States, using variation in the distribution of land ownership and educational expenditure, across states and over time, for the period 1900-1940. Deininger and Squire (1998) use cross-country data and panel data for 103 countries in the period 1960-1990 and find a negative relationship between initial inequality in land distribution and long-run growth. Land ownership is negatively correlated with education, while high levels of education lead to more investment. Policies that promote investments and equal access of land are conducive to growth and poverty reduction.

Erickson and Vollrath (2004) use two new measures of land ownership that capture the level of landlessness within the countries. The first one is land inequality Gini and the second is the agricultural population divided by the total number of holdings. They do not find a significant relationship between land ownership, institutions and development but they support that high land ownership across agricultural populations leads to low levels of education. However, their results contend that land ownership does not impinge on the relationship between geographic endowments and income inequality, which is a very surprising inference because previous studies support the opposite.

Easterly (2007) confirms the idea that inequality affects development, schooling and institutions using cross-country data for developing countries with different agricultural endowments- specifically the land suitability for wheat to that suitability for sugarcane. High inequality in land ownership constitutes the most important drawback to welfare, high quality institutions and high schooling. In addition, high structural inequality obstructs the mechanisms which promote economic development and argues that this structural inequality is brought about by agrarian conditions.

A panel data analysis of Wegenast (2009) suggests that the agricultural production system in Asia and Latin America constitutes a major factor of different educational outcomes. Countries with higher agricultural plantations provide less broadly based educational policies in contrast to countries organized around family farming. The panel data analysis and OLS regressions conclude that exports of plantation crops, as proxy for the political strength of the agrarian elite, reduce

secondary education attainment levels and government's investments in secondary schooling but these same exports are associated with higher tertiary education levels. Other agricultural products such as cereals, animals or forestry do not affect the different types of educational categories. These results explain that educational and political inequalities, especially in Latin America, are due to the historical influence of landlords on schooling. Low plantation crops can lead to harmful consequences for social equality.

Baten and Hippe (2017) examine the relationship between geographic factors and literacy in more than 300 European regions in the 19th century. The indicator of numeracy is the ABCC index which is a proxy for basic numerical skills. The main geographic variables are: soil suitability (cereal, pasture, potato, sugar) temperature, precipitation, land size, altitude and ruggedness. They argue that human capital is negatively correlated with land ownership, which is also related to geographic factors. Specifically, capitalists benefit more from an increase in human capital of their workers than landholders. For this reason landowners do not promote policies that promote education. Consequently, land ownership plays an important role in human capital formation, industrialization and economic growth.

Moreover, Adamopoulos and Restuccia(2018) argue that low agricultural productivity in poor countries is not due to poor land endowments. They examine the role of geography and land quality for differences in agricultural productivity using micro-geography data for 162 countries from Global Agro-Ecological Zones (GAEZ) and spatial analysis and try to answer if these differences exist because of geography or economic choices. If all farmers worldwide use the same productivity

procedures then land quality is not a restriction for poor countries but the issue is what crops are produced, where they are produced within the country and how efficiently they are produced in each cell of land.

Some papers analyze this negative relationship for specific countries. Specifically, Beltrán Tapia and Martínez-Galarraga (2015) confirm the negative relationship between the fraction of farm laborers and literacy rates (female and male) by using a large historical Spanish dataset in mid-19th century, i.e. before industrialization. They examine if demand or supply factors are responsible for this negative relationship and find that in places where there are landowning elites that obstruct education development, demand factors are also crucial. For societies with high land ownership, the economic situation of rural poor is uncertain and there are only few investments on education because of the budget constraints of large segments of the population. In rural areas large landowners obstruct public support of education but in urban areas this does not happen because of the existence of other elites. Moreover, more equal societies are characterized by better educational attainments of girls and women. They conclude that high land ownership obstructs human capital accumulation. The example of Prussia is examined by Cinnirella and Hornung (2016), using a unique country-level census database. They examine the negative relationship between landownership concentration and primary education taking into account agricultural features, demand and supply factors in 19th century. In order to deal with endogeneity, they use an instrumental variable (exogenous variation in farm size because of differences in soil texture), country and time fixed effects, and confirm the negative effect of landownership concentration on

education and suggest that the effect is indeed causal and weakens over time. Regions with poorer soil quality have a lower population density, a weaker demand for land and are characterized by higher average farm sizes.

Other historical studies yield insignificant and sometimes contradictory results. Goni (2018) supports the same idea that high landownership concentration has a significant negative effect on education by using a dataset on 1387 School Boards and 32 counties in the period 1871-1899 in England. This effect is significant only for changes that began after the Industrial Revolution in England. High land ownership reduces the ratio of state to private schools, the number and salaries of teachers and the facilities per pupil and as a consequence the attribution of children declines. He argues that the relationship between land ownership and state education is causal and uses two instrumental variables: soil texture and the massive redistribution of land after the Norman conquest of England in 1066. The estimations with these instrumental variables provide two different results: first-stage estimations provide a strong persistence in land ownership over eight centuries, from 1066 to the late-nineteenth century and second-stage estimations support that the effect of landownership concentration on state education is causal. The consequences of the negative effects of land ownership are confined in areas that landowners are political figures that promote land elites. In addition, he observes a positive relationship between people that work in manufacturing and state education. On the other hand, private education is not influenced by landownership concentration.

Overall, some studies support the significant impact of land ownership, geography and other factors on human capital formation whereas some others do not. However, there is no a study using Greek data, which investigates the historical determinants of human capital formation taking into account the effect of land ownership in this process. I am based on the literatures of human capital formation, geography and land ownership, by studying two cross-sections of all provinces and municipalities in Greece in 1907 and 1928, i.e. I use the earliest available data on educational attainment before the establishment of compulsory primary education in Greece. Finally, I construct data which allow me to examine the impact of land ownership on educational outcomes before the implementation of a comprehensive land reform which started in the mid-1920s and redistributed large land holdings to landless peasants.

1.3 Data and Methodology

1.3.1 Data

The empirical analysis aims to test the hypothesis that land ownership and geographic factors have an impact on human capital formation in a cross-section of Greek provinces and municipalities around 1900. I examine the validity of GMV theory (Galor et al., 2009) in a regional Greek sample for the first time. This differs from much of the literature, because Greece was at the early stages of the transition to the industrial era during the period studied. I build an entirely novel data set with Greek historical data, regarding literacy rates (the dependent variable) at the provincial and municipal levels based on the Censi of 1907 and 1928 of the Hellenic Statistical Authority. The first sample consists of 69 provinces in 1907 and 142

provinces in 1928 and the second sample consists of 445 municipalities in 1907. I use these two periods because they are the first ones including an extended coverage at the provincial and municipal levels around 1900 (Galor et al., 2009; Cinnirella and Hornung, 2016; use also Census data for literacy rates). I use male, female, total and refugee literacy rates. Refugee literacy rates are only for year 1928, since refugees had not come to Greece in 1907. I use refugees to control for possible human capital externalities from the refugees to the natives and make sure that the estimated effect of land inequality and geography on literacy rates does not include the impact of refugee literacy. Moreover, I construct maps that illustrate the geographical distribution of literacy rates in Greek provinces in 1907 and 1928. I combine the data of literacy rates from the Hellenic Statistical Authority with the coordinates of Greek provinces which are available in Geodata, which is a Greek geospatial database.

The first three maps illustrate male, female and total literacy rates in Greek provinces in 1907.

(Insert here maps 1, 2 and 3)

From these maps it is evident that province “Attica” in Central Greece has the highest literacy rates (male, female, total), which is expected because “Attica” contains the capital city Athens. Province “Valtou” in Central Greece has the lowest literacy rates (male, female, total). In addition, Thessaly is characterized by low literacy rates and Peloponnese is characterized by high literacy rates.

The next three maps present male, female and total literacy rates in Greek provinces in 1928 for native people (refugees not included). As expected, province “Attica” has again one of the highest literacy rates (male, female, total). The lowest

literacy rates are detected in Western Thrace and Epirus, and the rest highest literacy rates are detected in Peloponnese, Cyclades and Ionian Islands regions.

(Insert here maps 4, 5 and 6)

Furthermore, I create maps for Greek provinces in 1928 including the total population (native and refugees) and maps for the distribution of literacy rates of refugees (male, female, total) in 1928.

For total literacy rates (refugees included) in 1928, "Attica" in Central Greece has again one of the highest literacy rates. Territories with high literacy rates are Peloponnese, Ionian Islands and Cyclades and regions with low literacy rates are Western Thrace and Epirus.

(Insert here maps 7, 8 and 9)

The distribution of refugee literacy rates is very different from the distribution of native literacy rates. High literacy rates are detected in Crete and Cyclades.

(Insert here maps 10, 11 and 12)

The main control variable of interest is land ownership which captures the part of land owned by large landholders for the whole Greece based on relevant historical sources. The assumption is that land ownership is one of the main driving forces of the divergence in terms of human capital accumulation between Greek regions. This in turn is expected to have serious long-run implications for the spatial distribution of income per capita across regions in modern Greece. Galor et al. (2009) and Baten and Hippe (2017) use quantitative Census data for land ownership. Baten and Hippe (2017) calculate the share of large holdings by dividing the total

area of holdings larger than 50 ha by the total area of all holdings. For Greece this is impossible because only few quantitative data exist but at higher regional levels (i.e. territory level) and it is not sure if these data are valid, because there is no a credible definition of large land ownership in 1907 and 1928. As a result I have studied economic history sources in order to construct a qualitative variable that measures land ownership. According to these sources (Vergopoulos, 1975; MCGrew, 1987; Petmezas, 2003; Petmezas and Kostis, 2006; Kontogiorgi, 2006; Petmezas, 2012) large landholdings in Greece were concentrated in areas where *chiftliks (*large manors farmed for the market) existed, while these sources focus on the different forms of land tenure in different regions of Greece. *Chiftliks* first appeared in 16th century and became widespread in 18th century because of the weakening of the central political authority of the Ottoman Empire and the large expansion in international trade. According to Ottoman law there is no private ownership and theoretically all regions of the Empire belong to the Sultan. The *chiftlik* system is associated with the development of local administrators, beys and aghas. In 1858 by Land Code aghas and beys had the right to full ownership of estates which they had been cultivating for ten years. By the end of the nineteenth century the land tenure system divided the land into two categories: the *chiftliks*, the large freehold estates, and the head villages, Christian villages which were under the supervision of the central government and not of a local bey or agha. *Chiftliks* consisting of twenty to thirty families of tenant farmers—rarely up to fifty—were located in the fertile lowlands close to main roads and were conferred as the estates of a landlord. Head villages were mainly mountain villages or at best situated in the foothills, where there were better hygienic conditions and cultivation of land was easier. They are

characterized by local autonomy, since the major intervention of the state in village life was in the field of tax-farming. Another important social group was that of the transhumant shepherds. They had lots of sheep and goats, spent the summer on the highland pastures on the mountain ranges and migrated to the valleys, where they spent the winter on rented pastures. They offered more to the existing money economy than the farmers. The *chiftlik* system developed as a result of attempts to commercialize Ottoman agriculture and deal with the demands from Europe for products such as wheat, cotton and wool. *Chiftlik* was a combination of agriculture and pasture and not only a large agricultural venture. The landowner ensured his annual income by the combination of winter grains (wheat, barley, etc.), spring crops and pasture land rented by shepherds. During and after the Balkan Wars there is a defective transition from the Ottoman system of land rights to one based on private ownership principles. The *chiftliks* started getting under the control of the Greek state, a state that depended on guarantees. On the grounds that the majority of large estates had remained in the hands of Muslim landlords, Greek government was afraid of expropriating this land because of the diplomatic complications with Turkey. During the first decade of Greek administration, the position of the chiftlik planters impaired roughly. In addition after liberation, the state also rented the abandoned *chiftliks* of Muslim landlords who had migrated. Regional differences in land ownership can generally be viewed as very stable over time according to these sources until approximately 1922, i.e. the arrival of mass refugee inflows. This is why mass expropriations of large landholdings started soon afterwards in order to distribute this land to the refugees. In this study land ownership is a dummy variable taking values 0 and 1, where 0 corresponds to regions with low share of large land

holdings and 1 corresponds to regions with high share of large land holdings. According to the theoretical framework (see literature review) I expect the sign of this variable to be negative. In this respect, I construct a map for the geographical distribution of land ownership in Greek provinces referring to the period before the expropriation of large properties by the state.

(Insert here map 13)

Provinces with high share of large land holdings are found in Central Greece/Euboea (e.g. provinces of “Attikis, “Thivon”, “Istiaias”), Thessaly, Epirus, Macedonia and in province “Didymoteichou” in Western Thrace. Provinces characterized by low share of large land holdings are mainly located in Peloponnese, Crete, Ionian Islands, Cyclades, Aegean Islands and Western Thrace.

I enrich the estimated specifications with control variables consistent with the literature (e.g. León, 2004; Clark and Gray, 2014; Baten and Hippe, 2017; Adamopoulos and Restuccia, 2018; Goñi, 2018).

First, I use geographic factors as the channel behind the correlation of land ownership with literacy rates. The geographic factors are soil yield (wheat, potato, sugar, olive, rice, and pasture), temperature, altitude, ruggedness and precipitation. Geographic factors are very stable over time. I use soil yield factors instead of soil suitability factors, because the latter are highly correlated with each other and I cannot use them all together in my estimates. Altitude (the elevation above sea level in meters) is collected from the Census of 1951 from Hellenic Statistical Authority and ruggedness is the standard deviation of altitude. The remaining eight geographic variables are obtained using GIS techniques. Specifically, I have downloaded Greek

maps that contain the data on each variable from the database IPUMS Terra (Integrated Population and Environmental Data). Then I have taken the coordinates of Greek municipalities from Geodata, a Greek geospatial database and combined maps and coordinates using the QGIS software (open source geographic information system) in order to extract the geographic data at municipal level. Finally, in order to collect these geographic data at the provincial level I have found the municipalities belonging to every province and taken the median of the data of these municipalities. Following the relevant literature (Wegenast, 2009; Beltrán Tapia and Martínez-Galarraga, 2015; Baten and Hippe, 2017) I expect the sign of the variables to be negative for some factors and positive for others.

Another strand of the literature studies the relationship between fertility and education. The basic theoretical idea is that people face a trade-off between quantity and quality of children in terms of human capital when they decide about the number of children. Perotti (1996) shows that more equal societies are characterized by lower fertility rates and higher educational investments and consequently higher growth rates. But he does not find that these societies with more democratic features grow faster because of low redistribution. A higher female educational attainment is associated with a higher fertility rate. Moreover, secondary school enrollment is negatively correlated with fertility and positively with growth. Except from schooling, fertility rates are also influenced by cultural factors. Cygan-Rehm and Maeder (2012) import the same idea that education is negatively correlated with fertility rates over the life cycle under inflexible labor market conditions (e.g. high wage penalties for motherhood, very limited supply of public childcare, high opportunity costs of childrearing, policies that encourage mothers to

stay out of the labor market or to work part time). They examine three different outcomes: the number of children, the probability of remaining childless, and the timing of births. They use two complementary German datasets: the German Microzensus and the German Socio-Economic Panel, to investigate fertility of West German women born between 1937 and 1961. Education reduces the number of children and increases the probability of never having children. Specifically, one additional year of schooling reduces the number of children by more than 0.1, and increases the probability of childlessness by about 2-5 percentage points. The following studies focus on specific countries. Regarding developed countries, León (2004) uses U.S. Census data in the period 1950-1990 and finds that women who attend more years of schooling have on average one less child than they would have otherwise. Increases on wages lead to higher opportunity cost of having children and spending a lot of time in work. In order to generate exogenous variation in individual schooling choices and estimate the effect of education on fertility he uses changes in state compulsory attendance and child labor laws over time. Education increases the probability that a woman do not have children in her life, a point that confirms the shocking demographic changes in the last few decades in many countries. IV estimations suggest a stronger negative relationship between education and fertility than OLS regressions because of the presence of measurement error in schooling.

Hoem et al. (2006) examine the association between education (field and level) and fertility of Swedish women for sixty educational groups. There is a negative relationship between fertility and educational level, but this result is impressive when they focus more on the field of education, e.g. women who have

teaching or healthcare occupations have higher fertility rates than women who deal with esthetic, administrative or humanist professions.

Concerning developing countries, Axinn (1993) uses data from a rural community in Nepal and concludes that children's schooling has a significant impact on fertility preferences but the influence of parental schooling on fertility is weaker, and also inconsistent in different models. These results support theories that associate mass education to the onset of fertility reduction through children's schooling experience. According to Lam and Duryea (1999) significant negative effects exist between women's schooling and fertility rates over the first eight years of schooling in Brazil. Also, they find strong effects of parental schooling on children's schooling. The rapid fertility decline in Brazil is explained by the increases in male and female schooling. A four years increase in wife's schooling attendance drives to a decline of almost one birth in fertility up to age 30. Their results indicate that there is a weak link between declines in fertility and increases in women's labor force participation at low levels of schooling. Despite this weak relationship, significant positive effects of schooling on measures of child quality exist. In addition, increases in schooling at low levels give the opportunity to parents to have healthier and better educated children by increasing investments in child quality and reducing the number of children.

Roach (2009) deals with the relationship between fertility levels and children educational attainment levels in Uganda controlling for other factors that influence education, e.g. parent's educational attainment, region, age. The data are collected from two databases: Demographic and Health Surveys (DHS) and Integrated Public

Use Microdata Series International (IPUMS-I) for the years 2000-2001 and 2002, respectively. Although the results are statistically significant for both datasets, they have opposite signs and are not quantitatively important. In the long run, the decrease on educational attainment because of high fertility levels is not as significant as we expect and especially after controlling for these factors, the effect of fertility is almost zero.

Günes (2013) follows León (2004) and uses the compulsory schooling law (CSL) in Turkey to show that an extra year of female schooling reduces teenage fertility by 0.03 births, a reduction of 33%. This result emerges from the IV estimations and OLS regressions. For the estimations he uses data from Turkish Demographic and Health Survey in 2008 (TDHS-2008). The TDHS survey is a demographic health survey conducted every five years by Hacettepe University Institute of Population Studies (HUIPS) since 1993. Some of the basic variables at provincial level are: male and female education, number of teachers, number of schools, different age groups. This negative relationship is confirmed in provinces with higher initial fertility and lower population density. Furthermore, the effect of the CSL on female education and fertility depends on various characteristics: initial levels of fertility and education, income, urbanization, population density and agricultural activity.

According to these studies there is also a relationship between fertility rates and education. The main idea is that more equal societies are characterized by lower fertility rates and higher educational investments. I use male, female and total birth rates as control variables. According to this research, I expect a negative relationship

between birth rates and education. These birth rates are obtained from the Census of 1884 of the Hellenic Statistical Authority. Birth rates are the quotients of the division between number of births and population. I use the year 1884 for births in order to have people who are already educated or not in 1907 and 1928, for which I use literacy rates.

Finally, I use four additional control variables. These are taken from the Census of the Hellenic Statistical Authority. The first one is the land area for Greek provinces which is the same in both 1907 and 1928. The second is the population density which is calculated by the division of population with land area. The third is the urbanization rate, a proxy for regions with large population to control for urbanization effects on literacy (Baten and Hippe, 2017 use also this variable). Specifically, it is a ratio calculated by the division of five thousands with the population in 1907 and by division of ten thousands with the population in 1928[1]. The last control variable is the ratio of Bulgarians/Turks in 1920 which is the division of the number of Bulgarians/Turks with the population in 1920. I use this variable only for the estimations in 1928.

For Greek municipalities I use data only for year 1907 because there are no available data at municipal level in the Census of 1928. Moreover, for Greek municipalities I have no data for altitude, ruggedness and land area which are available only for Greek provinces.

Table 1 presents the description of the data, the period and the data sources.

[1]These are the thresholds set by the Hellenic Statistical Authority in the 2 Census to define that a settlement constitutes an urban center.

(Insert here Table 1)

In addition, summary statistics of the variables of Greek provinces in 1907 and 1928 are reported in Table 2 and Table 3 presents summary statistics for Greek municipalities in 1907. The literacy rates take values between zero and one because they are ratios.

(Insert here Table 2 and Table 3)

Tables 4 and 5 provide the correlations between geographic factors in Greek provinces and municipalities respectively. Variables with high correlation, such as wheat and sugar yield, are not included together in the estimations to avoid multicollinearity.

(Insert here Table 4 and Table 5)

1.3.2 Methodology

I pursue empirical analysis taking into account regional (department and prefecture) fixed effects because the regressions are likely to suffer from endogeneity in the form of omitted variable bias. If there are omitted variables, and these variables are correlated with the variables in the model, then fixed effects models provide a means for controlling for this bias. The idea is that whatever effects the omitted variables have on the subject at one time, they will also have the same effect at a later time; hence their effects will be constant or fixed over time. In addition, I allow for clustered standard errors at the regional level (department and prefecture levels) to account for spatial error dependence. For example, clustered standard errors at the department levels assume that errors of the spatial units are correlated within departments, but independent between departments.

I also use treatment effects (regression adjustment, nearest-neighbor matching and propensity-score matching) to address endogeneity. I use the average treatment effects (ATEs) to estimate treatment effects from observed data and the potential-outcome means (POMs) to estimate the distribution of individual-level treatment effects. A potential-outcome model specifies the potential outcomes that each individual would obtain under each treatment level, the treatment assignment process, and the dependence of the potential outcomes on the treatment assignment process. The term treatment effect is defined as the average causal effect of a variable (land ownership) on an outcome variable (literacy rates). A treatment effect is the change in an outcome caused by a subject, often an individual, getting one treatment instead of another. The defining characteristic of observational data is that treatment status is not randomized and it is not possible to observe a specific subject having received the treatment and having not received the treatment. Moreover, that implies that the outcome and treatment are not necessarily independent. A classic solution to this problem is to randomize the treatment. High costs or ethical issues rule out this solution in many observational datasets. The treatment effect estimations allow me to estimate the efficacy of treatments using observational data. All the estimators require some form of the following three assumptions. The independent and identically distributed (i.i.d.) sampling assumption, ensures that the outcome and treatment status of each individual are unrelated to the outcome and treatment status of all the other individuals in the population. The conditional-independence (CI) assumption means once we control for all observable variables, the potential outcomes are independent of treatment assignment. The third assumption is the overlap

assumption. This assumption ensures that each individual could receive any treatment level and more formally this assumption states that each individual have a positive probability of receiving treatment.

The regression adjustment (RA) method extends the idea of using sample means to estimate treatment effects by using a regression model to predict potential outcomes adjusted for covariates. Regression adjustment fits separate regressions for each treatment level and uses averages of the predicted outcomes over all the data to estimate the POMs. The estimated ATEs are differences in the estimated POMs.

Matching estimators use an average of the outcomes of the nearest individuals to impute the missing potential outcome for each sampled individual. Matching estimators are based on the idea of comparing the outcomes of subjects that are as similar as possible with the sole exception of their treatment status. Nearest-neighbor matching (NNM) is accomplished by calculating the "distance" between pairs of observations with regard to a set of covariates and then "matching" each subject to comparable observations that are closest to it. NNM is nonparametric in that no explicit functional form for either the outcome model or the treatment model is specified. This flexibility comes at a price; the estimator needs more data to get to the true value than an estimator that imposes a functional form. More formally, the NNM estimator converges to the true value at a rate slower than the parametric rate, which is the square root of the sample size, when matching on more than one continuous covariate.

Propensity-score matching (PSM) is an alternative to NNM. PSM does not require bias correction, because it uses a model for the treatment. Instead of

performing bias correction to handle the case of more than one continuous covariate, a common solution is to combine all the covariate information into estimated treatment probabilities, known as propensity scores, and use this single continuous covariate as the matching variable. In effect, the PSM estimator parameterizes the bias-correction term in the treatment probability model. One advantage of matching on the estimated treatment probabilities over the bias-correction method is that one can explore the fit of different treatment probability models using standard methods before performing the nonparametric matching. For my PSM estimations, I test for the overlap assumption and it holds.

1.4 Empirical Results

To better understand the relationship between land ownership, geographic factors and literacy rates, I estimate the following regression model:

$$LT_{itg} = \beta_0 + \beta_1 landown + \beta_2 popdens + \beta_3 land + \beta_4 urban + \beta_5 geography + \beta_6 birth + \mu_i + \varepsilon_i, \quad (1)$$

where *i* denotes the specific region, *t* denotes the period, *g* denotes the gender, *LT* denotes the different literacy rates, *landown* is a dummy variable for land ownership, *popdens* is the population density, *land* is the land area, *urban* is the urbanization rate, *geography* denotes the different geographic factors, *birth* denotes the birth rates, μ_i represents regional fixed effects and ε comprises the non-observed influences on the *LT*.

I start with fixed effect estimation techniques. Table 6 provides fixed effect estimations of land ownership on geographic factors (without literacy rates) for Greek provinces in 1907 and 1928. I use fixed effects at department level and cluster standard errors at department level (Columns 1, 3) and prefecture level (Columns 2,

4). Land ownership is highly correlated with many geographic factors. For example, rice yield, wheat yield and land area have positive and statistically significant relationships with land ownership. Potato yield and precipitation are negatively correlated with land ownership.

(Insert here Table 6)

Table 7 presents fixed effect estimations of literacy rates in Greek provinces in 1907. I use geographic factors and other control variables as independent variables. Consequently, land ownership is not included in control variables. I use fixed effects at department level and cluster standard errors at department level (Columns 1, 2, 3) and prefecture level (Columns 4, 5, 6). Population density and urbanization have positive and statistically significant signs. Precipitation, pasture yield and birth rates are negatively correlated with literacy rates.

(Insert here Table 7)

Table 8 presents fixed effect estimations of native literacy rates in Greek provinces in 1928. I use geographic factors and other control variables as independent variables without land ownership. I use fixed effects at department level and cluster standard errors at department level (Columns 1, 2, 3) and prefecture level (Columns 4, 5, 6). Population density and urbanization have positive and statistically significant signs. Pasture yield, precipitation, birth rates and Bulgarians/Turks are negatively correlated with literacy rates.

(Insert here Table 8)

The next two Tables provide fixed effects estimations for Greek provinces in 1907, including land ownership as independent variable to investigate if land

ownership is one factor behind the correlation of geography and literacy. This is consistent with a large number of papers (see for example Easterly, 2007; Galor et al., 2009; Cinnirella and Hornung, 2016; Baten and Hippe, 2017), which find a negative association between land ownership and literacy rates. I present fixed effect estimations because the main independent variable (land ownership) is a dummy variable. Table 9 presents fixed effect estimations of literacy rates on land ownership for Greek provinces in 1907. I use fixed effects at department level and cluster standard errors at department level (Columns 1, 2, 3) and prefecture level (Columns 4, 5, 6). In Table 10, I use the same estimations but fixed effects are at prefecture level and cluster standard errors at department level (Columns 1, 2, 3) and prefecture level (Columns 4, 5, 6). The results are almost the same for these two Tables except from geographic factors. Land ownership is negatively correlated with all literacy rates (at 1% and 5% levels of significance) in 1907. So, the presence of large land holdings in a province reduces female literacy on average by 8.8% (Column 2, Table 9). Pop density and urbanization are strongly and positively correlated with literacy rates. Birth rates have negative relationships with literacy rates. In Table 9 pasture yield and precipitation have negative and statistically significant signs.

(Insert here Tables 9 and 10)

Table 11 provides fixed effect estimations of literacy rates on land ownership for Greek provinces in 1928. I use fixed effects at department level and cluster standard errors at department level (Columns 1, 2, 3) and prefecture level (Columns 4, 5, 6). Table 12 presents again the same estimations but I use fixed effects at prefecture level and cluster standard errors at department level (Columns 1, 2, 3) and prefecture level (Columns 4, 5, 6). The results are almost the same for these two

Tables. Land ownership is negatively correlated with literacy rates. Pop density and urbanization are strongly and positively correlated with literacy rates. Pasture yield and precipitation have strong negative relationships with literacy rates.

(Insert here Tables 11 and 12)

The majority of the related studies start empirical analysis with OLS estimations in order to interpret the negative relationship between land ownership and literacy rates (Easterly, 2007; Wegenast, 2009; Cinnirella and Hornung, 2016; Baten and Hippe, 2017). The next four Tables provide OLS estimations of male, female and total literacy rates of Greek provinces in 1907 and 1928. Table 13 (fixed effects at department level) and Table 14 (fixed effects at prefecture level) present the same OLS results in Greek provinces in 1907. Table 15 (fixed effects at department level) and Table 16 (fixed effects at prefecture level) present OLS estimations in Greek provinces in 1928. Cluster standard errors are at department level (Columns 1, 2, 3) and prefecture level (Columns 4, 5, 6). Regions with higher land ownership tend to have lower male, female and total literacy rates in 1907 and 1928, i.e. the presence of land ownership results in a 0.069% decrease in male literacy rates (Column 4, Table 13). Population density and urbanization continue to have the same signs (positive) and statistical significance. Precipitation and pasture yield have significant negative impacts on literacy rates. Birth rates are negatively correlated with literacy rates. In Table 16, the impact of Bulgarians/Turks on all literacy rates is negative and statistically significant (at the 5% and 10% levels of significance). The results for birth rates are in line with most of the previous literature (Perotti, 1996; Hoem et al., 2006; Cygan-Rehm and Maeder, 2012). The result that geographic factors have an impact on literacy rates is supported by the

related literature but the signs of the coefficients are different. For example, Baten and Hippe (2017) support that pasture yield has a positive impact on literacy and population density has a negative impact on literacy. The signs of the coefficients of these variables are in contrast with my results. The inference that I can draw from these estimations is that the signs of the main variables are the same regardless of the estimation technique.

(Insert here Tables 13, 14, 15 and 16)

For Greek municipalities in 1907 I employ use the same estimation analysis. In Table 17 I present fixed effect estimations of land ownership on geographic factors (without literacy rates) in Greek municipalities in 1907. Fixed effects are at department level. Cluster standard errors are at department level (Column 1) and prefecture level (Column 2). Population has statistically significant relationship with land ownership at 1% level of significance. Moreover, wheat yield and olive yield are positively correlated with land ownership.

(Insert here Table 17)

Table 18 presents fixed effect estimations of literacy rates in Greek municipalities in 1907. I use geographic factors and other control variables as independent variables without land ownership. I use fixed effects at department level and cluster standard errors at department level (Columns 1, 2, 3) and prefecture level (Columns 4, 5, 6). Population and urbanization have again strong positive relationship with literacy rates but birth rates have no statistically significant coefficients. Pasture yield has strong negative relationship with male and total literacy rates. Precipitation has strong negative impact on female literacy rates.

(Insert here Table 18)

Table 19 and Table 20 illustrate fixed effect regressions with male, female and total literacy rates as dependent variables, including land ownership as independent variable in Greek municipalities in 1907. I use fixed effects at department level (Table 19) and at prefecture level (Table 20). Cluster standard errors are at department level (Columns 1, 2, 3) and prefecture level (Columns 4, 5, 6). Land ownership is negatively correlated with male and total literacy rates but only in estimations with prefecture fixed effects (Table 20). Pasture yield and precipitation have negative and statistically significant coefficients. Population and urbanization have strong positive signs. Olive yield is positively correlated with male literacy rates at the 10% level of significance.

(Insert here Tables 19 and 20)

The last two estimations at municipal level are the ordinary OLS regression. In Table 21 fixed effects are at department level and in Table 22 are at prefecture level. Cluster standard errors are at department level (Columns 1, 2, 3) and prefecture level (Columns 4, 5, 6) In these Tables there are the same variables with important effects on literacy rates (land ownership, population, urbanization, pasture yield, precipitation) as in the previous Tables. The main results are the same for all the above Tables. Specifically, land ownership, pasture yield, precipitation and birth rates are negatively correlated with literacy rates. On the other hand, population density and urbanization are positively correlated with literacy rates.

(Insert here Tables 21 and 22)

For Greek provinces in 1928 except from the native literacy rates I have also data for total literacy rates (native and refugee) and only for refugee. To this end I include in my analysis some Tables with estimations for total literacy rates (native and refugee). For these estimations I have total literacy rate as dependent variable and also use refugee literacy rate as independent variable. For land ownership, population density and urbanization the signs and the statistical significance are the same as the estimations only with native literacy rates but not so strong. Table 23 illustrates fixed effect regressions with male, female and total literacy rates as dependent variables in Greek provinces in 1928 but without land ownership. Fixed effects are at department level and cluster standard errors are at department level (Columns 1, 2, 3) and prefecture level (Columns 4, 5, 6). Table 24 (fixed effects at department level) and Table 25 (fixed effects at prefecture level) provide OLS estimations with land ownership as independent variable. Cluster standard errors are at department level (Columns 1, 2, 3) and prefecture level (Columns 4, 5, 6). Land ownership, pasture yield and precipitation have statistically significant negative relationship with literacy rates. Pop density and urbanization have positive and statistically significant signs.

(Insert here Tables 23, 24 and 25)

Next, I use treatment effects (regression adjustment, nearest-neighbor matching and propensity-score matching) to address endogeneity. In these estimations the effect of geographic factors on literacy comes through land ownership. Table 26 provides the regression adjustment model of male, female and total literacy rates in Greek provinces in 1907. I use the average treatment effect (ATE) and the potential-outcome means with robust standard errors. The dependent

outcome variable is literacy rate and the independent outcome variables are population density, urbanization rate and birth rates. The treatment variable is land ownership. The results for land ownership are not statistically significant maybe because of low degrees of freedom.

(Insert here Table 26)

Table 27 presents the regression adjustment model of native male, female and total literacy rates in Greek provinces in 1928. I use the average treatment effect (ATE) and the potential-outcome means with robust standard errors. For all these estimations the dependent outcome variables are male, female and total literacy rates and the independent outcome variables are population density, urbanization rate, birth rates and the rate of Bulgarians/Turks. The treatment variable is land ownership. Land ownership has negative statistically significant relationships with female and total literacy rates in 1928. For example, the female literacy is 34% with low share of large landholdings and the strong presence of large land holdings reduce female literacy by 5.8%.

(Insert here Table 27)

Table 28 provides the nearest-neighbor matching model with the average treatment effect for native male, female and total literacy rates for Greek provinces in 1928. I use robust standard errors and bias adjustment. For these estimations the dependent outcome variables are male, female and total literacy rates and the independent outcome variables are population density, urbanization rate, birth rates and the rate of Bulgarians/Turks. The treatment variable is land ownership. Landownership is negatively correlated with male, female and total literacy rates.

These negative correlations are also statistically significant at the 1% level of significance for female and total population and at the 5% level of significance for male population. So, the presence of large land holdings in a province reduces total literacy on average by 5.8%.

(Insert here Table 28)

The next three Tables provide the treatment effects (regression adjustment, nearest-neighbor matching and propensity-score matching) of Greek municipalities in 1907. Table 29 presents the regression adjustment model of male, female and total literacy rates in Greek municipalities in 1907. I use the average treatment effect (ATE) and potential-outcome means with robust standard errors. For all these estimations the dependent outcome variables are male, female and total literacy rates and the independent outcome variables are birth rates. The treatment variable is land ownership. Land ownership has negative statistically significant relationships with all literacy rates for Greek municipalities in 1907. The male literacy rate is 47% with low share of large landholdings and the strong presence of large land holdings reduce male literacy by 4.6%. The female literacy rate is 13.7% with low share of large landholdings and the strong presence of large land holdings reduce female literacy by 4.9%. The total literacy is 30% with low share of large landholdings and the strong presence of large land holdings reduce female literacy by 3.5%.

(Insert here Table 29)

Table 30 provides the nearest-neighbor matching model with the average treatment effect for male, female and total literacy rates for Greek municipalities in 1907. I use robust standard errors and bias adjustment. For these estimations the

dependent outcome variables are male, female and total literacy rates and the independent outcome variables are population, urbanization rate and birth rates. The treatment variable is land ownership. Land ownership is negatively correlated with male, female and total literacy rates. These negative correlations are also statistically significant at the 1% levels of significance for all literacy rates.

(Insert here Table 30)

The last Table presents the propensity-score matching model with the average treatment effect for male, female and total literacy rates for Greek municipalities in 1907 with robust standard errors. The outcome variables are male, female and total literacy rates. The dependent treatment variable is land ownership and the independent treatment variables are geographic factors. Land ownership is negatively correlated with male and total literacy rates at the 5% level of significance and at the 1% level of significance for female literacy rates. The presence of large land holdings in a municipality reduces male literacy rate on average by 2.9% and female literacy rate by 5.5%.

(Insert here Table 31)

1.5 Long-Run Development

The last few decades there is a motivation of researching the evolutionary roots of comparative economic development across regions and countries. Specifically, many researchers focus on the influences of human evolution and the composition of human traits on comparative economic development across societies. Moreover, there is a literature (Nunn, 2009; Caicedo, 2018) that examines the long-term consequences of a historical human capital intervention. Following

this literature, I provide an informal inquiry on whether historical literacy rates in 1907 and 1928 still have an impact on educational attainment (approximately sixty years later (1981) and economic development about eighty years later (2000)). I use 1981 Census data for male, female and total secondary and tertiary education attainment for provinces which correspond to provinces in 1907 and 1928. Regarding, economic development I use 2000 Census data for GDP per capita. I use 1981 and 2000 Census data because they are the latest available data at provincial level. Using scatter plots I examine if historical human capital formation influences long-run outcomes. The first six figures illustrate male, female and total secondary education attainment versus male, female and total literacy rates in 1907 and 1928. The lines are linear trends and depict the positive influence of literacy rates in secondary education attainment. Correlations between literacy in 1907 and secondary education attainment in 1981 for provinces are 0.53, 0.51 and 0.58 for males, females and total population respectively and are statistically significant at the 1% level. Correlations between literacy in 1928 and secondary education attainment in 1981 for provinces are 0.42, 0.53 and 0.54 for males, females and total population respectively and are statistically significant at the 1% level.

(Insert here Figures 1-6)

Next, I present male, female and total tertiary education attainment versus male, female and total literacy rates in 1907 and 1928 (Figures 7-12). The lines are again a linear trend and the figures provide positive influence of literacy rates on tertiary education attainment. Correlations between literacy in 1907 and tertiary education attainment in 1981 for provinces are 0.48, 0.52 and 0.59 for males, females and total population respectively and are statistically significant at the 1%

level. Correlations between literacy in 1928 and tertiary education attainment in 1981 for provinces are 0.38, 0.45 and 0.49 for males, females and total population respectively and are statistically significant at the 1% level.

(Insert here Figures 7-12)

The last six figures provide GDP per capita versus male, female and total literacy rates in 1907 and 1928 (Figures 13-18). The lines are linear trends and depict positive influence between literacy rates and GDP per capita except for figures with male literacy rates (Figures 13 and 16). Correlations between literacy in 1907 and GDP per capita in 2000 for provinces are 0.46 and 0.40 for females and total population and are statistically significant at the 5% level. Correlations between literacy in 1928 and GDP per capita in 2000 for provinces are 0.35 and 0.25 for females and total population and are statistically significant at the 1% level of significance for females and at the 5% level of significance for total population. There is no statistically significant correlation between male literacy rates and GDP per capita.

(Insert here Figures 13-18)

I show that human capital transmission across generations is a main driver of economic outcomes. More generally, the findings that I present underscore the importance of historical institutions and interventions for driving economic growth in the long run.

1.6 Concluding Remarks

Theory suggests that the concentration of landownership has been a major factor of human capital formation and economic growth (Galor et al., 2009; Baten

and Hippe, 2017; Easterly, 2017; Adamopoulos and Restuccia, 2018). The increase in the demand for human capital in the process of industrialization and on the onset of the demographic transition is the main force in the transition from stagnation to growth. As the demand for human capital emerged, differences in the concentration of landownership generated variations in human capital formation, technological changes and the timing of industrialization.

This study examines and analyses the relationship between land ownership and literacy, as proposed by Unified Growth Theory. Empirical results suggest that unequal distribution of land has a substantial negative correlation with literacy. These results were found to be robust across alternative specifications. This negative relationship is based on theoretical grounds. I find a relationship between geographic factors and literacy in Greek provinces and municipalities around 1900. Pasture yield and precipitation are always negatively correlated with literacy rates in estimations with literacy as dependent variable. Rice yield is positively correlated with literacy rates in estimations with land ownership as dependent variable and also in estimations with land ownership as independent variable. Birth rates have a significant negative relationship with literacy rates for Greek provinces around 1900. I also control for several other explanatory factors (urbanization, population density, land area). Population density has positive, statistically significant and strong relationship with literacy rates for all estimation techniques. Land ownership has a negative, statistically significant relationship with all literacy rates. Urbanization is strongly positively correlated with literacy rates in Greek provinces and municipalities in 1907 and 1928.

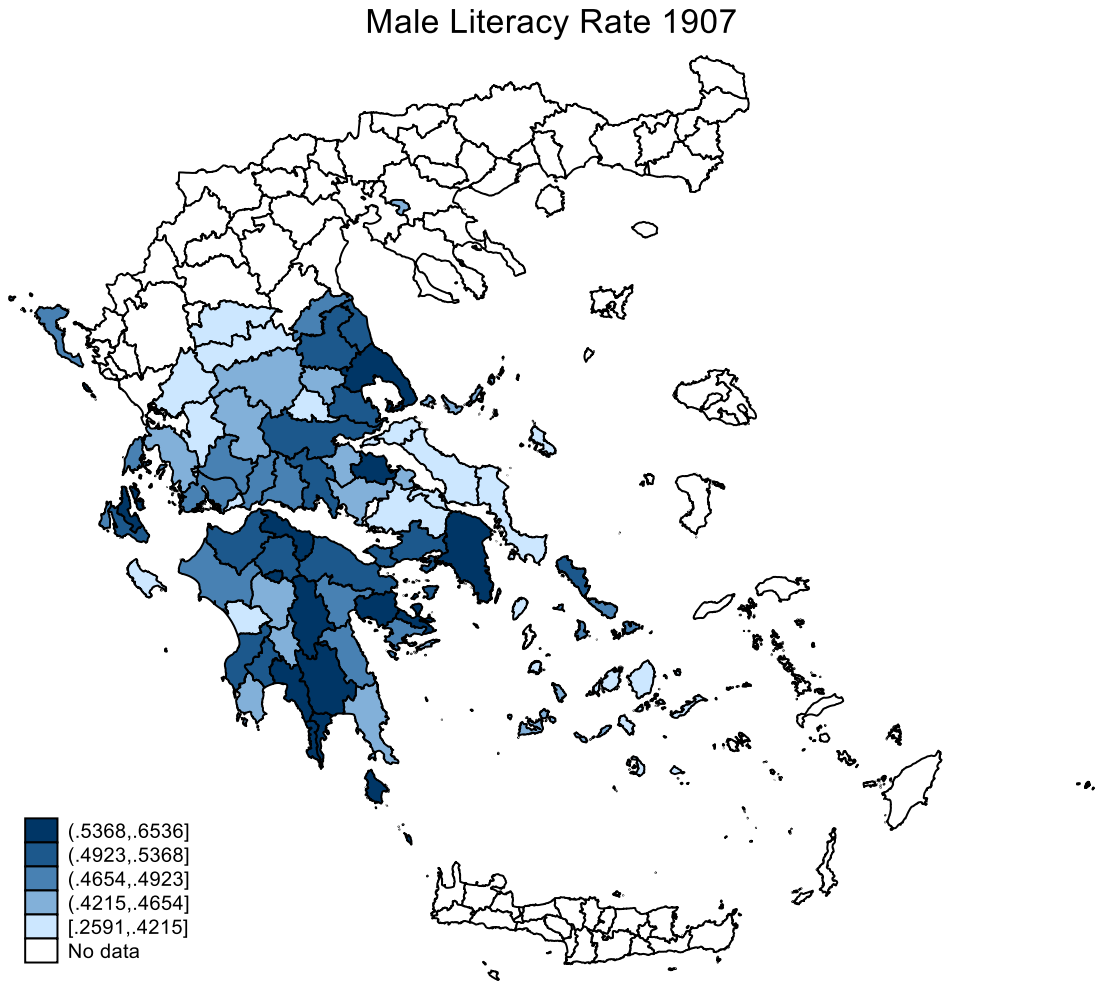
Although I include in my analysis a large set of geographic variables (soil yield,

temperature, altitude, precipitation etc.) the effect of land ownership on literacy rates does not change. I suggest that, in earlier phases of industrialization, the distribution of land has an important negative correlation with the development of literacy. A more equal distribution of land may help to foster educational attainment, economic growth and income distribution. Despite the fact that the agrarian reform dismantling large land holdings took place in the period between 1920 and 1930, the effects of early land ownership on literacy and development seem to persist in the long-run.

Appendix of Chapter 1

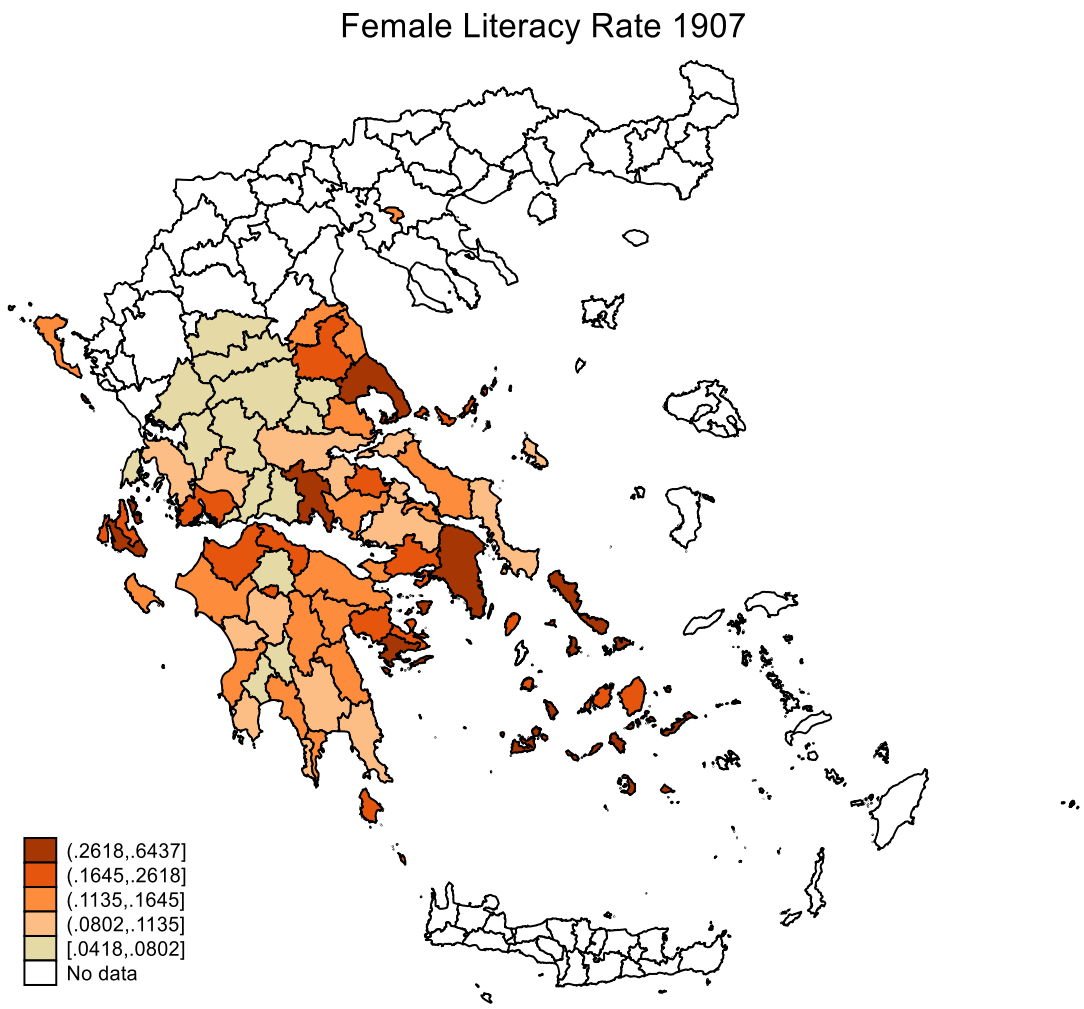
Maps

Map 1



Lowest literacy rates: Valtou(Central Greece/Euboea), Kalampakas(Thessaly), Istiaias(Central Greece/Euboea), Naxou(Cyclades). **Highest literacy rates**: Aigialeias(Peloponnese), Paxon(Ionian Islands) Kythiron(Peloponnese), Attikis(Central Greece/Euboea).

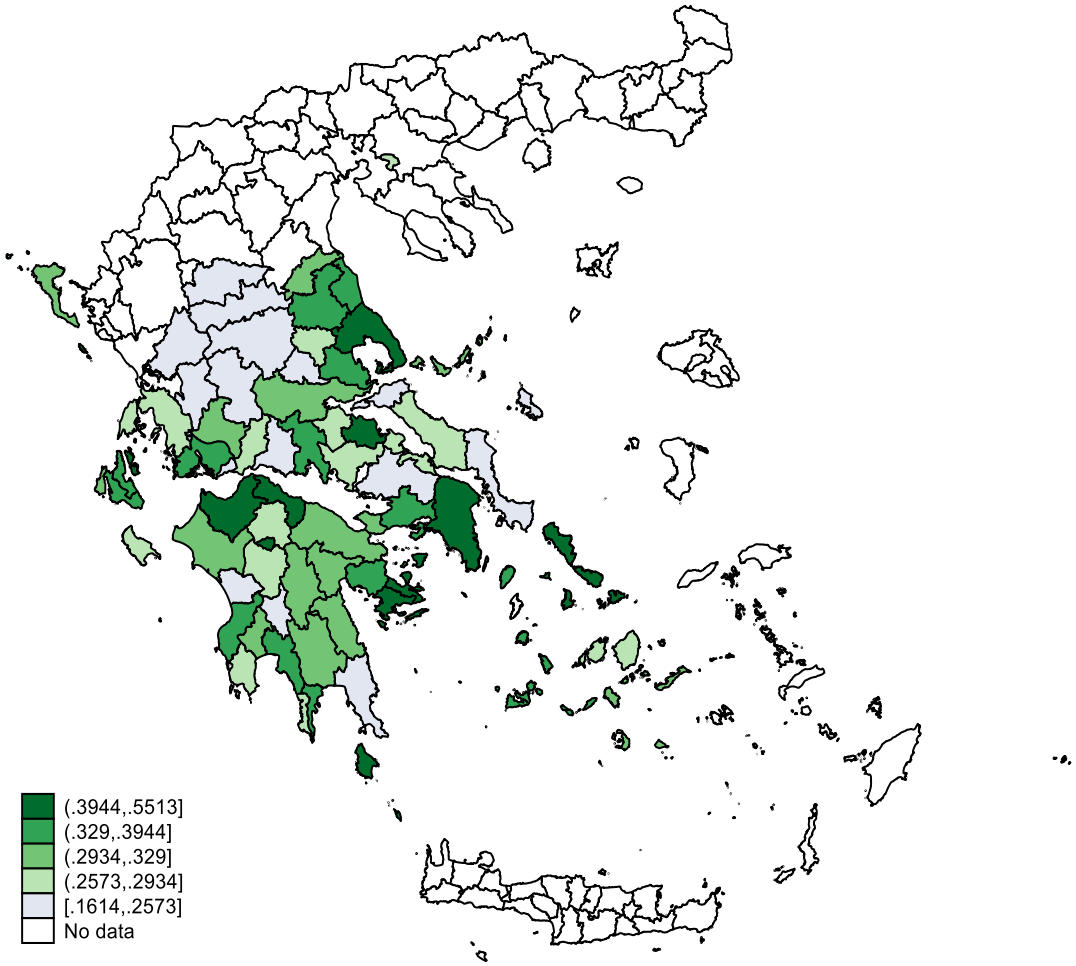
Map 2



Lowest literacy rates: Karditsis(Thessaly), Kalampakas(Thessaly), Doridos(Central Greece/Euboea), Valtou(Central Greece/Euboea),). **Highest literacy rates**: Syrou(Cyclades), Ydras(Peloponnese), Attikis(Central Greece/Euboea), Parnasidos(Central Grrece/Euboea).

Map 3

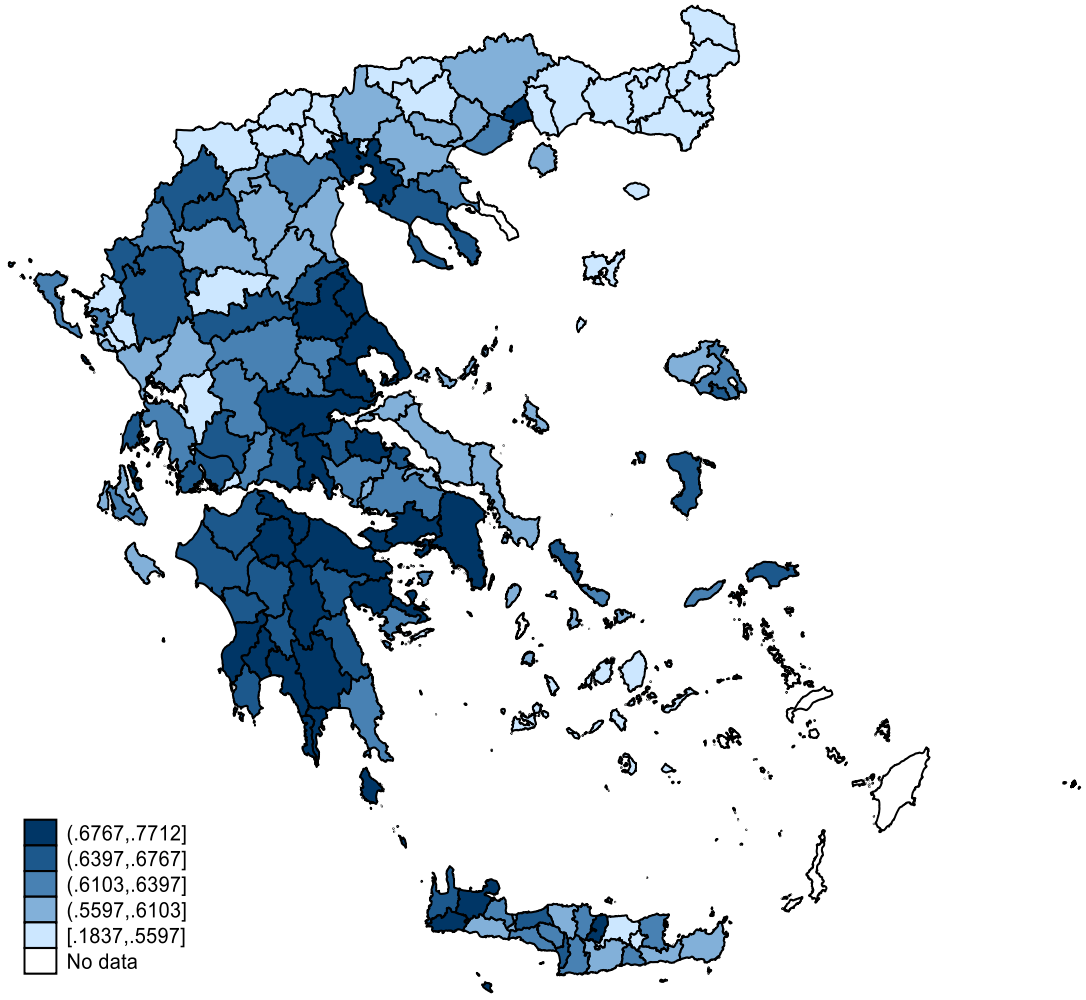
Total Literacy Rate 1907



Lowest literacy rates: Valtou(Central Greece/Euboea), Kalampakas(Thessaly), Istiaias(Central Greece/Euboea), Megalopoleos(Peloponnese). **Highest literacy rates:** Syrou(Cyclades), Kythiron(Peloponnese), Paxon(Ionian Islands), Attikis(Central Greece/Euboea).

Map 4

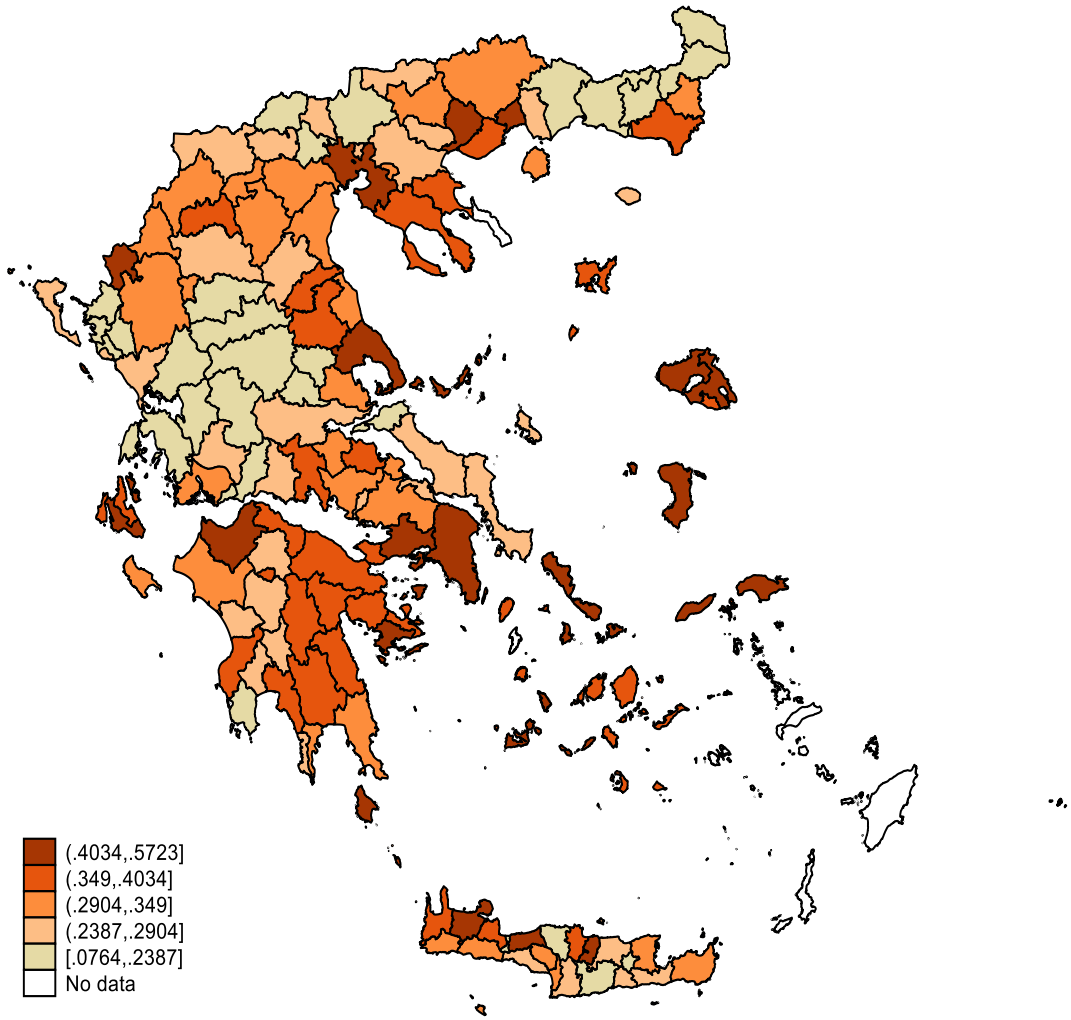
Native Male Literacy Rate 1928



Lowest literacy rates: Sapon(Western Thrace), Komotinis(Western Thrace), Orestiados(Western Thrace), Xanthis(Western Thrace). **Highest literacy rates:** Aigialeias(Peloponnese), Kythiron(Peloponnese), Attikis(Central Greece/Euboea), Paxon(Ionian Islands).

Map 5

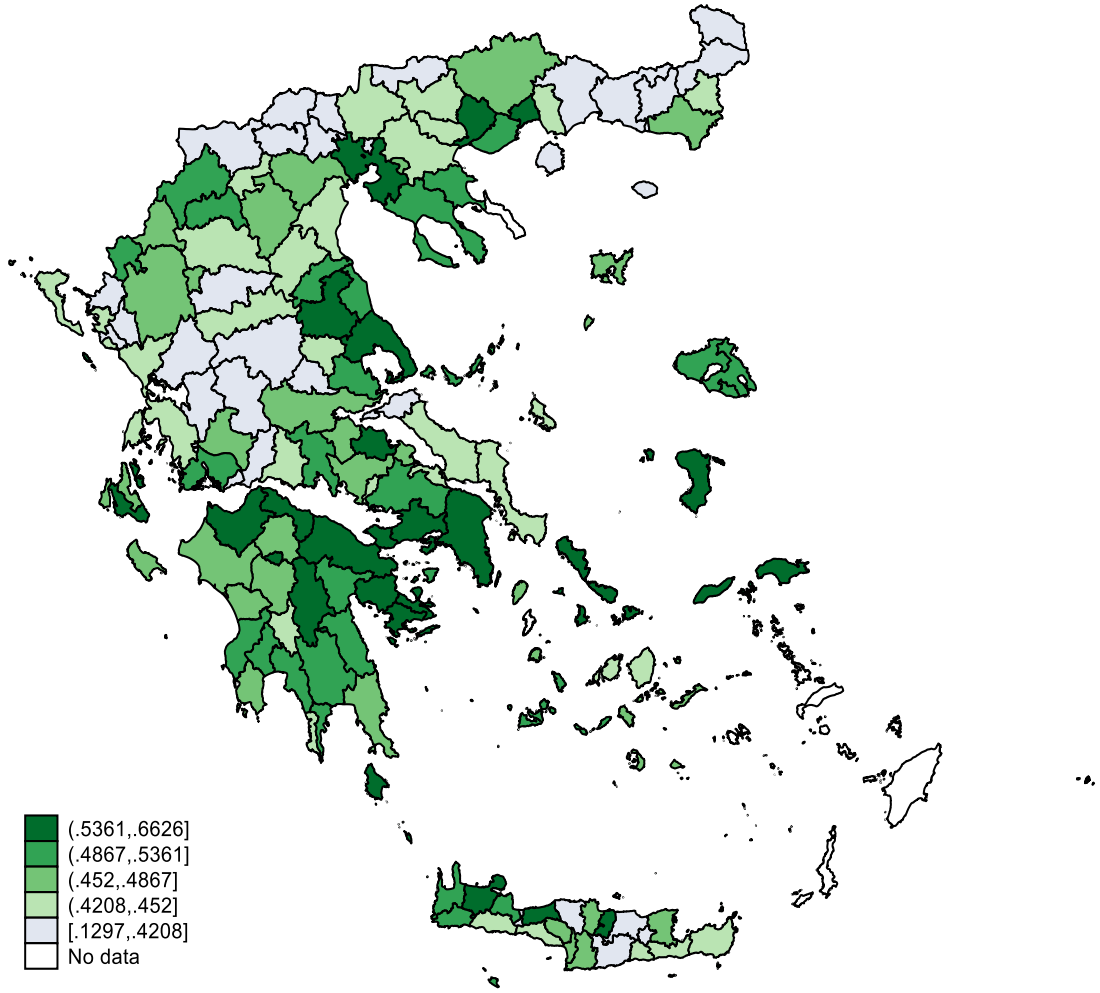
Native Female Literacy Rate 1928



Lowest literacy rates: Sapon(Western Thrace), Orestiados(Western Thrace), Margaritiou(Epirus), Paramythias(Epirus). **Highest literacy rates:** Ydras(Peloponnese), Attikis(Central Greece/Euboea), Tinou(Cyclades), Zichnis/Fylidos(Macedonia).

Map 6

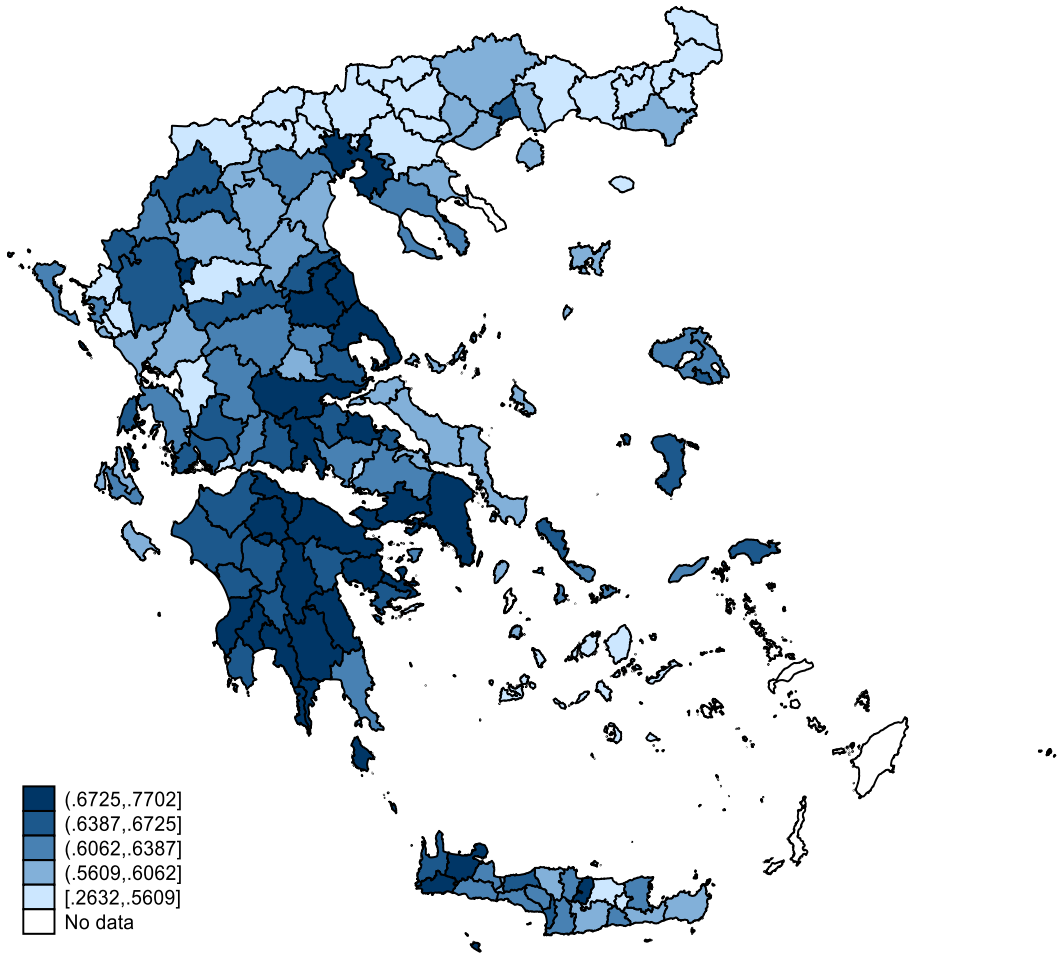
Native Total Literacy Rate 1928



Lowest literary rates: Sapon(Western Thrace), Orestiados(Western Thrace), Komotinis(Western Thrace), Margaritiou(Epirus). **Highest literacy rates:** Tinou(Cyclades), Ithakis(Ionian Islands), Paxon(Ionian Islands), Attikis(Central Greece/Euboea).

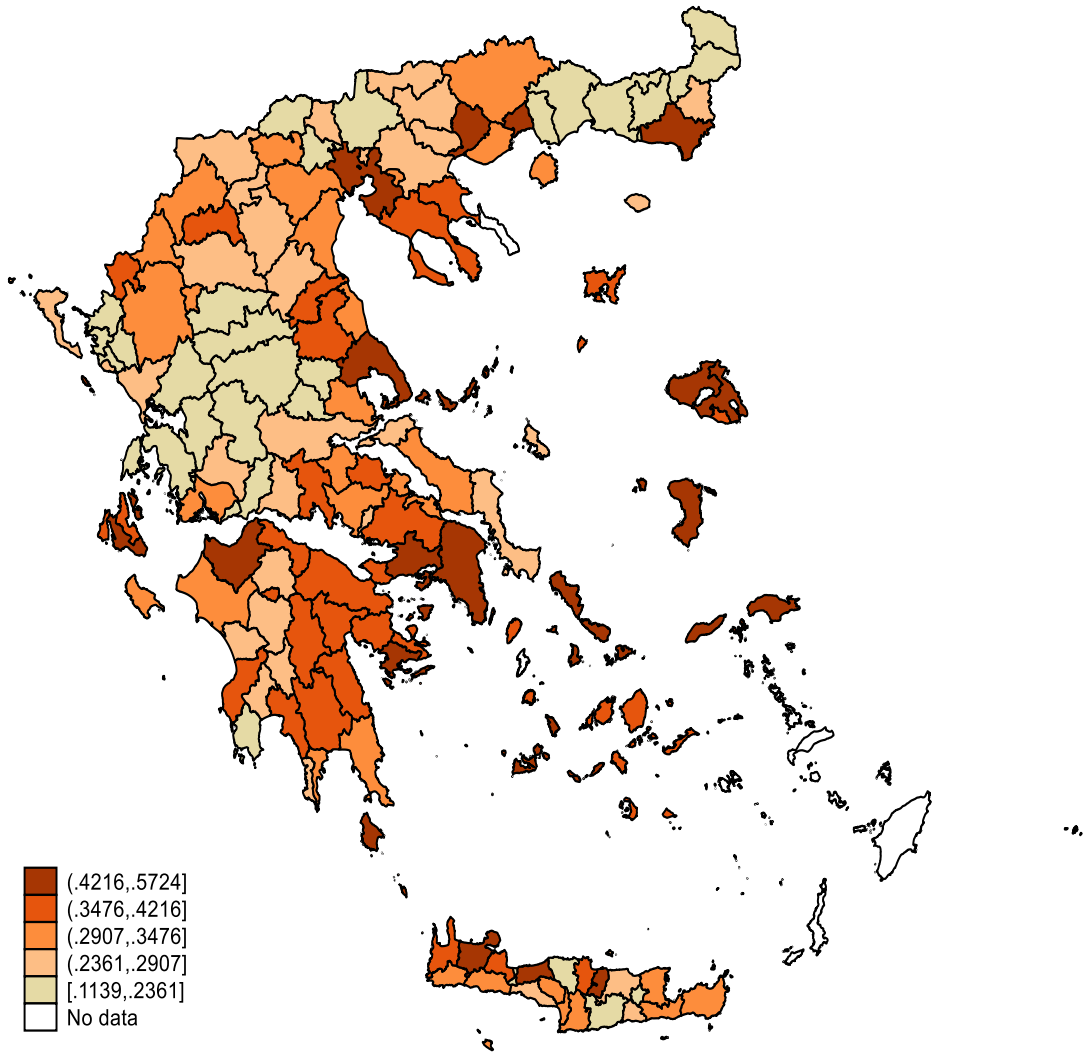
Map 7

Male Literacy Rate 1928



Lowest literacy rates: Sapon(Western Thrace), Margaritiou(Epirus), Orestiados(Western Thrace), Didymoteichou(Western Thrace). **Highest literacy rates:** Mantineias(Peloponnese), Attikis(Central Greece/Euboea), Kythiron(Peloponnese), Paxon(Ionian Islands).

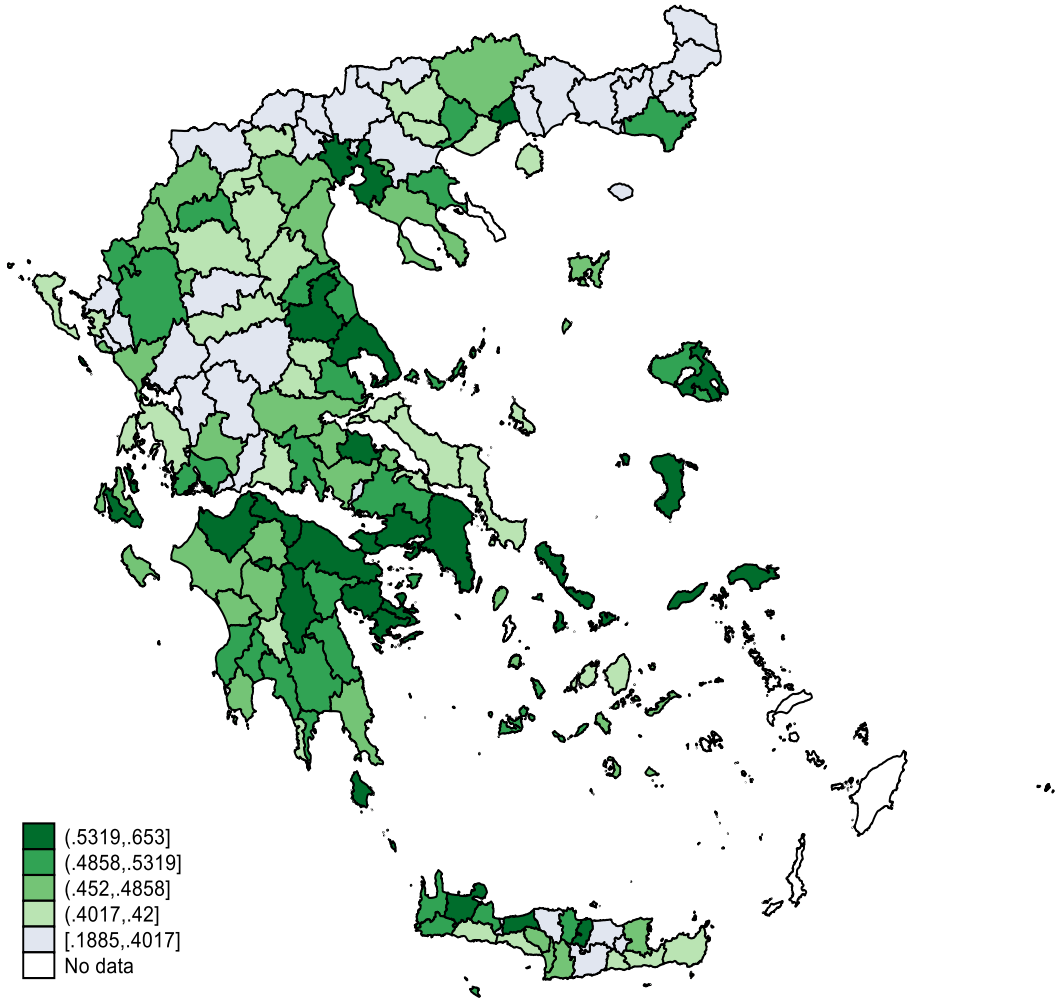
Map 8



Lowest literacy rates: Sapon(Western Thrace), Paramythias(Epirus), Margaritiou(Epirus), Valtou(Central Greece/Euboea). **Highest literacy rates:** Ithakis(Ionian Islands), Ydras(Peloponnese), Attikis(Central Greece/Euboea), Tinou(Cyclades).

Map 9

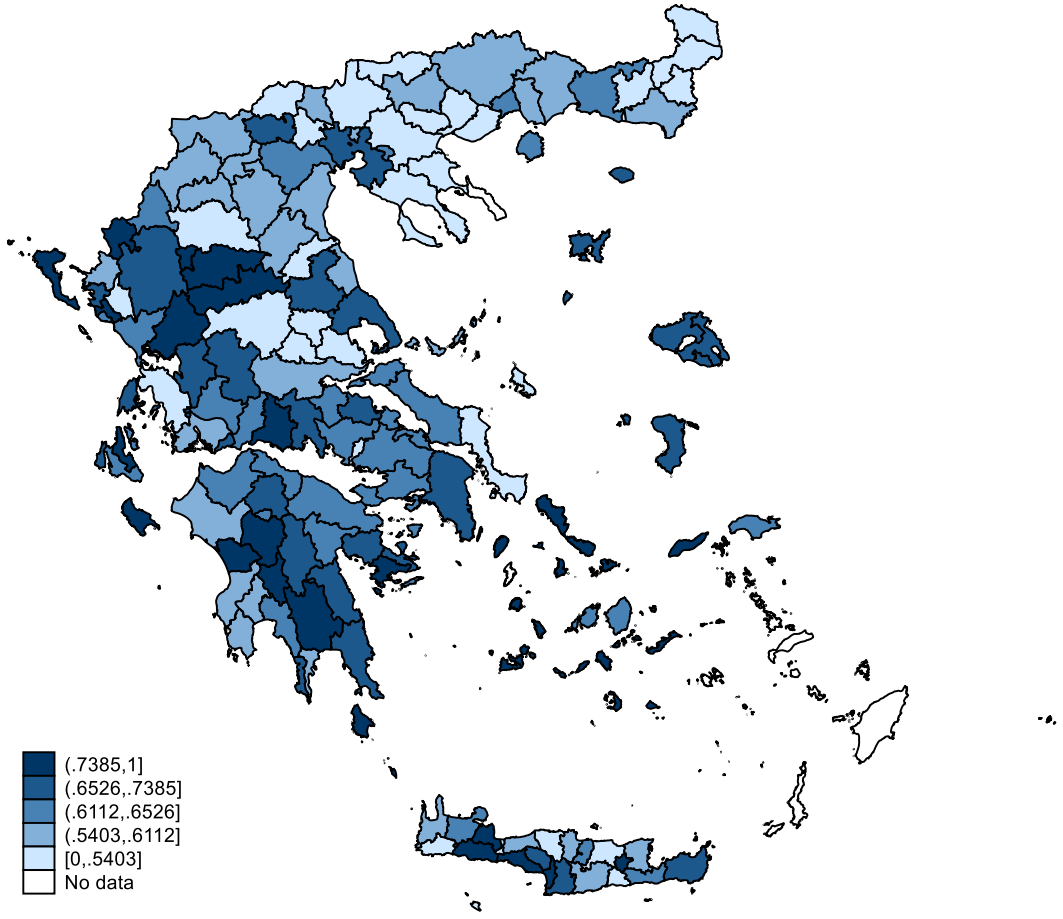
Total Literacy Rate 1928



Lowest literacy rates: Sapon(Western Thrace), Margaritiou(Epirus), Filiaton(Epirus), Orestiados(Western Thrace). **Highest literacy rates:** Tinou(Cyclades), Ithakis(Ionian Islands), Paxon(Ionian Islands), Attikis(Central Greece/Euboea).

Map 10

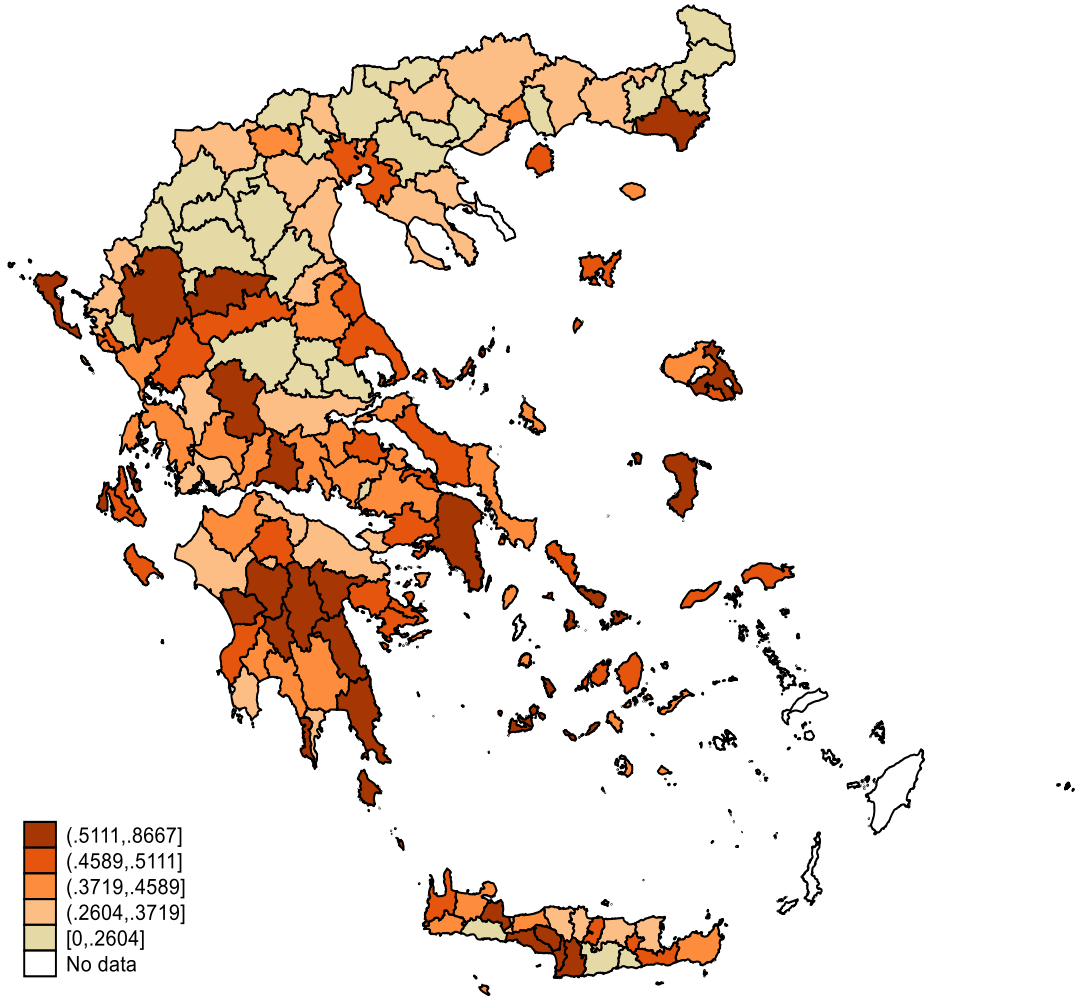
Male Refugee Literacy Rate 1928



Lowest literacy rates: Paxon(Ionian Islands), Didymoteichou(Western Thrace), Domokou(Central Greece/Euboea), Farsalon(Thessaly). **Highest literacy rates:** Syrou(Cyclades), Milou(Cyclades), Metsovou(Epirus), Sfakion(Crete).

Map 11

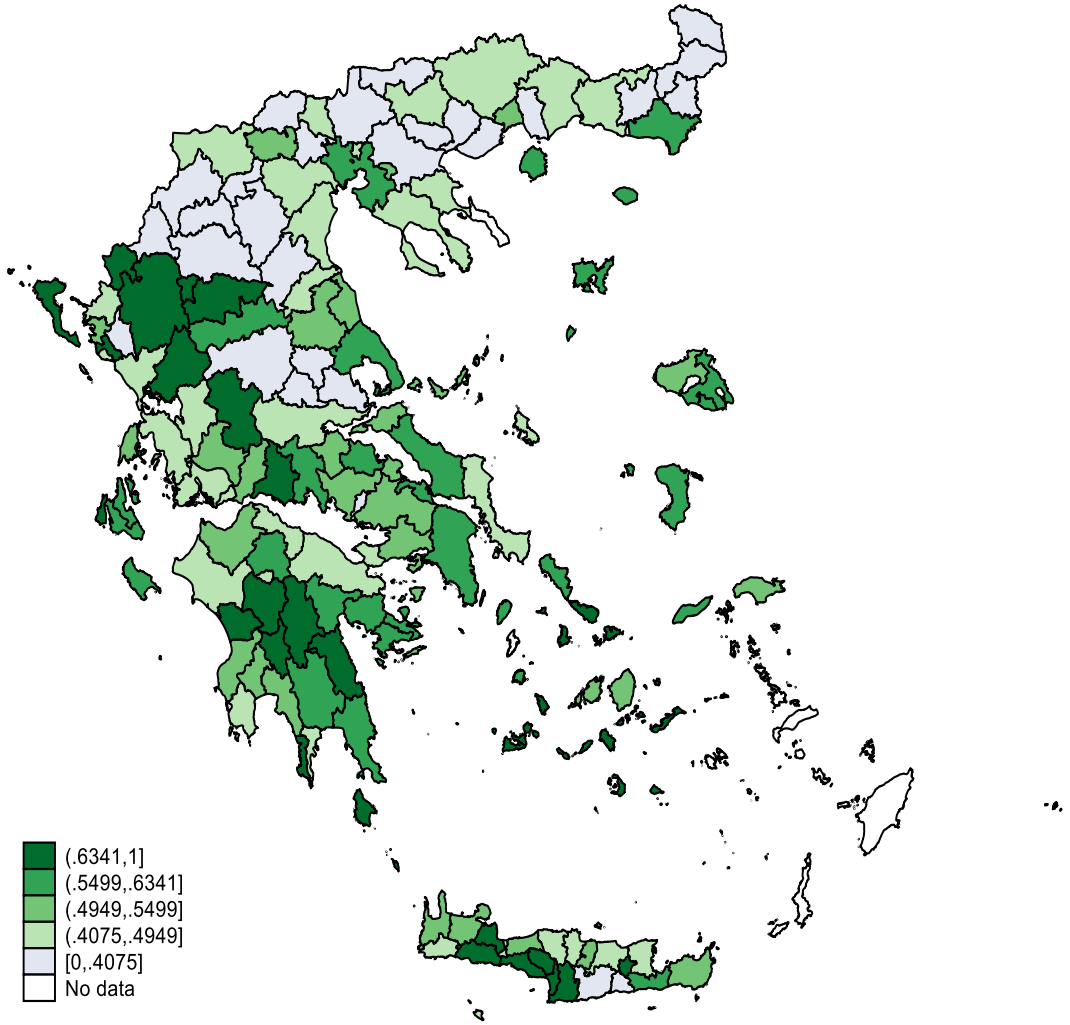
Female Refugee Literacy Rate



Lowest literacy rates: Domokou(Central Greece/Euboea), Konitsis(Epirus), Grevenon(Macedonia), Elassonos(Thessaly). **Highest literacy rates:** Tinou(Cyclades), Agiou Vasileiou(Crete), Evrytanias(Central Greece/Euboea), Apokoronou(Crete).

Map 12

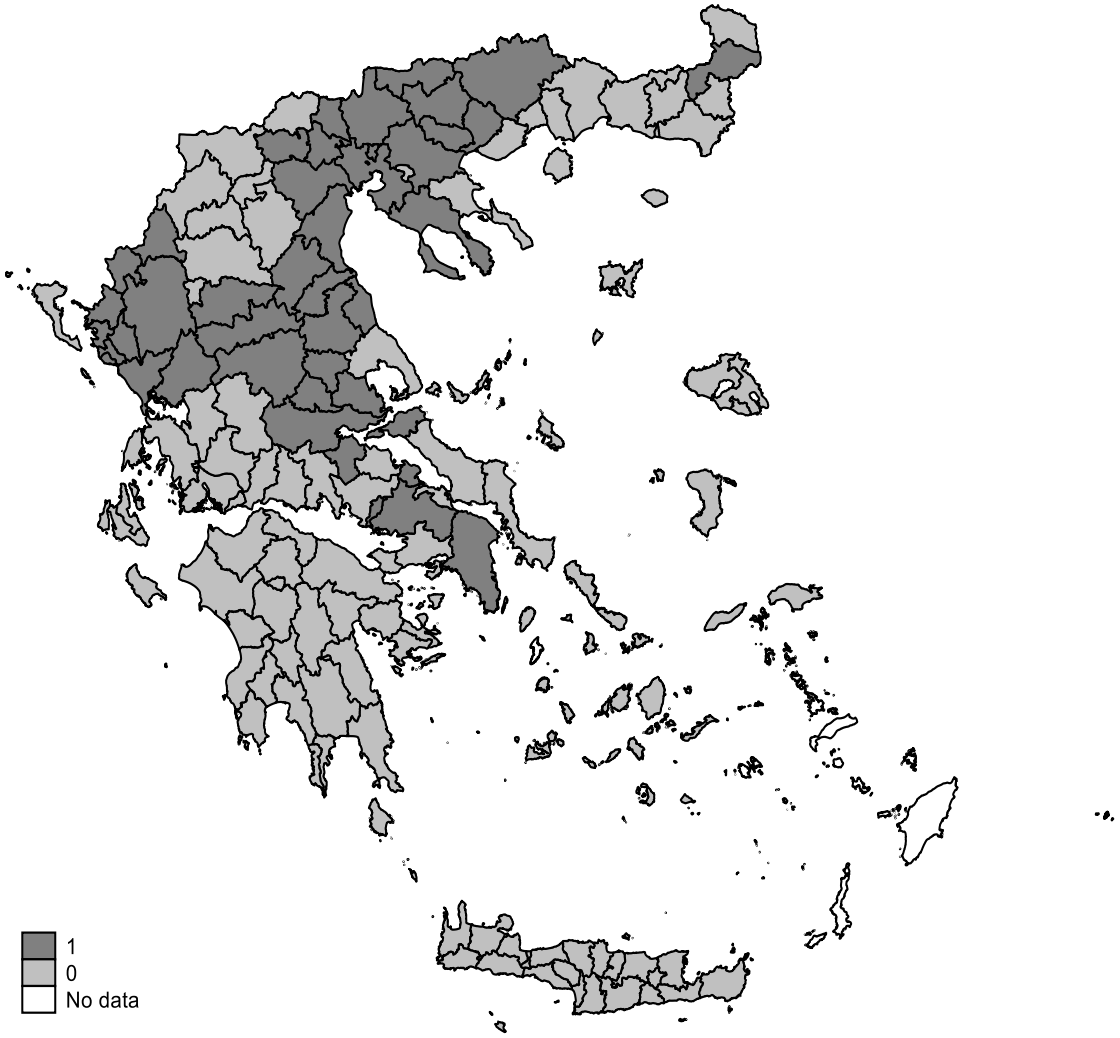
Total Refugee Literacy Rate 1928



Lowest literacy rates: Domokou(Central Greece/Euboea), Didymoteichou(Western Thrace), Grevenon(Macedonia), Paramythias(Epirus). **Highest literacy rates**: Agiou Vasileiou(Crete), Apokoronou(Crete), Sfakion(Crete), Metsovou(Epirus).

Map 13

Land Ownership in 19th century



Number ¨0¨ corresponds to provinces with small land ownership and number ¨1¨ corresponds to provinces with large land ownership. Provinces with dark gray colour characterized by large landholdings.

Tables

Table 1

Definition of Variables

Variable	Description	Period	Source
Human Capital	Literacy rates(native and refugee)	1907, 1928	Census(1907, 1928), Hellenic Statistical Authority
Land Ownership	Dummy variable("0" for regions with small land ownership and "1" for regions with large land ownership)	late 19^{th} century	Petmezas(2003,2006,2012) Vergopoulos(1975)
Urbanization	Urbanization rate (% of population of province (municipality) living in settlements with population>5,000 in 1907 and settlements with population>10,000 in 1928)	1907 ,1928	Census(1907, 1928), Hellenic Statistical Authority
Population density	The quotient of population with land area	1907, 1928	Census(1907, 1928), Hellenic Statistical Authority
Altitude	Altitude in meters	1951	Hellenic Statistical Authority
Ruggedness	Standard deviation of altitude	1951	Calculated from altitude
Precipitation	Median precipitation in mm	1950-2000	IPUMS Terra
Temperature	Median Temperature in ^{0}C	1950-2000	IPUMS Terra
Soil yield	Wheat, sugar, pasture, olive, rice, potato yields		IPUMS Terra
Birth Rate	The quotient of births(male, female, total) with population	1884, 1921	Births/Deaths Census(1884, 1921), Hellenic Statistical Authority
Bulgarians/Turks	The quotient of Bulgarians/Turks with population	1920	Census(1928), Hellenic Statistical Authority

Table 2

Descriptive Statistics of Greek Provinces

Variable	Mean	Standard Deviation	Min	Max	Observations
male literacy rate 1907	.4778	.0787	.2591	.6535	69
female literacy rate 1907	.1721	.1131	.0418	.6437	69
total literacy rate 1907	.3205	.0740	.1613	.5513	69
native male literacy rate 1928	.6095	.0902	.1837	.7712	141
native female literacy rate 1928	.3222	.1063	.0763	.5723	141
native total literacy rate 1928	.4647	.0856	.1297	.6626	141
male literacy rate 1928	.6104	.0814	.2632	.7702	141
female literacy rate 1928	.3230	.1032	.1139	.5724	141
total literacy rate 1928	.4647	.0799	.1884	.6530	141
refugee male literacy rate 1928	.6398	.1197	.3333	1	140
refugee female literacy rate 1928	.4144	.1591	.0560	.8667	138
refugee total literacy rate 1928	.5305	.1379	.2389	1	140
land area	921.0504	671.2756	31	3494	141
population density 1907	49.3539	34.2648	15.0940	158.6355	69
population density 1928	51.4635	48.7827	9.3051	445.6735	141
land ownership	.2766	.4489	0	1	141
urbanization 1907	.1525	.1953	0	.9731	69
urbanization 1928	.1103	.1838	0	.8193	141
altitude	280.5245	223.5981	28	1045	141
ruggedness	176.0746	74.5371	13.8914	363.9166	139
olive yield	.0716	.1048	0	.5164	137
pasture yield	7352337	7871879	0	5.26e+07	136
potato yield	.0038	.0041	4.67e-06	.0198	137
precipitation	114.4734	41.3183	47	196.5	141
rice yield	.0012	.0035	0	.0188	137
temperature	151.2589	21.7937	79.5	185	141
wheat yield	.0456	.0643	.0004	.2760	137
sugar yield	.0023	.0042	0	.0233	137
birth male 1884	.0313	.0074	.0096	.0545	69
birth female 1884	.0280	.0079	.0034	.0483	69
birth total 1884	.0297	.0075	.0063	.0494	69
birth male 1884	.0282	.0085	.0060	.0545	136
birth female 1884	.0241	.0084	.0034	.0483	136
birth total 1884	.0261	.0082	.0048	.0494	136
Bulgarians/Turks	.0833	.1795	0	.9147	142

Table 3

Descriptive Statistics of Greek Municipalities

Variable	Mean	Standard Deviation	Min	Max	Observations
male literacy rate	.4593	.1054	.1411	.7217	445
female literacy rate	.1277	.1082	.0020	.5131	445
total literacy rate	.2904	.0914	.0751	.6249	445
population	5914.519	9746.397	529	175430	445
urbanization	.0641	.2126	0	1	445
land ownership	.2180	.4133	0	1	445
temperature	151.2948	22.2106	59	192	405
precipitation	120.2031	38.7764	45	220	405
wheat yield	.0462	.0691	0	.3531	398
potato yield	.0044	.0065	0	.0475	398
sugar yield	.0019	.0042	0	.0267	398
pasture yield	7516514	9055733	0	6.38e+07	387
olive yield	.0679	.0970	0	.5747	398
rice yield	.0012	.0033	0	.0205	377
birth male 1884	.0330	.0133	.0035	.1012	437
birth female 1884	.0299	.0134	.0008	.0807	437
birth total 1884	.0314	.0124	.0020	.0816	437

Table 4

Correlation of Geographic Factors in Greek Provinces

	altitude	ruggedness	olive yield	pasture yield	potato yield	precipitation	rice yield	temperature	wheat yield	sugar yield
altitude	1.0000									
ruggedness	0.2528*	1.0000								
olive	-0.0972*	-0.2115*	1.0000							
pasture yield	-0.0852*	-0.1234*	0.3221*	1.0000						
potato yield	-0.0998*	-0.1785*	0.7825*	0.4189*	1.0000					
precipitation	0.2220*	-0.0067	0.3536*	0.4430*	0.3745*	1.0000				
rice yield	-0.2019*	0.0993*	-0.1570*	0.0111	-0.1307*	-0.4684*	1.0000			
temperature	-0.6510*	-0.3112*	0.4217*	0.3192*	0.3731*	0.2255*	-0.1011*	1.0000		
wheat yield	-0.0998*	0.0554	-0.2196*	-0.1571*	-0.1221*	-0.6313*	0.7054*	-0.3307*	1.0000	
sugar yield	-0.2049*	0.0024	-0.2715*	-0.2144*	-0.1324*	-0.5639*	0.5307*	-0.3314*	0.8757*	1.0000

Table 5

Correlation of Geographic Factors in Greek Municipalities

	olive yield	pasture yield	potato yield	precipitation	rice yield	temperature	wheat yield	sugar yield
olive yield	1.0000							
pasture yield	0.1647*	1.0000						
potato yield	0.5633*	0.1804*	1.0000					
precipitation	0.1897*	0.3593*	0.2476*	1.0000				
rice yield	-0.0639	0.0118	0.0698	-0.4147*	1.0000			
temperature	0.2946*	0.2137*	0.2929*	0.0542	-0.0054	1.0000		
wheat yield	-0.0375	-0.0376	0.0341	-0.5223*	0.6442*	-0.0877	1.0000	
sugar yield	-0.1628*	-0.0415	-0.0956	-0.4956*	0.5205*	-0.1547*	0.8885*	1.0000

Table 6

Fixed Effect Estimations of Land Ownership in Greek provinces in 1907 and 1928

	(1)	(2)	(3)	(4)
	land ownership	land ownership	land ownership	land ownership
land area	0.000	0.000	0.000**	0.000***
	(0.98)	(0.81)	(3.22)	(2.73)
pop density 1907	0.002	0.002		
	(0.76)	(0.97)		
urbanization 1907	-0.252	-0.252		
	(0.91)	(1.13)		
altitude	-0.000	-0.000	-0.000	-0.000
	(0.95)	(0.56)	(1.63)	(0.87)
ruggedness	0.000	0.000	0.000	0.000
	(0.70)	(0.70)	(0.20)	(0.17)
olive yield	0.124	0.124	0.500	0.500
	(0.11)	(0.11)	(1.46)	(1.48)
pasture yield	-0.000	-0.000	-0.000	-0.000
	(0.78)	(1.08)	(0.17)	(0.13)
potato yield	-11.770	-11.770	-18.657*	-18.657**
	(0.79)	(0.90)	(2.09)	(2.05)
precipitation	-0.004*	-0.004***	-0.000	-0.000
	(2.05)	(3.29)	(0.00)	(0.01)
rice yield	156.593	156.593	40.603***	40.603***
	(1.64)	(1.26)	(5.64)	(3.06)
temperature	-0.005	-0.005	0.000	0.000
	(0.94)	(0.65)	(0.11)	(0.08)
wheat yield	2.437**	2.437*	1.420	1.420
	(2.82)	(1.89)	(1.33)	(1.68)
pop density 1928			0.001*	0.001*
			(2.08)	(1.93)
urbanization 1928			-0.259	-0.259
			(1.32)	(1.32)
constant	1.387	1.387	0.047	0.047
	(1.46)	(0.96)	(0.08)	(0.05)
R^2	0.66	0.66	0.69	0.69
Observations	64	64	135	135

Notes: Fixed effect regressions with land ownership as dependent variable and geographic factors as independent variables in Greek provinces in 1907(Columns (1), (2)) and in 1928 (Columns (3), (4)). Fixed effects at department level and cluster standard errors at department level (Columns (1), (3)) and at prefecture level (Columns (2), (4)). Some coefficients are zero because the values are rounded off to three decimal points. *, **, *** Significant at 10%, 5%, 1% respectively.

Table 7

Fixed Effect Estimations of Literacy Rates on Geographic Factors in Greek provinces in 1907

	(1)	(2)	(3)	(4)	(5)	(6)
	male literacy	female literacy	total literacy	male literacy	female literacy	total literacy
land area	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.50)	(1.06)	(0.50)	(1.05)	(1.11)	(0.83)
pop density	0.001	0.001	0.001*	0.001*	0.001*	0.001**
	(2.00)	(1.50)	(2.26)	(2.01)	(2.03)	(2.32)
urbanization	0.134	0.257***	0.186**	0.134*	0.257***	0.186***
	(1.83)	(4.06)	(3.09)	(1.90)	(3.48)	(3.40)
altitude	-0.000	0.000	-0.000	-0.000	0.000	-0.000
	(0.46)	(0.88)	(0.08)	(0.54)	(0.84)	(0.06)
ruggedness	-0.000	0.000	-0.000	-0.000	0.000	-0.000
	(0.16)	(1.22)	(0.54)	(0.18)	(0.88)	(0.54)
olive yield	0.174	-0.003	0.056	0.174	-0.003	0.056
	(0.78)	(0.01)	(0.21)	(0.53)	(0.01)	(0.22)
pasture yield	-0.000**	-0.000	-0.000**	-0.000**	-0.000	-0.000**
	(3.87)	(0.63)	(3.30)	(2.24)	(0.82)	(2.91)
potato yield	-5.126	-2.723	-2.496	-5.126	-2.723	-2.496
	(1.24)	(0.37)	(0.47)	(0.89)	(0.48)	(0.49)
precipitation	0.000	-0.002	-0.001	0.000	-0.002**	-0.001
	(0.14)	(1.80)	(0.84)	(0.07)	(2.22)	(0.89)
rice yield	19.665	49.754*	32.215	19.665	49.754	32.215
	(1.41)	(2.30)	(1.90)	(0.49)	(1.40)	(0.97)
temperature	-0.001	0.001	0.000	-0.001	0.001	0.000
	(0.56)	(1.04)	(0.05)	(0.70)	(0.89)	(0.04)
wheat yield	-0.134	-0.271	-0.237	-0.134	-0.271	-0.237
	(0.89)	(0.83)	(1.05)	(0.67)	(0.94)	(1.19)
birth male	-3.718***			-3.718***		
	(4.36)			(3.20)		
birth female		-1.421			-1.421	
		(1.07)			(0.97)	
birth total			-2.865**			-2.865**
			(3.06)			(2.62)
constant	0.750*	0.090	0.451*	0.750**	0.090	0.451**
	(2.04)	(0.45)	(2.46)	(2.72)	(0.41)	(2.26)
R^2	0.48	0.53	0.65	0.48	0.53	0.65
Observations	64	64	64	64	64	64

Notes: Fixed effect estimations with male, female and total literacy rates as dependent variables in Greece in 1907 (without land ownership as independent variable). Fixed effects at department level and cluster standard errors at department level (Columns (1), (2), (3)) and at prefecture level (Columns (4), (5), (6). Some coefficients are zero because the values are rounded off to three decimal points.

*, **, *** Significant at 10%, 5%, 1% respectively.

Table 8

Fixed Effect Estimations of Native Literacy Rates on Geographic Factors in Greek provinces in 1928

	(1)	(2)	(3)	(4)	(5)	(6)
	native male literacy	native female literacy	native total literacy	native male literacy	native female literacy	native total literacy
land area	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.48)	(1.24)	(0.84)	(0.73)	(1.39)	(1.28)
pop density	0.000**	0.000*	0.000**	0.000*	0.000*	0.000**
	(2.43)	(1.87)	(2.27)	(1.85)	(1.89)	(2.11)
urbanization	0.081***	0.154***	0.124***	0.081***	0.154***	0.124***
	(4.44)	(5.07)	(5.84)	(3.88)	(4.95)	(5.14)
altitude	0.000	0.000	0.000	0.000	0.000	0.000
	(1.33)	(0.67)	(0.80)	(1.65)	(0.76)	(0.94)
ruggedness	0.000	-0.000	-0.000	0.000	-0.000	-0.000
	(0.32)	(1.01)	(0.58)	(0.31)	(1.10)	(0.61)
olive yield	0.032	0.066	0.040	0.032	0.066	0.040
	(0.27)	(0.50)	(0.38)	(0.27)	(0.53)	(0.35)
pasture yield	-0.000	-0.000***	-0.000**	-0.000**	-0.000***	-0.000***
	(1.75)	(3.36)	(2.56)	(2.20)	(3.55)	(3.20)
potato yield	-1.571	-0.450	-0.510	-1.571	-0.450	-0.510
	(0.46)	(0.11)	(0.16)	(0.48)	(0.13)	(0.16)
precipitation	-0.000	-0.001**	-0.001	-0.000	-0.001***	-0.001*
	(0.38)	(2.55)	(1.59)	(0.40)	(3.20)	(1.80)
rice yield	-2.764	-5.200	-4.023	-2.764	-5.200	-4.023
	(1.28)	(1.76)	(1.67)	(0.81)	(1.27)	(1.19)
temperature	0.001	0.001	0.001	0.001	0.001	0.001
	(0.82)	(0.52)	(0.70)	(0.95)	(0.71)	(0.90)
wheat yield	0.056	0.007	0.030	0.056	0.007	0.030
	(0.40)	(0.02)	(0.14)	(0.37)	(0.03)	(0.16)
birth male	-0.744			-0.744		
	(0.69)			(0.88)		
Bulgarians/Turks	-0.048	-0.072*	-0.044	-0.048	-0.072	-0.044
	(1.34)	(1.90)	(1.26)	(1.26)	(1.37)	(1.03)
birth female		-3.041**			-3.041**	
		(2.58)			(2.33)	
birth total			-1.766			-1.766*
			(1.67)			(1.70)
constant	0.557**	0.479*	0.507**	0.557***	0.479***	0.507***
	(3.18)	(2.08)	(2.58)	(3.78)	(3.06)	(3.51)
R^2	0.57	0.61	0.56	0.57	0.61	0.56
Observations	129	129	129	129	129	129

Notes: Fixed effect estimations with male, female and total literacy rates as dependent variables in Greece in 1928 (without land ownership as independent variable). Fixed effects at department level and cluster standard errors at department level (Columns (1), (2), (3)) and at prefecture level (Columns (4), (5), (6). Some coefficients are zero because the values are rounded off to three decimal points.

*, **, *** Significant at 10%, 5%, 1% respectively.

Table 9

Fixed Effect Estimations of Literacy Rates in Greek provinces in 1907

	(1)	(2)	(3)	(4)	(5)	(6)
	male literacy	female literacy	total literacy	male literacy	female literacy	total literacy
land ownership	-0.069**	-0.088**	-0.046**	-0.069***	-0.088	-0.046***
	(3.39)	(3.59)	(2.57)	(4.00)	(1.46)	(3.37)
land area	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.30)	(0.98)	(0.26)	(0.63)	(1.08)	(0.41)
pop density	0.001	0.001	0.001*	0.001**	0.001*	0.001**
	(1.93)	(1.31)	(2.05)	(2.15)	(1.94)	(2.26)
urbanization	0.116	0.236**	0.175**	0.116*	0.236***	0.175***
	(1.83)	(3.31)	(2.89)	(1.83)	(3.64)	(3.38)
altitude	-0.000	0.000	-0.000	-0.000	0.000	-0.000
	(0.64)	(0.58)	(0.30)	(0.85)	(0.52)	(0.28)
ruggedness	-0.000	0.000	-0.000	-0.000	0.000	-0.000
	(0.01)	(1.21)	(0.38)	(0.01)	(0.99)	(0.38)
olive yield	0.183	0.008	0.062	0.183	0.008	0.062
	(0.97)	(0.02)	(0.24)	(0.56)	(0.02)	(0.24)
pasture yield	-0.000***	-0.000	-0.000***	-0.000**	-0.000	-0.000***
	(12.54)	(1.27)	(4.55)	(2.60)	(1.43)	(3.19)
potato yield	-5.838	-3.776	-3.018	-5.838	-3.776	-3.018
	(1.56)	(0.54)	(0.58)	(0.99)	(0.59)	(0.58)
precipitation	-0.000	-0.002**	-0.001	-0.000	-0.002**	-0.001
	(0.57)	(2.58)	(1.23)	(0.25)	(2.35)	(1.15)
rice yield	28.910	63.804*	39.109*	28.910	63.804	39.109
	(1.87)	(2.41)	(2.14)	(0.75)	(1.49)	(1.20)
temperature	-0.001	0.001	-0.000	-0.001	0.001	-0.000
	(0.72)	(0.64)	(0.21)	(1.02)	(0.50)	(0.19)
wheat yield	0.048	-0.061	-0.120	0.048	-0.061	-0.120
	(0.20)	(0.15)	(0.40)	(0.22)	(0.16)	(0.49)
birth male	-3.981***			-3.981***		
	(4.42)			(3.56)		
birth female		-1.351			-1.351	
		(0.96)			(1.00)	
birth total			-2.941**			-2.941**
			(2.75)			(2.75)
constant	0.862*	0.208	0.520**	0.862***	0.208	0.520**
	(2.29)	(1.06)	(2.72)	(3.37)	(0.87)	(2.87)
R^2	0.53	0.57	0.68	0.53	0.57	0.68
Observations	64	64	64	64	64	64

Notes: Fixed effect estimations with male, female and total literacy rates as dependent variables in Greece in 1907. Fixed effects at department level and cluster standard errors at department level (Columns (1), (2), (3)) and at prefecture level (Columns (4), (5), (6). Some coefficients are zero because the values are rounded off to three decimal points. *, **, *** Significant at 10%, 5%, 1% respectively.

Table 10

Fixed Effect Estimations of Literacy Rates in Greek provinces in 1907

	(1)	(2)	(3)	(4)	(5)	(6)
	male literacy	female literacy	total literacy	male literacy	female literacy	total literacy
land ownership	-0.076***	-0.113**	-0.048***	-0.076***	-0.113	-0.048**
	(6.02)	(3.29)	(4.17)	(3.40)	(1.53)	(2.88)
land area	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.45)	(1.13)	(0.43)	(0.52)	(1.55)	(0.46)
pop density	0.001*	0.002	0.001**	0.001**	0.002	0.001**
	(2.37)	(1.29)	(2.80)	(2.43)	(1.71)	(2.55)
urbanization	0.071	0.202*	0.130*	0.071	0.202**	0.130**
	(1.68)	(2.56)	(2.33)	(0.97)	(2.88)	(2.31)
altitude	-0.000	0.000	-0.000	-0.000	0.000	-0.000
	(0.46)	(0.47)	(0.22)	(0.49)	(0.49)	(0.18)
ruggedness	0.000	0.000	-0.000	0.000	0.000	-0.000
	(0.56)	(0.83)	(0.47)	(0.35)	(0.58)	(0.40)
olive yield	0.015	-0.055	-0.110	0.015	-0.055	-0.110
	(0.03)	(0.07)	(0.28)	(0.03)	(0.09)	(0.27)
pasture yield	-0.000	0.000	-0.000	-0.000	0.000	-0.000
	(1.12)	(1.28)	(0.59)	(1.34)	(0.72)	(0.72)
potato yield	-4.177	-13.024	-4.157	-4.177	-13.024	-4.157
	(0.43)	(0.70)	(0.45)	(0.44)	(0.77)	(0.44)
precipitation	0.000	-0.001	-0.001	0.000	-0.001	-0.001
	(0.10)	(0.99)	(0.85)	(0.13)	(1.00)	(1.02)
rice yield	-11.621	102.687	19.758	-11.621	102.687	19.758
	(0.42)	(1.11)	(0.73)	(0.31)	(1.11)	(0.56)
temperature	-0.001	0.002***	0.001	-0.001	0.002	0.001
	(0.24)	(4.80)	(0.45)	(0.32)	(1.17)	(0.59)
wheat yield	0.337	-0.419	-0.055	0.337	-0.419	-0.055
	(1.13)	(0.73)	(0.18)	(0.95)	(0.69)	(0.22)
birth male	-4.034**			-4.034**		
	(2.66)			(2.46)		
birth female		1.318			1.318	
		(0.32)			(0.38)	
birth total			-2.789*			-2.789*
			(2.14)			(2.04)
constant	0.697	-0.093	0.391	0.697*	-0.093	0.391*
	(1.90)	(0.43)	(2.00)	(2.02)	(0.37)	(2.09)
R^2	0.72	0.69	0.82	0.72	0.69	0.82
Observations	64	64	64	64	64	64

Notes: Fixed effect estimations with male, female and total literacy rates as dependent variables in Greece in 1907. Fixed effects at prefecture level and cluster standard errors at department level (Columns (1), (2), (3)) and at prefecture level (Columns (4), (5), (6). Some coefficients are zero because the values are rounded off to three decimal points.

*, **, *** Significant at 10%, 5%, 1% respectively.

Table 11

Fixed Effect Estimations of Native Literacy Rates in Greek provinces in 1928

	(1)	(2)	(3)	(4)	(5)	(6)
	native male literacy	native female literacy	native total literacy	native male literacy	native female literacy	native total literacy
land ownership	-0.025	-0.007	-0.014	-0.025	-0.007	-0.014
	(1.30)	(0.21)	(0.68)	(1.41)	(0.26)	(0.83)
land area	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.13)	(1.61)	(0.75)	(0.20)	(1.21)	(0.94)
pop density	0.000***	0.000	0.000*	0.000**	0.000*	0.000**
	(3.41)	(1.52)	(2.12)	(2.46)	(1.74)	(2.22)
urbanization	0.074***	0.151***	0.120***	0.074***	0.151***	0.120***
	(5.14)	(4.78)	(6.03)	(4.01)	(4.69)	(5.05)
altitude	0.000	0.000	0.000	0.000	0.000	0.000
	(1.17)	(0.63)	(0.69)	(1.47)	(0.72)	(0.84)
ruggedness	0.000	-0.000	-0.000	0.000	-0.000	-0.000
	(0.31)	(0.98)	(0.57)	(0.32)	(1.09)	(0.61)
olive yield	0.042	0.068	0.045	0.042	0.068	0.045
	(0.38)	(0.50)	(0.43)	(0.36)	(0.54)	(0.40)
pasture yield	-0.000	-0.000***	-0.000**	-0.000**	-0.000***	-0.000***
	(1.79)	(3.39)	(2.59)	(2.24)	(3.54)	(3.21)
potato yield	-1.912	-0.544	-0.694	-1.912	-0.544	-0.694
	(0.61)	(0.14)	(0.22)	(0.60)	(0.15)	(0.22)
precipitation	-0.000	-0.001**	-0.001	-0.000	-0.001***	-0.001*
	(0.41)	(2.53)	(1.60)	(0.44)	(3.21)	(1.82)
rice yield	-1.947	-4.956	-3.570	-1.947	-4.956	-3.570
	(0.87)	(1.72)	(1.54)	(0.54)	(1.13)	(1.01)
temperature	0.001	0.001	0.001	0.001	0.001	0.001
	(0.72)	(0.52)	(0.66)	(0.87)	(0.70)	(0.87)
wheat yield	0.101	0.019	0.053	0.101	0.019	0.053
	(0.66)	(0.06)	(0.25)	(0.70)	(0.08)	(0.29)
birth male	-0.878			-0.878		
	(0.79)			(1.04)		
Bulgarians/Turks	-0.061	-0.076	-0.051	-0.061	-0.076	-0.051
	(1.39)	(1.78)	(1.26)	(1.54)	(1.47)	(1.19)
birth female		-3.052**			-3.052**	
		(2.59)			(2.34)	
birth total			-1.818			-1.818*
			(1.71)			(1.75)
constant	0.571**	0.481*	0.513**	0.571***	0.481***	0.513***
	(2.97)	(2.09)	(2.52)	(3.75)	(3.08)	(3.47)
R^2	0.57	0.61	0.57	0.57	0.61	0.57
Observations	129	129	129	129	129	129

Notes: Fixed effect estimations with native male, female and total literacy rates as dependent variables in Greece in 1928. Fixed effects at department level and cluster standard errors at department level (Columns (1), (2), (3)) and at prefecture level (Columns (4), (5), (6). Some coefficients are zero because the values are rounded off to three decimal points. *, **, *** Significant at 10%, 5%, 1% respectively.

Table 12

Fixed Effect Estimations of Native Literacy Rates in Greek provinces in 1928

	(1)	(2)	(3)	(4)	(5)	(6)
	native male literacy	native female literacy	native total literacy	native male literacy	native female literacy	native total literacy
land ownership	-0.066***	-0.045	-0.051**	-0.066***	-0.045	-0.051***
	(4.27)	(1.33)	(3.20)	(3.72)	(1.45)	(3.33)
land area	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.78)	(1.40)	(1.04)	(0.78)	(1.34)	(1.14)
pop density	0.000	0.000	0.000	0.000*	0.000*	0.000*
	(1.65)	(1.47)	(1.36)	(1.99)	(1.84)	(1.71)
urbanization	0.082**	0.143***	0.122***	0.082***	0.143***	0.122***
	(3.15)	(3.51)	(3.97)	(3.30)	(3.57)	(3.93)
altitude	0.000	0.000	0.000	0.000	0.000	0.000
	(1.05)	(0.30)	(0.66)	(1.10)	(0.29)	(0.64)
ruggedness	0.000	0.000	0.000	0.000	0.000	0.000
	(0.95)	(0.13)	(0.64)	(0.73)	(0.07)	(0.45)
olive yield	-0.004	0.010	-0.004	-0.004	0.010	-0.004
	(0.02)	(0.05)	(0.02)	(0.02)	(0.05)	(0.02)
pasture yield	-0.000	-0.000*	-0.000	-0.000*	-0.000	-0.000*
	(1.44)	(1.99)	(1.66)	(1.95)	(1.42)	(1.82)
potato yield	1.331	0.817	1.472	1.331	0.817	1.472
	(0.19)	(0.14)	(0.24)	(0.21)	(0.14)	(0.24)
precipitation	-0.000	-0.001*	-0.001	-0.000	-0.001**	-0.001
	(0.56)	(2.19)	(1.35)	(0.58)	(2.40)	(1.49)
rice yield	4.091	-4.316	0.071	4.091	-4.316	0.071
	(0.94)	(0.55)	(0.01)	(1.01)	(0.33)	(0.01)
temperature	0.000	0.001	0.000	0.000	0.001	0.000
	(0.33)	(0.61)	(0.60)	(0.24)	(0.67)	(0.56)
wheat yield	0.118	0.023	0.056	0.118	0.023	0.056
	(0.38)	(0.04)	(0.14)	(0.44)	(0.05)	(0.20)
birth male	-2.206*			-2.206**		
	(2.04)			(2.08)		
Bulgarians/Turks	-0.082**	-0.065*	-0.052**	-0.082**	-0.065	-0.052
	(2.73)	(2.09)	(2.28)	(2.36)	(1.23)	(1.33)
birth female		-2.951			-2.951	
		(1.56)			(1.69)	
birth total			-2.545			-2.545*
			(1.73)			(1.88)
constant	0.689***	0.443*	0.550***	0.689***	0.443**	0.550***
	(4.47)	(2.24)	(3.36)	(3.80)	(2.70)	(3.50)
R^2	0.77	0.77	0.77	0.77	0.77	0.77
Observations	129	129	129	129	129	129

Notes: Fixed effect estimations with native male, female and total literacy rates as dependent variables in Greece in 1928. Fixed effects at prefecture level and cluster standard errors at department level (Columns (1), (2), (3)) and at prefecture level (Columns (4), (5), (6). Some coefficients are zero because the values are rounded off to three decimal points. *, **, *** Significant at 10%, 5%, 1% respectively.

Table 13

OLS Estimations of Literacy Rates on Land Ownership in Greek provinces in 1907

	(1)	(2)	(3)	(4)	(5)	(6)
	male literacy	female literacy	total literacy	male literacy	female literacy	total literacy
land ownership	-0.069**	-0.088**	-0.046**	-0.069***	-0.088	-0.046***
	(3.39)	(3.59)	(2.57)	(4.00)	(1.46)	(3.37)
land area	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.30)	(0.98)	(0.26)	(0.63)	(1.08)	(0.41)
pop density	0.001	0.001	0.001*	0.001**	0.001*	0.001**
	(1.93)	(1.31)	(2.05)	(2.15)	(1.94)	(2.26)
urbanization	0.116	0.236**	0.175**	0.116*	0.236***	0.175***
	(1.83)	(3.31)	(2.89)	(1.83)	(3.64)	(3.38)
altitude	-0.000	0.000	-0.000	-0.000	0.000	-0.000
	(0.64)	(0.58)	(0.30)	(0.85)	(0.52)	(0.28)
ruggedness	-0.000	0.000	-0.000	-0.000	0.000	-0.000
	(0.01)	(1.21)	(0.38)	(0.01)	(0.99)	(0.38)
olive yield	0.183	0.008	0.062	0.183	0.008	0.062
	(0.97)	(0.02)	(0.24)	(0.56)	(0.02)	(0.24)
pasture yield	-0.000***	-0.000	-0.000***	-0.000**	-0.000	-0.000***
	(12.54)	(1.27)	(4.55)	(2.60)	(1.43)	(3.19)
potato yield	-5.838	-3.776	-3.018	-5.838	-3.776	-3.018
	(1.56)	(0.54)	(0.58)	(0.99)	(0.59)	(0.58)
precipitation	-0.000	-0.002**	-0.001	-0.000	-0.002**	-0.001
	(0.57)	(2.58)	(1.23)	(0.25)	(2.35)	(1.15)
rice yield	28.910	63.804*	39.109*	28.910	63.804	39.109
	(1.87)	(2.41)	(2.14)	(0.75)	(1.49)	(1.20)
temperature	-0.001	0.001	-0.000	-0.001	0.001	-0.000
	(0.72)	(0.64)	(0.21)	(1.02)	(0.50)	(0.19)
wheat yield	0.048	-0.061	-0.120	0.048	-0.061	-0.120
	(0.20)	(0.15)	(0.40)	(0.22)	(0.16)	(0.49)
birth male	-3.981***			-3.981***		
	(4.42)			(3.56)		
birth female		-1.351			-1.351	
		(0.96)			(1.00)	
birth total			-2.941**			-2.941**
			(2.75)			(2.75)
constant	0.877*	0.168	0.520**	0.877***	0.168	0.520***
	(2.55)	(0.86)	(3.22)	(3.87)	(0.65)	(3.13)
R^2	0.53	0.57	0.68	0.53	0.57	0.68
Observations	64	64	64	64	64	64

Notes: OLS estimations with male, female and total literacy rates as dependent variables in Greek provinces in 1907. Fixed effects at department level and cluster standard errors at department level (Columns (1), (2), (3)) and at prefecture level (Columns (4), (5), (6). Some coefficients are zero because the values are rounded off to three decimal points. *, **, *** Significant at 10%, 5%, 1% respectively.

Table 14

OLS Estimations of Literacy Rates on Land Ownership in Greek provinces in 1907

	(1)	(2)	(3)	(4)	(5)	(6)
	male literacy	female literacy	total literacy	male literacy	female literacy	total literacy
land ownership	-0.076***	-0.113**	-0.048***	-0.076***	-0.113	-0.048**
	(6.02)	(3.29)	(4.17)	(3.40)	(1.53)	(2.88)
land area	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.45)	(1.13)	(0.43)	(0.52)	(1.55)	(0.46)
pop density	0.001*	0.002	0.001**	0.001**	0.002	0.001**
	(2.37)	(1.29)	(2.80)	(2.43)	(1.71)	(2.55)
urbanization	0.071	0.202*	0.130*	0.071	0.202**	0.130**
	(1.68)	(2.56)	(2.33)	(0.97)	(2.88)	(2.31)
altitude	-0.000	0.000	-0.000	-0.000	0.000	-0.000
	(0.46)	(0.47)	(0.22)	(0.49)	(0.49)	(0.18)
ruggedness	0.000	0.000	-0.000	0.000	0.000	-0.000
	(0.56)	(0.83)	(0.47)	(0.35)	(0.58)	(0.40)
olive yield	0.015	-0.055	-0.110	0.015	-0.055	-0.110
	(0.03)	(0.07)	(0.28)	(0.03)	(0.09)	(0.27)
pasture yield	-0.000	0.000	-0.000	-0.000	0.000	-0.000
	(1.12)	(1.28)	(0.59)	(1.34)	(0.72)	(0.72)
potato yield	-4.177	-13.024	-4.157	-4.177	-13.024	-4.157
	(0.43)	(0.70)	(0.45)	(0.44)	(0.77)	(0.44)
precipitation	0.000	-0.001	-0.001	0.000	-0.001	-0.001
	(0.10)	(0.99)	(0.85)	(0.13)	(1.00)	(1.02)
rice yield	-11.621	102.687	19.758	-11.621	102.687	19.758
	(0.42)	(1.11)	(0.73)	(0.31)	(1.11)	(0.56)
temperature	-0.001	0.002***	0.001	-0.001	0.002	0.001
	(0.24)	(4.80)	(0.45)	(0.32)	(1.17)	(0.59)
wheat yield	0.337	-0.419	-0.055	0.337	-0.419	-0.055
	(1.13)	(0.73)	(0.18)	(0.95)	(0.69)	(0.22)
birth male	-4.034**			-4.034**		
	(2.66)			(2.46)		
birth female		1.318			1.318	
		(0.32)			(0.38)	
birth total			-2.789*			-2.789*
			(2.14)			(2.04)
constant	0.691*	0.024	0.458	0.691*	0.024	0.458*
	(2.10)	(0.07)	(1.93)	(1.93)	(0.09)	(2.09)
R^2	0.72	0.69	0.82	0.72	0.69	0.82
Observations	64	64	64	64	64	64

Notes: OLS estimations with male, female and total literacy rates as dependent variables in Greek provinces in 1907. Fixed effects at prefecture level and cluster standard errors at department level (Columns (1), (2), (3)) and at prefecture level (Columns (4), (5), (6). Some coefficients are zero because the values are rounded off to three decimal points.

*, **, *** Significant at 10%, 5%, 1% respectively.

Table 15

OLS Estimations of Native Literacy Rates on Land Ownership in Greek provinces in 1928

	(1)	(2)	(3)	(4)	(5)	(6)
	native male literacy	native female literacy	native total literacy	native male literacy	native female literacy	native total literacy
land ownership	-0.025	-0.007	-0.014	-0.025	-0.007	-0.014
	(1.30)	(0.21)	(0.68)	(1.41)	(0.26)	(0.83)
land area	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.13)	(1.61)	(0.75)	(0.20)	(1.21)	(0.94)
pop density	0.000***	0.000	0.000*	0.000**	0.000*	0.000**
	(3.41)	(1.52)	(2.12)	(2.46)	(1.74)	(2.22)
urbanization	0.074***	0.151***	0.120***	0.074***	0.151***	0.120***
	(5.14)	(4.78)	(6.03)	(4.01)	(4.69)	(5.05)
altitude	0.000	0.000	0.000	0.000	0.000	0.000
	(1.17)	(0.63)	(0.69)	(1.47)	(0.72)	(0.84)
ruggedness	0.000	-0.000	-0.000	0.000	-0.000	-0.000
	(0.31)	(0.98)	(0.57)	(0.32)	(1.09)	(0.61)
olive yield	0.042	0.068	0.045	0.042	0.068	0.045
	(0.38)	(0.50)	(0.43)	(0.36)	(0.54)	(0.40)
pasture yield	-0.000	-0.000***	-0.000**	-0.000**	-0.000***	-0.000***
	(1.79)	(3.39)	(2.59)	(2.24)	(3.54)	(3.21)
potato yield	-1.912	-0.544	-0.694	-1.912	-0.544	-0.694
	(0.61)	(0.14)	(0.22)	(0.60)	(0.15)	(0.22)
precipitation	-0.000	-0.001**	-0.001	-0.000	-0.001***	-0.001*
	(0.41)	(2.53)	(1.60)	(0.44)	(3.21)	(1.82)
rice yield	-1.947	-4.956	-3.570	-1.947	-4.956	-3.570
	(0.87)	(1.72)	(1.54)	(0.54)	(1.13)	(1.01)
temperature	0.001	0.001	0.001	0.001	0.001	0.001
	(0.72)	(0.52)	(0.66)	(0.87)	(0.70)	(0.87)
wheat yield	0.101	0.019	0.053	0.101	0.019	0.053
	(0.66)	(0.06)	(0.25)	(0.70)	(0.08)	(0.29)
birth male	-0.878			-0.878		
	(0.79)			(1.04)		
Bulgarians/Turks	-0.061	-0.076	-0.051	-0.061	-0.076	-0.051
	(1.39)	(1.78)	(1.26)	(1.54)	(1.47)	(1.19)
birth female		-3.052**			-3.052**	
		(2.59)			(2.34)	
birth total			-1.818			-1.818*
			(1.71)			(1.75)
constant	0.366**	0.313	0.328*	0.366**	0.313**	0.328**
	(2.33)	(1.62)	(1.92)	(2.65)	(2.33)	(2.51)
R^2	0.57	0.61	0.57	0.57	0.61	0.57
Observations	129	129	129	129	129	129

Notes: OLS estimations with native male, female and total literacy rates as dependent variables in Greek provinces in 1928. Fixed effects at department level and cluster standard errors at department level (Columns (1), (2), (3)) and at prefecture level (Columns (4), (5), (6). Some coefficients are zero because the values are rounded off to three decimal points. *, **, *** Significant at 10%, 5%, 1% respectively.

Table 16

OLS Estimations of Native Literacy Rates on Land Ownership in Greek provinces in 1928

	(1)	(2)	(3)	(4)	(5)	(6)
	native male literacy	native female literacy	native total literacy	native male literacy	native female literacy	native total literacy
land ownership	-0.066***	-0.045	-0.051**	-0.066***	-0.045	-0.051***
	(4.27)	(1.33)	(3.20)	(3.72)	(1.45)	(3.33)
land area	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.78)	(1.40)	(1.04)	(0.78)	(1.34)	(1.14)
pop density	0.000	0.000	0.000	0.000*	0.000*	0.000*
	(1.65)	(1.47)	(1.36)	(1.99)	(1.84)	(1.71)
urbanization	0.082**	0.143***	0.122***	0.082***	0.143***	0.122***
	(3.15)	(3.51)	(3.97)	(3.30)	(3.57)	(3.93)
altitude	0.000	0.000	0.000	0.000	0.000	0.000
	(1.05)	(0.30)	(0.66)	(1.10)	(0.29)	(0.64)
ruggedness	0.000	0.000	0.000	0.000	0.000	0.000
	(0.95)	(0.13)	(0.64)	(0.73)	(0.07)	(0.45)
olive yield	-0.004	0.010	-0.004	-0.004	0.010	-0.004
	(0.02)	(0.05)	(0.02)	(0.02)	(0.05)	(0.02)
pasture yield	-0.000	-0.000*	-0.000	-0.000*	-0.000	-0.000*
	(1.44)	(1.99)	(1.66)	(1.95)	(1.42)	(1.82)
potato yield	1.331	0.817	1.472	1.331	0.817	1.472
	(0.19)	(0.14)	(0.24)	(0.21)	(0.14)	(0.24)
precipitation	-0.000	-0.001*	-0.001	-0.000	-0.001**	-0.001
	(0.56)	(2.19)	(1.35)	(0.58)	(2.40)	(1.49)
rice yield	4.091	-4.316	0.071	4.091	-4.316	0.071
	(0.94)	(0.55)	(0.01)	(1.01)	(0.33)	(0.01)
temperature	0.000	0.001	0.000	0.000	0.001	0.000
	(0.33)	(0.61)	(0.60)	(0.24)	(0.67)	(0.56)
wheat yield	0.118	0.023	0.056	0.118	0.023	0.056
	(0.38)	(0.04)	(0.14)	(0.44)	(0.05)	(0.20)
birth male	-2.206*			-2.206**		
	(2.04)			(2.08)		
Bulgarians/Turks	-0.082**	-0.065*	-0.052**	-0.082**	-0.065	-0.052
	(2.73)	(2.09)	(2.28)	(2.36)	(1.23)	(1.33)
birth female		-2.951			-2.951	
		(1.56)			(1.69)	
birth total			-2.545			-2.545*
			(1.73)			(1.88)
constant	0.557***	0.372*	0.452**	0.557***	0.372**	0.452***
	(4.59)	(1.93)	(3.01)	(3.86)	(2.58)	(3.54)
R^2	0.77	0.77	0.77	0.77	0.77	0.77
Observations	129	129	129	129	129	129

Notes: OLS estimations with native male, female and total literacy rates as dependent variables in Greek provinces in 1928. Fixed effects at prefecture level and cluster standard errors at department level (Columns (1), (2), (3)) and at prefecture level (Columns (4), (5), (6). Some coefficients are zero because the values are rounded off to three decimal points. *, **, *** Significant at 10%, 5%, 1% respectively.

Table 17

Fixed Effect Estimations of Land Ownership on Geographic Factors in Greek municipalities in 1907

	(1)	(2)
	land ownership	land ownership
population	0.000***	0.000***
	(7.17)	(6.04)
urbanization	-0.144	-0.144
	(1.81)	(1.72)
wheat yield	0.847	0.847*
	(1.56)	(1.96)
potato yield	-0.923	-0.923
	(0.47)	(0.40)
pasture yield	-0.000	-0.000
	(1.67)	(1.47)
temperature	0.000	0.000
	(0.61)	(0.39)
precipitation	0.000	0.000
	(0.49)	(0.41)
olive yield	0.399	0.399*
	(1.38)	(2.09)
rice yield	-8.064	-8.064
	(1.23)	(1.62)
constant	0.032	0.032
	(0.23)	(0.16)
R^2	0.45	0.45
Observations	357	357

Notes: Fixed effect regressions with land ownership as dependent variable and geographic factors as independent variables in Greek municipalities in 1907. Fixed effects at department level and cluster standard errors at department level (Column 1) and at prefecture level (Column 2). Some coefficients are zero because the values are rounded off to three decimal points. *, **, *** Significant at 10%, 5%, 1% respectively.

Table 18

Fixed Effect Estimations of Literacy Rates on Geographic Factors in Greek municipalities in 1907

	(1)	(2)	(3)	(4)	(5)	(6)
	male literacy	female literacy	total literacy	male literacy	female literacy	total literacy
population	0.000***	0.000**	0.000***	0.000**	0.000*	0.000***
	(4.50)	(2.87)	(5.26)	(2.65)	(1.93)	(3.36)
urbanization	0.114***	0.217***	0.177***	0.114***	0.217***	0.177***
	(5.67)	(5.96)	(16.69)	(4.15)	(7.53)	(8.64)
wheat yield	-0.012	-0.131	-0.061	-0.012	-0.131	-0.061
	(0.10)	(0.81)	(0.46)	(0.11)	(1.04)	(0.58)
potato yield	-0.919	-0.475	-0.676	-0.919	-0.475	-0.676
	(0.72)	(0.34)	(0.54)	(0.97)	(0.46)	(0.75)
pasture yield	-0.000***	-0.000	-0.000**	-0.000***	-0.000	-0.000*
	(4.33)	(0.46)	(3.10)	(3.10)	(0.42)	(1.81)
temperature	0.000	0.000	0.000	0.000	0.000	0.000
	(0.05)	(0.24)	(0.41)	(0.06)	(0.30)	(0.37)
precipitation	0.000	-0.001**	-0.000	0.000	-0.001**	-0.000
	(0.73)	(2.60)	(1.62)	(0.59)	(2.29)	(1.03)
olive yield	0.060	0.060	0.063	0.060	0.060	0.063
	(1.18)	(0.48)	(0.73)	(1.50)	(0.67)	(1.13)
rice yield	2.788	-0.390	0.966	2.788	-0.390	0.966
	(1.21)	(0.28)	(0.59)	(1.52)	(0.16)	(0.47)
birth male	-0.663			-0.663		
	(1.11)			(0.99)		
birth female		-0.516			-0.516	
		(0.74)			(1.22)	
birth total			-0.642			-0.642
			(0.97)			(1.20)
constant	0.453***	0.177**	0.307***	0.453***	0.177***	0.307***
	(12.30)	(2.61)	(9.75)	(7.23)	(3.08)	(6.22)
R^2	0.21	0.40	0.30	0.21	0.40	0.30
Observations	349	349	349	349	349	349

Notes: Fixed effect estimations with male, female and total literacy rates as dependent variables in Greek municipalities in 1907 (without land ownership as independent variable). Fixed effects at department level and cluster standard errors at department level (Columns (1), (2), (3)) and at prefecture level (Columns (4), (5), (6). Some coefficients are zero because the values are rounded off to three decimal points. *, **, *** Significant at 10%, 5%, 1% respectively.

Table 19

Fixed Effect Estimations of Literacy Rates in Greek municipalities in 1907

	(1)	(2)	(3)	(4)	(5)	(6)
	male literacy	female literacy	total literacy	male literacy	female literacy	total literacy
land ownership	-0.036	-0.031	-0.026	-0.036	-0.031	-0.026
	(0.87)	(0.78)	(0.67)	(0.94)	(1.07)	(0.83)
population	0.000***	0.000***	0.000***	0.000**	0.000*	0.000**
	(9.11)	(4.21)	(7.89)	(2.32)	(2.06)	(2.65)
urbanization	0.108***	0.213***	0.173***	0.108***	0.213***	0.173***
	(5.48)	(5.66)	(15.28)	(3.63)	(6.85)	(7.42)
wheat yield	0.020	-0.105	-0.039	0.020	-0.105	-0.039
	(0.21)	(0.81)	(0.36)	(0.21)	(0.96)	(0.43)
potato yield	-0.952	-0.512	-0.701	-0.952	-0.512	-0.701
	(0.74)	(0.37)	(0.55)	(1.00)	(0.50)	(0.77)
pasture yield	-0.000***	-0.000	-0.000**	-0.000***	-0.000	-0.000*
	(4.21)	(0.57)	(3.05)	(3.10)	(0.50)	(1.86)
temperature	0.000	0.000	0.000	0.000	0.000	0.000
	(0.09)	(0.26)	(0.44)	(0.10)	(0.33)	(0.41)
precipitation	0.000	-0.000**	-0.000	0.000	-0.000**	-0.000
	(1.10)	(2.65)	(2.00)	(0.69)	(2.14)	(0.94)
olive yield	0.074	0.072	0.073	0.074*	0.072	0.073
	(1.38)	(0.59)	(0.85)	(1.87)	(0.84)	(1.34)
rice yield	2.517	-0.597	0.780	2.517	-0.597	0.780
	(1.10)	(0.47)	(0.49)	(1.34)	(0.24)	(0.38)
birth male	-0.656			-0.656		
	(1.14)			(0.97)		
birth female		-0.460			-0.460	
		(0.67)			(1.02)	
birth total			-0.615			-0.615
			(0.95)			(1.12)
constant	0.453***	0.176**	0.306***	0.453***	0.176***	0.306***
	(12.21)	(2.59)	(9.62)	(7.43)	(3.11)	(6.36)
R^2	0.22	0.41	0.31	0.22	0.41	0.31
Observations	349	349	349	349	349	349

Notes: Fixed effect estimations with male, female and total literacy rates as dependent variables in Greek municipalities in 1907. Fixed effects at department level and cluster standard errors at department level (Columns (1), (2), (3)) and at prefecture level (Columns (4), (5), (6). Some coefficients are zero because the values are rounded off to three decimal points. *, **, *** Significant at 10%, 5%, 1% respectively.

Table 20

Fixed Effect Estimations of Literacy Rates in Greek municipalities in 1907

	(1)	(2)	(3)	(4)	(5)	(6)
	male literacy	female literacy	total literacy	male literacy	female literacy	total literacy
land ownership	-0.063***	-0.050	-0.042*	-0.063**	-0.050	-0.042
	(4.93)	(1.54)	(2.14)	(2.78)	(1.73)	(1.74)
population	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
	(13.74)	(5.08)	(7.93)	(4.70)	(2.94)	(4.25)
urbanization	0.098***	0.194***	0.159***	0.098***	0.194***	0.159***
	(13.07)	(5.12)	(8.76)	(4.14)	(6.16)	(7.13)
wheat yield	0.134*	-0.028	0.047	0.134*	-0.028	0.047
	(2.12)	(0.26)	(0.65)	(1.94)	(0.27)	(0.63)
potato yield	-0.920	-0.406	-0.845	-0.920	-0.406	-0.845
	(0.63)	(0.27)	(0.63)	(0.90)	(0.38)	(0.88)
pasture yield	-0.000***	-0.000	-0.000**	-0.000***	-0.000	-0.000**
	(7.63)	(0.84)	(2.67)	(3.79)	(0.89)	(2.41)
temperature	0.000	0.000	0.000	0.000	0.000	0.000
	(0.27)	(0.34)	(0.47)	(0.26)	(0.40)	(0.44)
precipitation	0.000	-0.000	-0.000	0.000	-0.000	-0.000
	(1.26)	(1.98)	(0.24)	(0.99)	(1.10)	(0.09)
olive yield	0.080	0.014	0.052	0.080*	0.014	0.052
	(1.47)	(0.12)	(0.63)	(1.85)	(0.17)	(0.92)
rice yield	1.891	-0.857	0.544	1.891	-0.857	0.544
	(1.10)	(0.74)	(0.46)	(1.31)	(0.35)	(0.30)
birth male	-0.491			-0.491		
	(0.72)			(0.65)		
birth female		-0.344			-0.344	
		(0.51)			(0.72)	
birth total			-0.428			-0.428
			(0.64)			(0.69)
constant	0.436***	0.141*	0.280***	0.436***	0.141**	0.280***
	(10.19)	(2.27)	(8.44)	(8.00)	(2.55)	(6.24)
R^2	0.36	0.49	0.41	0.36	0.49	0.41
Observations	349	349	349	349	349	349

Notes: Fixed effect estimations with male, female and total literacy rates as dependent variables in Greek municipalities in 1907. Fixed effects at prefecture level and cluster standard errors at department level (Columns (1), (2), (3)) and at prefecture level (Columns (4), (5), (6). Some coefficients are zero because the values are rounded off to three decimal points. *, **, *** Significant at 10%, 5%, 1% respectively.

Table 21

OLS Estimations of Literacy Rates on Land Ownership in Greek municipalities in 1907

	(1)	(2)	(3)	(4)	(5)	(6)
	male literacy	female literacy	total literacy	male literacy	female literacy	total literacy
land ownership	-0.036	-0.031	-0.026	-0.036	-0.031	-0.026
	(0.87)	(0.78)	(0.67)	(0.94)	(1.07)	(0.83)
population	0.000***	0.000***	0.000***	0.000**	0.000*	0.000**
	(9.11)	(4.21)	(7.89)	(2.32)	(2.06)	(2.65)
urbanization	0.108***	0.213***	0.173***	0.108***	0.213***	0.173***
	(5.48)	(5.66)	(15.28)	(3.63)	(6.85)	(7.42)
wheat yield	0.020	-0.105	-0.039	0.020	-0.105	-0.039
	(0.21)	(0.81)	(0.36)	(0.21)	(0.96)	(0.43)
potato yield	-0.952	-0.512	-0.701	-0.952	-0.512	-0.701
	(0.74)	(0.37)	(0.55)	(1.00)	(0.50)	(0.77)
pasture yield	-0.000***	-0.000	-0.000**	-0.000***	-0.000	-0.000*
	(4.21)	(0.57)	(3.05)	(3.10)	(0.50)	(1.86)
temperature	0.000	0.000	0.000	0.000	0.000	0.000
	(0.09)	(0.26)	(0.44)	(0.10)	(0.33)	(0.41)
precipitation	0.000	-0.000**	-0.000	0.000	-0.000**	-0.000
	(1.10)	(2.65)	(2.00)	(0.69)	(2.14)	(0.94)
olive yield	0.074	0.072	0.073	0.074*	0.072	0.073
	(1.38)	(0.59)	(0.85)	(1.87)	(0.84)	(1.34)
rice yield	2.517	-0.597	0.780	2.517	-0.597	0.780
	(1.10)	(0.47)	(0.49)	(1.34)	(0.24)	(0.38)
birth male	-0.656			-0.656		
	(1.14)			(0.97)		
birth female		-0.460			-0.460	
		(0.67)			(1.02)	
birth total			-0.615			-0.615
			(0.95)			(1.12)
constant	0.464***	0.160*	0.307***	0.464***	0.160**	0.307***
	(13.51)	(2.05)	(8.20)	(5.63)	(2.32)	(4.54)
R^2	0.22	0.41	0.31	0.22	0.41	0.31
Observations	349	349	349	349	349	349

Notes: OLS estimations with male, female and total literacy rates as dependent variables in Greek municipalities in 1907. Fixed effects at department level and cluster standard errors at department level (Columns (1), (2), (3)) and at prefecture level (Columns (4), (5), (6). Some coefficients are zero because the values are rounded off to three decimal points. *, **, *** Significant at 10%, 5%, 1% respectively.

Table 22

OLS Estimations of Literacy Rates on Land Ownership in Greek municipalities in 1907

	(1)	(2)	(3)	(4)	(5)	(6)
	male literacy	female literacy	total literacy	male literacy	female literacy	total literacy
land ownership	-0.063***	-0.050	-0.042*	-0.063**	-0.050	-0.042
	(4.93)	(1.54)	(2.14)	(2.78)	(1.73)	(1.74)
population	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
	(13.74)	(5.08)	(7.93)	(4.70)	(2.94)	(4.25)
urbanization	0.098***	0.194***	0.159***	0.098***	0.194***	0.159***
	(13.07)	(5.12)	(8.76)	(4.14)	(6.16)	(7.13)
wheat yield	0.134*	-0.028	0.047	0.134*	-0.028	0.047
	(2.12)	(0.26)	(0.65)	(1.94)	(0.27)	(0.63)
potato yield	-0.920	-0.406	-0.845	-0.920	-0.406	-0.845
	(0.63)	(0.27)	(0.63)	(0.90)	(0.38)	(0.88)
pasture yield	-0.000***	-0.000	-0.000**	-0.000***	-0.000	-0.000**
	(7.63)	(0.84)	(2.67)	(3.79)	(0.89)	(2.41)
temperature	0.000	0.000	0.000	0.000	0.000	0.000
	(0.27)	(0.34)	(0.47)	(0.26)	(0.40)	(0.44)
precipitation	0.000	-0.000	-0.000	0.000	-0.000	-0.000
	(1.26)	(1.98)	(0.24)	(0.99)	(1.10)	(0.09)
olive yield	0.080	0.014	0.052	0.080*	0.014	0.052
	(1.47)	(0.12)	(0.63)	(1.85)	(0.17)	(0.92)
rice yield	1.891	-0.857	0.544	1.891	-0.857	0.544
	(1.10)	(0.74)	(0.46)	(1.31)	(0.35)	(0.30)
birth male	-0.491			-0.491		
	(0.72)			(0.65)		
birth female		-0.344			-0.344	
		(0.51)			(0.72)	
birth total			-0.428			-0.428
			(0.64)			(0.69)
constant	0.455***	0.208**	0.323***	0.455***	0.208***	0.323***
	(8.92)	(3.74)	(9.04)	(8.03)	(3.76)	(6.98)
R^2	0.36	0.49	0.41	0.36	0.49	0.41
Observations	349	349	349	349	349	349

Notes: OLS estimations with male, female and total literacy rates as dependent variables in Greek municipalities in 1907. Fixed effects at prefecture level and cluster standard errors at department level (Columns (1), (2), (3)) and at prefecture level (Columns (4), (5), (6). Some coefficients are zero because the values are rounded off to three decimal points. *, **, *** Significant at 10%, 5%, 1% respectively.

Table 23

Fixed Effect Estimations of Literacy Rates on Geographic Factors in Greek provinces in 1928

	(1)	(2)	(3)	(4)	(5)	(6)
	male literacy	female literacy	total literacy	male literacy	female literacy	total literacy
refugee male literacy	0.105*			0.105*		
	(2.00)			(1.84)		
land area	-0.000	-0.000	-0.000	-0.000	-0.000*	-0.000
	(0.81)	(1.49)	(0.98)	(1.11)	(1.91)	(1.56)
pop density	0.000	0.000	0.000	0.000	0.000	0.000
	(1.50)	(1.10)	(1.27)	(1.26)	(1.03)	(1.15)
urbanization	0.084***	0.182***	0.135***	0.084***	0.182***	0.135***
	(5.25)	(5.58)	(6.96)	(3.97)	(5.37)	(5.44)
altitude	0.000	0.000*	0.000	0.000	0.000*	0.000
	(1.47)	(2.16)	(1.23)	(1.66)	(1.96)	(1.35)
ruggedness	0.000	-0.000	-0.000	0.000	-0.000	-0.000
	(0.78)	(0.83)	(0.23)	(0.63)	(0.95)	(0.25)
olive yield	-0.016	-0.001	-0.010	-0.016	-0.001	-0.010
	(0.14)	(0.01)	(0.11)	(0.14)	(0.01)	(0.09)
pasture yield	-0.000	-0.000***	-0.000**	-0.000**	-0.000***	-0.000***
	(1.77)	(3.81)	(2.77)	(2.13)	(3.58)	(3.13)
potato yield	-0.315	0.716	0.616	-0.315	0.716	0.616
	(0.09)	(0.20)	(0.21)	(0.10)	(0.20)	(0.20)
precipitation	-0.000	-0.002***	-0.001**	-0.000	-0.002***	-0.001***
	(0.85)	(3.73)	(2.63)	(0.91)	(3.97)	(2.77)
rice yield	-2.418	-3.837	-3.508	-2.418	-3.837	-3.508
	(1.21)	(1.31)	(1.48)	(0.96)	(1.09)	(1.26)
temperature	0.001	0.001	0.001	0.001	0.001*	0.001
	(1.11)	(1.42)	(1.25)	(1.08)	(1.69)	(1.46)
wheat yield	0.098	0.004	0.073	0.098	0.004	0.073
	(0.69)	(0.02)	(0.34)	(0.68)	(0.02)	(0.41)
birth male	-1.191			-1.191		
	(1.15)			(1.45)		
Bulgarians/Turks	-0.051	-0.055	-0.051	-0.051	-0.055	-0.051
	(1.58)	(1.15)	(1.41)	(1.51)	(0.97)	(1.28)
refugee female literacy		0.085**			0.085**	
		(2.28)			(2.04)	
birth female		-2.971**			-2.971**	
		(2.54)			(2.56)	
refugee total literacy			0.113**			0.113***
			(2.74)			(3.35)
birth total			-1.953*			-1.953**
			(1.89)			(2.09)
constant	0.531***	0.350	0.440**	0.531***	0.350**	0.440***
	(4.43)	(1.71)	(2.96)	(4.93)	(2.38)	(3.92)
R^2	0.60	0.67	0.62	0.60	0.67	0.62
Observations	128	126	128	128	126	128

Notes: Fixed effect estimations with male, female and total literacy rates as dependent variables in Greek provinces in 1928 (without land ownership as independent variable). Fixed effects at department level and cluster standard errors at department level (Columns (1), (2), (3)) and at prefecture level (Columns (4), (5), (6). Some coefficients are zero because the values are rounded off to three decimal points. Refugees included. *, **, *** Significant at 10%, 5%, 1% respectively.

Table 24

OLS Estimations of Literacy Rates on Land Ownership in Greek provinces in 1928

	(1)	(2)	(3)	(4)	(5)	(6)
	male literacy	female literacy	total literacy	male literacy	female literacy	total literacy
refugee male literacy	0.095			0.095		
	(1.81)			(1.68)		
land ownership	-0.031**	-0.014	-0.017	-0.031**	-0.014	-0.017
	(2.62)	(0.55)	(1.38)	(2.03)	(0.61)	(1.19)
land area	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.24)	(1.35)	(0.74)	(0.36)	(1.64)	(1.17)
pop density	0.000*	0.000	0.000	0.000*	0.000	0.000
	(2.25)	(1.00)	(1.31)	(1.94)	(1.06)	(1.37)
urbanization	0.077***	0.179***	0.131***	0.077***	0.179***	0.131***
	(5.53)	(5.06)	(6.29)	(3.98)	(5.06)	(5.30)
altitude	0.000	0.000*	0.000	0.000	0.000*	0.000
	(1.22)	(1.98)	(1.05)	(1.45)	(1.89)	(1.22)
ruggedness	0.000	-0.000	-0.000	0.000	-0.000	-0.000
	(0.76)	(0.83)	(0.23)	(0.69)	(0.96)	(0.25)
olive yield	0.001	0.005	0.000	0.001	0.005	0.000
	(0.01)	(0.04)	(0.00)	(0.01)	(0.04)	(0.00)
pasture yield	-0.000	-0.000***	-0.000**	-0.000**	-0.000***	-0.000***
	(1.83)	(4.04)	(2.85)	(2.19)	(3.58)	(3.15)
potato yield	-0.856	0.532	0.311	-0.856	0.532	0.311
	(0.27)	(0.15)	(0.11)	(0.28)	(0.15)	(0.10)
precipitation	-0.000	-0.002***	-0.001**	-0.000	-0.002***	-0.001***
	(0.91)	(3.70)	(2.65)	(0.95)	(3.97)	(2.78)
rice yield	-1.334	-3.281	-2.882	-1.334	-3.281	-2.882
	(0.63)	(1.34)	(1.31)	(0.51)	(0.90)	(1.03)
temperature	0.001	0.001	0.001	0.001	0.001	0.001
	(0.83)	(1.37)	(1.14)	(0.93)	(1.64)	(1.38)
wheat yield	0.150	0.023	0.097	0.150	0.023	0.097
	(1.06)	(0.09)	(0.46)	(1.11)	(0.10)	(0.55)
birth male	-1.333			-1.333		
	(1.32)			(1.68)		
Bulgarians/Turks	-0.068*	-0.063	-0.060	-0.068*	-0.063	-0.060
	(1.87)	(1.23)	(1.56)	(2.01)	(1.10)	(1.55)
refugee female literacy		0.081**			0.081*	
		(2.44)			(2.02)	
birth female		-2.994**			-2.994**	
		(2.56)			(2.59)	
refugee total literacy			0.107**			0.107***
			(2.62)			(3.16)
birth total			-2.005*			-2.005**
			(1.97)			(2.18)
constant	0.381***	0.242	0.307**	0.381***	0.242**	0.307***
	(3.34)	(1.48)	(2.41)	(4.28)	(2.16)	(3.40)
R^2	0.61	0.67	0.63	0.61	0.67	0.63
Observations	128	126	128	128	126	128

Notes: OLS estimations with male, female and total literacy rates as dependent variables in Greek provinces in 1928. Fixed effects at department level and cluster standard errors at department level (Columns (1), (2), (3)) and at prefecture level (Columns (4), (5), (6). Some coefficients are zero because the values are rounded off to three decimal points. Refugees included. *, **, *** Significant at 10%, 5%, 1% respectively.

Table 25

OLS Estimations of Literacy Rates on Land Ownership in Greek provinces in 1928

	(1)	(2)	(3)	(4)	(5)	(6)
	male literacy	female literacy	total literacy	male literacy	female literacy	total literacy
refugee male literacy	0.065			0.065		
	(1.37)			(0.96)		
land ownership	-0.062***	-0.051*	-0.042*	-0.062***	-0.051	-0.042**
	(4.68)	(1.88)	(2.17)	(3.32)	(1.60)	(2.54)
land area	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.73)	(1.63)	(1.01)	(0.67)	(1.67)	(1.10)
pop density	0.000	0.000	0.000	0.000	0.000	0.000
	(1.11)	(1.03)	(0.83)	(1.31)	(1.17)	(1.02)
urbanization	0.087***	0.169***	0.134***	0.087***	0.169***	0.134***
	(3.34)	(3.69)	(4.36)	(3.24)	(3.84)	(4.22)
altitude	0.000	0.000	0.000	0.000	0.000	0.000
	(0.94)	(0.98)	(0.58)	(0.89)	(0.65)	(0.50)
ruggedness	0.000	0.000	0.000	0.000	0.000	0.000
	(1.25)	(0.07)	(1.03)	(0.97)	(0.05)	(0.83)
olive yield	-0.026	-0.057	-0.020	-0.026	-0.057	-0.020
	(0.12)	(0.27)	(0.10)	(0.13)	(0.26)	(0.10)
pasture yield	-0.000	-0.000**	-0.000	-0.000*	-0.000	-0.000*
	(1.46)	(2.29)	(1.55)	(1.85)	(1.45)	(1.71)
potato yield	1.939	1.455	1.546	1.939	1.455	1.546
	(0.29)	(0.23)	(0.25)	(0.32)	(0.21)	(0.26)
precipitation	-0.000	-0.002**	-0.001*	-0.000	-0.002**	-0.001*
	(0.86)	(3.07)	(1.87)	(0.84)	(2.64)	(2.00)
rice yield	3.677	-2.103	0.692	3.677	-2.103	0.692
	(0.94)	(0.35)	(0.16)	(0.99)	(0.22)	(0.13)
temperature	0.000	0.001	0.001	0.000	0.001	0.001
	(0.46)	(1.16)	(0.97)	(0.30)	(1.06)	(0.82)
wheat yield	0.120	0.018	0.038	0.120	0.018	0.038
	(0.41)	(0.04)	(0.11)	(0.45)	(0.05)	(0.14)
birth male	-2.258*			-2.258**		
	(2.12)			(2.16)		
Bulgarians/Turks	-0.091**	-0.073	-0.066	-0.091**	-0.073	-0.066
	(3.15)	(1.27)	(1.79)	(2.61)	(1.04)	(1.45)
refugee female literacy		0.085**			0.085*	
		(2.30)			(2.03)	
birth female		-2.883			-2.883*	
		(1.65)			(1.75)	
refugee total literacy			0.089***			0.089***
			(3.54)			(3.30)
birth total			-2.429			-2.429*
			(1.73)			(1.88)
constant	0.532***	0.352*	0.411***	0.532***	0.352**	0.411***
	(5.75)	(2.06)	(3.64)	(4.34)	(2.14)	(3.55)
R^2	0.76	0.80	0.78	0.76	0.80	0.78
Observations	128	126	128	128	126	128

Notes: OLS estimations with male, female and total literacy rates as dependent variables in Greek provinces in 1928. Fixed effects at prefecture level and cluster standard errors at department level (Columns (1), (2), (3)) and at prefecture level (Columns (4), (5), (6). Some coefficients are zero because the values are rounded off to three decimal points. Refugees included. *, **, *** Significant at 10%, 5%, 1% respectively.

Table 26

Regression Adjustment Model in Greek provinces in 1907 for literacy rates

	(1)	(2)	(3)
	male literacy	female literacy	total literacy
land ownership(ATE)	-0.017	-0.029	-0.003
	(0.99)	(1.60)	(0.23)
land ownership (POmean)	0.485***	0.185***	0.327***
	(48.08)	(11.92)	(37.06)
Observations	69	69	69

Table 27

Regression Adjustment Model in Greek provinces in 1928 for literacy rates

	(1)	(2)	(3)
	male literacy	female literacy	total literacy
land ownership(ATE)	-0.026	-0.058***	-0.037**
	(1.62)	(3.18)	(2.44)
land ownership (POmean)	0.622***	0.342***	0.479***
	(78.34)	(34.45)	(63.33)
Observations	135	135	135

Table 28

Nearest-Neighbor Matching Model in Greek provinces in 1928 for literacy rates

	(1)	(2)	(3)
	male literacy	female literacy	total literacy
land ownership(ATE)	-0.050**	-0.074***	-0.058***
	(2.00)	(3.21)	(2.58)
Observations	135	135	135

Table 29

Regression Adjustment Model in Greek municipalities in 1907 for literacy rates

	(1)	(2)	(3)
	male literacy	female literacy	total literacy
land ownership(ATE)	-0.046***	-0.049***	-0.035***
	(4.21)	(5.31)	(4.10)
land ownership (POmean)	0.471***	0.137***	0.299***
	(84.85)	(23.82)	(62.96)
Observations	437	437	437

Table 30

Nearest-Neighbor Matching Model in Greek municipalities in 1907 for literacy rates

	(1)	(2)	(3)
	male literacy	female literacy	total literacy
land ownership(ATE)	-0.045***	-0.052***	-0.033***
	(3.59)	(5.44)	(3.71)
Observations	437	437	437

Table 31

Propensity-Score Matching Model in Greek municipalities in 1907 for literacy rates

	(1)	(2)	(3)
	male literacy	female literacy	total literacy
land ownership(ATE)	-0.029**	-0.055***	-0.032**
	(1.96)	(3.98)	(2.41)
Observations	357	357	357

Figures

Figure 1

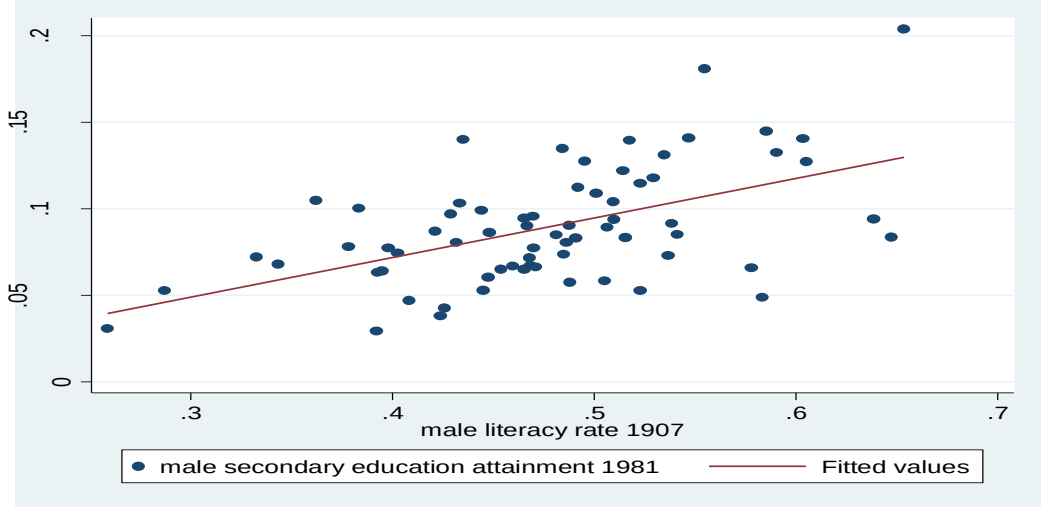


Figure 2

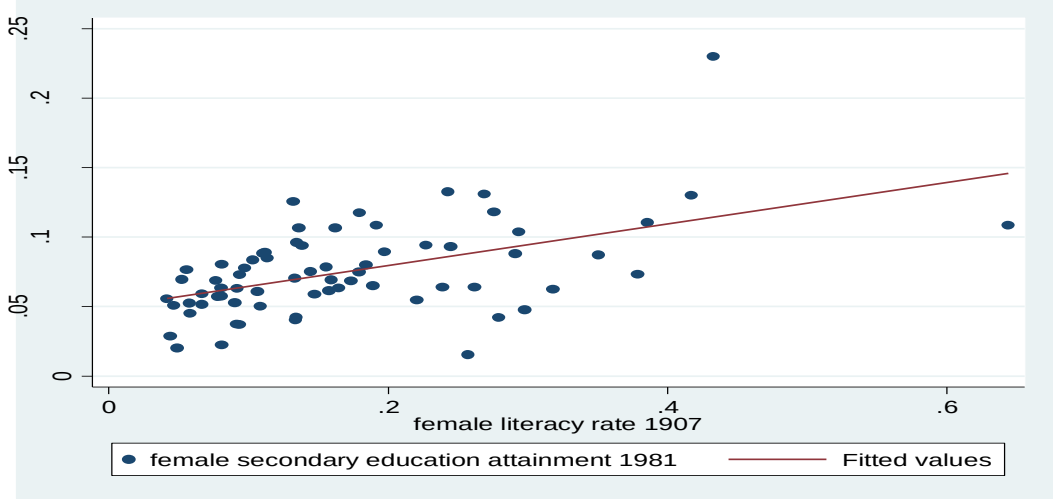


Figure 3

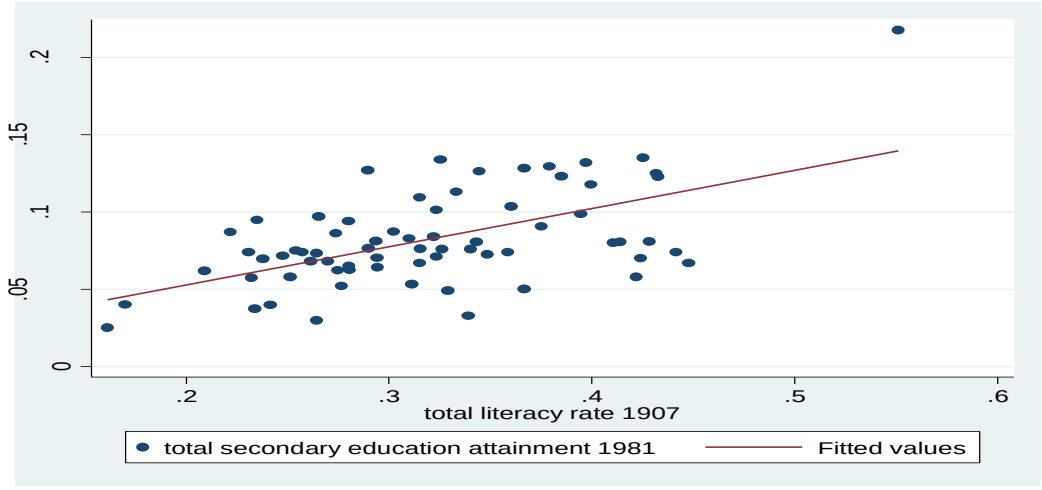


Figure 4

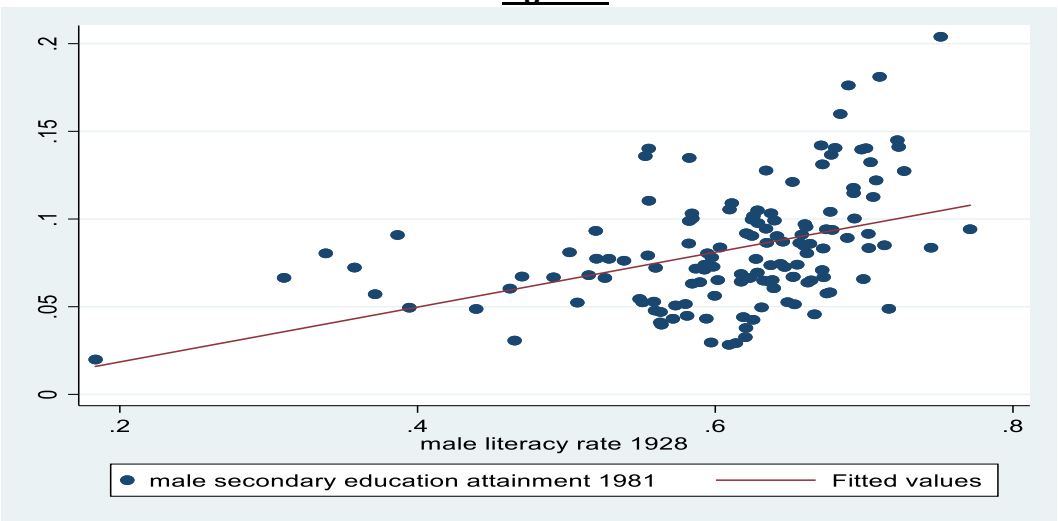


Figure 5

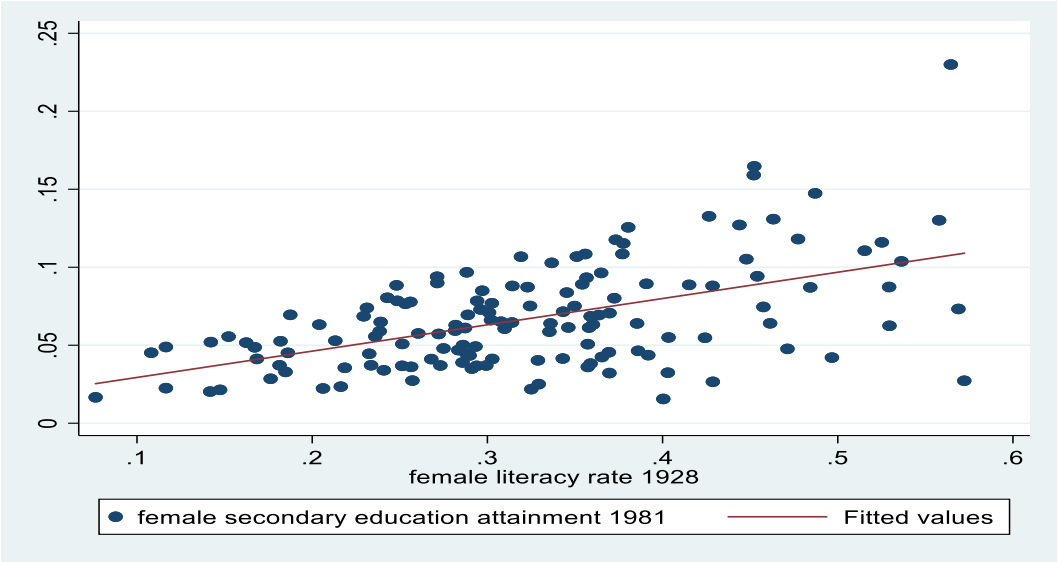


Figure 6

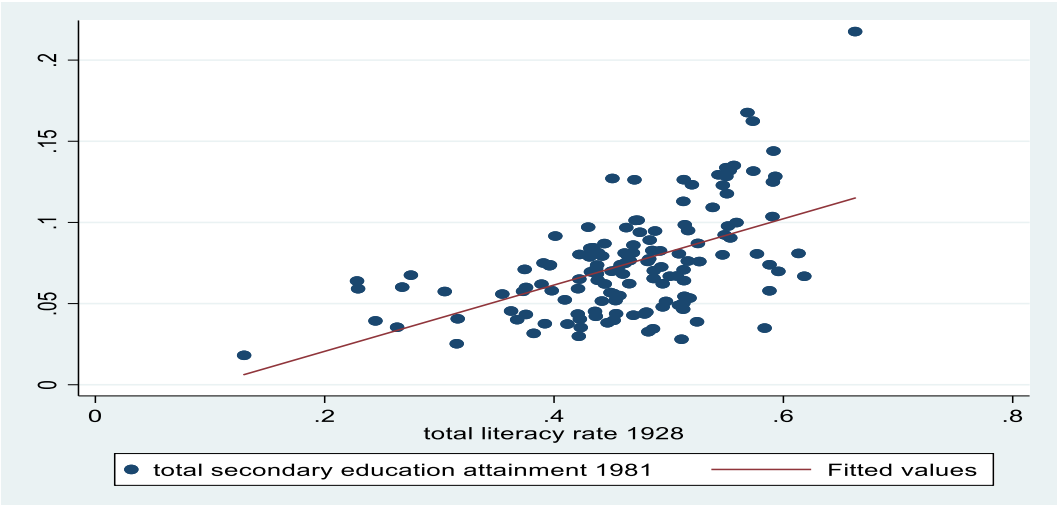


Figure 7

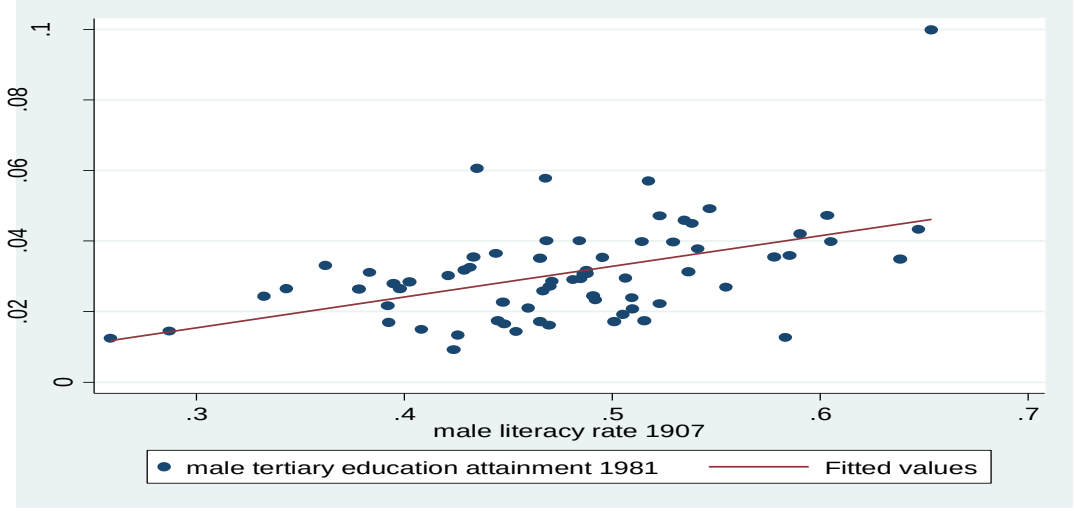


Figure 8

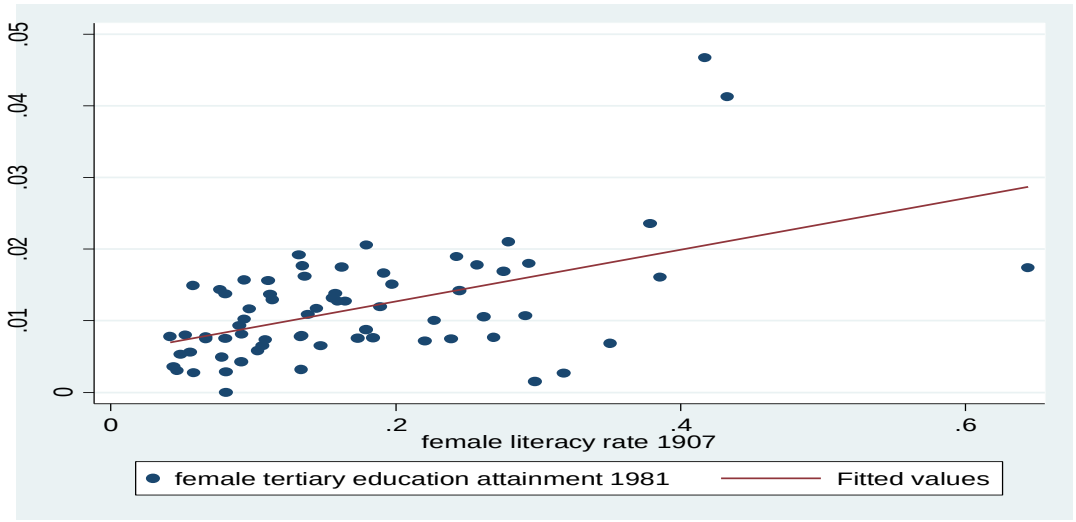


Figure 9

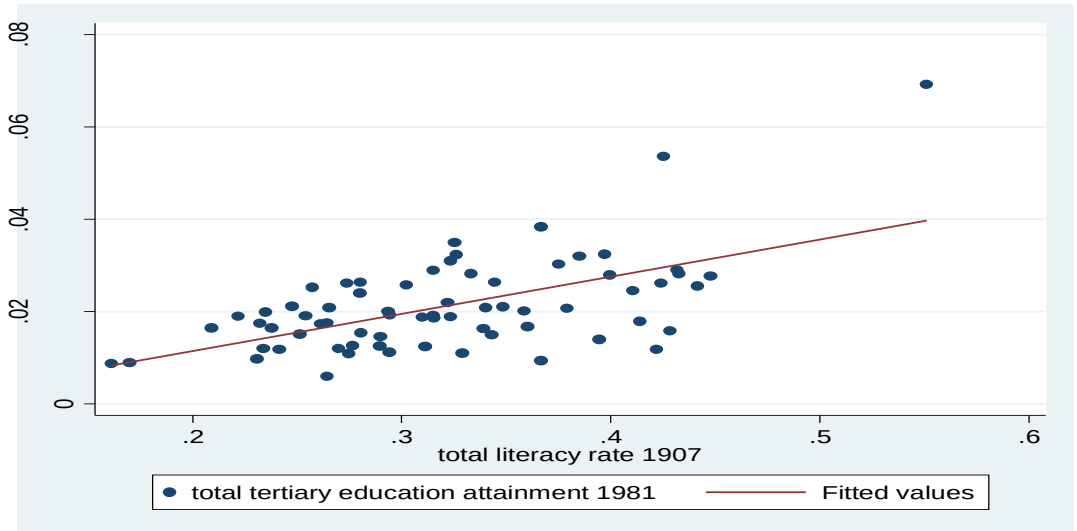


Figure 10

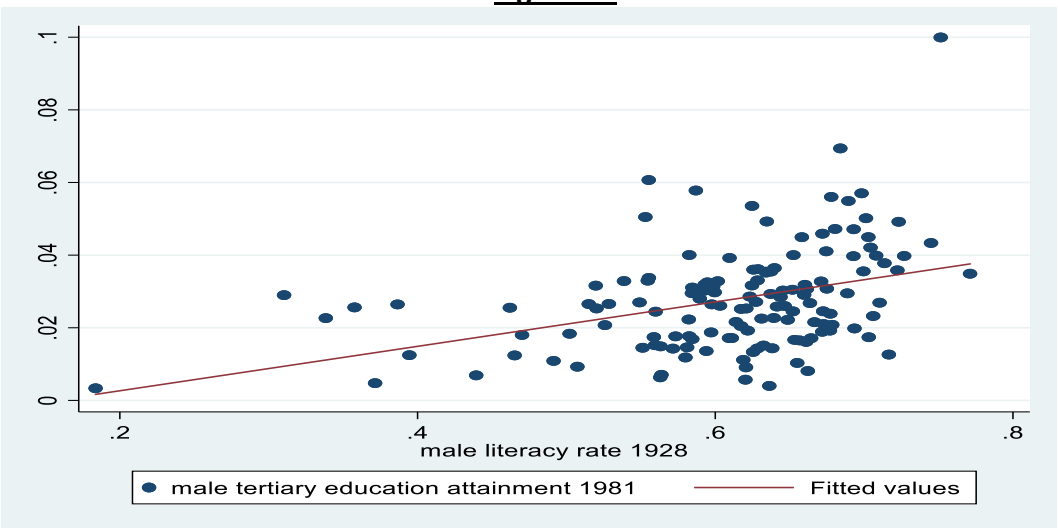


Figure 11

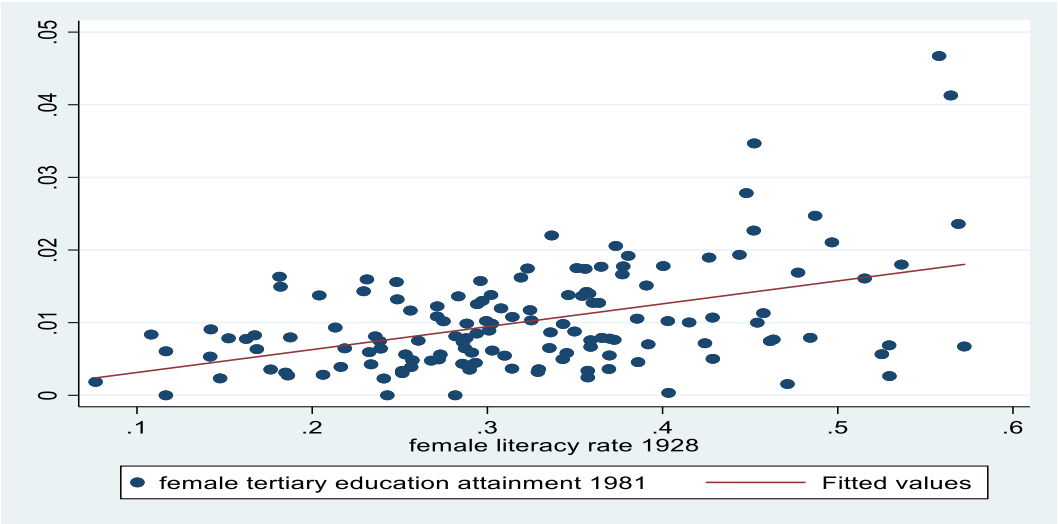


Figure 12

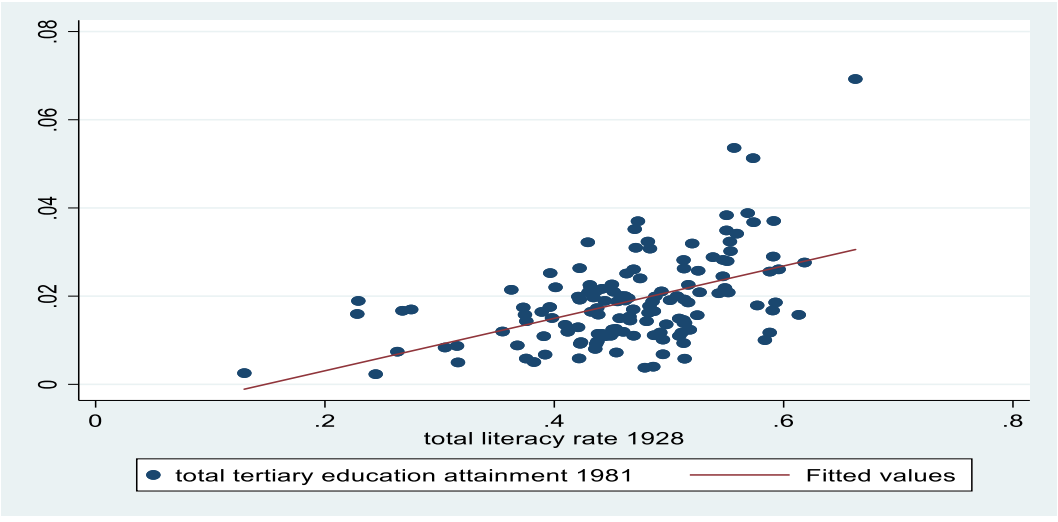


Figure 13

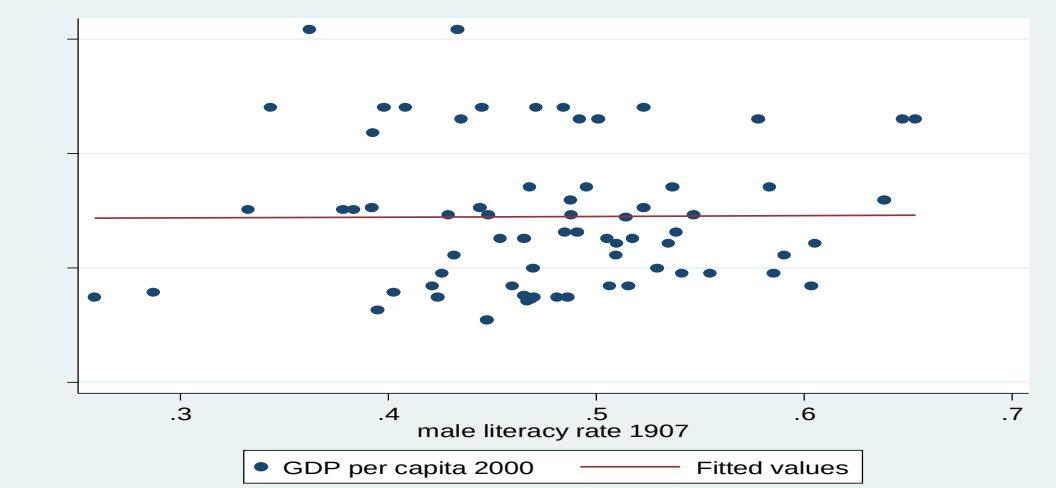


Figure 14

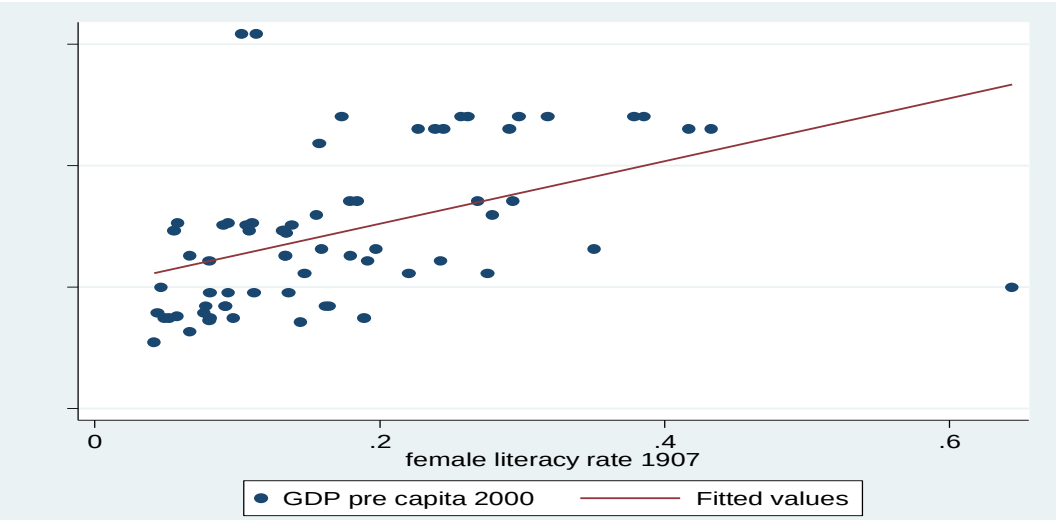


Figure 15

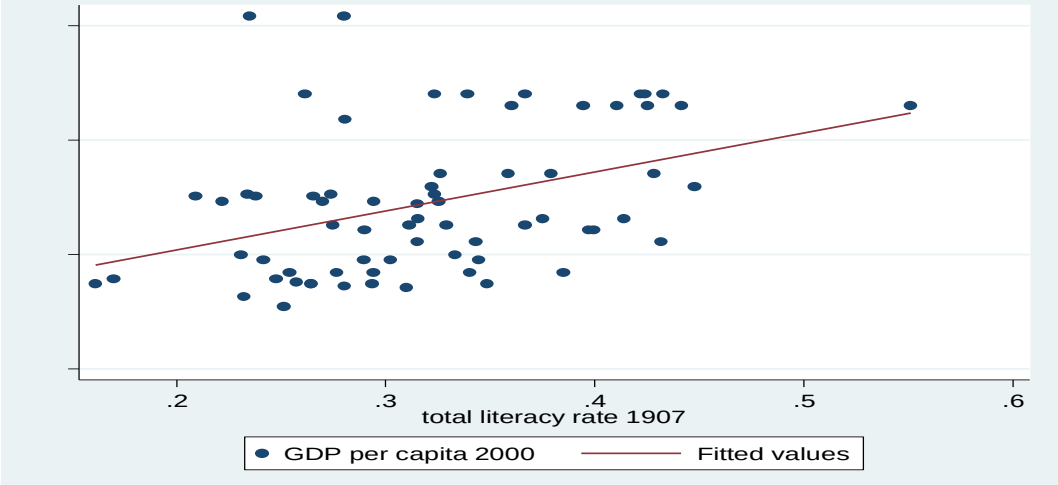


Figure 16

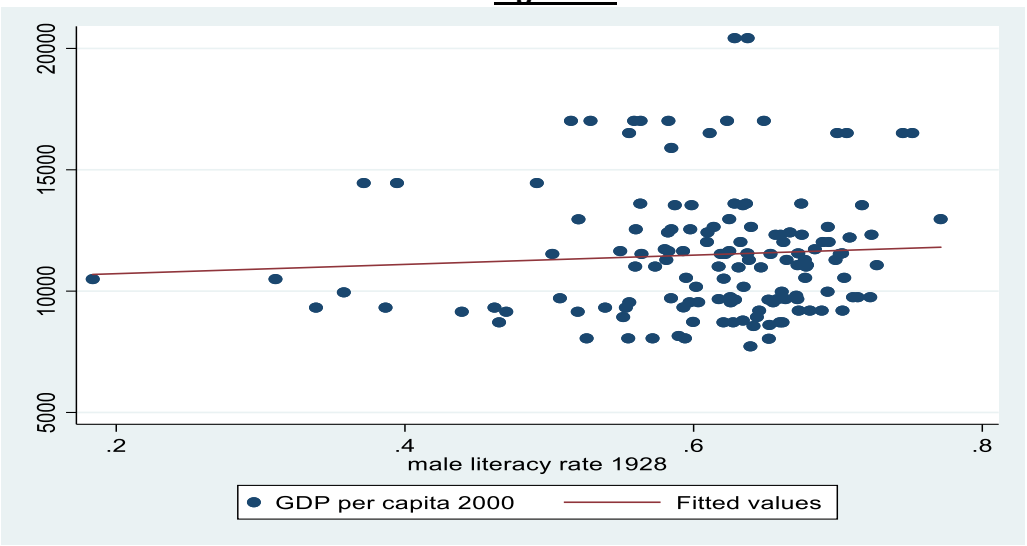


Figure 17

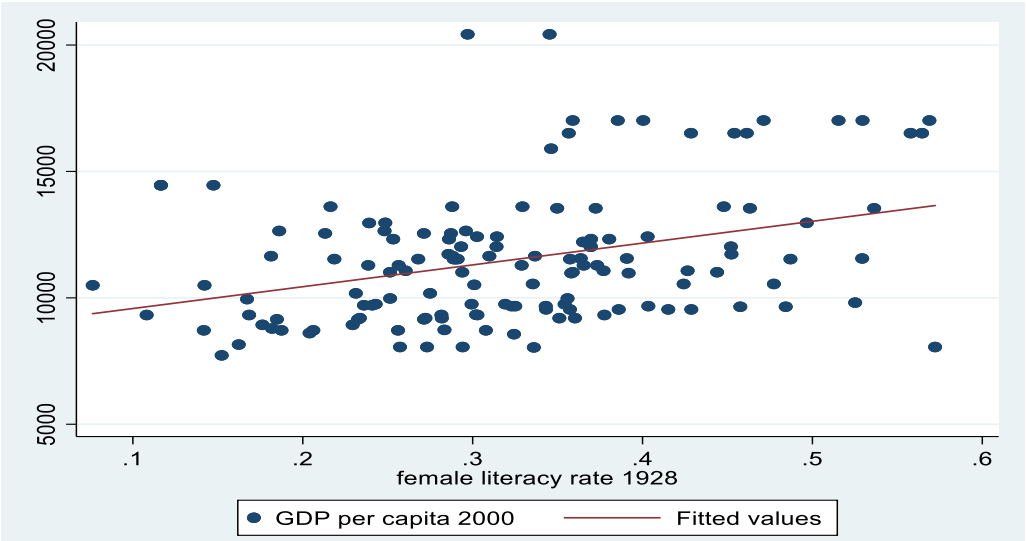
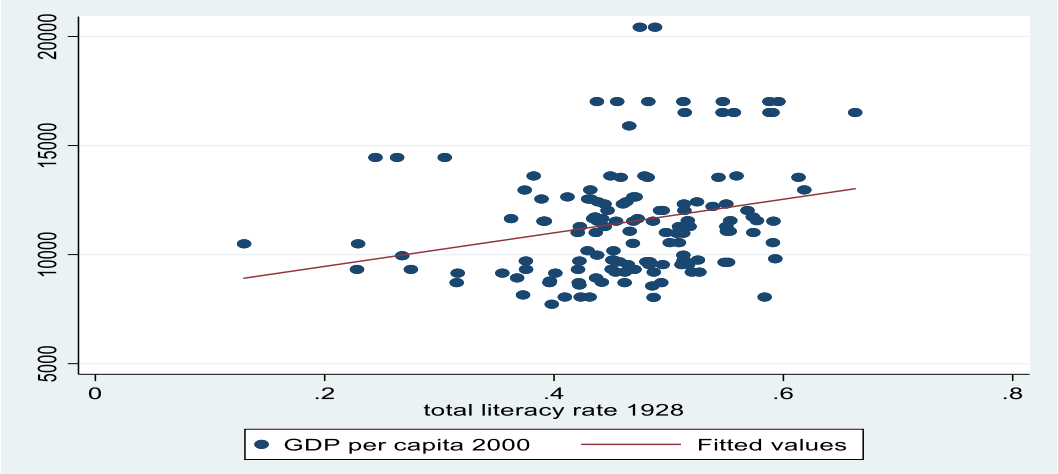


Figure 18



Chapter 2: The Implications of the Population Exchange between Greece and Turkey on Literacy and Employment Rates of the Native Population

2.1 Introduction

This chapter establishes the importance of a major historical event, i.e., the population exchange between Greece and Turkey that was officially agreed in 1923, on the economy of Greece. On the 30th January 1923, a bilateral agreement was signed between Greece and Turkey for the exchange of approximately 1.5 million Greeks and 500.000 Muslims currently living in Greece. Despite the fact that it was the first such agreement to be signed, nevertheless it was still a forced displacement practice that involved millions of people and imposed enormous social and economic cost for the then Greek government.

Based on the existing literature on forced displacement it is a plausible speculation to make that this event had long-lasting implications for Greece. My research project aspires to study the consequences of this event in the short-run (in this chapter) and the long-run (in future work) on the economic and the social domain.

Collecting a very rich dataset from various historical and archival sources, I compile a dataset of 141 provinces (the totality of provinces at the time period under examination) for which I have data for two periods, i.e., for the interval 1920 and 1928. My data includes rich information on literacy rates and employment rates at a very disaggregated level, i.e., by gender, by status (refugee or native), by age groups (for literacy), by sector (for employment) and all available at the province

level (as well as at the municipal level). I also have data on various other economics and social aspects of Greece at the time, which allows me to account for potentially confounding factors.

Besides the unique historical dataset, I also exploit this "natural experiment" to implement a differences-in-differences approach. Using as a treatment the difference in the fraction of refugee population in a province before and after the exchange, I explore how this massive inflow has affected literacy and employment.

Overall, my findings suggest that the presence of the refugee population has led to an increase in literacy population for both men and women. However, the effect on men is much stronger, both qualitatively and quantitatively, and concerns most age groups. Whereas the effect on women is significant for fewer age groups (especially the ones around 14-25) and is also less economically significant. This perhaps reflects the inferior position of women with respect to literacy.

Concerning employment, I find that upon the arrival of refugees, the male native population turned away from agriculture to the industry, transportation, stock farming and trade sectors. This could reflect either the fact of freed resources of the native population or their preference for more skilled jobs now that the refugee population, which was also a cheaper source of labour, was employed in agriculture. The latter could also account for increased rates of literacy. As to women, there was no immediate effect on the female native population.

In all the analysis, it is essential to remember that my period of examination is only 8 years. Therefore, any effects that I capture are immediate effects that could

be further magnified in the years to come, which is also part of my future research agendas. I thus anticipate that what I report is the lower bound of the estimated effects.

Hypothesizing about the potential mechanisms that could explain this reduced form effect I come up with various potential explanations that remain to be tested in future work. These are an increase in growth rates of Greece driven by the presence of skilled and cheap labour force, competition between natives and refugees, positive externalities of the presence of the refugee population driven either by them between interaction or the increased government spending to mention only a few.

The remainder of the chapter is structured as follows. Section 2 discusses the related literature. Section 3 provides the historical background. Section 4 discusses the compilation of the data. Section 5 elaborates on the empirical approach and methodology and discusses about the elements that render this historical even a "natural experiment". Section 6 reports the benchmark result. Following, I discuss the associated mechanisms that could explain this reduced form effect. Section 8 concludes.

2.2 Related Literature

My study aspires to contribute to three distinct strands of literature. The literature on the implications of forced displacement, the literature on the historical determinants of education, as well as to elaborate on the implications of this major historian event on the long-run evolution of the Greek economy and society.

Forced Displacement

There is a large literature in economics and demography that analyses the determinants of forced migration, as well as the deep and long-lasting effect on preferences, behavioural and economic outcomes of forced migration. Forced migration is a consequence of wars, civil conflicts, or natural disasters with significant implications for host locations (Ruiz and Vargas-Silva, 2016; Becker and Ferrara, 2019). Each year millions of people are forced to leave their homelands because of violence, conflict or oppression (Ruiz and Vargas-Silva, 2016). Forced migration is the result of forces outside the control of the migrants. It may be temporary, such as when refugees find a transitory abode in a safe country while waiting to return to their home countries, or permanent, as in the case of forced population movements after WWII when European borders were redrawn (Becker and Ferrara, 2019).

Bauer et al. (2013) analyse the mass flight and expulsion of millions of Germans from Eastern Europe to West Germany during and after World War II. Most of these Germans were displaced from the former eastern territories of the German Reich. This expulsion had a huge impact on the population structure of West Germany since about 8 million expellees arrived in the late 1940s. The population share of expellees in West Germany was about 17 per cent in 1950 while it was zero 5 years earlier (Wyrwich, 2020). Expellees were not selected on the basis of their skills or labour market prospects and, as ethnic Germans, were close substitutes to native West Germans. Expellees were forced to relocate and few of them could choose their initial destination in West Germany (Braun and Mahmoud, 2014).

Migrants either fled to regions close to their old homelands or, after the end of the war, were brought to West Germany in compulsory transfers. Once they had arrived in the West, their geographic mobility was severely restricted by law. Moreover, expellees were not allowed to resettle in the French occupation zone in the first years after the War while there was no such legislation in the other occupation zones (USA; UK; Soviet Union) (Wyrwich, 2020).

Another case of forced displacement in the aftermath of WWII is analysed by Becker et al. (2020). Over 2 million Poles were expelled from the Kresy territories of eastern Poland and resettled (primarily) in the newly acquired Western Territories, from which the Germans were expelled. Similarly, Bharadwaj et al. (2015) examine the immediate demographic impact of partition of India in 1947, in which approximately 17 million people displaced within four years into the countries of India, Pakistan, and what became Bangladesh. Murard and Sakalli (2019) examine the effects of the mass refugee inflow after the Greco-Turkish war of 1919–1922. The Greek Orthodox population of the Ottoman Empire was forcibly resettled from their homeland in Turkey to Greece, increasing the host population by more than 20% within a few months. More recently, civil wars and ethno-religious conflicts displaced millions of people and many more refugees are to be expected as environmental change continues.

Labanca (2020) analyses the unique characteristics of the migration to Italy resulting from the Arab Spring, a migration event that was forced by civil unrest. The Arab Spring caused a large spike in immigrants to Italy within a relatively brief period. In the first six months of 2011, the share of immigrants residing in Italy and

originating from the Arab Spring countries of Egypt, Libya, Tunisia and Yemen increased by 23%. In addition, the civil conflict in Syria, which ignited in mid-2011, lead to a forced displacement of Syrian refugees into many neighbours (e.g., Turkey, Lebanon, Iraq, and Jordan) and affected the most host locations (Tumen, 2016).

Forced migration can cause significant short-term and long-term costs for receiving communities but also, there are several beneficial effects of refugees, especially in terms of agglomeration economies. Mass refugee influxes affect labour market outcomes, education and generally economic outcomes of host regions. The inflow of German expellees after World War II significantly reduced employment rates of natives. A 10-percentage point increase in the share of expellees in the labour force decreased the employment rate of natives by 4 percentage points (Braun and Mahmoud, 2014). Braun and Kvasnicka (2014) find that this inflow of German expellees reduced low-productivity agricultural employment and led to higher employment rates of the non-agricultural sector (Bharadwaj et al, 2015-for the case of India). A one standard deviation increases in the 1950 population share of German expellees is associated with an increase in non-agricultural employment by 2.4 percentage points. The forced displacement of Germans also increased blue-collar employment and reduced self-employment among migrants (Bauer et al., 2013). In the case of Poland respondents with ancestors from Kresy are more likely to work in white-collar occupations, and are less likely to be unemployed (Becker et al., 2020). Benos et al. (2021) indicate that refugees living in Greek settlements close to native farming settlements accumulate more education and as a result enter manufacturing more than the nearby natives. In addition, Murard and Sakalli (2019)

find that regions with a higher share of Greek refugees in 1923 display higher level of prosperity and industrialization because of the provision of new agricultural know-how and the transfer of technological knowledge in textile. Also, Hornung (2014) supports that refugee inflows in Prussia led to a long-term productivity increase in the textile sector. Ruiz and Vargas-Silva (2016) support that Tanzanians after the forced refugee inflows had a higher likelihood of working in household shambas or cared for household livestock and a lower likelihood of working outside the household as employees. Particularly, they were less likely to be agricultural employees and also, many of the natives who were casual workers before the shock changed to other activities, including self-employment in the post-shock period.

On the other hand, Bharadwaj and Mirza (2019) suggest that regions with more refugees in India have higher average yields, are more likely to take up high yielding varieties of seeds, and are more likely to use agricultural technologies employment. In the case of the unexpected flows of Syrian refugees into Turkey around 43 percent of natives who lost their jobs as a consequence of refugee inflows stayed unemployed, while the remaining 57 percent left the labour force. Men preferred to stay unemployed, while females chose to leave the labour force (Tumen, 2016). Regarding human capital, several studies provide effects of refugee inflows on education. After the mass expulsion of Germans, the migrant children acquired more education than their native peers (Bauer et al., 2013). Bharadwaj et al. (2015) conclude that partition related flows in India resulted in an increase in literacy rates of natives. Due to higher education levels amongst migrants, districts with 10 per cent greater inflows saw their literacy rates increase by 3 percentage

points, while a 10 per cent increase in outflows reduced literacy by 1.2 percentage points. Similarly, Kreibaum (2016) suggests that Ugandans living near refugee settlements are more likely to have access to primary schools run by NGOs or other private organizations. There is also evidence that descendants of displaced Poles invested more in education as a result of a shift in preferences towards human capital (Becker et al., 2020). Specifically, while there were no pre-WWII differences in educational attainment, Poles with a family history of forced migration are significantly more educated today than other Poles.

My study contributes to the empirical historical literature discussing about how Greece evolved after the Greco-Turkish war of 1919–1922, in the short run. This historical forced migration led to substantial territorial adjustments and large-scale population movements in the wider region. I exploit a quasi-natural experiment and focus on the impact of refugee inflows on the total natives but also on the basis of gender. My results provide that refugee inflows, as a historical determinant of human capital and employment, increase all native literacy rates but on the other hand, there is no significant impact on employment except for male industrial and agricultural employment.

Long-run Determinants of Human Capital Formation

The role of human capital formation has been identified as the origin of the differential timing of the transition from stagnation to growth (e.g., Galor et al., 2009; Ashraf et al, 2020). According to historical studies several factors seem to influence the formation of human capital. Forced migration, landownership, religion, culture, urbanization, climatic factors and geographic factors have significant effects

on education formation. According to Evangelidis (1936), education did not develop mainly in places with moderate climate, geography favourable to human settlement, or suitable for cultivation. Higher climatic risk that fosters institutions may lead to higher literacy rates in regions with stronger past climate variability, to the extent that local authorities were partly responsible for the education system (Buggle and Durante, 2021).

Other historical studies suggest that religion and culture influence education outcomes. Becker and Woessmann (2009) find that the key channel which can account for the whole association between Protestantism and economic prosperity is the acquisition of literacy, in late 19th-century Prussia. Botticini and Eckstein (2007) find that change in the religious norm at the beginning of the first millennium, which made every male Jewish child learn Hebrew for purely religious purposes, brought long-term economic returns in the form of general education. The Judaism religion made literacy and education the main requirement for belonging to the Jewish community.

In terms of landownership as historical determinant of education, Galor et al. (2009) provide that inequality in landownership had an adverse effect on human capital promoting educational institutions (e.g., public schooling and child labour regulations) in the United States, using variation in the distribution of landownership and educational expenditure, across states, for 1900-1940. Return on land declines as wages of workers rise due to higher education that individuals obtain, while educated workers have stronger incentives to migrate to industrial areas than less educated ones. Landowners inhibit human capital accumulation through the supply

side. Alternatively, Cinnirella and Hornung (2016) and Ashraf et al. (2020) propose that demand is the mechanism behind the link between land inequality and human capital via labour coercion. This is because serfdom introduces an expropriation risk of the returns to human capital investments, discouraging peasants from acquiring education. In countries characterized by unequal distribution of landownership, land abundance led to under-investment in education, an unskilled-intensive industrial sector, and slower growth. Deininger and Squire (1998) find that large landownership is negatively correlated with education, while high levels of education lead to more investment. On the other hand, Sokoloff and Engerman (2000) show that a higher degree of inequality in landownership distribution was reflected in lower human capital investment. Erickson and Vollrath (2004) and Wegenast (2009) argue that high concentration of landownership across agricultural populations leads to low levels of education. Moreover, Beltrán Tapia and Martínez-Galarraga (2015) confirm the negative relationship between the fraction of farm laborers and literacy rates. Regarding urbanization as a factor of education formation, several studies support that urbanized region incur low education expenditures per child due to high population intensity, so this fosters educational attainment (Deininger and Squire, 1998; Galor et al., 2009; Baten and Hippe, 2017).

My Greek historical analysis contributes to the previous studies by analysing the impacts of the Greek forced displacement in 1922 on the literacy rates of natives and suggests that the arrival of refugees increases both male and female native literacy rates.

The Evolution of Greek Society and Economy

A significant part of previous literature focuses on the historical analysis of several issues related to Greece historically. A number of historical characteristics make the case of Greece unique both from a theoretical and an empirical point of view. Kammas et al. (2021), analyse the transformation of the agricultural Greek economy following historical exogenous shock of the increased demand of currants in the second half of the 19th century. During this period Greece was a typical agrarian economy with more than 76% of the workforce employed in agriculture, faced an incredible rise in the international demand for currants. Their findings support that higher exposure of the agricultural population in the shock, had a negative effect on the formation of human capital.

Kammas and Sarantides (2020) explore whether the extension of the voting franchise in the Greek agrarian economy, was the driving force behind the shift in the implemented tax policy. After the independence of Greece, the major priority of the Greek governments was the legitimization of their authority. The Greek governments in 1864 changed the structure of tax system in order to meet the preferences of the electorate, which mainly constituted of peasants and farmers who were illiterate during that period. As a result, the authorities reduced the share of direct to indirect taxes to satisfy them. Murard and Sakalli (2019) analyses the historical shock after the Greco-Turkish war of 1919–1922, when 1.2 million Greek Orthodox population of the Ottoman Empire were expelled from their homeland in Turkey to Greece, increasing the host population by more than 20% within a few months. Greece as an agrarian economy with most of its population living in rural

and semi-urban areas could not afford this large refugee inflow because of limited resources. As a result, the League of Nations intervened and founded the Refugee Settlement Commission (RSC) in September 1923. The RSC had full authority over the distribution of funds and resettlement of refugees. They argue that regions with a higher share of Greek refugees in 1923 display higher level of prosperity and industrialization sixty years after the event. By the end of nineteenth century, many European countries successfully transformed into industrial economies but for Greece a number of factors impeded from such transition (Chatziiosif, 1993). First, cultural and political forces didn't allow the adaptation of a market-based model, giving a central role to the state (Zolotas, 1926). Second, the country was experiencing a severe lack of skilled human capital, because human resources were mainly concentrated in agricultural production during the early days of the New Greek state (Dertilis, 1984). Finally, the small internal market did not facilitate the development of domestic industry. After 1922 Greece entered the new era of industrialization and specifically after the beginning of WWII when the country had drastically enlarged geographically, the structure of the economy was really starting to change (Louri and Minoglou, 2002), This spectacular boost, combined with the development of railways and the rapid monetization of the economy, created the fundamental preconditions for the development of indigenous industry (Mouzelis, 1978).

A more recently historical Greek issue analysed by Efthyvoulou et al. (2020), explores gender political differences during the transition from a rural economy to an industrial economy between 1953 and 1954. Specifically, their analysis identifies

the causal relationship between women's enfranchisement and party vote shares. During the decades of 1950s and 1960s, a large share of population moved from the countryside to the cities and as a result there was an increase in the number of small and medium-sized firms in the industrial sector. On the other hand, female labour force participation was affected negatively by this increase in urbanization.

My empirical analysis endows to these historical Greek studies by investigating the historical forced displacement after the Greek-Turkish War of 1920-1922 and the impacts of this war on human capital and employment.

2.3 Historical Background

2.3.1 The Turkish-Greek War of 1920–1922

The Turkish-Greek War of 1920–1922 marked the last phase of the historical confrontation between the two nations and ended a long history of substantial territorial adjustments and large-scale population movements in the wider region. The direct result of the Greek military defeat in August 1922 was the abrupt forced displacement of over a million civilians. In the autumn of 1922, hundreds of thousands of Anatolian Greeks fled the region for Greece. Their number was soon augmented by the arrival of Thracian Greeks (more than 250,000 persons) who evacuated Eastern Thrace (Psomiades, 1968). Also, earlier violent conflicts in 1907 that took place mainly in Bulgaria had resulted in similar population movements towards Greek territories (Gounaris, 1989). Overall, the Greek territory almost doubled from 63,606 sq. km in 1912 to 121,794 sq. km at the end of 1922.

The precise number of the Orthodox refugees who entered into Greece is difficult to estimate because reliable statistical records do not exist (Kontogiorgi, 2006). According to the general population Census of 1928, the total number of refugees was 1,221,849 persons, of which 673,025 were urban refugees and 578,824 rural. As a result, the population of Greece, which was around 2,530,000 at the beginning of the first Balkan War (1912) nearly doubled by the end of 1922, reaching 4,730,000 just before the mass refugee inflow (Pallis, 1928).

2.3.2 Refugees: Conditions on Arrival and Allocation over Greece

The wave of refugees fleeing Turkey and landing in Greece at the end of 1922 caused extraordinary pressure on the local economy and, given the end of an almost 20-year period of conflicts, the country's needs for housing, feeding and medical care immediately skyrocketed. The shortage of financial resources, the lack of public infrastructure and the enormous number of refugees were the main issues to be addressed, in an era where the political sentiment was rather conservative in Greece. The immediate needs of the refugees were tackled by the Hellenic Ministry of Health and Social Protection and foreign NGOs (American Red Cross, American Near East Relief, Swedish Red Cross, YMCA, Union de Secours aux Enfants among others) (see Pentzopoulos, 1962).

Much of the settlement of the refugees was accomplished through the work of the Refugee Settlement Commission (RSC), which in exchange of the securing a large loan was given significant powers from the Greek government. The purpose of the settlement programme was to help refugees to become self-supporting in the minimum required time (Kontogiorgi, 2006). The work of the Commission can be

divided into two main categories: that of agricultural colonization and that of urban settlement (Pentzopoulos, 1962; Krimpas, 1999). The settlement of refugees depended on a number of factors, namely the availability of land, the supply of housing and the associated need for occupancy and employment opportunities in agriculture in rural areas – mainly Macedonia and Thrace. The Commission considered the settlement of the population in rural areas of primary importance for two reasons: a) to increase agricultural production as rapidly as possible so as to satisfy the needs of incoming refugees; b) to avoid the congregation in the towns of large segments of the refugee population, which would very rapidly become assimilated with the urban population, and would find it difficult to leave the towns for the countryside.

The relocation of refugees in rural and agricultural regions was massive and challenging. The settlement of the refugees in vacant rural areas was crucial because: first, it would help them to become self-supporting very quickly; second, it would fill the demographic vacuum which emerged as a result of the departure of Muslims and Bulgarians especially from Macedonia and Thrace, and, third, it would increase agricultural productivity and contribute to the recovery of the economy of the region, which was impaired by the departure of Muslim and Bulgarian farmers (Kontogiorgi, 2006). Finally, Macedonia and Thrace were lagging in terms of economic development, a successful plan of exploiting their natural resources could sustain the demographic inflow of refugees.

Refugees were disproportionately located in previously thinly populated rural places. Many of these places were filled with extensive marshes, where the local

population was massively hit by malaria. As a result, many refugees got sick and finally died. Overall, the Greek state allocated to the Commission 8.4 million acres of land, originating mainly from properties of Muslims and Bulgarians that had left the country, land expropriated from agricultural reform and monasteries and empty areas close to urban centers. Almost 2,000 such settlements were created, either completely new or attached to already existing villages in Macedonia and Thrace.

On the work of the RSC related to the urban settlement of the refugees, almost 27,000 housing units were built in 118 suburban communities (Giannakopoulos, 1992). The respective urban centers that refugees were settled included Athens, Piraeus, Thessaloniki (cumulatively reaching almost 60%), Volos (Nea Ionia) and, to a smaller extent in Ioannina, Corfu and Patra.

2.3.3 Composition of the Refugee Population

As a result of the mass losses in Greek male population in Anatolia, especially in the productive ages of 18 and 45, the refugees from Asia Minor and Pontus consisted principally of women, children and old men. Thus, many refugee families lacked their male companions and had even greater difficulty sustaining themselves and consequently surviving (Klapsis, 2014). These are supported by the data related to the family status of the refugee population, where the percentage of widowed refugee women was almost 50 per cent higher in comparison to that of the general population in Greece (17.5% versus 12.1%). This disequilibrium was even more severe in the 25-39 age group since the male population of these ages had been most severely affected by the Turkish persecutions: for example, among those aged 25–29, 30–34 and 35-39 the percentage of widowhood among refugee women was

more than double in comparison to that of the general population in Greece (6% vs 2.8%; 12.9% vs 6.3% and 19% vs 10.5%, respectively). In addition, the refugee population suffered high mortality rates as a result of poor living conditions.

2.3.4 Education of Refugees and Natives

Regarding education in Greece, policy formulation and implementation was handled by the central government, but the majority of funding was coming from municipal authorities (Kalafati, 1988). There were very few significant policy changes since the establishment of the first Governor of Greece (Kapodistrias) in 1828 until 1929, when the most important policy reform took place. The Kapodistrias administration organized primary, secondary and professional education in order to reduce illiteracy and provide agriculture as well as light industry with trained labor force. 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.

The landslide victory of the Liberal party marked the most significant turning point in the history of modern Greek education since the formation of the Hellenic state in 1830 (Stefanidis, 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 proposed system intended to direct the predominantly lower middle-class and rural youth towards acquiring practical and professional skills.

The picture outlined above overlooks significant gender differences. In 1870, 28.7% of men and only 6.3% of women were literate. After Athens becomes capital of Greece in 1834, women from the Istanbul took the lead in promoting female education, which was left mainly to private initiative. Female education was organized according to western standards. For instance, most female middle education schools were created by foreigners, mainly missionaries. In Egina, the first public intermediate female school (allilodadaktiko) was created in 1830. Ermoupoli, Egina, Nafplio, Argos were the places with the first female schools. If we consider attendance rates, these were steadily increasing for both genders, reaching 75% for boys but only 33% for girls in 1907 (Kalafati, 1988). The lower attendance of girls is attributed to parents keeping them at home or employing them in agricultural works. Another possible reason is the spatial allocation of schools relative to the allocation of population. Specifically, 76% of the population resided in settlements with less than 2,000 inhabitants and 70% of settlements had no school in late 19th century (Kalafati, 1988). Accordingly, urbanization is positively correlated with female students. Most literate women resided in the province of Attica (29%), Siros (28%) and Tinos (22%). This is line with the fact that only in three provinces 57%-70% of females aged 5-10 attend school (i.e., in the urban centers of Siros, Corfu and Attica where female schools were mainly located), while in thirty-four provinces the attendance rate is less than 10% and in four provinces no female school exists. The gender gap in attendance rates was evident mainly in public schools, which is a manifestation of the fact that the upper class devoted the necessary resources to female education (Kalafati, 1988).

A sizeable gender gap in terms of literacy was still evident in early 20th century. In 1920, just before the arrival of refugees, literacy rate was 56.2% for men and 27.4% for women, despite the fact that Greek state had already mandated primary education of boys, girls up to age 12. After the mass refugee inflow of 1922-1924, the literacy rates were similar for native and refugee males (60.9%, 63.5% respectively), while it was significantly higher for female refugees relative to natives (40.5% vs 32.2%).

2.3.5 Labor Force Participation and Sectoral Allocation for Refugee and Native Population

In terms of employment, there were important gender inequalities, as was the case with literacy. Paid employment was even lower for women than men in Greece during the 19th century, compared to other countries for several reasons (Avdela, 1990). First, due to the structure of Greek economy, its late industrialization and the slow and limited urbanization. Second, the type of employment heavily relied on the class in which native women belonged to as well as on the province of residence. Most women who lived in rural areas worked in agriculture and were occupied within the household. Women living in urban centers and belonging to the middle class were mainly housewives and mothers. Third, working class women were employed as housemates, midwifes, workers or teachers. Overall, in late 19th century, for most women employment and professional development was very unlikely, except for midwifery and teaching. Although the law mandated female education till the age of 12, this was most often perceived as luxury by the rural

population, who would rather utilize females in agricultural or household's duties (Klapsis, 2014).

Just before the refugee inflow in 1920, males and females were mainly employed in agriculture (56.3% and 45.2% respectively). The second most important sector was manufacturing, where 13.8% of men and 24.7% of women were active. Following the mass refugee inflow, there was a sizeable increase in agricultural and a corresponding decline of industrial employment mainly for native females in 1928. Refugee women were forced to work in order to survive with their children. Housewives became servants, laundresses, workers. Young girls worked into factories, stores and any job they could find (Chrysochoou, 1981).

Regarding refugees, males were employed mainly in agriculture (43.4%), industry (25.7%) and commerce (9.7%). Females were occupied mostly in agriculture (45.3%), industry (27.1%) and personal services (12.7%). Interestingly, 83% of the labor force in textile, 72% in tobacco, 71% in clothing industries were comprised by females (Giannakopoulos, 2003). Literate single women found work into offices, stores, businesses. Other women worked in construction along with men (Iordanidou, 1990). Many refugee women were high school graduates speaking Greek, English, French, and knew typing (Chrysochoou, 1981).

2.4 The Data

My research reveals the long-lasting effect of the 1923 population exchange between Greece and Turkey on literacy and employment rates of the native population. The richness of my data and the availability of the pre- and post-exchange period allows me to conduct the analysis at a fine spatial level

(provinces/eparchies), distinguishing between rural and urban regions, industrialized and high-skilled ones and explore how the sudden inflow of refugee population impacted native labour force participation and education.

I therefore build a unique Greek historical dataset, with data derived from the 1929 Census of the Hellenic Statistical Authority at the provincial level for the years 1920 and 1928. The sample consists of 138 provinces for 1920 and of 141 provinces for 1928. This difference between the two periods is due to the fact that in 1920 there were three provinces that were embedded in other provinces (provinces of Arnaias, Sapon, Filippiados_Thiamidos). These three cases broke up into smaller independent provinces in 1928. The 1929 Census of the Hellenic Statistical Authority provides data for both 1920 and 1928 for the same variables, thus enabling a better comparison between the variables. I use this Census because it is an extensive approach to the events that took place since the 1923 population exchange and affected domestic affairs. All the data were manually collected by archival census sources.

I have also collected the corresponding data at the municipal level. As this approach comes at the expense of the number of observations, I use those data only for robustness purposes.

Outcome Variables: Literacy

First, I collect literacy rates (dependent variable) at the provincial level for both natives and refugees by gender. I compute literacy rates by dividing literates with the sum of literates plus illiterates. I use these two periods because they are the first ones including an extended coverage at the provincial level around the 1923 population exchange between Greece and Turkey. My estimations are based on the

literacy of the natives in order to examine the possibility that the 1922-1924 massive refugee inflow caused significant changes on literate natives in areas with high concentration of refugees. Moreover, I use literacy rates by 15 age groups: under 5 years old, 5 years old, 6 years old, 7 years old, 8 years old, 9 years old, 10-14 years old, 15-19 years old, 20-24 years old, 25-29 years old, 30-34 years old, 35-39 years old, 40-49 years old, 50-59 years old and older than 60 years old, to examine if there is a difference in the literacy between the different age groups.

The Attica province, which includes the capital city of Athens, together with the major urban centres of Thessaloniki, Kavala, Volos, Larissa and many provinces in Peloponnesus exhibit the highest male native literacy rates in 1928. The same applies to 1920 with the difference that the province of Ioannina has also high literacy rate. The lowest male native literacy rates are present in Thrace and parts of north Macedonia in 1920 and 1928. The Attica province, urban centres of Thessaloniki, Kavala, Volos, Patra and Aegean Islands exhibit the highest female native literacy rates in 1928. The same applies to 1920 except urban regions of Thessaloniki and the province of Kavala. The lowest female native literacy rates are present in Thrace and parts of Macedonia and Central Greece in 1920 and 1928.

Broadly speaking, urban areas manifest the highest literacy rates for both men and women, while rural areas manifest the lowest rates of literacy.

Outcome Variables: Employment

Second, I include total employment and employment by sector (dependent variables) at the provincial level for both natives and refugees by gender for the years 1920 and 1928. I have collected data for 11 sectors: agriculture, stock-

farming/hunting, fishing, mines, industry, transport/transportation, credit/exchange/mediation, trade, staff services, self-employed and public services.

Moreover, I compute two different employment shares for the 11 sectors: the quotient of sectoral employment with population aged older than 10 years old and the quotient of sectoral employment with total employment. My estimations are based on the employment of the natives in order to examine how the inflow of refugee population impacted male and female labour force participation of natives.

The Attica province exhibits the highest male and female native employment in 1920 and 1928. The lowest male native employment is detected in province Samothraki for the year 1920 and in province Hydra for the year 1928. The lowest female native employment is detected in provinces of Crete for the years 1920 and 1928.

Treatment Variables: Refugee Shares

The main treatment variables of interest are the total refugee share and the difference between total refugee share before 1922 and after 1922. Specifically, I examine the possibility that the 1922-1924 massive refugee inflow induces significant effects on native population both in terms of literacy and employment. I anticipate that this effect might be stronger in areas with high concentration of refugees, therefore I use the share of refugees to construct my treatment variable that I will lengthy discuss below.

I control for externalities between refugees who came to Greece from Asia Minor, Eastern Thrace, Pontus etc. and natives. This investigation is important, because refugees constituted around 20% of population in 1928 and their spatial

allocation was very skewed towards Macedonia, Thrace and Attica, where approximately 87% of them settled. Here I should note that according to historical accounts, the only movements of natives took place from rural to urban areas, because they felt that they would have better employment opportunities in manufacturing and services there (Riginos, 1987).

Control Variables

I enrich the estimated specifications with a set of potential confounders. A first important control is population density and urbanization rates (which I extract from the 1929 Census of the Hellenic Statistical Authority)[1]. Population density is obtained by dividing total population with the land area of each province. Urbanization rate is a proxy for regions with large population. Specifically, it is a ratio calculated by the division of the population of the largest city of each province with the total population in 1920 and 1928. Both controls aim to capture scale and urbanization effects on native literacy and employment.

Another important aspect to capture is what we call "initial conditions" with respect to refugee population in each region. Most of the regions were already populated by a high percentage of Bulgarians and Turks. Those people left the provinces as part of the population exchange. Especially in rural places, the incoming refugee population settled in places and even in residences that this population initially lived. Therefore, Turkish and Bulgarian population is a crucial initial

[1] A crucial aspect of the 1929 Census of the Hellenic Statistical Authority is that it provides information about the same variables in 1920 in order to allow for comparisons. Therefore, the 1920 Census of the Hellenic Statistical Authority is quite often the source of my 1920 data as well.

condition, as it partly reflects past composition and labour force structure of each province as well as potential “infrastructure” for the incoming refugee population. I calculate this share in the 1920 population from the 1929 Census of the Hellenic Statistical Authority. Bulgarians/Turks share is calculated by the division of the total number of Bulgarians/Turks who left Greece with the total population in 1920.

In addition, I include in my analysis four employment control variables derived from the 1929 Census for the periods 1920 and 1928. I use total labour force, the share of labour in agriculture/industry/commerce to account for structural differences of provincial economies in the outcome variables.

2.4.1 Descriptives

Literacy

Figures 1 and 2 show the maps of literacy rates, one of my two main outcome variables, in the full sample of provinces included in my analysis, for the years 1920 and 1928 respectively.

Darker colours are associated with higher literacy rates, for both men and women.

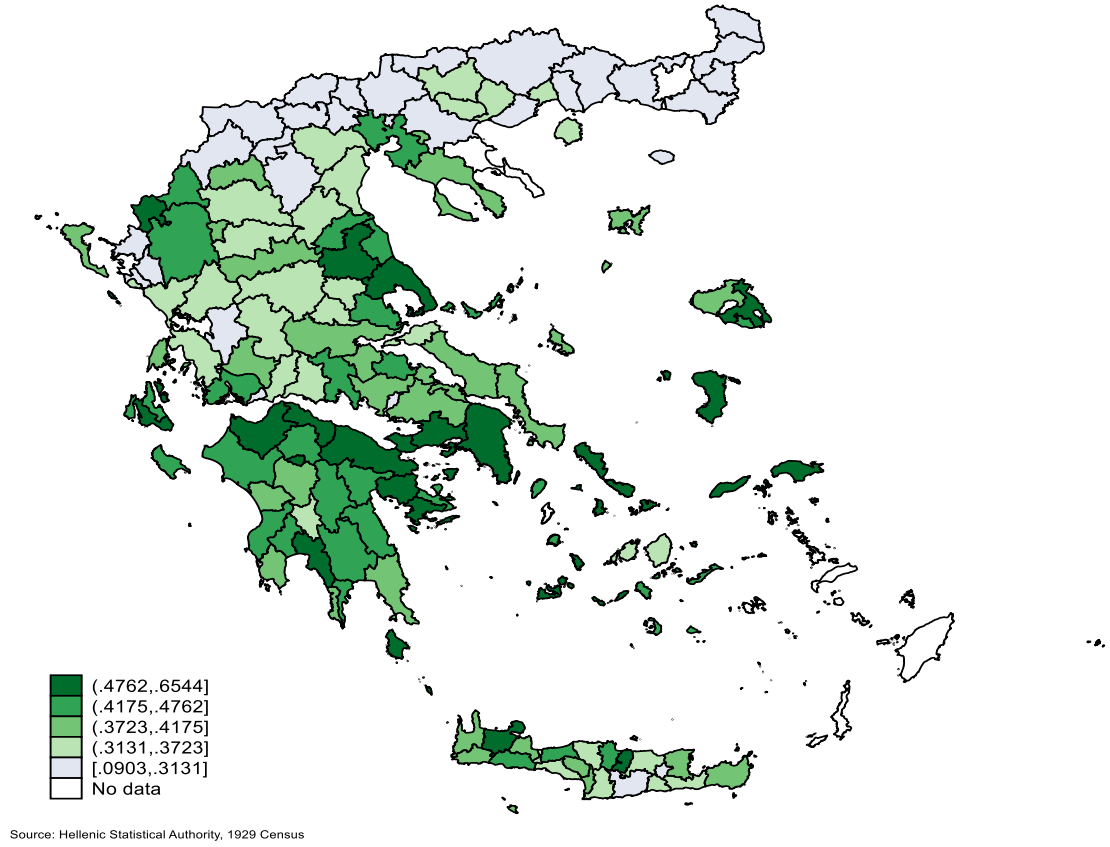


Figure 1: Native Total Literacy Rate in 1920

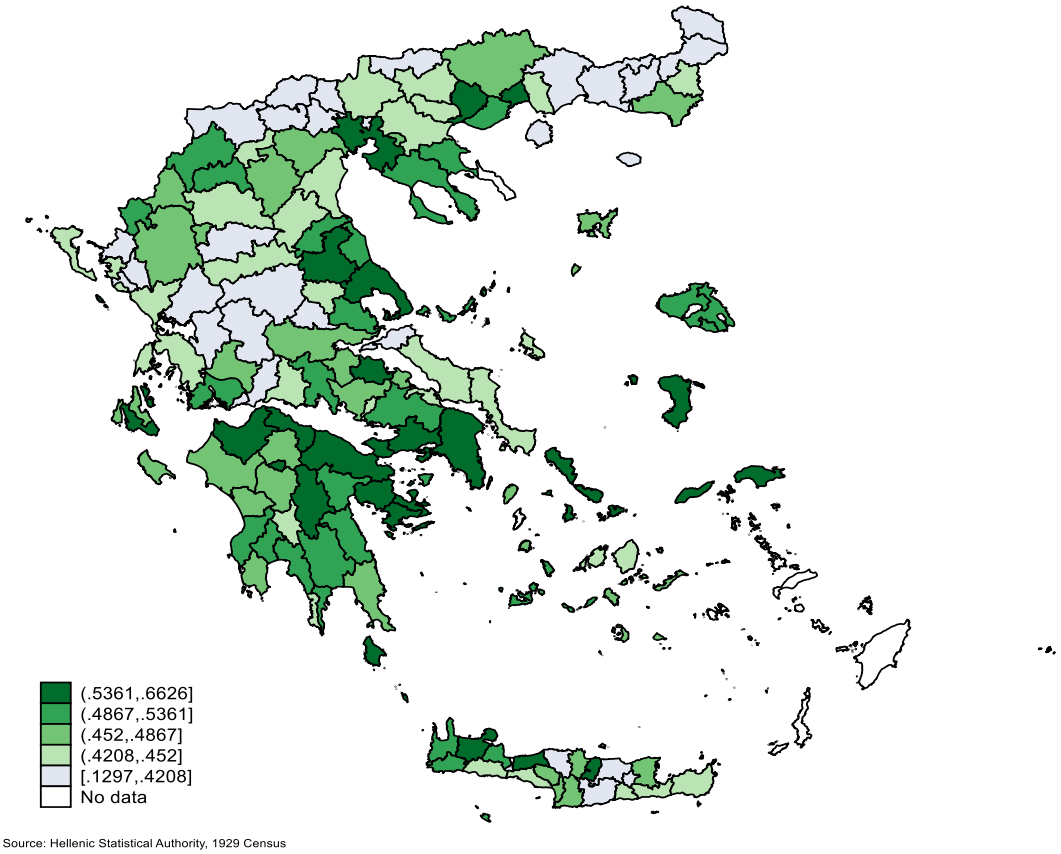


Figure 2: Native Total Literacy Rate in 1928

Comparing the two time periods, I can directly observe that total literacy rate has increased in several areas between the years 1920-1928. It is essential to note that what I illustrate in those figures is the literacy rates of the native population. As such, what I capture is the change that occurred during these eight years in the native population. Given the short time period, and the fact that there was no major institutional change that could have such an impact on the overall literacy rates, I attribute this effect on the high inflow of a largely educated group of refugees from Turkey that had positive spillovers on the native population.

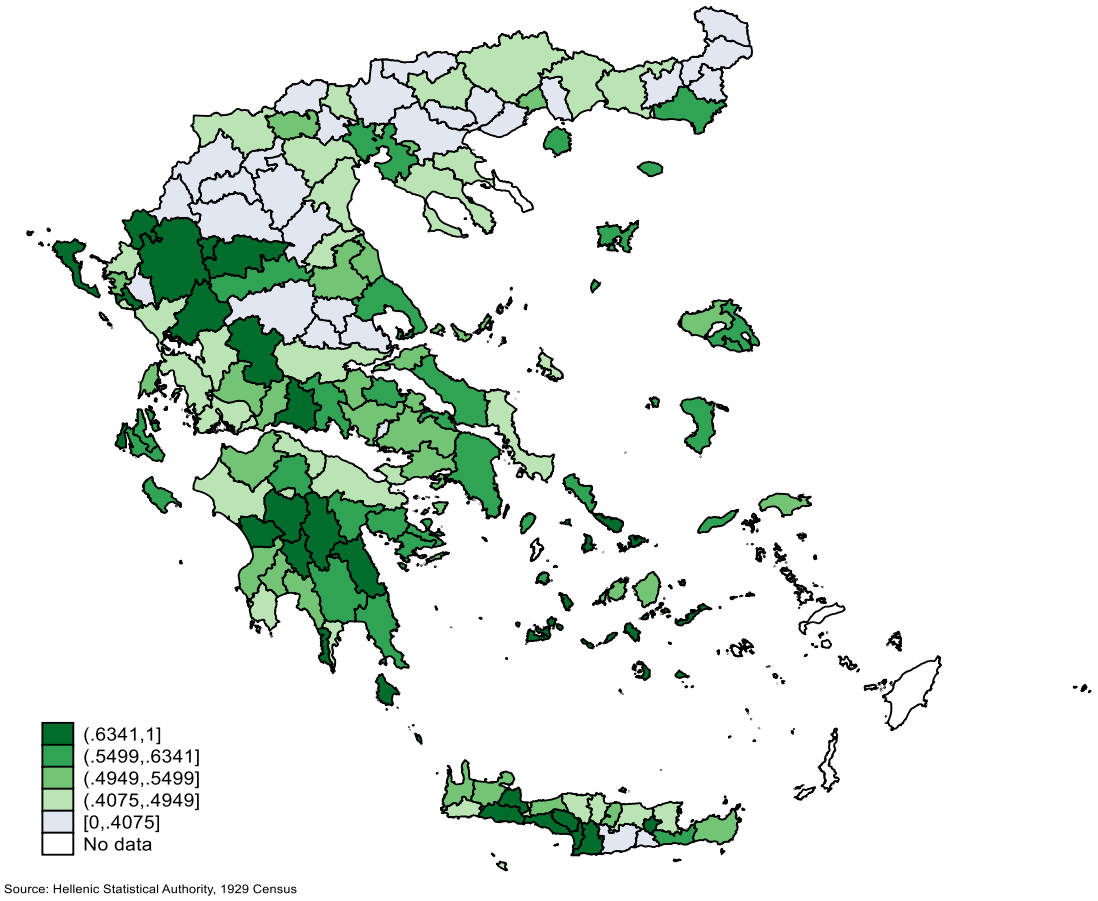


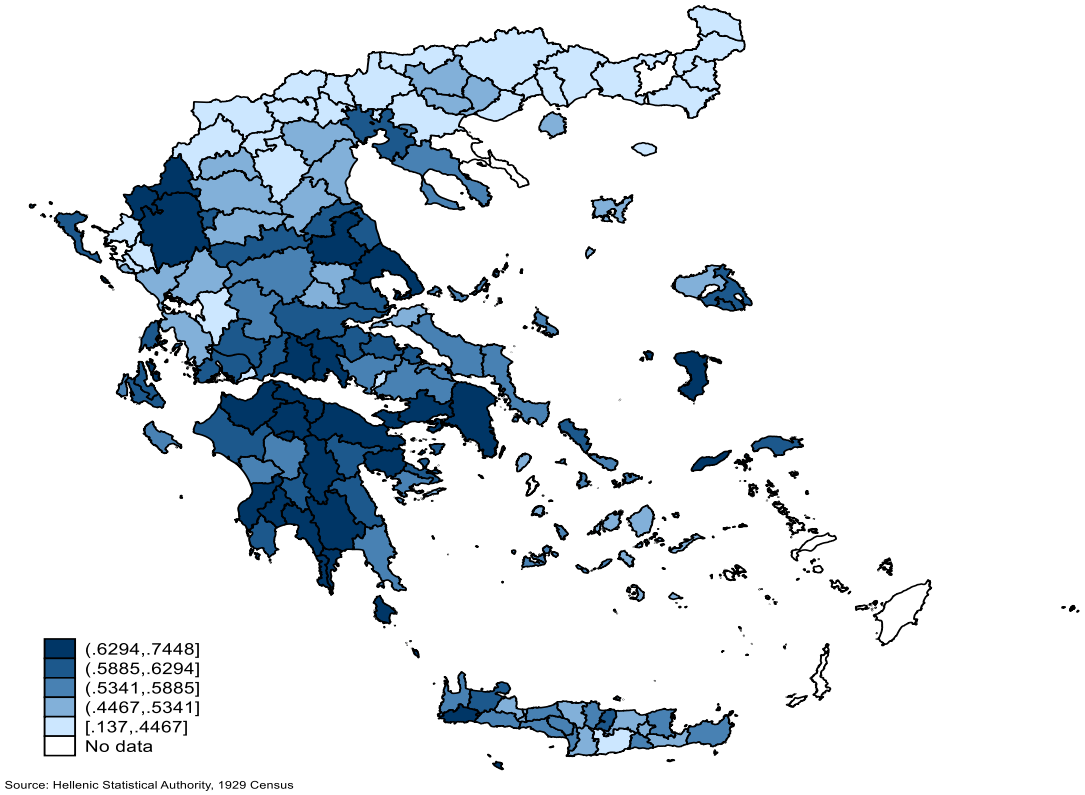
Figure 3: Refugee Total Literacy Rate in 1928

Figure 3 illustrates the literacy rates for the refugee population in 1928. Crucially, I observe that the allocation of refugees was unrelated to their educational

level. The allocation of the refugee population all over Greece was unrelated to their educational skill. The inflow was so massive and the associated expenses to sustain the incoming population so immense that the then Greek government allocated refugees to places where either past settlement from the departing population (as part of the population exchange) existed or where the population could be employed in agricultural activities. This element uncovered in Figure 3, is crucial for my identification strategy, i.e., for the need to preclude that educated refugees clustered into places with highly educated natives.

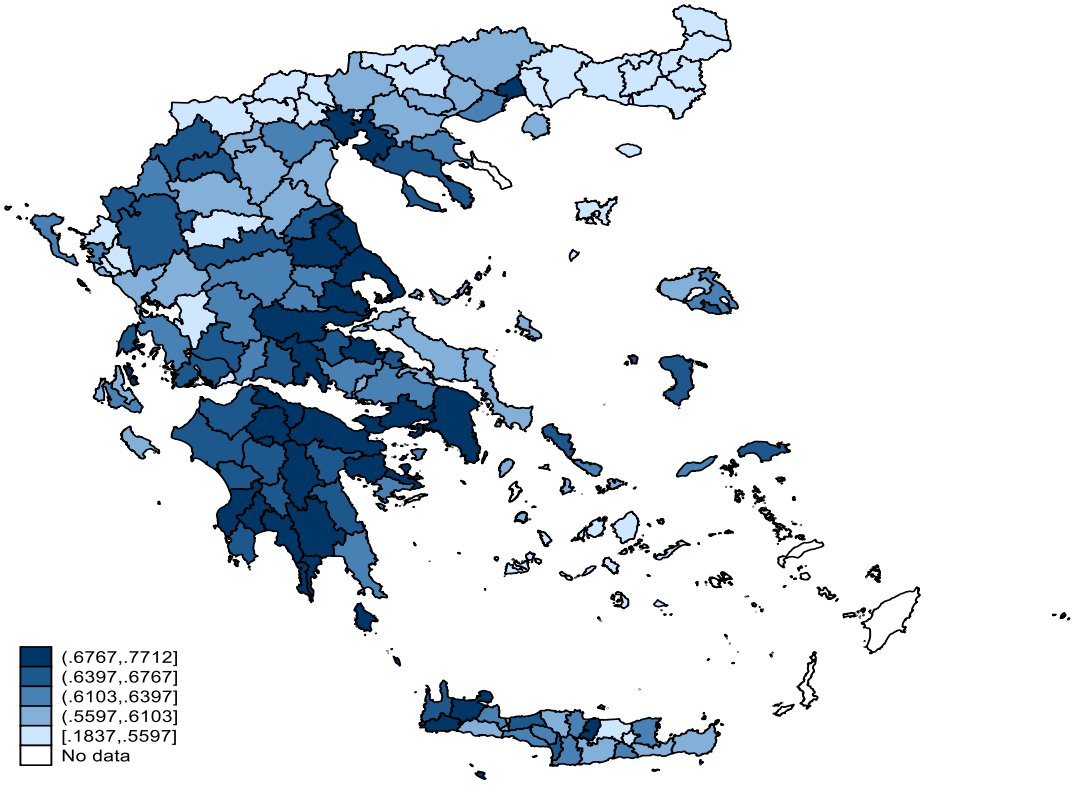
A second crucial element of the data is the decomposition of native and refugee literacy by gender.

Figures 4 and 5 illustrate native male literacy for the years 1920 and 1928 correspondingly, while Figures 6 and 7 do the same for native female literacy.



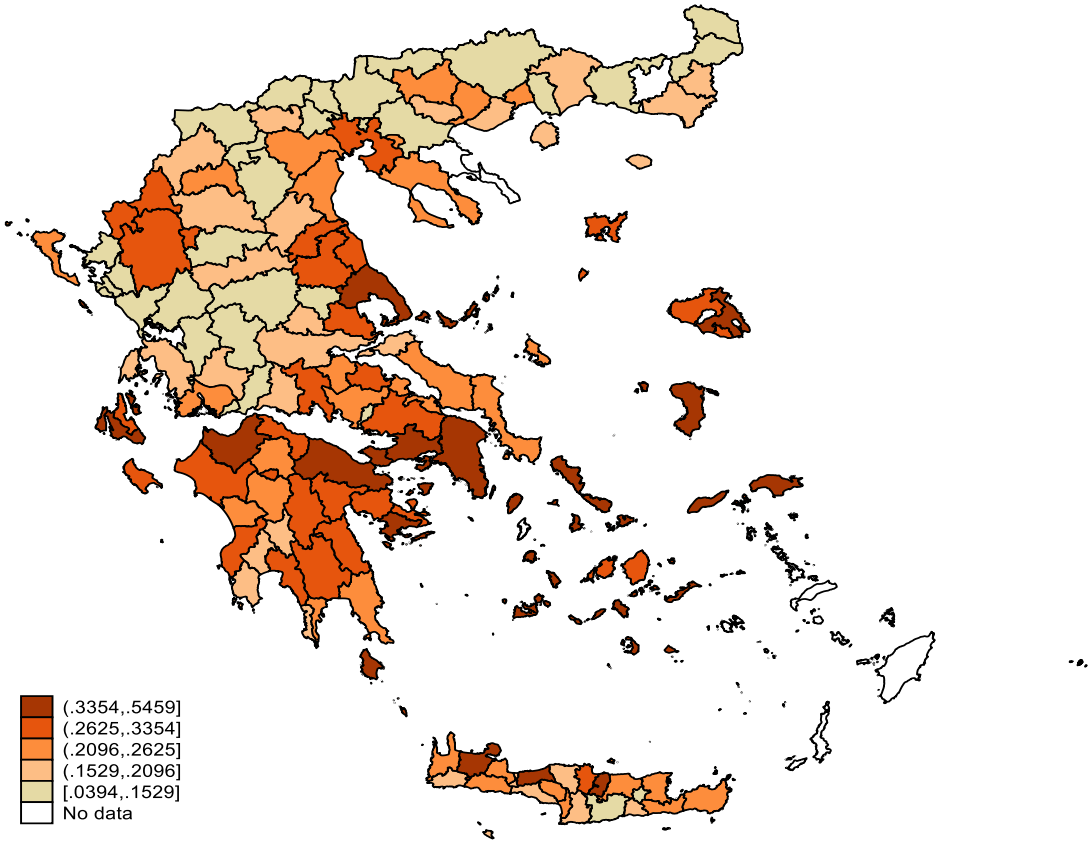
Source: Hellenic Statistical Authority, 1929 Census

Figure 4: Native Male Literacy Rate in 1920



Source: Hellenic Statistical Authority, 1929 Census

Figure 5: Native Male Literacy Rate in 1928



Source: Hellenic Statistical Authority, 1929 Census

Figure 6: Native Female Literacy Rate in 1920

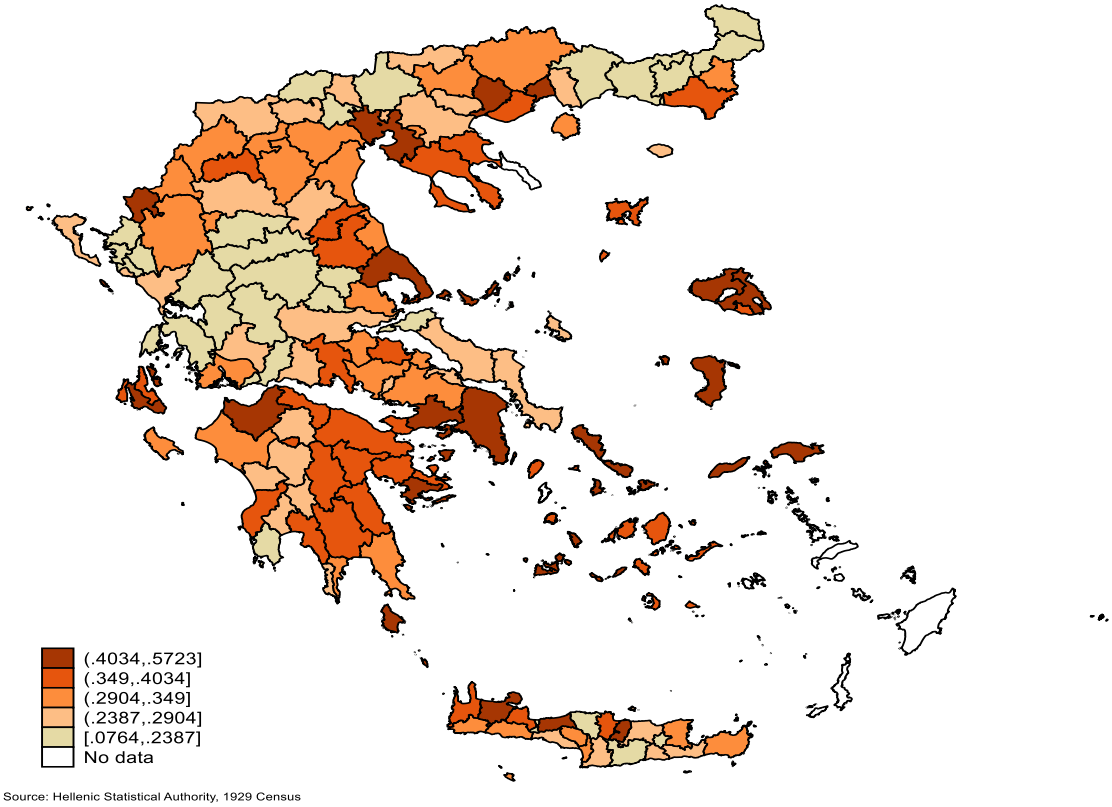
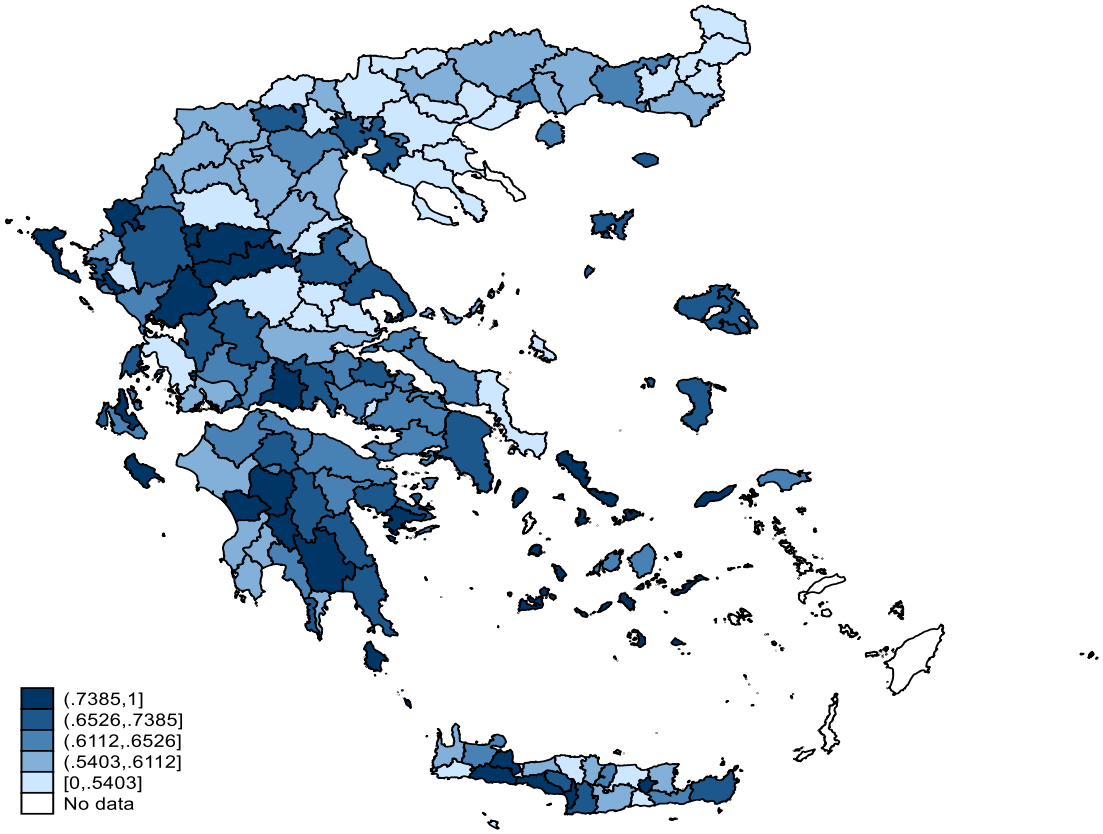


Figure 7: Native Female Literacy Rate in 1928

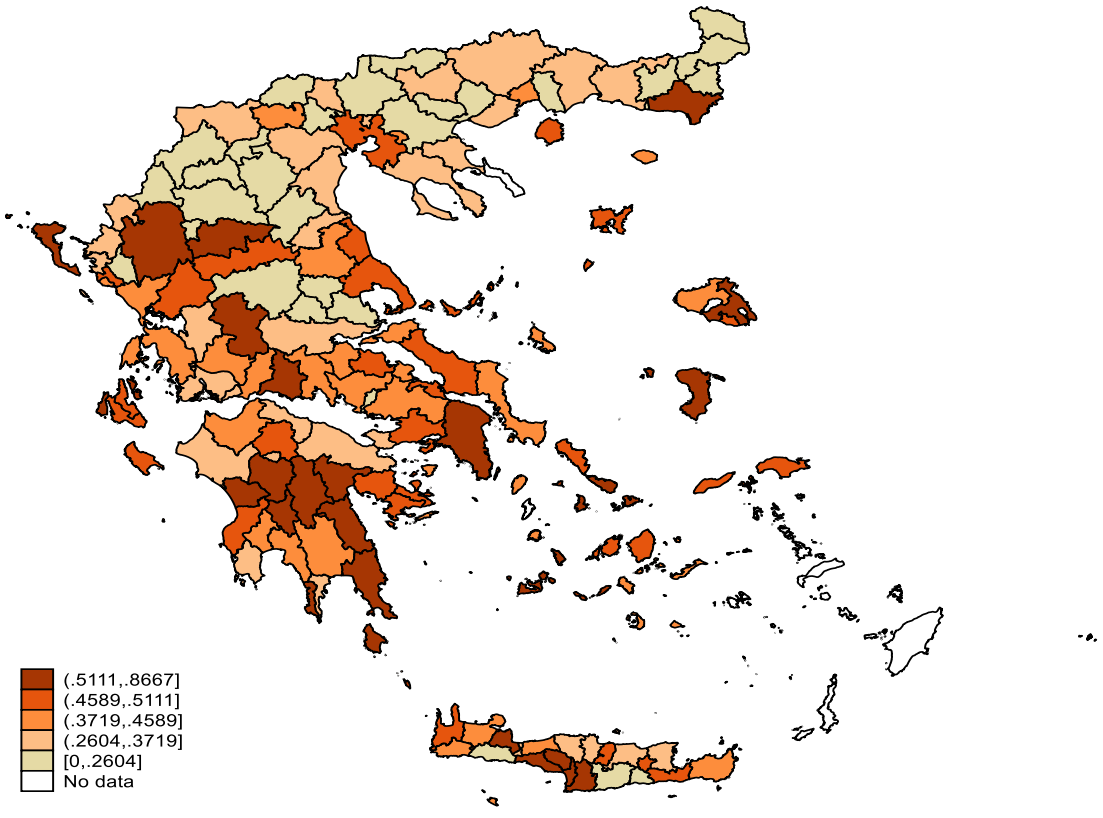
A first inspection of the two groups reveals very general patterns. Literacy increased in both groups, therefore both groups seem to have been affected by the inflow of refugees. Second, the allocation of literacy across provinces (and with the exception of the major urban centres) the allocation and especially the evolution of literacy within each of the two groups differs across provinces. Overall, a larger fraction of the male population seems to be literate, however, the female population that becomes literate accelerates over time.

Inspecting the corresponding graphs for refugees in 1928, suggests that a high fraction of refugee females was literate, a fraction proportionally larger that native females in 1920. Moreover, both male and female literacy rates seemed to be quite close to each other.



Source: Hellenic Statistical Authority, 1929 Census

Figure 8: Refugee Male Literacy Rate in 1928



Source: Hellenic Statistical Authority, 1929 Census

Figure 9: Refugee Female Literacy Rate in 1928

Table 1 shows the summary statistics related to literacy for native and refugee population, males and females, for each period (1920 and 1928 for natives and 1928 only for refugees). The numbers confirm the intuition derived from Figures 1-9.

Variable	Mean	Standard Deviation	Min	Max	Observations
Native Total Literacy Rate 1920	0.386	0.107	0.090	0.655	138
Native Male Literacy Rate 1920	0.531	0.128	0.137	0.745	138
Native Female Literacy Rate 1920	0.248	0.112	0.394	0.546	138
Native Total Literacy Rate 1928	0.465	0.086	0.129	0.663	141
Native Male Literacy Rate 1928	0.610	0.091	0.183	0.772	141
Native Female Literacy Rate 1928	0.323	0.107	0.076	0.573	141
Refugee Total Literacy Rate 1928	0.527	0.145	0	1	141
Refugee Male Literacy Rate 1928	0.636	0.131	0	1	141
Refugee Female Literacy Rate 1928	0.406	0.169	0	0.867	141

Table 1: Summary statistics for literacy by status/gender/year

Briefly, a higher fraction of native males is literate compare to native females in 1920 and this gap does not close in 1928, despite the fact that both groups have higher numbers in 1928. Second, the refugee literacy rate also manifests a gap, though somewhat smaller compared to that of natives in 1928 (when comparing the means). When comparing natives in 1920 with refugees in 1928, I observe that the total literacy rates of each group as well as the comparisons by gender, indicate without doubt that refuges were on average more educated for both genders. This gap somewhat closed in 1928, yet the refugees were still on average more educated than natives in 1928. Interestingly, my data is also available by age groups, and there I realize more clearly that some age groups changed faster compared to other age groups.

Given the short time period, as well as the financial difficulties that the country was faced with, I test whether this effect was driven by externalities related to the presence of the relatively more educated refugee population.

Literacy Shares	**Natives**				**Refugees**	
	1920		**1928**		**1928**	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
Total	0.531	0.248	0.609	0.322	0.635	0.527
Age 7	0.440	0.321	0.667	0.564	0.507	0.466
Age 9	0.702	0.499	0.871	0.757	0.628	0.573
Age 10-14	0.742	0.484	0.892	0.702	0.811	0.690
Age 15-19	0.726	0.423	0.817	0.552	0.685	0.524
Age 20-24	0.755	0.362	0.839	0.488	0.659	0.489
Age 25-29	0.720	0.289	0.835	0.405	0.727	0.457
Age 30-34	0.672	0.228	0.806	0.330	0.683	0.414
Age 35-39	0.609	0.192	0.755	0.276	0.663	0.345
Age 40-49	0.568	0.152	0.671	0.198	0.643	0.281
Age 50-59	0.494	0.115	0.603	0.150	0.596	0.225
Age 60 plus	0.415	0.084	0.485	0.102	0.477	0.159

Table 2: Summary statistics for literacy by status/gender/year/age groups

Employment

My second outcome variable that could potentially be affected by the inflow of refugees is labour supply. My data on labour supply are quite rich and provide information about employment shares of both natives (for the years 1920 and 1928) and the refugee population (for the year 1928) at the following sectors: agriculture, stock-farming, fishing, mining, industry, transportation, exchange, trade, staff services, self-employed and public services. My information is also available by gender.

Combining all the above data I can capture both the inter-sectoral and the intra-sectoral allocation after the inflow of the refugee population.

As a result of the inflow of the refugee population I could speculate on three potentially opposing effects to take place. On the one hand, the refugee population was a highly skilled population on average. Therefore, refugees could in principle be employed in skilled jobs. However, this was not a major concern of the native population especially during the first years that the living conditions of the refugees were bad and the degree of integration to the society very low. Second, the direct effect of the refugee inflow was translated into cheap and abundant labour force, therefore a sectoral reallocation was not unlikely. A reallocation that would lead more native men to sectors less labour-intensive compared to agriculture. Third, there is a gender-related dimension of the employment part of the analysis. This is related to the fact that native women, whose participation in the labour market was very limited at the time, could potentially be affected by the fact that refugee women were forced to work. Especially given the fact that many of those women were employed in white collar job, as secretaries and in administration positions, could set up a process of destigmatizing female labour force participation.

Variable	Mean	Standard Deviation	Min	Max	Observations
Native Total Employment 1920	11633.23	19011.79	728	207082	138
Native Male Employment 1920	10044.72	15763.68	668	171213	138
Native Female Employment 1920	1588.28	3415.98	22	35869	138
Native Total Employment 1928	13806.48	20868.72	713	225117	141
Native Male Employment 1928	10356.11	17026.51	635	187706	141

Native Female Employment 1928	3450.30	4268.83	26	37411	141
Refugee Total Employment 1928	3289.96	8976.24	1	84154	141
Refugee Male Employment 1928	2247.92	6595.09	1	63839	141
Refugee Female Employment 1928	1042.04	2523.32	0	20315	141

Table 3: Summary statistics for labour by status/gender/year

In Table 3 I present summary statistics for the total labour statistics. Overall, the mean values of native male population do not seem to increase between the years 1920 and 1928. The picture is somewhat different for the female population, where I observe an increase in the mean number of women employed. Comparatively, men outnumber women in labour force participation by large numbers and this remains unchanged between the year 1920 and 1928.

Shares of Employment by Sector	**Natives**				**Refugees**	
	1920		**1928**		**1928**	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
Agriculture	0.409	0.052	0.418	0.189	0.335	0.177
Stock Farming	0.068	0.005	0.080	0.019	0.016	0.002
Fishing	0.008	0.000	0.008	0.000	0.022	0.000
Mining	0.003	0.000	0.002	0.000	0.004	0.000
Industry	0.098	0.020	0.106	0.025	0.185	0.059
Transportation	0.035	0.000	0.035	0.000	0.046	0.000
Exchange	0.002	0.000	0.003	0.000	0.006	0.000
Trade	0.050	0.001	0.055	0.001	0.068	0.002
Staff Services	0.008	0.010	0.005	0.007	0.013	0.027
Self-Employed	0.021	0.004	0.023	0.006	0.035	0.025
Public Services	0.018	0.001	0.014	0.000	0.013	0.000

Table 4: Mean Employment by Status/Gender/Year/Sector

Table 4, presents the mean values by status/sector/gender/year. The dominating sectors for the Greek economy are agriculture, followed by industry, stock farming and trade. The picture remains similar for the year 1928 as well. It is interesting to highlight that the fraction of native women is extremely low in all sectors in both periods, though the numbers seem somewhat higher during 1928. As to refugee women, with the exception of agriculture, they dominate every other sector, in terms of female labour force participation. The participation rates though are still much lower compare to those of male refugees.

2.5 Empirical specification (Diff-in-Diff)

I attempt to identify the effect of the massive inflow of the refugee population on literacy rates and the intra/inter sectoral allocation of native men and women.

Simply controlling for the share of the refugee population allows for many unobservables to kick-in especially given the lack of historical data for the period. I therefore use a diff-in-diff approach. On the one hand the abrupt massive inflow of refugees and on the other hand the availability of precise and detailed data for the two-periods right preceding and succeeding the inflow allows me to use this event as a “natural experiment” and study the immediate implications of forced displacement on native literacy and employment. In future research I plan to explore the long-shadow of this event on modern-day norms and economic outcomes.

2.5.1 Forced Displacement as a Natural Experiment and Identification Concerns

Three are the various elements that render this event a "natural experiment' ideal to study the economic implications of forced displacement.

Sudden and Unexpected Event

First, it was a not predetermined event. The mutual agreement was agreed in 1923. Though it was the first officially agreed and signed displacement in history it was obligatory, nevertheless it was still an enforced movement for more than 2.000.000 people.

Random Allocation of Refugees Limited Movement across Provinces

Second, the allocation of refugees to various places all over Greece was unrelated to their educational or professional skills. Most of the work related to the settlement of the refugees was accomplished through the work of the Refugee Settlement Commission (RSC). This Commission was given significant powers from the Greek government in exchange of the securing a large loan was given. Its purpose was to help refugees to become self-supporting in the minimum required time (Kontogiorgi, 2006). The population of refugees was very high and constituted around 20% of population in 1928. The inflow was so massive and the associated expenses to sustain the incoming population so immense that the then Greek government allocated refugees to places where either past settlement from the departing population (as part of the population exchange) existed or where the population could be employed in agricultural activities. As a result, their spatial

allocation was very skewed towards Macedonia, Thrace and Attica, where approximately 87% of them settled.

Moreover, especially during the first years, there was not much movement of either refugees or natives. It was mostly natives took occasionally moved from rural to urban areas, in search of better employment opportunities in manufacturing and services (Riginos, 1987).

Short Interval under Examination

Third, the period I explore is so short that I can only trace the first direct implications of such a major event. As the literature suggests, such a massive inflow of refugees could have long-lasting implications for the Greek economy and society. In this study, I restrict my period of analysis to the years 1920-1928, i.e., right before and right after the displacement. As a result, my analysis is rather unaffected by other potentially confounding events of the period.

2.5.2 Treatment

Based on the above, I create my treatment variable. I construct a dummy variable that takes the value of 0 if the difference in the refugee population between the years 1920 and 1928 does not exceed 3% of the population in 1928 and the value 1 otherwise.

It is essential to note that there are hardly any provinces that did not have any refugees either before or after 1928, and this is the reason why I am forced to pick the difference in order to construct my treatment variable. Second, the reason

for which I chose 3% as the threshold value for the construction of the treatment variables is because it seems to be supported by the summary statistics.

The histogram of the 1928 data on refugee population in all provinces suggests that many provinces had a fraction of refugee population less than 5%. I obtain similar results though if I assume a stricter threshold at the level of 3% in the robustness section (which is the median value).

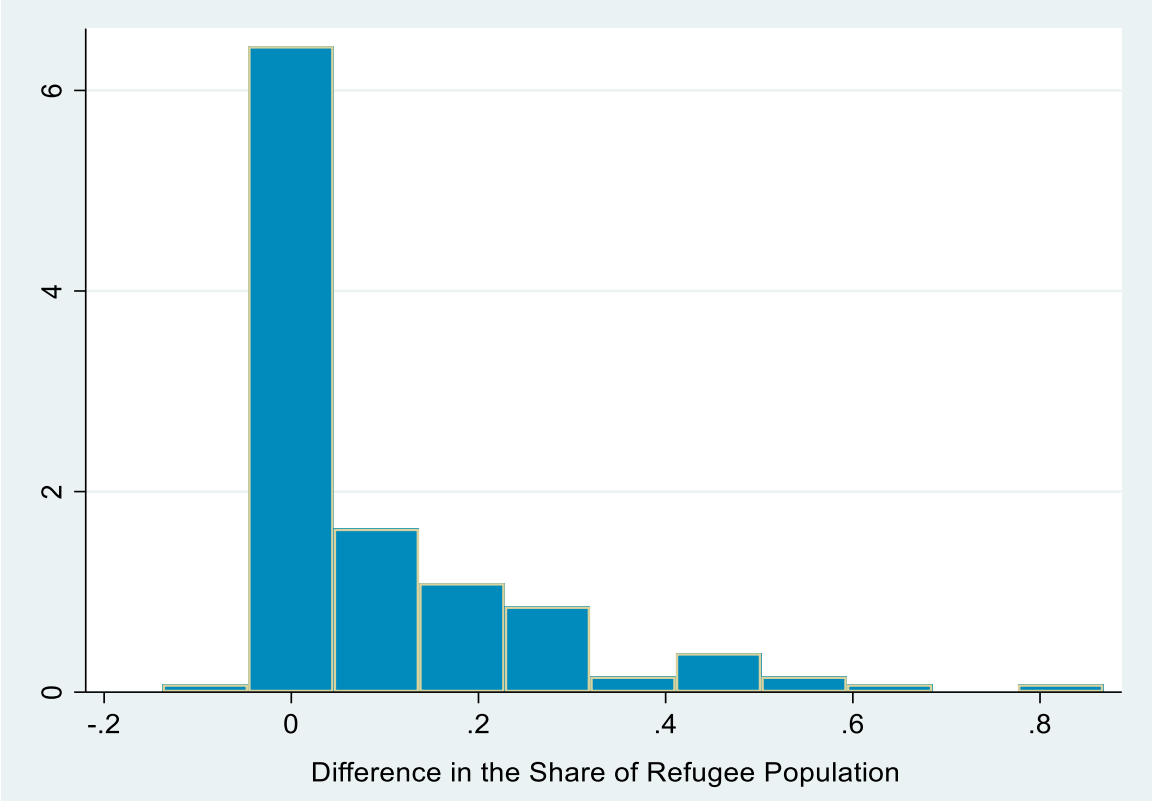


Figure 10: Difference in the Share of Refugee Population in 1928

2.5.3 Specification

I use the following specification as my benchmark specification of the chapter.

$$Y_{jt} = a_0 + a_1 * T1 + a_2 * Period + a_3 * T * Period + a_4 C_{jt} + a_5 * P + e_{jt}$$

where ***j*** refers to province j (138 provinces[2]) in period ***t*** (either 1920 or 1928), ***Y*** refers to either education outcomes (aggregate literacy rates, and rates by gender/age group) or employment outcomes (aggregate employment and employment by gender/sector). ***T*** is the treatment variable that is constructed as described above. I also include a full set of controls such as urbanization, population density, labour force, share of labour in industry/agriculture/commerce (in the employment about employment I only control for total labour force participation rate), share of Bulgarian/Turkish population and big city FE. Last, I include prefecture fixed effects to capture a large number of unobservable associated with the geopolitical division of Greece at the time and ***e*** is the error term clustered at the province level.

Henceforth, I will refer to this specification as the benchmark specification.

2.5.4 Parallel Trend Assumption

Though the parallel trend assumption is a quite critical step in employing a diff-in-diff approach, this is quite challenging in the context of this study as there are two major constraints. First, the geographic units have been very volatile during this period. And second, there is a severe lack of historical data. In Figure 11 I show that for Greece in total there has been a more or less parallel evolution in the male and

[2] The total number of provinces for the period 1928 is 141. However, for 1920 I have 138. Some of them broke up into smaller units in 141 which explain the higher number of provinces in 1928.

female literacy (note though that even the borders of Greece in whole have changed during this period).

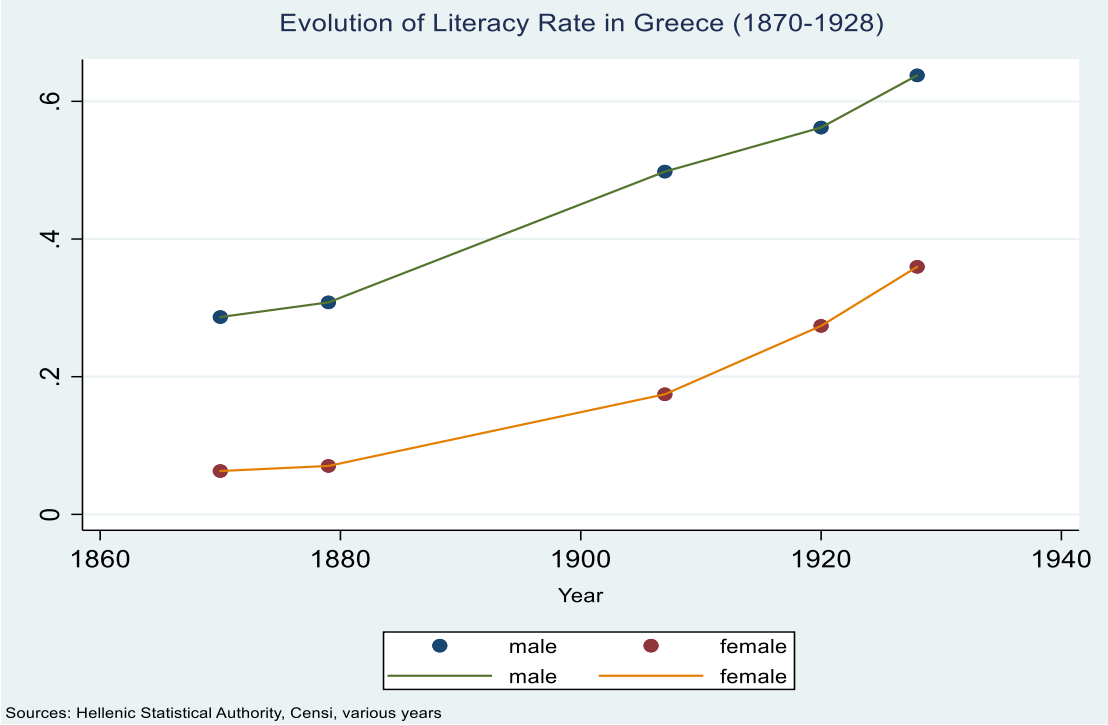


Figure 11: Evolution of Historical Literacy Rate in Greece

Though the whole graphs should be very cautiously interpreted there appears to be a slope change after 1923, especially for the case of men in terms of literacy.

2.6 Results

2.6.1 Literacy

I present my results graphically, as I have a large number of outcome variables.

Figure 12 presents the benchmark estimates for the case of native male literacy. I present the results for total literacy as well as by age group which gives me a more detailed picture as to which age group drives the result.

Overall, my findings suggest that the massive inflow of refugees has led to a statistically significant increase in the male literacy. It is interesting to note that this shift is relevant for all the age groups above the age of 7 and it is particularly strong in terms of magnitude for the age groups 10-14 and 15-17. This finding is quite interesting as it suggests that the groups that were particularly affected were the ones that would in principle abandon school especially in agricultural areas and especially after the age of 15. This could potentially hint to two potential sources of the effect. One reason could be that that competition has increased after the inflow of a large number of people that were mostly highly educated and therefore extending formal education was an optimal strategy. Second, it would denote that the presence of cheaper labour force, released resources for the natives who eventually allocated these extra resources to extending their education.

It is also interesting to note however, that similar qualitative results were obtained for all the age groups, even for older people. However, in older ages these results were weaker in terms of magnitude.

Figure 13 illustrates the corresponding results for female natives, for the total literacy and by age group. Overall, the picture is quite similar qualitatively, however one can trace significant differences that reflect the fact that women at the time did not have equal access to education. First, the effect is present from the age 9 onwards. It is overall much weaker in terms of significance, especially for women

above the age of 24. It is also interesting to note that the magnitude of the coefficients is quantitatively much smaller for all age groups, thereby suggesting that though women were affected, the biases related to women's education were quite strong.

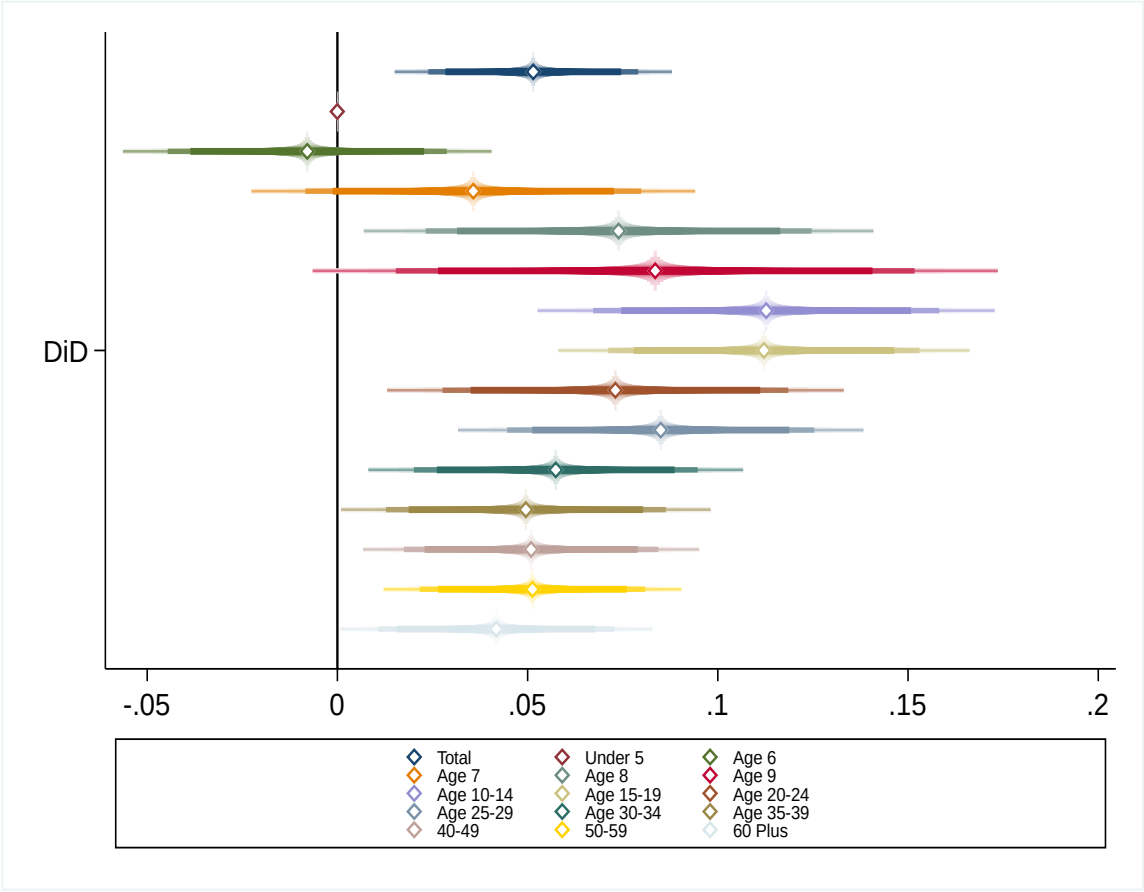


Figure 12: Benchmark Estimates for Native Male Literacy

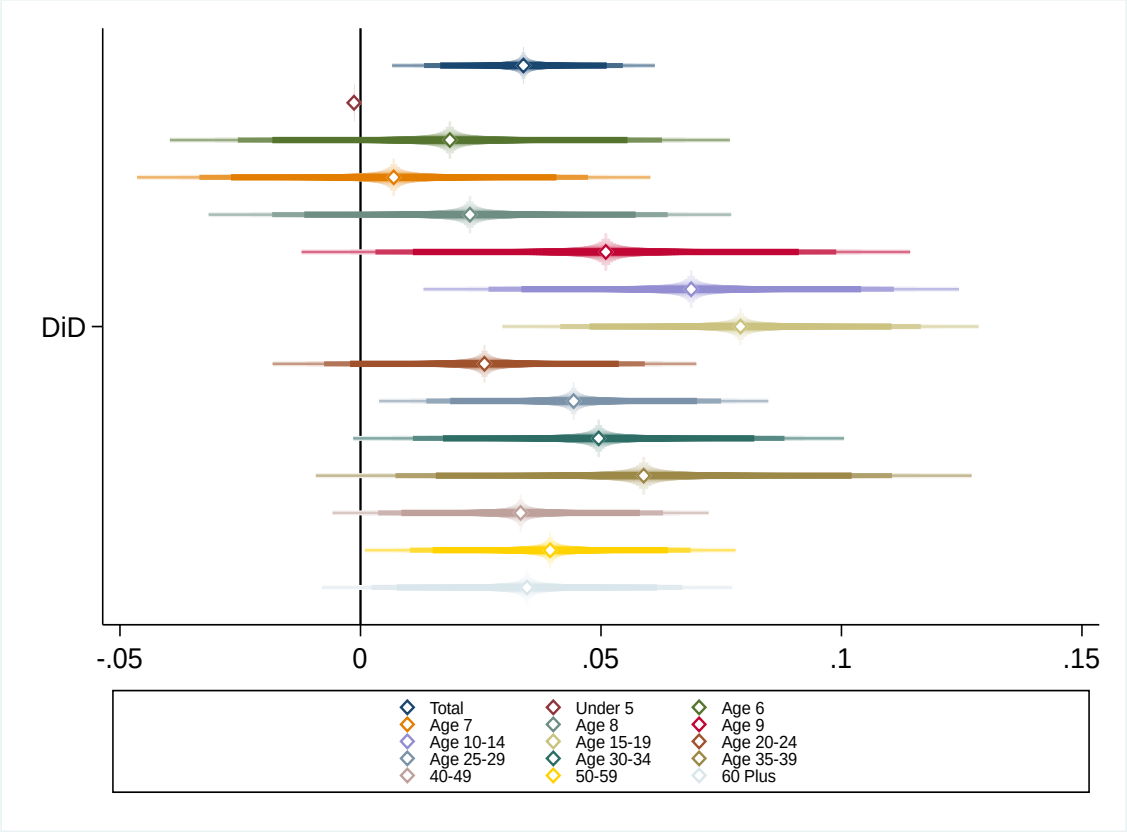


Figure 13: Benchmark Estimates for Native Female Literacy

2.6.2 Employment Allocation

Figure 14 presents the results for allocation of male employment. Unlike the case of literacy, the results are not overly strong in the case of employment. Still though, even for weaker results, the results seem plausible, i.e., it appears that after the massive inflow of refuges, the share of men employed in agriculture has reduced, whereas there is an increase in the share of men employed in stock farming, industry, transportation and trade. This could hint to the fact that the "heavier" agricultural jobs were undertaken by the cheaper labour force, i.e., the refugee population, which allowed more native men to turn to other activities. Given

that these activities are also more human capital intensive, this finding goes hand-in-hand with the findings about literacy.

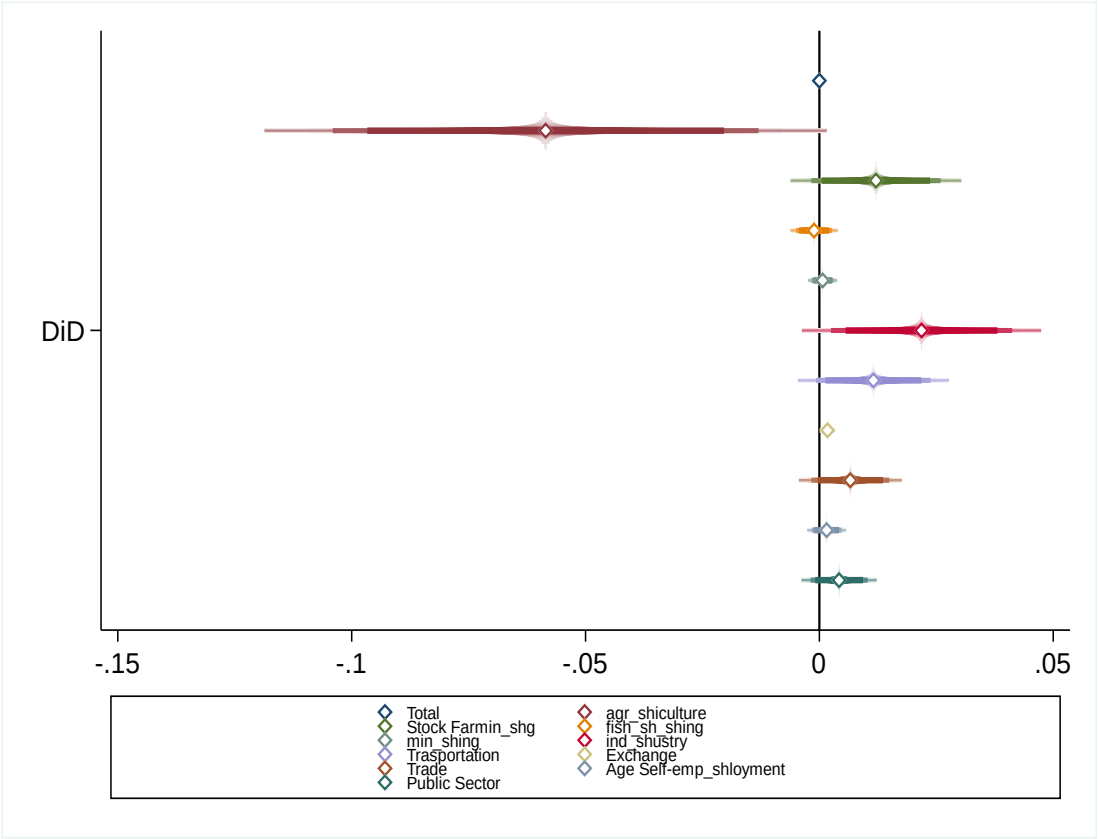


Figure 14: Benchmark Estimates for Native Male Allocation

The picture about women is completely different as Figure 15 illustrates. First, their participation in the labour market is already trivial as was already clear from the summary statistics. And second, their labour force participation seems to be totally unaffected, at least in such a short period, by the fact that refugee women are forced (by necessity) to participate in the labour market.

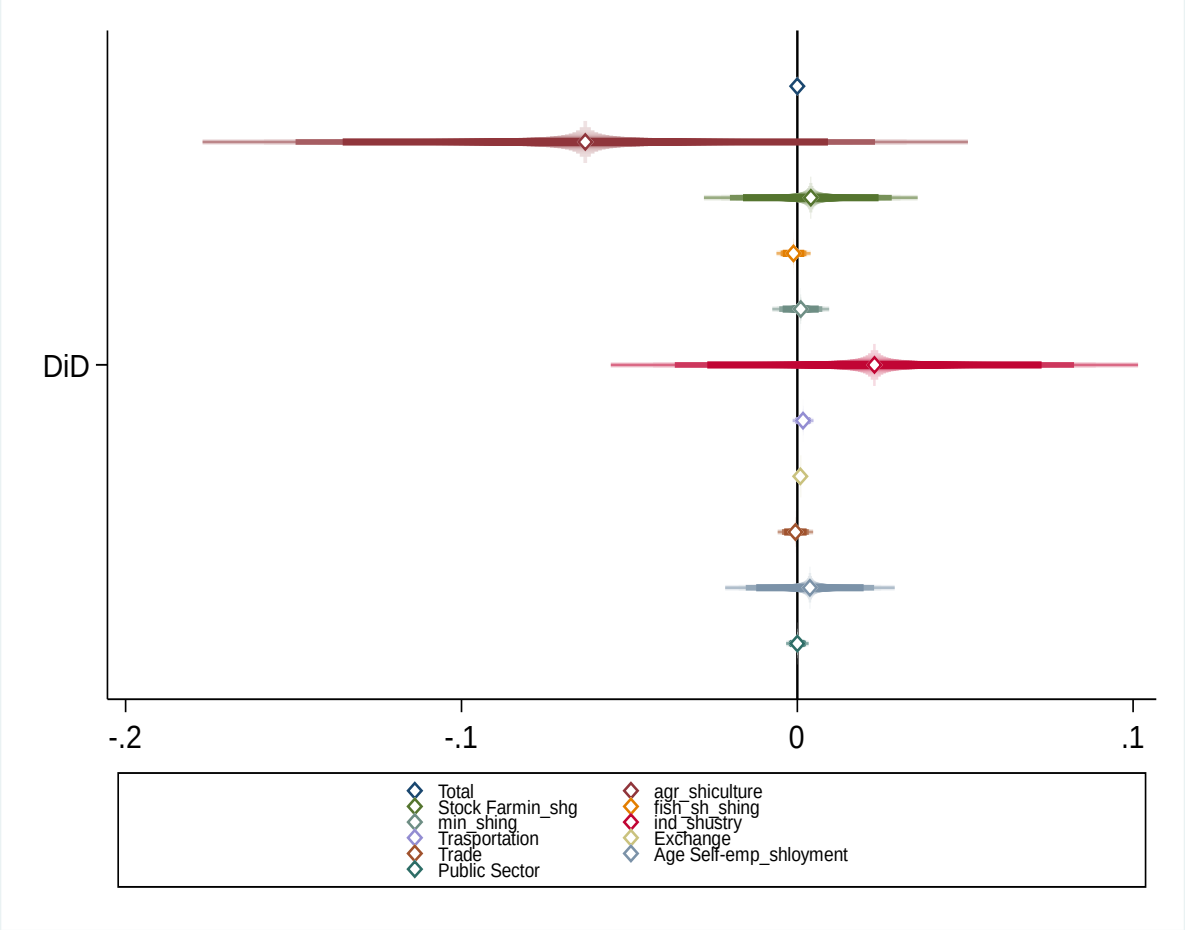


Figure 15: Benchmark Estimates for Native Female Allocation

The findings are quite similar if I consider the measure of inter-sectoral allocation, although these are two different but related measures of employment. The intra-sectoral employment allocation comes of the division between sectoral employment and total employment, thus giving how labour force participation is distributed among different sectors. On the other hand, inter-sectoral allocation of employment is the division of sectoral employment with the population aged older than 10 years old. This measure provides how potentially active population is distributed among labour force participation.

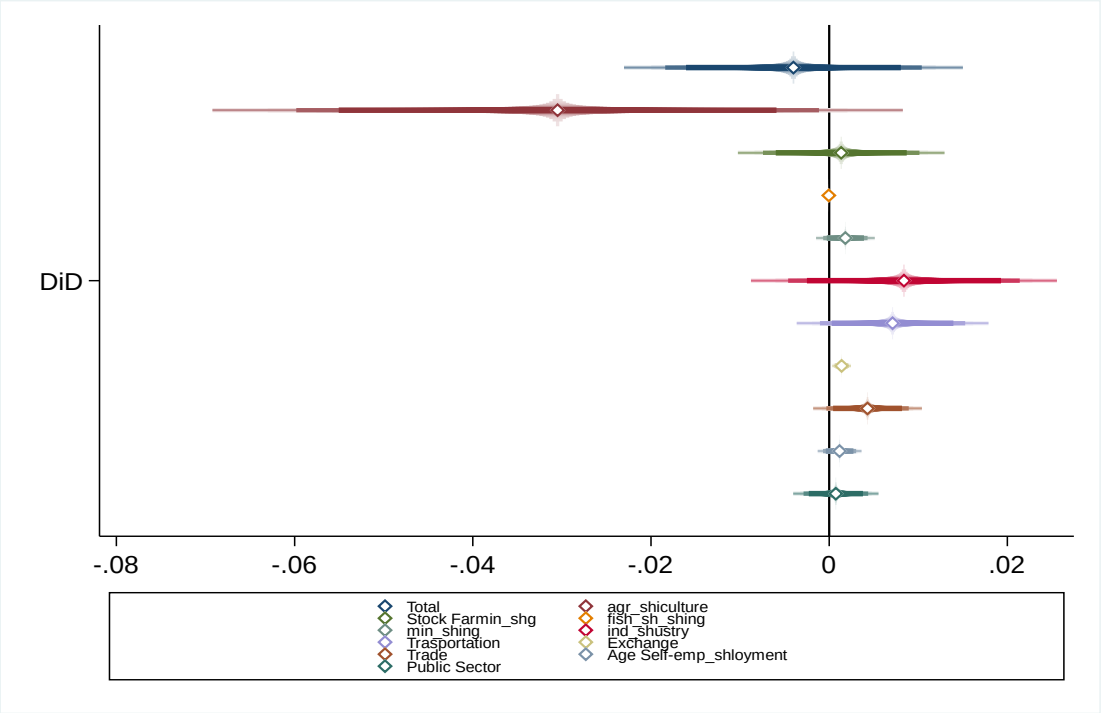


Figure 16: Benchmark Estimates for Native Male Intersectoral Allocation

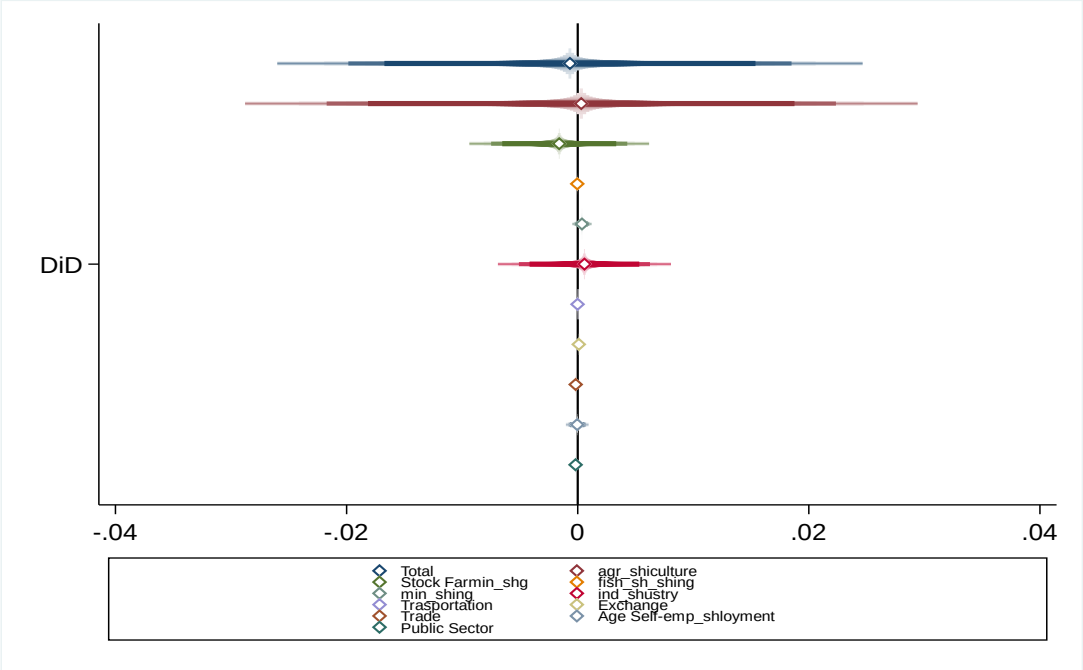


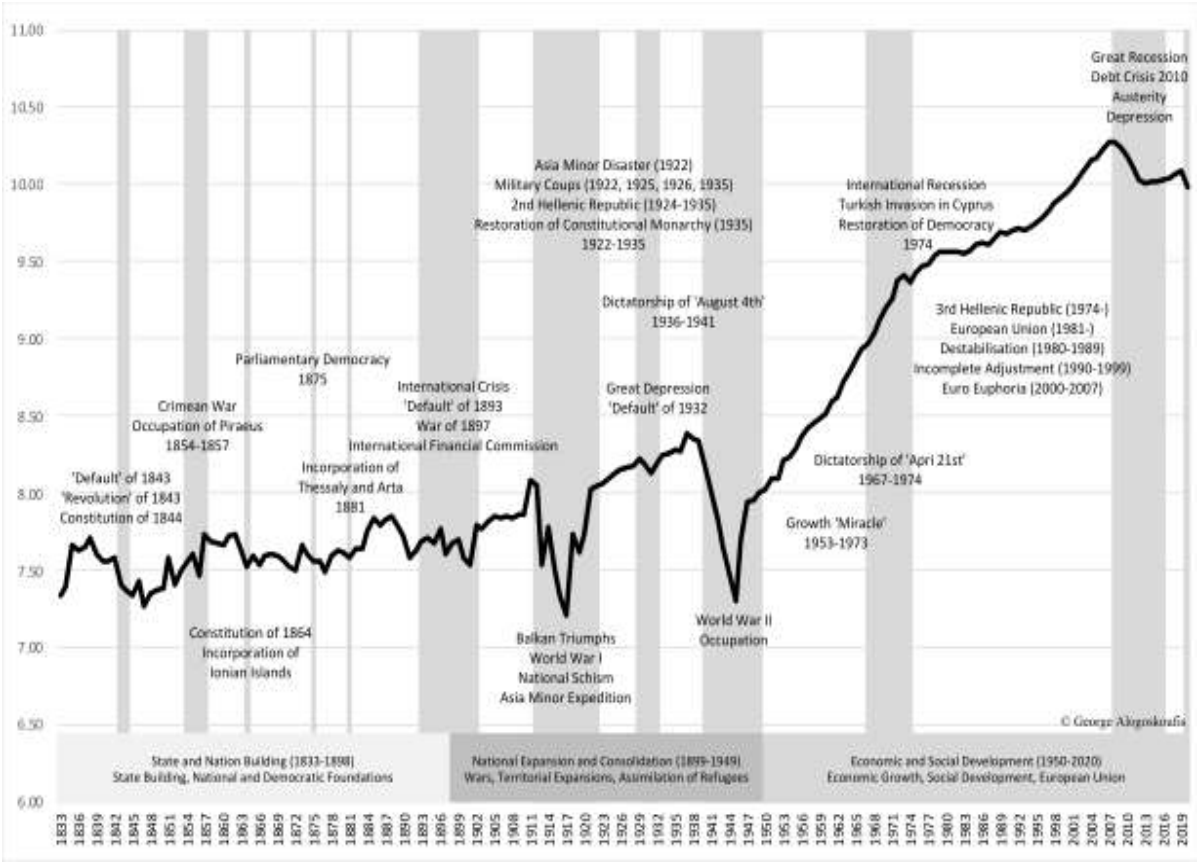
Figure 17: Benchmark Estimates for Native Female Intersectoral Allocation

2.7 Mechanisms

In this study I remain agnostic as to what as the potential mechanisms though which this effect operates. I intend to expand my analysis to explicitly test for potential mechanisms associated with my result. Below I list potential candidates that could explain the reduced form effect of the massive inflow of refugees on literacy and sectoral allocation in employment.

Rise in the growth rate of Greece

One plausible explanation could be associated with an increase in growth rates. A large inflow of cheap and educated labour force could potentially drive up the growth rates and the productivity of the Greek economy. Despite the fact that the economic needs were enormous during the first years, in order to be able to sustain the refugee population, nevertheless it cannot be ignored that they were directly introduced to the labour market of the country. GDP estimates dating back to the period 1920-1928 suggest a remarkable increase in GDP during that period and a period of sustained growth till the onset of WWII. Therefore, growth driven partly by relative political stability and partly by the presence of cheap labour force, could foster education for both gender groups and the rise of sectors other than the agricultural sector and the associated allocation of labour force into those sectors.



Source: Alogoskoufis (2021) and Maddison Project and European Commission. Real GDP per capita is measured in PPP adjusted US dollars, at 2011 prices and is presented in a logarithmic scale.

Figure 18: Real GDP per capita

Michalopoulos and Papaioannou (2017) examine the link between current GDP per capita and pre-WWI output and find that the correlation remains quite high. The evidence on persistence hints supports that historical factors shaping comparative economic development. Moreover, Murard and Sakalli (2019) find beneficial effects of forced displacement on long-run economic development. These long-run benefits of refugees appear to be driven by the provision of new agricultural know-how and the transfer of technological knowledge in textile, which fostered growth through higher diversity in complementary skills.

Competition in labour market between natives and refugees

During the early years upon the arrival of refugees their living conditions were not ideal and the degree of integration into the economy and the society was rather limited. As a result, the native population did not immediately consider the incoming population as a threat. However, the vast majority of the refugee population was comprised educated people and people with a different, more open and advanced set of norms. This population was quite well-off in the origin region and it was simply the forced displacement that placed them into a disadvantaged position. This was soon quite visible and as a result, the native population soon felt antagonized and threatened. This was especially the case for men in the labour marked and for women in "marriage markets" and the social domain. A potential other mechanism though could be related to the attempts of the native to not fall short of the refugee population in all aspects of economic and social life.

Spillover effects from refugees to natives

Following the rational of the competition argument, the interplay between the refugee and the native population could also lead to spillover effects. The creation of schools for the refugee population all over Greece as well as the new jobs associated with the presence of a skilled and massive population, could bear positive externalities in these domains for the native population. These positive externalities could be driven either informally by the plain interaction of the two groups, or by increased government spending in those regions.

All of the above mechanisms could be in place simultaneously reinforcing each other and the final outcome. I primarily believe that different mechanisms were in place in different time periods. For instance, as time went by, the increased interaction of the two groups (which eventually led even to mixed marriages) started becoming increasingly important.

2.8 Conclusions

Using a novel historical dataset about the totality of 141 Greek provinces during the period 1920-1928 I establish that the sudden inflow of more than 1 million refugees in the then Greek territory had a positive impact on both male and female literacy rates, even within a short period of 8 years. The results were stronger for men and relevant for more age groups compared to women, yet, the implications for women were also qualitatively and quantitatively significant. As to the implications for employment, I establish that native male employment was redirected from agriculture to more human capital-intensive sectors such as industry, trade and transportation. A similar effect was not observed for women. Potential mechanisms associated with these reduced form effects are growth effects as well as competition and/or spillover effects between the native and the refugee population.

I plan to extend my research to more recent outcomes and see whether this historical event has cast a long-shadow on the economic and social life of Greece.

Appendix of Chapter 2

Definitions of Variables at the Provincial Level

Human Capital: The quotient of literates with the sum of literates plus illiterates. I include literacy rates for both natives and refugees, by gender (male, female and total), computed for the periods 1920 and 1928. Also, I use literacy rates by 15 age groups: under 5 years old, 5 years old, 6 years old, 7 years old, 8 years old, 9 years old, 10-14 years old, 15-19 years old, 20-24 years old, 25-29 years old, 30-34 years old, 35-39 years old, 40-49 years old, 50-59 years old and older than 60 years old.

Source: Census (1929), Hellenic Statistical Authority, available at http://dlib.statistics.gr/portal/page/portal/ESYE/yeararticles?p_topic=10007862&p_cat=10007862&p_catage=1928.

Employment: Total employment and employment by 11 sectors (agriculture, stock-farming/hunting, fishing, mines, industry, transport/transportation, credit/exchange/mediation, trade, staff services, self-employed and public services) for both natives and refugees, by gender (male, female and total), for the periods 1920 and 1928. Also, I include two different employment shares. I compute employment for the 11 sectors by total employment and by population aged older than 10 years old.

Source: Census (1929), Hellenic Statistical Authority, available at http://dlib.statistics.gr/portal/page/portal/ESYE/yeararticles?p_topic=10007862&p_cat=10007862&p_catage=1928

Labour Force: The quotient of total employment with population aged older than 10 years old. I include labour force for both natives and refugees, by gender (male, female and total), computed for the periods 1920 and 1928.

Source: Census (1929), Hellenic Statistical Authority, available at http://dlib.statistics.gr/portal/page/portal/ESYE/yeararticles?p_topic=10007862&p_cat=10007862&p_catage=1928

Refugee Shares: The quotient of refugees with population by gender (male, female and total) before 1922 and after 1922 (Asia Minor Disaster in 1922). I also use the difference between refugee share before 1922 and after 1922.

Source: Census (1929), Hellenic Statistical Authority, available at http://dlib.statistics.gr/portal/page/portal/ESYE/yeararticles?p_topic=10007862&p_cat=10007862&p_catage=1928.

Population Density: The quotient of total population with land area computed for the periods 1920 and 1928.

Source: Census (1929), Hellenic Statistical Authority, available at http://dlib.statistics.gr/portal/page/portal/ESYE/yeararticles?p_topic=10007862&p_cat=10007862&p_catage=1928.

Urbanization: Urbanization rate is the percentage of total population of province living in settlements with population >10,000 for the years 1920 and 1928.

Source: Census (1929), Hellenic Statistical Authority, available at http://dlib.statistics.gr/portal/page/portal/ESYE/yeararticles?p_topic=10007862&p_cat=10007862&p_catage=1928.

Bulgarians/Turks share: The quotient of Bulgarians/Turks who left Greece with total population for the years 1920 and 1928. I use the same values for the two periods (1920 and 1928).

Source: Census (1929), Hellenic Statistical Authority, available at http://dlib.statistics.gr/portal/page/portal/ESYE/yeararticles?p_topic=10007862&p_cat=10007862&p_catage=1928.

Chapter 3: A European metropolitan analysis of enterprise births

3.1 Introduction

The common knowledge is that entrepreneurship and new firms are important drivers of job creation and long-term economic growth. An economy with high entrepreneurship tends to be more progressive, competitive and better at adapting. Despite the fact that entrepreneurship has always remained in the agenda of researchers and policymakers, it is not considered a topic belonging to the core. However, in recent years the industrial organization and firm dynamics literature has focused on studying firm entry, its determinants, as well as its economic impact.

Business dynamism in the United States has been declining in the last several decades. There is a huge literature analyzing various aspects of this decline in U.S. business dynamism, some of which, interestingly, have also emerged in other economies (Decker et al., 2016; Gourio et al., 2016; Pugsley and Sahin, 2019; Bijnens and Konings, 2018 for the case of Belgium). The 2007-2009 Great recession exhibited an unprecedented 5 percent decline in the number of firms, which was driven by a 25 percent decline in the number of entrants between 2007 and 2010 (Decker et al., 2016). The entry rate of new businesses has decreased, productivity growth has slowed down, and the labor share of output has decreased, while market concentration and the corporate profit share of Gross Domestic Product (GDP) have increased. Industries that suffered the most included agriculture, mining, logging, durable goods, construction, and automobiles. Nearly 800,000 manufacturing jobs were lost, and 630,000 construction jobs disappeared as home-building slowed. Jobs also dried up in the financial sector, in publishing houses and trucking companies,

department stores and hotels. One in five employees lost their jobs at the beginning of the Great Recession. Many of those people never recovered; they never got real work again. On the other hand, Heyman et al. (2019) observe no decline or a smaller and later decline in Sweden for different measures of entrepreneurial activity over the period of 1990–2013. In particular, they find an increase in start-up activity and a likely explanation for the high entrepreneurial activity was a set of policy reforms that initiated in the late 1980s but mostly implemented in the 1990s.

Business creation is a vital source of innovation, economic growth and employment creation. Policy makers around the world are increasingly trying to promote policies that foster local entrepreneurship and more innovation-based industries. Regions showing higher levels of entrepreneurship and employment are characterized by better local governance, spend more on R&D activities and have a more educated workforce. Similarly, regions have on average higher firm creation rate when they have business-friendly regulations (ease of doing business) and a higher quality of governance (i.e. low levels of corruption) (OECD, 2017). New firms can reshape markets and replace firms that are unable or slow to adjust. New firms account for a significant share of activity in a given year, and young firms make a long-lasting contribution to aggregate employment despite high failure rates (Haltiwanger et al., 2013). New firms and young firms create a disproportionately high number of jobs in Brazil (Brummund and Connolly, 2019). Business births account for about 20 percent of US gross (total) job creation while high-growth businesses (which are disproportionately young) account for almost 50 percent of gross job creation. Most new enterprises exit within their first ten years and most

surviving young businesses do not grow but remain small (Decker et al., 2014). A one standard deviation shock to the number of enterprise births leads ultimately to an increase of 1–1.5 percent in real GDP and lasts ten years or possibly longer (Gourio et al., 2016). Clementi and Palazzo (2016) support that entry and exit are responsible for almost one fifth of output growth over the ten years following the shock. Young firms appear to play a critical role in innovative activity that also contributes to productivity growth (including within-firm productivity growth).

Knowledge is a key source for enterprise births, particularly in innovative industries. Farre-Mensa et al. (2020), show that patent approvals have a substantial and long-lasting impact on births, especially the first time enterprise births apply for patent protection: births whose first patent application is approved create more jobs, enjoy faster sales growth and are more innovative than births with only randomly different inventions that fail to win patent protection**.** New businesses in general and innovative births in particular, can be regarded as manifestations of knowledge spillovers from extant knowledge sources (Fritsch and Wyrwich, 2018). Businesses, for example, buy patents to use the technology covered by the patent, which could be vital for their production, and the buyer's willingness-to-pay depends on the technological knowledge contained in the patent. Further, the buyer of a patent can develop connections with the seller in order to acquire the “how-to” knowledge to implement the patented technology (Drivas et al., 2016).

This research aims to conduct a systematic examination to extract and integrate the major factors for higher regional enterprise births across firms in European metropolitan regions for the period 2008-2013. So far, the lack of a

consistent and comprehensive database for European metropolitan regions has been a major impediment for assessing differences in entrepreneurship across space. This project contributes to fill that void. Understanding economic growth or decline requires to know that economies are territorially partitioned, first into regional and then into local economies. Regional analysis includes understanding of the similarities and differences relative to the relationships between people and places. The term region is multilateral because the region makes a complex interconnection between different geographical, economic, cultural, political and other factors. The notion of the region includes functional and formal regions. Formal regions exhibit high levels of homogeneity relative to their distinguishing features. Such features might include affluence, dominant land uses, or religious preferences. Functional regions may be based on distinguishing characteristics that occur throughout the region, but that vary in intensity and importance. I find that the size of the population and labor market, total patents and human capital factors (e.g. human resources in science and technology) stimulate enterprise births in all sectors. I incorporate these measures and employ cross-country and panel fixed effects econometric analysis of enterprise births in all sectors in EU metropolitan areas. Metropolitan areas with larger, growing, denser markets as well as more foreign-born and higher labour turnover have higher enterprise births in all sectors (Behrens et al., 2014; Motoyama and Malizia, 2017).

The value-added of my work is that it is the first among empirical studies on business dynamism which focuses on European metropolitan regions, while most literature focuses on specific countries and especially in the U.S. Particularly, I use

more disaggregated (metropolitan-level) data than related literature, which better capture the spatial variation of the phenomena analyzed.

I pursue empirical analysis taking into account numerous empirical challenges. I use births of enterprises (total sector, manufacturing, construction and services) as dependent variable and a large number of confounding factors affecting the dependent variable as explanatory variables: total patents divided by human resources in science and technology, population density, gross domestic product (GDP), purchasing power standard (PPS), gross value added (GVA), urbanization and physical infrastructure factors (e.g. roads, transport goods). I utilize alternative techniques to deal with endogeneity. These include fixed-effects and Arellano-Bond dynamic panel estimator (difference GMM). To further strengthen my results, I also provide spatial estimations because it is widely recognized that sample data collected from geographically close entities are not independent but spatially correlated. In addition, I allow for clustered standard errors at the country level to account for spatial error dependence.

I empirically demonstrate that: i) total patent applications in combination with human resources have a positive effect on enterprise births; ii) demographic factors and economic factors exhibit positive association with the dependent variable except for purchasing power standard; iii) some physical infrastructure factors are positively correlated with enterprise births and some other are negatively correlated.

My study has, of course, a number of limitations. I have no information on a number of issues, including: 1) number of enterprise births for the years before

2008; 2) data about capital stock and R&D expenditure that might provide important insights about their effect on the economy of their region.

The rest of the chapter is structured as follows: in Section 2 I revisit the theoretical framework on business dynamics and innovation. In Section 3 I describe the data and the methodology used in this study. In Section 4 I present my empirical results and Section 5 concludes.

3.2 Literature Review

A huge part of the literature analyses the decline in USA business dynamics, the impacts of this decline and the possible factors led to this decrease. The decline in the entry rate has an immediate effect on aggregate employment. The share of US employment accounted for by young firms has declined by almost 30 percent over the last 30 years in all 50 states but also there is a large increase in the share of mature employer businesses. An analysis from Decker et al. (2014) suggests that the decline in business dynamics is due to the lower rate of business startups and the decreasing role of dynamic young businesses in the economy. They show that most firm startups exit within their first ten years, and most surviving young businesses do not grow but remain small. Nevertheless, a small fraction of young firms display very high growth and contribute significantly to job creation. These high-growth firms compensate for almost all the job losses associated with shrinking and exiting firms within their cohort. The implication is that each entering cohort of startups makes a long-lasting contribution to net job creation. The share of employment in young firms has declined in all three sectors but the decline is much bigger for young retail and services firms than it is for young manufacturing firms. Two possible causes for

this decline in entrepreneurial activity are policy frictions that reduce incentives for labor adjustment, such as unlawful discharge regulations or occupational licensing, and financial frictions that might arise as a result of bank consolidation or perceived changes in financial risk. Another possible reason for this declining productivity in the high-tech sector is the transition from “general-purpose” to “special-purpose” equipment manufacturing in the United States. In addition, a strong explanatory reason is the changing nature of network externalities faced by certain types of young firms (Decker et al., 2014; Decker et al., 2016). They present evidence that the contribution of high-growth (young) firms to U.S. job creation and the patterns of positive skewness in the firm growth rate distribution are changing. They have seen a decline in high-growth firms and especially high-growth young firms. The overall decline reflects a sharp drop in the 90th percentile of the growth rate distribution accounted for by the declining share of young firms without an equivalent increase in the 10th percentile and the declining propensity for young firms to be high-growth firms. Other key groups with similar declines in skewness include the information sector, the high tech sector and publicly traded firms. While they have described in detail several aspects of declining skewness in the U.S. using industry, size, and age comparisons, they have not identified the underlying causes of these trends. Theory suggests that firms may sometimes operate below optimal scale because of credit constraints or imperfect information about their own quality and high growth follows the loosening of these constraints as the firm size distribution evolves to better reflect the underlying firm productivity distribution. In their last study Decker et al. (2017) using firm-level data on labor productivity, show that dampened growth in allocative efficiency can account for much of the decline in aggregate productivity

growth between the late 1990s and the mid-2000s. This slowdown reflects inefficient allocation of productive resources as well as the interaction between allocation and slowing within-firm productivity growth. Their findings imply that, consistent with the conclusions of Decker et al. (2014, 2016), declining business dynamism since 2000 is likely a drag on American living standards.

Pugsley and Sahin (2019) support that the direct effects of the continued decline in the entry rate and its delayed effects on the firm age distribution together play a major role in the slowing trend of employment growth and the emergence of jobless recoveries in the USA. The decline in the entry rate has an immediate effect on aggregate employment because of the direct contribution of new firms to net employment growth and a delayed but growing effect on the aggregate cyclical elasticity of employment of older firms. These two effects slow the cyclical employment recovery from a recession, a permanent change in employment dynamics refer to as “grown-up business cycles”. Alternately, changes in worker demographic structure and secular increases in import competition are leading factors behind the declines in entry within both sectors and geography. Consequently, changes in the age distribution have been driven only by the declines in the startup rate and not by a change in the dynamics of new and extant businesses. The decline in the startup employment share and shift in the firm age distribution is referred as a “startup deficit”. They find statistically significant declines in startup rates in industries whose exposure to import competition increased the most. The change in employer age composition has brought with it the introduction of grown-up business cycles, with lower trend employment growth,

jobless recoveries and generally a reduction in the elasticity of employment to the business cycle.

Another analysis of Karahan et al. (2019) proposes that the reason for the long-run decline in the startup rate is a slowdown in labor supply growth since the late 1970s, largely pre-determined by demographics. There are two channels between supply growth and the startup rate: the first is the direct effect of the change in labor force growth that requires a change in the net entry rate to maintain balanced growth and the second is an indirect amplification through its effects on the aggregate exit rate. Demographic changes explain credibly the declining startup rate and in particular, across all specifications labor supply growth explains between 10 and 60 percent of the decline in the startup rate. They find that shifts in labor supply growth have no statistically significant effect on startup size or on survival and employment growth rates, conditional on firm age. For average startup size, exit rate of young firms and their conditional growth rate, the coefficients are statistically indistinguishable from zero.

Other studies support this decline in business dynamics, as the previous literature, but for other countries except for USA. Bijnens and Konings (2018) support that business dynamism and entrepreneurship have been declining over recent decades by using unique data with characteristics (dispersion and skewness) of the distribution of firm growth rates from all for-profit firms incorporated in Belgium for the years 1985-2014. This decline starts around the year 2000 following a decade of declining start-up rates and as a result the firm landscape consists of older and less volatile firms. This decline also occurred in the USA, where firms face a

more flexible institutional environment than in Belgium. This explains that global trends are at the basis of this evolution and not country-specific changes. They observe a decreasing share of young firms that become high-growth firms and a declining tendency for small (not necessarily young) firms to experience fast growth. This might be due to the fact that resources are already allocated to the most productive (larger) firms. If this happens, then the declining dynamism is not necessarily negative for overall productivity growth. A driving force for productivity growth is the reallocation of (human) resources in the most productive firms. A similar research from Bakhtiari (2019) concludes that the Australian entrepreneurial landscape has become less dynamic and more hazardous over the years 2002-2015. This downturn follows a similar trend to that of the United States but it is also due to forces specific to Australia. Firm's startup rates decrease and those firms that enter the market are more likely to exit than their counterparts that entered in earlier years, a fact that probably happens because of a spike after the global financial crisis. Despite fewer firms entering, the share of young firms from job creation has been steadily high.

On the other hand, Heyman et al. (2019) suggest that young firms are more prominent in the Swedish business sector than in the U.S. business sector, using a Swedish comprehensive employer-employee database from Statistics Sweden over the period 1990-2013. Young Swedish firms, aged five years or less, account for more than half of all firms during this period. They observe no decline or a smaller and later decline in Sweden for different measures of entrepreneurial activity. Moreover, most of the job creation by young firms occurs in the expanding service

sector. However, there is an increase in job destruction rates for young firms, which giving rise to a small decrease in employment share for young firms in the business sector. They argue that a possible explanation for why Sweden appears to have high entrepreneurial activity is the economic reforms that were implemented in Sweden in the 1990s which mitigated several hurdles to entrepreneurship. Sweden transformed from having one of the most regulated, static business sectors in the developed world in the late 1980s to one of the more pro-entrepreneurial business regulation systems in the developed world. Finally, the reform pace in business policy has substantially slowed in the Swedish business sector the last decades, suggesting that the recent trend of less dynamism in the Swedish business sector might be here to stay.

The importance of the creation of new businesses, which mostly consist of small and medium enterprises, for local employment, productivity growth and job creation is highlighted by many researchers. Haltiwanger et al. (2013) using data from the Census Bureau's Business Dynamics Statistics and Longitudinal Business Database and find that there is a strong relationship between firm size and employment growth. Their findings enhance the important role of business start-ups and young businesses in U.S. job creation. When they do not control for firm age, they find an inverse relationship between net growth rates and firm size. Instead, the negative relationship between firm size and net growth disappears once they control for firm age and may even reverse sign as a result of relatively high rates of exit among the smallest firms. Firm births contribute significantly to both gross and net job creation. More generally, young firms are more volatile and exhibit higher

rates of gross job creation and destruction than their older counterparts. Their results suggest that policies targeting firms of a certain size without taking account of the role of firm age are likely to have limited impact on overall job creation. Lawless (2014) confirms that smaller firms contribute significantly to new job creation and younger firms are consistently more dynamic than older firms but these results arise more from the firm age than the firm size. There is also a strong inverse relationship between employment growth and size for young firms-according to Gibrat's law prediction that size and growth are independent- but this declines very remarkably for older age groups (Haltiwanger et al., 2013).

An OECD (2017) study on regional business demography (active firms, births, deaths and survival rates) across a large set of OECD countries for the years 2007-2014 analyses the variation in and the importance of entrepreneurial activities for regional development. Regions with higher levels of entrepreneurship have on average better local governance, spend more on R&D activities, and have a more educated local workforce. Furthermore, regions with business-friendly regulations (ease of doing business) and a higher quality of governance (i.e. low levels of corruption) have on average higher firm startups. Financing restrictions of firms related with higher rates of firm deaths and lower rates of new firm creation, while additional resources via EU Cohesion Funds can increase both the births and deaths of businesses. Urban regions have the largest levels of business dynamics, both in business creation and destruction rates, while new firms, small and medium enterprises contribute significantly to regional employment growth and can be stimulated by the right set of regional conditions pertaining to local governance,

financing availability and education. This study distinguishes employer firms (those with at least one employee) from the set of all firms and that allows for a much stronger comparability of actual business dynamics, as it mitigates the bias emerging from institutional, taxation and regulatory differences across countries.

Motoyama and Malizia (2017) by using a multivariate cross-sectional analysis of start-up rates in all sectors and in high-tech sectors in 366 metropolitan areas, as defined by the US Census Bureau in 2009, examine the factors that related to higher regional start-up activity. Two hypotheses are defined to analyze start-ups: the demand-pull hypothesis argues that the amount, growth and density of aggregate demand will stimulate start-ups in any sectors; and the supply-push hypothesis argues that factors including high-tech industry concentrations, patent generation, industrial and university research activities, and government funding will stimulate high-tech start-ups. Highly educated or skilled workforce and thick labour markets are important human capital factors for both hypotheses. They find very strong support for the demand-pull hypothesis, but only humble support for the supply-push hypothesis, which provide substantial caveats for public policy to promote start-up activities. Metropolitan regions with larger and denser markets, with more foreign-born and higher labour turnover have higher start-up rates in all sectors. The same applies to high-tech start-up rates, but there is no correlation with other supply-push factors: patents, Research I universities and NIH or SBIR spending. Tertiary education and high-tech industry concentrations are statistically related to high-tech start-up rates.

Adelino et al. (2017) develop an empirical strategy in which cross-sectional variation in local demand shocks-some regions hit harder from a national shock because of their preexisting economic structure-gives rise to variation in sector-specific investment opportunities, which in turn induces firms of different ages to respond by increasing employment. They focus on a region's nontradable sector and consider how the amount of job creation through new firm formation and through the expansion of existing firms responds to changes in local income. Regional variation in income gives rise to variation in investment opportunities, especially for firms that are dependent on local demand and this exogenous variation resolves the reverse causality problem in analyzing the correlation between firm creation and economic growth. New firms account for the bulk of net employment creation in response to local investment opportunities and net employment creation by older firms is heavily concentrated among large older firms: older small firms shed jobs on net as a result of these shocks. The balance of net job creation in response to economic shocks comes from small, new firms and large, old ones. Finally, they find significant gross job creation and destruction by existing firms, suggesting that positive local shocks accelerate churn.

Heyman et al. (2018) examine job creation and productivity dynamics in the Swedish business sector covering the period 1996–2013. When they take firm age into account, they find that it is small, young firms that create most of the net jobs, whereas it is the old, large firms that create most of the increase in productivity. On the other hand, small firms have contributed to the increase in productivity in the service sector. They also find that relatively small changes in net value added

creation mask large values of gross value added creation and gross value added destruction. Moreover, Liang et al. (2018) investigate the correlation between entrepreneurship and population age using the Global Entrepreneurship Monitor data and conclude that older communities have lower entrepreneurship rates at every age. Older workers do not possess the advantages of youth, but older workers with key positions may block younger workers from acquiring business skills. The results imply that a one-standard deviation decrease in the median age of a country increases the rate of new business formation by 2.5 percentage points, which is about forty percent of the mean rate. This effect is statistically and economically significant and robust across different specifications. Workers in countries with a younger workforce will be more entrepreneurial than workers in countries with an older workforce. Younger countries have higher entrepreneurship rates because at any given age, countries with young populations have more entrepreneurship than countries with old populations. Consequently, the entrepreneurship variation is owing to within-age group differences across countries, rather than differences in age-group proportions.

Calvino et al. (2018) explore the characteristics of employment dynamics of new firms, using comparable and highly representative data for 19 OECD countries. They apply a novel analytical decomposition of the contribution of new firms to the creation of new jobs, which allows to separately focusing on different margins. The contribution of new firms to job creation can be expressed as a combination of the following: start-up ratio, survival share, average size of firms at entry and average post-entry growth rate of survivors. The greater cross-country differences are along start-up ratios and post-entry growth, as a result policies might play a significant role

in determining both of these channels for the net job creation by young firms. Brummund and Connolly (2019) examine the association of firm age and size with job creation and destruction in Brazil, by using linked employer–employee RAIS data for 2004–2013. Employment, job creation and job destruction are most concentrated in small-young firms. Firm age is a more important determinant of job creation and employment growth in Brazil than firm size; young firms and start-ups create a disproportionately high number of jobs. However, young firms are also more likely to have much higher exit rates and higher levels of employment volatility. Firm age is also an important determinant of stable employment growth in Brazil. Young firms and large firms have relatively higher stable employment growth rates in Brazil. Furthermore, Eriksson and Rataj (2019) analyse how start-up rates moderate across municipalities in Sweden 2002–2012 by focusing on spatial differences of human capital, social capital, entrepreneurial culture and industrial specialization. They show how the degree of rurality and peripherality moderates the role of different regional resources. Human capital is a significant and positive factor explaining the regional variation of start-up levels but on the other hand, there is no evidence that the role of human capital is stronger in the regions where unemployment levels are low. Moreover, the degree of urbanization (population density) does not constrain the main effect of human capital in both metropolitan and peripheral regions. There is only a significant negative correlation between human capital and population density in regions located near a middle-sized city. Social capital has a substitutional effect on regions with unfavorable economic conditions or with relatively scarce resources and this effect is stronger in more densely populated peripheral settings. Entrepreneurial culture has a positive effect

that remains the same across regions independent of other regional characteristics. On the other hand, this effect does not remain statistically significant for separate categories of regions (metropolitan, close to a city, peripheries). Finally, they confirm the negative effects of industrial specialization on start-up rates and these effects would be less detrimental in regions with low unemployment rates.

Brown et al. (2019) document a large skewness in the employment size distribution by analyzing data on all US employers in a cohort of entering firms. The largest 5% account for over half of cohort employment at firm birth and more than two-thirds at firm age 7. They find that female, African–American, Hispanic, Asian owners and younger founders are initially less likely to start-up large firms but when they control for other variables the coefficients are insignificant. The gender gap persists through firm age 7, while racial and age gaps do not. Immigrants have no significant differences in any specification and military experience is negatively associated with large size. Prior entrepreneurship and founding team size are positively associated with large size, but team diversity is not. They support that large entrants are more likely to be large firms at age 7 and they are also more dynamic than their smaller counterparts. Education is positively correlated with start-up size and at firm age 0 bachelor's degrees are associated with a much higher (31%) probability of large size than either high school or graduate degrees. Controlling for detailed industry raises the graduate education coefficient, implying that this group tends to choose sectors with relatively small firms but in any case, their results do not support an important role for graduate education in producing founders of large start-ups.

Moreover, two other studies address the importance of enterprise births focusing on high growth firms. First, Huber et al. (2013) support that a decrease in average firm age increases the exit hazard of young firms and reduces the probability of observing high-growth firms. They use a three-equation econometric model that simultaneously examines initial firm size, firm survival, and (average) firm growth to estimate firm-specific transition probabilities between size classes (such as the probability to be a high-growth firm) and to explore the determinants of (net) job creation. An increase in the entry rate of new firms and an increase in market growth have no obvious impact on the intra-distribution dynamics of the firm size distribution and, thus, are not able to successfully increase the frequency of high-growth firms. Finally, an increase in a firm's birth size positively affects the smallest and youngest firms. From a policy perspective by investigating the impact of small business and entrepreneurship policies, competition policy and growth-oriented policies they find that all of these policies have important structural implications. Second, Li et al. (2016) examine the self-reported INC5000 data of fastest-growing firms across US counties and conclude that the local conditions fostering high-growth firms are likely to be different from those fostering new firms in general, and that these conditions differ significantly in urban and rural areas and by industrial sectors. The data show that rapidly growing firms also are hosted in non-metro areas and high-growth firms are also found in traditional sector. State-level policy variables have no significant impact but local governments can play major roles in encouraging entrepreneurship and in helping firms succeed. They support that a college-educated work force is crucial to the presence of high-growth firms, whereas a higher high school dropout rate is a statistically significant discouragement in all counties except

those that are non-metro. Greater population density is positively linked to the presence of INC5000 firms and it is significant in both metro and non-metro counties. Faster per capita income growth is associated with a lower presence of high-growth firms in all counties as well as in metro counties. Moreover, the population growth rate turns out to be positive but insignificant except in non-metro counties. They find a positive correlation between the presence of larger existing firms and the number of high-growth firms in all and metro counties as well as in most industries.

Another strand of previous literature analyzes the relationship between enterprise births and innovation because entrepreneurship appears as an important factor boosting innovation and making enterprises more efficient while also generating new employment in regions. Chatterji et al. (2014) review the spatial concentration of entrepreneurship and innovation in the US. They discuss the link between entrepreneurship, innovation and local economic growth and find a very strong connection between initial entrepreneurship and employment growth. They review the spatial concentration of economic activity and the factors that have been found important at the city level: education, age structure, local entrepreneurial culture and physical infrastructure. Finally, they discuss local policy initiatives that encourage entrepreneurship and innovation with three classes of justification: externalities, redistribution, and credit constraints. Even though entrepreneurship is a powerful force for local and economic growth, it is not obvious that government policy can create entrepreneurship. Moreover, Santoleri (2020) examines the link between innovation and employment for startups using the Kauffman Firm Survey, a

dataset for a single cohort of US firms founded in 2004. The findings indicate that R&D and patents seem to be important factors for enhancing employment growth of US startups and especially for new high growth firms. Shrinking firms do not generally benefit from innovation activities but on the contrary, the fastest growing firms indicate that innovation proxies are significantly and positively associated with their performance. This result support the idea that entrepreneurship is not important per se: it is a small subset of highly innovative new firms that foster aggregate economic dynamism.

Faggio and Silva (2014) study the relationship between self-employment, business creation and innovation by combining individual and firm-level data for Britain aggregated at the Travel-to-Work Area level. Their results show that self-employment is positively and significantly correlated with business creation and innovation in urban areas, but not in rural TTWA areas. Rural workers become more self-employed in areas with comparably poor labour market opportunities than urban workers. In order to explain the differences between urban and rural they construct proxies for the lack of employment opportunities. These proxies are negatively and significantly associated with entrepreneurship in both urban and rural areas. This negative link is evident for self-employment in urban areas – but this association turns flat in rural areas. Moreover, the heterogeneity between rural and urban disappears once they study the relationship between self-employment and entrepreneurship controlling for local labour market conditions.

Daunfeldt et al. (2016) examine the association among high-tech firms (HGFs), innovation and growth across all four-digit NACE industries, using data

covering all limited liability firms of all sizes in Sweden during the period 1997–2008. The results indicate that R&D intensity has a negative or no effect on the share of HGFs, regardless of how HGFs were defined. They find that HGFs are overrepresented in knowledge-intensive service industries, i.e., service industries with a high share of human capital. This challenges the prevailing view that R&D is beneficial for high growth. Furthermore, the share of HGFs does not necessarily justify policy interventions because it is not clear if an increase in the share of HGFs will increase industry growth or economic growth more generally.

Zábojník (2020) supports that startup creation depends on firm characteristics, the stage of the business cycle and the availability of financing for entrepreneurs. He studies how a firm's reputation for rewarding innovative employees affects innovation and startup creation and builds a theory in which startup creation is a consequence of a reputational failure. New ideas are more efficiently developed in established firms than in startups, but sale of ideas suffers from the property rights problem. Inventors therefore offer their ideas to an established firm only if the firm has a reputation for rewarding ideas of similar value. Moreover, Farre-Mensa et al. (2020) indicate that startups that win the patent "lottery" by drawing lenient examiners have, on average, 55% higher employment growth and 80% higher sales growth five years later, using unique data on all first-time applications filed at the U.S. Patent Office since 2001. They suggest that startups whose first patent application is approved create more jobs, enjoy faster sales growth and are more innovative than startups with only randomly different inventions that fail to win patent protection. Successful second and third

applications only significantly boost follow-on innovation but not growth. Patents are particularly beneficial for startups founded by inexperienced entrepreneurs, for those located in states with many startups competing for funding and for IT startups. The results show that patents convey substantial economic benefits on startups, not least by facilitating contracting between them and their investors.

3.3 Data and Methodology

3.3.1 Data

I build a data set with European data for 18 countries (see Table 1) at the metropolitan level and examine the major factors that lead to higher regional enterprise births in European metropolitan regions for the period 2008-2013. Metropolitan European regions are defined as urban agglomerations (NUTS level 3 regions or a combination of NUTS level 3 regions) where at least 50% of the population lives inside a functional urban area (Urban Audit's Functional Urban Area-FUA) that is composed of at least 250,000 inhabitants. First, I collect enterprise births as the dependent variable for the total private sector based on the database of Eurostat and at a later stage I include also enterprise births by sector (manufacturing, construction and services) as dependent variable to examine if there is any difference between sectors. A birth amounts to the creation of a combination of production factors with the restriction that no other enterprises are involved in the event. Births do not include entries into the population due to mergers, break-ups, split-off or restructuring of a set of enterprises. It does not include entries into a sub-population resulting only from a change of activity. A birth occurs when an enterprise starts from scratch and actually starts activity. An enterprise creation can

be considered an enterprise birth if new production factors, in particular new jobs, is created. If a dormant unit is reactivated within two years, this event is not considered a birth. The sample consists of 230 metropolitan regions for the years 2008-2013. I use this period because I have no information on the number of enterprise births for the years before 2008. Higher enterprise births for total private sector and services are detected in the metropolitan regions of France, Spain and Portugal. In manufacturing sector Bulgaria has the highest enterprise births and in construction sector France and Lithuania have the highest enterprise births.

The main category of control variables is intellectual property rights. Specifically, I use total patent applications to examine the effect on business births. Total patents are divided by human resources in science and technology, thus taking into account the human capital factor on enterprise births. The source of these data is Eurostat database. Regions showing higher levels of births are more innovative and have a more educated workforce. Patent approvals have a substantial and long-lasting impact on enterprise births (Drivas et al., 2016; Motoyama and Malizia, 2017; Liang et al., 2018; Farre-Mensa et al., 2020). New businesses and especially innovative births can be regarded as manifestations of knowledge spillovers from extant knowledge sources (Fritsch and Wyrwich, 2018). According to this theoretical framework I expect the sign of total patents to be positive. Metropolitan regions of Germany and Netherlands have the highest total patent applications.

I enrich the estimated specifications with control variables in line with related literature (Motoyama and Malizia, 2017; OECD, 2017; Pugsley and Sahin, 2019). All control variables are collected from the databases of Eurostat and European

Commission. I use population density as a demographic factor, which is the measure of agglomeration economies. It is the quotient of total population with land area. The use of metropolitan region population density corresponds to the notion that these regions are defined as being economically integrated and thus firms may take advantage of the broader region's agglomeration effects. I expect the sign of this factor to be positive for enterprise births (Li et al., 2016; Motoyama and Malizia, 2017; OECD, 2017). According to OECD (2017) urbanization factors have positive effect on enterprise births. I use two urbanization factors: the population of the largest city from each metropolitan area divided by total population and the gross value added. Gross value added (GVA) is defined as output (at basic prices) minus intermediate consumption (at purchaser prices); it is the balancing item of the national account's production account. Urban regions are dynamic environments with higher enterprise births, larger and more knowledge-intensive firms. In urban regions, firms find conditions to exist, but also face more competition, especially in large or capital cities. Cities are subject to economies of agglomeration but also to forces of creative destruction. In addition, I use two economic factors, the gross domestic product and the purchasing power standard. A growing economy will naturally create higher enterprise births. The GDP is the total of all value added created in an economy. The value added means the value of goods and services that have been produced minus the value of the goods and services needed to produce them; the so called intermediate consumption. Gross domestic product seems to be a good proxy for stage of development and a reflection of investment potential. Gross domestic product is positively correlated with enterprise births (OECD, 2017; Liang et al., 2018). Purchasing power standard (PPS) is an artificial

currency unit. Theoretically, one PPS can buy the same amount of goods and services in each country. However, price differences across borders mean that different amounts of national currency units are needed for the same goods and services depending on the country. PPS is derived by dividing any economic aggregate of a country in national currency by its respective purchasing power parities. Purchasing power standard gives a more accurate picture about a country's overall standard of living than GDP. Purchasing power is important because, all else being equal, inflation decreases the number of goods or services you would be able to purchase. A decrease in a currency's purchasing power due to high inflation have serious negative economic consequences, including rising costs of goods and services contributing to a high cost of living and high interest rates that affect the global market and falling credit ratings as a result. All of these factors can contribute to an economic crisis. Finally, according to previous studies infrastructure factors are important tools for generating positive externalities (Chatterji et al., 2014). The infrastructure factors included in my sample are the number of motorways and the loading/unloading transport goods. Infrastructure factors seem to affect enterprise births either positively or negatively. Some industries require specialized infrastructure, including physical highways and information highways. As a result many governments have long taken on the role of supporting infrastructure provision.

Table 2 presents the description of the data, the period and the data sources.

(Insert here Table 2)

In addition, summary statistics of the variables of metropolitan regions are

reported in Table 3. All variables are logarithms and also control variables have a time lag of one period. Values of enterprise births range between 4 and 12 for both total and services sector. For manufacturing the values of births range between 1 and 9 and for construction the values range between 1 and 10. The lowest values are detected in manufacturing and construction. Also, construction sector has the highest variation.

(Insert here Table 3)

3.3.2 Methodology

I pursue empirical analysis taking into account country fixed effects because the regressions are likely to suffer from endogeneity in the form of omitted variable bias. If there are omitted variables, and these variables are correlated with the variables in the model, then fixed effects models provide a means for controlling for this bias. The idea is that whatever effects the omitted variables have on the subject at one time, they will also have the same effect at a later time; hence their effects will be constant or fixed over time. Moreover, I use lagged explanatory variables to control for potential reverse causality from the dependent variable to the control variables. In addition, I allow for clustered standard errors at the country level to account for spatial error dependence. Clustered standard errors are measurements that estimate the standard error of a regression parameter in settings where observations may be subdivided into smaller-sized groups ("clusters") and where the sampling is correlated within each group. Specifically, clustered standard errors are identical for situations where observations within each group are not independently and identically distributed. A classic example is if you have many observations for a

panel of firms across time. With cluster standard errors I assume independence across clusters but correlation within clusters.

However, these estimates might still capture reverse causation if there is strong serial correlation in my dependent variables. In light of this, besides cross-section and panel *OLS* estimations I employ *GMM* estimates (Arellano and Bond, 1991; Roodman, 2009). The Arellano–Bond dynamic panel estimator starts by transforming all regressors, usually by differencing, and uses the generalized method of moments (GMM) designed for situations with few time periods and many individuals, with independent variables that are not strictly exogenous, meaning they are correlated with past and possibly current realizations of the error, with fixed effects and with heteroscedasticity and autocorrelation within individuals but not across them (Hansen, 1982; Roodman, 2009). The Sargan/Hansen test for joint validity of the instruments is standard after GMM estimation. The null hypothesis of the test implies all instruments are valid, so if p-value is greater than 5% (0.05) I accept the null hypothesis (Ho). The higher the p-value is of the Sargan/Hansen statistic the better. However, according to Roodman (2006), it is recommended that Sargan/Hansen p-value should be greater than 0.25. One more contribution of Arellano and Bond is a test (AR(2)) for autocorrelation appropriate for linear regressions on panels, which is especially important when lags are used as instruments. The null hypothesis of the test implies no second-order serial correlation in disturbances.

Furthermore, I conduct spatial analysis because it is widely recognized that sample data collected from geographically close entities are not independent but

spatially correlated, which means observations of closer units tend to be more similar than observations of further units and spatial panel-data models have become a popular tool for measuring spatial spillovers (Belotti et al., 2017). Spatial clustering is often observed for a large number of variables i.e. economic variables, demographic variables. Spatial models allow for spillover effects from nearby areas. In spatial data, observations are areas. I use a spatial panel fixed effect model with a spatial inverse distance weighting matrix. Fixed effects implement the quasi–maximum likelihood (QML) estimator in Lee and Yu (2010a) to fit the model. Fixed-effects estimator does not give consistent estimates for the levels of the panel fixed effects nor for their standard deviation. Particularly, fixed effect estimator removes the panel-level effects from the estimation and no distributional assumptions are made about them and also it allows the panel-level effects to be correlated with the observed covariates. Only covariates that vary across both panels and time can be fit with this estimator. The weighting matrix is based on the inverse distance between metropolitan regions and this spatial weighting matrix defines a spatial lag of the dependent variable. In order to create this matrix I use the coordinates of each metropolitan region. Inverse distance makes the assumption that the greater the distance between regions the smaller the effect. The measured values closest to the prediction location have more influence on the predicted value than those further away. Inverse distance assumes that each measured point has a local influence that diminishes with distance. In addition, I check for endogeneity through a spatial IV model and endogeneity does not exist.

3.4 Empirical Results

3.4.1 Baseline Results

In the empirical analysis, I first estimate the following regression model:

$$log(births)_{itg} = \beta o + \beta_1 log(popdens) + \beta_2 log(patents) + \beta_3 log(pps) + \mu_i + \varepsilon_i, \quad (1)$$

where *births* denotes enterprise births, *popdens* is population density, *patents* denotes the total patent applications divided by human resources in science and technology, *pps* is the purchasing power standard, *i s*tands for metropolitan region, *t* denotes the period, *g* denotes sector (total, manufacturing, construction, services), μ_i represents regional (country) fixed effects and ε_j comprises the non-observed influences on enterprise births. All variables are logarithms and also control variables have a time lag of one period.

I begin with cross-section and panel *OLS* estimations. Table 4 provides fixed effect estimations of enterprise births for all sectors on population density, total patents and purchasing power standard for European metropolitan regions in 2008-2013. I use fixed effects and cluster standard errors at the country level. Enterprise births are highly correlated with all three independent variables. For example, population density and total patents have positive and statistically significant relationships with enterprise births, while purchasing power standard is negatively correlated with enterprise births. These results are in line with the literature (see for example Drivas et al., 2016; Motoyama and Malizia, 2017; Liang et al., 2018; Farre-Mensa et al., 2020). Regions with high innovation and highly educated workforce have more enterprise births. The case of the purchasing power standard is excluded because it does not exist in previous studies as a control variable of enterprise births.

The results remain the same across all sectors. A one-standard deviation in total patents increases enterprise births by 14% (Column 2, Table 4). A 1% increase in population density in a metropolitan region increases total enterprise births on average by 42% (Column 1, Table 4) and a 1% increase in purchasing power standard decreases enterprise births by approximately 57% (Column 8, Table 4).

(Insert here Table 4)

I continue my analysis with the same cross-section and panel *OLS* fixed effect estimations as in equation (1) but using more control variables. Particularly, I use as control variables total patent applications, purchasing power standard or GDP, motorways, gross value added of secondary and tertiary sector divided by total gross value added, urbanization (population of the largest city divided by total population) and loading/unloading transport goods. Tables 5-8 provide these cross-section and panel *OLS* estimates of enterprise births in the different sectors (total, manufacturing, construction and services) with total patent applications as control variable. I use fixed effects and cluster standard errors at country level. Total patents are positively correlated with all enterprise births (at 1% level of significance). So, a one-standard deviation in total patents increases total enterprise births by 94% (Column 1, Table 5), manufacturing enterprise births by 86% (Column 1, Table 6), construction enterprise births by 83% (Column 1, Table 7) and services enterprise births by 96% (Column 1, Table 8). It is obvious that these percentages differ by estimate and mainly between cross-section and panel estimates where in the case of cross-section they are higher but the coefficients remain the same across all sectors. Urbanization factors are strongly and positively correlated with enterprise births (Li et al., 2016; Motoyama and Malizia, 2017; OECD, 2017). Urban regions are dynamic

environments with higher enterprise births where firms find conditions to exist. A 1% increase in urbanization in a metropolitan region increases total enterprise births on average by 44% (Column 6, Table 5). Motorways and GDP are positively correlated with enterprise births (Chatterji et al., 2014; OECD, 2017; Liang et al., 2018) while loading and unloading transport goods have negative relationships with enterprise births. This study is the first one that uses loading/unloading transport goods as control variable on enterprise births for European metropolitan regions. A 1% increase in motorways increases manufacturing enterprise births on average by 47% (Column 2, Table 6), while a 1% increase in GDP increases construction enterprise births by 70% (Column 5, Table 7). Better infrastructure factors foster enterprise births. A one-standard deviation in loading transport goods decreases services enterprise births by 35% (Column 1, Table 8).

(Insert here Tables 5-8)

3.4.2 GMM Results

Next, I use difference *GMM* estimates to address endogeneity. I apply and estimate the following equation:

$$log(births)_{itg} = \alpha log(births_{t-1}) + \beta_0 + \beta_1 log(popdens) + \beta_2 log(patents) + \beta_3 log(pps) + \mu_i + \varepsilon_i, \quad (2)$$

where *births* denotes enterprise births, $births_{t-1}$ denotes the first lag of enterprise births, *popdens* is population density, *patents* denotes total patents divided by human resources in science and technology, *pps* is the purchasing power standard, *i* *st*ands for metropolitan region, *t* denotes the period, *g* denotes the sector (total, manufacturing, construction, services), μ_i represents regional (country) fixed effects and ε_j comprises the non-observed influences on enterprise births. All variables are

logarithms and I use the first two lags of all control variables as instruments.

Table 9 presents difference *GMM* estimates with enterprise births as dependent variable. Population density, total patent applications and purchasing power standard are used as control variables. Total patent applications have positive and statistically significant signs. A one-standard deviation in total patents in a metropolitan region increases total enterprise births by 37% (Column 1, Table 9). Population density has a positive effect on enterprise births, while purchasing power standard has a negative effect on enterprise births. The first lags of enterprise births have positive and statistically significant coefficients. Arellano-Bond's test (AR(2)) null hypothesis implies no second-order serial correlation in disturbances. In my estimates the p-value of AR(2) test is greater than 5% (0.05) for all cases, so I accept the null hypothesis that there is no second-order. Moreover, the Hansen test null hypothesis implies that the over-identified restrictions are satisfied. Here the p-value of all estimates is greater than 5% (0.05), so I accept the null hypothesis that over-identified restrictions are satisfied.

(Insert here Table 9)

3.4.3 Spatial Results

I construct maps of enterprise births (total private sector, manufacturing, construction and services) in European metropolitan regions from 2008 to 2013. I combine the data of enterprise births from Eurostat with the coordinates at the European level of NUTS3 which are available in Eurostat, JRC and European Commission Directorate. I construct these maps to illustrate virtually the spatial clustering between neighboring metropolitan regions because it is widely known

that geographically close regions are spatially correlated and the observations of neighboring metropolitan regions tend to be more similar than observations in non-adjacent metropolitan regions. I present four maps, one for each enterprise birth per sector. It is evident that neighboring metropolitan regions are spatially correlated because the values of enterprise births in these regions are similar to each other.

(Insert here Maps 1, 2, 3 and 4)

I provide a spatial panel fixed-effect analysis (quasi–maximum likelihood (QML) estimator) with a spatial inverse distance weighting matrix in order to take into account the spatial clustering of the variables:

$$log(births)_{ntig} = \lambda W(log(births)) + \beta o + \beta_1 log(popdens) + \beta_2 log(patents) + \beta_3 log(pps) + \mu_i + \varepsilon_i \quad (3)$$

where *births* is an *n* X 1 vector of observations for the dependent variable (enterprise births) for time period *t* with *n* number of panels, *popdens* is a matrix of population density, *patents* is a matrix of the total patents divided by human resources in science and technology, *pps* is a matrix of the purchasing power standard, *i st*ands for metropolitan region, *g* denotes sector (total, manufacturing, construction, services), μ_i represents fixed effects, ε_j comprises the non-observed influences on enterprise births and *W* is the spatial inverse distance weighting matrix. All variables are logarithms and also control variables have a time lag of one period.

Table 10 provides fixed effect spatial estimates of enterprise births for all sectors on population density, total patents and purchasing power standard for European metropolitan regions for the period 2008-2013. Independent variables have no statistically significant coefficients but the spatial term of the dependent

variable is positive and statistically significant (at 1% level of significance) for all cases. This means that enterprise births of a neighboring metropolitan region increase the enterprise births for all sectors. Particularly, a 1% increase in neighboring enterprise births in a metropolitan region increases total enterprise births on average by 76% (Column 1, Table 10). I end up that my data are not independent but spatially correlated, which means observations of closer metropolitan regions tend to be more similar than observations of further metropolitan regions.

(Insert here Table 10)

I continue my analysis with the same spatial fixed effect estimates as in equation (3) but more control variables. I use as control variables total patent applications, purchasing power standard or GDP, motorways, gross value added of secondary and tertiary sector divided by total gross value added, urbanization (population of the largest city divided by total population) and loading/unloading transport goods. Tables 11 and 12 provide these spatial fixed effect estimates of enterprise births in the different sectors (total, manufacturing, construction and services) with total patent applications as control variable. Many control variables have statistically significant coefficients. Specifically, motorways, urbanization and gross value added are strongly and positively correlated with enterprise births as in the baseline estimates (Chatterji et al., 2014; Li et al., 2016; OECD, 2017). A one-standard deviation in motorways increases manufacturing enterprise births by 46% (Column 4, Table 11), while a one-standard deviation in urbanization increases total enterprise births by 89% (Column 1, Table 11). On the other hand, GDP have negative relationships with enterprise births which contradicts the baseline

estimates. The spatial term of enterprise births is positive and statistically significant (at 1% level of significance) for all cases and for all sectors, which confirms spatial clustering. Enterprise births of a neighboring metropolitan region have a positive effect on the births. A 1% increase in neighboring enterprise births in a metropolitan region increases total enterprise births by approximately 60% (Column 1, Table 11), while a 1% increase in neighboring enterprise births in a metropolitan region increases construction enterprise births by 50% (Column 1, Table 12).

(Insert here Tables 11 and 12)

3.5 Concluding Remarks

Business dynamism in the United States and in many countries has been declining in the last several decades because of financial crisis. Theoretical and empirical studies support that entrepreneurship and particularly enterprise births are important drivers of job creation and long-term economic growth. Business startups account for about 20 percent of US gross (total) job creation while high-growth businesses (which are disproportionately young) account for almost 50 percent of gross job creation. Recently literature focuses on firm births, its determinants, and its economic impacts (Decker et al., 2016; Gourio et al., 2016; OECD, 2017; Bijnens and Konings, 2018; Pugsley and Sahin, 2019). Regions with higher levels of business births are characterized by better local governance, spend more on intellectual property rights and have a more educated population. Business creation is a vital source of innovation as patent applications have a substantial and long-lasting impact on enterprise births (Farre-Mensa et al., 2020). Innovative

enterprise births can be observed as manifestations of knowledge spillovers extant knowledge sources (Fritsch and Wyrwich, 2018).

This study conducts a systematic examination of how different factors (i.e. demographic, intellectual property rights) affect enterprise births in all sectors. Most literature on firm's births focuses on USA areas and on specific countries, while this study examines metropolitan regions of 18 European countries. Although I include in my analysis a large set of control variables the effect of these variables on enterprise births does not change. My empirical results suggest that intellectual property rights (total patent applications) divided by human resources in science and technology have a substantial positive correlation with enterprise births. This positive relationship is based on theoretical grounds. Demographic factors such as population density and urbanization foster enterprise births. Gross domestic product and gross value added are positively correlated with enterprise births except in spatial analysis where GDP has negative coefficients. Purchasing power standard and loading/unloading transport goods hamper enterprise births in all sectors while motorways has an important positive correlation with the development of enterprise births in all sectors. These findings were found to be robust across alternative specifications.

My work highlights that metropolitan areas with larger, innovative, denser markets and higher human resources have higher enterprise births in all sectors. The evidence underlines the importance of innovation in the development process of enterprise births.

Appendix of Chapter 3

Tables

Table 1

List of Countries in the sample

Country
Austria
Bulgaria
Croatia
Czech Republic
Denmark
Estonia
Finland
France
Germany
Hungary
Italy
Lithuania
Netherlands
Poland
Portugal
Romania
Slovakia
Spain

Table 2

Definition of Variables and Data Sources

Variable	Description	Period	Source
Enterprise births	Births of enterprises (total private sector, manufacturing, construction, services)	2008-2016	Eurostat
Patents	Total patents divided by human resources in science and technology	1999-2012	Eurostat
Population density	The quotient of population with land area	1990-2016	Eurostat
Urbanization	Urbanization rate (population of the largest cities divided by total population)	1990-2016	Eurostat
Purchasing power standard	Purchasing power standard	2000-2018	Eurostat
GDP	Gross domestic product	1980-2018	European Commission-Ardeco online
Gross value added	Gross value added secondary and tertiary sectors divided by total gross value added	1980-2018	European Commission-Ardeco online
Roads	Motorways(kilometers per thousand square kilometers)	1990-2018	Eurostat
Loading transport goods	Loading transport goods divided by total population or GDP	2008-2018	Eurostat
Unloading transport goods	Unloading transport goods divided by total population or GDP	2008-2018	Eurostat

Table 3

Descriptive Statistics of European Metropolitan Regions

Variable	Mean	Standard Deviation	Min	Max	Observations
log(total enterprise births)	7.1004	1.5197	3.8501	11.7479	2,575
log(manufacturing enterprise births)	4.2391	1.5241	1.0986	8.6917	2,573
log(construction enterprise births)	4.8952	1.5983	1.0986	9.5352	2,585
log(services enterprise births)	6.8907	1.5231	3.6109	11.6282	2,585
lag1(log(total patents/human resources)	.2883	.3898	.0002	4.2823	3,497
lag1(log(population density))	5.6162	1.0029	3.3068	9.6461	5,230
lag1(log(purchasing power standard))	10.1210	.3846	8.2940	11.2516	3,843
lag1(log(GDP))	9.4992	.9249	6.1228	12.3239	10,667
lag1(log(motorways))	3.0015	1.2013	0	5.3279	5,644
lag1(log (gross value added secondary and services/total gross value added))	4.5755	.0646	3.8313	4.6052	6,339
lag1(log(city population/total population))	3.6470	.4821	1.1641	4.6052	4,164
lag1(log(loading transport goods/total population))	3.1940	.5619	.4965	4.4071	2,432
lag1(log(unloading transport goods/total population))	3.2081	.5376	.4737	4.4201	2,432
lag1(log(loading transport goods/GDP))	7.0879	.7927	3.9117	10.2938	2,464
lag1(log(unloading transport goods/GDP))	7.1025	.7742	4.3904	10.2525	2,464

Table 4

Baseline Estimations

	OLS	Panel	OLS	Panel	OLS	Panel	OLS	Panel
Dependent variable: log(enterprise births)	total private sector	total private sector	manufacturing	manufacturing	construction	construction	services	services
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
lag1(log(population density))	0.427***	1.041	0.400***	0.268	0.402***	0.949	0.427***	1.079**
	(3.76)	(1.55)	(5.01)	(0.17)	(7.27)	(0.49)	(3.44)	(2.21)
lag1(log((total patents/human resources))	0.734***	0.148**	0.632***	0.235***	0.572***	0.052	0.766***	0.148**
	(18.22)	(2.55)	(17.32)	(3.91)	(16.80)	(0.90)	(17.49)	(2.55)
lag1(log(purchasing power standard))	0.287	-0.645***	-0.136	-1.329***	0.173	-0.812***	0.323	-0.570**
	(1.43)	(3.07)	(0.90)	(4.20)	(1.31)	(4.53)	(1.36)	(2.75)
constant	2.448	7.601*	3.881*	15.988**	1.219	7.565	1.935	6.434
	(1.48)	(1.90)	(2.15)	(2.27)	(0.84)	(0.80)	(0.99)	(1.75)
R^2	0.84	0.18	0.81	0.17	0.81	0.10	0.83	0.15
Observations	1,164	1,164	1,163	1,163	1,166	1,166	1,166	1,166

Notes: Cross-section (Columns 1, 3, 5, 7) and panel (2, 4, 6, 8) *OLS* regressions with enterprise births in all sectors as dependent variable in EU metropolitan regions. Fixed effects and cluster standard errors at country level. *, **, *** denote significance at 10%, 5%, 1% levels, respectively.

Table 5

Additional Estimations

	OLS	Panel	OLS	Panel	OLS	Panel
Dependent variable: log(enterprise births)	total private sector	total private sector	total private sector	total private sector	total private sector	total private sector
	(1)	(2)	(3)	(4)	(5)	(6)
lag1(log(total patents/human resources))	0.945***	0.068	0.996***	0.064	0.182***	0.167
	(7.46)	(1.01)	(7.25)	(0.95)	(3.33)	(1.16)
lag1(log(purchasing power standard))	0.413	0.104	0.068	0.057		
	(0.95)	(0.17)	(0.13)	(0.10)		
lag1(log(motorways))	0.102	0.570***	0.124	0.578***	0.061	0.553***
	(1.09)	(5.34)	(1.37)	(5.35)	(1.18)	(3.73)
lag1(log(gross value added 2nd and 3rd/total gross value added))	10.668***	12.511**	11.590***	12.486**	1.114*	10.872*
	(5.81)	(2.93)	(4.42)	(2.92)	(2.10)	(2.12)
lag1(log(city population/total population))	0.216	1.076	0.158	1.090	0.044	0.441***
	(1.10)	(0.97)	(0.99)	(0.97)	(0.62)	(3.11)
lag1(log(loading transport goods/total population))	-0.324***	-0.070				
	(5.67)	(0.81)				
lag1(log(loading transport goods/GDP))			-0.239***	-0.065		
			(3.20)	(0.78)		
lag1(log(GDP))					0.698***	-0.129
					(10.47)	(0.42)
lag1(log(unloading transport goods/total population))					-0.356**	-0.097
					(2.30)	(0.85)
constant	-45.819***	-56.624*	-45.875***	-55.869*	-3.202	-44.073
	(6.65)	(2.12)	(5.19)	(2.14)	(1.37)	(1.67)
R^2	0.87	0.36	0.87	0.35	0.93	0.32
Observations	393	393	393	393	489	489

Notes: Cross-section (Columns 1, 3, 5) and panel (2, 4, 6) *OLS* regressions with enterprise births in total private sector as dependent variable in EU metropolitan regions. Fixed effects and cluster standard errors at country level. *, **, *** denote significance at 10%, 5%, 1% levels, respectively.

Table 6

Additional Estimations

	OLS	Panel	OLS	Panel	OLS	Panel
Dependent variable: log(enterprise births)	manufacturing	manufacturing	manufacturing	manufacturing	manufacturing	manufacturing
	(1)	(2)	(3)	(4)	(5)	(6)
lag1(log(total patents/human resources))	0.860***	0.374***	0.888***	0.371***	0.138***	0.390***
	(8.87)	(4.57)	(8.86)	(4.48)	(4.09)	(3.78)
lag1(log(purchasing power standard))	0.202	0.021	-0.074	-0.017		
	(0.68)	(0.05)	(0.21)	(0.04)		
lag1(log(motorways))	0.122	0.478***	0.136*	0.487***	0.044	0.454***
	(1.62)	(6.42)	(1.87)	(6.44)	(1.12)	(4.53)
lag1(log(gross value added 2nd and 3rd/total gross value added))	9.504***	10.364**	10.020***	10.338**	0.630*	8.668*
	(6.53)	(2.55)	(6.02)	(2.53)	(1.78)	(1.93)
lag1(log(city population/total population))	0.050	0.696	-0.054	0.710	-0.118	0.536***
	(0.24)	(0.92)	(0.32)	(0.92)	(1.37)	(5.59)
lag1(log(loading transport goods/total population))	-0.260***	-0.058				
	(5.61)	(0.72)				
lag1(log(loading transport goods/GDP))			-0.287**	-0.051		
			(2.90)	(0.64)		
lag1(log(GDP))					0.641***	-0.223
					(11.02)	(0.94)
lag1(log(unloading transport goods/total population))					-0.290**	-0.057
					(2.43)	(0.56)
constant	-41.123***	-47.319*	-39.269***	-46.707*	-3.194*	-36.216
	(5.70)	(1.98)	(5.04)	(2.00)	(1.87)	(1.60)
R^2	0.86	0.11	0.86	0.11	0.91	0.12
Observations	393	393	393	393	489	489

Notes: Cross-section (Columns 1, 3, 5) and panel (2, 4, 6) *OLS* regressions with enterprise births in manufacturing as dependent variable in EU metropolitan regions. Fixed effects and cluster standard errors at country level. *, **, *** denote significance at 10%, 5%, 1% levels, respectively.

Table 7

Additional Estimations

	OLS	Panel	OLS	Panel	OLS	Panel
Dependent variable: log(enterprise births)	construction	construction	construction	construction	construction	construction
	(1)	(2)	(3)	(4)	(5)	(6)
lag1(log(total patents/human resources))	0.832***	0.150	0.836***	0.155	0.013	0.202
	(6.22)	(0.74)	(6.35)	(0.75)	(0.17)	(0.99)
lag1(log(purchasing power standard))	0.132	-0.071	0.001	-0.015		
	(0.35)	(0.13)	(0.00)	(0.03)		
lag1(log(motorways))	0.121*	0.344***	0.125*	0.335***	0.062	0.331***
	(1.86)	(5.49)	(1.99)	(4.99)	(1.71)	(3.13)
lag1(log(gross value added 2nd and 3rd/total gross value added))	12.555***	11.870***	12.643***	11.897***	0.861	10.404***
	(5.12)	(4.01)	(5.28)	(4.03)	(1.51)	(3.03)
lag1(log(city population/total population))	0.257	0.745	0.167	0.729	0.043	0.356***
	(1.14)	(1.04)	(0.81)	(1.02)	(0.46)	(3.66)
lag1(log(loading transport goods/total population))	-0.124*	0.084				
	(1.98)	(1.12)				
lag1(log(loading transport goods/GDP))			-0.204*	0.079		
			(1.87)	(1.05)		
lag1(log(GDP))					0.704***	0.098
					(11.82)	(0.20)
lag1(log(unloading transport goods/total population))					-0.221	0.075
					(1.41)	(0.86)
constant	-55.036***	-52.736***	-52.816***	-53.649***	-5.110*	-45.838**
	(6.12)	(3.19)	(6.68)	(3.36)	(2.08)	(2.54)
R^2	0.87	0.14	0.87	0.14	0.92	0.12
Observations	393	393	393	393	489	489

Notes: Cross-section (Columns 1, 3, 5) and panel (2, 4, 6) *OLS* regressions with enterprise births in construction as dependent variable in EU metropolitan regions. Fixed effects and cluster standard errors at country level. *, **, *** denote significance at 10%, 5%, 1% levels, respectively.

Table 8

Additional Estimations

	OLS	Panel	OLS	Panel	OLS	Panel
Dependent variable: log(enterprise births)	services	services	services	services	services	services
	(1)	(2)	(3)	(4)	(5)	(6)
lag1(log(total patents/human resources))	0.965***	0.046	1.024***	0.040	0.201***	0.153
	(7.49)	(0.64)	(7.22)	(0.53)	(3.37)	(1.08)
lag1(log(purchasing power standard))	0.462	0.135	0.085	0.069		
	(1.00)	(0.22)	(0.16)	(0.12)		
lag1(log(motorways))	0.100	0.599***	0.125	0.609***	0.060	0.583***
	(0.98)	(5.13)	(1.27)	(5.17)	(1.05)	(3.64)
lag1(log(gross value added 2nd and 3rd/total gross value added))	10.444***	12.787**	11.499***	12.756**	1.174**	11.093*
	(5.44)	(2.81)	(4.17)	(2.79)	(2.22)	(2.04)
lag1(log(city population/total population))	0.227	1.114	0.175	1.132	0.058	0.445**
	(1.15)	(0.95)	(1.11)	(0.95)	(0.81)	(2.96)
lag1(log(loading transport goods/total population))	-0.354***	-0.098				
	(6.07)	(1.09)				
lag1(log(loading transport goods/GDP))			-0.241***	-0.092		
			(3.41)	(1.07)		
lag1(log(GDP))					0.707***	-0.124
					(10.18)	(0.38)
lag1(log(unloading transport goods/total population))					-0.377**	-0.128
					(2.38)	(1.10)
constant	-45.356***	-58.530*	-45.825***	-57.467*	-3.675	-45.335
	(6.18)	(2.06)	(5.01)	(2.07)	(1.58)	(1.63)
R^2	0.87	0.36	0.86	0.36	0.92	0.33
Observations	393	393	393	393	489	489

Notes: Cross-section (Columns 1, 3, 5) and panel (2, 4, 6) *OLS* regressions with enterprise births in services as dependent variable in EU metropolitan regions. Fixed effects and cluster standard errors at country level. *, **, *** denote significance at 10%, 5%, 1% levels, respectively.

Table 9

Difference GMM estimations

	GMM	GMM	GMM	GMM
Dependent variable: log(enterprise births)	total private sector	manufacturing	construction	services
	(1)	(2)	(3)	(4)
lag2(log(population density))	2.506***	6.915***	-0.896	2.438***
	(3.94)	(4.03)	(0.78)	(3.66)
lag2(log(total patents/ human resources)	0.374***	0.749***	0.110	0.402***
	(5.39)	(3.24)	(0.88)	(5.39)
lag2(log(purchasing power standard))	-0.172**	-0.863***	-0.702***	-0.081
	(2.15)	(4.88)	(4.68)	(1.06)
lag(enterprise births)	0.522***	0.269***	0.283***	0.546***
	(10.35)	(5.59)	(4.77)	(11.86)
Observations	887	887	887	887
Arellano-Bond test AR(2)	z=1.15	z=-0.58	z=1.06	z=1.69
	Pr>z=0.249	Pr>z=0.561	Pr>z=0.288	Pr>z=0.092
Hansen test	chi2(88)=85.70	chi2(88)=91.78	chi2(88)=90.84	chi2(88)=87.68
	Prob>chi2=0.549	Prob>chi2=0.370	Prob>chi2=0.397	Prob>chi2=0.490

Notes: Difference *GMM* estimates with enterprise births in all sectors as dependent variable in EU metropolitan regions. Fixed effects and cluster standard errors at country level. *, **, *** denote significance at 10%, 5%, 1% levels, respectively. Arellano-Bond's test AR(2) null hypothesis implies no second-order serial correlation in disturbances. Hansen's test null hypothesis implies all instruments are valid.

Table 10

Spatial estimations

	Spatial	Spatial	Spatial	Spatial
Dependent variable: log(enterprise births)	total private sector	manufacturing	construction	services
	(1)	(2)	(3)	(4)
lag1(log(population density))	0.332	0.886	-0.297	0.460
	(0.71)	(1.09)	(0.42)	(0.94)
lag1(log(patents/human resources))	-0.046	-0.122	-0.342***	0.016
	(0.53)	(0.80)	(2.59)	(0.18)
lag1(log(purchasing power standard))	-0.121	-0.022	0.082	-0.132
	(0.93)	(0.10)	(0.42)	(0.97)
neighboring enterprise births	0.763***	0.561***	0.474***	0.812***
	(8.27)	(4.34)	(2.64)	(10.60)
constant	0.130***	0.226***	0.197***	0.136***
	(34.32)	(34.43)	(34.42)	(34.29)
Observations	714	714	714	714
Wald test of spatial terms	chi2(1)=68.47	chi2(1)=18.86	chi2(1)=6.98	chi2(1)=112.36
	Prob>chi2=0.0000	Prob>chi2=0.0000	Prob>chi2=0.0082	Prob>chi2=0.0000

Notes: Fixed effects spatial panel (quasi-maximum likelihood) regressions with enterprise births in all sectors as dependent variable in EU metropolitan regions. *, **, *** denote significance at 10%, 5%, 1% levels, respectively.

Table 11

Spatial estimations

	Spatial	Spatial	Spatial	Spatial	Spatial	Spatial
Dependent variable: log(enterprise births)	total private sector	total private sector	total private sector	manufacturing	manufacturing	manufacturing
	(1)	(2)	(3)	(4)	(5)	(6)
lag1(log(total patents/human resources))	-0.097	-0.084	-0.196	0.319	0.335	0.217
	(0.30)	(0.26)	(0.62)	(0.66)	(0.70)	(0.46)
lag1(log(purchasing power standard))	-0.257	-0.293		-0.183	-0.216	
	(0.95)	(1.09)		(0.46)	(0.55)	
lag1(log(motorways))	0.483***	0.494***	0.450***	0.464***	0.475***	0.425***
	(4.58)	(4.69)	(4.42)	(2.98)	(3.07)	(2.77)
lag1(log(gross value added 2nd and 3rd/total gross value added))	12.466***	12.422***	13.036***	9.518***	9.472***	10.089***
	(6.38)	(6.35)	(6.97)	(3.30)	(3.29)	(3.58)
lag1(log(city population/total population))	0.890***	0.891***	0.842***	0.739	0.739	0.696
	(2.86)	(2.86)	(2.80)	(1.61)	(1.61)	(1.54)
lag1(log(loading transport goods/total population))	-0.061			-0.057		
	(1.11)			(0.69)		
lag1(log(loading transport goods/GDP))		-0.045			-0.039	
		(0.81)			(0.47)	
lag1(log(GDP))			-1.344***			-1.147**
			(3.61)			(2.04)
lag1(log(unloading transport goods/total population))			-0.060			-0.076
			(1.02)			(0.85)
neighboring enterprise births	0.606***	0.612***	0.612***	0.468***	0.474***	0.469***
	(4.04)	(4.11)	(4.13)	(2.94)	(2.99)	(2.94)
constant	0.157***	0.158***	0.152***	0.232***	0.232***	0.229***
	(18.50)	(18.49)	(18.49)	(18.64)	(18.64)	(18.64)
Observations	220	220	220	220	220	220
Wald test of spatial terms	chi2(1)=16.29	chi2(1)=16.88	chi2(1)=17.09	chi2(1)=8.63	chi2(1)=8.91	chi2(1)=8.62
	Prob>chi2=0.0001	Prob>chi2=0.0000	Prob>chi2=0.0000	Prob>chi2=0.0033	Prob>chi2=0.0028	Prob>chi2=0.0033

Notes: Fixed effects spatial panel (quasi-maximum likelihood) regressions with enterprise births in total private sector and manufacturing as dependent variable in EU metropolitan regions. *, **, *** denote significance at 10%, 5%, 1% levels, respectively.

Table 12

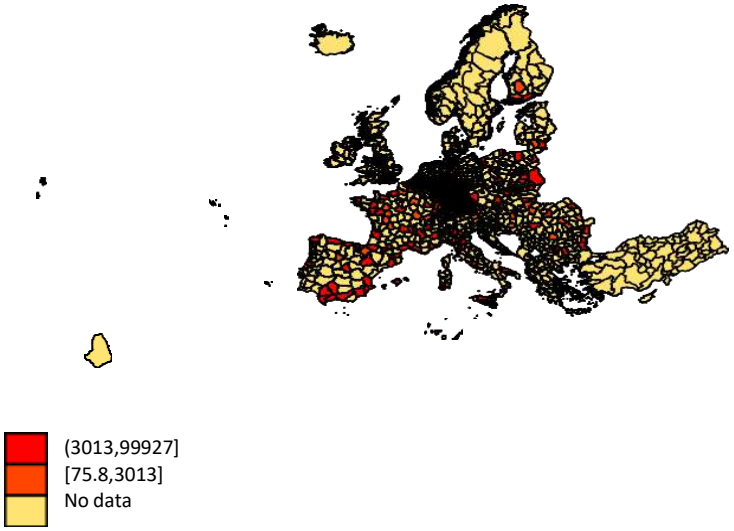
Spatial estimations

	Spatial	Spatial	Spatial	Spatial	Spatial	Spatial
Dependent variable: log(enterprise births)	construction	construction	construction	services	services	services
	(1)	(2)	(3)	(4)	(5)	(6)
lag1(log(total patents/human resources))	-0.394	-0.388	-0.435	-0.094	-0.081	-0.199
	(1.01)	(1.00)	(1.12)	(0.27)	(0.23)	(0.60)
lag1(log(purchasing power standard))	-0.065	-0.081		-0.335	-0.379	
	(0.20)	(0.25)		(1.18)	(1.34)	
lag1(log(motorways))	0.311**	0.315**	0.303**	0.503***	0.515***	0.467***
	(2.48)	(2.52)	(2.43)	(4.54)	(4.66)	(4.38)
lag1(log(gross value added 2nd and 3rd/total gross value added))	11.168***	11.151***	11.525***	13.017***	12.968***	13.579***
	(4.78)	(4.77)	(5.00)	(6.34)	(6.31)	(6.93)
lag1(log(city population/total population))	0.808**	0.809**	0.788**	0.893***	0.894***	0.837***
	(2.18)	(2.18)	(2.14)	(2.73)	(2.73)	(2.66)
lag1(log(loading transport goods/total population))	-0.027			-0.074		
	(0.41)			(1.27)		
lag1(log(loading transport goods/GDP))		-0.020			-0.057	
		(0.31)			(0.96)	
lag1(log(GDP))			-0.684			-1.491***
			(1.49)			(3.82)
lag1(log(unloading transport goods/total population))			-0.009			-0.073
			(0.13)			(1.18)
neighboring enterprise births	0.501***	0.502***	0.495***	0.602***	0.608***	0.605***
	(2.81)	(2.82)	(2.77)	(4.00)	(4.08)	(4.06)
constant	0.188***	0.188***	0.187***	0.165***	0.166***	0.159***
	(18.57)	(18.57)	(18.58)	(18.51)	(18.50)	(18.51)
Observations	220	220	220	220	220	220
Wald test of spatial terms	chi2(1)=7.92	chi2(1)=7.96	chi2(1)=7.66	chi2(1)=16.03	chi2(1)=16.64	chi2(1)=16.51
	Prob>chi2=0.0049	Prob>chi2=0.0048	Prob>chi2=0.0056	Prob>chi2=0.0001	Prob>chi2=0.0000	Prob>chi2=0.0000

Notes: Fixed effects spatial panel (quasi-maximum likelihood) regressions with enterprise births in construction and services as dependent variable in EU metropolitan regions. *, **, *** denote significance at 10%, 5%, 1% levels, respectively.

Map 1

Total Private Sector Startups

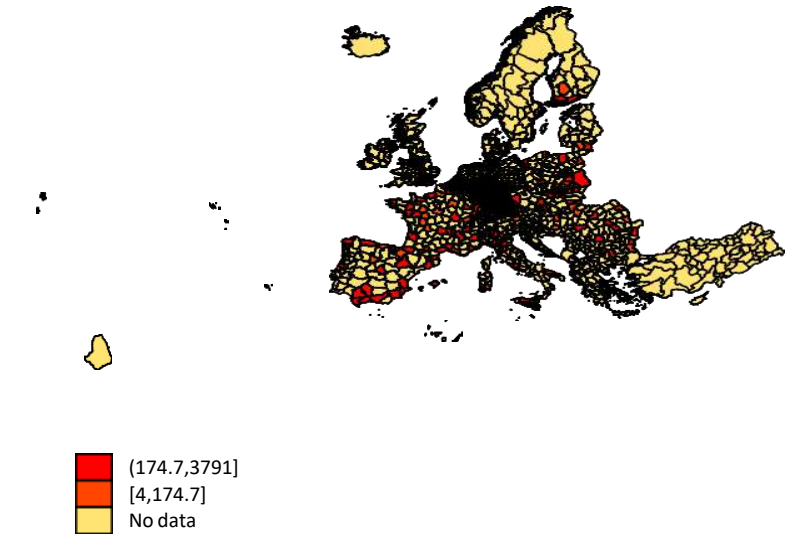


Note: Based on total startups in EU metropolitan regions from 2008 to 2013 and at the level of NUTS3 2016.

Source: Eurostat, JRC and European Commission Directorate.

Map 2

Manufacturing Startups

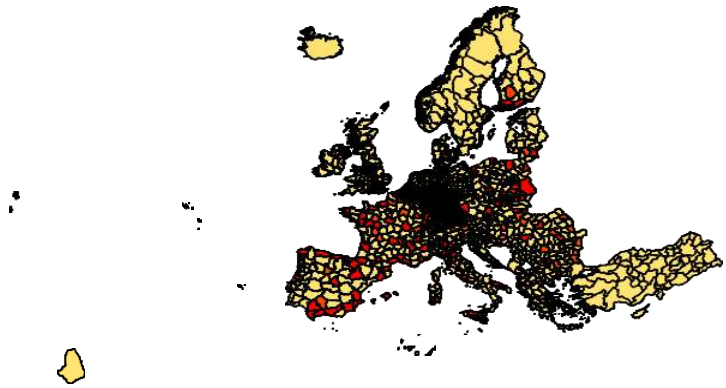


Note: Based on manufacturing startups in EU metropolitan regions from 2008 to 2013 and at the level of NUTS3 2016.

Source: Eurostat, JRC and European Commission Directorate.

Map 3

Construction Startups



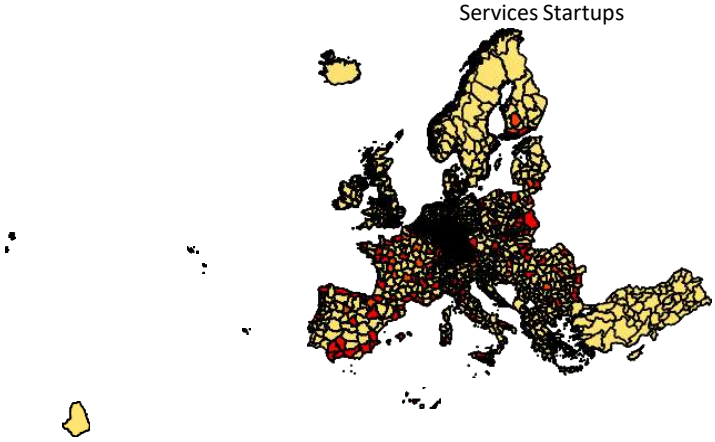
(381,12776.9]
[4.6,381]
No data

Note: Based on construction startups in EU metropolitan regions from 2008 to 2013 and at the level of NUTS3 2016.

Source: Eurostat, JRC and European Commission Directorate

Map 4

Services Startups



(2438.7,83841.4]
[56.9,2438.7]
No data

Note: Based on services startups in EU metropolitan regions from 2008 to 2013 and at the level of NUTS3 2016.

Source: Eurostat, JRC and European Commission Directorate.

Conclusions

To sum up, in the present thesis I seek to explore the determinants of human capital and the impacts of human capital on economic outcomes in geographically specific regions. I attempt to approach this complex field of human capital from a different perspective in each of the three chapters comprising this thesis. Forced migration, landownership, religion, culture, urbanization, climatic factors and geographic factors affects the development of human capital formation, while in turn human capital formation affects economic growth of specific regions. In this dissertation I examine in which way historical aspects of forced migration, landownership, urbanization, culture and geography determine the quantity and quality of human capital formation in a specific geographically defined area. In addition, I provide evidence on the causal relationship between human capital and economic development in terms of starting new businesses. Specifically, I do not only examine the effect of human capital formation on economic outcomes, but I also try to figure out several key determinants of human capital formation which in turn affect economic outcomes.

In this thesis, I first examine the historical determinants of human capital formation using Greek historical data on literacy, landownership, geography, fertility and demography at the provincial and municipal levels based on the years 1907 and 1928. I conclude that inequality in landownership has a substantial negative effect on human capital formation in Greek Regions. Moreover, geographic factors, fertility and demographic factors have significant impacts on literacy in Greek provinces and municipalities. Second, I explore the effect of the historical forced migration in

Greece after the Greco-Turkish war of 1919–1922, in the short run. My findings suggest that forced migration, as a historical determinant of human capital and employment, increase all native literacy rates but on the other hand, there is no significant impact on employment except for male industrial and agricultural employment. Last but not least, I conduct a systematic examination of how human capital and other major factors stimulate regional enterprise births in European metropolitan regions for the period 2008-2013. . I empirically demonstrate that total patent applications in combination with human capital factors (e.g. human resources in science and technology) lead to higher enterprise births in all sectors.

My research has, of course, a number of limitations because I have no information on a number of issues. As a result, further extensions can take many directions and enrich the analysis of human capital by accounting for more indicators that comprise different levels and forms of human capital. An interesting extension of chapter two is to examine if the historical results of human capital are valid not only in the short-run but also in the long-run. Moreover, a political analysis for political parties with an agenda in favor of human-capital could enhance the results for all parts of this thesis. Regarding the third part of this study other explanatory variables can enhance the results. An interesting indicator that might provide important insights about the results is the R&D expenditure. In addition, the extension of the examined period before 2008 could stimulate my findings.

References

Abadie, A., & Imbens, G. W., (2006). Large sample properties of matching estimators for average treatment effects. Econometrica 74, 235–267.

Adamopoulos, T., & Restuccia, D., (2018). Geography and agricultural productivity: cross-country evidence from micro plot-level data. NBER Working Paper 24532. Cambridge, MA 02138.

Adelino, M., Ma, S., & Robinson, D., (2017). Firm Age, Investment Opportunities, and Job Creation. The Journal of Finance, 72(3), 999-1038.

Aghion, P., & Jaravel, X., (2015). Knowledge Spillovers, Innovation and Growth. The Economic Journal, 125 (March), 533–573.

Alogoskoufis, G., (2021). Historical Cycles of the Economy of Modern Greece From 1821 to the Present. Working Paper 01-2021.

Arellano, M., & Bond, S., (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Review of Economic Studies, 58, 277–297.

Ashraf, Q., & Galor, O., (2013). The "Out of Africa" hypothesis, human genetic diversity, and comparative economic development. The American Economic Review, 103(1), 1–46.

Ashraf, Q., & Galor, O., (2018). The Macrogenoeconomics of Comparative Development. Journal of Economic Literature, 56(3), 1119-1155.

Ashraf, Q., Galor, O., & Klemp, M., (2020). The Evolutionary Origins of the Wealth of Nations. Bravo Working Paper 2020-022.

Avdela, E., (1990). Female public servants. Disaggregation of employment by gender in the public sector, 1908-1955. Foundation of Research and Education of the Commercial Bank of Greece, Athens (in Greek).

Axinn, W. G., (1993). The effects of children's schooling on fertility limitation. Population Studies 47(3), 481-493.

Bakhtiari, S., (2019). Entrepreneurship Dynamics in Australia: Lessons from Micro-data. Economic Record, 95(308), 114-140.

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

Baten, J., & Hippe, R., (2017). Geography, land inequality and regional numeracy in Europe in historical perspective. J Econ Growth, 1-31.

Bauer, T. K., Braun, S., & Kvasnicka, M., (2013). The Economic Integration of Forced Migrants: Evidence for Post-War Germany. The Economic Journal, 123(571), 998–1024.

Becker, G. S., (1962). Investment in Human Capital: A theoretical Analysis. Journal of Political economy, 70(5), 9-49.

Becker, S. O., & Woessmann, L., (2009). Was Weber Wrong? A human capital theory of protestant economic history. The Quarterly Journal of Economics, 124(2), 531-596.

Becker, S. O., & Ferrara, A., (2019). Consequences of forced migration: A survey of recent findings. Labour Economics, 59, 1-16.

Becker, S. O., Grosfeld, I., Grosjean, P., Voigtländer, N., & Zhuravskaya, E., (2020). Forced Migration and Human Capital: Evidence from Post-WWII Population Transfers. American Economic Review, 110(5), 1430–1463.

Behrens, K., Duranton, G., & Robert-Nicoud, F., (2014). Productive Cities: Sorting, Selection, and Agglomeration. Journal of Political Economy, 122(3), 507-553.

Belotti, F., Hughes, G., & Mortari, A. P., (2017). Spatial panel-data models using Stata. The Stata Journal (2017), 17(1), 139–180.

Beltrán Tapia, F. J., & Martínez-Galarraga, J., (2015). Land access inequality and education in pre-industrial Spain. University of Oxford discussion papers in Economic and Social History, 137.

Benos, N., Michalopoulos, S., Papaioannou, E., Karagiannis, S., Murard, E., & Sakalli, S. O., (2021). Refugees in the Mediterranean: Economic Consequences. Work in progress.

Bharadwaj, P., Khwaja, A. I., & Mian, A., (2015). Population Exchange and its Impact on Literacy, Occupation and Gender – Evidence from the Partition of India. International Migration, 53(4), 90-106.

Bharadwaj, P., & Mirza, R. A., (2019). Displacement and development: Long term impacts of population transfer in India. Explorations in Economic History, 73, 101273.

Bijnens, G., & Konings, J., (2018). Declining business dynamism in Belgium. Small Business Economics, 54, 1201-1239.

Botticini, M., & Eckstein, Z., (2007). From Farmers to Merchants, Conversions and Diaspora: Human Capital and Jewish History. Journal of the European Economic Association, 5(5), 885–926.

Boucekkine, R., de la Croix, D., & Peeters, D., (2007). Early literacy achievements, population density, and the transition to modern growth. Journal of the European Economic Association, 5, 183–226.

Braun, S., & Mahmoud, T. O., (2014). The Employment Effects of Immigration: Evidence from the Mass Arrival of German Expellees in Postwar Germany. The Journal of Economic History, 74(1), 69-108.

Braun, S., & Kvasnicka, M., (2014). Immigration and structural change: Evidence from post-war Germany. Journal of International Economics, 93(2), 253-269.

Brown, J. D., Earle, J. S., Kim, M. J., & Lee, K. M., (2019). Start-ups, job creation, and founder characteristics. Industrial and Corporate Change, 28(6), 1637-1672.

Brummund, P., & Connolly, L., (2019). Who Creates Stable Jobs? Evidence from Brazil. Oxford Bulletin of Economics and Statistics, 81(3), 540-563.

Buggle, J., C., & Durante, R., (2021). Climate Risk, Cooperation, and the Co-Evolution of Culture and Institutions. The Economic Journal, 637, 1947–1987.

Caicedo, FV., (2018). The Mission: Human Capital Transmission, Economic Persistence, and Culture in South America. The Quarterly Journal of Economics, 134, 507-556.

Calvino, F., Criscuolo, C., & Menon, C., (2018). A cross-country analysis of start-up employment dynamics. Industrial and Corporate Change, 27(4), 677–698.

Cameron, A. C., & Trivedi, P. K., (2005). Microeconometrics: Methods and Applications. New York, Cambridge University Press.

Cattaneo, M. D., (2010). Efficient semiparametric estimation of multi-valued treatment effects under ignorability. Journal of Econometrics 155, 138–154.

Chatterji, A., Glaeser, E., & Kerr, W., (2014). Clusters of Entrepreneurship and Innovation. NBER Working Paper 19013, Published in Innovation Policy and the Economy, 14.

Chatziiosif, C., (1993). Old moon. Industry in Greek Economy, 1830-1940. Themelio editions (in Greek).

Chrysochoou, I., (1981). Uprooted generation. The chronicle of the immigration in Thessaloniki. Filippotis editions, 2nd edition (in Greek), Athens, 51-52.

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

Clark, G., & Gray, R., (2014). Geography is not destiny: geography, institutions and literacy in England, 1837–1863. Oxford Economic Papers, 66(4), 1042–1069.

Clementi, G. L., & Palazzo, B., (2016). Entry, Exit, Firm Dynamics, and Aggregate Fluctuations. American Economic Journal, Macroeconomics, 8(3), 1-41.

Corrado, L., & Fingleton, B., (2016). The W Matrix in Network and Spatial Econometrics: Issues Relating to Specification and Estimation. CEIS Tor Vergata, 14(3), 369, 2-27.

Cygan-Rehm, K., & Maeder, M., (2012). The Effect of education on fertility: Evidence from a compulsory schooling reform. SOEP papers on Multidisciplinary Panel Data Research 528, University of Erlangen-Nuremberg, 1-49.

Daunfeldt, S. O., Elert, N., & Johansson, D., (2016). Are high-growth firms overrepresented in high-tech industries? Industrial and Corporate Change, 25(1), 1-21.

Deininger, K., & Squire, D., (1998). New ways of looking at old issues: Inequality and growth. Journal of Development Economics, 57, 259–287.

Decker, R. A., Haltiwanger, J., Jarmin, R. S., & Miranda, J., (2017). Declining Dynamism, Allocative Efficiency, and the Productivity Slowdown. American Economic Review: Papers & Proceedings, 107(5), 322-326.

Decker, R. A., Haltiwanger, J., Jarmin, R. S., & Miranda, J., (2016). Declining Business Dynamism: What We Know and the Way Forward. American Economic Review: Papers & Proceedings, 106(5), 203-207.

Decker, R. A., Haltiwanger, J., Jarmin, R. S., & Miranda, J., (2016). Where has all the skewness gone? The decline in high-growth (young) firms in the U.S. European Economic Review, 86, 4-23.

Decker, R., Haltiwanger, J., Jarmin, R., & Miranda, J., (2014). The Role of Entrepreneurship in US Job Creation and Economic Dynamism. Journal of Economic Perspectives, 28(3), 3-24.

Dertilis, G., (1984). Greek Economy (1830-1910) and Industrial Revolution. Sakkoula Editions, Athens (in Greek).

Djankov, S., Ganser, T., McLiesh, C., Ramalho, R., & Schleifer, A., (2010). The effect of corporate taxes on investment and entrepreneurship. American Economic Journal: Macroeconomics, 2(3), 31-64.

Drivas, K., Economidou, C., Karkalakos, S., & Tsionas, E. G., (2016). Mobility of knowledge and local innovation activity. European Economic Review, 85, 39-61.

Duranton, G., & Puga, D., (2020). The Economics of Urban Density. Journal of Economic Perspectives, 34(3), 3–26.

Easterly, W., (2007). Inequality does cause underdevelopment: Insights from a new instrument. Journal of Development Economics, 84, 755–776.

Eddie, S. M., (1967). The changing pattern of landownership in Hungary, 1867–1914. The Economic History Review, 20, 293–310.

Efthyvoulou, G., Kammas, P. & Sarantides, V., (2020). Gender voting gap in the dawn of urbanization: evidence from a quasi-experiment with Greek special elections. GreeSE Papers, 146.

Elhorst, J. P., Gross, M., & Tereanu, E., (2018). Spillovers in space and time: where spatial econometrics and Global VAR models meet. ECB Working Paper Series, 2134, 1-37.

Erickson, L., & Vollrath, D., (2004). Dimensions of land inequality and economic development. In IMF working paper WP/04/158. Washington, D.C.

Eriksson, R., & Rataj, M., (2019). The geography of starts-ups in Sweden. The role of human capital, social capital and agglomeration. Entrepreneurship & Regional Development, 31(9-10), 735-754.

Evangelidis, T., (1936). Education under Ottoman Occupation, Greek Schools from Constantinople Conquest until Kapodistrias. Halkiopoulos Publications, v. 1-2.

Faggio, G., & Silva, O., (2014). Self-employment and entrepreneurship in urban and rural labour markets. Journal of Urban Economics, 84, 67–85.

Farre-Mensa, J., Hegde, D., & Ljungqvist, A., (2020). What Is a Patent Worth? Evidence from the U.S. Patent "Lottery". The Journal of Finance, 75(2), 639-682.

Fritsch, M., & Wyrwich, M., (2018). Regional knowledge, entrepreneurial culture, and innovative start-ups over time and space—an empirical investigation. Small Business Economics, 51, 337–353.

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

Giannakopoulos, G., (2003). Greece with the refugees, The History of New Hellenism. Panayotopoulos (Ed.), Chapter 7, Ellinika Grammata Publishers.

Gibbons, S., Overman, H. G., & Patacchini, E., (2014). Spatial Methods. SERC discussion paper, 162, 1-68.

Goñi, M., (2018). Landed Elites and education provision in England: Evidence from school boards, 1870-99. Working Paper, University of Vienna.

Gounaris, B. C., (1989). Emigration from Macedonia in the Early Twentieth Century. Journal of Modern Greek Studies, 7(1), 133-153.

Gourio, F., Messer, T., & Siemer, M., (2016). Firm Entry and Macroeconomics Dynamics: A State-Level Analysis. American Economic Review, 106(5), 214-218.

Güneş, P. M., (2013). The impact of female education on fertility: Evidence from Turkey. Working Paper Series, University of Pennsylvania Scholarly Commons.

Haltiwanger, J., Jarmin, R. S., & Miranda, J., (2013). Who Creates Jobs? Small versus Large versus Young. The Review of Economics and Statistics, 95(2), 347-361.

Hansen, L. P., (1982). Large sample properties of generalized method of moments estimators. Econometrica, 50, 1029-1054.

Heyman, F., Norbäck, P. J., Persson, L., & Andersson, F., (2019). Has the Swedish business sector become more entrepreneurial than the US business sector? Research Policy, 48(7), 1809-1822.

Heyman, F., Norbäck, P. J., & Persson, L., (2018). Who creates jobs and who creates productivity? Small versus large versus young versus old. Economics Letters, 164, 50-57.

Hippe, R., (2012a). How to measure human capital? The relationship between numeracy and literacy. Economies et Sociétés, 45, 1527–1554.

Hippe, R., (2012b). Spatial clustering of human capital the European regions. Economies et Sociétés, 46, 1077–1104.

Hoem, J.M., Neyer, G., & Andersson, G., (2006). Educational attainment and ultimate fertility among Swedish women born in 1955-59. Demographic Research, Vol14/16, 381-404.

Hoover, E. M., & Giarratani, F., (2020). An Introduction to Regional Economics. The Web Book of Regional Science, The Research Repository @ WVU, 1-304.

Hopenhayn, H. A., (1992). Entry, Exit and Firm Dynamics in Long Run Equilibrium. Econometrica, 60(5), 1127-1150.

Hornung, E., (2014). Immigration and the Diffusion of Technology: The Huguenot Diaspora in Prussia. American Economic Review, 104(1), 84–122.

Huber, P., Oberhofer, H., & Pfaffermayr, M., (2013). Job creation and the intra-distribution dynamics of the firm size distribution. Industrial and Corporate Change, 23(1), 171-197.

Iordanidou, M., (1990). Like crazy birds. 15th edition (in Greek), Athens, 97-98.

Kalafati, E., (1988). School buildings of primary education, 1821-1929. From specification to planning. Historical Archive of Greek Youth, General Secretariat of Youth, Athens.

Kammas, P., & Sarantides, V., (2020). Democratisation and tax structure in the presence of home production: Evidence from the Kingdom of Greece. Sheffield Economic Research Paper Series, SERPS Papers, 2019010.

Kammas, P., Sakalis, A., & Sarantides, V., (2021). Pudding, plague and education: Trade and human capital formation in an agrarian economy. GreeSE Papers, 164.

Karahan, F., Pugsley, B. & Şahin, A., (2019). Demographic Origins of the Startup Deficit. FRBN Staff Reports 888.

Klapsis, A., (2014). Violent Uprooting and Forced Migration: A Demographic Analysis of the Greek Populations of AsiaMinor, Pontus and Eastern Thrace. Middle Eastern Studies, 50(4), 622-639.

Kontogiorgi, E., (2006). Population Exchange in Greek Macedonia: The Rural Settlement of Refugees 1922–1930. Oxford University Press Inc., New York 1-395

Kostis, K., & Petmezas, S., (2006). 'I anaptyxi tis Ellinikis oikonomias ton 19o aiwna' (The growth of Greek economy in 19th century). Ekdoseis Alexandreia, Athens, Greece.

Kreibaum, M., (2016). Their Suffering, Our Burden? How Congolese Refugees Affect the Ugandan Population. World Development, 78, 262–287.

Krimpas, V., (1999). The Work of Agricultural Settlement of the Reffugee Resettlement Commission. Bulletin of Minor Asia Studies, 13, 225-268. doi:http://dx.doi.org/10.12681/deltiokms.148

Labanca, C., (2020). The effects of a temporary migration shock: Evidence from the Arab Spring migration through Italy. Labour Economics, 67, 101903.

Lam, D., & Duryea, S., (1999). Effects of schooling on fertility, labor supply, and investments in children, with evidence from Brazil. The Journal of Human Resources, XXXIV 1, 1-34.

Lane, P. W., & Nelder, J. A., (1982). Analysis of covariance and standardization as instances of prediction. Biometrics 38, 613–621.

Lawless, M., (2014). Age or size? Contributions to job creation. Small Business Economics, 42, 815-830.

Lee, L. F., & Yu, J., (2010a). Estimation of spatial autoregressive panel data models with fixed effects. Journal of Econometrics, 154, 165–185.

León, A., (2004). The Effect of education on fertility: Evidence from compulsory schooling laws. Working Paper, University of Pittsburgh.

Li, M., Goetz, S. J., Partridge, M., & Fleming, D. A., (2016). Location determinants of high-growth firms. Entrepreneurship & Regional Development, 28(1-2), 97-125.

Liang, J., Wang, H., & Lazear, E. P., (2018). Demographics and Entrepreneurship. Journal of Political Economy, 126(1), 140-196.

Louri, H., & Pepelasis Minoglou, I., (2002). A Hesitant Evolution: Industrialization and De-industrialization in Greece over the Long Run. Journal of European Economic History, 31(2), 321-348.

MCGrew, W., (1987). Land and Revolution in Modern Greece, 1800–1881: The Transition in the Tenure and Exploitation of Land from Ottoman Rule to Independence. The American Historical Review, Volume 92, 3, 707–708.

Michalopoulos, S., & Papaioannou, E., (2017). The Long Economic and Political Shadow of History. CEPR E-book, Vol. 1-3.

Motoyama, Y., & Malizia, E., (2017). Demand pull or supply push? Metro-level analysis of start-ups in the United States. Regional Studies, Regional Science, 4(1), 232-246.

Mouzelis, N., (1978). Class and Clientelistic Politics: The Case of Greece. The Sociological Review, 26(3), 471-497.

Murard, E., & Sakalli, S. O., (2019). Mass Refugee Inflow and Long-run Prosperity: Lessons from the Greek Population Resettlement. CReAM Discussion Paper Series, Centre for Research and Analysis of Migration (CReAM).

Nunn, N., (2009). The Importance of History for Economic Development. Annual Review of Economics, 1.1, 65-92.

OECD, (2017). The Geography of Firm Dynamics: Measuring Business Demography for Regional Development. OECD Publishing, Paris, 1-154.

Özyurt, S., & Dees, S., (2015). Regional dynamics of economic performance in the EU: To what extent spatial spillovers matter? ECB Working Paper, 1870, 1-39.

Pallis, A. A., (1928). Les Effets de la Guerre sur la population de la Grèce. In A. Andreades "Les Effets economiques et sociaux de la Guerre en Grèce". Paris and New Haven, 133–35.

Pentzopoulos, D., (1962). The Balkan Exchange of Minorities and its Impact upon Greece, Hurst and Company, London.

Perotti, R., (1996). Growth, income distribution, and democracy: What the data say. Journal of Economic Growth, 1, 149–187.

Petmezas, S., (2003). 'I Elliniki agrotiki oikonomia kata ton 19o aiwna' (The Greek agricultural economy in the 19th century). Panepistimiakes Ekdoseis Kritis, Crete, Greece.

Petmezas, S., (2012). 'Prolegomena stin istoria tis Ellinikis agrotikis oikonomias tou Mesopolemou' (Introduction in the history of Greek agricultural economy in interwar period). Ekdoseis Alexandreia, Athens, Greece.

Psomiades, H. J., (1968). The Eastern Question—The Last Phase: A Study in Greek-Turkish Diplomacy. Institute of Balkan studies, Thessaloniki.

Pugsley, B. W., & Sahin, A., (2019). Grown-up Business Cycles. The Review of Financial Studies, 32(3), 1102-1147.

Riginos, M., (1987). Productive Structures and Wages in Greece, 1909-1936, Industry-Craftmanship. Modern Greek History Studies, Institute of Research and Education of the Commercial Bank of Greece, (in Greek).

Roach, T., (2009). The effect of fertility levels on the educational attainment of children in Uganda. Working Paper, University of California.

Rodríguez-Pose, A., & Crescenzi, R., (2008). Research and development, spillovers, innovation systems, and the genesis of regional growth in Europe. Regional studies, 42(1), 51-67.

Roodman, D., (2009). How to do xtabond2: An introduction to difference and system GMM in Stata. The Stata Journal 9, 1, 86–136.

Rubin, D. B., (1973). Matching to remove bias in observational studies. Biometrics 29, 159–183.

Ruiz, I., & Vargas-Silva, C., (2016). The labour market consequences of hosting refugees. Journal of Economic Geography, 16(3), 667–694.

Santoleri, P., (2020). Innovation and job creation in (high-growth) new firms. Industrial and Corporate Change, 29(3), 731–756.

Sokoloff, K. L., & Engerman, S., (2000). Institutions, Factor Endowments, and Paths of Development in the New World. Journal of Economic Perspectives, 14, 217-232.

Stata Treatment Effects Reference Manual: Potential outcomes/counterfactual outcomes, Release 14. Stata Press Publication, College Station, Texas.

Stefanidis, I. D., (2006). Reconstructing Greece as a European State: Venizelos' Last Premiership 1928-32. Eleftherios Venizelos: The Trials of Statesmanship, Edinburgh University Press, 193-233.

Sung, N., & Kim, J., (2020). Entry and exit of small self-employed businesses in Korea's service industries. Small Business Economics, 54, 303–322.

Tumen, S., (2016). The Economic Impact of Syrian Refugees on Host Countries: Quasi-Experimental Evidence from Turkey. American Economic Review: Papers & Proceedings, 106(5), 456–460.

Vergopoulos, K., (1975). 'To agrotiko zitima stin Ellada- to provlima tis koinwnikis enswmatwsis tis georgias' (The agricultural issue in Greece -the problem of social integragion of agriculture). Exantas 1975, Athens, Greece.

Vittinghoff, E., Glidden, D. V., Shiboski, S. C., & McCulloch, C. E., (2012). Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models. 2nd edition New York: Springer.

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

Wooldridge, J. M., (2002). Inverse probability weighted M-estimators for sample selection, attrition, and stratification. Portuguese Economic Journal 1, 117–139.

Wyrwich, M., (2020). Migration restrictions and long-term regional development: evidence from large-scale expulsions of Germans after World War II. Journal of Economic Geography, 20(2), 481–507.

Zábojník, J., (2020). Firm Reputation, Innovation and Employee Startups. The Economic Journal, 130(627), 822–851.

Zolotas, X., (1926). Greece at the stage of industrialization. Athens, Elefteroudakis, 157. Translation of the PhD Dissertation: Griechenland auf dem Wege zur Industrialisierung, Leipzig: B. G. Teubner, 144.