SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Three Essays on Technological Progress, Natural Resources and Economic Growth

Department of Economics and Statistics

PhD Program in Economics   Cycle XXXII

Curriculum in DEFAP

A DISSERTATION PRESENTED
BY
MICHELE MAURIZIO MALPEDE

SUPERVISOR: LUCIA CORNO
CO-SUPERVISOR: ROSARIO CRINO'

PHD COORDINATOR: MATTEO MANERA

ACADEMIC YEAR: 2019

Thesis advisor: Professor Lucia Corno Michele Maurizio Malpede

Three Essays on Technological Progress, Natural Resources and Economic Growth

Abstract

This dissertation examines how medical progresses, allowing for the cure of particularly diseases, and technological progress, such as the discovery of modern electric batteries, affect socio-economic development in the USA and in Sub-Saharan Africa. The dissertation also combines agricultural progresses, such as the discovery of a new staple crop with low health conditions, such as malaria prevalence to assess how much of the progress might be absorbed by poor environmental and health conditions.

The first chapter of the dissertation examines technological progresses, such as the creation of modern lithium-ion electrical batteries and their implications for socio-economic development. I examine the cobalt mining boom in the Democratic Republic of Congo, which occurred in 2007, and was caused by the advent of modern smartphones, PCs, tablets, and electric vehicles had on child labor, through reduced education attainment and subsequently on parental fertility choices. This is achieved first, by combining geo-referenced data on education attainment and wealth conditions of individuals across different villages or towns with the location of all cobalt mine deposits in the DRC in a differences-in-differences strategy. The procedure compares education attainment later in life of those individuals who, during their childhood, lived within 10 kilometers away from a cobalt mine deposit before and after the cobalt mining boom. Moreover, the first chapter of the thesis shows that the increase in child labor lowers the opportunity cost for parents of having an additional child, thus resulting in a higher fertility rate. Effects of cobalt mining on wealth conditions, later in life are also quantified.

The second chapter of the dissertation examines progresses in medical field, such as the eradication of malaria and related vector-borne diseases and their implications for historical economic development. I examine the eradication of malaria which happened in the US during early 1900s. This is achieved by comparing agricultural productivity levels of highly malarious counties with those of less malarious counties before and after the eradication of malaria in the US which was achieved as a result of the understanding that malaria was transmitted by the bite of specific species of mosquitoes and in turn of newly discovered drugs, such as quinine first and chemical components such as the DDT later on. Using a difference-in-difference (DID) estimation comparing historical agricultural productivity levels between US counties that had climatic conditions more suitable for the transmission of malaria with counties that were less suitable for a stable transmission of the disease, I present causal evidence on the effects of higher malaria prevalence on agricultural productivity and economic development.

Finally, the third chapter of this dissertation builds on the seminal paper of Nunn and Qian (2011) to examine if the positive impacts of the discovery of a new staple crop in the Old World, such as potato, on population and urbanization were partially absorbed by exogenous poor health conditions. The exogenous variations of weather conditions for the transmission of malaria allow for the comparison between potato suitable areas which had more or less prevalence of malaria. We employ two different estimation strategies. The first estimation entirely relies on that adopted in Nunn and Qian (2011). The second estimation procedure allows us to compare population and urbanization levels at a 0.5° latitude by 0.5° longitude level. We find that the presence of weather conditions suitable for a stable transmission of malaria counteracted the significant benefits on population and urbanization observed during the eighteenth and nineteenth centuries due to the introduction of potato.

Contents

List of Figures

List of Tables

Acknowledgments

First, I am very grateful for the abundance of guidance, inspiration and helpful comments from my two supervisors: Lucia Corno and Rosario Crino'. None of these three chapters would have been possible without your support, insights and patience. Both of you have greatly shaped the way I think about economic development and you are true role models, in work and in life.

I am also thankful to my two discussant of the first two chapters of the present thesis: Mara Squicciarini and Massimiliano Onorato for their helpful suggestions and advice. I thank the numerous participants of the European Environmental and Resource Economic Association. I am also thankful to seminar participants of the EIEE.

I especially thank those who overlapped with me while I was Junior Researcher at FEEM, Valentina Bosetti, Massimo Tavoni, Elena Verdolini, Johannes Emmerling and Cristina Cattaneo for their guidance on my very first (and uncertain) steps on economic research. Cate, Susy, Soheil, Jack, Simo, Teo, Lara, Laurent for always reaching out, grabbing dinner and drinks, and always helping me keep things in perspective. Soheil, once again, for our amazing biking trips (and many others to come).

I also, want to thank all Professors who I met while at Bocconi University, and shaped greatly my research interests and knowledge. This is where I initially decided to embrace my interest in economics.

Alla mia intera famiglia a Lavello. Grazie mille per il vostro supporto e amore. Il mio dottorato non sarebbe stato possibile senza il vostro appoggio. Un enorme abbraccio.

A mio fratello gemello, Alfonso: il nostro percorso e' appena iniziato.

A mia madre, Antonietta: grazie per tutto il tuo amore e a mio padre Francesco: un grazie immenso perchè ci sei sempre.

A Claudia. Per la tua presenza. For how important and essential you are for me. And for teaching me to love, dogs!

Thank you. For everything.

0
Introduction

This dissertation focuses on understanding the possible positive and negative implications of the technological progress both on the historical economic development of the US and the comparative development of Sub-saharan African region. The principal question that motivates this research and links the three essays presented is understanding the true effects of all kinds of technological progress on different socio-economic contexts. The dissertation approaches this question from the lens of development economics and environmental and resources economics by focusing on how different technological shocks, such as the adoption of modern and more efficient lithium-ion electric batteries, or breakthrough discoveries in medical field that allow for a cure of a particularly deadly disease, affect comparative development. Moreover, this dissertation assess the interaction of a technological shock, such as the discovery of a new staple crop and poor health conditions, provided by climatic conditions particularly suitable for the transmission of malaria, to understand if poor health conditions absorb part of the positive effects provided by the positive technological shocks. The research papers in this dissertation use varying insights from development economics, health economics and environmental and resource economics to better understand historical and comparative economic development and its determinants.

The first chapter in my dissertation focuses on the consequences of technological innovation on the socio economic development of regions rich of crucial minerals. I use the boom in cobalt mining which occurred in the Democratic Republic of Congo (henceforth the DRC) from 2007 to 2010 and was caused by the advent of modern lithium-ion electric batteries to assess the impact of the electrification boom on child labor through the reduced education attainment of those individuals who during the cobalt boom grew up in cobalt mining areas. Consequently, since child labor is a family decision, the present study shows if the boom in production of cobalt from mining has resulted in a decrease in the opportunity cost of having a new child for families living close to a cobalt deposit.

To examine the long-run impacts of the cobalt mining boom on local communities, I combine geo-referenced data on education attainment and wealth conditions of individuals with the exogenous location of cobalt mine deposits in the DRC in a differences-in-differences set up. Individuals who were born since 1993 (i.e. were aged between 6 and 14 at the time of cobalt mining boom) and during their childhood lived "close enough"* to a cobalt deposit constitute the treatment group. On the other hand, those individuals living in areas beyond 10 kilometers from a cobalt mine deposit

*The definition of close is within 10 kilometers from an active deposit of cobalt serve as a first comparison group.

The estimates obtained suggest that after the cobalt mining boom occurred in 2007 a considerable number of children was sent out from school, plausibly to work as cobalt miners. Completed years of education of individuals who grew up within 10 kilometres from the nearest cobalt mine deposit decreased by almost 0.4 years of baseline within seven years from the start of the cobalt mining boom. A reduction of 0.4 completed years of education constitutes a large a significant impact given that the average number of completed years of education in the DRC is four. Therefore, the boom of cobalt mine has caused a 10% reduction in education attainment of individuals who at the time of the boom where living in areas surrounding cobalt deposits. The analysis shows that the effects are concentrated within 10 kilometres from a cobalt mine deposit. These estimates support the well documented reports of the International Labor Organization along with Amnesty International and UNICEF which argue that around 40,000 children aged between 6 and 14 are currently sent to work in cobalt mines (Amnesty International, 2017).

The second chapter in my dissertation focuses on the consequences of climatic conditions suitable for the transmission and prevalence of malaria on the agricultural productivity. I use the eradication of malaria which occurred in the USA from early 1900s to 1920s as a natural experiment to understand the impact that poor health conditions, affecting mostly farmers living in rural areas, on the agricultural productivity.

There are two principal mechanisms through which the transmission of malaria might be tackled. One mechanism is land use conversion. In the U.S. drainage of swamps and wetland is one of the oldest and commonest forms of land modification undertaken to improve health conditions and lower the transmission of vector borne diseases such as yellow fever and malaria. For these reasons surface water removal was a predominant public policy objective in United States during the 20^{th} century. This conversion of unused wetland into arable and more productive land might have increased the agricultural output of endemic areas and in turn agricultural productivity. I test this hypothesis. A second channels through which the transmission of malaria might impact on agriculture productivity is linked to the poor health conditions of farmers. The whole harvest indeed will be negatively affected if a farmer caches malaria during the harvesting. It is therefore reasonable to conclude that more endemic lands were likely to be less productive than non endemic ones. The adopted strategy allows to show whether this is the case. Hence, the primary goal of this study is to clearly assess the impact that eradication of malaria might have had on agriculture productivity in terms of possible greater amount of arable land and in terms of increased labour productivity due to better health conditions of workers, in particular farmers.

I use a spatial time-invariant malaria ecology index created by Kiszewski et al. (2004) and based upon climatic conditions which are more or less suitable for the reproduction of two particular species of mosquitoes, namely: Plasmodium falciparum and Plasmodium vivax.

Results of the second chapter reveal that the eradication campaigns which in the U.S. took place between the 1900s and the 1940s (with the administration of quinine and drainage of wetlands first and the development of new effective drugs and chemical components later) are estimated to have had positive and significant effects on the historical agricultural productivity growth in the US. In particular, a 0.1 increase in the Malaria Stability Index (MSI) explains the 20% of the differential in agricultural productivity between more or less malarious counties.

Two are the main contributions of this study: The first is that historical agricultural productivity growth is crucial to understand the causes of historical economic development both between countries and within nations (Gollin et al., 2014, Bustos et al., 2016). Therefore, assessing the historical effects of malaria on agricultural productivity within the US is crucial to understand differences in the economic development of different US areas.

The second contribution of this study id that estimating the relationship between malaria eradication and agricultural productivity is also crucial to predict how economies which nowadays rely predominantly on agriculture (i.e. Sub-Saharan African, Latin American and South-East Asian countries) will be affected from government policies aimed at eradicating vector borne diseases. In addition to this, future climatic conditions will likely be more favorable to the transmission of malaria (Caminade et al., 2014, Medlock and Leach, 2015).

Finally, the third chapter of this dissertation focuses on the interactions between technological progress and poor health conditions. By using the introduction of the cultivation of potato in the Old World, which occurred in 1700, I interact time invariant climate and soil conditions which make a land more or less suitable to the cultivation of potato, with time invariant weather conditions which are more favourable to the reproduction of mosquitoes larvae and in turn to stable prevalence of malaria to quantify if the positive impact of the cultivation of potato on population and urbanization as shown in the seminal paper by Nunn and Qian (2011) was partly absorbed by greater prevalence of malaria.

Our results at the grid level first confirm the conclusion obtained with the country-level analysis in Nunn and Qian (2011) that is that the diffusion of potatoes to the Old World has positively impacted population and urbanization of grids which were more suitable for growing the staple crop. Specifically, grid-level results show that the grids highly suitable for potato experienced a 3.2% increase in urbanization compared to those which were not suitable for growing potato. Nevertheless, the positive impact of growing potatoes decreases as the endemicity of malaria increases. For instance, the grids that were suitable for growing potatoes but mildly malaria endemic experienced an increase in urbanization levels by only 1.4%. As the endemicity of malaria increases the impact of potato suitability decreases until it becomes null, suggesting that the diffusion of malaria fully offset the benefits of cultivating potato.

1

The Dark Side of Batteries: Education, Fertility and Cobalt Mining in the DRC

Being rich of crucial minerals might be a leading determinant of local economic development. On the other hand, intensive extraction might prove to negatively affect the health and wealth of people living close to mineral deposits and the surrounding environment. I examine the effects that local cobalt mining had on child labor and subsequently on fertility rates in the DRC by exploiting geographic variation of cobalt deposits prior to the boom of modern electric batteries and using both education attainment and fertility data. I find that the boom in cobalt mining led to a reduction in educational achievements of children who were between 6 and 14 at the time of the boom, and it was accompanied with higher fertility rates. Moreover, the analysis shows that treated people do not compare worse both in terms of wealth and health compared to those in the control groups once they become adult. The results are robust to spatial spillover effects and selective migration.

Child labor and its associated reduction in education attainment are crucial problems for the socio-economic development of countries*. This paper focuses on the effects that the boom of Cobalt mining had on completed education attainment and fertility rates by exploiting the effects of plausibly exogenous cobalt boom occurred in the Democratic Republic of Congo in 2007 generated by the increase in modern lithium-ion batteries contained in high-tech devices such as smartphones, PCs, wireless headphones and electric vehicles. The DRC is the country where 65% of the total world deposit of cobalt is contained US Geological Survey (2019). Figure 1.3 shows the countries where cobalt is mined. In particular, the objective of the present study is to provide a rigorous quantitative assessment to the question of whether the boom of cobalt mining production caused children who were between 6 and 14 years of age and live close to cobalt mine deposits to drop out from school and achieve a lower level of education to go to work as cobalt miners. Consequently, since child labor is a family decision, the present study shows if the boom in production of cobalt from mining has resulted in a decrease in the opportunity cost of having a new child for families living close to a cobalt deposit. As a result, fertility rate in areas around active cobalt deposits might have

*Child labor is reported to be a major determinant of low development of Sub-Saharan African countries (International Labor Organization, 2015). Child labor subtracts children from school, thus negatively affecting human capital (Hazan and Berdugo, 2002) increased compared to areas far from cobalt mine since 2007. Finally, the paper compares the wealth of those children who were between 6 and 14 and lived in villages surrounding a cobalt deposit in 2007 in their young adulthood compared to those children who at the time of the cobalt boom were living far from a cobalt mine deposit.

The boom of cobalt mining has occurred very fast and came overwhelmingly from outside the DRC. As Figure 1.5 shows, the production of cobalt has been always constant at around 20 Mt per year, with the DRC contributing for almost 70% of total production of cobalt. However, starting from 2007 until 2010 the production of cobalt tripled to around 60 Mt per year to keep those levels until 2017. Production of cobalt in the DRC has followed the same path of the global demand of the mineral. Figure 1.5 shows that since 2007 its mining sector has seen a steady increase in industrial cobalt production. A cobalt production peak was reached in 2011 when the DRC mine production amounted to more then 60 Mt (Figure 1.5). Since 2011, cobalt production has stabilized to then same levels of 2010[†] Both the natural presence of cobalt in the DRC and the boom originated in 2007 outside the DRC, created a shock which is independent of local education and fertility choices.

To examine the impacts of the cobalt mining boom, I combine two source of variation: i. geographic variations in the exposure to cobalt mining activities in the DRC, ii. time variation in the production of cobalt induced by the world-wide adoption of lithium-ion batteries, and iii. age specific exposure to cobalt deposits. The geographical variation comes from the presence of cobalt deposits in the DRC. Sub Saharan Africa is naturally abundant of crucial minerals and the presence of

[†]Data accuracy of cobalt mining statistics of the DRC is an important issue. There exist different sources on cobalt production from mining in the DRC each showing slightly different numbers. For instance, the DRC Chamber of Mines (Chambre des Mines 2015) reported a cobalt production of 69,328 tons for 2015, whereas CRU (2016) indicated a production of 66,120 tons. US Geological Survey (2019) instead reported a total cobalt production of 63,000 tons. However, all of those figures do not consider artisanal produced cobalt which is considered to account for 15-20% of the total production of cobalt in the DRC. CRU (2016) estimated that artisanal cobalt production in 2015 amounted to 10,500 tons. For this reason, the DRC Ministry of Mines reported a cobalt production of 84,400 tons considering also artisanal based production.

cobalt is plausibly exogenous with respect to the education achievements of individuals and fertility of women in the DRC[‡]. The boom of cobalt mining occurred in the DRC in 2007 constitutes the source of time variation. The assumption of this study is that individuals who were aged between 6 and 14 at the time of cobalt mining boom[§] and live in cobalt mining villages or towns (within 10 kilometers from a cobalt deposit) should be more affected by the boom in cobalt mining. These individuals constitute the treatment group. On the other hand, those individuals living in areas beyond 10 kilometers from a cobalt mine deposit serve as a first comparison group. In addition, since the production of other minerals such as zinc, copper, diamonds, silver was not affected by the discovery of modern lithium-ion batteries in 2007, I can rule out any confounding factors related to other types of mine.

The empirical strategy therefore compares birth-year cohorts based on the proximity to a cobalt mine deposit in their current place of residence, interacted with a post 2007 indicator variable. This strategy identifies an intention-to-treat effect under the assumption that trends in outcomes would have been similar in areas close to a cobalt mine deposit in the absence of the sudden boom of cobalt production from mining in the DRC. I provide evidence supporting the plausibility of this assumption. Similarly, the procedure compares different measures of women's fertility rates during their fertile period (i.e. aged between 15 and 39 years of age) who lived within 10 kilometers from a cobalt mine deposit before and after the cobalt mining boom. Analysis on the fertility rate is done to verify if the increase in child labor lowers the opportunity cost for parents of having an additional child, thus resulting in a higher fertility rate[¶]. Data sources include geocoded cobalt deposit locations, place of residence during childhood of individuals surveyed, their education attainment and wealth

[‡]Some parts of the country are naturally "more suitable" for the extraction of the critical mineral, while others are not. Notably the eastern part of the DRC is rich of gold, diamonds, zinc and silver while the western side of the country is rich of petroleum and cement. See Figure ?? for a visual representation of the location of mineral deposits in the Democratic Republic of Congo.

[§]The ILO defines child labor as whoever between 6 and 14 years of age is working

[¶]This result would confirm the theoretical literature which focused on the relationship between child labor and fertility (Doepke and Zilibotti, 2005, Hazan and Berdugo, 2002).

index. Moreover, I use data on the number of births and the associated year for each women in the DRC. These data allow me to define those individuals that during their childhood were exposed to the surge in cobalt production from mining across areas more or less close to a cobalt mine deposit.

I first show that education achievements of children between 6 and 14 are negatively affected by cobalt mining activities during the period 2007-2014. In particular, I show that children between 6 and 14 living in cobalt mining areas, after the cobalt boom occurred in 2007, achieve 0.5 year of education less compared to their peers not exposed to cobalt mining areas. Short term effects of cobalt mining on education demonstrate that the largest effects on education appeared for children between 6 and 14 and thus enrolled in primary education. No effect was shown for secondary-school age children (14-18 years old). Furthermore, I find a positive effect of cobalt mining on wealth of treated children. The relative increase in earning, thus induces parents to opt for sending their children to the mines rather than going to school.

Secondly, I show that education attainment of individuals, who grew up in cobalt-mining areas was also negatively affected by the exposition to cobalt mining activities. For example, completed years of education of those individuals who grew up within 10 kilometres from the nearest cobalt mine deposit decreased by almost 0.5 years of baseline within seven years from the start of the cobalt mining boom. A reduction of 0.4 completed years of education constitutes a large a significant impact given that the average number of completed years of education in the DRC is four. Therefore, the boom of cobalt mine has caused a 10% reduction in education attainment of individuals who at the time of the boom where living in areas surrounding cobalt deposits. This study, also addresses the possible effects that a reduction in education attainment caused by the boom in cobalt production might have had on the wealth conditions of those individuals, later in life. I find that the initial increase in wealth shown for children living in cobalt mining areas disappear after seven years.

The paper also examines the potential changes that the child labor choice might have had on fertility choices of women. I consider different measures of fertility rates of women between 15 and 39 years of age obtained from the Demographic Health Surveys data, which were also used to find the effects of cobalt mining on education. I find that after the cobalt mining boom occurred in 2007, women living in cobalt mining villages had 0.3 more children during the five years preceding the date of the interview. This translates in an increase in 0.06 children per year per woman. I show that the increase in fertility rates in cobalt mining areas are due to the initial wealth gains due to children working. Parents are therefore more prone to send their children to work as cobalt miners. Moreover, these results are robust to alternative specifications addressing a number of potential concerns.

I report the aforementioned set of results suggesting that the loss in education attainment accompanied by the increase in fertility rates in cobalt mining areas of the Democratic Republic of Congo is mainly driven by the initial wealth gains due to the child labor associated to cobalt mining. Parents are generally reluctant to send their children to work as miners since working underground to search for commodities such as gold and diamonds is dangerous for children. However, what characterizes cobalt mining from other types of mines in the DRC is that cobalt is largely mined on the surface rather than underground and it is found in the dust. Thus, in a cobalt mine children typically wash the tiny cobalt matters from the dust. To perform this relatively not dangerous job small hands are needed (Amnesty International, 2017, International Labor Organization, 2015). Furthermore, although farming and family business management are generally associated with the use of child labor[‖] those children have time to go to school and the average hours worked by a typical child worker are not necessarily incompatible with schooling[**]. On the other hand children of the same age who work outside their family business in paid market work tend to work considerably more hours (31 hours per week on average). As a consequence, those children drop out from school earlier.

[‖] Edmonds and Pavcnik (2005) show that in early 2000s more than 70% of children in the DRC were helping their families in running their business, farm and all sort of housework.

[**] Edmonds and Pavcnik (2005) reports that children between 4 and 16 years of age working in domestic work or helping with their parents or relatives typically allocate 16 hours per week

A closely related consequence of an increase in child labor is fertility choice by families. This because child labor is a family decision. When deciding whether to send the child to a paid job rather to invest their resources by sending their child to school, families generally take into account their valuation of both factors that raise the future return to schooling of the child and present child's wages and choose accordingly. Thus, the choice to send a child to work outside the domestic environment may be encouraged by a shock that raises the family's valuation of the present child's wages (Doepke and Zilibotti, 2005, Hazan and Berdugo, 2002). The cobalt mining boom occurred from 2007 in the Democratic Republic of Congo represents a good example of such a shock and serve as an instrument to fully comprehend the extent to which the boom of cobalt production has affected child labor and in turn how families have responded to it by modifying their fertility choices. This represents a novelty of this paper since no previous economic study has empirical investigated to which extent an increase in child labor translates in fertility rates of women's in sub-saharan african economies.

The method controls for subregional[††] changes and trends within the DRC between 2007 and 2014. A set of relevant individual-specific controls is included in the empirical strategy such as gender, mother level of education, if the village of residence is in a rural or urban area, if the individual has ever migrated and each individual's year of birth. In addition, the model also controls for the year of which the DHS survey was conducted[‡‡].

Additionally, I use two sets of controls. The first control is defined as all individuals living beyond 10kilometres from a cobalt mine. This limit is based on existing economic studies which focused on the effects of mining activities on health and local socio-economic indicators. For instance Benshaul-Tolonen (2018) finds that gold mines have decreased infant mortality in areas within 10 kilometres from the nearest gold mine. Aragón and Rud (2015) finds that the effects of mining activities on

[††] DRC subregions are defined at the administrative level 2 as reported by the DRC government

[‡‡] Two waves of DHS surveys were used in the analysis. The first wave was conducted in 2007, and is regarded as pre cobalt boom survey and the second survey in 2014, and is considered as post boom survey.

the agricultural productivity extends up to 20 kilometres. However, I do not limit the analysis to compare individuals who grew up in areas "more or less close" to a cobalt mine deposit. In fact, in order to clearly identify that the effects on child labor is entirely due to cobalt mines and not being a consequence of any other type of mine, I identify a second set of control group which is constituted by children being far from a cobalt mine but within 10 kilometres to *any other* mine in the DRC.

Finally, the analysis does not limit to the comparison of individuals living more or less close to cobalt mine deposit, but questions if all cobalt mines are equal in terms of unethical use of child labor and in turn on education attainment of those children. Those possible differences within cobalt mines are addressed by considering the different ownership. Various reports from the ILO, Amnesty International (Amnesty International, 2017, International Labor Organization, 2015) show how illegal child labor practices consistently occur in artisanal based cobalt mining deposits and no controls on health and age of miners are implemented in DRC and Chinese owned mines. On the other hand, European and American owned cobalt mines are believed to implement more severe measure to contrast child labor practices. Nevertheless, there are reports and video footage raising concerns on illegal child employment in those mines too.

I show that these results are robust with respect to set of different econometric specifications, each of them addressing a potential concern. First the procedure differentiates all cobalt mines according to their ownership. Results show that although the effects on education attainment are due to all cobalt mines, those owned by Chinese and DRC based companies appear to have a greater negative effects on children education attainment. This last result, confirms concerns raised by a multitude of reports of International Organizations and NGOs Amnesty International (2017), International Labor Organization (2015).

Another concern is if the effects of the cobalt mining boom on the education attainment of children along with their wealth and women's fertility living in the surrounding areas are really due to cobalt mining or they are a direct consequence of any mining activities. The mere use of proximity analysis does not address this concern. To overcome this issue I employ a second set of controls, which consist of all those children who at the time of cobalt mining boom were between 6 and 14 and lived within 10 kilometres from *any other mine* in the DRC. If the differences in terms of education attainment and fertility persist, then the possible concerns of this issue would be alleviated and one would be sure about the impact of cobalt mining uniquely.

Concerns that education attainment for those exposed to the cobalt mining boom during their childhood might reflect preexisting trends are alleviated by the robustness of the differences-in-differences estimates to geographic controls along with an array of different specifications. For example, I examine education attainments and wealth by birth cohort over time. If the exposure to the cobalt mining boom during childhood had a negative treatment effect on education attainment, and no other event interfered to it, then we would expect to see the negative impact of the proximity to a cobalt mine deposit only for post-cobalt boom cohorts relative to pre-boom cohorts. In other words we would expect the impact of living within 10 kilometres from a cobalt mine deposit to be negative and statistically significant only for those individuals born after 1993 (i.e. who were at most 14 years old at the time of the boom). While there should be no significant effect of proximity to a cobalt mine to education attainment for those individuals born until 1992 (i.e. who were at least 15 years old in 2007). The cohort analysis shows significant negative effects in education attainment for those post-cohorts born within 10 kilometres from a cobalt mine deposit. The results obtained with the cohort analysis alleviate the concern that any spurious trends across areas surrounding cobalt deposits and those beyond 10 kilometres would have to be reflected not only at the national level, but within district as well.

An additional concern which might raise with the implementation of proximity analysis is that some other change might have caused the education attainment of individuals in the control group to increase after 2007. For example, the number of schools in the control group might have increase after 2007, driving the education achievements upwards as a result. Our concern is that the effects of cobalt mining boom on education attainment might be biased upward. If no other shock beside the cobalt mining boom which has affected school achievement of people in the control group occurred after 2007 then we would expect the impact of cobalt mining boom to be only limited to those living within 10 kilometres from the nearest cobalt mine deposit, while no effect should be for those people living beyond 10 kilometres. To check for this additional concern, I use a spatial lag model that allows for non-linear effects with distance from the cobalt mine. This method is further explained in the robustness section.

Importantly, the results are robust to alternative strategies which control for all possible migration waves in and out the treatment group. Indeed, another threat to identification is the endogenous selective migration. An active cobalt mine might indeed represent a source of job and attract more people as a result. Some people, indeed might have migrated from the control group (i.e. beyond 10 kilometres from a cobalt mine) to the treatment group (i.e. within 10 kilometres) with the intent to work in a cobalt mine. Poor wealth indicators are often associated to low education levels. Thus, if only the poorest people migrated from control group to the treatment group after the cobalt mining boom in 2007, in order to find a job, the results obtained from the empirical procedure might be overestimated. The procedure adopted takes into account this direction of migration by simply imposing a sample limitation and drop those individuals who at the time of the survey lived in the treatment group but had previously migrated and are in the last quintile in terms of the wealth index (i.e. the poorest individuals who are generally associated with low education attainment). Imposing this sample limitation reduces the treatment effect from 0.37 years of completed education less to 0.33 years, showing that there are not significant spillover effects of selective migration from the control group to the treatment group and vice-versa. This last result, considering endogenous migration, estimates a possible lower bound of the relationship between living in a vil-

The empirical strategy also considers the opposite direction (that is from the treatment group to the control group) lage within 10 kilometers from a cobalt mine and education achievements later in life.

Finally, the model also controls for a set of both geographical and standard of living controls at district level.

By adopting a rigorously quantitative strategy, first, this study confirms the associations between cobalt mining and child labor made public by various reports by Amnesty International, UNICEF and the ILO (Amnesty International, 2017, 2016, International Labor Organization, 2015). Moreover, the results of this paper provide novel evidence on the close relationship between child labor and fertility decisions. Additional evidence on the long-run effects of child labor on wealth is also emphasized. In a contest of low socio-economic development as that represented by the Democratic Republic of Congo, educational decisions made by parents for their children do not result in significantly different wealth conditions later in life.

This paper is connected to several previous studies documenting the relationship between economic booms and human capital formation. Carrillo (2019) focuses on both short and long-run effects of coffee booms in education achievements and future wealth conditions of young adults in Colombia. The study concludes that individuals who, during their childhood, faced higher returns to coffee related work completed fewer years of education and had negative effects on their future wealth. A similar paper is Atkin (2016) who documents that the arrival of formal jobs during years of substantial expansions in export-manufacturing industries in Mexico led to reduced school attendance and lower educational attainment, although it had no overall impacts on subsequent labor market income. This paper differentiates from Atkin (2016) and Carrillo (2019) in different ways. First, this paper focuses and provides also evidence on the relationship between child labor and fertility decisions of families. None of the above mentioned studies focuses on the possible effects on fertility decisions and their implications. Second, the context greatly differs from the previous studies. While Atkin (2016) and Carrillo (2019) focus on middle income countries such

Twenty-one sub-regional districts compose the DRC. See Figure ??

as Mexico and Colombia respectively, this paper focuses on individuals living in contexts of low socio-economic growth such as the Democratic Republic of Congo. As a result, this study shows no effects of fewer years of schooling on future wealth conditions, in line with theoretical models (Eckstein and Wolpin, 1999). Moreover, the greater availability of jobs along with less effective government institutions in preventing unethical labor practices determine a lower cost opportunity of having a child. As a result fertility might increase Doepke and Zilibotti (2005). Moreover, Atkin (2016) focuses on large formal firms, which provide valuable on-the-job-training opportunities and skill accumulation that may offset income losses from reduced formal schooling. On the other hand Carrillo (2019) focuses in informal market shocks which are likely to have different long-run implications.

Another paper, similar in spirit of this study, is Charles et al. (2018) which investigate the relationship between college attendance of young adults in the US and the housing booms. Using the housing boom occurred in the US during the 2000s Charles et al. (2018) conclude that young adults were more likely to drop from college, since the housing boom represented an additional source of available jobs. Although, in the years following the housing boom young adults appear to re-attend college, Charles et al. (2018) show that the trend was not completely reversed. Thus, providing evidence that these shocks may have permanently affected college education in the United States.

However, two important differences should be addressed between this paper and Charles et al. (2018). The first difference concerns the educational level considered. While Charles et al. (2018) focuses on college attendance, this paper on the other hand, focuses on primary and secondary education. As a results, children who drop their primary education to work might lack basic knowledge and cognitive skills to perform well in more complicated labor tasks during their adulthood. The second difference concerns the two socio-economic contexts analyzed in the two papers. While Charles et al. (2018) focuses on the US, an advanced economy with a highly efficient education system. As a results, a reduction in college attendance might significantly worsen the wealth of those individuals compared to those who continue to pursue their degree. This paper focuses on a context of low socio-economic development with a highly inefficient education system. Therefore, the relationship between reduction in education attainment and wealth later in life is not trivial and deserves a deeper understanding. These results are also confirmed by Sviatschi (2019) which uses variations in return to cocaine, to address the use of child labor in coca suitable areas in Peru. Children who have been exposed to increases in return of cocaine are more likely to be incarcerated for violent and drug-related crimes as adults. Interestingly, Sviatschi (2019) also finds no long-term effects of those individuals who, during their childhood, worked in coca plants going to the legal sector. On the other hand Beegle et al. (2004) evaluates the causal effect of child labor participation on education achievements using panel data from Vietnam and an instrumental variables strategy to alleviate endogeneity issues. They show that from age 30 onward the forgone earnings attributable to lost schooling exceed any earnings gain associated with child labor. This scientific evidence proves that existing research suggests that the long term effects of a reduction in the number of completed years of schooling are not obvious and greatly differentiate according to the context of the study.

Moreover, there exists evidence on the negative effects of economic crises on education attainment. An example of such is Thomas et al. (2004) who focus on the 1998 Indonesian economic crisis and show how it resulted in a decline in household spending on education, and in turn in significant reductions in school enrollment.

There exists also an important literature which investigates the effects of general mining activities on health of individuals living in nearby areas both in the short-term and in the long-run. De Putter et al. (2011), Banza et al. (2009) provide evidence about the adverse effects of the pollution due to general mining activities on the health of population living nearby. Xia and Deininger (2019) shows that children living in tobacco farms are shown to have higher likelihood of suffering from illnesses related to green tobacco sickness. Moreover, exposure to large-scale tobacco cultivation reduced the height-for-age z-score also of children too young to work. A recent study (Nkulu et al., 2018) finds that people living in a villages surrounding an artisanal cobalt mine had much higher levels of cobalt in their urine and blood than people living in a nearby control area. Moreover, the authors observe that the differences were most pronounced for children, in whom they found evidence of exposure related DNA damage.

However, potential negative health effects of mining could be offset from the positive impact of the mining boom on the local economies surrounding mining sites, as shown in a recent paper by Benshaul-Tolonen (2018). Specifically, Benshaul-Tolonen (2018) finds that the booms in gold mining in the sub-saharan african region was associated to a reduction in infant mortality rates, although some limitations arise on the possible mechanisms of such effect. Moreover, geo-localized data on health of infants and individuals is scarce and long-term implications of mineral exposition while certainly a crucial issue is so far difficult to assess.

The main goal of the present study is, on the other hand, to clearly assess the effects of cobalt mining on school achievements, through a unique mechanism that is cobalt-related child labor, as well as on future wealth of those individuals who grew up in villages surrounding a cobalt mine deposit after its boom, and to evaluate how families respond to the increase availability of paid jobs for children through women's fertility rates.

Finally, the so called "resource curse" economic literature also relates to the present study. The term "resource curse" suggests that the abundant presence of mineral and agricultural resources is generally associated with low economic development (Hausmann and Rigobon, 2003, Sachs and Warner, 1995). Manifold reasons have been exploited, such as conflicts which might be fostered by the concentration of minerals in one region and the relative scarcity of them in a neighboring region (Berman et al., 2017, Dube and Vargas, 2013, Angrist and Kugler, 2008). Subramanian and Sala-i Martin (2003) focuses on the deleterious long run impact of oil abundance on institutional quality for the case of Nigeria. Similarly, Tsui (2010) focuses on high quality oil discoveries and their negative impact on democracy in Iraq.

The main limitation of the paper is linked to data availability. A real long-run analysis of the impact of cobalt mining boom on individuals' wealth in the DRC is still not possible since the cobalt mining boom occurred relatively recently, (i.e in 2007), and the last wave of the DHS in the DRC was conducted in 2014. Thus, the present data make possible to assess the effects of cobalt mining after a period of seven years. As a result, those individuals who at the time of the boom were between 6 and 14, in 2014 aged between 15 and 22 years old. This data limitation constitutes an issue when assessing the long term effects of cobalt mining on future wealth. This, because possible long run effects of a decreased education attainment on wealth could not be soon visible, rather they might need a few more years to show.

On the other hand, this is not an important issue for what concern both fertility decisions and education attainment of those children who at the time of cobalt mining boom were between 6 and 14. The reason is that DHS data show that about 87% of those people who are 15 years old state that they had already completed their education. Therefore, it is reasonable to consider a 15 year old boy living in the DRC as an adult. That said, a new wave of surveys implemented in 2017/2018 might soon be publicly available. This, would reduce possible concerns of too short time span between the onset of the cobalt boom and the evaluation of the long-term effects on wealth.

The remainder of the paper is organized as follows. Section 2 provides background information about the boom in cobalt production from mining in the DRC, caused by the sharp increase in the global demand of electric batteries. In Section 3, I present the data on education attainment, along with measures of wealth conditions used in this study. Fertility data is also described along with sources of gps data of cobalt mine deposits in the DRC. Section 4 presents a brief conceptual framework showing the relationship between child labor, fertility rates and its long run effects on wealth of individuals. In Section 5, the empirical strategy examining the effects of cobalt mining boom on education attainment, fertility rates and future wealth is described. Section 6 shows the baseline results from the analysis. In Section 7, all possible estimation concerns are presented and robustness checks are described. Finally, Section 8 concludes. In the Appendix, I provide further results on children along with additional robustness checks.

1.1 Background and Institutional Context

1.1.1 Electric Batteries and Cobalt

Currently the most popular technology of the battery sector is that of the lithium-ion battery. Compared to Nickel-based battery types, which were greatly used in the past, lithium-ion batteries have superior energy, power density and superior cycling ability Cobalt Institute (2019). These major advantages of the lithium-ion batteries make them first choice for the manufacture of new electronic vehicles.

Figure 1.1 highlights the global demand breakdown of modern lithium-ion batteries by type. Figure 1.1 shows how cobalt based batteries dominate worldwide demand of lithium-ion batteries, with Lithium Cobalt Oxide(LCO), lithium-nickel-cobalt-aluminium-oxide (NCA) and lithium-nickel-manganese-cobalt-oxide (NMC) technology supplying almost 70% of the global lithium-ion battery market.

The possible applications of modern lithium-ion batteries are manifold and depend upon the minerals used. Specifically, the most popular lithium-ion technology is the lithium-cobalt oxide (LCO) battery. The main feature of the LCO battery is the high energy density translating into a long run-time. This characteristics makes the LCO particularly suitable for portable devices such as cell phones, tablets, laptops, wireless headphones and cameras (Cobalt Institute, 2019). LCO batteries contain approximately 60% cobalt which accounts for 50% of the total weight of the cathode. The LCO batteries are not particularly suitable for large and heavy devices since their principal disadvantage consists in having a low life span, and a relatively low safety performance at hot and cold temperatures. However, Cobalt Institute (2019) reckons that the majority of research which nowadays focuses on electric batteries is particularly focusing on the LCO and constant improvements are being made in terms of greater durability and safety.

The second most popular lithium-ion battery type is the lithium-nickel-manganese-cobalt-oxide (NMC) which contains about 20% of cobalt. Although the NMC battery has a lower capacity than that of the LCO, they have a high cycling rate which, combined with a high capacity, high power and their particularly long-life makes them ideally suited to be used in the new generation of electric cars and buses. NMC also suits its use for power tools, modern e-bikes and electric motor vehicles. An additional feature of the NMC battery is that it has the lowest self-heating rate out of the different types of lithium-ion batteries. Hence, it is also particularly safe.

Finally, the lithium-nickel-cobalt-aluminum-oxide (NCA). This particular type of lithium-ion battery shares similar qualities with NMC batteries in that it also has a high specific energy, specific power and a long life span. Downsides to the technology of NCA is that it is expensive and has poor safety. For these reasons NCA batteries are mainly used in industry, medical devices and also in electric power trains.

Where does the cobalt produced go? In Figure 1.2, I show that about half of all cobalt produced is used to fuel modern batteries with the other half being used for the construction of superalloys, hard metals (capable of withstanding high temperatures) tires and magnets. Where does cobalt come from? Darton Commodities (2017) estimates that 65% of total worldwide cobalt supply comes from the Democratic Republic of Congo (DRC). Figure 1.3 shows how the second biggest supplier of cobalt in the world is Russia with just 6% of the total cobalt production. This highlights the importance of the DRC as producer of such a critical material for our modern era.

As what concerns the global demand of cobalt US Geological Survey (2019) shows that in 2015

Considering no major breakthrough invention in battery technology, Cobalt Institute (2019) forecasts that each electric-vehicle battery will need about 18 pounds of cobalt. This is over 1,000 times as much as the current quantity of cobalt needed in a battery of a smartphone (i.e. approximately 7 grams). A recent report by Euractiv (2019) states that Volkswagen, for example, expects it will need to build six giant battery factories within a decade simply to supply its electric-car plants.

global cobalt demand increased to 90 Mt. According to the CRU commodity consulting agency, the global demand of cobalt will further accelerate in the next years to reach 155 Mt in 2025. I also show that the DRC dominates global reserves of cobalt, with approximately 9,000 Mt of cobalt available on the existing mining sites (see Figure 1.4).

1.1.2 Cobalt Mining and Child Labor in the DRC

Almost 70% of total lithium-ion batteries are made of cobalt. The Democratic Republic of Congo is naturally rich of such a critical mineral, with roughly 65% of the world's cobalt production coming from the DRC's southeastern province of Lualaba, near the border with Zambia (see Figure **??**). Mining accounts for about 80% of the DRC's earnings. However, despite intense mining activities of tin, gold, nickel, copper, and now cobalt, the World Bank reckons that the average person in the DRC earns just 700 USD a year. The Sub-Saharan African country therefore suits as a good example of "mineral resource curse".

Figure 1.5 shows that since data on cobalt production from mining were available for the DRC, the production of the core minerals was stable at around 20 million tonnes per year. This despite the noteworthy country political instability. However, starting from 2007 with the great usage of modern lithium-ion batteries China entered in the market of cobalt by importing cobalt supplies directly from the DRC. The production of the african country remarkably increased, tripling to around 60 Mt per year in 2010. These levels remained constant until 2016 when the increasing pressure for modern Electric Vehicles (EVs) has determined a second cobalt boom which is not over yet. US Geological Survey (2019) forecasts that production of cobalt from mining in the DRC will be 90 Mt in 2018 and will reach 125 Mt by 2025 putting even more pressure to the already unstable national institutions. Nowadays, around 70 Mt of the total world cobalt production is mined in the DRC (US Geological Survey, 2019) which makes the unstable country the largest world supplier of of the critical mineral. Other producers are China which accounts for the 8% of the total world production of cobalt, Canada (6%), Russia and Cuba (4% each). Remarkably, the top four cobalt mining sites are located in the DRC and account for 80% of the total cobalt production of the country and 43% of the total cobalt production in the world.

The Democratic Republic of Congo is predicted to remain by far the world's leading producer of mined cobalt and is estimated to represent one-half of global production (US Geological Survey, 2019). However, the vast mineral wealth, and in particular cobalt which will be even more important in the next decade, has failed to contribute to increase the overall standard of living of the katanga province in the specific case as well as that of the country as a whole. Various reasons concerning the failure of cobalt mining activities to increase the standard of living of the katanga province have been proposed. Cuvelier (2017), for instance points at the unskilled migration waves which occurred since late 2000s. The unskilled labors has contributed to depress wages in cobalt rich regions.

As previously stated, the economy of the DRC as well as that of the Katanga province extensively relies on mining activities and in particular in cobalt mining since the last two decades. However, raw cobalt is only mined in the DRC, it is not refined. Olivetti et al. (2017) show the only 0.4% of the total production of cobalt from mining is refined in the DRC, while the vast majority of the total production is sold to Chinese owned companies which refine the raw product in China.

In 2017 a large portion of the country's cobalt mine production was from cobalt ores mined by industrial or mechanized methods, and a small portion, around 20% was gathered by tens of thousands of artisanal miners by handpicking cobalt-rich ores. Spencer (2016) estimated that artisanal mines in the Democratic Republic of Congo produced approximately 10,500 tons of cobalt in 2015. China represents the leading destination for the DRC cobalt exports US Geological Survey (2019).

Eight of the 14 largest cobalt mines in Congo are now Chinese-owned, accounting for almost

2017 is the latest year for which public data is available.

half of the country's output. In a recent article, the Financial Times (2019) also shows consistent evidence that chinese owned cobalt mines hire informal miners. In particular, the report estimates that there are around 200,000 informal cobalt miners in the DRC who in large part work in chinese and DRC owned mines.

The artisanal and small-scale mining (ASM) is generally defined as 'labour-intensive, low-tech mineral exploration and processing activities' Hilson (2011). ASM usually employ independent diggers who have converged on the cobalt rich area to search for the critical mineral, often with primitive tools. The precious mineral is then sold to chinese companies as recently reported by the Financial Times (2019). Various reports show that although companies are not allowed to buy cobalt from unknown sources, since the unethical use of child labor cannot be excluded, chinese owned companies keep this practice. Hence, contributing to the use of children in cobalt mines. Amnesty International (2017) reckons that the cobalt diggers include an unknown number of children.

The artisanal mining sector has witnessed a dramatic expansion since the late 2000s. A report by Center for International Forestry Research (2013) provides detailed information on how the Artisanal and Small-scale Mining (ASM) works. Artisanal miners work either legally, in concessions earmarked for ASM by the DRC government. These zones are known as Zones d'exploitation artisanale (artisanal exploitation zones). They are created by the National Minister of Mines. However, a large portion of artisanal miners works illegally. They use shovels, pickaxes and other rudimentary tools to excavate copper and cobalt ores. According to the most recent reports of Amnesty International and the International labor Organization (ILO) approximately 20% of total cobalt mined in the DRC comes from those artisan-based mining operations (roughly 16 Mt per year) in which labor supervision is scarce if null, and the risk of unethical use of child labor is relevant (International Labor Organization, 2015, Turcheniuk et al., 2018). Once mined by hand, these minerals find their

Reports also include cctv cameras by the CNN news, Sky news, the Financial Times

Amnesty International (2017) interviews different children working in cobalt mines, such as the 15-year-old Lukasa, who support their families by digging small quantities of cobalt by hand way to the international market through the mediation of local mineral buyers who buy them from miners and sell them to authorized trading houses in big cities such as Lubumbashi, Likasi and Kolwezi.

According to the most recent report of Amnesty International (2017) all children interviewed in locations close to cobalt deposits, namely Kasulo, Kolwezi, Malo Lake, Kambove and Kapata report working above ground, either collecting the mineral from the mountains of tailings in active and inactive industrial mining concessions, or working in streams and lakes close to the concessions where they washed and sorted the stones or in some case even carrying bags weighing between 20 to 40 kg to earn between one and two dollars a day. UNICEF estimated in 2014 that approximately 40,000 boys and girls work in cobalt mines across the whole of Katanga province. This is more than the double the number of children reported working in rest of the DRC provinces. Finally, Faber et al. (2017) conducted a set of representative large-scale surveys in artisanal mining communities of cobalt-rich areas of the (DRC). They estimate that 11% of children in these communities work outside of the home, of which 23% (or an estimated 4,714 children in the entire population of the 426 communities) work in the cobalt mining sector.

Assessing the long term effects of child labor has been a crucial task for social scientists as well as for clinical researchers. Nevertheless, while the health effects of child labor in cobalt mines are well known (Nkulu et al., 2018), assessing the economic consequences of this phenomenon has proven to be a difficult and uncertain task. For example, Basu and Van (1998) child labor depresses adult wages, making child labor necessary and more persistent over time. The effects of being rich of critical minerals on child labor and child cognitive development is an empirical question. It will depend not only on the intensity to which the minerals are extracted but also on the general wealth of households surrounding the mining sites and the effectiveness of institutions and policies to prevent the children to be sent to the mines thus subtracting time and energies from school. For this reason, I also examine the effects of cobalt mining on the wealth of those individuals who grew up close to active cobalt deposits. Additionally, as a recent survey conducted by Faber et al. (2017) shows, a higher number of children working in cobalt mines is associated to higher fertility rates. Hence, this paper wants to clearly assess the relationship between cobalt mining, child labor and fertility choices.

1.2 Data

1.2.1 Education Attainment Data

The dependent variables, i.e. education achievements, individual's wealth and women's fertility rates used in the analysis come from the Demographic Health Surveys (henceforth DHS). Specifically the DHS provide data on health, education attainments, as well as the evaluation of cognitive skills of children. More details on the data can be found in the relevant appendix. The wealth index computed in the DHS divided the households into 5 wealth quintiles. The indicator ranges from 1 (bottom quintile) to 5 (top quintile). Wealth index shows relationship between wealth, population and health indicators of households.

The data relative to the highest completed year of education of individuals were constructed based on data from the Demographic Health Surveys (henceforth called DHS). Specifically, two waves of DHS were used in the analysis: the first wave which took place in 2007 constitutes the period pre cobalt mining boom, while the second wave constructed in 2014 represents the period post cobalt extraction boom. All gps provided by the DHS have an error ranging from two to five kilometers, due to privacy reasons. For this reason the DHS does not recommend researchers to restrict their studies to less than 5 kilometers distance.

All DHS data record the gps of each individual interviewed. The analysis considers all individuals who during their childhood period (i.e. between 6 and 14 years of age) grew up close to a cobalt

A third set of surveys conducted in 2018 is expected to be available soon

See the relevant appendix on DHS data and privacy restrictions of the corresponding gps locations to have more information of the gps errors.

mine and evaluates them later in time, in 2014, when their education is completed. Therefore, individuals who were born from 1993 onward (who at the time of the boom of cobalt were at most 14) and lived within 10 kilometers from a cobalt mine deposit constitute the treatment group, while those born until 1992 (who were already 15 years old in 2007, therefore not falling in the child labor definition of the International Labor Organization) are part of the control group. I construct a database of a total of 43,385 individuals in the DRC who were between 6 and 14 years of age at the time of cobalt mining boom. Among these, 1,904 lived within 100 kilometers from the nearest active cobalt mine deposit at the time of the interview, while 548 individuals lived within 10 kilometres from a cobalt mine deposit. Figure **??** shows the gps of the individuals interviewed in the DHS.

1.2.2 Women's Fertility Data

Data for fertility rates of women were also provided by the DHS. The analysis considers all women during their fertile period (i.e. between 15 and 39 years old). Each of the interviewed woman states the number of total children and their respective birth year. I do not differentiate from the total number of children ever born and the number of stillbirths, since the aim of the empirical investigation is to assess the effect that cobalt mining boom had on the willingness to have more children and not on the effective health of the children. Health effects of cobalt mining go beyond the scope of this paper. Moreover, differently from Benshaul-Tolonen (2018) this paper does not consider infant mortality rates and other measures of health, since one of the goals of the following study is to assess the relationship between child labor in cobalt mines and women's fertility in those areas through a lower opportunity cost of having a new child.

I consider two measure of fertility for each woman: the five-year fertility rate and three-year fertility rate. The first measure of fertility considers the number of children ever born for each woman in the last five year preceding the interview, that is for all women surveyed in the 2007, the number of children born in the period between 2002 and 2007, while for all women surveyed in 2014, the number of children ever born during the period 2009-2014. The second measure of fertility considers the number of children ever born from each woman during the last three year before the interview, that is for all women surveyed in the 2007, the number of children born in the period between 2004 and 2007, while for all women surveyed in 2014, the number of children ever born during the period 2011-2014. Since, DHS for the DRC is available in two waves, all women surveyed in the 2007 wave constitute the pre treatment group. As for children, I assume that the effect of the cobalt mining boom should be concentrated with 10 kilometers. Therefore those women, surveyed in 2014 and living within 10 kilometres from a cobalt deposit constitute the treatment group. If fertility choice do not respond immediately to a particular shock, then we would expect that although the five year fertility rate would have been higher for women within 10 kilometers from the nearest cobalt mine deposit, this would have been lower than the three year fertility rate. I construct a database of a total 21,816 women from the DRC between during their fertile period (i.e. aged between 15 and 39), among which 844 women reported living within 100 kilometres from a cobalt mine deposit at the time of the survey and 277 lived within 10 kilometres from a cobalt mine.

1.2.3 Cobalt Mines Data

Different sources have been used to identify the location of all deposits of cobalt and other minerals mined in the DRC. Specifically, I use the most recent version of all cobalt deposits obtained from the digitization of data from Ghosh Banerjee et al. (2014) and from the US Geological Survey (2019) which is publicly available. Moreover, data for the actual production of cobalt for different mines were retrieved from the Cobalt Cobalt Institute (2019). More information on the data used in the analysis are found in the Appendix. The aforementioned sources contain geo-referenced data on all mining deposits in the DRC. Combining all locations of mining sites of minerals in the DRC, Figure **??** was obtained.

Approximately 80% of total mining capacity in the DRC is owned by large non African companies (in particular Chinese, European and Australian companies), while the remaining 20% (i.e. approximately 15 Mt of cobalt mined every year) is constituted by artisan-based mining operations. In addition, various reports from CNN news, Sky news, the Financial Times show that although companies are not allowed to buy cobalt from unknown sources, since the unethical use of child labor cannot be excluded, chinese owned companies keep this practice. Hence, contributing to the use of children in cobalt mines.

Therefore, I also combine gps data with the ownership of the cobalt mines Cobalt Institute (2019). To this regard the aim of this study is not only to provide the effects of cobalt mining on the education attainment of children but also if this phenomenon is circumcised on artisan-based cobalt mining activities or also those mines owned by Chinese or European companies. In other words the goal of the paper is also to understand if artisanal and small scale cobalt mining practices are associated with illegal child labor activities whereas large-scale cobalt mining, being more controlled might disincentivize unethical uses of labor and therefore being beneficial for the economy of the areas surrounding cobalt deposits in the DRC. I combine data on the location of all mine deposits in the DRC, including cobalt mining sites, and gps of all individuals interviewed in the DHS to calculate the distance of each individual interviewed to the closest mining deposit. Figure 1.6 is thus obtained by combining the two sources of data. The analysis will compare education attainment, fertility rates and wealth of individuals living within 10 kilometres from a cobalt mine, against those beyond 10 kilometers and within 100 kilometres.

Treatment is based on the proximity of the village of residence of each individual interviewed and the closest active cobalt deposit. The distance is computed using GIS software and gps coordinates of individuals recorded during the surveys and the exact position of the mine. The definition of treatment group closely follows relevant economic and medical literature focusing on the economic and medical effects of mining activities, respectively. For instance Cust (2015) examines the labor market effects of various industrial mines in Indonesia. He finds that the effects of mining activities propagate to approximately 10 kilometres within each mine. Aragón and Rud (2015) find that the effects of pollution due to mining are felt up to 20 kilometres with a mineral mine, while Benshaul-Tolonen (2018) finds that gold mines within 10 kilometres have positive effects on infant health due to better local development. In the particular case of cobalt mines a sufficient number of observations (i.e. children and women) is located within 10 kilometres from a cobalt mine. Therefore, treatment group is defined as all children who were born and live within 10 kilometres from the closest active cobalt mine and were between 6 and 14 at the time of the cobalt boom (i.e. were born since 1993).

Additionally, beside comparing individuals living within 10 kilometers from a cobalt mine deposit with those living beyond 10 kilometres within 100 kilometres from a cobalt mine, the analysis also uses a second set of control which is composed of individual living within 10 kilometres from any other mine in the DRC. Therefore I compare living within 10 kilometers from a cobalt mine deposit with those living with those living within 10 kilometers from any other mine. See Figure **??** to have a visual representation of treatment and control groups.

We expect that by increasing the distance from the individual to the nearest cobalt mine the effect of cobalt mining of our variables of interest would decrease. As argued in Benshaul-Tolonen (2018) I limit the individuals living in villages within 100 kilometres from a cobalt mine, as well as controlling for sub-regional changes over time, in order to exclude confounding factors along with any other change occurred in 2007 beside the boom in cobalt mining.

Table 1.1 reports summary statistics of relevant variables pre and post 2007 for villages within 10 kilometres from the closest cobalt mine deposit and between 10 and 100 kilometers. From Table 1.1 we observe that the average number of years of education completed by young adults in the DRC is less then four. In particular DHS data reveals that six years of education, represents the highest education level for adults in the DRC and it is reached by about 18% of the total adult population (i.e. DHS shows that 761 out of 9,452 adult individuals complete six years of schooling in their life).

Data also shows that 83% of those individual typically complete their highest education level before they turn 15. For this reason it is reasonable to assume that the vast majority of individuals do not keep studying after they turn 15.

Table 1.2 reports summary statistics of relevant fertility related variables pre and post 2007 for villages within 10 kilometres from the closest cobalt mine deposit and between 10 and 100 kilometers. From Table 1.2 we observe that, on average, each women living within 10 kilometers from a cobalt mine, and were surveyed in 2007 had approximately 0.77 children born in last five years preceding the interview, 0.45 during the last three years and 0.14 in the year of the interview. These statistics compare to an average of 1.02 children born during the last five years preceding the interview for those women living within 10 kilometers from a cobalt mine, who were surveyed in 2014 DHS wave; 0.68 children were born on average for each woman during the last three years and 0.26 in the year of the interview. In general, Table 1.2 shows that while five-year fertility rate of women living beyond 10 kilometers from a cobalt mine surveyed in 2014 increase by approximately 10% compared to those surveyed in 2007 that for those women living within 10 kilometers from a cobalt mine deposit increased by over 32%. Similarly, the three-year fertility rate of women living beyond 10 kilometers from a cobalt mine surveyed in 2014 increase by approximately 6% compared to those surveyed in 2007; while that for those women living within 10 kilometers from a cobalt mine deposit increased by over 50%. Summary statistics on children are included in Table A.1.

1.3 Conceptual Framework

An important additional question is also addressed in the following study, that is: are families living close to a cobalt mining site more prone to have children? The rationale behind this question is that families might take into account the lower opportunity cost of having a child since the latter might be sent out to work in a cobalt mine as soon as he turns 6. The methodology addresses this question by comparing fertility rates as well as willingness of having more children of women between 15 and 39 living in villages more or less close to a cobalt deposit. Results from the comparison of fertility rates of women before and after 2007 reveal that since the opportunity cost of having an additional child decreases, those families leaving close to a cobalt mine had a higher fertility rate after 2007 compared to those living beyond 10 kilometres from a cobalt mine of close to any other mine in the DRC. Those important results are also confirmed by the cohort analysis.

1.4 Empirical Strategy

The study focuses on the effects of cobalt mining boom which occurred in the Democratic Republic of Congo after 2007 on education attainments and cognitive development of children who at the time of the boom in cobalt demand were between 6 and 14 years of age and were living in areas surrounding a cobalt mining deposit.

The boom in the demand of cobalt was caused by the electrification revolution which allowed the diffusion of modern lithium-ion batteries. Lithium-ion batteries are principally contained in modern smartphones, electric vehicles and PCs and heavily rely on cobalt as primary material to function properly.

In the analysis, children who were between 6 and 14 and lived within 10 kilometres away from a cobalt mine define the treatment group. Therefore, those children in the second wave of the DHS conducted in 2014 were between 15 and 22 years of age.

To this regard, the boom of cobalt mining which started since 2007 due to the worldwide increase in demand of modern lithium-ion batteries with smartphones, PC and EVs serves as a quasi-experiment to understand how education attainment, cognitive development, wealth and consequently fertility rates have changed with intensive cobalt extraction. In particular, the effects on wealth are not trivial. On one hand children might be subtracted from school to work in the mines (or in the best case scenario they might still be able to attend school, but dedicate less hours to studying, consequently they will have lower cognitive skills compared to their peers who attend school full time). However, on the other hand, by working they might contribute to their family standard of living or alleviate the conditions of poverty they live in. Moreover, child labor also depresses wages since children are generally paid less than adults. To this regard, the economic burden of artisan-based cobalt mining is ambiguous.

The sudden boom in cobalt mining has fostered artisan based mines. These artisan mines constitute provide for the 30% of the total production of cobalt of the DRC and are not adequately supervised. Thus, the risk of unethical labour practices is high as highlighted in the last report of the Amnesty International (2017).

Almost 70% of total lithium-ion batteries are made of cobalt. For this reason, the boom of modern lithium-ion batteries which happened since 2007 constitutes a quasi-experiment to assess if cobalt mining impacted on child education achievements and their cognitive abilities and if so in which measure it did. The recent boom in cobalt mining in the DRC also provides for a good quasi-experiment since it came overwhelmingly from outside the DRC, notably from the most advanced economies. Therefore, it is reasonable to assume the exogeneity of cobalt mining boom with any other variable affecting child labor trends in the DRC. Additionally, as shown by International Labor Organization (2015) 80% of cobalt mines in the is owned by non African companies. The additional 20% of total production of cobalt from mining is instead artisan based where little is known.

I first compare school achievement rates for cohorts of children who were between 6 and 14 before and after the boom of cobalt mining and lived in villages within 10 kilometers and between 10 and 100 kilometers from the closest cobalt deposit. Those differences in differences estimates show that school achievement of children living close to cobalt mines decreased after 2007 compared to

Glencore, a mining company from Switzerland controls 29% of total cobalt production in the DRC; while 45% of total supply of cobalt is Chinese owned. The rest 6% instead is controlled by companies based in Luxemburg and Canada.

those children living beyond 10 kilometers from any cobalt mine. However, by simply comparing school achievement rates we will miss those children who live within 10 kilometers from a cobalt mine and find the time to also attend school after work. Those children, indeed, could attend school but it is likely that their performance and cognitive skills could be lower than those of their peers who instead are full time students. In order not to miss the true effect of cobalt mining on children cognitive development, I also use a dataset with cognitive assessment for children between 6 and 14 before and after the cobalt mining boom in the DRC. A limitation of the present economic strategy is represented by the fact that evaluating the effects of cobalt mining on the education attainment we will miss those children who would not go to school other than working as cobalt miners. Those children generally help their families in running their business, farms and any type of domestic work activities. This data however, do not allow to quantify how many of those children "switch" from not harmful child labor to harmful child labor.

Nevertheless, this is not an issue since the inclusion of those children will likely increase the effect of cobalt mining on child labor.

The main specification is presented as follows:

$$\text{Outcome}_{i,c,d,t} = \alpha + \beta_1 \ (Post)_t \times (\text{Cobalt Mine})_c + \beta_2 \ (\text{Cobalt Mine})_c +$$

$$+ \text{"}\gamma \ \mathbf{X}_i' + \delta_t + \sigma_{1,d} + \sigma_{2,dtrend} + \varepsilon_{\ i,c,d,t} \quad (1.1)$$

where the outcome variable $Outcome_{i,c,d,t}$ represents either the number of completed school years and the evaluation of cognitive development of individual i, of birth year cohort t, living in DHS cluster c, in the sub-regional district d. $(Post)_t$ indicates whether the individual i's birth year was later than 1992 (in this case the child was at most 14 years old in 2007 and therefore a possible cobalt miner) or earlier than 1992 (in this case the individual i was more than 14 years old and could not be considered as a child at the time of cobalt boom). Variable $(Post)_t$ is interacted with

(Cobalt Mine)$_c$ which represents the measure of distance between the cluster of residence c of individual i and the nearest active cobalt deposit. If the nearest cobalt mine was distant within 10 kilometres from cluster c and the individual i's birth year was 1992 or later than he would be considered in the treatment group. Otherwise he would compose the control group. The regressions control for differences among different sub-regional districts in the DRC between the seven years of the two DHS waves $\sigma_{2,dtrend}$; birth year cohort-fixed effects δ_t and a vector of parents education levels, along with individual-specific characteristics $\mathbf{X}'_{i,c}$ such as, his/her gender, if the individual i lives in a rural of urban area, if he or she ever migrated from their native village. Additionally, since a limitation of the present study is constituted by the relatively short time period between the onset of the cobalt mining boom in the DRC (i.e. 2007) and the last DHS wave conducted in the Sub-Saharan African country (i.e 2014), those individuals who at the time of the boom were between 6 and 14, in 2014 aged between 15 and 22 years old. As a result, a possible concern that individuals between 15 and 22 years of age might still be enrolled in a school arises. To alleviate this concern, I introduce an individual specific control variables which takes into account if the individual "i" is still in school. Moreover, DHS data show that about 83% of the people interviewed who are 15 years old state that they have already completed their education. Therefore, it is reasonable to consider a 15 year old boy living in the DRC as an adult.

Different factors such as business environment, policies, corruption, education expenses might vary at district level and over time. For this reason the specification 1.1 also controls for sub-regional districts time trends. Finally, opening a cobalt mine requires that a stable deposit of cobalt must be present underground. This is exogenous to our outcome variables such as education achievement, cognitive skills and women's fertility rates. Our coefficient of interest is β which represents the difference-in-difference estimate of cobalt boom on children school achievements.

There are eleven regions in the DRC, identifies with the code "admn1". Twenty-one sub-regional districts are instead represented as "admn2" and correspond to provinces

1. If the boom in cobalt production from mining has caused children in the DRC who were going to school to drop out earlier and hamper their cognitive abilities;
2. if cobalt mining boom has caused individuals who grew up in cobalt mining areas during the boom of cobalt to show worse wealth conditions later in life.

Finally, this paper investigates a possible consequence of the child labor increase in villages surrounding cobalt mines, that is changes in fertility rates observed in those villages after the cobalt mining boom. In other words since results obtained with rigorous quantitative analysis show that children in areas within 10 kilometres to a cobalt mine were sent to the mine as soon as they turn 6, then a direct consequence of that effect would be a lower opportunity cost for families to have a child. Thus, families would find more convenient to have an additional child since the latter might soon actively contribute to his family's income levels by working in the nines. The possible increase in fertility rates in areas within 10 kilometres to a cobalt mine might have crucial implications on the long-term socio-economic growth of those areas. Families might choose to invest less in their children education, opting for sending them to mine cobalt stones.

1.4.1 Parallel Trends Assumption

The difference in difference strategy presented above relies on the parallel trends assumption. In other words, in the case of no occurrence of the cobalt mining boom we would expect that the control groups, defined as those individuals living between 10 and 100 kilometers away from a cobalt mine and the treatment group, composed of those individuals living within 10 kilometers away from cobalt mine, would have shared the same pattern in our outcomes of interest (i.e. education attainment and fertility rates). Therefore, by verifying the validity of the parallel trends assumption we want to exclude that any other event which occurred before the cobalt boom affected school achievements in the control or in the treatment group.

I use both linear, non-parametric techniques and regressing the dependent variables on the full sample using the control variables, sub-regional district time trends and fixed effects defined in the main specification. First, I limit the sample until the onset of the cobalt boom and subsequently limiting the sample to years following the cobalt boom. Figure 1.7 shows the local polynomial smooth estimates of education attainment on the y-axis of each birth-year cohort, represented on the x-axis. Individuals living in cobalt mining villages and those living in non-cobalt mining areas are shown to be on similar trends in education attainment for all birth-cohorts until 1992 (graph on the left). This is confirmed considering those individuals living in other-mining villages as a second control group (graph on the right).

Figure 1.8 shows the average number of children per woman for each village on the y-axis at a given year, from 2000 to 2013, on the x-axis. Women living in cobalt mining villages and those living in non-cobalt mining areas are shown to have similar trends of number of children born per year until 2010 (approximately 3 years since the boom of cobalt). This is confirmed considering those women living in other-mining villages as a second control group. In addition, I also regress the dependent variables on the full sample using the same control variables, sub-regional district time trends and fixed effects defined in the baseline equation 1.1. Results are shown in Figures 1.9 and 1.10.

In Figure 1.9 I run the main specification provided in equation 1.1 limiting the sample of children between 6 and 14 surveyed in 2007 only (pre boom cohorts) and in the 2014 wave only. The only difference with the baseline specification is that I consider current schooling year instead of completed education since children in those ages are still going to school. I use the same control variables, sub-regional district time trends and fixed effects defined in the baseline equation 1.1.

The regressions estimated are presented as follows:

$$\text{Outcome}_{i,c,d,t} = \alpha + \beta \,(\text{Cobalt Mine})_c + \gamma \,\mathbf{X}'_{i,c} + \delta_t + \sigma_{1,d} + \varepsilon_{\,i,c,d,t} \tag{1.2}$$

Since, I first limit the sample to only individuals surveys in the 2007 DHS wave, and to those interviewed in the 2014 wave after, the indicator variable $(\text{Cobalt Mine})_c$ indicating the presence of a cobalt deposit within 10 kilometers from the village where individual i is living, is not interacted with a Post indicator variable. Therefore, this time the coefficient of interest β reveals the relationship of being 10 kilometers away from a cobalt deposit and the education level of children or women's fertility rates.

If nothing occurred in cobalt-mining areas prior to 2007, then we would expect the coefficient β to be not significantly different from zero when limiting the sample to individuals surveyed in 2007 only, while to be significantly different from zero when considering only individuals interviewed in 2014.

Figure 1.9 shows the effects of proximity to a cobalt mine deposit on current education attainment of children aged between 6 and 14. On a similar note, I run equation 1.2 considering only women surveyed in 2007 (i.e. pre cobalt boom) first, and only those surveyed in 2014 (i.e. post cobalt boom), then. As for children's education we would expect that being within 10 kilometres from a cobalt mine deposit should have no effect on fertility rates before 2007, while a positive and statistically significant effect on fertility only for women interviewed in 2014. Figure 1.10 shows the effects of proximity to a cobalt mine deposit on 5 year and 3 year fertility rates.

Obviously, no conclusions can be drawn about the cobalt mining boom as the unique determinant of such differences. This section merely wants to address that prior to 2007 (year of which the cobalt boom occurred in the DRC) being close to a cobalt mine deposit was not associated with a reduction in education attainment and with an increase in fertility rates. Further robustness checks as cohort analysis and spatial lag models are implemented in section 2.5.

1.5 Results

1.5.1 Education Attainment

I begin by examining the results of the baseline equation 1.1. The beta coefficient resulting from this specification compares education attainment of individuals who grew up in cobalt mining areas with those individuals, who during their childhood, were living in non mining areas, before and after the cobalt boom.

Table 3 reports beta coefficients of the effect of the boom of cobalt on educational attainment based on equation 1.1. Column (1) presents results from a specification with no individual controls, sub-regional district time trends, or fixed effects. In this first specification, I find a significant effect of cobalt mining on schooling achievements later in life, with a coefficient of -0.569 (and a standard error of 0.226). Columns (2)-(3) add other controls and fixed effects sequentially. The addition of individual-specific controls such as gender, type of residence along with a dummy variable indicating if the individual surveys has ever migrated, in column (2) have small effects on the estimated coefficient, which is now -0.505 (and a standard error of 0.197). Finally, specification in column (3) includes a set of fixed effects, i.e. survey year and district by time fixed effect. This specification further controls for the year of birth of each individual surveyed. Results, confirm what shown in columns (1)-(2), with a coefficient of -0.496 (and a standard error of 0.185). Taken together, these results imply that the cobalt mining boom during school-going years led to an average of 0.5 fewer years of completed schooling in areas surrounding a cobalt mine deposit.

Overall, I observe that the cobalt mining activities reduce educational attainment in areas surrounding a cobalt deposit. The estimates imply that the boom of cobalt led to a decline in completed schooling of approximately 0.5 year of those individuals born since 1993 (post boom cohorts) and living in areas surrounding a cobalt deposit, compared to those individuals who were born until

1992 (pre boom cohorts). Additionally, results show no effect of cobalt mining on future wealth of individuals who grew up within 10 kilometres away from the nearest cobalt mine compared to those who during their childhood lived between 10 kilometres and 100 kilometres from a mine.

Important to notice is that some children did not go to school before the cobalt mining boom, for example because they were helping their parents in housework activities or with their farm. Those children might have shifted from not harmful child labor to the cobalt mines (i.e. harmful child labor). The methodology presented does not consider those children. For this reason the results presented above constitute a lower bound since

As a robustness check the methodology additionally excludes individuals who were born beyond 10 kilometres from a cobalt mine and moved closer to a cobalt mine later in their life. In this way, I control for migration patterns which might be due to the boom of cobalt mine in the DRC. This further restriction is done since families in the DRC might migrate because of the greater availability of mining jobs in areas surrounding deposits of cobalt. Results of this restriction are shown in section 2.5 and further confirm the negative impact that cobalt boom had on education achievement of children living in areas around cobalt mines.

1.5.2 Fertility

I now move to examine the results of the baseline equation 1.1 in which the outcome variables are defined as five year fertility and three year fertility rates. This time, the beta coefficients resulting from this specification compare the two measure of fertility of women living in cobalt mining areas with those women, living in non mining areas, before and after the cobalt boom.

Table 4 reports beta coefficients of the effect of the boom of cobalt on women's fertility rates

Individuals who were born in 1992 were 15 years old in 2007, when the cobalt mining boom occurred. Therefore, people born *until* 1992 were considered as pre-boom cohort. On the other hand individuals who were born *since* 1993 were at most 14 years old in 2007, when the cobalt mining boom occurred. These are considered as post-boom birth years cohorts based on equation 1.1. Columns (1) and (4) presents results on 5 year and three year fertility, respectively from a specification with only woman's age fixed effects and no woman-specific controls, sub-regional district time trends, or fixed effects. In these first specifications, I find positive and significant effects of cobalt mining on the two measures of women's fertility, with a coefficient of 0.407 (and a standard error of 0.138) for five year fertility and a coefficient of 0.333 (and a standard error of 0.105 for three year fertility). Columns (2) and (5) and add other controls and fixed effects sequentially. The addition of woman-specific controls such as woman's education, type of residence along with a dummy variable indicating if the woman surveyed has ever migrated, in columns (2) and (5) have small effects on the estimated coefficients, which are now 0.363 (and a standard error of 0.134) and 0.3 (and a standard error of 0.103) on five year and three year fertility rates, respectively.

Finally, specification in columns (3) and (6) include a set of fixed effects, i.e. survey year and district by time fixed effect. Results, partially confirm what shown in columns (1)-(2) and (5)-(6), with a coefficient of 0.325 (and a standard error of 0.198) on five year fertility rate and a coefficient of 0.276 (and a standard error of 0.137) on three year fertility rate.

Taken together, these results imply that the cobalt mining boom during woman's fertile period led to an average of 0.3 more children during the last five years preceding the interview and an average of 0.28 more children during the last three years preceding the interview in areas surrounding a cobalt mine deposit. This means that almost all the increase in fertility experienced in cobalt-mining areas after 2007 started from 2010 (i.e. three years after the onset of the boom in cobalt mining). Little of the total impact of cobalt mining on fertility is due to years five and four preceding the interview. This expected result is explained by the time period families need to adapt to the shock. Results suggest that women's were more likely to have approximately 0.3 children during the five years preceding the date of the interview and after the cobalt mining boom. This translates in an increase of approximately 0.06 children per year per woman.

1.6 Addressing Potential Concerns

In this section, I address potential concerns that arise from the baseline strategy. Finally. I present a series of robustness checks that address the potential e endogenous migration waves in and out treatment group after the boom of cobalt.

1.6.1 Artisanal and Chinese VS European and American Cobalt Mines

Unregulated labor conditions are strictly dependent on the supervision which is made into place by the company that owns the mine. As mentioned earlier about 20% of the total production of cobalt in the DRC comes from artisan-based mines which are not owned by any company and therefore constitute the biggest threat of child labor. Artisan-based mines born due to the sudden boom in the demand of cobalt. While 70% of total production of cobalt in the DRC is owned by chinese companies were active actions and controls to prevent child labor are officially states although no data are provided by those companies. Finally the rest 10% of the total cobalt mined in the DRC comes from mines owned by European, Canadian and Australian companies.

The methodology takes into account the ownership of each mine by interacting the distance from a cobalt mine with its ownership country. This because DRC owned and artisan based cobalt mines might be less compelling with the law and therefore might incentivize illegal child labor activities.

Once again results suggest that education attainments decreased more in areas surrounding artisan based and DRC owned mines compared to cobalt mines owned by countries outside Africa.

Tables 5 and 6 report beta coefficients of the effect of the boom of cobalt on individual's education attainment and women's fertility rates, respectively based on equation 1.1.

Column (1) of Tables (5) and (6) presents results of the effects of living within 10 kilometers away from *any* cobalt mine deposit on the outcome variable of interest from a specification including a set of fixed effects, i.e. survey year, sub-regional district by time fixed effect, individual specific fixed effects such as woman's level of education, age, birth cohort, a indicator variable showing if the individual has ever migrated and type of residence. On the other hand, Column (2) of Tables (5) and (6) presents results of the effects of living within 10 kilometers away from *an artisanal or Chinese owned* cobalt mine deposit on the outcome variable of interest from the same specification used in Column (1) including all controls and fixed effects.

In these specifications, I find a negative and significant effects of cobalt mining on the individual's education attainment, with a coefficient of 0.496 (and a standard error of 0.185) and a coefficient of 0.502 (and a standard error of 0.176) for artisanal cobalt mining effect.

Concerning the effect of cobalt mining and artisanal cobalt mining in fertility rates: I find a positive and significant effects of cobalt mining on the three year women's fertility (Column (3) of Table 6), with a coefficient of 0.276 (and a standard error of 0.137) and a coefficient of 0.295 (and a standard error of 0.131) for artisanal cobalt mining effect.

Taken together, these results imply that the artisanal cobalt mining activities, have similar but slightly greater effects on both education attainment and women's fertility rates compared to all types of cobalt mining. In other words, the results from this sample restriction seem to further confirm that the although the unethical child labor uses in cobalt mines are practices diffuses in all cobalt mines, those which are completed unsupervised such as the artisanal-based ones, are associated with ever lower education attainments of those individuals who during their childhood were exposed to the boom in the production of cobalt. Consequently, fertility rates in areas surrounding artisanal-based cobalt mines seem to be slightly higher compared to any type of cobalt-mining areas.

1.6.2 Cobalt Mines vs Any Mine

In order to clearly identify that the effects on child labor is entirely due to cobalt mines and not being a consequence of any other type of mine, I identify a second set of control group which is constituted by children being far from a cobalt mine but within 10 kilometres to *any other* mine in the DRC.

This robustness check was possible since mines in the DRC are sparsely located as shown in Figure **??**. Comparison from children within 10 kilometres from a cobalt mine and children within 10 kilometres from any other mine leads to similar results of those obtained in the benchmark regression. The cobalt mining boom has caused children in areas surrounding cobalt mines to achieve lower education rates and cognitive skills to decrease with respect to children and households living in areas surrounding any other mine in the DRC.

Tables 7 and 8 report beta coefficients of the effect of the boom of cobalt on individual's education attainment and women's fertility rates, respectively based on baseline equation 1.1 where individuals living in villages within 10 kilometers away from *any* other types of mine now compose the control group.

Column (1) of Tables 7 and 8 presents results of the effects of living within 10 kilometers away from a cobalt mine deposit on the outcome variables of interest from a specification with only birth year fixed effects and no individual-specific controls, sub-regional district time trends, or fixed effects. In this specification, I find a negative and significant effects of cobalt mining on the education attainment of individuals who grew up in cobalt-mining areas compared to those who spent their childhood in general-mining areas.

I obtain a coefficient of -0.452 (and a standard error of 0.173). Columns (2)-(3) add other controls and fixed effects sequentially. The addition of individual-specific controls such as gender, type of residence along with a dummy variable indicating if the individual surveys has ever migrated. In column (2) I have small effects on the estimated coefficient, which is now -0.41 (and a standard error of 0.176). Finally, specification in column (3) I further control for the year of birth of each individual surveyed. Results, confirm what shown in columns (1)-(2), with a coefficient of -0.433 (and a standard error of 0.179).

Taken together, these results imply that the cobalt mining boom during school-going years led to an average of 0.433 fewer years of completed schooling of those individuals born since 1993 (post boom cohorts) and living in areas surrounding a cobalt mine deposit compared to individuals living in other-mining areas. Once again, results show no effect of cobalt mining on future wealth of individuals who grew up within 10 kilometres away from the nearest cobalt mine compared to those who during their childhood lived between 10 kilometres and 100 kilometres from a mine.

Table 8 reports beta coefficients of the effect of the boom of cobalt on women's fertility rates based on equation 1.1 where individuals living in villages within 10 kilometers away from *any* other types of mine now compose the control group. Here, columns (1) and (4) presents results on 5 year and three year fertility, respectively from a specification with only woman's age fixed effects and no woman-specific controls, sub-regional district time trends, or fixed effects. In these first specifications, once again, I find positive and significant (although slightly lower in magnitude compared to those obtained when comparing individuals within 10 kilometers away from a cobalt mine with those living between 10-100 kilometers from a cobalt mine deposit) effects of cobalt mining on the two measures of women's fertility, with a coefficient of 0.389 (and a standard error of 0.145) for five year fertility and a coefficient of 0.315 (and a standard error of 0.105 for three year fertility). Columns (2) and (5) add other controls and fixed effects sequentially. Finally, results presented in columns (3) and (6), partially confirm what shown in columns (1)-(2) and (4)-(5), with a coefficient of 0.433 (and a standard error of 0.171) on five year fertility rate and a coefficient of 0.255 (and a standard error of 0.104) on three year fertility rate.

Overall, these results imply that the cobalt mining boom during woman's fertile period led to an average of 0.4 more children during the last five years preceding the interview and an average of 0.25 more children during the last three years preceding the interview in areas surrounding a cobalt mine deposit compared to women of the same age living within 10 kilometers away from any other mine in the DRC.

1.6.3 Cohort Analysis

A cohort-specific relationship between pre-cobalt boom and children education achievements is also included in the paper. The cohort-specific relationship allows for a visual and clear representation of the effects of cobalt booming. Hence, the following equation is estimated:

$$\text{Outcome}_{i,c,d,k} = \alpha + \sum_k \beta_k \times (\text{Cobalt Mine})_c + \gamma_k \, \mathbf{X}_i^{'} + \sigma_{1,d} + \sigma_{2,dtrend} + \varepsilon_{\,i,c,d,k} \qquad (1.3)$$

In this specification the parameters of interest are β_k which give the results of the cohort-specific relationships between the measure of the distance of each individual surveyed and the nearest cobalt deposit and either education attainment and cognitive skills. Treatment group is unchanged, i.e. those children born since 1993, thus being at most 14years old at the time of cobalt boom in 2007. On the other hand children born until 1992, were older than 14and thus constitute the control group. If the boom of cobalt effectively subtracted children from school, then a break from the pre-existing trend of β_k should start only from children born after 1993. In other words the analysis must not yield any statistically significant relationship between distance from a cobalt mine and education attainment of all children born before 1992. While, for children born since 1993 a negative and statistically significant relationship should appear.

I now examine graphically the relationship between childhood exposure to cobalt mining and completed years of education. The coefficients obtained from equation 1.3 compare the trends in schooling over time in villages with different distance from a cobalt mine deposit. Figure 1.11 plots the coefficients and respective 95 percent confidence intervals. To have a robust estimates, I group birth cohorts in three-year groups. Therefore, for example I group individuals born in 1960, 1961 and 1962 in one cohort and so on until individuals born in 1998 and 1999. I group birth cohorts in order to avoid the limited data sample available. To focus on individuals of relevant birth cohorts (i.e. born since 1993), the last four birth cohorts are two years long .

Figure 1.11 shows no differential trends in education attainment among cohorts who were born between 1960 and 1992 across villages more or less distant from a cobalt mine deposit. Given these cohorts were exposed to relatively stable and similar trends in cobalt production from mining during their childhood, this lack of association provides reassuring evidence that there were no pre-existing differential trends in schooling achievements across areas within and beyond 10 kilometers from a cobalt mine and. For the cobalt boom cohorts, those born since 1993, there is a statistically significant decline in schooling in cobalt mining areas compared to non cobalt mining areas.

1.6.4 Spatial Analysis

To exclude the possibility that some other change might have caused the education attainment of individuals in the control group (i.e. beyond 10 kilometers from a cobalt mine) to increase after 2007, such as any change in the number of schools in the control group, I use a spatial lag model that allows for non-linear effects on education attainment and fertility with distance from the cobalt mine.

Our concern here, is that the effects of cobalt mining boom on education attainment might be biased upward. If no other shock beside the cobalt mining boom which has affected education attainment of people in the control group occurred after 2007 then we would expect the impact of cobalt mining boom to be only limited to those living within 10 kilometres from the nearest cobalt mine deposit, while no effect should be for those people living beyond 10 kilometres. Hence, the following equation is estimated:

The cohort of people born in 1994 actually groups individuals born in 1993 and 1994. Cohort 1996 considers individuals born in 1995 and 1996. Cohort 1998, instead considers those born in 1997 and 1998. Finally, cohort 2000 considers individuals born in 1999 and 2000

$$\text{Outcome}_{i,c,d,t} = \alpha + \sum_{b} \beta_b (\text{Post})_t \times (\text{Cobalt Mine})_c + \sum_{b} \beta_b \times (\text{Cobalt Mine})_c +$$

$$+ \text{"}\gamma\ \mathbf{X}_i' + \delta_t + \sigma_{1,d} + \sigma_{2,dtrend} + \varepsilon_{\ i,c,d,t} \quad (1.4)$$

for $b \in \{0-10, 10-20, ..., 40-50, 50-70\}$.

This spatial lag model allows for non-linear effects with distance from the nearest cobalt mine. Each individual is recorded to a distance bin: 0–10 kilometres, 10–20 kilometres, etc. and compared with the reference category 70–100 kilometres away. The specification controls for the same fixed effects, trends and individual level controls as the baseline specification. The results from this alternative model can be seen in Figures 1.12 and 1.13 for education attainment and fertility, respectively.

In addition, I also consider a further cohort-specific relationship between pre-cobalt boom and children education attainment in a spatial lag model. This combination of a cohort analysis in a spatial lag model allows for a robust comparison between children born in the same year in different locations without assuming linear effects with distance from a cobalt deposit. This further specification test allows for a better understanding of both temporal and geographic distribution of the effects on education attainment. See Figure A.8 in the Appendix.

1.6.5 Selective Migration

As Tables 9 and 10 below show, those individuals living within 10 kilometers from a cobalt mine were actually born in the same village (cluster) where they were surveyed and never migrated. However, we cannot exclude that endogenous migration did not play an important role in explaining the

The adoption of a spatial lag model to show evidence of the spatial diffusion of the effects of mining activities is in spirit very similar to Benshaul-Tolonen (2018) reduction of the education achievement post cobalt boom. Next I control for selective migration.

The analysis also takes into account the possibility that some people migrated to the treatment group to work in a cobalt mine. This, might bias either upward of downward the impact of being close to a cobalt deposit and education attainment. Consider the following cases:

1. Consider the extreme case in which only the poorest people migrated from control group (i.e. $> 10km$ from the nearest cobalt deposit) to the treatment group (i.e. $< 10km$ from the nearest cobalt deposit) after the cobalt mining boom occurred in 2007. This scenario is actually very plausible since those people without a job might see the opening of a mine as an opportunity to work. If this is the case then, since poorest people are often associated with low education levels, this might result in a reduction in education attainment in the treatment group and at the same time in an increase in the control group. The result of this first migration flow would be an overestimation of the negative impact of cobalt mining activities on the education achievements.

2. The opposite direction of migration might also play a crucial role. Consider that only richest individuals migrated from the treatment group to the control group after 2007. Richest people are associated with higher education attainment. If this is the case, then following the same reasoning of above, the education of those individuals who stayed in the treatment group would decrease while those in the control group would have completed more years of schooling. This second migration flow would again overestimate the negative impact of cobalt mining activities on the education achievements.

Combining the two cases we obtain the worst scenario in terms of selective migration from the treatment group and to the treatment group. Therefore, by dropping only those individuals who were surveyed in a village belonging in the control group who migrated and show high values of wealth and at the same time dropping those individuals who were surveyed in a village in the treatment group, who migrated and show low values of wealth we would estimate the lower bound of the relationship between living in a village within 10 kilometers from a cobalt mine and education achievements later in life. In other words, if by considering the above mentioned two selective migration flows we still obtain a significant negative effect of distance from a cobalt mine and education rate, then this relationship would be robust to any selective migration flow in and out the treatment group.

Table 11 reports beta coefficients of the effect of the boom of cobalt on individual's education attainment and wealth based on equation 1.1. This time, I exclude from the sample the poorest 20% of people who migrated *to* the treatment group (i.e. $< 10km$ from the nearest cobalt deposit) from the control group (i.e. between 10 and 100 kilometers away from the nearest cobalt deposit) after the cobalt mining boom occurred in 2007 and only the richest 20% (i.e. with a wealth index equal to 5) of those individuals who migrated *from* the treatment group to the control group after 2007.

Here, column (1) presents results on highest completed education year from a specification with only individual's birth cohort fixed effects and no other individual-specific controls, sub-regional district time trends, or fixed effects. Columns (2) adds other controls and fixed effects sequentially. Finally, column (3) shows results from the specifications considering all individual-specific control variables, district linear trends and survey year fixed effects. Results from these specifications show that the patterns observed for individuals education attainment later in life when excluding selective migrants are similar to those in the full sample. With a coefficient of -0.443 (and a standard error of

The wealth index computed by the DHS Program has values ranging from 1 (representing the poorest quintile of the population) to 5 (the highest quintile of the population). Excluding the poorest 20% of the individuals who migrated means that I exclude those individuals with a wealth index equal to 1 out of 5.

I also run a further specification in which I exclude the poorest 40% (i.e. with wealth index ≤ 2) of people who migrated *to* the treatment group (i.e. $< 10km$ from the nearest cobalt deposit) from the control group (i.e. between 10 and 100 kilometers away from the nearest cobalt deposit) after the cobalt mining boom occurred in 2007 and only the richest 40% ((i.e. with wealth index ≥ 4)) of those individuals who migrated *from* the treatment group to the control group after 2007. Results are shown in the Appendix and confirm the similar pattern observed in the baseline specification

0.149) for the restricted sample, compared to the beta coefficient of -0.496 (and a standard error of 0.185) for the full sample.

Table 12 instead, reports beta coefficients of the effect of the boom of cobalt on women's fertility rates based on equation 1.1 where I consider the extreme scenario in which only the poorest people migrated *to* the treatment group (i.e. $< 10km$ from the nearest cobalt deposit) from the control group (i.e. between 10 and 100 kilometers away from the nearest cobalt deposit) after the cobalt mining boom occurred in 2007 and only richest individuals migrated *from* the treatment group to the control group after 2007.

Here, columns (1) and (4) present results on five year and three year fertility, respectively from a specification with only woman's age fixed effects and no woman-specific controls, sub-regional district time trends, or fixed effects. Columns (2) and (5) add other controls and fixed effects sequentially. Finally, columns (3) and (6) show results from the specifications considering all woman-specific control variables, district linear trends and survey year fixed effects. Once again, I find positive and significant effects of cobalt mining on the two measures of women's fertility (although slightly lower in magnitude compared to the baseline specification, considering the full sample of women), with a coefficient of 0.397 (and a standard error of 0.188) for five year fertility and a coefficient of 0.301 (and a standard error of 0.136 for three year fertility). Thus, we conclude that the patterns observed for women's fertility rates when excluding selective migrants are similar to those in the full sample.

1.7 Conclusions

This paper provides evidence that exposure to cobalt mining activities during childhood, in the DRC subtracts children from school, thus leading to lower education attainment later in life. As a result of the increasing use of child labor due to cobalt mining, I show that the cost-opportunity of having an additional child for families living in cobalt-mining areas decreases, hence pushing fertility rates upwards. Moreover, I show that dropping out from school to work as cobalt miners does not lead to significantly different wealth conditions later in life, in a context of low socio-economic development with a highly inefficient education system such as that of the Democratic Republic of Congo.

I contribute to the literature by showing that geographic conditions, naturally richer of critical minerals, and technological advancements leading to modern lithium-ion electric batteries can generate, through the increase in child labor practices, higher fertility rates providing an explanation for the persistence of low education attainment of individuals living in cobalt mining areas

I then provide evidence that these effects are generally greater in artisanal based cobalt mining areas, where there is a substantial lack of labor supervision. Moreover, the results indicate that the use of children is strictly associated to cobalt mines rather than any type of mine. In a cobalt mine, children who are not sent underground to search for cobalt, wash the tiny cobalt matters which is mixed with dust. To perform this relatively not dangerous job small hands are needed (Amnesty International, 2017).

The results shows that the effects are concentrated within 10 kilometres from a cobalt mine deposit. The results further suggest that in the long-term, parental fertility choices adapt as consequence of child labor.

Though the socio-economic and natural suitability of minerals of the Democratic Republic of Congo is unique in some ways, there are also other examples of artisanal small-scale cobalt mining activities which might lead to similar consequences for children and the economic growth of the local communities. For example, children are heavily involved in cobalt mining activities in Zambia, and in some cases even in South Africa.

Reports indicate the use of child labor also outside the DRC are provided from various news agencies such as Sky News, BBC and CBS News

The second half of the present study provides evidence that more efficient labor supervision might mitigate the effects of the exposure to cobalt mining activities on both education and fertility choices. Moreover, I argue that the root reason pushing parents to opt for sending their children to work in a cobalt mine rather than going to schools lies on the relative low future returns of schooling in the DRC.

Overall, this paper provides a first step at understanding how the production of cobalt from mining might effects parental fertility decisions through the use of children to perform particular labor practices, motivating not the mere implementation of child labor related policies but also investing more on the education system in a context such as that of the DRC in order to address the root causes of child labor.

References

Acemoglu, D. and Johnson, S. (2007). Disease and development: the effect of life expectancy on economic growth. *Journal of political Economy*, 115(6):925–985.

Allen, S., Badiane, O., Sene, L., and Ulimwengu, J. (2014). Government expenditures, health outcomes and marginal productivity of agricultural inputs: The case of tanzania. *Journal of agricultural economics*, 65(3):637–662.

Alsan, M. (2015). The effect of the tsetse fly on african development. *American Economic Review*, 105(1):382–410.

Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of political economy*, 113(1):151–184.

Alves Dias, P., Blagoeva, D., Pavel, C., and Arvanitidis, N. (2018). Cobalt: demand-supply balances in the transition to electric mobility. *European Commission, Joint Research Centre, EUR-Scientific and Technical Research Reports. Publications Office of the European Union. DOI*, 10:97710.

Amnesty International (2016). This is What We Die For. Human Rights Abuses in the Democratic Republic of the Congo Power the Global Trade in Cobalt. [illegible]62/3183/2016/[illegible]. [Online; accessed 9-July-2019].

Amnesty International (2017). *Time to Recharge: Corporate Action and Inaction to Tackle Abuses in the Cobalt Supply Chain* . Amnesty International.

Angrist, J. D. and Kugler, A. D. (2008). Rural windfall or a new resource curse? coca, income, and civil conflict in colombia. *The Review of Economics and Statistics*, 90(2):191–215.

Aragón, F. M. and Rud, J. P. (2015). Polluting industries and agricultural productivity: Evidence from mining in ghana. *The Economic Journal*, 126(597):1980–2011.

Ashraf, Q. H., Lester, A., and Weil, D. N. (2008). When does improving health raise gdp? *NBER macroeconomics annual*, 23(1):157–204.

Atkin, D. (2016). Endogenous skill acquisition and export manufacturing in mexico. *American Economic Review*, 106(8):2046–85.

Banza, C. L. N., Nawrot, T. S., Haufroid, V., Decrée, S., De Putter, T., Smolders, E., Kabyla, B. I., Luboya, O. N., Ilunga, A. N., Mutombo, A. M., et al. (2009). High human exposure to cobalt and other metals in katanga, a mining area of the democratic republic of congo. *Environmental research*, 109(6):745–752.

Basu, K. and Van, P. H. (1998). The economics of child labor. *American economic review*, pages 412–427.

Beegle, K., Dehejia, R., and Gatti, R. (2004). Why should we care about child labor? the education, labor market, and health consequences of child labor. Technical report, National Bureau of Economic Research.

Benshaul-Tolonen, A. (2018). Local industrial shocks and infant mortality. *The Economic Journal*, 129(620):1561–1592.

Berman, N., Couttenier, M., Rohner, D., and Thoenig, M. (2017). This mine is mine! how minerals fuel conflicts in africa. *American Economic Review*, 107(6):1564–1610.

Bleakley, H. (2003). Disease and development: Evidence from the american south. *Journal of the European Economic Association*, 1(2-3):376–386.

Bleakley, H. (2007). Disease and development: evidence from hookworm eradication in the american south. *The quarterly journal of economics*, 122(1):73–117.

Bleakley, H. (2009). Economic effects of childhood exposure to tropical disease. *American Economic Review*, 99(2):218–23.

Bleakley, H. (2010a). Health, human capital, and development. *Annu. Rev. Econ.*, 2(1):283–310.

Bleakley, H. (2010b). Malaria eradication in the americas: A retrospective analysis of childhood exposure. *American Economic Journal: Applied Economics*, 2(2):1–45.

Bleakley, H., Costa, D., and Lleras-Muney, A. (2014). Health, education, and income in the united states, 1820–2000. In *Human Capital in History: The American Record*, pages 121–159. University of Chicago Press.

Bleakley, H. and Hong, S. C. (2017). Adapting to the weather: lessons from us history. *The journal of economic history*, 77(3):756–795.

Bowden, S., Michailidou, D. M., and Pereira, A. (2008). Chasing mosquitoes: An exploration of the relationship between economic growth, poverty and the elimination of malaria in southern europe in the 20th century. *Journal of International Development: The Journal of the Development Studies Association*, 20(8):1080–1106.

Brown, G. (1986). Control and eradication of malaria: Past, present and future.

Bustos, P., Caprettini, B., and Ponticelli, J. (2016). Agricultural productivity and structural transformation: Evidence from brazil. *American Economic Review*, 106(6):1320–65.

Caminade, C., Kovats, S., Rocklov, J., Tompkins, A. M., Morse, A. P., Colón-González, F. J., Stenlund, H., Martens, P., and Lloyd, S. J. (2014). Impact of climate change on global malaria distribution. *Proceedings of the National Academy of Sciences*, 111(9):3286–3291.

Carrillo, B. (2019). Present bias and underinvestment in education? long-run effects of childhood exposure to booms in colombia. *Long-Run Effects of Childhood Exposure to Booms in Colombia (March 16, 2019)*.

Center for International Forestry Research (2013). The formalisation of artisanal mining in the democratic republic of the congo and rwanda. Technical report, Center for International Forestry Research.

Charles, K. K., Hurst, E., and Notowidigdo, M. J. (2018). Housing booms and busts, labor market opportunities, and college attendance. *American Economic Review*, 108(10):2947–2994.

Cobalt Institute (2019). Data on Production and Supply of Cobalt. [illegible] [Online; accessed 19-July-2019].

Conley, D., McCord, G. C., and Sachs, J. D. (2007). *Africa's lagging demographic transition: Evidence from exogenous impacts of malaria ecology and agricultural technology*.

Cust, J. (2015). The spatial effects of resource extraction: mining in indonesia. Technical report, Oxcarre Discussion Paper, University of Oxford.

Cutler, D., Fung, W., Kremer, M., Singhal, M., and Vogl, T. (2010). Early-life malaria exposure and adult outcomes: Evidence from malaria eradication in india. *American Economic Journal: Applied Economics*, 2(2):72–94.

Cuvelier, J. (2017). Money, migration and masculinity among artisanal miners in katanga (dr congo). *Review of African Political Economy*, 44(152):204–219.

Darton Commodities (2017). Annual market review on cobalt demand. [illegible]. Accessed: 2018-11-12.

De Herdt, T. and Titeca, K. (2016). Governance with empty pockets: the education sector in the democratic republic of congo. *Development and Change*, 47(3):472–494.

De Putter, T., Decrée, S., Banza, C. L. N., and Nemery, B. (2011). Mining the katanga (drc) copperbelt: geological aspects and impacts on public health and the environment—towards a holistic approach. In *Mining and the Environment in Africa. Proceedings of the Inaugural Workshop, IGCP/SIDA*, number 594, pages 14–17.

Deaton, A. (2013). *The great escape: health, wealth, and the origins of inequality*. Princeton University Press.

Demographic and Health Surveys (2014). The dhs program, usaid. *Calverton: Measure DHS*.

Derryberry, O. and Gartrell, F. (1952). Trends in malaria control program of the tennessee valley authority1. *The American journal of tropical medicine and hygiene*, 1(3):500–507.

Doepke, M. and Zilibotti, F. (2005). The macroeconomics of child labor regulation. *American Economic Review*, 95(5):1492–1524.

Dube, O. and Vargas, J. F. (2013). Commodity price shocks and civil conflict: Evidence from colombia. *The Review of Economic Studies*, 80(4):1384–1421.

Easterly, W. and Levine, R. (2016). The european origins of economic development. *Journal of Economic Growth*, 21(3):225–257.

Eckstein, Z. and Wolpin, K. I. (1999). Why youths drop out of high school: The impact of preferences, opportunities, and abilities. *Econometrica*, 67(6):1295–1339.

Edmonds, E. V. and Pavcnik, N. (2005). Child labor in the global economy. *Journal of Economic Perspectives*, 19(1):199–220.

EPA (1975). Ddt: A review of scientific and economic aspects of the decision to ban its use as a pesticide.

Euractiv (2019). Europe will have at least ten gigafactories. [illegible]. Accessed: 2019-04-05.

Faber, B., Krause, B., and De La Sierra, R. S. (2017). Artisanal mining, livelihoods, and child labor in the cobalt supply chain of the democratic republic of congo.

Financial Times (2019). Congo, child labour and your electric car. Technical report, Financial Times.

Fischer, G., Nachtergaele, F. O., Prieler, S., Teixeira, E., Toth, G., van Velthuizen, H. T., Verelst, L., and Wiberg, D. (2012). Global agro-ecological zones (gaez v3. 0).

for Disease Control Prevention, U. C. (2010). The history of malaria, an ancient disease.

Gallup, J. L. and Sachs, J. D. (2001). The economic burden of malaria. *The American journal of tropical medicine and hygiene*, 64(1_suppl):85–96.

Galor, O. and Weil, D. N. (2000). Population, technology, and growth: From malthusian stagnation to the demographic transition and beyond. *American economic review*, 90(4):806–828.

Ghosh Banerjee, S., Romo, Z., McMahon, G., Toledano, P., Robinson, P., and Pérez Arroyo, I. (2014). *The power of the mine: A transformative opportunity for sub-Saharan Africa*. The World Bank.

Giuliano, P. and Nunn, N. (2013). The transmission of democracy: From the village to the nation-state. *The American Economic Review*, 103(3):86.

Gollin, D., Lagakos, D., and Waugh, M. E. (2014). Agricultural productivity differences across countries. *American Economic Review*, 104(5):165–70.

Gooch, E. (2017). The impact of reduced incidence of malaria and other mosquito-borne diseases on global population. *Journal of Development Economics*, 124:214–228.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017). Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202:18–27.

Haines, M. R. et al. (2005). *Historical, demographic, economic, and social data: The United States, 1790-2002*. Inter-university Consortium for Political and Social Research.

Hansen, C. W. and Lønstrup, L. (2015). The rise in life expectancy and economic growth in the 20th century. *The Economic Journal*, 125(584):838–852.

Hausmann, R. and Rigobon, R. (2003). An alternative interpretation of the'resource curse': Theory and policy implications. Technical report, National Bureau of Economic Research.

Hay, S. I., Guerra, C. A., Tatem, A. J., Noor, A. M., and Snow, R. W. (2004). The global distribution and population at risk of malaria: past, present, and future. *The Lancet infectious diseases*, 4(6):327–336.

Hazan, M. and Berdugo, B. (2002). Child labour, fertility, and economic growth. *The Economic Journal*, 112(482):810–828.

Henderson, J. V., Squires, T., Storeygard, A., and Weil, D. (2017). The global distribution of economic activity: nature, history, and the role of trade. *The Quarterly Journal of Economics*, 133(1):357–406.

Hijmans, R., Garcia, N., and Wieczorek, J. (2010). Gadm database of global administrative areas. *V ersion*.

Hilson, G. (2011). Artisanal mining, smallholder farming and livelihood diversification in rural sub-saharan africa: An introduction. *Journal of international development*, 23(8):1031–1041.

Humphreys, M. (2001). *Malaria: poverty, race, and public health in the United States*. JHU Press.

International Labor Organization (2015). *World report on child labour 2015: paving the way to decent work for young people*. International Labour Organization.

Kiszewski, A., Mellinger, A., Spielman, A., Malaney, P., Sachs, S. E., and Sachs, J. (2004). A global index representing the stability of malaria transmission. *The American journal of tropical medicine and hygiene*, 70(5):486–498.

Klein Goldewijk, K., Beusen, A., Van Drecht, G., and De Vos, M. (2011). The hyde 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. *Global Ecology and Biogeography*, 20(1):73–86.

Lysenko, A. and Semashko, I. (1968). Geography of malaria. a medico-geographic profile of an ancient disease. *Itogi Nauki: Medicinskaja Geografija*, pages 25–146.

Majori, G. (2012). Short history of malaria and its eradication in italy with short notes on the fight against the infection in the mediterranean basin. *Mediterranean journal of hematology and infectious diseases*, 4(1).

McCarthy, D., Wolf, H., and Wu, Y. (2000). The growth costs of malaria. Technical report, National Bureau of Economic Research.

Medlock, J. M. and Leach, S. A. (2015). Effect of climate change on vector-borne disease risk in the uk. *The Lancet Infectious Diseases*, 15(6):721–730.

Michalopoulos, S. and Papaioannou, E. (2013). Pre-colonial ethnic institutions and contemporary african development. *Econometrica*, 81(1):113–152.

Millimet, D. L. (2011). The elephant in the corner: a cautionary tale about measurement error in treatment effects models. In *Missing Data Methods: Cross-Sectional Methods and Applications*, pages 1–39. Emerald Group Publishing Limited.

Newberry. Atlas of historical county borders. [illegible] Accessed: 2018-09-30.

Nkulu, C. B. L., Casas, L., Haufroid, V., De Putter, T., Saenen, N. D., Kayembe-Kitenge, T., Obadia, P. M., Mukoma, D. K. W., Ilunga, J.-M. L., Nawrot, T. S., et al. (2018). Sustainability of artisanal mining of cobalt in dr congo. *Nature sustainability*, 1(9):495.

Nunn, N. and Puga, D. (2012). Ruggedness: The blessing of bad geography in africa. *Review of Economics and Statistics*, 94(1):20–36.

Nunn, N. and Qian, N. (2011). The potato's contribution to population and urbanization: evidence from a historical experiment. *The Quarterly Journal of Economics*, 126(2):593–650.

Olivetti, E. A., Ceder, G., Gaustad, G. G., and Fu, X. (2017). Lithium-ion battery supply chain considerations: analysis of potential bottlenecks in critical metals. *Joule*, 1(2):229–243.

Percoco, M. (2013). The fight against disease: Malaria and economic development in italian regions. *Economic Geography*, 89(2):105–125.

Picard, J. and Mills, A. (1992). The effect of malaria on work time: analysis of data from two nepali districts. *The Journal of tropical medicine and hygiene*, 95(6):382–389.

Ramankutty, N., Foley, J. A., Norman, J., and McSweeney, K. (2002). The global distribution of cultivable lands: current patterns and sensitivity to possible climate change. *Global Ecology and biogeography*, 11(5):377–392.

Ross, R. (1897). On some peculiar pigmented cells found in two mosquitos fed on malarial blood. *British medical journal*, 2(1929):1786.

Ruggles, S., Genadek, K., Goeken, R., Grover, J., and Sobek, M. (2015). Integrated public use microdata series: Version 6.0 [dataset]. minneapolis: University of minnesota.

Sachs, J. and Malaney, P. (2002). The economic and social burden of malaria. *Nature*, 415(6872):680.

Sachs, J. D. and Warner, A. M. (1995). Natural resource abundance and economic growth. Technical report, National Bureau of Economic Research.

Saul, A. (2002). Malaria: Poverty, race, and public health in the united states.

SP Global Market Intelligence (2018). SP Global Market Intelligence Annual Report. [illegible] 2018/2018-[illegible] [Online; accessed 19-July-2018].

Spencer, E. (2016). Cobalt supply and demand—a global perspective. In *Proceedings from The Cobalt Conference, Seoul, Korea*.

Subramanian, M. A. and Sala-i Martin, X. (2003). *Addressing the natural resource curse: An illustration from Nigeria*. Number 3-139. International Monetary Fund.

Sviatschi, M. M. (2019). Making a narco: Childhood exposure to illegal labor markets and criminal life paths.

Tatem, A. J., Gething, P. W., Smith, D. L., and Hay, S. I. (2013). Urbanization and the global malaria recession. *Malaria journal*, 12(1):133.

Thomas, D., Beegle, K., Frankenberg, E., Sikoki, B., Strauss, J., and Teruel, G. (2004). Education in a crisis. *Journal of Development economics*, 74(1):53–85.

Tsui, K. K. (2010). More oil, less democracy: Evidence from worldwide crude oil discoveries. *The Economic Journal*, 121(551):89–115.

Turcheniuk, K., Bondarev, D., Singhal, V., and Yushin, G. (2018). Ten years left to redesign lithium-ion batteries.

US Geological Survey (2019). Mineral commodity summaries 2019. Technical report, United States Geological Survey.

Weil, D. N. (2014). The impact of malaria on african development over the longue durée. *Africa's Development in Historical Perspective*, pages 89–130.

Williams, R. C. (1952). The united states public health service, 1798–1950. *AJN The American Journal of Nursing*, 52(10):1279.

Wouterse, F. and Badiane, O. The role of health, experience, and educational attainment in agricultural production: Evidence from smallholders in burkina faso. *Agricultural Economics*, 0(0).

Xia, F. and Deininger, K. (2019). Spillover effects of tobacco farms on the labor supply, education, and health of children: Evidence from malawi. *American Journal of Agricultural Economics*.

1.8 Figures and Tables

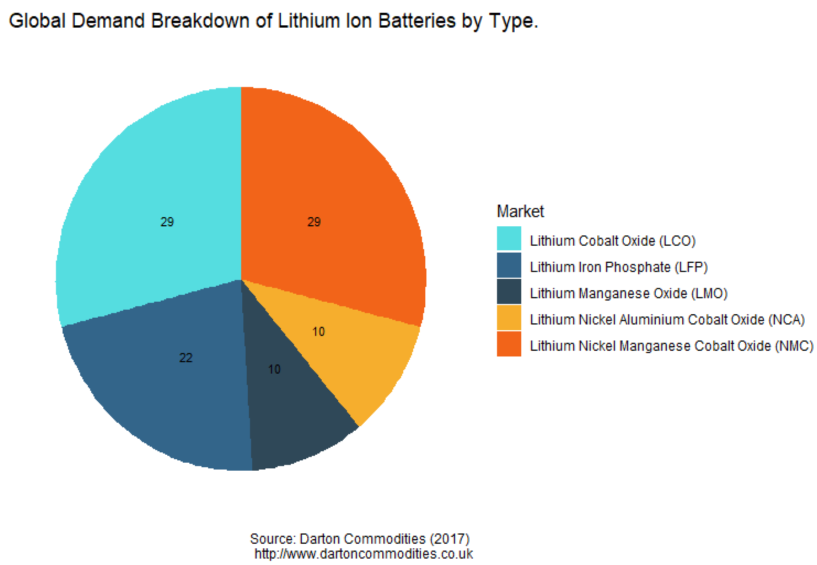


Figure 1.1: Global Demand Breakdown of Lithium-ion Batteries by Type

Notes: The data in this figure is retrieved from Darton Commodities (2017).

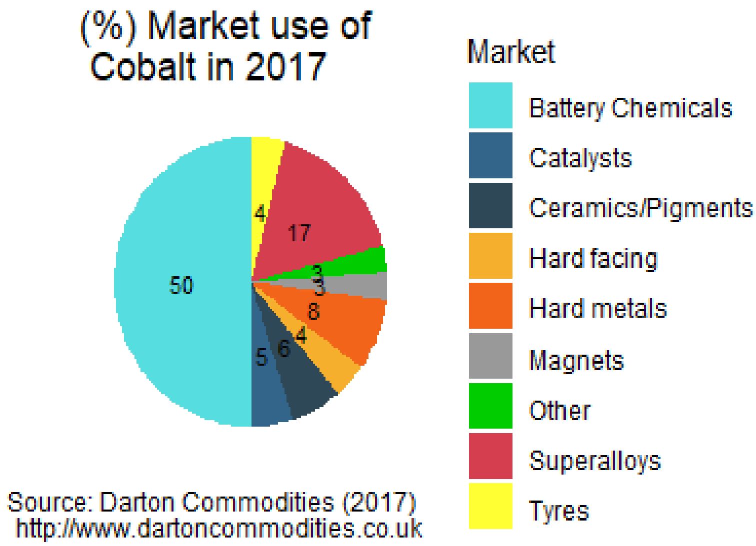


Figure 1.2: Market Use of Cobalt in Percentage

Notes: The data in this figure is retrieved from Darton Commodities (2017) and Alves Dias et al. (2018).

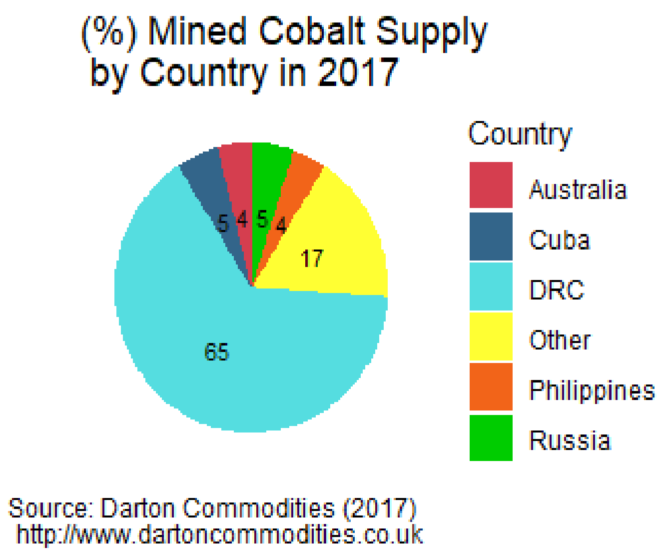


Figure 1.3: Mined Cobalt Supply by Country in Percentage

Notes: The data in this figure is retrieved from Darton Commodities (2017) and Alves Dias et al. (2018).

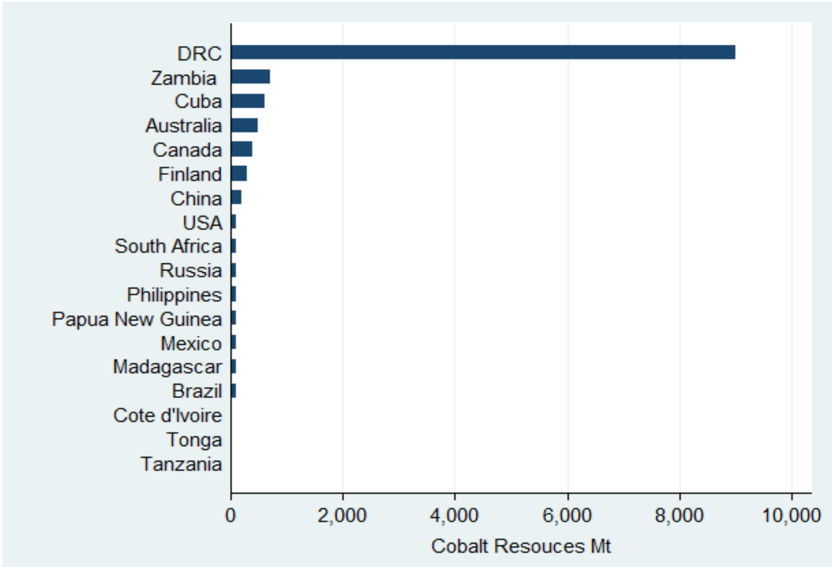


Figure 1.4: Estimated Cobalt Reserves by Country in 2018. Million Tons

Notes: The data in this figure is retrieved from Alves Dias et al. (2018) and SP Global Market Intelligence (2018).

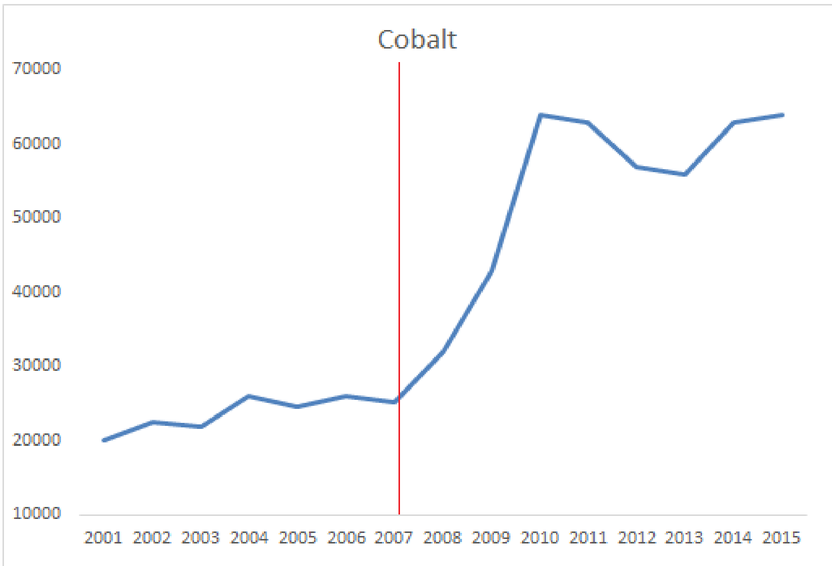


Figure 1.5: Total Cobalt Production from Mining in DRC. Metric Tons

Notes: The data in this figure is retrieved from US Geological Survey (2019) and Cobalt Institute (2019).

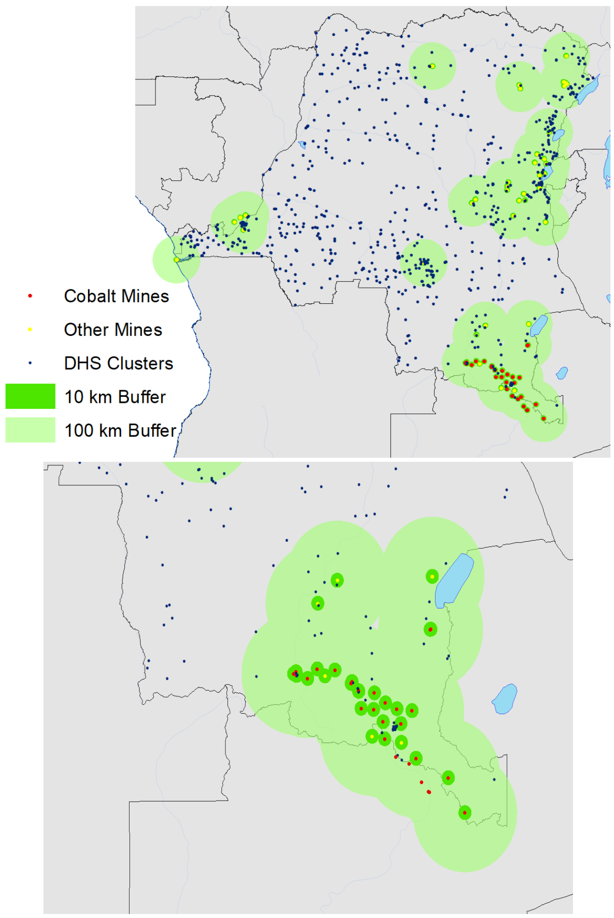


Figure 1.6: Location of all Mining Deposits in the DRC and gps of all Individuals Surveyed in the 2007 and 2014 DHS waves

Notes: The data in this figure is a combination of author's calculations, using a GIS software and US Geological Survey (2019) and Demographic and Health Surveys (2014).

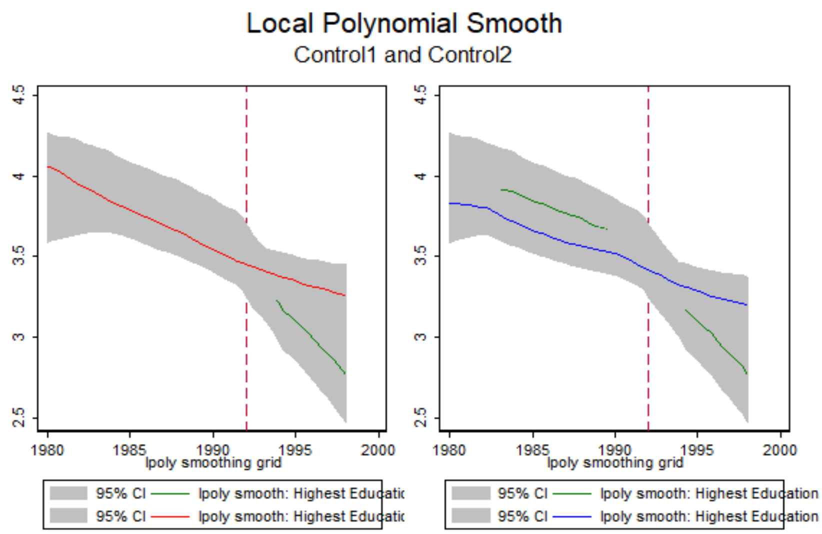


Figure 1.7: Local Polynomial Smooth Function

Notes: Birth cohorts and education attainment are shown on the horizontal and vertical axis, respectively. Birth cohorts range from 1960 to 1999. The treatment group is defined as individuals, at the time of the survey, living within 10 kilometres away from the closest active cobalt mine deposit. The first control group is composed of individuals, at the time of the survey, living between 10 and 100 kilometres away from a cobalt mine (red line, left graph). The second control group considers individuals, at the time of the survey, living within 10 kilometres away from any closest mine except cobalt (blue line, right graph). The Figure provides 95% confidence intervals. No control variables, fixed effects or sub-regional district linear trends were considered.

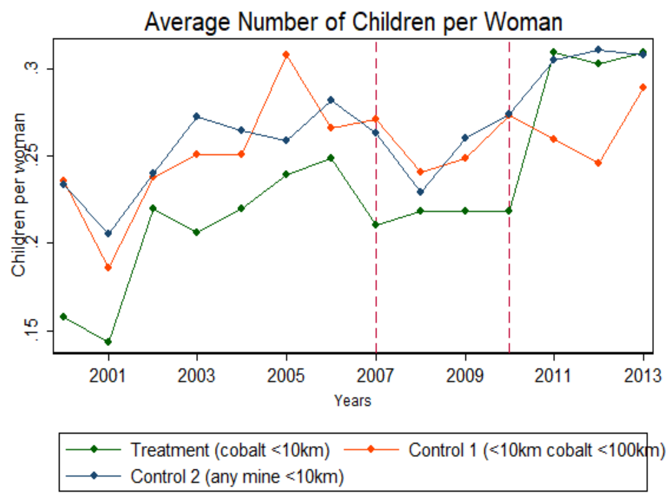


Figure 1.8: Average Number of Children born per Woman at a Given Year

Notes: Children birth years and average number of children per woman are shown on the horizontal and vertical axis, respectively. Children birth years range from 2000 to 2013. The treatment group is defined as women surveyed in 2014, who at the time of the survey, were living within 10 kilometres away from the closest active cobalt mine deposit. The first control group is composed of women surveyed in 2007, who at the time of the survey, were living between 10 and 100 kilometres away from a cobalt mine (red line). The second control group considers women surveyed in 2014, who at the time of the survey, were living within 10 kilometres away from any closest mine except cobalt (blue line). No control variables, fixed effects or sub-regional district linear trends were considered.

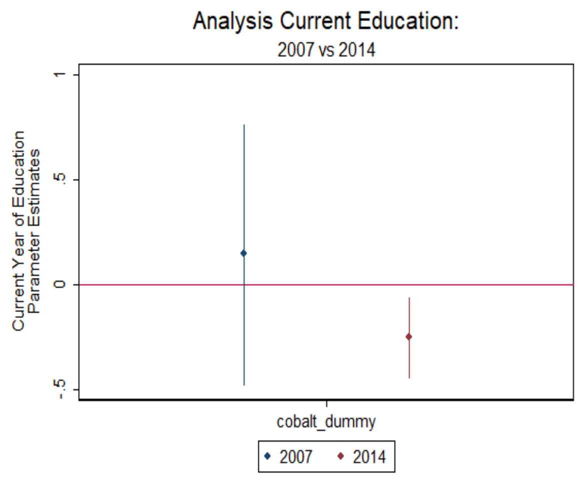


Figure 1.9: Pre-Cobalt Boom and Children's Education Attainment

Notes: This Figure shows the relationship between living within 10 km from a cobalt deposit and children's current year of education, for all individuals surveyed pre cobalt boom, in 2007 (left) and for all individuals surveyed post cobalt boom, in 2014 (right) using the baseline set of control variables and 95% confidence intervals. The sample is all individuals who at time of the DHS surveys were between 6 and 14. Regressions also included all individual specific controls, birth year and sub-regional district fixed effects

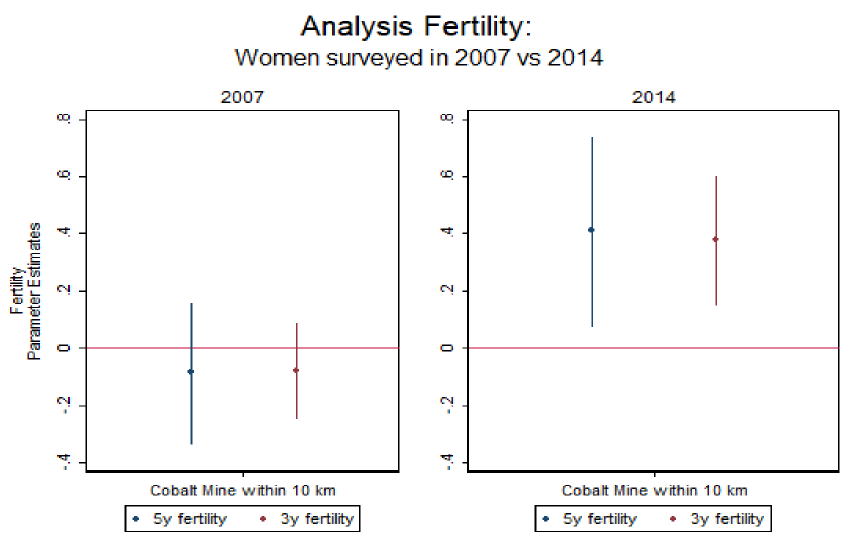


Figure 1.10: Pre-Cobalt Boom and Women's Fertility Rates

Notes: This Figure shows the relationship between living within 10 km from a cobalt deposit and women's fertility rates, for all women surveyed pre cobalt boom, in 2007 (left) and for all women surveyed post cobalt boom, in 2014 (right) using the baseline set of control variables and 95% confidence intervals. The sample is all women who at time of the DHS surveys were between 15 and 39. Regressions also included all individual specific controls and sub-regional district fixed effects

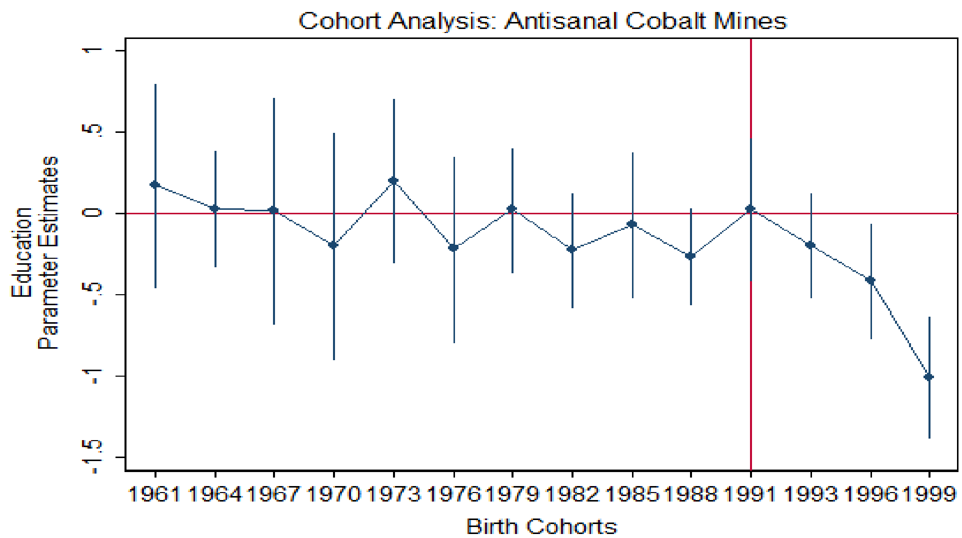


Figure 1.11: Pre-Cobalt Boom and Completed Years of Education; Three-Year Cohort-Specific relationships for all Individuals Born between 1960 and 1999

Notes: This figure reports estimated birth year (birth cohort) fixed effects in completed years of education for all individuals born between 1960 and 1999 using the baseline set of control variables and 95% confidence intervals. The sample is all individual born between 1960 and 1999, pooling DHS datasets of 2007 and 2014. Regressions also included all individual specific controls, survey year and sub-regional district by time fixed effects. To focus on individuals of relevant ages (< 14 y.o. at the time of the cobalt mining boom), the last four birth cohorts are two years long.

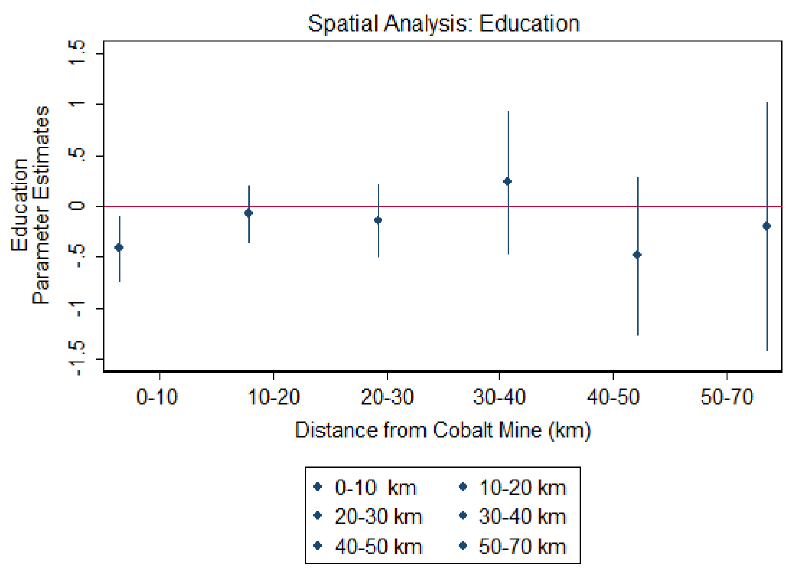


Figure 1.12: Non Linear Distance from Cobalt Mine: Education Attainment

Notes: This Figure shows the results from a spatial lag model with 10 kilometres distance bins using the baseline set of control variables and 95% confidence intervals. The sample is all individual born between 1960 and 1999, pooling DHS datasets of 2007 and 2014. Regressions also included all individual specific controls, survey year and sub-regional district by time fixed effects

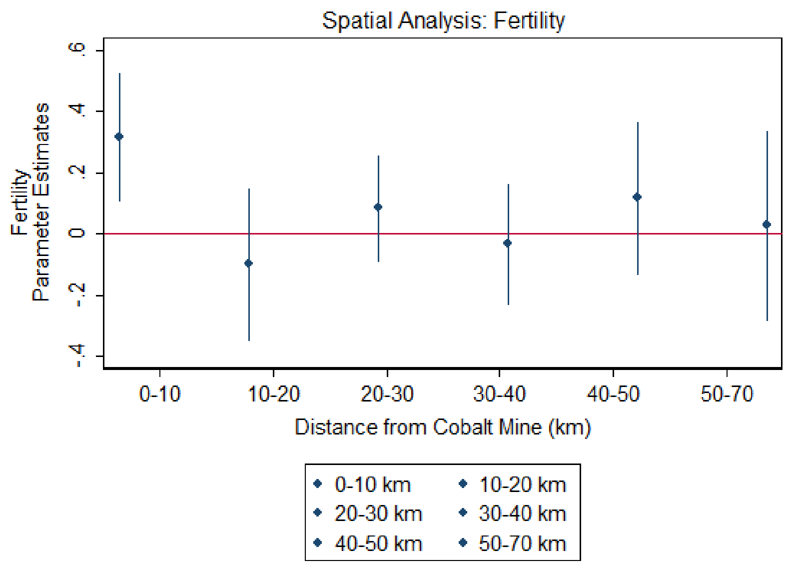


Figure 1.13: Non Linear Distance from Cobalt Mine: Women's Fertility Rate

Notes: This Figure shows the results from a spatial lag model with 10 kilometres distance bins using the baseline set of control variables and 95% confidence intervals. The sample is all women who at time of the DHS surveys were between 15 and 39, pooling DHS datasets of 2007 and 2014. Regressions also included all women specific controls, survey year and sub-regional district by time fixed effects

Table 1.1: Extensive Descriptive Statistics - adults born from 1960 to 1999

	Pre 1993 < 10 km Mean	Pre 1993 > 10 km Mean	Pre 1993 any mine Mean	Post 1993 < 10 km Mean	Post 1993 > 10 km Mean	Post 1993 any mine Mean
Education						
Highest Year of Education	4.00	3.82	3.71	3.01	3.26	3.21
Controls						
Ever Migrated	0.99	0.98	0.98	0.99	0.99	0.99
Female	1.47	1.51	1.54	1.53	1.49	1.48
Urban	1.06	1.33	1.30	1.16	1.29	1.28
Current Student	1.63	1.01	0.81	5.91	4.92	5.07
Age	28.23	29.32	28.62	18.01	17.94	18.01
Wealth						
Wealth Index	4.72	4.11	3.99	4.52	4.27	4.24
Health						
Hospitalized	0.10	0.08	0.11	0.10	0.06	0.07
Received Treatment	0.12	0.11	0.13	0.14	0.05	0.07
BMI	23.15	23.28	23.09	21.83	21.70	21.67
Observations	846	625	1180	303	290	488

Table 1.2: Extensive Descriptive Statistics - Women between 15 and 39

	Pre < 10 km Mean	Pre > 10 km Mean	Pre any mine Mean	Post < 10 km Mean	Post > 10 km Mean	Post any mine Mean
Fertility						
Births in the last 5 years	0.77	0.98	1.04	1.02	1.10	0.94
Births in the last 3 years	0.45	0.63	0.63	0.67	0.68	0.60
Births in the past year	0.14	0.23	0.24	0.24	0.27	0.22
Wealth						
Wealth Index	4.78	3.60	3.86	4.58	3.85	4.20
Controls						
Woman's Education	3.66	3.41	3.55	3.37	3.51	3.46
Urban	1.00	1.37	1.32	1.13	1.42	1.27
Age	28.73	27.93	28.24	26.77	27.83	27.59
Observations	121	401	371	180	689	589

Table 1.3: Childhood Cobalt Mining Exposure and Education Attainment: Benchmark Results

	Education Coef./SE	Education Coef./SE	Education Coef./SE	Wealth Coef./SE	Wealth Coef./SE	Wealth Coef./SE
Post x Cobalt Deposit	-0.569**	-0.505**	-0.496**	0.033	-0.070	-0.065
	(0.226)	(0.197)	(0.185)	(0.305)	(0.130)	(0.126)
Cobalt Mine within 10 km	0.186	0.087	0.071	0.556**	0.086	0.097
	(0.161)	(0.128)	(0.124)	(0.259)	(0.092)	(0.100)
Female	No	Yes	Yes	No	Yes	Yes
Urban	No	Yes	Yes	No	Yes	Yes
Ever migrated	No	Yes	Yes	No	Yes	Yes
Current Student	No	Yes	Yes	No	Yes	Yes
Survey Year FE	No	No	Yes	No	No	Yes
District x Time FE	No	No	Yes	No	No	Yes
Year of Birth FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1828	1822	1822	1904	1898	1898

Table 1.4: Cobalt Mining Exposure and Fertility Rate: Benchmark Results

	5y Fertility Coef./SE	5y Fertility Coef./SE	5y Fertility Coef./SE	3y Fertility Coef./SE	3y Fertility Coef./SE	3y Fertility Coef./SE
Post x Cobalt Deposit	0.407***	0.363**	0.325	0.333***	0.300***	0.276*
	(0.138)	(0.134)	(0.198)	(0.105)	(0.103)	(0.137)
Cobalt Mine within 10 km	-0.287*	-0.182	-0.074	-0.186*	-0.101	-0.025
	(0.151)	(0.149)	(0.139)	(0.099)	(0.097)	(0.080)
Woman's Age FE	Yes	Yes	Yes	Yes	Yes	Yes
Survey Year FE	No	No	Yes	No	No	Yes
District x Time FE	No	No	Yes	No	No	Yes
Urban	No	Yes	Yes	No	Yes	Yes
Woman's Education	No	Yes	Yes	No	Yes	Yes
Ever Migrated	No	Yes	Yes	No	Yes	Yes
Observations	844	790	790	844	790	790

Table 1.5: Childhood Cobalt Mining Exposure and Education Attainment: Artisanal Mining

	Education Coef./SE	Education Coef./SE	Wealth Coef./SE	Wealth Coef./SE
Post x Cobalt Deposit	-0.496**		-0.066	
	(0.185)		(0.127)	
Post x Art. Cobalt Deposit		-0.502***		-0.009
		(0.176)		(0.117)
Female	Yes	Yes	No	No
Urban	Yes	Yes	Yes	Yes
Ever migrated	Yes	Yes	Yes	Yes
Current Student	Yes	Yes	Yes	Yes
Survey Year FE	Yes	Yes	Yes	Yes
District x Time FE	Yes	Yes	Yes	Yes
Year of Birth FE	Yes	Yes	Yes	Yes
Observations	1822	1822	1898	1898

Table 1.6: Cobalt Mining Exposure and Fertility Rate: Artsanal Mining

	5y Fertility Coef./SE	5y Fertility Coef./SE	3y Fertility Coef./SE	3y Fertility Coef./SE
Post x Cobalt Deposit	0.325		0.276*	
	(0.198)		(0.137)	
Post x Art. Cobalt Deposit		0.380**		0.295**
		(0.184)		(0.131)
Woman's Age FE	Yes	Yes	Yes	Yes
District x Time FE	Yes	Yes	Yes	Yes
Survey Year FE	Yes	Yes	Yes	Yes
Woman's Education	Yes	Yes	Yes	Yes
Urban	Yes	Yes	Yes	Yes
Ever Migrated	Yes	Yes	Yes	Yes
Observations	790	790	790	790

Table 1.7: Childhood Cobalt Mining Exposure and Education Attainment: Cobalt Mines vs All Mines

	Education Coef./SE	Education Coef./SE	Education Coef./SE	Wealth Coef./SE	Wealth Coef./SE	Wealth Coef./SE
Post x Cobalt Deposit	-0.465**	-0.444**	-0.481***	-0.400*	-0.277*	-0.255*
	(0.174)	(0.172)	(0.176)	(0.233)	(0.143)	(0.150)
Cobalt Mine within 10 km	0.163	0.141	0.231	0.789***	0.395***	0.326
	(0.122)	(0.118)	(0.144)	(0.201)	(0.103)	(0.229)
Female	No	Yes	Yes	No	Yes	Yes
Urban	No	Yes	Yes	No	Yes	Yes
Ever migrated	No	Yes	Yes	No	Yes	Yes
Current Student	No	Yes	Yes	No	Yes	Yes
Survey Year FE	No	No	Yes	No	No	Yes
District x Time FE	No	No	Yes	No	No	Yes
Year of Birth FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2016	2000	2000	2149	2133	2133

Table 1.8: Cobalt Mining Exposure and Fertility Rate: Benchmark Results. Cobalt Mines vs Any Mine

	5y Fertility Coef./SE	5y Fertility Coef./SE	5y Fertility Coef./SE	3y Fertility Coef./SE	3y Fertility Coef./SE	3y Fertility Coef./SE
Post x Cobalt Deposit	0.389**	0.337**	0.433**	0.315***	0.290**	0.255**
	(0.145)	(0.145)	(0.171)	(0.109)	(0.110)	(0.104)
Cobalt Mine within 10 km	-0.292**	-0.227	0.785***	-0.217**	-0.175*	0.463***
	(0.140)	(0.143)	(0.153)	(0.088)	(0.090)	(0.108)
Woman's Age FE	Yes	Yes	Yes	Yes	Yes	Yes
Survey Year FE	No	No	Yes	No	No	Yes
District x Time FE	No	No	Yes	No	No	Yes
Urban	No	Yes	Yes	No	Yes	Yes
Woman's Education	No	Yes	Yes	No	Yes	Yes
Ever Migrated	No	Yes	Yes	No	Yes	Yes
Observations	1013	927	927	1013	927	927

Table 1.9: Childhood Cobalt Mining Exposure and Migration

	Migration Coef./SE	Migration Coef./SE	Migration Coef./SE
Post x Cobalt Deposit	-0.011	-0.011	-0.010
	(0.008)	(0.009)	(0.009)
Cobalt Mine within 10 km	0.009	0.010	0.012
	(0.010)	(0.011)	(0.010)
Gender	No	Yes	Yes
Type of Residence	No	Yes	Yes
Current Student	No	Yes	Yes
Survey Year FE	No	No	Yes
District Linear Trend	No	No	Yes
Birth Cohort FE	Yes	Yes	Yes
Observations	1796	1790	1790

Table 1.10: Cobalt Mining Exposure and Selective Women Migration

	Ever Migrated Coef./SE	Ever Migrated Coef./SE	Ever Migrated Coef./SE
Post x Cobalt Deposit	0.015	0.016	-0.011
	(0.010)	(0.011)	(0.016)
Cobalt Mine within 10 km	-0.012*	-0.014	0.010
	(0.007)	(0.009)	(0.007)
Woman's Age FE	Yes	Yes	Yes
Survey Year FE	No	No	Yes
District x Time FE	No	No	Yes
Urban	No	Yes	Yes
Woman's Education	No	Yes	Yes
Observations	844	790	790

Table 1.11: Childhood Cobalt Mining Exposure and Education Attainment: Selective Migration

	Education	Education	Education	Wealth	Wealth	Wealth
	Coef./SE	Coef./SE	Coef./SE	Coef./SE	Coef./SE	Coef./SE
Post x Art. Cobalt Deposit.	-0.343**	-0.444***	-0.443***	0.227	-0.081	-0.079
	(0.167)	(0.153)	(0.149)	(0.303)	(0.102)	(0.101)
Cobalt Mine within 10 km	0.114	0.044	0.028	0.515*	0.073	0.085
	(0.132)	(0.107)	(0.107)	(0.275)	(0.084)	(0.094)
Gender	No	Yes	Yes	No	Yes	Yes
Type of Residence	No	Yes	Yes	No	Yes	Yes
Migrant	No	Yes	Yes	No	Yes	Yes
Current Student	No	Yes	Yes	No	Yes	Yes
Survey Year FE	No	No	Yes	No	No	Yes
District Linear Trend	No	No	Yes	No	No	Yes
Birth Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1705	1699	1699	1776	1770	1770

Table 1.12: Cobalt Mining Exposure and Fertility Rate: Selective Migration

	5y Fertility Coef./SE	5y Fertility Coef./SE	5y Fertility Coef./SE	3y Fertility Coef./SE	3y Fertility Coef./SE	3y Fertility Coef./SE
Post x Art. Cobalt Deposit	0.274*	0.399***	0.397**	0.224*	0.323***	0.301**
	(0.159)	(0.121)	(0.188)	(0.127)	(0.097)	(0.136)
Cobalt Mine within 10 km	-0.191	-0.183	-0.090	-0.105	-0.097	-0.022
	(0.173)	(0.136)	(0.126)	(0.124)	(0.088)	(0.078)
Woman's Age FE	Yes	Yes	Yes	Yes	Yes	Yes
Survey Year FE	No	No	Yes	No	No	Yes
District x Time FE	No	No	Yes	No	No	Yes
Urban	No	Yes	Yes	No	Yes	Yes
Woman's Education	No	Yes	Yes	No	Yes	Yes
Observations	810	756	756	810	756	756

Table 1.13: Childhood Cobalt Mining Exposure and Education Attainment: Placebo Test on Zambia

	Education Coef./SE	Education Coef./SE	Education Coef./SE	Wealth Coef./SE	Wealth Coef./SE	Wealth Coef./SE
Post x Cobalt Deposit	0.024	0.079	0.074	0.089	-0.008	0.003
	(0.182)	(0.175)	(0.175)	(0.182)	(0.123)	(0.117)
Cobalt Mine within 10 km	0.148	0.223**	0.180*	0.619***	0.014	-0.057
	(0.106)	(0.104)	(0.103)	(0.171)	(0.125)	(0.127)
Female	No	Yes	Yes	No	Yes	Yes
Urban	No	Yes	Yes	No	Yes	Yes
Ever migrated	No	Yes	Yes	No	Yes	Yes
Current Student	No	Yes	Yes	No	Yes	Yes
Survey Year FE	No	No	Yes	No	No	Yes
District x Time FE	No	No	Yes	No	No	Yes
Year of Birth FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4515	4451	4451	4663	4598	4598

Table 1.14: Childhood Mining Exposure and Education Attainment: Placebo Treatment

	Education Coef./SE	Education Coef./SE	Education Coef./SE	Wealth Coef./SE	Wealth Coef./SE	Wealth Coef./SE
Post x Mine Deposit	-0.133	-0.137	-0.048	0.230	0.144	0.016
	(0.116)	(0.105)	(0.101)	(0.180)	(0.155)	(0.139)
Mine within 10 km	-0.078	-0.165**	-0.082	0.394**	0.120	0.710***
	(0.087)	(0.081)	(0.075)	(0.189)	(0.115)	(0.122)
Female	No	Yes	Yes	No	Yes	Yes
Urban	No	Yes	Yes	No	Yes	Yes
Ever migrated	No	Yes	Yes	No	Yes	Yes
Current Student	No	Yes	Yes	No	Yes	Yes
Survey Year FE	No	No	Yes	No	No	Yes
District x Time FE	No	No	Yes	No	No	Yes
Year of Birth FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	15931	15800	15800	17893	17757	17757

Table 1.15: Childhood Cobalt Mining Exposure and Secondary Education: Placebo Test

	Education Coef./SE	Education Coef./SE	Education Coef./SE
Post x Cobalt Deposit	0.118	0.035	0.039
	(0.144)	(0.030)	(0.031)
Cobalt Mine within 10 km	0.022	-0.021**	-0.019**
	(0.071)	(0.008)	(0.009)
Female	No	Yes	Yes
Urban	No	Yes	Yes
Ever migrated	No	Yes	Yes
Current Student	No	Yes	Yes
Survey Year FE	No	No	Yes
District x Time FE	No	No	Yes
Year of Birth FE	Yes	Yes	Yes
Observations	1904	1898	1898

2

Climate, Disease & Development: Malaria Control and Historical Agricultural Productivity in the US

This study provides an estimation of the causal relationship between reduction in malaria transmission and agricultural productivity in the U.S. by exploiting exogenous geographic variations in the stability of malaria and using historical disaggregated county data for the U.S. together with a robust quasi-experimental approach. The two principal channels through which malaria control policies could affect agriculture i.e. conversion of wetland to arable land and higher farmer productivity due to better health conditions are also investigated. Results show that while the eradication of malaria led to approximately one fifth of the agricultural productivity growth in the U.S. counties, the positive effect was entirely due to better health conditions, since wetland conversion had little or no effect on the amount of arable land in highly endemic counties. Robustness checks from geographic variations in malaria stability within neighboring counties along with placebo treatments reinforce the positive effect of malaria eradication on historical growth rates of agricultural productivity.

Understanding the deep causes of divergent economic development has always been fascinating for economists and historians. One important role in explaining comparative socio-economic development is constituted by health conditions, the transmission of diseases and in particular malaria

(Deaton, 2013) and Bleakley et al. (2014). Micro empirical studies have examined disease prevention and malaria eradication on human capital finding a positive relationship at the individual level (Bleakley, 2003, 2009, Cutler et al., 2010, Percoco, 2013). However, aggregate effects of disease eradication on the economy are still debated (Acemoglu and Johnson, 2007, Ashraf et al., 2008) and more recently Hansen and Lønstrup (2015) and Gooch (2017).

Nevertheless, what has yet to be examined empirically is the direct effect of the eradication of malaria on historical agricultural productivity along with the identification of the main channels through which vector-borne disease control might impact on socio-economic development. In this respect, the principal contribution of this study is to provide a rigorous quantitative analysis of the historical impact of the eradication of malaria on the agricultural productivity growth of each US county and to investigate the underlying mechanisms. Figure 2.1 below shows the average annual agricultural productivity* growth rates per US county for the period 1870 to 1900 on the left and for the period 1900 to 1920 on the right †. In Figure 2.2 below the respective correlation coefficients are displayed. The graph on the left indicates a marked negative correlation between agricultural productivity growth prior to 1900 and malaria prevalence‡. The correlation between agricultural productivity growth and malaria prevalence appears to dramatically switch sign after 1900 as shown on the left hand side of Figure 2.2. In particular, correlation between agricultural productivity growth between 1870 and 1900 and the index of malaria prevalence§ was -0.36. This suggests that counties where malaria was more prevalent experienced lower agricultural productivity growth prior to 1900. However, the sign of the correlation between agricultural productivity growth and

*County agricultural productivity is defined as the total farm output value per farmer for each county. More details on data and variables are provided in Section 2.2.2

†The map shown in Figure 2.1 was created using the geographic information system ArcGis software. Green areas have experienced a greater increase in farm productivity growth while areas in red have shown a decrease in farm productivity over time

‡1870 is the first year for which farm productivity data per US county is available in Haines et al. (2005)

§Detailed description of the indices used to measure the prevalence and transmission of malaria disease is provided in Section 2.2.1

the prevalence of malaria appears to reverse to +0.17 after 1900 as shown in the right hand side of Figure 2.2. Specifically the average annual agricultural productivity growth between 1900 and 1920 was higher in counties where malaria was historically more prevalent. Thus, suggesting that consequently to the active fight against malaria thanks to the discovery of new drugs and chemical components such as the wide use of quinine at each state hospital and spraying of larvicides (Williams, 1952, Humphreys, 2001) counties which where historically more affected by malaria experienced larger growth rates in agricultural productivity compared to US areas less affected by malaria. Together, Figures 2.1 and 2.2 suggests that something happened in 1900 that has caused agricultural productivity to grow more in areas historically more affected by malaria compared to areas in which malaria was less prevalent. Potentially the eradication of malaria, which happened between 1900s and 1920s, could possibly be a determinant of such increase in agricultural productivity occurred in more "malarious" US counties.

To this regard it is important to emphasize that the purpose of the following study is not to identify the determinants of agricultural productivity in the US. The assessment of all agricultural and policy determinants of agricultural productivity goes beyond the scope of the following empirical analysis. Rather, the goal of this paper is to quantify the impact that a climate-related vector-borne disease such as malaria had on historical agricultural productivity growth in the US. This is achieved by comparing agricultural productivity levels of highly malarious counties with those of less malarious counties before and after the eradication of malaria in the US which was achieved as a result of the understanding that malaria was transmitted by the bite of specific species of mosquitoes (Ross, 1897) and in turn of newly discovered drugs, such as quinine first and chemical components such as the DDT later on (Williams, 1952, Humphreys, 2001).

Establishing a clear direction in the relationship between malaria and agricultural productivity growth is crucial to understand the causes of historical economic development both between countries and within nations. An important part of the economic literature has examined the rela-

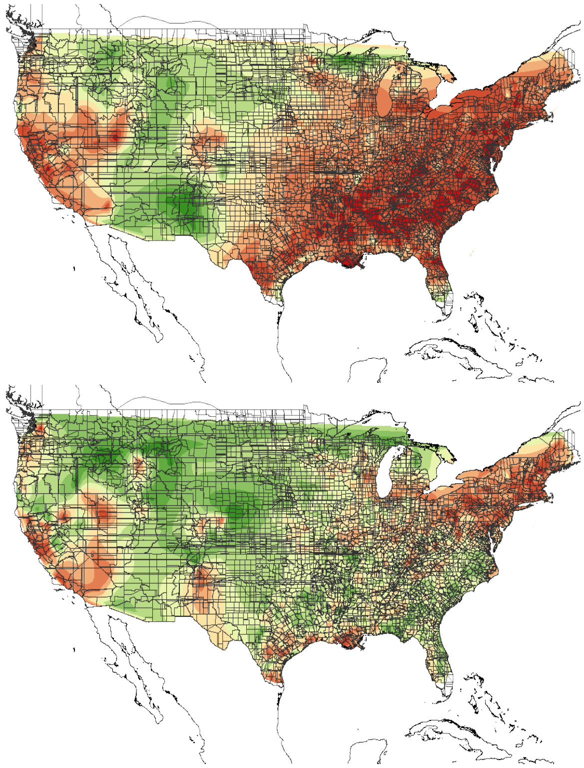


Figure 2.1: Average annual agricultural productivity growth rates per US county from 1870 to 1900 (left) and from 1900 to 1920 (right)

Notes: Source: Author's calculation using the geographical information system software ArcGis based on two historical records: Historical, Demographic and Social Data. The United States, 1790-2002" (Haines et al., 2005) and the IPUMS dataset (Ruggles et al., 2015)

tionship between agricultural productivity and economic growth. In particular Gollin et al. (2014) examined world agricultural productivity rates using new disaggregated data concluding that understanding agricultural productivity is at the heart of understanding world income inequality. Similar conclusions were reached in Bustos et al. (2016) who studied the effects of the adoption of new agricultural technologies on structural transformation in Brazil, finding that exogenous higher agricultural productivity shock caused by technological change led to industrial growth. Therefore, assessing the historical effects of malaria on agricultural productivity within the US could also be crucial to understand differences in the economic development of different US areas.

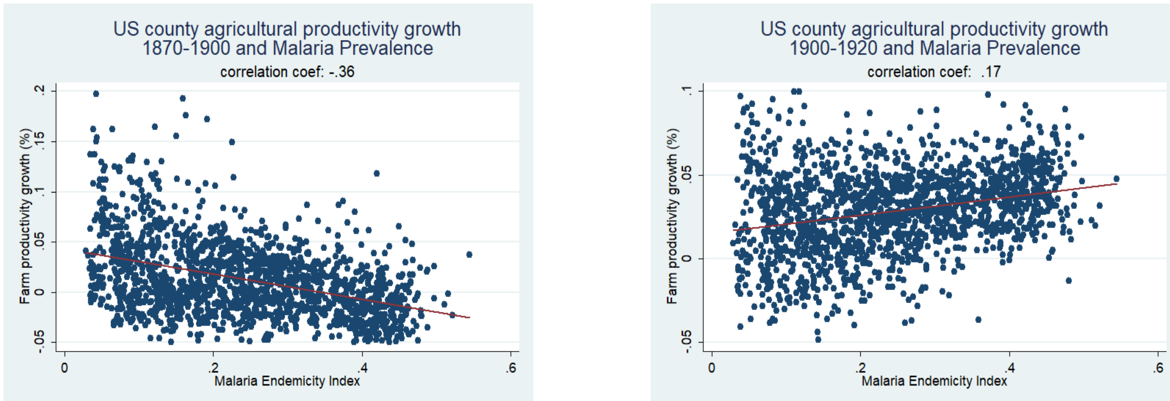


Figure 2.2: Correlation between average annual agricultural productivity growth rates per US county and malaria prevalence in the US

Notes: Figure on the left shows agricultural productivity growth rates from 1870 to 1900, while figure on the right shows agricultural productivity growth rates from 1900 to 1920. Source: Author's calculation using the geographical information system software ArcGis based on two historical records: Historical, Demographic and Social Data. The United States, 1790-2002" (Haines et al., 2005) and the IPUMS dataset (Ruggles et al., 2015)

Another reason of the importance of estimating the relationship between malaria eradication and agricultural productivity is that it is also crucial to predict how economies which nowadays rely predominantly on agriculture (i.e. Sub-Saharan African, Latin American and South-East Asian countries) will be affected from government policies aimed at eradicating vector borne diseases. In addition to this, future climatic conditions will likely be more favorable to the transmission of malaria (Caminade et al., 2014, Medlock and Leach, 2015). Thus, not only the already high infection rates of countries affected by the disease today are predicted to worsen in the future but countries which had never been interested by malaria will be affected by it in the decades.

The empirical analysis relies on a treatment/control strategy using a difference-in-difference (DID) estimation comparing agricultural productivity levels between US counties that had climatic conditions more suitable for the transmission of malaria to counties that were less suitable for a stable transmission of the disease. This approach is permitted since sufficient time has passed that we can evaluate the long-term consequences of the eradication. Furthermore, the United States are well-suited for this analysis for two reasons: first, the US show a remarkable degree of heterogeneity in weather variables which in turn allows for a sufficient variability in the malaria stability values; and second, reliable disaggregated data (i.e. at county level) on historical agricultural productivity variables are available for a long time.

The effects of malaria and its eradication during the 20^{th} century has been the object of several economic studies. Specifically, Bleakley (2003, 2009), Cutler et al. (2010) and Percoco (2013) have shown how improved health (due to vaccination and disease prevention programs) affects GDP indirectly through educational gains (better children's health translates to higher education levels and in turn to greater future income). These gains at individual level will directly translate to gain in GDP per capita unless the increase in population (also due to lower mortality) outweighs these productivity and educational gains. On the other hand, Acemoglu and Johnson (2007), Ashraf et al. (2008) and more recently Hansen and Lønstrup (2015) have shown that the decreased mortality rate due to scientific advancements has positively impacted population while negligible (if any) effects were found on GDP per capita. With particular focus on malaria, Gooch (2017) has shown the positive effect of its eradication campaigns on population and population density during the 20^{th} century. Thus, while micro studies on individual level have shown a positive impact of the eradication of malaria aggregate effects on the economy are still debated.

Malaria is defined as an intermittent and remittent fever transmitted by mosquitoes to humans through their bites. The World Health Organization estimates that in 2015, there were an estimated 211 million cases of malaria. Among them 450,000 people lost their lives because of malaria. The Institute of Health Metrics and Evaluation (IHME), Global Burden of Disease (GBD) revises these estimates upwards to 720,000 casualties directly related to malaria in 2015. More recent WHO reports on malaria show an upward trend in the number of cases of malaria, i.e. in 2017, there were an estimated 216 million cases of malaria, with an increment of about 5 million cases over 2015. The principal reason of such increase in the number of people affected is the more favorable climate for the reproduction and development of mosquitoes larvae. This is indeed one of the main reasons why the study of the socio-economic impacts of malaria is crucial. Malaria is in fact strictly correlated to hot and humid weather conditions which in turn are predicted to be more frequent in the next future.

One mechanism through which malaria transmission is tackled is land use conversion. In the U.S. drainage of swamps and wetland is one of the oldest and commonest forms of land modification undertaken to improve health conditions and lower the transmission of vector borne diseases such as yellow fever and malaria. For these reasons surface water removal was a predominant public policy objective in United States during the 20^{th} century. This conversion of unused wetland into arable and more productive land might have increased the agricultural output of endemic areas and in turn agricultural productivity. I test this hypothesis. A second channels through which the transmission of malaria might impact on agriculture productivity is linked to the poor health conditions of farmers. The whole harvest indeed will be negatively affected if a farmer caches malaria during the harvesting. It is therefore reasonable to conclude that more endemic lands were likely to be less productive than non endemic ones. The adopted strategy allows to show whether this is the case. Hence, the primary goal of this study is to clearly assess the impact that eradication of malaria might have had on agriculture productivity in terms of possible greater amount of arable land and in terms of increased labour productivity due to better health conditions of workers, in particular farmers.

The identification strategy compares highly malarious counties to less endemic counties within the US. It is, therefore crucial that the identification of malaria endemic counties is exogenous (i.e. it does not depend upon socio-economic conditions which might affect the transmission of the latter). For this reason, I use a spatial time-invariant malaria ecology index created by Kiszewski et al. (2004) and based upon climatic conditions which are more or less suitable for the reproduction of two particular species of mosquitoes, namely: Plasmodium falciparum and Plasmodium vivax. These two species of mosquitoes are the natural vehicle of the most severe category of malaria, while the asymptomatic, uncomplicated categories of malaria are not considered in this study. Figure 2.1 shows how regions such as Sub-Saharan Africa, Latin America and South East Asia present the ideal climatic conditions for the reproduction of species of mosquitoes transmitting malaria to human beings. However, historically malaria was also prevalent in many other regions of the world. Notable examples of regions in which malaria was widespread are: the Western US, Southern Europe and Northern Australia. Figure 2.1 is at 0.5°x 0.5°gridded level which corresponds approximately to 56km x 56km at the equator. To get the average malaria stability value for each US county spatial data on malaria is then intersected with historical US county borders[¶]. This procedure will generate a time invariant malaria stability index for each county based upon climatic conditions only[‖].

In addition to the Malaria Stability Index (MSI) this study uses a second measure of endemicity of malaria i.e. the Malaria Endemicity Index (MEI) developed by Hay et al. (2004) who have digitized an old map produced by Lysenko and Semashko (1968) showing historical malaria geography and prevalence. Differently from the MSI, the MEI captures actual distribution of malaria in 1900, just before the onset of vector control as shown in Figure 2.3. The main difference between the MSI and the MEI is that while the former is obtained using exclusively weather conditions more or less suitable for a stable transmission of malaria throughout the year, the latter relies on actual historical presence of malaria and therefore capture the proportion of people affected by the disease in different areas.

The exogenous variations in suitability of weather conditions for the transmission of malaria allow for the comparison between highly malaria suitable and less malaria suitable counties within the U.S. to clearly determine the effects of the eradication campaigns on agricultural productivity.

[¶] Spatial data on historical US county borders are retrieved from the Newberry of historical county borders of the Newberry Library. The historical US county borders contains publicly available spatial data on counties from 1629 to 2000

[‖] Since some counties had an area smaller than 0.5°x 0.5°I first use the natural neighbor algorithm to interpolate the malaria stability values. The interpolation method allows to lower the scale to 0.1°x 0.1°corresponding to an area approximately equal to 5 km x 5 km at the equator.

Results obtained with the MSI are then compared with those obtained with the MEI index. Population, weather and agricultural controls are included in the regressions with the inclusion of county fixed effects and state specific linear trends. Furthermore, a series of robustness checks from geographic variations in malaria stability index *within neighboring counties* and placebo treatments are performed to prove the validity of the estimated coefficients.

Results reveal that: first, the eradication campaigns which in the U.S. took place between the 1900s and the 1940s (with the administration of quinine and drainage of wetlands first and the development of new effective drugs and chemical components later) are estimated to have had substantial effects on the historical agricultural productivity growth in the US. In particular, a 0.1 increase in the Malaria Stability Index (MSI) is associated with a 20 percentage point increase in total county agricultural productivity defined as the county farm output value per farmer[**]. Second, vector-borne control policies had no substantial effects on the amount of arable land of more endemic counties. Since the two main channels through which malaria control policies might affect agricultural output are the increased cropland and better health conditions of farmers, this additional result leads us to conclude that the increase in agricultural productivity of highly malaria suitable counties was entirely due to the increased labor productivity of farmers resulting from better health conditions after the elimination of malaria.

A potential threat to the validity of the empirical analysis comes from the possible endogeneity of the treatments to malaria to more endemic counties with respect to counties less suitable for the transmission of malaria. In other words more endemic counties could have pushed to receive the treatment before non endemic ones. First, the study does rely on exogenous malaria suitability index rather than data on infection rates. Second, Bleakley (2007), Cutler et al. (2010) and more recently

[**] Although the stability index developed by Kiszewski et al. (2004) ranges from 0 for malaria free areas to 39 for areas with weather conditions extremely suitable for a persistent reproduction of the Plasmodium falciparum and Plasmodium vivax species, values in the U.S. ranges mainly from 0 to 1 with values greater than 0.06 being defined as slightly malarious

Gooch (2017) show that the reduction in the burden of malaria clearly resulted from critical scientific innovations coming overwhelmingly from outside the highly endemic counties and which have culminated with better understanding of the origins of the disease and the discover of new drugs and chemical components[††]. This mitigates the usual concern about policy endogeneity. Furthermore, in the regressions I include state specific linear trends to control for changes in agricultural productivity resulting from state specific policies. Finally, and most importantly, the empirical strategy compares also neighboring counties belonging to the same state (and therefore subject to the same federal policies) which had a great degree of heterogeneity in malaria stability values[‡‡]. However, the evidence presented from the neighboring counties analysis reinforce the positive effect of malaria eradication on agricultural productivity.

The main limitation of the results is constituted by the fact that many endogenous factors affecting farm productivity cannot be considered in this study. Examples of such endogenous factors are government policies and investments. The use of potential additional control variables is limited by the lack of such historical variables or plausible instruments. That said however, the hypothetical inclusion of such control variables would absorb part of the effects of malaria treatment to agricultural productivity because government policies, for instance, are investments in adapting to circumstances, including the effects of malaria and other vector borne diseases. For this reason we would not want to remove such adaptation from the estimates, because they are part of the story. The neighboring counties analysis partially alleviates this issue since only counties belonging to the same state, being neighbors and showing substantially different stability values are considered.

The rest of the paper is organized as follows: Section 2.1 provides a brief review of the microeconomic and macroeconomic studies on the relationship between disease and economic development,

[††]The discovery of quinine followed by the improvements in its usage and the widespread availability of the DDT are examples of scientific progresses which led to the substantial reduction in malaria transmission

[‡‡]I specifically compare those counties highly suitable for malaria with neighboring counties which present weather conditions not suitable for the transmission of the disease.

discussing also about possible issues. Section 2.2 briefly presents the data which I use in the analysis, while Section 3.3 presents the empirical strategy. Section 2.4 presents the main historical results, possible threats to the validity of the estimates along with the workaround strategies, robustness checks and placebo tests. Finally, Section 2.6 concludes.

2.1 Historical Background and Related Literature

Malaria has always been an important factor of the economic development. As Figure 2.3 shows until early 1900s south western U.S. and Europe were largely affected by the disease. It was not until the 20^{th} century when technological advancements and the discovery of new drugs along with modern chemical components made possible to effectively prevent the transmission of malaria.

The first major discovery was quinine in the late 17^{th} century when it was imported first in Spain as a curative plant. Its beneficial effects were not fully understood for almost two centuries. In 1897 Ronald Ross succeeded in demonstrating the life-cycle of the parasites of malaria in mosquitoes, thus establishing that malaria came from the bite of mosquitoes, for this discovery he was awarded the nobel prize in medicine. Consequently the properties of quinine were fully appreciated and the plant started being used in hospitals both in Europe and the USA as an effective drug in curing the fever caused by malaria (Majori, 2012). The first interventions to contrast the diffusion of malaria were adopted during early 1900s which consisted mostly in the administration of quinine to people living in areas with high infection rates along with the drainage of swamps and wetlands. An example of the effectiveness of early control policies was the construction of the Panama Canal in 1910s. Yellow fever and malaria were a major cause of death and illness among workers in the area. The Center for Disease Control Prevention (2010) shows that in 1906, there were over 26,000 employ-

Due to its bitter taste people were reluctant to consider it as a real measure against malaria

Italy started an active fight against malaria in 1900 when the law ensuring the national production and sale of the antimalarial quinine, the so called "Chinino di Stato", was promulgated (Majori, 2012).

ees working on the Canal. Of these, over 21,000 (i.e. more than 80%) were hospitalized for malaria at some time during their work. By 1912, there were over 50,000 employees, and the number of hospitalized workers had decreased to approximately 5,600 (i.e. 11%).

In the USA quinine was made available at every state hospital in the early 1900s thanks to donations of the Rockefeller Foundation and the newly established United States Public Health Services (USPHS) along with new methods such as the spraying of larvicides and wetland conversion. By 1912 all board of health of each state followed the USPHS model (Williams, 1952, Humphreys, 2001, Saul, 2002). Thus, until early 1900s scientific advancements were not adequate to treat malaria. This caused more than 80% of the countries in the world to be affected by vector-borne diseases. Scientific progress which brought to the availability of modern drugs along with better sanitation and socio-economic conditions led to a shrink in the burden of malaria. In particular, the U.S. and southern Europe which were largely affected by vector borne diseases successfully eradicated the latter during the first half of the 20^{th} century. During the whole 20^{th} century to the present day the burden of malaria was constantly reduced. Bleakley (2010b) shows how during the twenty year period of early interventions against malaria (i.e. from 1900s to 1920s): mortality rates of malaria declined by more than 70%. However, this is only part of the story, since morbidity rates rather than mortality fully represents the true impact of malaria into the economy. After the first positive results of the fight against malaria until the 1920s, a small resurgence of the latter followed during the period ranging from the 1930s to the 1940s caused by the great depression and the second world war. Finally, the use of DDT and newly discovered chemical components caused a reduction of the burden of malaria to negligible levels during the early 1940s (Bleakley, 2003).

Digitizing historical maps Hay et al. (2004) have quantified the anthropogenic impact on the

---

Williams (1952) presents a thorough history of the US Public Health Service. Humphreys (2001) summarizes the history of malaria-control efforts in the United States. The annual reports of the Rockefeller Foundation's International Health Board (1919) provide information about its anti-malaria demonstration projects. Much of the historical detail for the United States is drawn from these sources distribution of malaria in the 20^{th} century at six intervals between 1900 and 2002 (i.e. 1900, 1946, 1965, 1975, 1994, 2002) with 1900 being the last year before the first adoption of eradication policies. These changes in the global distribution of malaria since 1900 are shown in Figure 2.3 below:

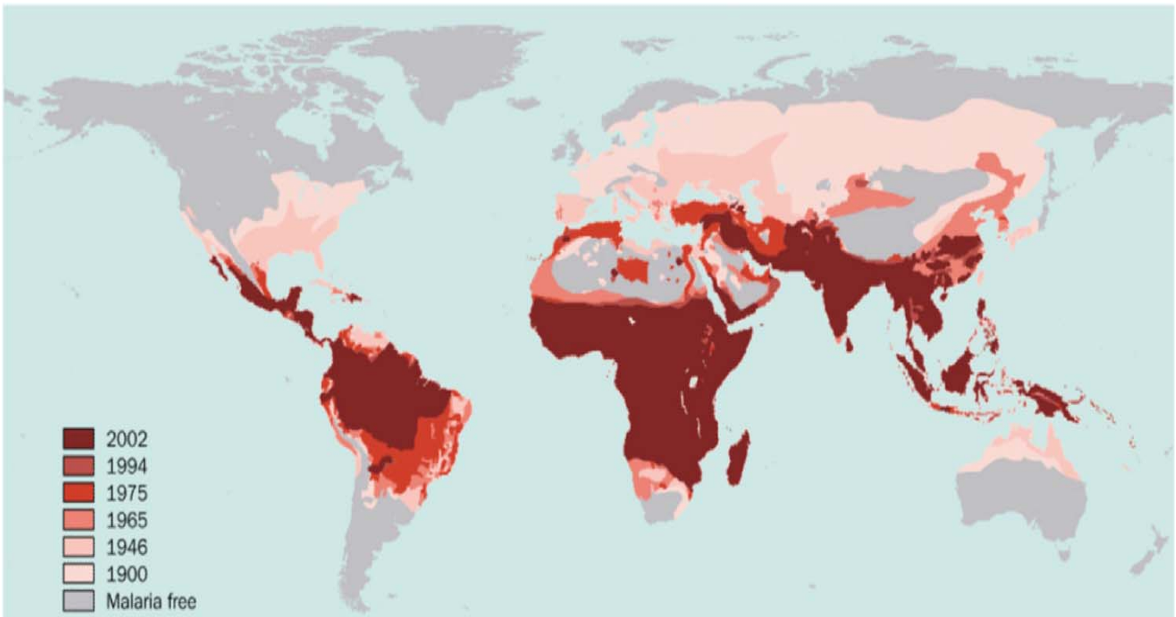


Figure 2.3: Changes in the global distribution of malaria since 1900

Notes: Areas of high and low risk are merged to establish all-cause malaria transmission limits (Hay et al., 2004)

As shown in Figure 2.3 the U.S. were also historically affected by vector borne diseases until the first half of the 20^{th} century. Before the implementation of vector control policies, the problem was so serious that, as shown in the 9^{th} Census of the United States realized in 1870, for some south western counties more than 10% of deaths from all causes were caused by malaria alone. In other words 1 out of 10 people died because of causes linked to malaria. Figure 2.4 below shows in detail the proportion of deaths caused by malaria to the total number of deaths.

In the U.S. early effective measures to prevent the transmission of vector borne diseases started

The map shown in Figure 2.4 has been created by Francis A. Walker using data from the Statistical Atlas from the 9^{th} Census of the United States 1870

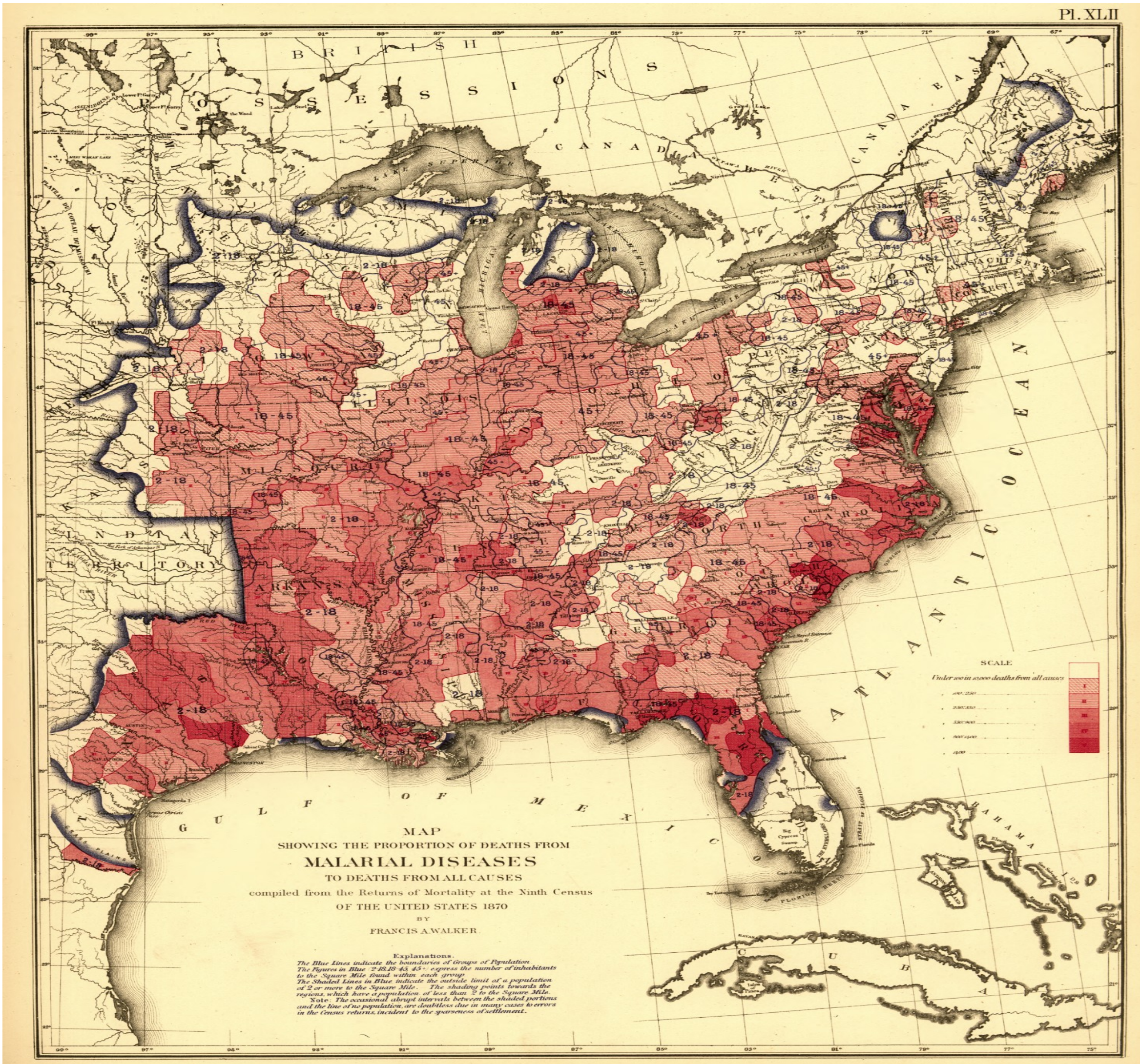


Figure 2.4: Proportion of deaths caused by malaria to deaths from all causes, US 1870

Map retrieved from the Statistical Atlas from the 9th Census of the United States 1870

with the establishment of the U.S. Public Health Service (USPHS) and the programs implemented by the Rockefeller Foundation in early 1900s. Such measures lasted until 1930s when new technological advancements made the discovery of the DDT possible along with new fertilizers. The newly discovered chemical component, originally developed as an insecticide for use as an agricultural and household pesticide, further reduced the reproduction rates of mosquitoes and improved agricultural production Derryberry and Gartrell (1952). DDT was extensively used in agriculture from 1940s to early 1970s when it was finally banned in the US because of the controversial effects it had on the environmental biodiversity and human health. The US Environmental Protection Agency (EPA, 1975) has estimated that a total of 1.8 million tonnes have been produced globally since the 1940s with more than 600,000 tonnes produced in the U.S. only before the 1972 ban. The usage of DDT reached its peak in 1959 at about 36,000 tonnes in the United States only.

Given the aforementioned evidence malaria has possibly played an important role in the socio-economic development of different U.S. areas both in terms of population and labour productivity. Hence one might ask whether more endemic areas have suffered in terms of human capital compared to less endemic ones.

To this regard, economic literature has extensively studied the relationship between disease with particular interest to malaria, and human capital loss for a long time. In particular, previous microeconomic literature has studied the impact of different disease control policies on human capital providing large evidence of the positive impact of disease prevention campaigns on *individual* education levels which are widely regarded as fundamentals of persistent economic growth Bleakley (2007), Bleakley (2010a), Cutler et al. (2010) and Bleakley et al. (2014). However, macroeconomic studies have generally emphasized the Malthusian view, arguing that the expansion in population in the short-term due to the increase in life expectancy could not be matched by an increment in the availability of natural resources, thus leading to modest (if any) improvements on GDP per capita (Hansen and Lønstrup, 2015, Acemoglu and Johnson, 2007, Ashraf et al., 2008).

No study, however has clearly addressed the role that malaria might indeed have played in explaining different historical economic development of US areas. This because some counties might have benefited from the eradication of malaria in terms of agricultural productivity, therefore generating positive effects on their economies compared to less endemic counties.

Gallup and Sachs (2001) and Sachs and Malaney (2002) first pointed out the correlation between malaria transmission and economic growth arguing that reversed causality would not pose a substantial problem. However, from their first work a large strand of economic studies on the causal impact of health and disease on economic growth has followed, showing that people living in areas with higher levels of malaria infection prior to some eradication campaign experienced greater increases in school attendance and literacy afterwards compared to people living in non malarious areas in the Americas, India and Italy (Bleakley, 2010a, Cutler et al., 2010, Percoco, 2013). Specifically, Bleakley (2003) and Bleakley (2010b) show that level of income of adults not exposed to tropical diseases during their childhood was higher compared to those exposed to weather conditions suitable for transmission of infectious diseases.

However, as argued by Acemoglu and Johnson (2007) and later by Hansen and Lønstrup (2015) microeconomic studies are likely overestimating the real impact of disease transmission reduction since they do not control for general equilibrium effects of the increased life expectancy due to less mortality. By instrumenting life expectancy with the predicted mortality rates of 15 major diseases of the 20^{th} century Acemoglu and Johnson (2007) find a negative relationship between the latter and GDP per capita growth. Finally, in a recent macroeconomic study, by using highly disaggregated data Gooch (2017) shows the positive effect that malaria eradication has had on world population and population density. That said, macroeconomic studies only consider mortality rates rather than data morbidity rates. As a consequence, the economic burden of the infectious disease is likely to be underestimated since only a fraction of people affected by malaria effectively die, while a lot more find themselves physically debilitated thus not being able to work efficiently.

To have a comprehension of the possible effect of malaria on agricultural productivity is important to notice that prior to any intervention against malaria malarious countries had large portions of their cultivable land actually uncultivated or barely cultivated because of the prevalence of mosquitoes in those regions. For instance, as reported by Brown (1986) in 1898 Fortunato and

Franchaneti wrote a letter to their sponsor indicating the devastation caused by this disease in Italy. "*Malaria disease leaves uncultivated 2 million hectares of land. It poisons every year about 2 million inhabitants and kills 15,000 of them. There is no other health problem so deeply linked to the prosperity of our country.*" This means that the eradication of malaria other than increasing life expectancy and in turn population could have made the amount of land which could not be cultivated before (or only marginally cultivated) possible to cultivate without the prevalence of malaria, thus providing an exogenous positive output shock which might have counterbalanced the negative effect of the increased population.

By using a robust Difference in Difference approach (DID), this study aims at identifying the effects of the successful eradication of malaria in the U.S. on agricultural productivity for each county. In other words since malaria historically strikes rural areas more than urban ones and mostly farmers, its eradication could have fostered farm activities and crop cultivation in those areas which were suitable for agriculture but highly endemic. Thus, a positive impact of malaria eradication on agricultural productivity would possibly counterbalance the increase in population due to the lower mortality rate, meaning that we could "escape" from the Malthusian trap as pointed out by Acemoglu and Johnson (2007).

None of the above mentioned related studies has yet attempted to evaluate the impact of the eradication of malaria on the historical agricultural productivity of each county in the US. As presented in Section **??** establishing a clear direction in the relationship between malaria and agricultural productivity is crucial to understand the causes of historical economic development of different areas of the US. Moreover, it is also important to evaluate if the eventual impact of malaria control on historical agricultural productivity was due to a mere increase in arable land (i.e. due to wetland conversion to cropland) or if the latter was explained by an increase in labour productivity

This corresponds to the equivalent area of the italian region Puglia

Which might have a negative impact on the GDP per capita as shown in Acemoglu and Johnson (2007) and Hansen and Lønstrup (2015) deriving from better health conditions of farmers whom were not affected by malaria any longer. To this regard the present study investigates the two principal channels through which malaria control might effect agricultural productivity. These are:

1. Higher labour productivity due to better health conditions of farmers;

2. Greater availability of cropland due to wetland conversion to arable land.

Therefore, this study aims at answering two questions: has the eradication of malaria increased agricultural productivity of endemic counties with respect to less endemic ones? And, how much of the eventual increase was due to the mere increment of arable land and to the higher labour productivity of farmers? The robust empirical strategy in this study allows to disentangle the two effects of malaria control policies on agricultural productivity.

2.2 Data

This study merges time invariant malaria stability index with historical disaggregated data on US agricultural productivity at the county level. This section briefly explains how each variable is constructed or obtained.

2.2.1 Measuring Spatial Prevalence of Malaria

First, highly malaria endemic counties were identified by using the Malaria Endemicity Index (henceforth, MSI). The MSI was first developed by Kiszewski et al. (2004) and is a spatially disaggregated time-invariant global index representing the stability of malaria transmission. The MSI index has been recently used in a number of recent papers in development economics (Alsan, 2015, Giuliano and Nunn, 2013, Michalopoulos and Papaioannou, 2013, Easterly and Levine, 2016, Henderson et al., 2017) for its advantage of being independent of socio-economic conditions. Using a time-invariant index representing the suitability of each county to the transmission of malaria and related vector borne diseases is crucial to alleviate potential endogeneity issues. Indeed, one of the benefits of the MSI is that it does not directly measures human infection rates which might in turn depend upon socio-economic conditions. The MSI is indeed constructed by using weather variables available at 0.5°by 0.5°level along with the estimation of the survival time of mosquitoes larvae based on regionally dominant species of Anopheles mosquitoes. Another benefit of the MSI is that it is precise, since it is based on information measured with contemporary GIS accuracy. This may explain why the stability index captures moderate malaria prevalence more effectively.

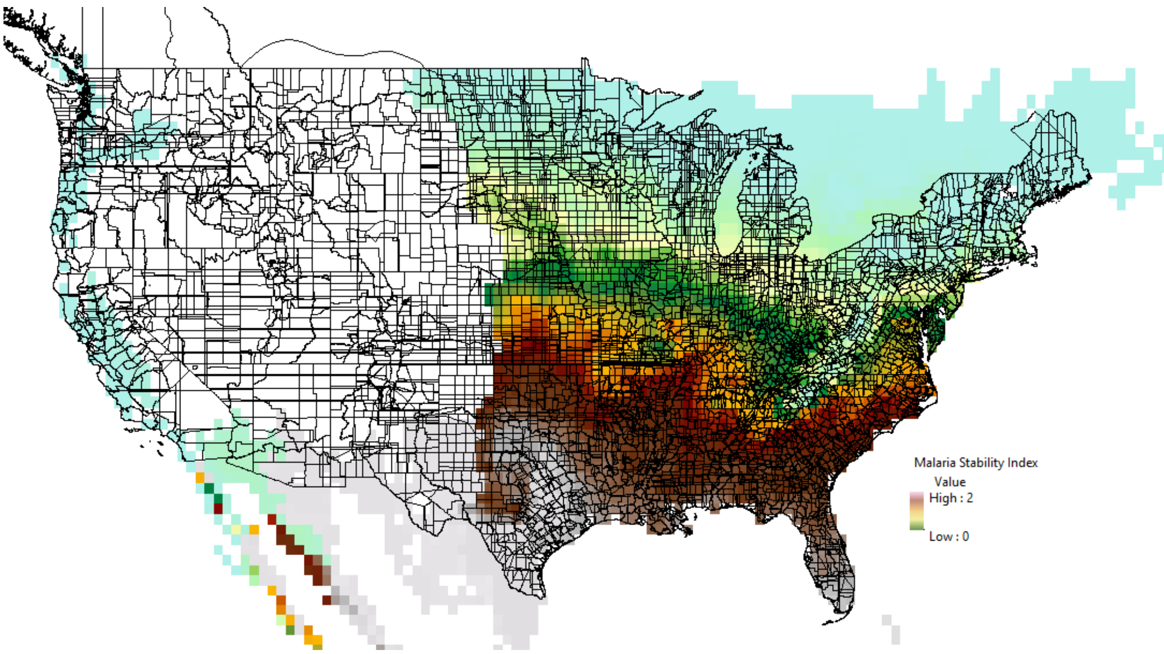


Figure 2.5: Malaria Suitability Index for the US

The MSI captures the potential stability of malaria transmission based on regionally dominant vector mosquitoes, temperature and precipitation data set (Kiszewski et al., 2004). This is intersected with US historical county borders.

A second measure of historical prevalence of malaria is the index developed by Hay et al. (2004) by digitizing historical maps shown first in Lysenko and Semashko (1968). However, this second

Weather data used for constructing the MSI are monthly average, minimum and maximum temperature, humidity and precipitations. Climatic data are exclusively provided by the Climate Research Unit at the University of East Anglia malaria stability index is less precise that the MSI, since it is constructed at 1°by 1°, and most importantly it uses actual infection rates data, thus potentially being influenced by human activity. For these reasons, this study uses both the MSI and the MEI as indexes of *suitability* of transmission of malaria and actual *prevalence* of the disease.

The MSI is a continuous variable whose values range from 0 to 39 with 0 being totally malaria free grid (i.e. the mosquito larva cannot survive with these climatic conditions) and 39 the grid with weather conditions most suitable for a stable and permanent transmission of malaria (i.e. mosquitoes larvae may survive and reproduce without obstacles of climatic nature any day of the year). However, the stability index comes as a categorical version too. Therefore, in the empirical analysis I use the MSI index as both a continuous and discrete variable.

The MEI is also a continuous variable capturing the observed distribution of malaria. MEI values range from 0 to 1 with 0 being a grid for which no case of malaria was observed until 1900 and 1 being a grid with the highest infection rate of malaria. The lowest endemicity level is hypoendemic with a value ≤ 0.1, followed by mesoendemic grid ≤ 0.5, hyperendemic with a value ranging from 0.5 to 0.75 and holoendemic with a value > 0.75. Although the MEI index ranges from 0 to 1 worldwide, values for the USA do not exceed 0.5 (in other words the USA ranged from being totally malaria free to mesoendemic, while cases of hyperendemic and holoendemic areas were not registered).

Since this study merges spatial data on historical county boundaries with 0.5°x 0.5°gridded data on malaria stability and endemicity values, some historical counties had an area which was less than that of a grid. To fix this issue the MSI and the MEI are interpolated at a very disaggregated geographic level (i.e. 0.2°x 0.2°) in order to have at least an stability value for each county. I then cal-

Grids with values from 0 to 0.05 are considered as malaria free while grids with values greater than 0.06 present climatic conditions suitable for a certain persistence of malaria transmission. In total Kiszewski et al. (2004) differentiated stability levels in 9 categories: 0-0.05, 0.06-1, 1.01-2, 2.01-5, 5.01-8, 8.01-12, 12.01-18, 18.01-25, 25.01-39

I have used the natural neighbor interpolation method automatically computed by the GIS software.

culate the mean of the index for each county. The results from the intersection of the MSI and the MEI and the historical US counties are shown in Figure 2.6. Important to notice is that the Malaria Stability Index (MSI) relative to the US territories ranges from 0 to about 1.5. This means that the transmission of malaria was largely unstable, implying that climatic conditions suitable for the survival and reproduction of mosquitoes larvae were mainly concentrated in one season, e.g. summer. Hence, rather than asking by how much historical agricultural productivity was affected by an increase in the MSI and the MEI by one unit, it is more relevant to evaluate how much an increase in the MSI and MEI by 0.1 unit affected agricultural productivity after the eradication of malaria.

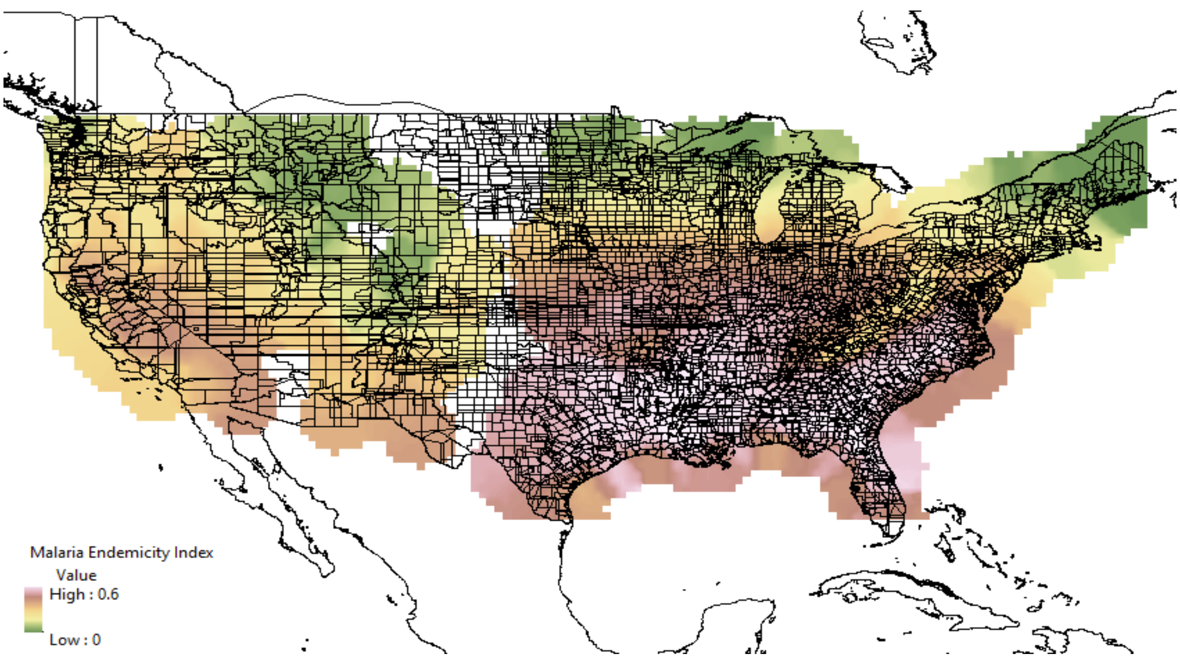


Figure 2.6: Malaria Endemicity Index for the US

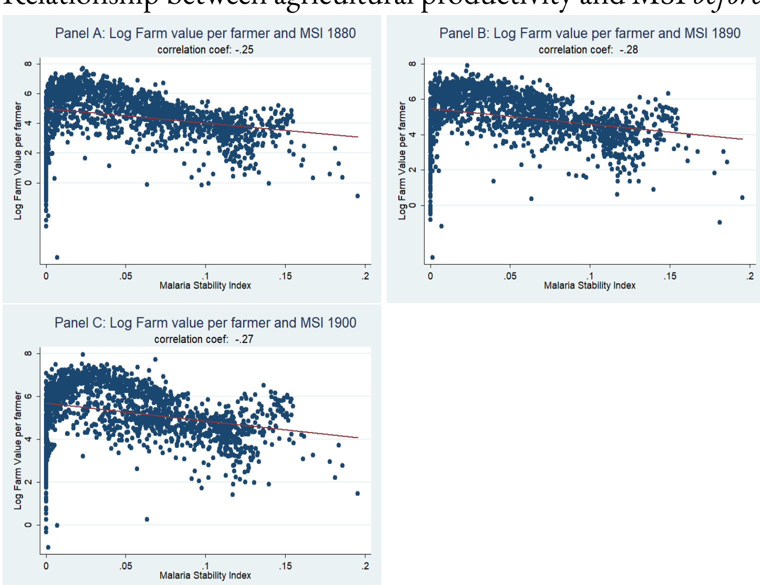
The MEI captures the distribution of Malaria in the USA in 1900. Areas of high and low risk are merged to establish all-cause malaria transmission limits (Hay et al., 2004). This is intersected with historical US county borders.

Higher MSI values do not necessary imply a large number of people affected but more stable climatic conditions for species of anopheles mosquitoes to live throughout the whole year.

2.2.2 Agricultural Productivity

The second variable of interest used in this study is agricultural productivity. Historical, Demographic, Economic, and Social Data: The United States, 1790-2002 (Haines et al., 2005) provides for two different farm productivity measures, namely total farm value and farm output per county acre of arable land and total farm value and farm output per county farmer. Farm value is defined as the value of all farmland, housing and outbuildings at the time of census enumeration. As an alternative measure of agricultural productivity I use farm output defined as the total value of all farm products, such as crop and livestock products, within the year prior to the enumeration day. Unfortunately, farm output value is not available in the census years of 1910, 1920, and 1930. Therefore, I consider the period 1870 to 1900 as pre-eradication of malaria while 1940 to 2000 as post eradication period. The reason why I consider the total farm value and output per acre of farmland other than total farm value and output per farmer is because in this way possible changes in cropland for each county (also due to conversion of swamps and wetland into cropland in order to eradicate malaria) would be taken into account. These two measures of productivity actually consider the productivity of land rather than farmers labor productivity. Alternatively, I consider farm value/output per number of farmers in order to investigate also labor productivity in the agricultural sector rather than land productivity only. Thus, this study considers a two measures of county land productivity (i.e. farm value and farm output) and two measures of farmer labor productivity per county (i.e. farm value and farm output per number of farmers). Historical, Demographic, Economic, and Social Data: The United States, 1790–2002 provides data on farm value per county registered at each agricultural census starting from 1850 and undertaken every decade until 2000. However, the quality of agricultural census was acceptable only starting from 1870. This determined the choice to use farm value and farm output per farmland county acre and farm value and farm output per farmer only starting from 1870. Therefore, pre-eradication period starts from 1870 until 1900 with 1910 being the starting year of vector control policies which took place until 1940s with the introduction of the DDT. A worthwhile consideration is the change in county boundaries during the period 1870-2000. These changes of county boundaries matter when county fixed effects are considered. As in Bleakley and Hong (2017) to partially fix the problem, I adjust farm value, along with all the other variables at county level that I control for, on the 1870 county boundary using the area-weighted average method. Figures 2.7 and 2.8 shows respectively the MSI index and the MEI joined with the log farm value per farmer in different decades before and after the introduction of vector control policies. Panel A, B and C show a significant negative correlation between the MSI and the MEI and the log of farm value per farmer which does not appear to change until 1900 (i.e. prior to start of the eradication campaigns) with associated slopes being -0.25, -0.28 and -0.27 for 1880, 1890 and 1900 respectively (-0.14, -0.19 and -0.22 for the MEI). This negative correlation decreased in 1920 as shown in Panel D and disappears from 1950 until the present day as shown in Panels E and F with associated slopes being -0.18, -0.12 and -0.07 for 1920, 1950 and 1970 respectively (-0.13, -0.09 and 0.09 for the MEI).

Relationship between agricultural productivity and MSI *before* vector control policies



Relationship between agricultural productivity and MSI *after* vector control policies

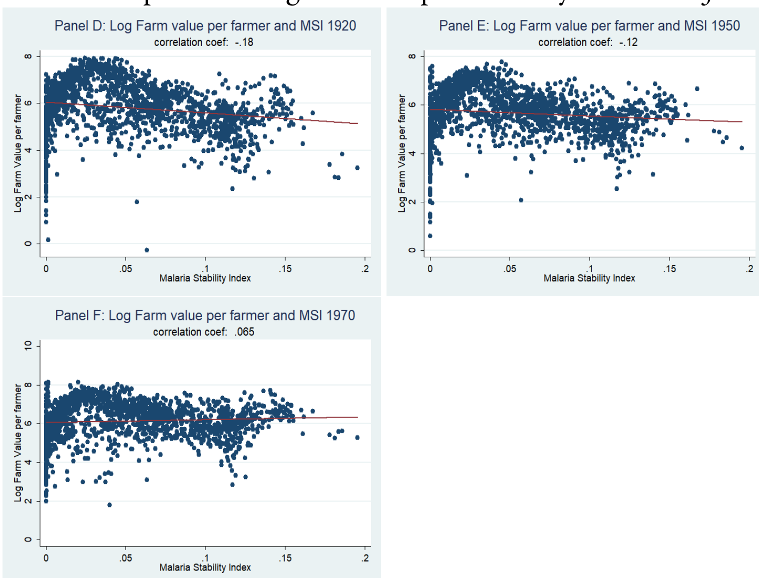
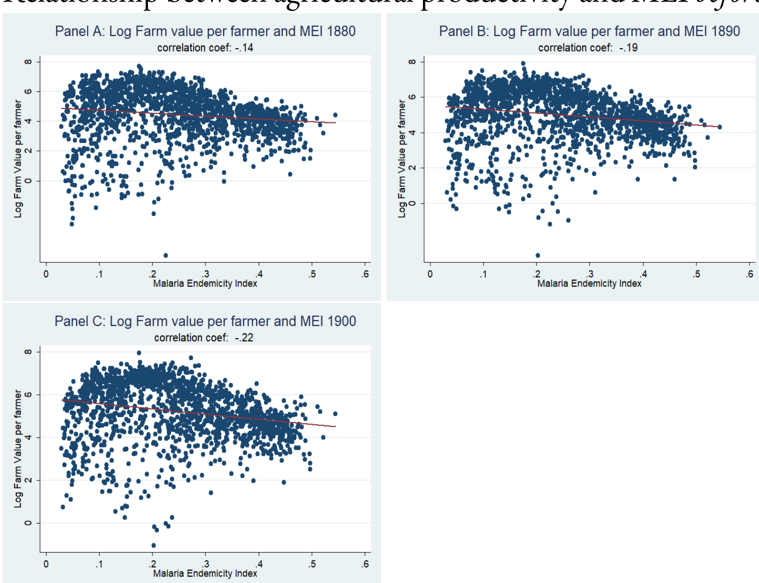


Figure 2.7: Correlation of malaria stability levels and log county farm value per farmer in 1880, 1890, 1900, 1920, 1950 and 1970

Relationship between agricultural productivity and MEI *before* vector control policies



Relationship between agricultural productivity and MEI *after* vector control policies

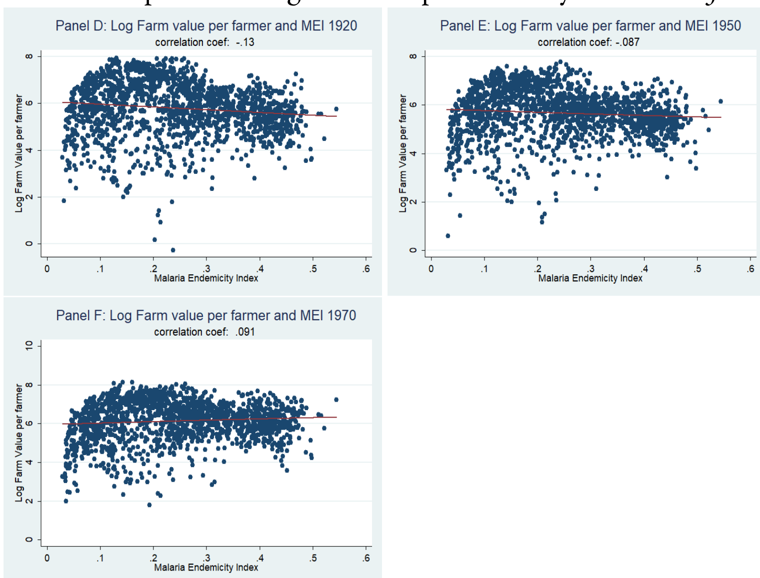


Figure 2.8: Correlation of malaria stability levels and log county farm value per farmer in 1880, 1890, 1900, 1920, 1950 and 1970

2.3 Empirical Framework

This study is intended to show the effects of vector control policies on historical farm productivity per farmers for each US county. Figures 2.7 and 2.8 in Section 2.2.1 provide a visual representation of the correlation between the two measures of malaria incidence and farm productivity throughout the 20^{th} century. The following section presents more formally the empirical strategy adopted to determine the relationship between the eradication of malaria and historical county agricultural productivity in the US. The estimation method presented below follows the same logic of a standard differences-in-differences (DID) strategy. This allows for the comparison of counties with high levels of stability of malaria with counties with low malaria stability and between counties with high historical incidence of malaria and counties with low incidence of the vector borne disease. The difference between the empirical strategy presented in this study and the standard DID is that the treatment (which in this specific case is represented by the MSI and the MEI) is not discrete but it is a continuous measure (namely I will not simply compare malarious counties with non malarious ones but more malarious with less malarious counties). Therefore, the procedure will capture more variation in the treatment. The regression presented below estimates the impact of the county average MSI and MEI on the two indices of farm productivity (i.e. farm value and farm output per farmer). In the next section I will further explore if the effects of the treatment of malaria on agricultural productivity of US counties was also in part due to the land conversion process implemented in those years i.e. from wetland and swamps into arable and therefore cultivable land.

$$y_{i,s,t} = \alpha \; Malaria_i \cdot I_t^{Post} + \sum_{t=1870}^{2000} \Omega \; \mathbf{X'}_{i,s} I_t + \sum_{t=1870}^{2000} \delta I_t + \sum_{p} \gamma_p I_t^p + \sum_{t=1870}^{2000} \sum_{q} \omega_q I_t^q + \varepsilon_{\; i,s,t} \quad (2.1)$$

where index i represents each US county and t indexes time periods considered in the analysis (i.e. 1870, 1890, 1900, 1910, 1920, 1930, 1940, 1950, 1960, 1970, 1980, 1990 and 2000). Years from 1870 to 1900 are part of the pre-eradication period, and since first national-wide policies were implemented soon after 1900, 1900 is considered the last year prior to the interventions. Post intervention period ranges from 1910 until the 1950s when the last round of the National Malaria Eradication Programm (1951) was undertaken (US Center for Disease Control and Prevention, 2010). By the end of 1950s malaria was considered eliminated in all US counties also thanks to the usage of newly discovered chemical components such as the DDT. The dependent variable is represented as $y_{i,s,t}$ which is the natural log of one of the two measures of agricultural productivity defined in section 2.2.2, i.e. farm value per farmer and farm output per farmer respectively, associated with county i, state s at time t. The variable $Malaria_i$ is one of the two indices representing the incidence of malaria, i.e. the average time invariant malaria stability index of county i and the average malaria endemicity index of county i. Variable I_t^{Post} is a post treatment dummy variable which takes value 1 for years after 1900 (i.e. 1910, 1920, 1930 and so on until 2000) while value 0 for years before the exogenous intervention against malaria (i.e. 1870, 1880, 1890 and 1900). This specification also includes county fixed effects $\sum_p \gamma_p I_t^p$, where p indicates the set of US counties; time period fixed effects $\sum_{t=1870}^{2000} \delta I_t$. $\mathbf{X}_{i,s}$ represents vectors of time invariant county-specific agricultural controls included in the regression. As county-level controls I use a set of relevant geographical and historical county specific characteristics which might have affected farm productivity during the 20^{th} century as in Bleakley and Hong (2017): county population density in 1870, the ratio of white population to county population in 1870, the ratio of farmland to total available county area in 1870, the proportion of farmers reporting the use of fertilizers in 1870, the proportion of farmland that reported have received drainage in 1870, the proportion of farmland with improved land in 1870 and county average agriculture suitability. A correlation matrix between the MSI and the MEI index and these population and agricultural controls is shown in Figure 2.0.2 in the Appendix. Figure 2.0.2 also shows the correlation between the MEI index and the MSI, the population density and the proportion of white farmers out of total farmers. To account for these differences, when comparing malaria-free counties with more endemic counties, I include the following covariates: (1) two interaction terms: ln(Population Density in 1870) interacted with the MSI and ln(Proportion of white farmers out of total farmers in 1870) interacted with the MSI, and the stability index interacted with the full set of time fixed effects. Finally, in order to address the concern that the econometric strategy might simply capture the fact that southern states of the US might on average be more suitable for agriculture than northern states of the US and that each state might have had agricultural productivity growth rates differently from others because of any other reason besides the eradication of malaria I include state fixed effects interacted with time period fixed effects to the baseline specification equation 2.1. State-by-time fixed effects are represented as follows: $\sum_{t=1870}^{2000} \sum_q \omega_q I_t^q$. Where q indicates the set of US states. As emphasized, the coefficient of interest is α which indicates by what extend malaria stability has contributed to the agricultural productivity of county i before and after the vector control intervention. With the inclusion of state-by-time fixed effects the coefficient of interest α is identified from within-state variation only, that is the empirical strategy compares counties within the same state and therefore subjected to the same federal policies. To this regard it is important to stress that decisions to tackle malaria along with other vector borne diseases (i.e. supply of free quinine, availability of newly discovered pesticides etc..) were made at the governmental level and implemented at federal level by each state. Counties had not power in deciding when and where to implement malaria eradication campaigns.

Although I am aware of the fact that other variables not included in the regressions could have explained the historical changes in the relationship between malaria stability and farm productivity, the main limitation to the use of a higher number of controls is due to the long time period of the analysis. That said however, the inclusion of other control variables such as government policies and investments which might have had a role in explaining changes in farm productivity would absorb part of the effects of malaria treatment to agricultural productivity because government policies, for instance, are investments in adapting to circumstances, including the effects of malaria and other vector borne diseases. For this reason we would not want to remove such adaptation from the estimates, because they are part of the story. I use the same set of control variables at the county level as in Bleakley and Hong (2017) including also the average county suitability for crop cultivation. Agricultural Suitability Index (henceforth ASI) is computed by the Center for Sustainability and Global Environment at the University of Wisconsin-Madison (Ramankutty et al., 2002). To compute the suitability for agriculture Ramankutty et al. (2002) rely on weather and environmental conditions required for crop cultivation. Weather conditions are taken from the global climatic database compiled by the Climate Research Unit at the University of East Anglia.

As with a standard difference-in-difference, the empirical strategy adopted in this study relies

The database includes monthly data on weather conditions (e.g. precipitations, cloud cover, minimum, maximum temperature, humidity etc.) from 1900 until 2016.

on the assumption that no other event, besides the availability of new drugs and effective chemical components in preventing the transmission of malaria and vector borne diseases, also occurred in early 1900s and affected farm productivity differently from each US county. This is a crucial assumption which should not be taken for granted since the US have experienced many changes during the 20^{th} century. This issue is partially fixed by the massive number of counties considered in this analysis (i.e. almost 2000 historical counties are part of the analysis). However, I implement a number of cautions in order to examine whether the patterns in the data are consistent with this assumption. First, as described in section 2.1 the historical evidence suggests that in the US early effective measures act to prevent the transmission of vector borne diseases started with the establishment of the US Public Health Service (USPHS) in early 1900s. Those measures consisted in the free availability of quinine in all hospitals in the US, the use of newly discovered pesticides along with the drainage of swamps and wetland. Given this evidence, the most reasonable cutoff date is 1900, and therefore 1910 is the first post-adoption time period.

Second, in order to be sure that no event other than the eradication of malaria happened from 1900s and affected agricultural productivity in the US I estimate a fully flexible estimating equation as in Nunn and Qian (2011) which takes the following form:

$$y_{i,s,t} = \sum_{j=1870}^{2000} \alpha_j \, Malaria_i \cdot I_t^j + \sum_{t=1870}^{2000} \Omega \, \mathbf{X}'_{i,s} I_t + \sum_{t=1870}^{2000} \delta I_t + \sum_p \gamma_p I_t^p + \sum_{t=1870}^{2000} \sum_q \omega_q I_t^q + \varepsilon_{\,i,s,t} \quad (2.2)$$

The only difference from equation 2.1 is that in equation 2.2, rather than interacting $Malaria_i$ with a post-adoption indicator variable, I interact respectively the malaria stability and endemicity measures with each of the time-period fixed effects. The estimated vectors of α_j reveal the relationship between the MSI (and the MEI) and the agricultural productivity measures in each time-period (e.g. we will have an estimate α_j for each decade from 1870 to 2000). If, for instance, no other event occurred in early 1900s and the eradication of malaria had a positive effect on farm productivity, then we would expect the estimated α_js not to be statistically significant over time for the years before the adoption of vector control policies while becoming positive and statistically significant after the eradication campaigns started in early 1900s. We would also expect the α_js to be constant after the eradication campaigns successfully ended (i.e. 1960s circa)

Figures 2.9 and 2.10 below show the estimated α_js compared to 1870 (i.e. the starting year of the analysis) along with the associate confidence intervals.

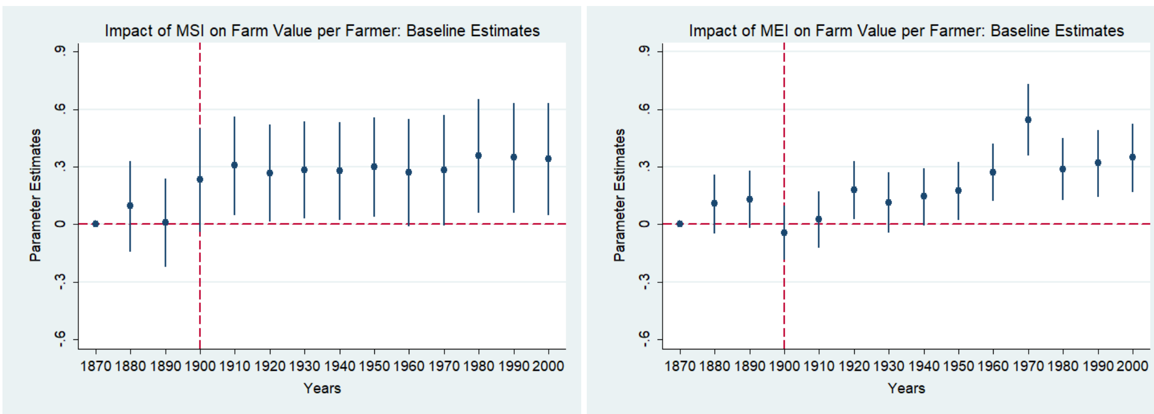


Figure 2.9: Flexible coefficients of the relationship between farm value per farmer and MSI (left) and MEI (right): All counties

Estimates of equation 2.2 along with their confidence intervals reported in Figures 2.9 and 2.10 show a clear pattern, i.e. the relationship between the time invariant MSI index per county and the MEI index and farm productivity per farmer is not statistically significant different from zero until 1900 (i.e. last year prior to the onset of the eradication policies). A spike between 1910 and 1920 emerges, followed by a small resurgence in during the 1930s and 1940s and then it steadily increases in magnitude from 1940 to 1970. Coefficients shown in Figures 2.9 and 2.10 are also crucial to exclude that any other event except the eradication of malaria occurred during the time periods immediately prior to the implementation of vector control policies. These results lead us to understand

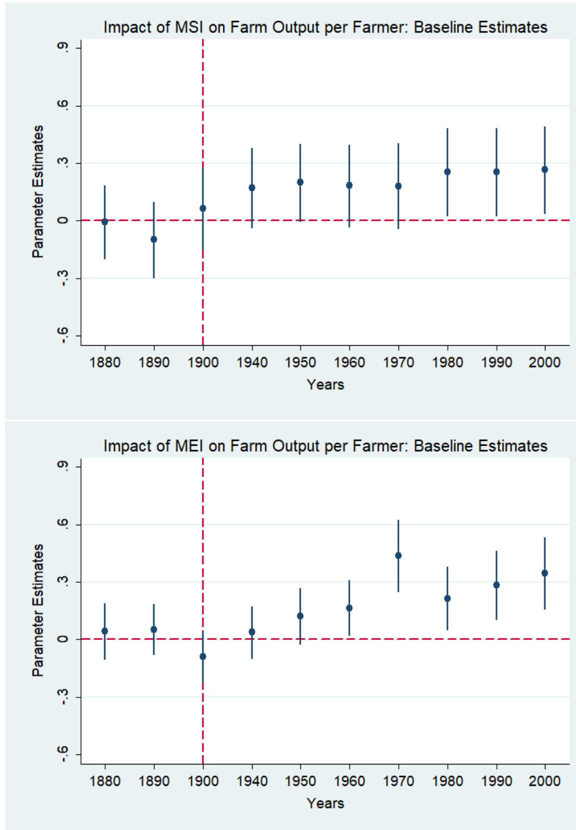


Figure 2.10: Flexible coefficients of the relationship between farm output per farmer and MSI (left) and MEI (right): All counties

that counties with a greater degree of prevalence of malaria had no statistical difference in farm productivity from counties with less prevalence of malaria prior to the eradication of the vector borne disease and that after the eradication of malaria more endemic counties showed a higher agricultural productivity relatively to less endemic ones. However, a limitation of this procedure is that from Figures 2.9 and 2.10 the mechanism underlying the increase in farm productivity of more endemic counties cannot be clearly established. In other words: how much of the latter was caused by the increase in cropland, in turn due to land conversion, and by the greater labour productivity of farmers, in turn due to better health conditions? This question will be addresses in the next section.

2.4 Results

2.4.1 Baseline Results

As Figures 2.9 and 2.10 show, after 1900 farm productivity of counties with higher incidence of malaria begins to increase relative to counties which were less endemic and less suitable for the transmission of vector borne diseases. The effect on farm productivity appears to begin immediately after 1900, this is probably due to better health conditions of farmers who would be cured from malaria by newly discovered drugs more effectively. This last assumption will be addressed more formally in the next section. The coefficients thereafter appear to slightly increase until 1970 when malaria was fully eradicated in the US.

Table 1 reports different specifications of equation 2.1 where the variable of interest is the suitability of malaria, i.e. the MSI. Therefore, the results reported in Table 1 show the effects that the eradication of malaria after 1900 had on the farm value per farmer of counties with high stability of malaria compared to counties with less stability of malaria. Column (1) controls only for the proportion of farmland to total county acre area other than including state-specific linear trends, time and county fixed effects. A 0.1 increase in the stability of malaria index is associated to an increase of almost 20 percentage points in farm productivity, measured as farm value per farmer, of counties with weather conditions more suitable to the transmission of malaria. Moving across columns more controls are added, including controls on population density which might have pushed for a greater agricultural productivity, the proportion of white people out of total county population, the use of fertilizers, drainage of land and the use of improved land along with the average agricultural suitability of county i. Nevertheless the coefficients indicating the impact of the MSI on farm productivity do not appear to considerably change in magnitude and of statistical significance thus, reducing concern for selection on unobservables (Altonji et al., 2005). The preferred specification is reported in column (4) of Table 1 and includes all population and agricultural controls other than

Table 2.1: The impact of MSI on county agricultural productivity: Farm value per farmer

	(1) b/se	(2) b/se	(3) b/se	(4) b/se
MSI x Post	0.834***	0.810***	0.819***	0.190***
	(0.242)	(0.256)	(0.276)	(0.031)
Farmland	-1.645***	-2.126***	-2.120***	-1.729***
	(0.167)	(0.189)	(0.182)	(0.102)
Pop density	-6.446***	-8.373***	-8.436***	-2.435***
	(1.391)	(1.658)	(1.656)	(0.532)
White people	2.115***	1.675***	1.697***	-0.335**
	(0.111)	(0.176)	(0.152)	(0.130)
Fertilizer		9.280***	8.864***	2.581***
		(1.382)	(1.972)	(0.915)
Land improved		1.072***	1.147***	-0.144
		(0.228)	(0.351)	(0.177)
Drainage		0.256***	0.303***	0.409***
		(0.072)	(0.088)	(0.076)
Agriculture suitability			-0.124	0.003
			(0.236)	(0.097)
State X Year	No	No	No	Yes
Observations	23783	23769	23769	23769
Adjusted R^2	0.505	0.511	0.511	0.846

Table 2.2: The impact of MEI on county agricultural productivity: Farm value per farmer

	(1) b/se	(2) b/se	(3) b/se	(4) b/se
MEI x Post	0.805***	0.582***	0.584***	0.140***
	(0.036)	(0.037)	(0.037)	(0.034)
White people x MSI	17.040***	3.190**	3.285**	-0.343
	(1.366)	(1.309)	(1.349)	(1.224)
Pop density x MSI	-50.166**	-57.392**	-57.218**	-25.939***
	(22.201)	(23.159)	(23.102)	(7.684)
Farmland	-1.453***	-3.108***	-3.110***	-1.858***
	(0.078)	(0.132)	(0.133)	(0.102)
Fertilizer		3.840***	3.681***	2.087**
		(0.753)	(0.851)	(0.924)
Land improved		2.582***	2.615***	-0.085
		(0.142)	(0.206)	(0.175)
Drainage		0.136*	0.152**	0.389***
		(0.077)	(0.076)	(0.076)
Agriculture suitability			-0.044	0.027
			(0.133)	(0.098)
State X Year	No	No	No	Yes
Observations	23783	23769	23769	23769
Adjusted R^2	0.475	0.513	0.513	0.846

including state-specific linear trends, time and county fixed effects. Results in column (4) show that after the implementation of malaria eradication policies a 0.1 increased in the MSI has increased farm productivity of highly malaria suitable counties by 19 percentage points with respect to less malaria suitable counties. Table 2 instead reports the results of the impact of the MSI on the second measure of farm productivity, i.e. farm output per farmer. Here the results observed in Table 1 are confirmed. Therefore, the positive impact of the stability of malaria after its eradication is consistent with both measures of county agricultural productivity i.e. farm value per farmer and farm output per farmer. Similarly Table 3 and Table 4 report the results of equation 2.1 where variable $Malaria_i$ is the average endemicity value (MEI) for each county i. Table 3 shows the impact of the MEI on county farm value per farmer, while in Table 4 results of the impact of the MEI on county farm output per farmer are reported. The preferred specification is reported in column (4) of both Table 3 and Table 4. Estimates, are consistent with those in Table 1 and Table 2. We, thus confirm that the eradication of malaria, which in the US took place in early 1900s, caused an increase in agricultural productivity in counties were malaria was more prevalent and stable.

2.4.2 Impact of land conversion

Was the increase in farm productivity the effect of the increment in the availability of cropland due to the conversion of wetland? Figure 2.11 below shows no correlation between the diffusion of malaria and the available cropland per county neither before the eradication campaigns nor after their implementation. This leads us to have a clear insight on the main channels through which the control of malaria has increased farm productivity since it appears that highly endemic counties did not see a significant increase in their total cropland. In other words the increase in agricultural productivity in malarious counties was not the result of a greater availability of arable land which did not increase with the eradication of malaria but the result of a minor number of farmers needed in one acre of farmland. This means that if until 1900 in more malarious counties more farmers

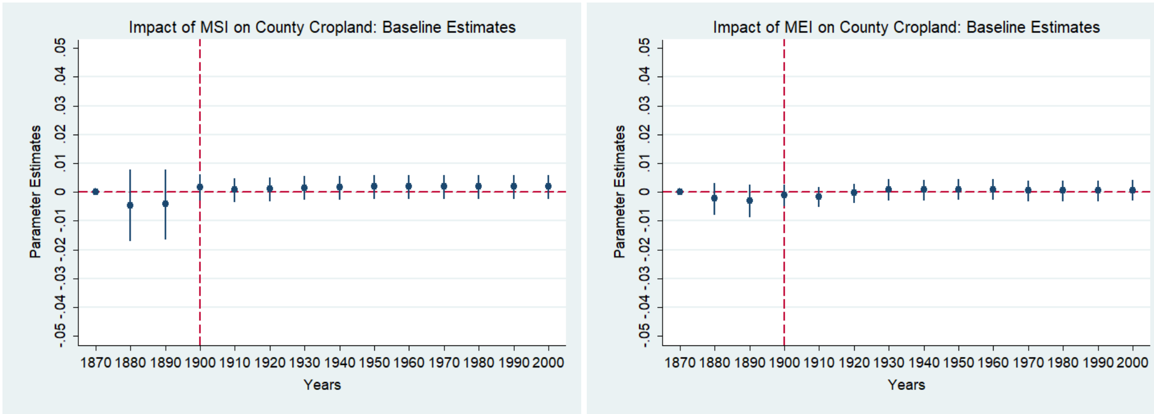


Figure 2.11: Flexible coefficients of the relationship between proportion of cropland out of total county area and MSI (left) and MEI (right): All counties

were needed in order to compensate for those who were affected by malaria, after its eradication the farm output per farmer increased compared to less malarious counties.

2.5 Threats to Validity and Robustness Checks

A potential threat to the validity of the empirical analysis comes from the possible endogeneity of the treatments to malaria which could be adopted by highly malarious counties first. In other words counties with greater diffusion of malaria could have received the treatment before counties with low or no infection rates. If this is the case, then having received the treatment before less malarious counties would bias the results upwards. As discussed in Section **??**, economic literature addressing the effect of diseases and health on economic outcomes (Bleakley, 2007, Cutler et al., 2010) and more recently Gooch (2017) has argued that the reduction in the burden of malaria clearly resulted from critical scientific innovations coming overwhelmingly from outside the highly endemic counties and which have culminated with better understanding of the origins of the disease and the discover of new drugs and chemical components. Therefore, major discoveries on malaria such as that it is transmitted to human beings through the bites of certain species of mosquitoes living in par-

Table 2.3: The impact of MSI on county amount of cropland

	(1) b/se	(2) b/se	(3) b/se	(4) b/se
MSI x Post	0.001	0.001	0.001	0.001
	(0.002)	(0.002)	(0.002)	(0.002)
Farmland	-0.004*	-0.004*	-0.003	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)
Pop density		0.007	0.014	0.013
		(0.009)	(0.011)	(0.011)
White people		0.001	0.003	0.003
		(0.004)	(0.005)	(0.005)
Fertilizer			-0.003	-0.008
			(0.024)	(0.025)
Land improved			-0.005	-0.004
			(0.005)	(0.006)
Drainage			0.000	0.001
			(0.001)	(0.001)
Agriculture suitability				-0.001
				(0.003)
State X Year	Yes	Yes	Yes	Yes
Observations	23776	23776	23763	23763
Adjusted R^2	0.999	0.999	0.999	0.999

Table 2.4: The impact of MEI on county amount of cropland

	(1)	(2)	(3)	(4)
	b/se	b/se	b/se	b/se
MEI x Post	-0.001	-0.001	-0.002	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)
Farmland	-0.003	-0.003	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)
White people x MSI		-0.011	-0.002	-0.001
		(0.027)	(0.027)	(0.028)
Pop density x MSI		-0.152	-0.100	-0.096
		(0.110)	(0.095)	(0.088)
Fertilizer			0.000	-0.002
			(0.022)	(0.023)
Land improved			-0.006	-0.006
			(0.004)	(0.006)
Drainage			0.001	0.001
			(0.001)	(0.001)
Agriculture suitability				-0.001
				(0.003)
State X Year	Yes	Yes	Yes	Yes
Observations	23776	23776	23763	23763
Adjusted R^2	0.999	0.999	0.999	0.999

ticular weather conditions or the finding of the curative properties of quinine and the diffusion of newly discovered pesticides and chemical components did not come from regions in which malaria was more prevalent. Moreover, as shown in Williams (1952) and Humphreys (2001) decisions on the adoption of campaigns addressed to the eradication of malaria were taken at the governmental level and implemented at the federal level by each state. Based on historical evidence we conclude that the eradication of malaria depended on discoveries from abroad and counties had no decision power in terms of availability and administration of quinine to people affected from malaria, use of pesticides and later on of the DDT and drainage of swamps and wetland. Hence, this attenuates possible concerns of endogeneity of the treatment by areas of the US more affected by vector borne disease. That said, I adopt a series of tests to check that:

1. The relationship between the two measures of prevalence of malaria and farm productivity per county did not change until 1900;

2. Counties with high prevalence of malaria are comparable to counties with low prevalence of the disease;

3. The cutoff date (i.e. 1900) is reasonably correct and no other cutoff produces statistically significant results.

First, since counties with higher stability in the transmission of malaria and high infection rates were predominantly located in the southern part of the US, I limit the analysis to those counties only. The results are shown in Figure 2.3 in the Appendix. By limiting the analysis to southeast US we are considering states and counties which had respectively weather conditions particularly suitable for a stable transmission of malaria, measured with the MSI index, and high infection rates of malaria, measured with the MEI. Thus, the econometric procedure will compare the effect of slight changes in the MSI and MEI on agricultural productivity of malarious counties.

Neighboring Counties Analysis

Similarly, I implement the same procedure, this time considering only a subset of counties rather than all of the available ones. The robust strategy compares *neighboring* counties belonging to the same state (and therefore subject to the same federal policies) which had substantially different MSI and MEI values. In particular, when considering the MSI as variable representing the diffusion of malaria, I compare each county showing weather conditions suitable for the transmission of malaria (i.e. with a MSI $\geq$ 0.06) with its neighboring counties that presented weather conditions not suitable for the reproduction of mosquitoes larvae (i.e. with a MSI $\approx$ 0). This restriction substantially reduces concerns of any difference between counties with more or less malaria stability. Therefore, the procedure allows to select only those neighborhoods containing at least one malarious and one malaria free county . The reasoning behind this restriction is that we would be sure to compare counties which are very similar, that is they are close enough, and belong to the same state but only differ in terms of malaria stability values.

For the same reason, when considering the MEI as variable representing the diffusion of malaria, I compare each county showing high endemicity of malaria (i.e. with a MEI $\geq$ 0.1, or in other words mesoendemic counties) with its neighboring counties that presented showed no diffusion of malaria (i.e. with a MEI $\approx$ 0 or malaria-free counties).

A total of 258 counties were obtained following the aforementioned restriction. I therefore, create an indicator variable I_n for each neighborhood n considered which is added in equation 2.1 so to compare endemic counties with non-endemic counties belonging to the same neighborhood. The estimated regression therefore becomes the following:

(Neighboring counties are defined as those counties having their centroids within a grid of 2° latitude by 2° longitude and belong to the same state)

For instance a neighborhood containing all malarious counties was not considered. The same reasoning was applied to neighborhoods with no malarious county. In this way we do not compare a malarious county in Alabama with a malaria free county in Montana

The 258 selected counties belonged to a total of 119 different neighborhoods

$$y_{i,s,t} = \alpha\ Malaria_i \cdot I_n \cdot I_t^{Post} + \sum_{t=1870}^{2000} \Omega\ \mathbf{X'}_{i,s} I_t + \sum_{t=1870}^{2000} \delta I_t + \sum_p \gamma_p I_t^p + \sum_{t=1870}^{2000} \sum_q \omega_q I_t^q + \varepsilon_{\ i,s,t} \quad (2.3)$$

Where $Malaria_i$ is again one of the two indexes, namely MSI based on weather conditions suitable for a stable reproduction of mosquitoes larvae and the MEI based on actual infection rates of malaria.

In equation 2.3 the coefficient of interest α is identified no more from within-state variations, but *within-neighborhood* variation only. With this restriction the econometric procedure not only compares counties which are belonging to the same state (and therefore subjected to the same federal policies) but also neighboring counties. That is, the empirical strategy compares counties within the same state whose centroids are within 2° latitude by 2° longitude. By comparing neighboring counties which had different degrees of diffusion of malaria we are substantially reducing concerns of possible differences within the same state.

If highly malarious counties were effectively different from less malarious counties we should expect the parameter estimates of years prior to the implementation of vector control policies to be statistically significant. However, as Figure 2.12 and 2.13 show this is not the case. Results obtained when considering the almost 2000 counties are thus confirmed by the restriction of neighboring counties.

Figures 2.12 and 2.12 shows respectively the estimates of the impact of the MSI and the MEI on the measures of agricultural productivity relative to a baseline time-period, which we take to be 1870. Therefore, the absolute level simply tells us the difference in the relationship relative to an arbitrarily chosen baseline. We then would expect the parameter estimates of 1880, 1890 and 1900 (i.e. years prior to the discovery of anti malaria methods) not to be statistically significant different from the baseline year, while being positive from 1910 onward. What we observe is that prior to the

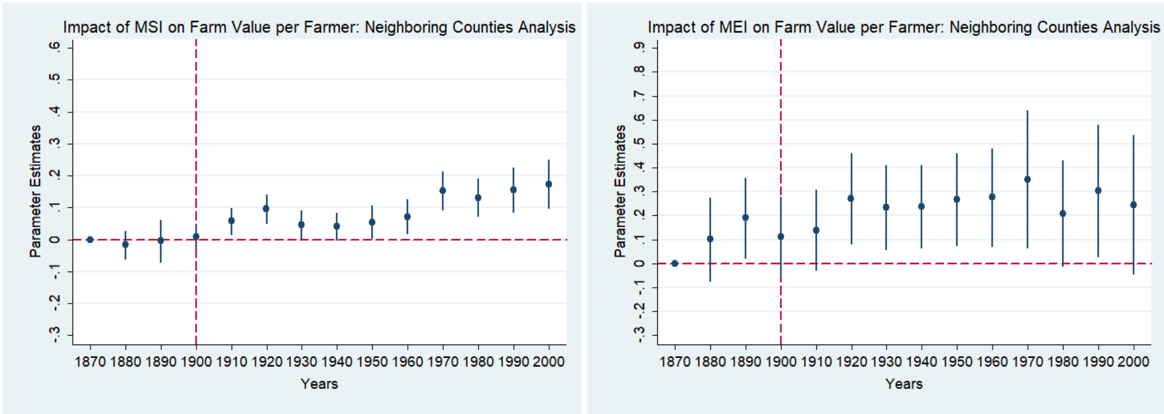


Figure 2.12: Flexible coefficients of the relationship between farm value per farmer and MSI (left) and MEI (right): Neighboring Counties

eradication of malaria the relationship between the diffusion of malaria and agricultural productivity was not statistically different from highly malarious counties compared to less malarious counties. After 1900 however, this relationship starts to be positive until 1920s, indeed during 1930s and 1940s we observe a resurgence of the transmission of malaria followed by a stable decrease in the number of infected people which translates into higher growth rates of agricultural productivity.

Estimates of equation 2.3 are reported in Table 5 and Table 6. Specifically, Table 5 reports the results of the estimation of equation 2.3 in which the dependent variable is represented by the natural logarithm of county average farm value per farmer, while Table 6 reports the results of the estimation of equation 2.3 in which the dependent variable is represented by the natural logarithm of county average farm output per farmer. Column (1) controls only for the proportion of farmland to total county acre area other than including state-specific linear trends, time and county fixed effects. Results here are very similar in magnitude compared to the estimates of equation 2.1 and show that counties with weather conditions more favorable to a stable transmission of malaria (i.e. counties whose MSI ≥ 0.06) experienced an increase of more than six percentage points in farm productivity compared to malaria free neighboring counties (i.e. counties whose MSI ≈ 0). As before, concerns

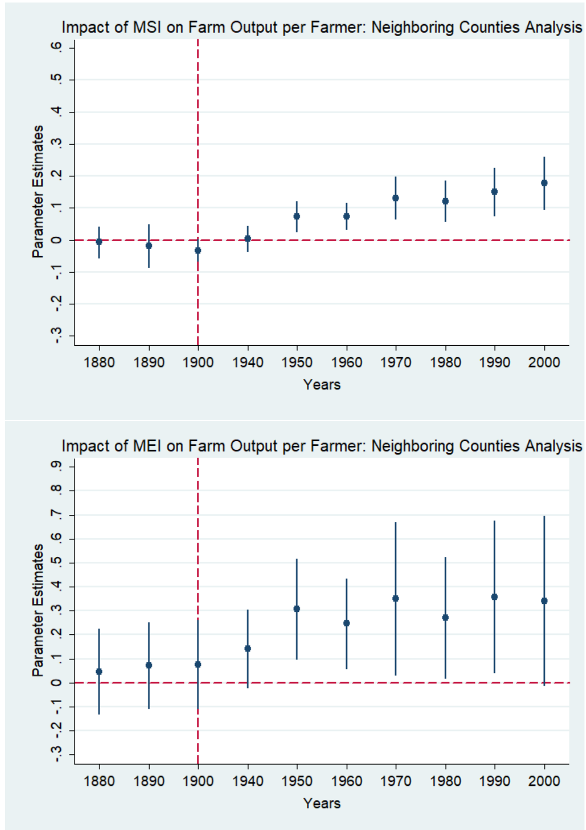


Figure 2.13: Flexible coefficients of the relationship between farm output per farmer and MSI (left) and MEI (right): Neighboring Counties

for selection on unobservables are also mitigated by the fact that moving across columns coefficients indicating the impact of the MSI on farm productivity do not appear to considerably change in magnitude and of statistical significance. The preferred specification is reported in column (4) and includes controls on county population density in 1870, the ratio of white population to the total county population in 1870, the ratio of farmland to total available county area 1870 the proportion of farmers reporting the use of fertilizers in 1870, the proportion of farmers that had their farm drained in 1870, the proportion of farms with improved land and the county average agriculture suitability index, other than including state-specific linear trends, time and county fixed effects. Results in column (4) show that counties with MSI $\geq$ 0.06 experienced an increase of about seven percentage points in farm productivity compared to neighboring counties with MSI $\approx$ 0 after the eradication of malaria.

Table 2.5: The impact of MSI on county agricultural productivity: Farm value per farmer. Neighboring Counties Analysis

	(1) b/se	(2) b/se	(3) b/se	(4) b/se
MSI x Post	0.103***	0.108***	0.099***	0.101***
	(0.027)	(0.026)	(0.028)	(0.026)
Farmland	-1.225***	-1.084***	-1.147***	-1.120***
	(0.118)	(0.122)	(0.131)	(0.131)
Pop density		-3.148***	-3.591***	-3.614***
		(0.672)	(0.849)	(0.775)
White people		0.085	0.053	0.147
		(0.110)	(0.143)	(0.155)
Fertilizer			1.639**	0.639
			(0.635)	(0.706)
Land improved			0.434*	0.722**
			(0.244)	(0.325)
Drainage			0.407***	0.506***
			(0.070)	(0.073)
Agriculture suitability				-0.413*
				(0.214)
Observations	7375	7375	7361	7361
Adjusted R^2	0.868	0.870	0.872	0.873

Similar tables are produced showing the results of the estimation of equation 2.3 considering the MEI and are included in the Appendix. Here we compare counties with high actual infection rates (i.e. counties with MEI $\geq$ 0.1) to again neighbor counties which had no cases of malaria (i.e with MEI $\approx$ 0). Results obtained considering the actual infection rates of malaria are consistent with those obtained with the time invariant malaria stability index.

Placebo Treatment Periods

The estimated coefficient for the interaction between the Malaria Endemicity Index and the post

Table 2.6: The impact of MEI on county agricultural productivity: Farm value per farmer. Neighboring Counties Analysis

	(1)	(2)	(3)	(4)
	b/se	b/se	b/se	b/se
MEI x Post	0.164***	0.160***	0.090	0.100
	(0.061)	(0.060)	(0.060)	(0.066)
Farmland	-1.300***	-1.244***	-1.274***	-1.253***
	(0.121)	(0.119)	(0.134)	(0.130)
White people x MSI		1.358	0.537	1.944
		(1.551)	(2.118)	(2.146)
Pop density x MSI		-22.211***	-24.518***	-23.854***
		(5.610)	(7.018)	(6.943)
Fertilizer			1.081*	0.188
			(0.565)	(0.774)
Land improved			0.372	0.618*
			(0.246)	(0.321)
Drainage			0.382***	0.472***
			(0.075)	(0.073)
Agriculture suitability				-0.366*
				(0.213)
Observations	7375	7375	7361	7361
Adjusted R^2	0.868	0.870	0.872	0.873

eradication dummy variable indicates the average increment in farm productivity between the pre and post periods for highly malarious counties relative to less malarious counties. As shown early, it is crucial that in order to be valid, estimates, are expected to be close to zero until the cutoff period. I have chosen 1900 as cutoff year since it is consistent with the historical description of the discovery that malaria was transmitted to humans through the bite of mosquitoes and of the curative properties of quinine with its subsequent mass distribution. Therefore, until 1900 there is no reason to expect counties with higher prevalence of malaria to have differential growth in farm productivity. To test this assumption I estimate equation 2.1 considering different placebo cutoff dates. The estimates are reported in Table 7. Column (1) reports estimated effects for farm productivity using a sample that includes forty-year pre-eradication periods, ranging from 1870 to 1900 (i.e. 1870, 1880, 1890 and 1900). For this regression, the post indicator variable I^{Post} takes on the value of zero in 1870 and the value of one in 1880, 1890 and 1900. Column (2) reports estimated effects for farm productivity which consider the same period of column (1) but with the post indicator variable I^{Post} taking on the value of zero in 1870 and 1880 and the value of one in 1890 and 1900. Therefore, in columns (1) and (2) I check the validity of 1880 and 1890 respectively as cutoffs. Since the major discoveries of the treatment against malaria and in turn, the first adoption of vector control policies in the US did not take place before 1900 the results from specifications in columns (1) and (2) can be interpreted as a placebo experiment. To this regard, results in columns (1) and (2) yield statistically not significant coefficient estimates confirming that endemic counties did not have differential growth in farm productivity throughout the period from 1870 to 1900. Moreover, columns (1) and (2) prove that 1890 and 1900 cannot be considered as plausible cutoffs. Column (3) instead reports estimates using a sample that spans across forty years, ranging from 1890 to 1920. For this regression, the post indicator variable I^{Post} takes on the value of zero in 1890 and the value of one in 1900, 1910 and 1920. Here all three of the post decades coincide with the postadoption period. The post indicator variable now coincides exactly with the postadoption period. This time, the results displayed in column (3) yield positive and statistically significant coefficient estimates. Similarly column (4) reports estimates using the same sample of column (3), ranging from 1890 to 1920. For this regression, the post indicator variable I^{Post} takes on the value of zero in 1890 and 1900 while the value of one in 1910 and 1920. As in column (3) results reported in column (4) yield positive and statistically significant coefficient estimates. Finally columns (5) and (6) report estimates using a sample that spans across forty years, ranging from 1940 to 1970. In column (5), the post indicator variable I^{Post} takes on the value of zero in 1940 while the value of one in 1950, 1960 and 1970 while in column (6) the post indicator variable I^{Post} takes on the value of zero in 1940 and 1950 and the value of one in 1960 and 1970. None of the two specifications reported in columns (5) and (6) yield to statistically significant results. Thus, confirming that the differential growth in agricultural productivity between highly malarious counties with less malarious counties happened after 1900 until 1940, and neither before or after that period.

The same exercise is shown in Table 8. Here I consider the MEI as a measure of the transmission of malaria for each county. Again, as in Table 7 columns (1) and (2) do not yield statistical significant results. This is crucial to clearly state that the relationship between the endemicity level and the agricultural productivity of US counties did not change prior to 1900. Columns (3) and (4) show instead a slight negative estimate and a positive estimate respectively. This suggests that again more malarious counties experienced higher agricultural productivity growth rates only after 1900 and not before also when considering the actual data on infection rates of malaria. Finally, columns (5) and (6) of Table 8 also yield positive estimates. Therefore, it appears that the positive effect of the eradication of malaria on agricultural productivity in counties with higher infection rates lasted also after until 1970s.

Tables 9 and 10 show the same robustness check performed in Tables 7 and 8, this time considering the natural logarithm of the average county farm output per farmer as dependent variable. Farm output per farmer data at county level are missing for the years 1910, 1920 and 1930 therefore in

Tables 9 and 10 I only check the non existence of any other event prior to 1900 which changed agricultural productivity of more malarious counties compared to less malarious one. Results shown in Tables 9 and 10 confirm the conclusions drawn in Tables 7 and 8.

Taken together, the results shown in Tables 7 and 8 confirm what shown in Figures 2.9 and 2.12, namely that the relationship between the prevalence of malaria measured both as stability and endemicity and agricultural productivity starts to changes only after 1900 and in no other previous year (i.e. 1870-1900).

Table 2.7: The impact of MSI on county farm value per farmer: Placebo Treatment Periods

	1870-1900 Post=1880, 1890, 1900	1870-1900 Post= 1890, 1900	1890-1920 Post=1900, 1910, 1920	1890-1920 Post=1910, 1920	1940-1970 Post= 1950, 1960 1970	1940-1970 Post= 1960, 1970
	b/se	b/se	b/se	b/se	b/se	b/se
MSI x Post	0.174	0.039	0.258***	0.166***	0.004	-0.013
	(0.111)	(0.028)	(0.085)	(0.028)	(0.022)	(0.028)
State X Year	Yes	Yes	Yes	Yes	Yes	Yes
Pfarm 1870 X Year	Yes	Yes	Yes	Yes	Yes	Yes
Popden 1870 X Year	Yes	Yes	Yes	Yes	Yes	Yes
Pwhite 1870 X Year	Yes	Yes	Yes	Yes	Yes	Yes
Fertilizer 1870 X Year	Yes	Yes	Yes	Yes	Yes	Yes
Drain 1870 X Year	Yes	Yes	Yes	Yes	Yes	Yes
Land Imp 1870 X Year	Yes	Yes	Yes	Yes	Yes	Yes
Suitability X Year	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6752	6752	6808	6808	6808	6808
Adjusted R^2	0.863	0.856	0.880	0.877	0.817	0.821

*The periods range from 1870 to 2000, and are observed every other decade. The dependent variable is the natural log of the county average farm value per county acre of farmland. The variable of interest is the average county MSI index. The Post Indicator variable equals zero for the periods 1870–1900 and one for the periods 1910–2000. All regressions include time periods fixed effects, county fixed effects, and state by time fixed effects. The inclusion of a control variable interacted with the full set of time-period fixed effects is indicated by a yes; no indicates that the control is not included in the specification. Coefficients are reported with standard errors. In the parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.*

Table 2.8: The impact of MEI on county farm value per farmer: Placebo Treatment Periods

	1870-1900 Post=1880, 1890, 1900 b/se	1870-1900 Post= 1890, 1900 b/se	1890-1920 Post=1900, 1910, 1920 b/se	1890-1920 Post=1910, 1920 b/se	1940-1970 Post= 1950, 1960 1970 b/se	1940-1970 Post= 1960, 1970 b/se
MEI x Post	-0.004	-0.044	-0.083**	0.059**	0.116***	0.163***
	(0.057)	(0.044)	(0.038)	(0.026)	(0.029)	(0.036)
State X Year	Yes	Yes	Yes	Yes	Yes	Yes
Pfarm 1870 X Year	Yes	Yes	Yes	Yes	Yes	Yes
MSI	Yes	Yes	Yes	Yes	Yes	Yes
Ln(pop den) X MSI	Yes	Yes	Yes	Yes	Yes	Yes
Ln(p white) X MSI	Yes	Yes	Yes	Yes	Yes	Yes
Fertilizer 1870 X Year	Yes	Yes	Yes	Yes	Yes	Yes
Drain 1870 X Year	Yes	Yes	Yes	Yes	Yes	Yes
Land Imp 1870 X Year	Yes	Yes	Yes	Yes	Yes	Yes
Suitability X Year	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6752	6752	6808	6808	6808	6808
Adjusted R^2	0.864	0.857	0.881	0.877	0.818	0.822

*Notes: The periods range from 1870 to 2000, and are observed every other decade. The dependent variable is the natural log of the county average farm value per county acre of farmland. The variable of interest is the average county MSI index. The Post Indicator variable equals zero for the periods 1870–1900 and one for the periods 1910–2000. All regressions include time periods fixed effects, county fixed effects, and state by time fixed effects. The inclusion of a control variable interacted with the full set of time-period fixed effects is indicated by a yes; no indicates that the control is not included in the specification. Coefficients are reported with standard errors. In the parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.*

2.6 Concluding Remarks

This study has investigated the effects of the successful eradication of malaria on historical agricultural productivity growth of US counties. Joining an exogenous time-invariant spatial index which measures stability of malaria across different geographical locations with historical US county data on agriculture productivity this paper shows that after the adoption of policies designed to eradicate malaria and other vector borne diseases, counties with higher stability index were correlated with higher growth levels in agricultural productivity. Moreover, finding suggests that the positive effects of the eradication of malaria on agricultural productivity of counties with high stability of malaria are not due to mere land conversion (from wetland to arable land) leading to the conclusion that the principal mechanism through which malaria eradication led to increased farm productivity has to do with better health conditions of farmers which in turn led to higher labor productivity.

Estimating the impact of the successful eradication of malaria on historical agricultural productivity within the US is crucial for two main reasons: first, an extensive economic literature has show that understanding agricultural productivity is at the heart of understanding world income inequality (Gollin et al., 2014, Bustos et al., 2016). Therefore, being able to assess a clear effect of malaria on historical agricultural productivity in the US is fundamental to understand the causes of the differential economic growth of different areas within the US. To this regard, Figure 2.14 shows the negative correlation between log income per capita and the average malaria stability index per county at 2000.

Second, nowadays economies largely dependent on agriculture and highly malarious are mostly located within Sub-Saharan Africa, Latin America and South-East Asia. The results shown in this paper could be useful to predict how these countries are economically affected by malaria and, more importantly, how their economies will respond to policies aimed at eradicating it.

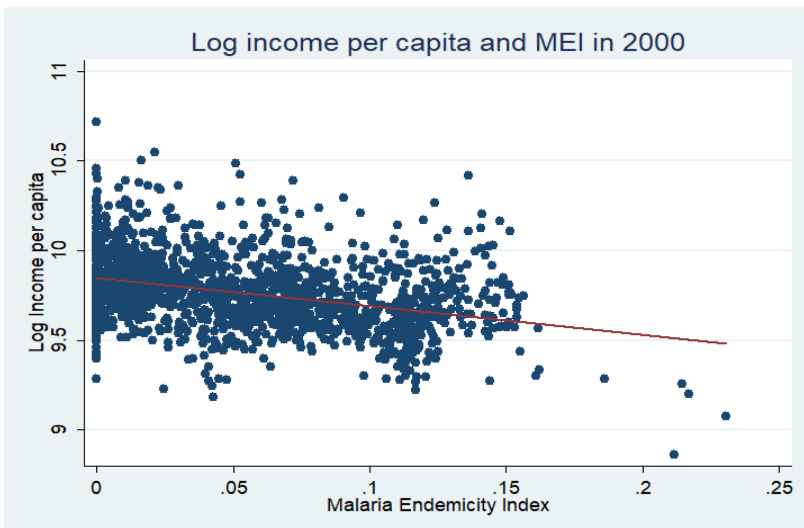


Figure 2.14: Correlation between malaria stability index and log income per capita per county in 2000

References

Acemoglu, D. and Johnson, S. (2007). Disease and development: the effect of life expectancy on economic growth. *Journal of political Economy*, 115(6):925–985.

Allen, S., Badiane, O., Sene, L., and Ulimwengu, J. (2014). Government expenditures, health outcomes and marginal productivity of agricultural inputs: The case of tanzania. *Journal of agricultural economics*, 65(3):637–662.

Alsan, M. (2015). The effect of the tsetse fly on african development. *American Economic Review*, 105(1):382–410.

Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of political economy*, 113(1):151–184.

Alves Dias, P., Blagoeva, D., Pavel, C., and Arvanitidis, N. (2018). Cobalt: demand-supply balances in the transition to electric mobility. *European Commission, Joint Research Centre, EUR-Scientific and Technical Research Reports. Publications Office of the European Union. DOI*, 10:97710.

Amnesty International (2016). This is What We Die For. Human Rights Abuses in the Democratic Republic of the Congo Power the Global Trade in Cobalt. p//wpgedaf62/3183/2016/e. [Online; accessed 9-July-2019].

Amnesty International (2017). *Time to Recharge: Corporate Action and Inaction to Tackle Abuses in the Cobalt Supply Chain* . Amnesty International.

Angrist, J. D. and Kugler, A. D. (2008). Rural windfall or a new resource curse? coca, income, and civil conflict in colombia. *The Review of Economics and Statistics*, 90(2):191–215.

Aragón, F. M. and Rud, J. P. (2015). Polluting industries and agricultural productivity: Evidence from mining in ghana. *The Economic Journal*, 126(597):1980–2011.

Ashraf, Q. H., Lester, A., and Weil, D. N. (2008). When does improving health raise gdp? *NBER macroeconomics annual*, 23(1):157–204.

Atkin, D. (2016). Endogenous skill acquisition and export manufacturing in mexico. *American Economic Review*, 106(8):2046–85.

Banza, C. L. N., Nawrot, T. S., Haufroid, V., Decrée, S., De Putter, T., Smolders, E., Kabyla, B. I., Luboya, O. N., Ilunga, A. N., Mutombo, A. M., et al. (2009). High human exposure to cobalt and other metals in katanga, a mining area of the democratic republic of congo. *Environmental research*, 109(6):745–752.

Basu, K. and Van, P. H. (1998). The economics of child labor. *American economic review*, pages 412–427.

Beegle, K., Dehejia, R., and Gatti, R. (2004). Why should we care about child labor? the education, labor market, and health consequences of child labor. Technical report, National Bureau of Economic Research.

Benshaul-Tolonen, A. (2018). Local industrial shocks and infant mortality. *The Economic Journal*, 129(620):1561–1592.

Berman, N., Couttenier, M., Rohner, D., and Thoenig, M. (2017). This mine is mine! how minerals fuel conflicts in africa. *American Economic Review*, 107(6):1564–1610.

Bleakley, H. (2003). Disease and development: Evidence from the american south. *Journal of the European Economic Association*, 1(2-3):376–386.

Bleakley, H. (2007). Disease and development: evidence from hookworm eradication in the american south. *The quarterly journal of economics*, 122(1):73–117.

Bleakley, H. (2009). Economic effects of childhood exposure to tropical disease. *American Economic Review*, 99(2):218–23.

Bleakley, H. (2010a). Health, human capital, and development. *Annu. Rev. Econ.*, 2(1):283–310.

Bleakley, H. (2010b). Malaria eradication in the americas: A retrospective analysis of childhood exposure. *American Economic Journal: Applied Economics*, 2(2):1–45.

Bleakley, H., Costa, D., and Lleras-Muney, A. (2014). Health, education, and income in the united states, 1820–2000. In *Human Capital in History: The American Record*, pages 121–159. University of Chicago Press.

Bleakley, H. and Hong, S. C. (2017). Adapting to the weather: lessons from us history. *The journal of economic history*, 77(3):756–795.

Bowden, S., Michailidou, D. M., and Pereira, A. (2008). Chasing mosquitoes: An exploration of the relationship between economic growth, poverty and the elimination of malaria in southern europe in the 20th century. *Journal of International Development: The Journal of the Development Studies Association*, 20(8):1080–1106.

Brown, G. (1986). Control and eradication of malaria: Past, present and future.

Bustos, P., Caprettini, B., and Ponticelli, J. (2016). Agricultural productivity and structural transformation: Evidence from brazil. *American Economic Review*, 106(6):1320–65.

Caminade, C., Kovats, S., Rocklov, J., Tompkins, A. M., Morse, A. P., Colón-González, F. J., Stenlund, H., Martens, P., and Lloyd, S. J. (2014). Impact of climate change on global malaria distribution. *Proceedings of the National Academy of Sciences*, 111(9):3286–3291.

Carrillo, B. (2019). Present bias and underinvestment in education? long-run effects of childhood exposure to booms in colombia. *Long-Run Effects of Childhood Exposure to Booms in Colombia (March 16, 2019)*.

Center for International Forestry Research (2013). The formalisation of artisanal mining in the democratic republic of the congo and rwanda. Technical report, Center for International Forestry Research.

Charles, K. K., Hurst, E., and Notowidigdo, M. J. (2018). Housing booms and busts, labor market opportunities, and college attendance. *American Economic Review*, 108(10):2947–2994.

Cobalt Institute (2019). Data on Production and Supply of Cobalt. [illegible] [Online; accessed 19-July-2019].

Conley, D., McCord, G. C., and Sachs, J. D. (2007). *Africa's lagging demographic transition: Evidence from exogenous impacts of malaria ecology and agricultural technology*.

Cust, J. (2015). The spatial effects of resource extraction: mining in indonesia. Technical report, Oxcarre Discussion Paper, University of Oxford.

Cutler, D., Fung, W., Kremer, M., Singhal, M., and Vogl, T. (2010). Early-life malaria exposure and adult outcomes: Evidence from malaria eradication in india. *American Economic Journal: Applied Economics*, 2(2):72–94.

Cuvelier, J. (2017). Money, migration and masculinity among artisanal miners in katanga (dr congo). *Review of African Political Economy*, 44(152):204–219.

Darton Commodities (2017). Annual market review on cobalt demand. p//wddifsokbi. Accessed: 2018-11-12.

De Herdt, T. and Titeca, K. (2016). Governance with empty pockets: the education sector in the democratic republic of congo. *Development and Change*, 47(3):472–494.

De Putter, T., Decrée, S., Banza, C. L. N., and Nemery, B. (2011). Mining the katanga (drc) copperbelt: geological aspects and impacts on public health and the environment—towards a holistic approach. In *Mining and the Environment in Africa. Proceedings of the Inaugural Workshop, IGCP/SIDA*, number 594, pages 14–17.

Deaton, A. (2013). *The great escape: health, wealth, and the origins of inequality*. Princeton University Press.

Demographic and Health Surveys (2014). The dhs program, usaid. *Calverton: Measure DHS*.

Derryberry, O. and Gartrell, F. (1952). Trends in malaria control program of the tennessee valley authority1. *The American journal of tropical medicine and hygiene*, 1(3):500–507.

Doepke, M. and Zilibotti, F. (2005). The macroeconomics of child labor regulation. *American Economic Review*, 95(5):1492–1524.

Dube, O. and Vargas, J. F. (2013). Commodity price shocks and civil conflict: Evidence from colombia. *The Review of Economic Studies*, 80(4):1384–1421.

Easterly, W. and Levine, R. (2016). The european origins of economic development. *Journal of Economic Growth*, 21(3):225–257.

Eckstein, Z. and Wolpin, K. I. (1999). Why youths drop out of high school: The impact of preferences, opportunities, and abilities. *Econometrica*, 67(6):1295–1339.

Edmonds, E. V. and Pavcnik, N. (2005). Child labor in the global economy. *Journal of Economic Perspectives*, 19(1):199–220.

EPA (1975). Ddt: A review of scientific and economic aspects of the decision to ban its use as a pesticide.

Euractiv (2019). Europe will have at least ten gigafactories. [illegible]. Accessed: 2019-04-05.

Faber, B., Krause, B., and De La Sierra, R. S. (2017). Artisanal mining, livelihoods, and child labor in the cobalt supply chain of the democratic republic of congo.

Financial Times (2019). Congo, child labour and your electric car. Technical report, Financial Times.

Fischer, G., Nachtergaele, F. O., Prieler, S., Teixeira, E., Toth, G., van Velthuizen, H. T., Verelst, L., and Wiberg, D. (2012). Global agro-ecological zones (gaez v3. 0).

for Disease Control Prevention, U. C. (2010). The history of malaria, an ancient disease.

Gallup, J. L. and Sachs, J. D. (2001). The economic burden of malaria. *The American journal of tropical medicine and hygiene*, 64(1_suppl):85–96.

Galor, O. and Weil, D. N. (2000). Population, technology, and growth: From malthusian stagnation to the demographic transition and beyond. *American economic review*, 90(4):806–828.

Ghosh Banerjee, S., Romo, Z., McMahon, G., Toledano, P., Robinson, P., and Pérez Arroyo, I. (2014). *The power of the mine: A transformative opportunity for sub-Saharan Africa*. The World Bank.

Giuliano, P. and Nunn, N. (2013). The transmission of democracy: From the village to the nation-state. *The American Economic Review*, 103(3):86.

Gollin, D., Lagakos, D., and Waugh, M. E. (2014). Agricultural productivity differences across countries. *American Economic Review*, 104(5):165–70.

Gooch, E. (2017). The impact of reduced incidence of malaria and other mosquito-borne diseases on global population. *Journal of Development Economics*, 124:214–228.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017). Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202:18–27.

Haines, M. R. et al. (2005). *Historical, demographic, economic, and social data: The United States, 1790-2002*. Inter-university Consortium for Political and Social Research.

Hansen, C. W. and Lønstrup, L. (2015). The rise in life expectancy and economic growth in the 20th century. *The Economic Journal*, 125(584):838–852.

Hausmann, R. and Rigobon, R. (2003). An alternative interpretation of the'resource curse': Theory and policy implications. Technical report, National Bureau of Economic Research.

Hay, S. I., Guerra, C. A., Tatem, A. J., Noor, A. M., and Snow, R. W. (2004). The global distribution and population at risk of malaria: past, present, and future. *The Lancet infectious diseases*, 4(6):327–336.

Hazan, M. and Berdugo, B. (2002). Child labour, fertility, and economic growth. *The Economic Journal*, 112(482):810–828.

Henderson, J. V., Squires, T., Storeygard, A., and Weil, D. (2017). The global distribution of economic activity: nature, history, and the role of trade. *The Quarterly Journal of Economics*, 133(1):357–406.

Hijmans, R., Garcia, N., and Wieczorek, J. (2010). Gadm database of global administrative areas. *V ersion*.

Hilson, G. (2011). Artisanal mining, smallholder farming and livelihood diversification in rural sub-saharan africa: An introduction. *Journal of international development*, 23(8):1031–1041.

Humphreys, M. (2001). *Malaria: poverty, race, and public health in the United States*. JHU Press.

International Labor Organization (2015). *World report on child labour 2015: paving the way to decent work for young people*. International Labour Organization.

Kiszewski, A., Mellinger, A., Spielman, A., Malaney, P., Sachs, S. E., and Sachs, J. (2004). A global index representing the stability of malaria transmission. *The American journal of tropical medicine and hygiene*, 70(5):486–498.

Klein Goldewijk, K., Beusen, A., Van Drecht, G., and De Vos, M. (2011). The hyde 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. *Global Ecology and Biogeography*, 20(1):73–86.

Lysenko, A. and Semashko, I. (1968). Geography of malaria. a medico-geographic profile of an ancient disease. *Itogi Nauki: Medicinskaja Geografija*, pages 25–146.

Majori, G. (2012). Short history of malaria and its eradication in italy with short notes on the fight against the infection in the mediterranean basin. *Mediterranean journal of hematology and infectious diseases*, 4(1).

McCarthy, D., Wolf, H., and Wu, Y. (2000). The growth costs of malaria. Technical report, National Bureau of Economic Research.

Medlock, J. M. and Leach, S. A. (2015). Effect of climate change on vector-borne disease risk in the uk. *The Lancet Infectious Diseases*, 15(6):721–730.

Michalopoulos, S. and Papaioannou, E. (2013). Pre-colonial ethnic institutions and contemporary african development. *Econometrica*, 81(1):113–152.

Millimet, D. L. (2011). The elephant in the corner: a cautionary tale about measurement error in treatment effects models. In *Missing Data Methods: Cross-Sectional Methods and Applications*, pages 1–39. Emerald Group Publishing Limited.

Newberry. Atlas of historical county borders. [illegible]. Accessed: 2018-09-30.

Nkulu, C. B. L., Casas, L., Haufroid, V., De Putter, T., Saenen, N. D., Kayembe-Kitenge, T., Obadia, P. M., Mukoma, D. K. W., Ilunga, J.-M. L., Nawrot, T. S., et al. (2018). Sustainability of artisanal mining of cobalt in dr congo. *Nature sustainability*, 1(9):495.

Nunn, N. and Puga, D. (2012). Ruggedness: The blessing of bad geography in africa. *Review of Economics and Statistics*, 94(1):20–36.

Nunn, N. and Qian, N. (2011). The potato's contribution to population and urbanization: evidence from a historical experiment. *The Quarterly Journal of Economics*, 126(2):593–650.

Olivetti, E. A., Ceder, G., Gaustad, G. G., and Fu, X. (2017). Lithium-ion battery supply chain considerations: analysis of potential bottlenecks in critical metals. *Joule*, 1(2):229–243.

Percoco, M. (2013). The fight against disease: Malaria and economic development in italian regions. *Economic Geography*, 89(2):105–125.

Picard, J. and Mills, A. (1992). The effect of malaria on work time: analysis of data from two nepali districts. *The Journal of tropical medicine and hygiene*, 95(6):382–389.

Ramankutty, N., Foley, J. A., Norman, J., and McSweeney, K. (2002). The global distribution of cultivable lands: current patterns and sensitivity to possible climate change. *Global Ecology and biogeography*, 11(5):377–392.

Ross, R. (1897). On some peculiar pigmented cells found in two mosquitos fed on malarial blood. *British medical journal*, 2(1929):1786.

Ruggles, S., Genadek, K., Goeken, R., Grover, J., and Sobek, M. (2015). Integrated public use microdata series: Version 6.0 [dataset]. minneapolis: University of minnesota.

Sachs, J. and Malaney, P. (2002). The economic and social burden of malaria. *Nature*, 415(6872):680.

Sachs, J. D. and Warner, A. M. (1995). Natural resource abundance and economic growth. Technical report, National Bureau of Economic Research.

Saul, A. (2002). Malaria: Poverty, race, and public health in the united states.

SP Global Market Intelligence (2018). SP Global Market Intelligence Annual Report. [illegible] 2018/2018-[illegible] [Online; accessed 19-July-2018].

Spencer, E. (2016). Cobalt supply and demand—a global perspective. In *Proceedings from The Cobalt Conference, Seoul, Korea*.

Subramanian, M. A. and Sala-i Martin, X. (2003). *Addressing the natural resource curse: An illustration from Nigeria*. Number 3-139. International Monetary Fund.

Sviatschi, M. M. (2019). Making a narco: Childhood exposure to illegal labor markets and criminal life paths.

Tatem, A. J., Gething, P. W., Smith, D. L., and Hay, S. I. (2013). Urbanization and the global malaria recession. *Malaria journal*, 12(1):133.

Thomas, D., Beegle, K., Frankenberg, E., Sikoki, B., Strauss, J., and Teruel, G. (2004). Education in a crisis. *Journal of Development economics*, 74(1):53–85.

Tsui, K. K. (2010). More oil, less democracy: Evidence from worldwide crude oil discoveries. *The Economic Journal*, 121(551):89–115.

Turcheniuk, K., Bondarev, D., Singhal, V., and Yushin, G. (2018). Ten years left to redesign lithium-ion batteries.

US Geological Survey (2019). Mineral commodity summaries 2019. Technical report, United States Geological Survey.

Weil, D. N. (2014). The impact of malaria on african development over the longue durée. *Africa's Development in Historical Perspective*, pages 89–130.

Williams, R. C. (1952). The united states public health service, 1798–1950. *AJN The American Journal of Nursing*, 52(10):1279.

Wouterse, F. and Badiane, O. The role of health, experience, and educational attainment in agricultural production: Evidence from smallholders in burkina faso. *Agricultural Economics*, 0(0).

Xia, F. and Deininger, K. (2019). Spillover effects of tobacco farms on the labor supply, education, and health of children: Evidence from malawi. *American Journal of Agricultural Economics*.

3

Revisiting the Contribution of Potato to Population and Urbanization in 1700-1900: The Effect of Malaria

We[*] exploit regional variation in suitability for cultivating potatoes along with exogenous variation in stability of malaria transmission, together with time variation arising from their introduction of potato to the Old World from the Americas, to estimate the impact of malaria endemic and potato suitable areas on Old World population and urbanization. This study shows that the presence of weather conditions suitable for a stable transmission of malaria counteracted the significant benefits of the introduction of potato to population and urbanization in Old World countries observed during the eighteenth and nineteenth centuries. Robustness checks from geographic variations in malaria stability and suitability for potato cultivation at a disaggregated level along with placebo treatments reinforce the positive effects of the eradication of malaria on population and urbanization in potato suitable areas after 1900.

Improving human health has a direct and positive impact on social welfare and economic output specially in the sectors like agriculture that are more labor intensive. Deaton (2013), Acemoglu and Johnson (2007) have shown that improvements in health conditions during the last century have positively affected world population and urbanization. The sharp increase in calorie intake and nu-

[*]Co-authored with Giacomo Falchetta and Soheil Shayegh

trients availability has played a key role in improving human health conditions specially after the introduction of new high-calorie crops like potato to the Old World Nunn and Qian (2011). The introduction of potato has been shown to contribute to as much as one-quarter of the growth in Old World population and urbanization between 1700 and 1900. There are two channels through which the introduction of potato has impacted population and urbanization. The first channel is *nutrition effect* where the introduction of a new, and more nutrient crop, has resulted in healthier and more productive population. The second channel is *productivity effect* where increased agricultural productivity has allowed more workers to migrate to the cities and work in industry (Galor and Weil, 2000). Therefore, countries with larger areas suitable for growing potato have experienced higher population and urbanization growth between 1700 and 1900.

In contrast to the positive impacts of health improvements on population growth and economic output, deteriorating health conditions due to infectious disease outbreaks such as malaria infection has negative effects on population and economic growth as shown in Bleakley (2003), Bleakley (2009), Cutler et al. (2010), and Gooch (2017). It is also shown that malaria outbreaks can be a decisive factor in hampering agricultural productivity through deteriorating farmers' health and ability to work (Malpede, 2019 mimeo). Therefore, any potential gain from introducing new crops can be offset by emergence of infectious disease threats that undermine human health and farmers' availability and productivity. For example, in the study presented by Nunn and Qian (2011) many areas which were highly suitable for the cultivation of potatoes, are those with climate conditions suitable for malaria transmission. Therefore, the net positive effect of the introduction of potatoes on population and urbanization in these areas might have been much lower than the areas suitable for potato cultivation but not for malaria transmission. In other words, the contribution to population and urbanization levels due to the introduction of a more efficient and nutrient staple crop such as potato could have been even higher in the absence of malaria. It is important to note that, in Nunn and Qian (2011)'s work, they assume that areas *suitable* for cultivating potatoes actually produce more potatoes. This important assumption must have been made due to the fact that there is no historical records on actual potato production. Therefore, actual productivity is assumed to be equal to potential suitability.

Our study contributes to this debate by comparing the impact of potato cultivation on population growth and urbanization in areas with and without malaria transmission risk. We expand the scope of Nunn and Qian (2011)'s analysis in several ways:

1. By relaxing the assumption that holds potato productions equal for areas with similar potato suitability. We assume the actual production is not only affected by suitability of each area for growing potatoes but also by its exposure to malaria transmission threat that can hamper the actual cultivation of potato through affecting farmers;

2. By expanding the country-level data used in Nunn and Qian (2011) and using historical disaggregated data on population and urbanization rates;

3. By including the effect of malaria transmission and comparing its impact *before and after* the malaria eradication program in 1900s.

The main contribution of this study is therefore, two-folded: First, we estimate the net effect of the introduction of potato to Old World population and urbanization considering the negative impact of malaria transmission threats. Second, we highlight the positive impact of malaria eradication on population growth and urbanization. We compare population and urbanization levels after 1700 of areas suitable for potato cultivation and at the same time not malaria endemic with those areas equally suitable for cultivating potato while presenting weather conditions suitable for a stable transmission of malaria. As Figure 3.1 shows, areas extremely suitable for potato[†] are likely to be malaria endemic as well. We argue that, this might have offset the positive impact of the introduction of potato in the Old World after 1700.

[†]Areas highly suitable for potato cultivation have a potato suitability index > 0.8 as defined by the FAO

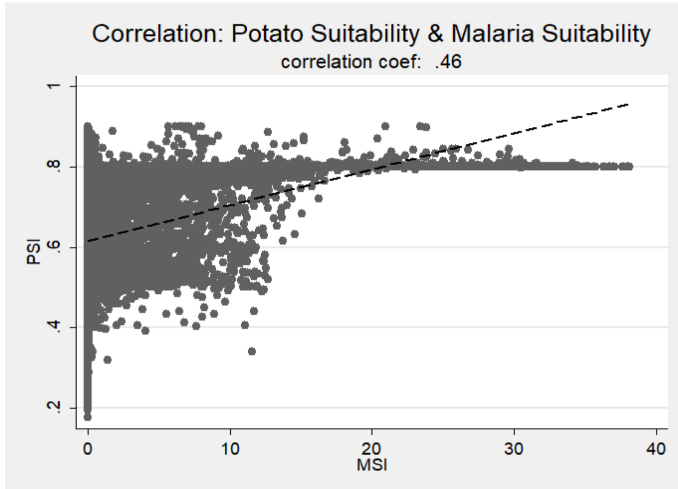


Figure 3.1: Correlation between Potato Suitability Index (PSI) and Malaria Stability Index (MSI)

Correlation between Potato Suitability Index (PSI) and Malaria Stability Index (MSI) in Old world, at 0.5° by 0.5° gridded level. Source: Authors' calculations based on Kiszewski et al. (2004) and Fischer et al. (2012).

The exogenous variations of weather conditions for the transmission of malaria allow for comparison between equally potato suitable areas with high and low endemic rates. We employ two different estimation strategies to account for the impacts of potato suitability in the presence or absence of malaria endemic.

The first estimation relies on the work by Nunn and Qian (2011). It exploits two sources of variation. The first is time variation arising from the introduction of potatoes as a field crop in the Old World. The introduction of potato in the Old World represents an exogenous agricultural shock since Potatoes did not exist and had never been cultivated in the Old World until 1650s. However, not all areas were suitable for growing potatoes and therefore, the cross-sectional differences in countries' suitability for cultivating potatoes represents the second source of variation of the estimation strategy employed in Nunn and Qian (2011). Differences in countries' suitability for cultivating potatoes are time invariant and obtained from the Food and Agriculture Organization

(FAO)'s Global Agro-Ecological Zones (GAEZ) 2012 database (Fischer et al., 2012)‡. In their estimation strategy Nunn and Qian (2011) compute the area suitable for cultivating potatoes for each Old World country and compare population and urbanization levels between Old World countries that had a greater proportion of suitable area for potato cultivation to regions with smaller proportion of potato-suitable area, before and after potatoes were adopted in the Old World. In our first estimation, we not only compare population and urbanization levels between Old World countries with more or less suitable area for potato cultivation before and after the diffusion of potatoes in the Old World, but also compute the role of malaria endemic in offsetting the positive impacts of potato cultivation on population and urbanization levels for each Old Wold country after 1700. Our results show that the positive effect of the diffusion of potato in the Old World found in Nunn and Qian (2011) is higher in countries with lower risk of malaria transmission.

In our second estimation strategy, we focus on the *local* impact of potato suitability rather than its country level impact. We compare population and urbanization levels between each grid, measuring 0.5° latitude by 0.5° longitude§. The use of disaggregated data allows for a more robust analysis of different areas within the same country with similar institutions, policies, and other country specific characteristics. Our results at the grid level confirm the conclusions obtained with the country-level analysis: the diffusion of potatoes to the Old World has positively impacted population and urbanization levels for grids which were more suitable for growing the staple crop. Specifically, the second estimation strategy shows that the grids highly suitable for potato and not malaria endemic experienced a 3.2% increase in urbanization compared to those which were not suitable for growing potato. Nevertheless, the positive impact of growing potatoes decreases as the endemicity of malaria increases. For example, the grids that were suitable for growing potatoes but mildly malaria endemic experienced an increase in urbanization levels by only 1.4%. As the endemicity of malaria increases

‡Nunn and Qian (2011) use an older version of GAEZ potato suitability, dated 2002. We use the updated version from ???.

§0.5° latitude by 0.5° longitude cover an area of approximately 56 km^2 at the equator the impact of potato suitability decreases until it becomes negative suggesting that the diffusion of malaria fully offset the benefits of cultivating potato.

3.1 History of Diseases and Development

Macroeconomic effects of health conditions have been investigated in the literature Acemoglu and Johnson (2007). This study is the first to consider an exogenous measure of life expectancy. Using an IV strategy the authors proxy life expectancy with mortality rates from the most common and deadly diseases during 1940-1980 period. They conclude that although the reduction in mortality rates has increased the population levels, the same does not hold for income per capita. In particular, the increase of health conditions does have a positive impact on the population levels in the short run. however, since the land is fix in the short run, the effects on income per capita are slightly negative.

3.1.1 Malaria

Malaria has been one of the major infectious disease throughout 16-20th century. It was not until the 20^{th} century when technological advancements and the discovery of new drugs along with modern chemical components made possible to effectively prevent the transmission of malaria. An example of the effectiveness of early control policies was the construction of the Panama Canal in the 1910s. Malaria was a major cause of death and illness among workers in the area. According to the Center for Disease Control Prevention (CDC), in 1906 there were over 26,000 employees working on the Canal. Of these, over 21,000 (i.e. more than 80%) were hospitalized for malaria at some time during their work. By 1912, there were over 50,000 employees, and the number of hospitalized workers had decreased to approximately 5,600 (i.e. 11%).

Bowden et al. (2008), Weil (2014) and Hansen and Lønstrup (2015) analyze the role of malaria on health conditions of African farmers and economic growth finding that whilst the control of malaria did have an economic effect, the escape from poverty was not explained by the eradication of malaria. More specifically on the role of malaria on population levels, Tatem et al. (2013) and Gooch (2017) employing new data at sub-national level find that malaria eradication programs which started in early 1900s in Europe and the USA increased population and urbanization rates of areas which were more suitable for the transmission of malaria. Whilst economic literature on the aggregate effects of malaria and health conditions on income per capita converges on the absence or negative effects of better health to economic growth, opposite results are obtained from micro analysis.

Conley et al. (2007) look for the underlying causes of the hitherto only mildly observed demographic transition in sub-Saharan Africa, with both mortality and fertility rates still being high. The authors exploit the exogenous variation in the ecology of malaria transmission and in agricultural productivity through the staggered introduction of high-yield seed varieties. They find an important role of farm productivity - instrumented on malaria ecology - on aggregate total fertility rates. Bleakley (2010a), Cutler et al. (2010) study the impact of different disease control policies on human capital providing large evidence of the positive impact of disease prevention campaigns on *individual* education levels which are widely regarded as fundamentals of persistent economic growth. Similar results were also obtained in Bleakley (2007) and Percoco (2013). More specifically, Bleakley (2003) and Bleakley (2010b) show that level of income of adults not exposed to tropical diseases during their childhood was higher compared to those exposed to weather conditions suitable for transmission of infectious diseases.

Early studies such as Picard and Mills (1992), McCarthy et al. (2000) and Sachs and Malaney (2002) have shown a strong and negative correlation between malaria endemic and per-capita gross domestic product (GDP). However, as addressed by the authors, the relationship between economic growth and malaria greatly suffers from reverse causality issues. More recent studies such as Allen et al. (2014) and Wouterse and Badiane apply parametric and non nonparametric methods focusing in Burkina Faso and Tanzania to show the positive effects of increasing health conditions of farmers on agricultural labor productivity. Finally, exploiting geographic variations in the stability of malaria, Malpede (2019) investigates the causal relationship between the reduction in malaria transmission occurred in the early 1900s and historical farm productivity in the United States. He concludes that the eradication of malaria led to approximately one fifth of the agricultural productivity growth in the U.S. counties.

3.2 Data

The main data sources used in the empirical framework are summarised in Table 3.1. For the country-level analysis, spatially-explicit datasets are summarised within country boundaries, as described in the corresponding Section below. Non-gridded control variables, namely the national extent of cropland and tropical area, the elevation, the distance from the equator and to the nearest coast, are drawn from Nunn and Qian (2011)'s original replication data. Figures 3.2 and 3.3 plot the spatial heterogeneity in the two main variables of interest, namely the potato suitability index (PSI) and the malaria stability index (MSI), respectively.

Table 3.1: Table of data inputs for the econometric identification

Dataset	Source	Time resolution	Spatial resolution
Country shapefiles	Hijmans et al. (2010)	-	-
Crop suitability indexes	Fischer et al. (2012)	Time-invariant	0.008°
Malaria stability index	Kiszewski et al. (2004)	Time-invariant	0.008°
Historical population and urbanisation	Klein Goldewijk et al. (2011)	100/50/10 years	0.008°
Terrain ruggedness index	Nunn and Puga (2012)		

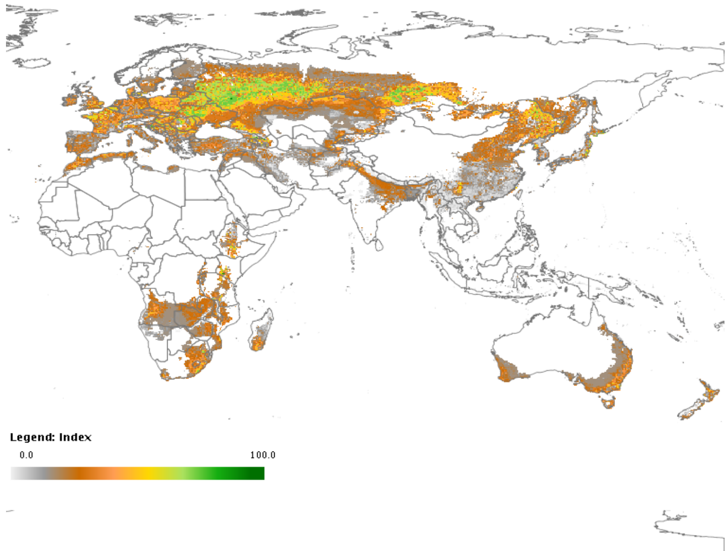


Figure 3.2: Potato Suitability Index

The PSI captures the potential suitability for cultivating potatoes based on regional weather and land conditions. (Fischer et al., 2012). This is intersected with Old World country borders.

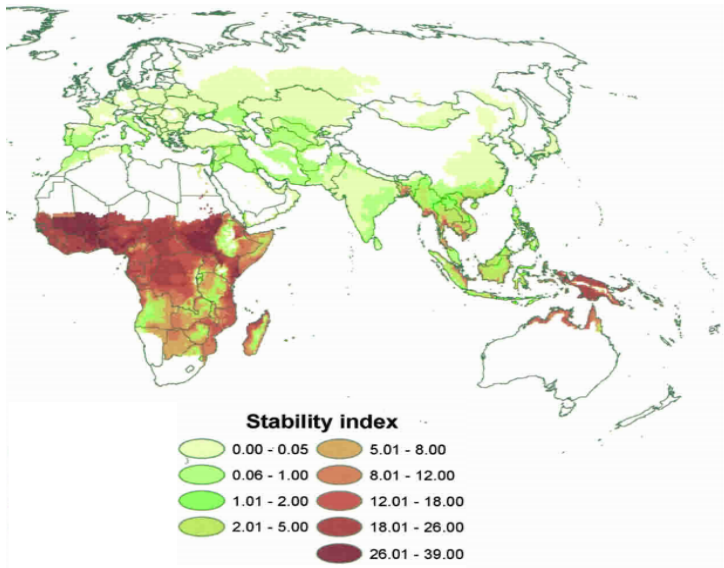


Figure 3.3: Malaria Suitability Index

The MSI captures the potential stability of malaria transmission based on regionally dominant vector mosquitoes, temperature and precipitation data set (Kiszewski et al., 2004). This is intersected with Old World country borders.

3.2.1 Country-level analysis

For country-level modelling, data is extracted with the following procedure. The potato suitability index layer is downloaded from the FAO Global Agro-Ecological Zones database (Fischer et al., 2012) with medium-input, rainfed, historical average parameters. Also the malaria stability index is retrieved. The two files are imported in Google Earth Engine (Gorelick et al., 2017), where the area of (i) zones with potato suitability greater than 40% (as in Nunn and Qian (2011)) and (ii) zones with potato suitability greater than 40% and a malaria Stability index lower than 1, 0.1, 0.06, 0.05, 0.01, and 0 is calculated. The GADM (Global Administrative Unit Layer, (Hijmans et al., 2010)) are used as the reference shapefile for national borders.

3.2.2 Grid-level analysis

Grid-level data on population and urbanization rates for years between 1100 and 1980 (at time steps of 100 years until 1700, 50 years from 1700 to 1900, and 10 years from 1900 to 1980) from the HYDE database (Klein Goldewijk et al., 2011) is extracted into a global regular grid shapefile with a resolution of 0.5°; a spatial join was also performed between the regular grid and the GADM shapefile to create a country attribute in the regular grid shapefile matching the underlying country name. Zonal statistics for the mean malaria Stability index and potato suitability, as well as for the suitability of an array of control crops, namely cassava, wheat, maize, and barley (again, with medium-input, rainfed, historical average parameters), are produced for each grid cell.

A possible relevant issue arising with the use of grid-level estimated historical population data from HYDE is the measurement error of the latter. For instance, estimated historical population and urbanization data for country A might be very highly inaccurate, thus having adverse effects on the performance of the indicators (Millimet, 2011, Gooch, 2017). We address this concern by including country specific time trends in the grid-level empirical strategy. This allows for a comparison of each grid within the same country. Second both the potato suitability and malaria ecology indexes are void of this problem being them time invariant and computed using weather and soil variables.

3.3 Empirical framework

Country-level estimation strategy:

$$y_{i,t} = \beta_1 \, ln \; Potato \; Area_i \cdot I_t^{Post} + \beta_2 \, ln \; Potato \; NoMal \; Area_i \cdot I_t^{Post} +$$

$$+ \text{”}\beta_3 \, ln \; Potato \; Mal \; Area_i \cdot I_t^{Post} + \sum_{j=1000}^{1900} \Omega_j \mathbf{X}'_i I_t^j + \sum_{j=1000}^{1900} \delta I_t^j + \sum_{c=1}^{130} \gamma_c I_i^c + \varepsilon_{i,t}, \quad (3.1)$$

where index i represents each Old World country. A total of 130 countries were considered in the analysis, indexed with *c*. Index *t* represents time periods which are part of the analysis (i.e. 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1750, 1800, 1850 and 1900). Years from 1000 to 1600 are part of the pre-potato introduction period as in Nunn and Qian (2011). Since potato became widely diffused in the Old World soon after 1700, 1700 is considered the start of the post intervention period, which ranges from 1700 until 1900[¶]. The dependent variable is represented as $y_{i,t}$ which is the natural log of one of the two proxies of historical socio-economic development, i.e. population and city population share for each Old World country *i* at time *t*.

The variable $ln\ Potato\ Area_i$ is the time invariant area suitable for potato cultivation for each country *i*. As in Nunn and Qian (2011) β_1 shows the estimated impact of potato suitable area on population and city population share after its adoption in the Old World. A positive and statistical significant value of β_1 is interpreted as an additional increase in population and urbanization of Old World countries with higher potato suitable land between 1700 and 1900 compared to countries with less amount of potato suitable land.

Beside computing the impact of the amount of the area suitable for potato cultivation for each country to population and urbanization, our procedure allows to disentangle the impact of the amount of the country area suitable for the cultivation of potato *but not malaria endemic* with the country area suitable for potato *and malaria endemic*. To this regard the variable $ln\ Potato\ NoMal\ Area_i$ is the time invariant area suitable for potato cultivation but not malaria endemic for each country *i*. The coefficient of our interest is β_2, which is estimated impact of the adoption of potato in non malaria endemic and potato suitable land rather than on mere potato suitable land. On the other hand the variable $ln\ Potato\ Mal\ Area_i$ is the time invariant area suitable for potato cultivation for each country *i* and at the same time also suitable for a stable transmission of malaria. We expect β_2

[¶]In the original paper, Nunn and Qian (2011) further test 1700 as the only valid cut off date by performing the empirical analysis using alternative cut offs.

to be positive and greater in magnitude then respectively β_1 and β_3. In particular we would expect that $\beta_2 > \beta_1 > \beta_3$. For concreteness, such situation would mean that the cultivation of potato after potatoes were introduced in 1700 had a greater impact on population and urbanization on countries that beside having a geographic environment more suitable for growing potatoes, were not malaria endemic as well. Moreover, a coefficient β_3 being the smallest in magnitude would mean that the adoption of potatoes had a small impact on those countries that had a large amount of potato suitable land which at the same time was malaria endemic.

Variable I_t^{Post} is a post treatment dummy variable which takes value 1 for years after 1700 (i.e. 1700, 1750, 1800, 1850 and 1900) while value 0 for years before the exogenous introduction of potato in the Old World (i.e. 1000, 1100, 1200, 1300, 1400, 1500 and 1600). This specification also includes country fixed effects $\sum_{c=1}^{130} \gamma_c I_t^c$, where c indicates the set of Old World countries constituted by a total of 130 countries and time period fixed effects $\sum_{t=1000}^{1900} \partial I_t$.

As argued in Nunn and Qian (2011) European countries have, on average, experienced a higher growth in population and urbanization between 1700 and 1900 for reasons related to socio-economic and technological progress which are beyond the cultivation of potato. However, European countries were also naturally more suitable to growing potatoes. It would be, therefore necessary to estimate the effects of introducing the cultivation of potatoes within continent variation only[‖]. Hence, in order to allow for a comparison of each country within the same continent we add to equation 3.1 continent fixed effects interacted with time-period fixed effects. The introduction of *Continent x Year FE* to equation 3.1 changes the interpretation of our coefficients of interest β_1, β_2 and β_3, which are now identified from within-continent variation only.

$\sum_{j=1000}^{1900} \Omega_j \mathbf{X}'_i I_t^j$ represents time invariant country-specific characteristics interacted with time period fixed effects to take into account other country-specific characteristics aside from the culti-

[‖]In other words we do not want to compare an European country such as Italy with an African country, but comparing an European country with another European country makes the estimation strategy more robust and allows to take into account all the other factors that may have caused European divergence vation of potato that might have affected population and urbanization growth between 1700 and 1900. As country-level controls we use the set of relevant geographical and historical country specific characteristics which might have affected population and urbanization between 1000 and 1900 used in Nunn and Qian (2011): i.e. country elevation, total cropland area, ruggedness, tropical area, distance from the equator and from the nearest coast, an indicator variable taking value one if a country is an exporter of potatoes along with total areas of other crops, i.e. maize, silage, sweet potato and cassava.

Tables S1 and S2 show the estimates from our main estimating equation 3.1 respectively without the inclusion of additional controls and with the inclusion of all country and geographical specific control variables. All columns include country and time-period fixed effects along with all control variables used in the analysis as described in section 3.3 i.e. controls for land suitable for Old World staple crops interacted with time-period fixed effects and controls for ruggedness, elevation, and tropics, each interacted with the time-period fixed effects

Results shown in Tables S1 and S2 confirm our prediction that suggested that overall the introduction of potatoes increased total population and urbanization, however the presence of weather conditions particularly favorable to the transmission of malaria counteracted the positive benefits of potatoes.

This is illustrated by the estimated coefficient of the potato suitability interaction term, $ln\ Potato\ Area_i \cdot I_t^{Post}$, which displays the average increase in population and urbanization levels respectively arising from the widely diffusion of potato in the Old World after 1700. According to the estimates in column (1) and (4), a 1 percent increase in the amount of land suitable for the cultivation of potato increased population by 0.014 percent on average and the urban population by 0.002 percentage points after 1700. This results is with no surprise very similar to what shown in Nunn and Qian (2011)[**]. However, the following two rows of tables S1 and S2 show that the positive impact of

[**]The slight differences are due to the fact that Nunn and Qian (2011) use a version of potato suitability

the introduction of potato in the Old World was entirely appreciated in non malaria endemic areas while the introduction of potato did not have any positive effect on malaria endemic areas.

index dated 2002 while the version used to perform the following analysis is the most recent available, dated 2011.

Table 3.2: The impact of non malarious potato suitable area on population and urbanization: Continent FE estimates

	Pop	Pop	Pop	Urb	Urb	Urb
	b/se	b/se	b/se	b/se	b/se	b/se
ln Potato Area x Post	0.014			0.002**		
	(0.011)			(0.001)		
ln Non Malarious Potato Area x Post		0.025*			0.002*	
		(0.015)			(0.001)	
ln Malarious Potato Area x Post			-0.008			0.001
			(0.015)			(0.001)
Old World	Yes	Yes	Yes	Yes	Yes	Yes
Elevation	Yes	Yes	Yes	Yes	Yes	Yes
Tropical	Yes	Yes	Yes	Yes	Yes	Yes
Rugged	Yes	Yes	Yes	Yes	Yes	Yes
Continent X Year	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1528	1528	1528	1528	1528	1528
Adjusted R^2	0.990	0.990	0.990	0.400	0.399	0.397

Table 3.3: The impact of non malarious potato suitable area on population and urbanization: Continent FE estimates with additional controls

	Pop	Pop	Pop	Urb	Urb	Urb
	b/se	b/se	b/se	b/se	b/se	b/se
ln Potato Area x Post	0.010			0.004***		
	(0.013)			(0.001)		
ln Non Malarious Potato Area x Post		0.019			0.004***	
		(0.015)			(0.001)	
ln Malarious Potato Area x Post			-0.001			-0.000
			(0.017)			(0.001)
Old World	Yes	Yes	Yes	Yes	Yes	Yes
Elevation	Yes	Yes	Yes	Yes	Yes	Yes
Tropical	Yes	Yes	Yes	Yes	Yes	Yes
Rugged	Yes	Yes	Yes	Yes	Yes	Yes
Coast distance	Yes	Yes	Yes	Yes	Yes	Yes
Equator distance	Yes	Yes	Yes	Yes	Yes	Yes
Exporter	Yes	Yes	Yes	Yes	Yes	Yes
Continent X Year	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1528	1528	1528	1528	1528	1528
Adjusted R^2	0.991	0.991	0.991	0.456	0.456	0.447

The results shown in Tables S1 and S2 emphasise the beneficial effects that the introduction of potatoes had on the increase of population and urbanization of countries with greater land naturally suitable to growing potatoes. This first strategy we employ, however does not allow to estimate the eventual impact of a higher suitability for growing potato on population and urbanization at a very disaggregated level. In the next section we employ a second econometric strategy that permits to estimate not only the effects of a higher suitability for potato cultivation on population and urbanization after 1700 *locally* but also if the relative level of stability of malaria transmission counteracted the positive effects of cultivating potatoes.

Grid-level estimation strategy: The second strategy that we employ allows to estimate

$$y_{i,c,t} = \beta \, \mathrm{PSI}_i \text{ x } \mathrm{MSI}_i \cdot I_t^{Post} + \sum_{t=1000}^{1900} \Omega \, \mathbf{X}'_{i,c} I_t + \sum_{t=1000}^{1900} \partial I_t + \sum_p \gamma_p I_t^p + \varepsilon_{\,i,c,t} \qquad (3.2)$$

Equation 3.2 is very similar to equation 3.1 except that we do not compare population and urbanization levels between countries but between grids measuring 0.5° altitude by 0.5° longitude. This specification not only allows for more robust results but also for understanding if the cultivation of potato after its diffusion in the Old World had a local impact on population and urbanization levels as well, and if the presence of weather conditions suitable for the transmission of malaria counteracted the positive effects of potato. In equation 3.2 index i represents each Old World grid and t indexes time periods considered in the analysis (i.e. 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1750, 1800, 1850 and 1900). Years from 1000 to 1600 are part of the pre-potato introduction period as in Nunn and Qian (2011). Once again 1700 is considered the start of the post diffusion of potato. Therefore the post intervention period ranges from 1700 until 1900. The dependent variable is represented as $y_{i,c,t}$ which is the natural log of one of the two proxies of historical socio-economic development, i.e. population and city population share for each Old World grid *i* in country *c* at time *t*.

The variable $\text{PSI}_i \times \text{MSI}_i$ is the combination of the time invariant suitability index for potato cultivation (PSI) and its degree of stability of malaria (MSI) for each grid i. Our variable of interest $\text{PSI}_i \times \text{MSI}_i$ shows the impact on population and urbanization levels of the suitability of potato in non malaria endemic grids versus malaria endemic grids. If malaria effectively counteracted the positive effects of potato on the development of Old World countries then we would expect to find a positive effect of the PSI in non malaria endemic grids while disappearing in grids with high MSI.

Again, the variable I_t^{Post} is a post adoption dummy variable which takes value 1 for years after the exogenous spreading of potato in the Old World (i.e. 1700, 1750, 1800, 1850 and 1900) while value 0 for years ex ante the exogenous introduction of potato in the Old World (i.e. 1000, 1100, 1200, 1300, 1400, 1500 and 1600). Equation 3.2 also includes grid fixed effects $\sum_p \gamma_p I_t^p$, where p indicates the set of Old World grids; time period fixed effects $\sum_{t=1000}^{1900} \delta I_t$ and country specific fixed effects which allow for a comparison of each grid within the same country. As before $\mathbf{X}_{i,s}$ represents vectors of time invariant grid-specific controls included in the regression. As grid-level controls I use a set of relevant geographical and historical characteristics which might have affected population and urbanization: i.e. total cropland area per grid, ruggedness level and total areas of other crops, i.e. maize, silage, sweet potato and cassava.

3.4 Estimated local impact

3.4.1 Grid-level Analysis, Pre Malaria Eradication

Results: Grid-level Analysis, Pre Eradication of Malaria: Old World

Table 3.4: The impact of malarious and non malarious potato suitability on urbanization: Old World, Pre Malaria Eradication, Gridded Analysis

	Urbanization b/se	Urbanization b/se	Urbanization b/se	Population b/se	Population b/se	Population b/se
potato_dummy_post	0.049***			1.531***		
	(0.004)			(0.053)		
potato_dummy_negmal_post	0.001***			-0.045***		
	(0.000)			(0.002)		
PSI x No Mal x Post		0.015***			-0.848***	
		(0.001)			(0.039)	
PSI x Mal x Post			-0.005***			0.714***
			(0.002)			(0.041)
Barley	Yes	Yes	Yes	Yes	Yes	Yes
Maize	Yes	Yes	Yes	Yes	Yes	Yes
Cassava	Yes	Yes	Yes	Yes	Yes	Yes
Wheat	Yes	Yes	Yes	Yes	Yes	Yes
Grid X Year	Yes	Yes	Yes	Yes	Yes	Yes
Observations	541178	521004	521004	473154	453200	453200

Results: Grid-level Analysis, Pre Eradication of Malaria: Africa and South East Asia

Table 3.5: The impact of malarious and non malarious potato suitability on urbanization: Africa, Pre Malaria Eradication, Gridded Analysis

	Urbanization b/se	Urbanization b/se	Urbanization b/se	Population b/se	Population b/se	Population b/se
potato_dummy_post	0.024***			2.008***		
	(0.008)			(0.111)		
potato_dummy_negmal_post	0.001***			-0.038***		
	(0.000)			(0.002)		
PSI x No Mal x Post		0.016***			-0.628***	
		(0.001)			(0.037)	
PSI x Mal x Post			-0.008***			0.813***
			(0.001)			(0.039)
Barley	Yes	Yes	Yes	Yes	Yes	Yes
Maize	Yes	Yes	Yes	Yes	Yes	Yes
Cassava	Yes	Yes	Yes	Yes	Yes	Yes
Wheat	Yes	Yes	Yes	Yes	Yes	Yes
Grid X Year	Yes	Yes	Yes	Yes	Yes	Yes
Observations	293810	291907	291907	275985	274148	274148

3.5 Robustness

3.5.1 The effect of malaria after its eradication (1900s)

$$y_{i,c,t} = \beta \, \mathrm{PSI}_i \text{ x } \mathrm{MSI}_i \cdot I_t^{Post} + \sum_{t=1000}^{1980} \Omega \, \mathbf{X}'_{i,c} I_t + \sum_{t=1000}^{1980} \delta I_t + \sum_p \gamma_p I_t^p + \sum_{t=1000}^{1980} \sum_q \omega_q I_t^q + \varepsilon_{\, i,c,t} \quad (3.3)$$

The specification expressed in equation 3.3 is identical to equation 3.2 except that in here time periods, indexed with t, range from 1000 to 1980[††]. Years from 1000 to 1900 are part of the pre-malaria eradication period while years from 1910 to 1980 are part of the post-malaria eradication period. Therefore, in equation 3.3, variable I_t^{Post} is a post treatment dummy variable which takes value 1 for years after 1900 (i.e. from 1910 to 1980) while value 0 for years before the exogenous intervention against malaria (i.e. 1000, 1100, 1200, 1300, 1400, 1500 and 1600). This specification also includes grid fixed effects $\sum_p \gamma_p I_t^p$, where p indicates the set of Old World countries; time period fixed effects $\sum_{t=1000}^{1900} \delta I_t$ and country specific fixed effects which allow for a comparison of each grid within the same country. As before $\mathbf{X}_{i,s}$ represents vectors of time invariant grid-specific controls included in the regression. As grid-level controls I use a set of relevant geographical and historical characteristics which might have affected population and urbanization: i.e. total cropland area per grid, ruggedness level and total areas of other crops, i.e. maize, silage, sweet potato and cassava.

[††]Time periods are respectively 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1750, 1800, 1850, 1900, 1910, 1920, 1930, 1940, 1950, 1960, 1970 and 1980

Results: Gridded Analysis Post Eradication of Malaria: Old World

Table 3.6: The impact of malarious and non malarious potato suitability on urbanization: Old World, post1900Malaria Eradication, Gridded Analysis

	Urbanization b/se	Urbanization b/se	Urbanization b/se	Population b/se	Population b/se	Population b/se
potato_dummy_post1900	0.070***			1.819***		
	(0.008)			(0.055)		
potato_dummy_negmal_post1900	-0.000			-0.046***		
	(0.000)			(0.002)		
PSI x Mal x Post		0.013***			-1.041***	
		(0.003)			(0.041)	
pot_mal_post1900			0.018***			0.682***
			(0.003)			(0.044)
Barley	Yes	Yes	Yes	Yes	Yes	Yes
Maize	Yes	Yes	Yes	Yes	Yes	Yes
Cassava	Yes	Yes	Yes	Yes	Yes	Yes
Wheat	Yes	Yes	Yes	Yes	Yes	Yes
Grid X Year	Yes	Yes	Yes	Yes	Yes	Yes
Observations	934762	899916	899916	816641	782175	782175

Results: Gridded Analysis Post Eradication of Malaria: Africa and South East Asia

Table 3.7: The impact of malarious and non malarious potato suitability on urbanization: Africa, post1900malaria eradication, Gridded Analysis

	Urbanization b/se	Urbanization b/se	Urbanization b/se	Population b/se	Population b/se	Population b/se
potato_dummy_post1900	0.074***			0.000		
	(0.015)			(.)		
potato_dummy_negmal_post1900	-0.000**			0.000		
	(0.000)			(.)		
PSI x Mal x Post		0.010***			0.000	
		(0.003)			(.)	
pot_mal_post1900			0.017***			0.000
			(0.003)			(.)
Barley	Yes	Yes	Yes	Yes	Yes	Yes
Maize	Yes	Yes	Yes	Yes	Yes	Yes
Cassava	Yes	Yes	Yes	Yes	Yes	Yes
Wheat	Yes	Yes	Yes	Yes	Yes	Yes
Grid X Year	Yes	Yes	Yes	Yes	Yes	Yes
Observations	507490	504203	504203	293810	291907	291907

3.6 Conclusion

This third chapter of this dissertation has focused on the interactions between technological progress and poor health conditions.

The main contribution of this study was that to rigorously quantify the counter effect of poor health conditions to an initial positive technological shock. We know from the seminal paper by Nunn and Qian (2011) that the introduction of potato in the Old World has contributed to the increase in population and urbanization of more potato suitable countries, from 1700 to 1900, however with this paper we wanted to know if in the absence of malaria the positive effects of such agricultural shock could have been even higher.

We, therefore used the introduction of the cultivation of potato in the Old World, which occurred in 1700, and interacted time invariant climate and soil conditions which make a land more or less suitable to the cultivation of potato, with time invariant weather conditions which are more favourable to the reproduction of mosquitoes larvae and in turn to stable prevalence of malaria to quantify if the positive impact of the cultivation of potato on population and urbanization as shown in the seminal paper by Nunn and Qian (2011) was partly absorbed by greater prevalence of malaria.

We compared population and urbanization levels of areas which were equally suitable for the cultivation of potato but different in terms of malaria ecology index.

Our results at the grid level first confirm the conclusion obtained with the country-level analysis in Nunn and Qian (2011) that is that the diffusion of potato to the Old World has positively impacted population and urbanization of grids which were more suitable for growing the staple crop. Specifically, grid-level results show that the grids highly suitable for potato experienced a 3.2% increase in urbanization compared to those which were not suitable for growing potato. Nevertheless, the positive impact of growing potatoes decreases as the endemicity of malaria increases. For instance, the grids that were suitable for growing potatoes but mildly malaria endemic experienced an increase in urbanization levels by only 1.4%. As the endemicity of malaria increases the impact of potato suitability decreases until it becomes null, suggesting that the diffusion of malaria fully offset the benefits of cultivating potato.

References

Acemoglu, D. and Johnson, S. (2007). Disease and development: the effect of life expectancy on economic growth. *Journal of political Economy*, 115(6):925–985.

Allen, S., Badiane, O., Sene, L., and Ulimwengu, J. (2014). Government expenditures, health outcomes and marginal productivity of agricultural inputs: The case of tanzania. *Journal of agricultural economics*, 65(3):637–662.

Alsan, M. (2015). The effect of the tsetse fly on african development. *American Economic Review*, 105(1):382–410.

Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of political economy*, 113(1):151–184.

Alves Dias, P., Blagoeva, D., Pavel, C., and Arvanitidis, N. (2018). Cobalt: demand-supply balances in the transition to electric mobility. *European Commission, Joint Research Centre, EUR-Scientific and Technical Research Reports. Publications Office of the European Union. DOI*, 10:97710.

Amnesty International (2016). This is What We Die For. Human Rights Abuses in the Democratic Republic of the Congo Power the Global Trade in Cobalt. p//wpgedaf62/3183/2016/e [illegible]. [Online; accessed 9-July-2019].

Amnesty International (2017). *Time to Recharge: Corporate Action and Inaction to Tackle Abuses in the Cobalt Supply Chain* . Amnesty International.

Angrist, J. D. and Kugler, A. D. (2008). Rural windfall or a new resource curse? coca, income, and civil conflict in colombia. *The Review of Economics and Statistics*, 90(2):191–215.

Aragón, F. M. and Rud, J. P. (2015). Polluting industries and agricultural productivity: Evidence from mining in ghana. *The Economic Journal*, 126(597):1980–2011.

Ashraf, Q. H., Lester, A., and Weil, D. N. (2008). When does improving health raise gdp? *NBER macroeconomics annual*, 23(1):157–204.

Atkin, D. (2016). Endogenous skill acquisition and export manufacturing in mexico. *American Economic Review*, 106(8):2046–85.

Banza, C. L. N., Nawrot, T. S., Haufroid, V., Decrée, S., De Putter, T., Smolders, E., Kabyla, B. I., Luboya, O. N., Ilunga, A. N., Mutombo, A. M., et al. (2009). High human exposure to cobalt and other metals in katanga, a mining area of the democratic republic of congo. *Environmental research*, 109(6):745–752.

Basu, K. and Van, P. H. (1998). The economics of child labor. *American economic review*, pages 412–427.

Beegle, K., Dehejia, R., and Gatti, R. (2004). Why should we care about child labor? the education, labor market, and health consequences of child labor. Technical report, National Bureau of Economic Research.

Benshaul-Tolonen, A. (2018). Local industrial shocks and infant mortality. *The Economic Journal*, 129(620):1561–1592.

Berman, N., Couttenier, M., Rohner, D., and Thoenig, M. (2017). This mine is mine! how minerals fuel conflicts in africa. *American Economic Review*, 107(6):1564–1610.

Bleakley, H. (2003). Disease and development: Evidence from the american south. *Journal of the European Economic Association*, 1(2-3):376–386.

Bleakley, H. (2007). Disease and development: evidence from hookworm eradication in the american south. *The quarterly journal of economics*, 122(1):73–117.

Bleakley, H. (2009). Economic effects of childhood exposure to tropical disease. *American Economic Review*, 99(2):218–23.

Bleakley, H. (2010a). Health, human capital, and development. *Annu. Rev. Econ.*, 2(1):283–310.

Bleakley, H. (2010b). Malaria eradication in the americas: A retrospective analysis of childhood exposure. *American Economic Journal: Applied Economics*, 2(2):1–45.

Bleakley, H., Costa, D., and Lleras-Muney, A. (2014). Health, education, and income in the united states, 1820–2000. In *Human Capital in History: The American Record*, pages 121–159. University of Chicago Press.

Bleakley, H. and Hong, S. C. (2017). Adapting to the weather: lessons from us history. *The journal of economic history*, 77(3):756–795.

Bowden, S., Michailidou, D. M., and Pereira, A. (2008). Chasing mosquitoes: An exploration of the relationship between economic growth, poverty and the elimination of malaria in southern europe in the 20th century. *Journal of International Development: The Journal of the Development Studies Association*, 20(8):1080–1106.

Brown, G. (1986). Control and eradication of malaria: Past, present and future.

Bustos, P., Caprettini, B., and Ponticelli, J. (2016). Agricultural productivity and structural transformation: Evidence from brazil. *American Economic Review*, 106(6):1320–65.

Caminade, C., Kovats, S., Rocklov, J., Tompkins, A. M., Morse, A. P., Colón-González, F. J., Stenlund, H., Martens, P., and Lloyd, S. J. (2014). Impact of climate change on global malaria distribution. *Proceedings of the National Academy of Sciences*, 111(9):3286–3291.

Carrillo, B. (2019). Present bias and underinvestment in education? long-run effects of childhood exposure to booms in colombia. *Long-Run Effects of Childhood Exposure to Booms in Colombia (March 16, 2019)*.

Center for International Forestry Research (2013). The formalisation of artisanal mining in the democratic republic of the congo and rwanda. Technical report, Center for International Forestry Research.

Charles, K. K., Hurst, E., and Notowidigdo, M. J. (2018). Housing booms and busts, labor market opportunities, and college attendance. *American Economic Review*, 108(10):2947–2994.

Cobalt Institute (2019). Data on Production and Supply of Cobalt. [illegible] [Online; accessed 19-July-2019].

Conley, D., McCord, G. C., and Sachs, J. D. (2007). *Africa's lagging demographic transition: Evidence from exogenous impacts of malaria ecology and agricultural technology*.

Cust, J. (2015). The spatial effects of resource extraction: mining in indonesia. Technical report, Oxcarre Discussion Paper, University of Oxford.

Cutler, D., Fung, W., Kremer, M., Singhal, M., and Vogl, T. (2010). Early-life malaria exposure and adult outcomes: Evidence from malaria eradication in india. *American Economic Journal: Applied Economics*, 2(2):72–94.

Cuvelier, J. (2017). Money, migration and masculinity among artisanal miners in katanga (dr congo). *Review of African Political Economy*, 44(152):204–219.

Darton Commodities (2017). Annual market review on cobalt demand. p//wddifsoKbl. Accessed: 2018-11-12.

De Herdt, T. and Titeca, K. (2016). Governance with empty pockets: the education sector in the democratic republic of congo. *Development and Change*, 47(3):472–494.

De Putter, T., Decrée, S., Banza, C. L. N., and Nemery, B. (2011). Mining the katanga (drc) copperbelt: geological aspects and impacts on public health and the environment—towards a holistic approach. In *Mining and the Environment in Africa. Proceedings of the Inaugural Workshop, IGCP/SIDA*, number 594, pages 14–17.

Deaton, A. (2013). *The great escape: health, wealth, and the origins of inequality*. Princeton University Press.

Demographic and Health Surveys (2014). The dhs program, usaid. *Calverton: Measure DHS*.

Derryberry, O. and Gartrell, F. (1952). Trends in malaria control program of the tennessee valley authority1. *The American journal of tropical medicine and hygiene*, 1(3):500–507.

Doepke, M. and Zilibotti, F. (2005). The macroeconomics of child labor regulation. *American Economic Review*, 95(5):1492–1524.

Dube, O. and Vargas, J. F. (2013). Commodity price shocks and civil conflict: Evidence from colombia. *The Review of Economic Studies*, 80(4):1384–1421.

Easterly, W. and Levine, R. (2016). The european origins of economic development. *Journal of Economic Growth*, 21(3):225–257.

Eckstein, Z. and Wolpin, K. I. (1999). Why youths drop out of high school: The impact of preferences, opportunities, and abilities. *Econometrica*, 67(6):1295–1339.

Edmonds, E. V. and Pavcnik, N. (2005). Child labor in the global economy. *Journal of Economic Perspectives*, 19(1):199–220.

EPA (1975). Ddt: A review of scientific and economic aspects of the decision to ban its use as a pesticide.

Euractiv (2019). Europe will have at least ten gigafactories. [illegible]. Accessed: 2019-04-05.

Faber, B., Krause, B., and De La Sierra, R. S. (2017). Artisanal mining, livelihoods, and child labor in the cobalt supply chain of the democratic republic of congo.

Financial Times (2019). Congo, child labour and your electric car. Technical report, Financial Times.

Fischer, G., Nachtergaele, F. O., Prieler, S., Teixeira, E., Toth, G., van Velthuizen, H. T., Verelst, L., and Wiberg, D. (2012). Global agro-ecological zones (gaez v3. 0).

for Disease Control Prevention, U. C. (2010). The history of malaria, an ancient disease.

Gallup, J. L. and Sachs, J. D. (2001). The economic burden of malaria. *The American journal of tropical medicine and hygiene*, 64(1_suppl):85–96.

Galor, O. and Weil, D. N. (2000). Population, technology, and growth: From malthusian stagnation to the demographic transition and beyond. *American economic review*, 90(4):806–828.

Ghosh Banerjee, S., Romo, Z., McMahon, G., Toledano, P., Robinson, P., and Pérez Arroyo, I. (2014). *The power of the mine: A transformative opportunity for sub-Saharan Africa*. The World Bank.

Giuliano, P. and Nunn, N. (2013). The transmission of democracy: From the village to the nation-state. *The American Economic Review*, 103(3):86.

Gollin, D., Lagakos, D., and Waugh, M. E. (2014). Agricultural productivity differences across countries. *American Economic Review*, 104(5):165–70.

Gooch, E. (2017). The impact of reduced incidence of malaria and other mosquito-borne diseases on global population. *Journal of Development Economics*, 124:214–228.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017). Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202:18–27.

Haines, M. R. et al. (2005). *Historical, demographic, economic, and social data: The United States, 1790-2002*. Inter-university Consortium for Political and Social Research.

Hansen, C. W. and Lønstrup, L. (2015). The rise in life expectancy and economic growth in the 20th century. *The Economic Journal*, 125(584):838–852.

Hausmann, R. and Rigobon, R. (2003). An alternative interpretation of the'resource curse': Theory and policy implications. Technical report, National Bureau of Economic Research.

Hay, S. I., Guerra, C. A., Tatem, A. J., Noor, A. M., and Snow, R. W. (2004). The global distribution and population at risk of malaria: past, present, and future. *The Lancet infectious diseases*, 4(6):327–336.

Hazan, M. and Berdugo, B. (2002). Child labour, fertility, and economic growth. *The Economic Journal*, 112(482):810–828.

Henderson, J. V., Squires, T., Storeygard, A., and Weil, D. (2017). The global distribution of economic activity: nature, history, and the role of trade. *The Quarterly Journal of Economics*, 133(1):357–406.

Hijmans, R., Garcia, N., and Wieczorek, J. (2010). Gadm database of global administrative areas. *V ersion*.

Hilson, G. (2011). Artisanal mining, smallholder farming and livelihood diversification in rural sub-saharan africa: An introduction. *Journal of international development*, 23(8):1031–1041.

Humphreys, M. (2001). *Malaria: poverty, race, and public health in the United States*. JHU Press.

International Labor Organization (2015). *World report on child labour 2015: paving the way to decent work for young people*. International Labour Organization.

Kiszewski, A., Mellinger, A., Spielman, A., Malaney, P., Sachs, S. E., and Sachs, J. (2004). A global index representing the stability of malaria transmission. *The American journal of tropical medicine and hygiene*, 70(5):486–498.

Klein Goldewijk, K., Beusen, A., Van Drecht, G., and De Vos, M. (2011). The hyde 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. *Global Ecology and Biogeography*, 20(1):73–86.

Lysenko, A. and Semashko, I. (1968). Geography of malaria. a medico-geographic profile of an ancient disease. *Itogi Nauki: Medicinskaja Geografija*, pages 25–146.

Majori, G. (2012). Short history of malaria and its eradication in italy with short notes on the fight against the infection in the mediterranean basin. *Mediterranean journal of hematology and infectious diseases*, 4(1).

McCarthy, D., Wolf, H., and Wu, Y. (2000). The growth costs of malaria. Technical report, National Bureau of Economic Research.

Medlock, J. M. and Leach, S. A. (2015). Effect of climate change on vector-borne disease risk in the uk. *The Lancet Infectious Diseases*, 15(6):721–730.

Michalopoulos, S. and Papaioannou, E. (2013). Pre-colonial ethnic institutions and contemporary african development. *Econometrica*, 81(1):113–152.

Millimet, D. L. (2011). The elephant in the corner: a cautionary tale about measurement error in treatment effects models. In *Missing Data Methods: Cross-Sectional Methods and Applications*, pages 1–39. Emerald Group Publishing Limited.

Newberry. Atlas of historical county borders. [illegible]. Accessed: 2018-09-30.

Nkulu, C. B. L., Casas, L., Haufroid, V., De Putter, T., Saenen, N. D., Kayembe-Kitenge, T., Obadia, P. M., Mukoma, D. K. W., Ilunga, J.-M. L., Nawrot, T. S., et al. (2018). Sustainability of artisanal mining of cobalt in dr congo. *Nature sustainability*, 1(9):495.

Nunn, N. and Puga, D. (2012). Ruggedness: The blessing of bad geography in africa. *Review of Economics and Statistics*, 94(1):20–36.

Nunn, N. and Qian, N. (2011). The potato's contribution to population and urbanization: evidence from a historical experiment. *The Quarterly Journal of Economics*, 126(2):593–650.

Olivetti, E. A., Ceder, G., Gaustad, G. G., and Fu, X. (2017). Lithium-ion battery supply chain considerations: analysis of potential bottlenecks in critical metals. *Joule*, 1(2):229–243.

Percoco, M. (2013). The fight against disease: Malaria and economic development in italian regions. *Economic Geography*, 89(2):105–125.

Picard, J. and Mills, A. (1992). The effect of malaria on work time: analysis of data from two nepali districts. *The Journal of tropical medicine and hygiene*, 95(6):382–389.

Ramankutty, N., Foley, J. A., Norman, J., and McSweeney, K. (2002). The global distribution of cultivable lands: current patterns and sensitivity to possible climate change. *Global Ecology and biogeography*, 11(5):377–392.

Ross, R. (1897). On some peculiar pigmented cells found in two mosquitos fed on malarial blood. *British medical journal*, 2(1929):1786.

Ruggles, S., Genadek, K., Goeken, R., Grover, J., and Sobek, M. (2015). Integrated public use microdata series: Version 6.0 [dataset]. minneapolis: University of minnesota.

Sachs, J. and Malaney, P. (2002). The economic and social burden of malaria. *Nature*, 415(6872):680.

Sachs, J. D. and Warner, A. M. (1995). Natural resource abundance and economic growth. Technical report, National Bureau of Economic Research.

Saul, A. (2002). Malaria: Poverty, race, and public health in the united states.

SP Global Market Intelligence (2018). SP Global Market Intelligence Annual Report. [illegible] 2018/2018-[illegible] [Online; accessed 19-July-2018].

Spencer, E. (2016). Cobalt supply and demand—a global perspective. In *Proceedings from The Cobalt Conference, Seoul, Korea*.

Subramanian, M. A. and Sala-i Martin, X. (2003). *Addressing the natural resource curse: An illustration from Nigeria*. Number 3-139. International Monetary Fund.

Sviatschi, M. M. (2019). Making a narco: Childhood exposure to illegal labor markets and criminal life paths.

Tatem, A. J., Gething, P. W., Smith, D. L., and Hay, S. I. (2013). Urbanization and the global malaria recession. *Malaria journal*, 12(1):133.

Thomas, D., Beegle, K., Frankenberg, E., Sikoki, B., Strauss, J., and Teruel, G. (2004). Education in a crisis. *Journal of Development economics*, 74(1):53–85.

Tsui, K. K. (2010). More oil, less democracy: Evidence from worldwide crude oil discoveries. *The Economic Journal*, 121(551):89–115.

Turcheniuk, K., Bondarev, D., Singhal, V., and Yushin, G. (2018). Ten years left to redesign lithium-ion batteries.

US Geological Survey (2019). Mineral commodity summaries 2019. Technical report, United States Geological Survey.

Weil, D. N. (2014). The impact of malaria on african development over the longue durée. *Africa's Development in Historical Perspective*, pages 89–130.

Williams, R. C. (1952). The united states public health service, 1798–1950. *AJN The American Journal of Nursing*, 52(10):1279.

Wouterse, F. and Badiane, O. The role of health, experience, and educational attainment in agricultural production: Evidence from smallholders in burkina faso. *Agricultural Economics*, 0(0).

Xia, F. and Deininger, K. (2019). Spillover effects of tobacco farms on the labor supply, education, and health of children: Evidence from malawi. *American Journal of Agricultural Economics*.

4

General Conclusions

The first chapter of this dissertation has focused on the local consequences of the electrification boom, with the advent of modern lithium-ion electric batteries on child labor and consequently on fertility decisions in the Democratic Republic of Congo.

This study provides evidence that the exposure to cobalt mining activities during childhood, in the DRC subtracts children from school, thus leading to lower education attainment later in life. As a result of the increasing use of child labor due to cobalt mining, I show that the cost-opportunity of having an additional child for families living in cobalt-mining areas decreases, hence pushing fertility rates upwards.

The second chapter of this dissertation has investigated the effects of the successful eradication of malaria on historical agricultural productivity growth of US counties.

This study shows that progresses in the medical field that made possible the eradication of malaria along with other vector borne diseases, caused an increase in the agricultural productivity of areas of the US with weather conditions highly suitable for the prevalence of malaria compared to those areas in which climatic conditions (i.e. temperature, humidity, precipitations etc.) were not particularly favorable for the reproduction of mosquitoes larvae. Moreover, results suggest that the positive effects of the eradication of malaria on agricultural productivity of US counties with high stability of malaria are not due to the mere land conversion process (from wetland to arable land) which took place in early 1900s, leading to the conclusion that the principal mechanism through which malaria eradication led to increased farm productivity has to do with better health conditions of farmers which in turn led to higher labor productivity. Agricultural

The third chapter of this dissertation has focused on the interactions between technological progress and poor health conditions.

We compared population and urbanization levels of areas which were equally suitable for the cultivation of potato but different in terms of malaria ecology index. This study shows that the diffusion of potato to the Old World has positively impacted population and urbanization of grids which were more suitable for growing the staple crop. Nevertheless, results show that as the endemicity of malaria increases the impact of potato suitability decreases until it becomes null, suggesting that the diffusion of malaria fully offset the benefits of cultivating potato.

A

Appendix to: The Dark Side of Batteries

A.0.1 Additional Results and Statistics

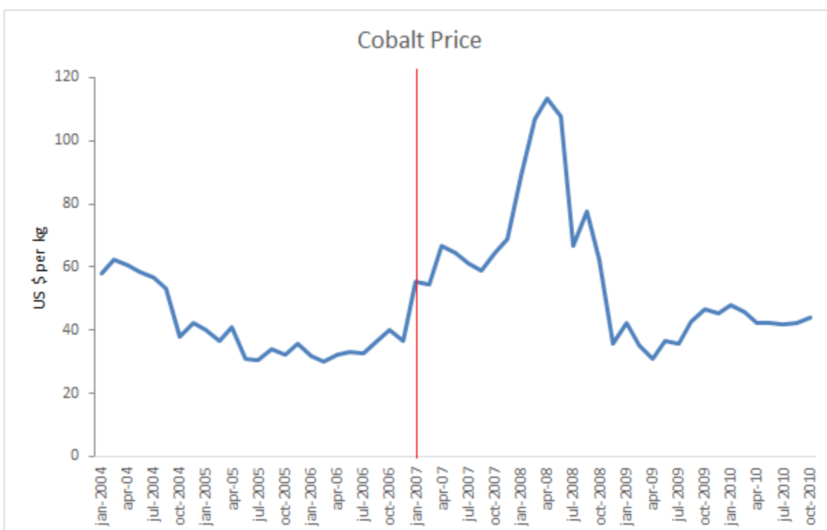


Figure A.1: Cobalt Price Trends. US dollars per kilogram.

Notes: The data in this figure is retrieved from **?** and **?**.

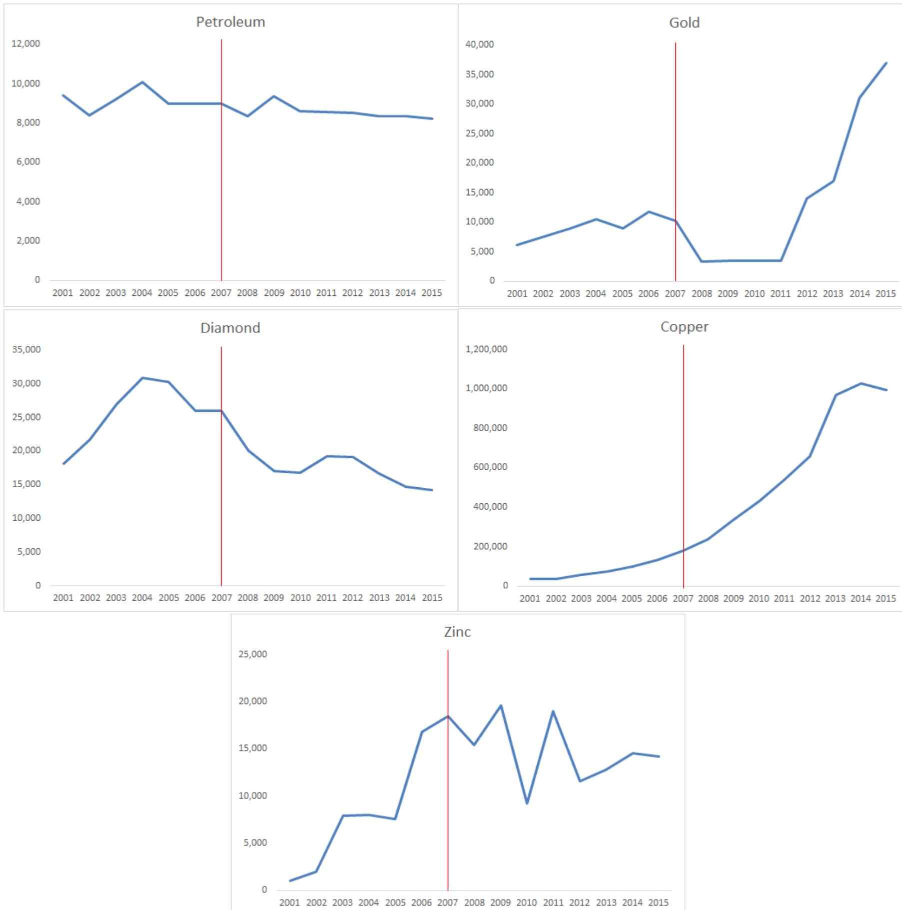


Figure A.2: Total Production of Major Minerals in DRC. Metric Tons

Notes: The data in this figure is retrieved from US Geological Survey (2019).

A.0.2 Cohort Analysis in a Spatial Lag Model

Here I consider a further cohort-specific relationship between pre-cobalt boom and children education attainment in a spatial lag model. This further specification test allows for a better understanding of both temporal and geographic distribution of the effects on education attainment. Differently from the spatial lag model presented in equation 1.4 in Section 1.4, this time each individual is recorded to a slightly different distance bin: 0–10 kilometres, 10–30 kilometres, 30-50 kilometers and 50-70 kilometers and compared with the reference category 70–100 kilometres away. I consider 20 kilometer distance bins to allow for a greater number of individuals to compare across birth years. The specification controls for the same fixed effects, trends and individual level controls as the baseline specification. The results from this alternative model can be seen in Figure A.8. The regression

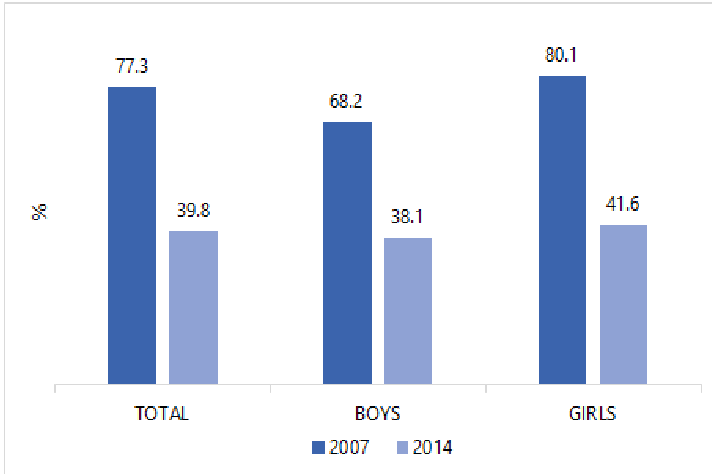


Figure A.3: Proportion of Children between 6 and 14, who reported working in DRC in the 2007 and 2014 DHS waves, by gender.

Notes: The data in this figure is retrieved from Demographic and Health Surveys (2014).

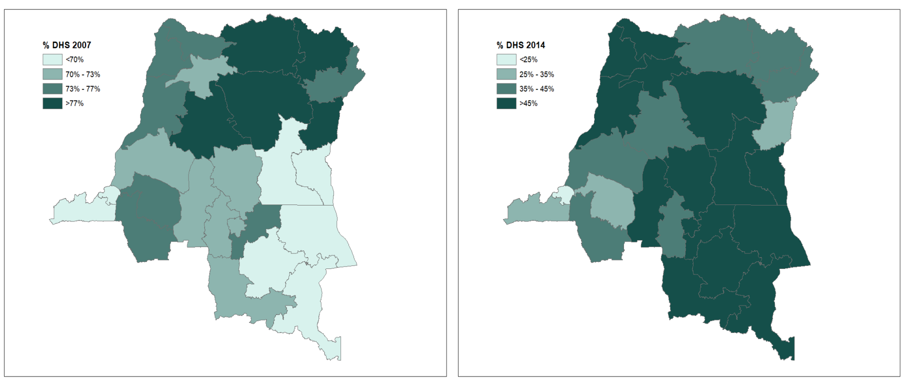


Figure A.4: Proportion of Children between 6 and 14, who reported working in DRC in the 2007 and 2014 DHS waves, per province

Notes: The data in this figure is retrieved from Demographic and Health Surveys (2014).

is specified as follows:

$$\text{Outcome}_{i,c,d,k} = \alpha + \sum_{b} \sum_{k} \beta_{b,k} \times (\text{Cobalt Mine})_c + \mathbf{X}_i' + \sigma_{1,d} + \sigma_{2,dtrend} + \varepsilon_{\,i,c,d,k} \qquad \text{(A.1)}$$

for $b \in \{0-10, 10-30, 30-50, 50-70\}$.

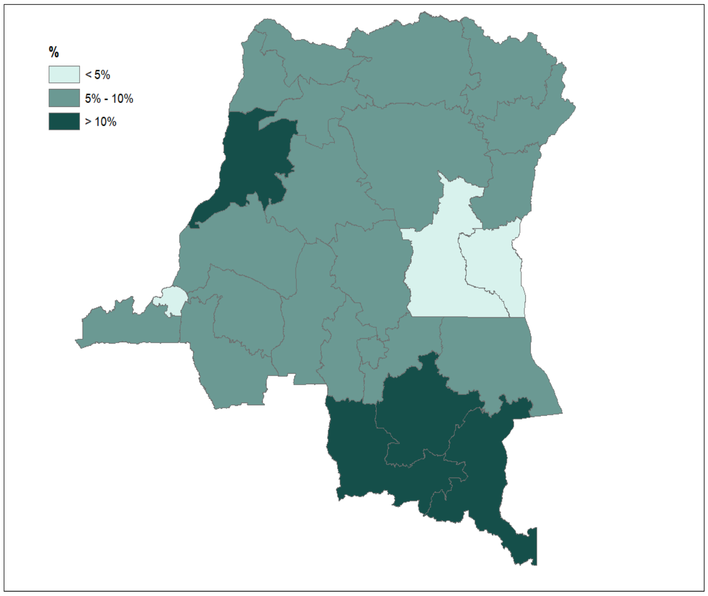


Figure A.5: Proportion of Children between 6 and 14, who are reported being engaged in the worst forms of child labor, per province in DRC in 2011.

Notes: The data in this figure is retrieved from **?**.

A.0.3 Short-Term Effects of Cobalt Mining Boom: Children (6-14 y.o.)

Table A.1 reports summary statistics of relevant variables for all children aged between 6 and 14 in the 2007 DHS survey wave and in the more recent 2014 wave of surveys. This is for all villages within 10 kilometres from the closest cobalt mine deposit and between 10 and 100 kilometers. From Table A.1 we observe that the current year of education by children in the DRC is around three.

Here, I examine the results of the baseline equation 1.1 in which the dependent variable is now the current year of education for all children aged between 6 and 14. The beta coefficient resulting from this specification compares current year of schooling of children who currently live in cobalt mining areas with those children who live in non mining areas, before and after the cobalt boom.

Column (1) of Table 14 presents results from a specification with no individual controls, sub-regional district time trends, or fixed effects. In this first specification, I find a significant effect of cobalt mining on current schooling , with a coefficient of -0.333 (and a standard error of 0.449). Columns (2)-(3) add other controls and fixed effects sequentially. The addition of individual-specific controls such as gender, type of residence, an indicator variable showing if the mother and the father are respectively alive, along with a dummy variable indicating if the children has ever migrated, in column (2) have significant effects on the estimated coefficient, which is now -0.628 (and

Table A.1: Extensive Descriptive Statistics - Children between 6 and 14

	DHS 2007 < 10 km Mean	DHS 2007 > 10 km Mean	DHS 2007 any mine Mean	DHS 2014 < 10 km Mean	DHS 2014 > 10 km Mean	DHS 2014 any mine Mean
Education						
Completed Year of Education	2.93	2.22	2.28	2.98	2.98	3.01
Wealth						
Wealth Index	4.75	4.24	3.97	4.56	4.28	4.13
Controls						
Age	10.43	10.08	10.42	10.20	10.38	10.46
Urban	1.00	1.26	1.28	1.14	1.26	1.25
Female	1.54	1.47	1.46	1.46	1.50	1.50
Ever Migrated	1.00	1.00	0.99	0.99	0.99	1.00
Mother Alive	0.98	0.96	0.96	0.96	0.95	0.96
Father Alive	0.94	0.95	0.90	0.98	0.89	0.92
Cog.Dev.						
Speech not normal				0.15	0.03	0.04
Retard				0.19	0.02	0.01
Work						
Heavyloads				0.24	0.17	0.28
Exposed to dust/fumes				0.21	0.07	0.20
Observations	122	278	401	206	612	751

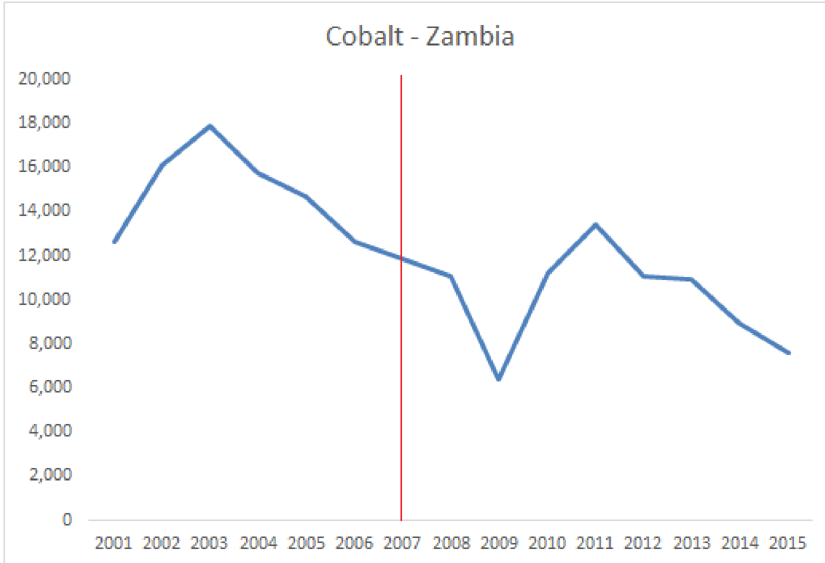


Figure A.6: Total Cobalt Production from Mining in Zambia. Metric Tons

Notes: The data in this figure is retrieved from US Geological Survey (2019).

a standard error of 0.314). Finally, specification in column (3) includes a set of fixed effects, i.e. survey year and district by time fixed effect. This specification further controls for the year of birth of each individual surveyed. Results, confirm what shown in columns (1)-(2), with a coefficient of -0.565 (and a standard error of 0.198). Taken together, these results imply that the cobalt mining boom during school-going years leads to an average of 0.5 fewer years of current schooling years in areas surrounding a cobalt mine deposit. This results on the short-term effects of cobalt mines confirm what shown in the long term analysis: children actively work in cobalt mines as soon as they turn 6.

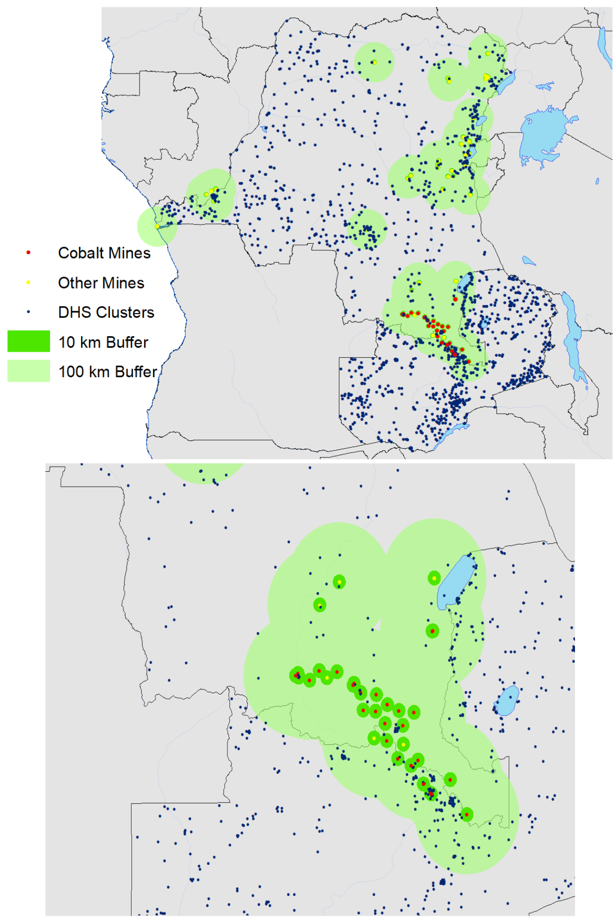


Figure A.7: Location of Cobalt Mines in DRC and Zambia

Notes: The data in this figure is retrieved from US Geological Survey (2019).

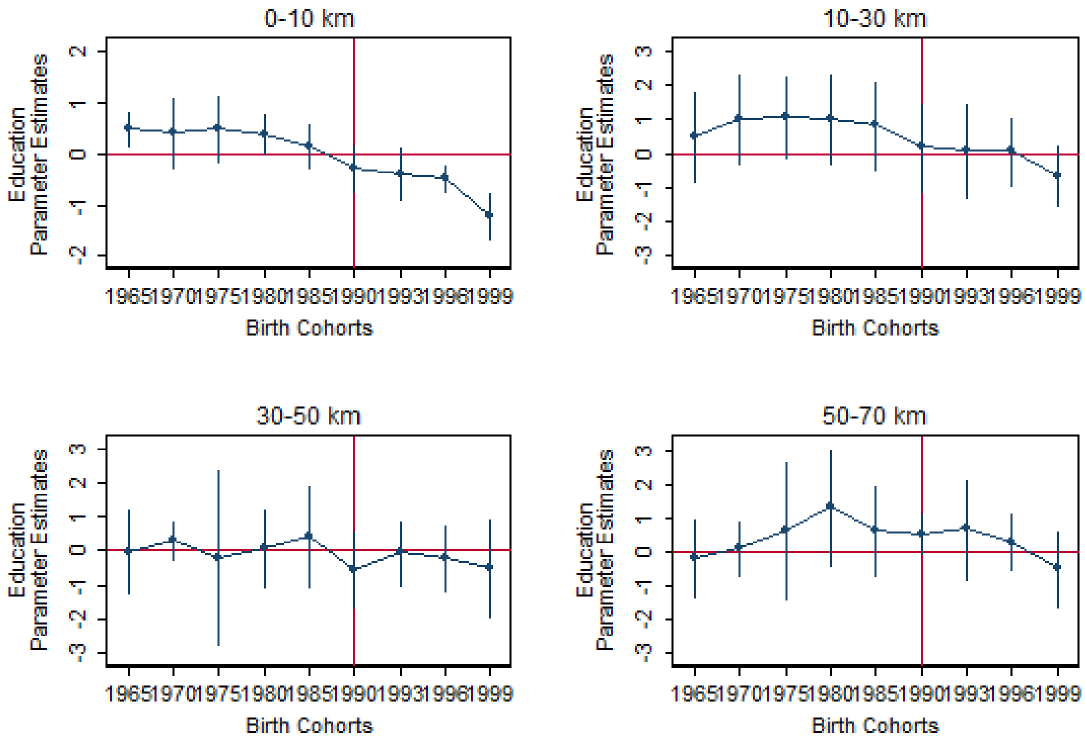


Figure A.8: Birth Cohorts in a Spatial Lag Model: Education Attainment. Five-Year Cohort-Specific relationships for all Individuals Born between 1960 and 1999

Notes: This figure reports estimated birth year (birth cohort) fixed effects in completed years of education for all individuals born between 1960 and 1999 using a spatial lag model with 20 kilometres distance bins using the baseline set of control variables and 95% confidence intervals. The sample is all individual born between 1960 and 1998 pooling DHS datasets of 2007 and 2014. Regressions also included all individual specific controls, survey year and sub-regional district by time fixed effects. To focus on individuals of relevant ages (< 14 y.o. at the time of the cobalt boom), the last four birth cohorts are three years long.

Columns (4-6) show the effects of cobalt mining exposure on current years of education of children. Results shown in these specifications strengthen the conclusions achieved in Hazan and Berdugo (2002) first and Doepke and Zilibotti (2005) later on the positive effect of child labor on the short term. Children who drop out from school to work, increase the income level of their families. This is found in Column (6) of Table 14. However, the greater wealth experienced in the short term fades away once they become adult.

Overall, I observe that the cobalt mining activities reduce current schooling in areas surrounding a cobalt deposit also in the short term. The estimates imply that children whose age falls within the ILO definition of child labor and live in cobalt mining areas complete on average, 0.5 years of education less than their peers who live in non-cobalt mining areas. Nevertheless, in the short run they achieve greater wealth conditions which fade away later in life.

Table A.2: Children Cobalt Mining Exposure and Current Education Attainment: Benchmark Results

	Current Edu.	Current Edu.	Current Edu.	Wealth	Wealth	Wealth
	Coef./SE	Coef./SE	Coef./SE	Coef./SE	Coef./SE	Coef./SE
Post x Cobalt Deposit	-0.333	-0.628*	-0.565***	0.525**	0.296*	0.301*
	(0.449)	(0.314)	(0.198)	(0.216)	(0.171)	(0.171)
Birth Cohort FE	No	No	Yes	No	No	Yes
Survey Year FE	Yes	Yes	Yes	Yes	Yes	Yes
District x Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Female	No	Yes	Yes	No	Yes	Yes
Urban	No	Yes	Yes	No	Yes	Yes
Mother Alive	No	Yes	Yes	No	Yes	Yes
Father Alive	No	Yes	Yes	No	Yes	Yes
Observations	1423	1418	1418	1424	1419	1419

Spatial Analysis: Cobalt Mining on Children (6-14 y.o.)

Here, I use a spatial lag model in order to capture spatial variations of the exposure to cobalt mining activities on current children in the DRC. As for the long term analysis, if no other shock other than the cobalt mining boom affected children's education, then we would expect the impact of cobalt mining to be only limited to those living within 10 kilometres from a cobalt mine deposit, while no effect for those people living beyond 10 kilometres. Hence, the following equation is estimated:

$$\text{Outcome}_{i,c,d,t} = \sum_b \beta_b (\text{Post})_t \times (\text{Cobalt Mine})_c + \sum_b \beta_b \ (\text{Cobalt Mine})_c +$$

$+ "\gamma \ \mathbf{X}'_{i,c} + \delta_k + \sigma_{1,d} + \sigma_{2,dtrend} + \varepsilon_{\ i,c,d,t}$ (A.2)
for $b \in \{0-10, 10-30, ..., 50-70\}$.

This spatial lag model allows for non-linear effects with distance from the nearest cobalt mine. Each children is recorded to a distance bin: 0–10 kilometres, 10–30 kilometres, etc. and compared with the reference category 70–100 kilometres away. The specification controls for the same fixed effects, trends and individual level controls as the baseline specification. The results from this alternative model can be seen in Figure 1.9.

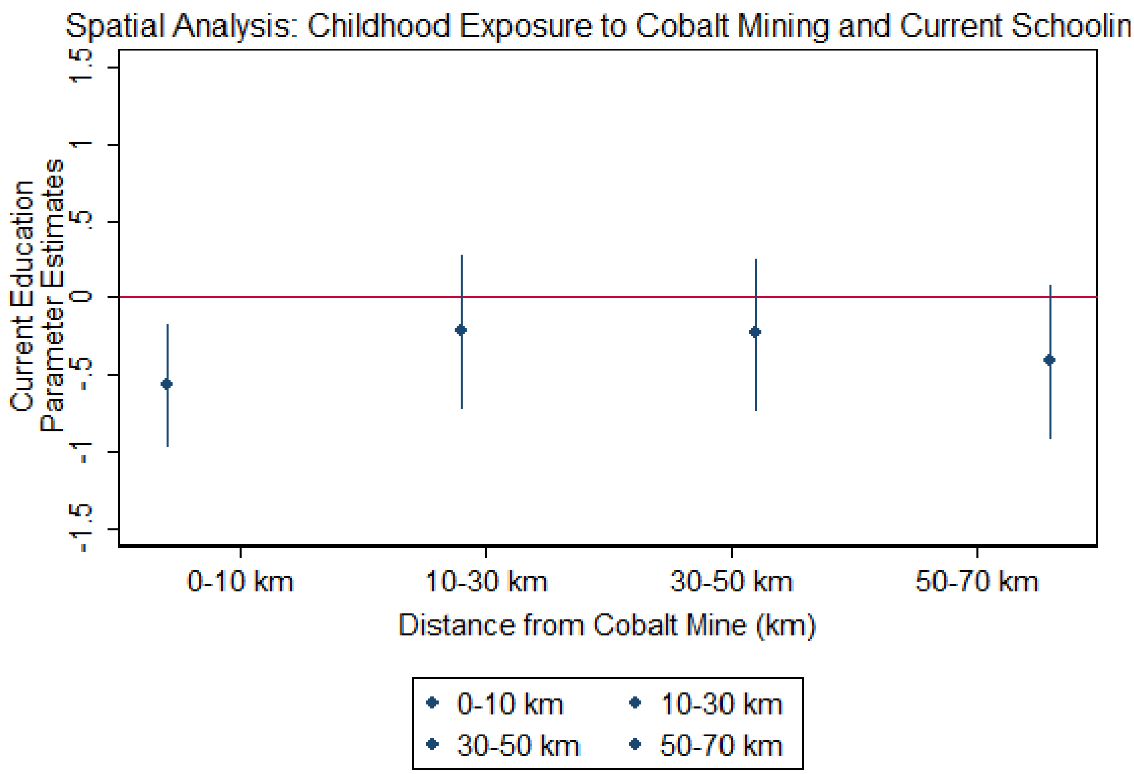


Figure 1.9: Current Year of Education for all Children between 6 and 14 years old

Notes: This Figure shows the results from a spatial lag model with 20 kilometres distance bins using the baseline set of control variables and 95% confidence intervals. The sample is all individual born between under the age of 14 , pooling DHS datasets of 2007 and 2014. Regressions also included all individual specific controls, survey year and sub-regional district by time fixed effects

2

Appendix to: Disease and Development

2.0.1 Useful Variables

Malaria Stability Index

Agricultural Suitability

2.0.2 Summary Statistics

	Pre 1991, <15 km			***Pre 1991, >15 km***			***Pre 1991, Mines***			***Post 1991, <15 km***			***Post 1991, >15 km***			***Post 1991, Mines***		
	Mean	Std.Dev.	Obs	Mean	Std.Dev.	Obs	Mean	Std.Dev.	Obs	Mean	Std.Dev.	Obs	Mean	Std.Dev.	Obs	Mean	Std.Dev.	Obs
Group 1																		
highest year of education	3.26	1.78	212	3.13	1.74	148	3.17	1.81	348	2.97	1.69	248	3.27	1.84	234	3.16	1.72	444
Group 2																		
wealth index	4.80	0.50	212	4.14	1.33	153	3.78	1.14	379	4.49	1.04	259	4.29	0.91	243	4.06	1.07	463

Extensive Summary Statistics of Relevant Variables Author's calculations based on two historical records: Historical, Demographic and Social Data:" "The United States, 1790-2002 (Haines et al., 2005) and the IPUMS dataset (Ruggles et al., 2015)

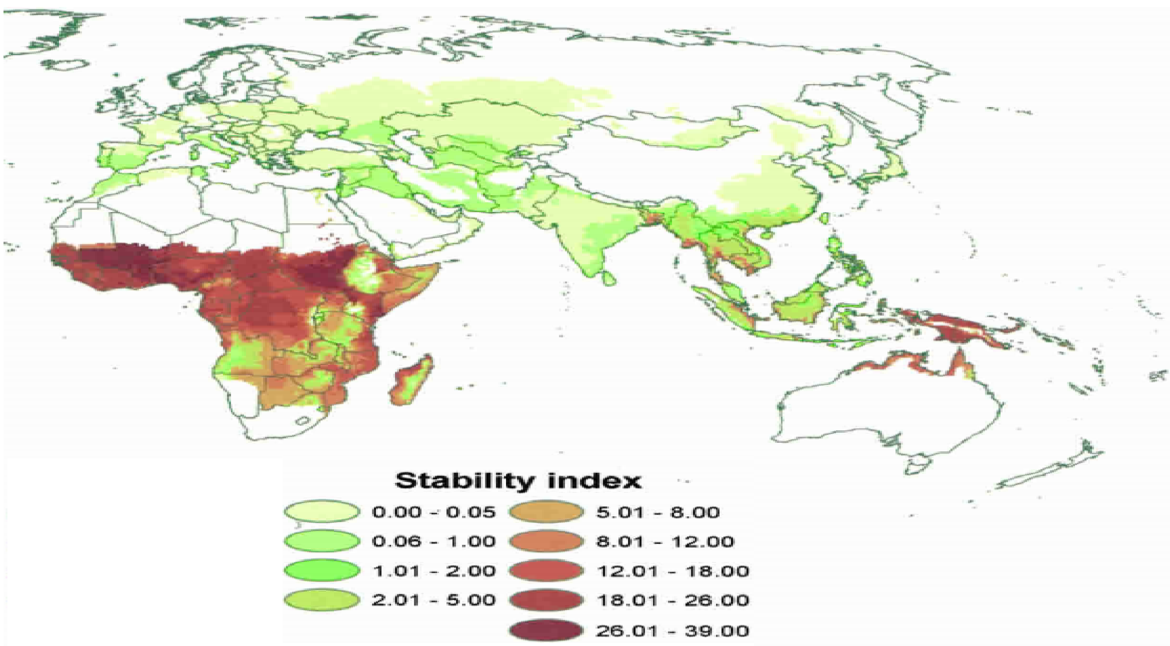


Figure 2.1: Distribution of the potential stability of malaria

Malaria transmission based on regionally dominant vector mosquitoes and a 0.5°gridded temperature and precipitation data set (Kiszewski et al., 2004)

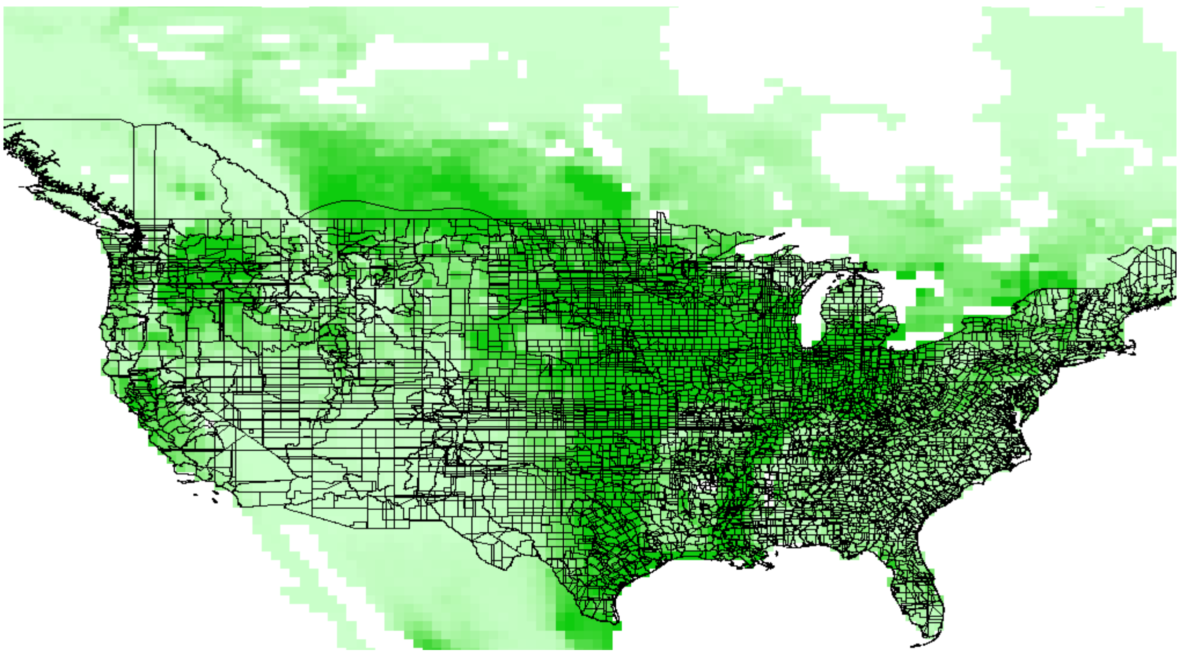


Figure 2.2: Distribution of the total crop suitability per county

Data retrieved from (Ramankutty et al., 2002)

Table 2.1: Pairwise Correlation of Useful Variables

	MSI	MEI	Ln fv/farmer	Ln fout/farmer	Pop density	Farmland	White people	Fertilizer	Drainage	Land i
MSI	1.00									
MEI	0.56	1.00								
Ln fv/farmer	-0.10	-0.14	1.00							
Ln fout/farmer	-0.03	-0.02	0.92	1.00						
Pop density	-0.15	-0.13	0.15	0.13	1.00					
Farmland	-0.11	0.09	0.22	0.23	0.49	1.00				
White people	-0.38	-0.66	0.21	0.10	0.08	-0.06	1.00			
Fertilizer	0.11	0.24	-0.09	-0.03	0.15	0.23	-0.35	1.00		
Drainage	-0.09	-0.06	0.22	0.21	0.05	0.05	0.15	-0.19	1.00	
Land improved	-0.31	-0.43	0.41	0.36	0.40	0.38	0.39	-0.17	0.24	1.
Agr. Suitability	-0.10	-0.16	0.42	0.37	0.10	0.11	0.31	-0.39	0.46	0.
Observations	23842									

Correlation Matrix of Regressors

2.0.3 Correlation between malaria endemicity/stability measures and control variables

Table 2.2: Correlation Table for each measure of malaria

	MEI	MSI
	b	b
Ln fout/farmer	0.005	0.003
MSI	0.629	
Pop density	-0.157	-0.043
Farmland	0.082	-0.013
White people	-0.249	-0.102
Fertilizer	0.068	0.048
Drainage	0.015	-0.014
Land improved	-0.160	-0.069
Agr. Suitability	0.040	0.029
Constant	0.399	0.145
Observations	18663	18663

Table 14 shows correlation between control variables and dependent with the two measures of malaria to check for the presence of multicollinearity. The primary concern is that as the degree of multicollinearity increases, the regression model estimates of the coefficients become unstable and the standard errors for the coefficients can get wildly inflated.

To check for multicollinearity I compute the variance inflation factor (VIF). As a rule of thumb, a variable whose VIF value is greater than 10 may merit further investigation. It means that the variable could be considered as a linear combination of other independent variables. The regression model predicting the natural logarithm of farm value and farm output per farmer from the MSI population density in 1870, proportion of farmland out of total county acre in 1870, the proportion of white people out of total number of farmers in 1870, the proportion of farmers reporting the use of fertilizers, have received drainage and have improved land in 1870, and agriculture suitability index is shown with the associated VIFs in Table 15. VIFs exclude any problems of multicollinearity in the model, thus the empirical model presented in equation 2.1 gives valid results about the predictor and the control variables.

2.0.4 Robustness

Robustness check 1: Effect of the eradication of malaria disease on the second farm productivity variable, i.e. farm output per county farmer, all counties:

Robustness check 2: Effect of the eradication of malaria disease on the second farm productivity variable, i.e. farm output per county farmer, neighboring counties analysis:

Table 2.3: Check for multicollinearity of predictors

	Ln fv/farmer VIF
MSI	1.25
Pop density	1.45
Farmland	1.54
White people	1.48
Fertilizer	1.31
Drainage	1.26
Land improved	1.97
Agr. Suitability	1.82
Constant	
Mean VIF	1.53

Table 2.4: Check for multicollinearity of predictors

	(1) Ln fv/farmer vif
MEI	2.571945
MSI	1.54874
Pop density	1.457751
Farmland	1.658303
White people	1.989921
Fertilizer	1.388199
Drainage	1.272055
Land improved	2.176409
Agr. Suitability	1.868958
Constant	
Observations	23769

Table 2.5: The impact of MSI on county agricultural productivity: Farm output per farmer

	(1) b/se	(2) b/se	(3) b/se	(4) b/se
MSI x Post	0.523***	0.523***	0.530***	0.221***
	(0.122)	(0.141)	(0.153)	(0.035)
Farmland	-1.802***	-2.403***	-2.398***	-1.747***
	(0.123)	(0.146)	(0.142)	(0.147)
Pop density	-4.297***	-6.885***	-6.934***	-3.248***
	(0.847)	(1.206)	(1.217)	(0.628)
White people	2.003***	1.294***	1.311***	-0.121
	(0.089)	(0.131)	(0.126)	(0.170)
Fertilizer		8.590***	8.263***	5.159***
		(1.153)	(1.406)	(1.317)
Land improved		1.600***	1.659***	0.244
		(0.195)	(0.264)	(0.251)
Drainage		0.486***	0.523***	0.628***
		(0.073)	(0.078)	(0.098)
Agriculture suitability			-0.097	0.287**
			(0.166)	(0.124)
State X Year	No	No	No	Yes
Observations	18674	18663	18663	18663
Adjusted R^2	0.507	0.523	0.523	0.775

Table 2.6: The impact of MEI on county agricultural productivity: Farm output per farmer

	(1) b/se	(2) b/se	(3) b/se	(4) b/se
MEI x Post	0.639***	0.386***	0.383***	0.144***
	(0.037)	(0.038)	(0.038)	(0.048)
White people x MSI	20.814***	5.812***	5.711***	2.311
	(1.571)	(1.448)	(1.476)	(1.656)
Pop density x MSI	-53.112**	-61.028**	-61.214**	-35.481***
	(24.804)	(26.008)	(26.134)	(12.621)
Farmland	-1.170***	-3.002***	-3.001***	-1.886***
	(0.087)	(0.133)	(0.135)	(0.148)
Fertilizer		5.828***	5.998***	4.616***
		(0.957)	(1.052)	(1.305)
Land improved		2.760***	2.725***	0.293
		(0.143)	(0.208)	(0.250)
Drainage		0.442***	0.425***	0.613***
		(0.077)	(0.079)	(0.098)
Agriculture suitability			0.047	0.325***
			(0.136)	(0.124)
State X Year	No	No	No	Yes
Observations	18674	18663	18663	18663
Adjusted R^2	0.447	0.515	0.515	0.776

Table 2.7: The impact of MSI on county agricultural productivity: Farm output per farmer. Neighboring Counties Analysis

	(1) b/se	(2) b/se	(3) b/se	(4) b/se
MSI x Post	0.132***	0.137***	0.115***	0.119***
	(0.035)	(0.033)	(0.033)	(0.030)
Farmland	-1.242***	-1.109***	-1.337***	-1.293***
	(0.187)	(0.206)	(0.214)	(0.207)
Pop density		-2.913***	-4.134***	-4.175***
		(0.785)	(1.067)	(0.937)
White people		0.094	0.004	0.162
		(0.188)	(0.241)	(0.247)
Fertilizer			2.660**	0.983
			(1.260)	(1.281)
Land improved			1.336***	1.819***
			(0.349)	(0.427)
Drainage			0.473***	0.639***
			(0.085)	(0.097)
Agriculture suitability				-0.692**
				(0.269)
Observations	5794	5794	5783	5783
Adjusted R^2	0.747	0.750	0.763	0.766

Table 2.8: The impact of MEI on county agricultural productivity: Farm output per farmer. Neighboring Counties Analysis

	(1) b/se	(2) b/se	(3) b/se	(4) b/se
MEI x Post	0.164*	0.170**	0.092	0.111
	(0.086)	(0.085)	(0.083)	(0.089)
Farmland	-1.317***	-1.226***	-1.447***	-1.409***
	(0.196)	(0.205)	(0.221)	(0.211)
White people x MSI		3.916	2.362	4.906
		(2.874)	(3.979)	(3.993)
Pop density x MSI		-25.326***	-31.708***	-30.511***
		(7.303)	(10.753)	(10.573)
Fertilizer			2.490**	0.875
			(1.199)	(1.293)
Land improved			1.263***	1.709***
			(0.354)	(0.426)
Drainage			0.449***	0.610***
			(0.087)	(0.096)
Agriculture suitability				-0.661**
				(0.266)
Observations	5794	5794	5783	5783
Adjusted R^2	0.748	0.752	0.764	0.766

Robustness check 3: Effect of the eradication of malaria disease on the second farm productivity variable, i.e. farm output per county farmer, placebo treatment periods:

Table 2.9: The impact of MSI on county farm output per farmer: Placebo Treatment Periods

	1870-1900 Post=1880, 1890, 1900 b/se	1870-1900 Post= 1890, 1900 b/se	1940-1970 Post= 1950, 1960 1970 b/se	1940-1970 Post= 1960, 1970 b/se
MSI x Post	0.043	-0.011	0.018	-0.002
	(0.097)	(0.028)	(0.030)	(0.021)
State X Year	Yes	Yes	Yes	Yes
Pfarm 1870 X Year	Yes	Yes	Yes	Yes
Popden 1870 X Year	Yes	Yes	Yes	Yes
Pwhite 1870 X Year	Yes	Yes	Yes	Yes
Fertilizer 1870 X Year	Yes	Yes	Yes	Yes
Drain 1870 X Year	Yes	Yes	Yes	Yes
Land Imp 1870 X Year	Yes	Yes	Yes	Yes
Suitability X Year	Yes	Yes	Yes	Yes
Observations	6752	6752	6808	6808
Adjusted R^2	0.808	0.803	0.804	0.806

*Notes: The periods range from 1870 to 2000, and are observed every other decade. The dependent variable is the natural log of the county average farm value per county acre of farmland. The variable of interest is the average county MSI index. The Post Indicator variable equals zero for the periods 1870–1900 and one for the periods 1910–2000. All regressions include time periods fixed effects, county fixed effects, and state by time fixed effects. The inclusion of a control variable interacted with the full set of time-period fixed effects is indicated by a yes; no indicates that the control is not included in the specification. Coefficients are reported with standard errors. In the parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.*

Table 2.10: The impact of MEI on county farm output per farmer: Placebo Treatment Periods

	1870-1900 Post=1880, 1890, 1900 b/se	1870-1900 Post= 1890, 1900 b/se	1940-1970 Post= 1950, 1960 1970 b/se	1940-1970 Post= 1960, 1970 b/se
MEI x Post	-0.030	-0.062	0.146***	0.147***
	(0.053)	(0.042)	(0.034)	(0.040)
State X Year	Yes	Yes	Yes	Yes
Pfarm 1870 X Year	Yes	Yes	Yes	Yes
MSI	Yes	Yes	Yes	Yes
Ln(pop den) X MSI	Yes	Yes	Yes	Yes
Ln(p white) X MSI	Yes	Yes	Yes	Yes
Fertilizer 1870 X Year	Yes	Yes	Yes	Yes
Drain 1870 X Year	Yes	Yes	Yes	Yes
Land Imp 1870 X Year	Yes	Yes	Yes	Yes
Suitability X Year	Yes	Yes	Yes	Yes
Observations	6752	6752	6808	6808
Adjusted R^2	0.809	0.804	0.804	0.805

*Notes: The periods range from 1870 to 2000, and are observed every other decade. The dependent variable is the natural log of the county average farm value per county acre of farmland. The variable of interest is the average county MSI index. The Post Indicator variable equals zero for the periods 1870–1900 and one for the periods 1910–2000. All regressions include time periods fixed effects, county fixed effects, and state by time fixed effects. The inclusion of a control variable interacted with the full set of time-period fixed effects is indicated by a yes; no indicates that the control is not included in the specification. Coefficients are reported with standard errors. In the parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.*

Robustness check 4: Discrete measure of the Malaria Endemicity Index. As in Kiszewski et al. (2004) the MSI is divided into 9 categories ranging from malaria free grids with a MSI between 0 and 0.05 to the highest persistence of malarious weather conditions with a MSI between 26 and 39*:

Table 2.11: The impact of malaria eradication on county agricultural productivity: Baseline estimates

	ln(county average farm value per acre)		
	b/se	b/se	b/se
MEI	0.168***	0.231***	0.200***
	(0.056)	(0.056)	(0.055)
State X Year	Yes	Yes	Yes
Pfarm 1870 X Year	Yes	Yes	Yes
Popden 1870 X Year	No	Yes	Yes
Pwhite 1870 X Year	No	Yes	Yes
Farmer 1870 X Year	No	No	Yes
Fertilizer 1870 X Year	No	No	Yes
Suitability X Year	No	No	Yes
Observations	23842	23842	23842
Adjusted R^2	0.939	0.941	0.943

Notes: The periods range from 1870 to 2000, and are observed every other decade.
The dependent variable is the natural log of the county average farm value per county acre of farmland.
The variable of interest is the average county MEI index.
The Post Indicator variable equals zero for the periods 1870 to 1900 and one for the periods 1910 to 2000.
All regressions include time periods fixed effects, county fixed effects, and state by time fixed effects.
The inclusion of a control variable interacted with the full set of time-period fixed effects is indicated by a yes.
In the parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Robustness check 5: Restriction to counties located in the south-west US.

*Since the MSI for the USA predominantly ranges from 0 to 1, the table shows the effect of eradicating malaria on the agricultural productivity of counties being in the second category compared to the counties belonging in the first category

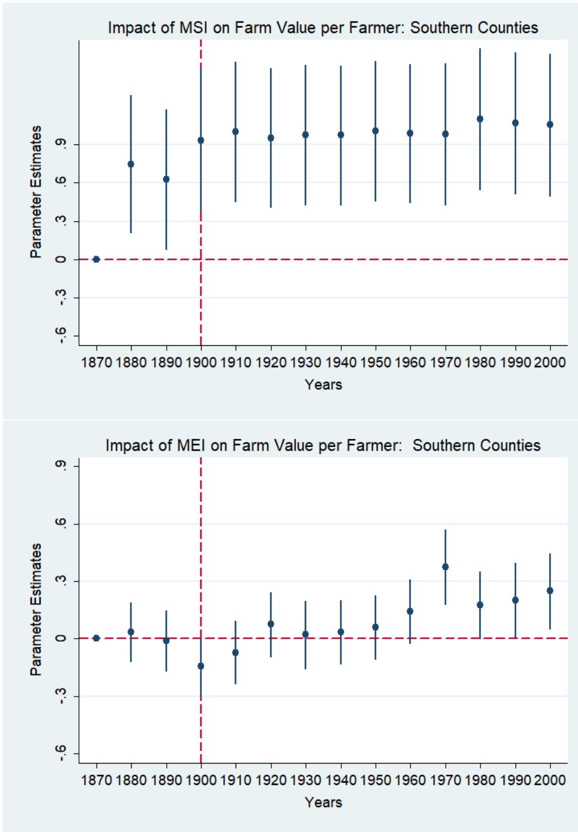


Figure 2.3: Farm value per farmer and MSI (left) and MEI (right): Southern Counties

Table 2.12: The impact of MSI on county agricultural productivity: Farm value per farmer. Southern Counties

	(1) b/se	(2) b/se	(3) b/se	(4) b/se
MSI	0.246***	0.240	0.241***	0.244***
	(0.055)	(.)	(0.047)	(0.046)
Farmland	-1.769***	-1.717	-1.547***	-1.545***
	(0.096)	(.)	(0.105)	(0.105)
Pop density		-2.748	-2.640***	-2.641***
		(.)	(0.709)	(0.717)
White people		-0.245	-0.139	-0.139
		(.)	(0.133)	(0.133)
Fertilizer			2.108**	2.154**
			(0.953)	(0.961)
Land improved			-0.182	-0.220
			(0.157)	(0.175)
Drainage			0.730***	0.710***
			(0.097)	(0.102)
Agriculture suitability				0.067
				(0.091)
State X Year	Yes	Yes	Yes	Yes
Observations	15162	15162	15148	15148
Adjusted R^2	0.835	0.835	0.837	0.837

Notes: The periods range from 1870 to 2000, and are observed every other decade.
The dependent variable is the natural log of the county average farm value per farmer.
The variable of interest is the average county MEI index.
The Post Indicator variable equals zero for the periods 1870 to 1900 and one for the periods 1910 to 2000.
All regressions include time periods fixed effects, county fixed effects, and state by time fixed effects.
The inclusion of a control variable interacted with the full set of time-period fixed effects is indicated by a yes.
In the parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

Table 2.13: The impact of MEI on county agricultural productivity: Farm value per farmer. Southern Counties

	(1) b/se	(2) b/se	(3) b/se	(4) b/se
MEI	0.121***	0.133	0.103***	0.106***
	(0.032)	(.)	(0.036)	(0.036)
Farmland	-1.847***	-1.797	-1.638***	-1.634***
	(0.096)	(.)	(0.101)	(0.101)
White people x MSI		-0.501	0.094	0.209
		(.)	(1.232)	(1.231)
Pop density x MSI		-24.142	-25.596***	-25.928***
		(.)	(7.712)	(7.799)
Fertilizer			1.861*	1.927**
			(0.960)	(0.969)
Land improved			-0.101	-0.164
			(0.157)	(0.173)
Drainage			0.727***	0.693***
			(0.095)	(0.099)
Agriculture suitability				0.120
				(0.089)
State X Year	Yes	Yes	Yes	Yes
Observations	15162	15162	15148	15148
Adjusted R^2	0.835	0.836	0.838	0.838

Notes: The periods range from 1870 to 2000, and are observed every other decade.
The dependent variable is the natural log of the county average farm value per farmer.
The variable of interest is the average county MEI index.
The Post Indicator variable equals zero for the periods 1870 to 1900 and one for the periods 1910 to 2000.
All regressions include time periods fixed effects, county fixed effects, and state by time fixed effects.
The inclusion of a control variable interacted with the full set of time-period fixed effects is indicated by a yes.
In the parentheses. ***, **, and * indicate significance at the 1, 5, and 10 percent levels, respectively.

3

Appendix to: Revisiting the Contribution of Potato

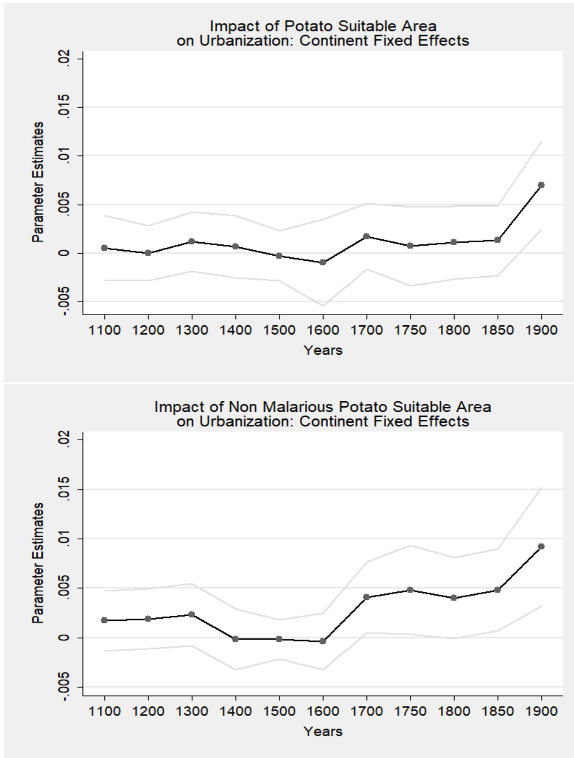


Figure 3.1: Potato Suitable Areas and Urbanization: Continent Fixed Effects, Flexible Estimates.

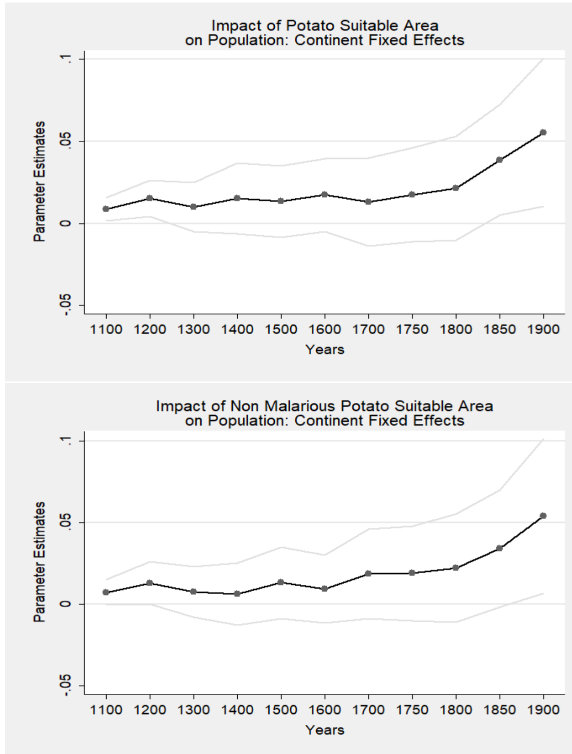


Figure 3.2: Potato Suitable Areas and Population: Continent Fixed Effects, Flexible Estimates.