Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN

ECONOMIA

Ciclo XXV

Settore Concorsuale di afferenza: 13/A1

Settore Scientifico disciplinare: SECS-P/01

---

# Essays on Geography and Diseases

---

Presentata da: Elena Esposito

*Coordinatore Dottorato:*
Prof. Matteo Cervellati

*Relatore:*
Prof. Matteo Cervellati

Esame finale anno: 2014

ALMA MATER STUDIORUM - UNIVERSITA' DI BOLOGNA

# *Abstract*

## Essays on Geography and Diseases

by Elena Esposito

This dissertation explores how diseases contributed to shape historical institutions and how health and diseases are still affecting modern comparative development. The overarching goal of this investigation is to identify the channels linking geographic suitability to diseases and the emergence of historical and modern institutions, while tackling the endogenenity problems that traditionally undermine this literature. I attempt to do so by taking advantage of the vast amount of newly available historical data and of the richness of data accessible through the geographic information system (GIS). The first chapter of my thesis, 'Side-Effects of Immunities: The African Slave Trade', proposes and test a novel explanation for the origins of slavery in the tropical regions of the Americas. I argue that Africans were especially attractive for employment in tropical areas because they were immune to many of the diseases that were ravaging those regions. In particular, Africans' resistance to malaria increased the profitability of slaves coming from the most malarial parts of Africa. In the second chapter, 'Caste Systems and Technology in Pre-Modern Societies', I advance and test the hypothesis that caste systems, generally viewed as a hindrance to social mobility and development, had been comparatively advantageous at an early stage of economic development. In the third chapter of my thesis, 'Malaria as Determinant of Modern Ethno-linguistic Diversity', I conjecture that in highly malarial areas the necessity to adapt and develop immunities specific to the local disease environment historically reduced mobility and increased isolation, thus leading to the formation of a higher number of different ethno-linguistic groups. In the final chapter, 'Malaria Risk and Civil Violence: A Disaggregated Analysis for Africa', I explore the relationship between malaria and violent conflicts. Using geo-referenced data for Africa, the article shows that violent events are more frequent in areas where malaria risk is higher.

# *Acknowledgements*

This work would not have been possible without the guidence of my supervisor, Matteo Cervellati.

# Contents

# List of Figures

# List of Tables

*Dedicated to Enrico, and to my family*

# Introduction

I am an **empirical economist** and all my current research investigates how diseases contributed to shape historical institutions and how health and diseases are still affecting modern comparative development.

The overarching goal of my **Ph.D. Dissertation** is to identify the channels linking geographic suitability to diseases and the emergence of several historical insitutions, while tackling the endogenenity problems that traditionally undermine this literature. I attempt to do so by taking advantage of the vast amount of newly available historical data and of the richness of data accessible through the geographic information system (GIS). In recent years, macroeconomists have become increasingly aware of the importance of history and historical events to explain a country's economic success today. In particular, economists started to explore the interplay between large historical processes, such as colonialism, and country-specific characteristics. The conceptual framework underlying my work can be summarized as follow: i) there seems to be a long shadow of history over modern growth, as there is evidence that historical events still explain a sizeable share of cross-country income differentials (Guiso, Sapienza, and Zingales [2008], Nunn [2008], Ashraf and Galor [2011b]); ii) history can matter today because it shaped deep, slow-moving, persistent 'institutions', defined in a broad sense (Guiso, Sapienza, and Zingales [2006], Nunn and Wantchekon [2009]); and, a less explored point, iii) throughout history, geographic physical characteristics permanently molded the social structure of societies, with consequences still visible today (Michalopoulos [2012]). My **Ph.D. Thesis** builds on these premises and contributes to this debate by exploring the role that health and diseases, notably malaria and geographic suitability to malaria, played in shaping relevant historical institutions.

The first chapter of my thesis, 'Side-Effects of Immunities: The African Slave Trade', advances a novel hypothesis about the **origins of slavery** in the southern part of the US. Why slavery was practiced in certain American regions and not in others, and why Africans, and certain Africans in particular, were transported to the New World so numerously and enslaved is an economically relevant question. In fact, there is ample evidence that the American regions that relied on slavery in colonial times, as well as the African regions from which the slaves came, are poorer today (Engerman and Sokoloff [1997], Nunn [2008], Dell [2010]). Borrowing insights from historians, such as Curtin [1968] and Wood [1974], I argue that the immunities of Africans from tropical regions to the disease environment that was prevalent in the American South made them especially attractive for employment in these regions. In other terms, Africans' resistances to malaria and other tropical diseases increased the profitability of African slave labor, and of African slaves from more malarious countries in particular. To verify the hypothesis, I exploit the time variation arising from the introduction of malaria falciparum into the US colonies, and the additional fact that malaria could thrive only in a subset of American colonies because of climatic characteristics that malaria requires. Thus, I compare the percentage of slaves in the US colonial states that were more suitable to malaria with states that were less suitable before and after the introduction of malaria falciparum. Finally, to show that resistances to malaria increased the profitability of African labor, I document the existence of a **malaria premium** using historical slave prices. Namely, I show that, among Africans, slaves born in African countries with more malaria commanded higher prices.

In the second chapter of my thesis, 'Caste Systems and Technology in Pre-Modern Societies' (which is the first article that I wrote for the phd thesis), I advance the hypothesis that **caste systems**, generally viewed as a hindrance to social mobility and development, had been comparatively advantageous at an early stage of economic development. Using data from the Murdock Ethnographic Atlas, I find robust evidence in support for the hypothesis. In order to rule out reverse causation, based on anthropological theories suggesting that caste systems represented a social device aiming at reducing disease contagion, I instrument castes with an index of disease exposure. Also because of data limitation, I am not able to fully exclude that additional channels linking the disease environment and technology violate the exclusion restriction I rely on for identification.

In the third chapter of my thesis, '**Malaria** as Determinant of Modern **Ethnolinguistic Diversity**', a joint work with Matteo Cervellati and Giorgio Chiovelli, I am better able to overcome problems of data limitation as the setting allows me to rely entirely on geolocalized disaggregated data. In this paper, we conjecture that in highly malarial areas the necessity to adapt and develop immunities specific to the local disease environment historically reduced mobility and increased isolation, thus leading to the formation of a higher number of different ethno-linguistic groups. First, we conduct the analysis at a disaggregated level by creating a grid of artificial countries of 1 x 1 degree of size (around 110 square km at the equator) to employ as units of observation. We use a rich set of new data on ethno-linguistic diversity constructed from geolocalized maps of ethno-linguistic groups around the world at several points in history, as well as new data on historical malaria endemicity. Results point to a strong positive correlation between historical malaria endemicity and the number of ethno-linguistic groups at all levels of spatial disaggregation. In the second part of the exercise, we hypothesize that the need to exploit and preserve location-specific immunities increased endogamous marriages in areas with higher geographic suitability to malaria. In order to test the existence of this channel, we exploit current data on marriage patterns in 22 African countries, retrieved from the Demographic and Health Survey. Regressions' results show a positive correlation between geographic suitability to malaria and the probability that an individual marries somebody from her same ethnic family.

In the final chapter of my dissertation, 'Malaria Risk and Civil Violence: A Disaggregated Analysis for Africa', a joint work with Matteo Cervellati, Simona Valmori (Università di Bologna) and Uwe Sunde (University of Munich), we explore the relationship between **malaria** and **violent conflicts**. Using geo-referenced data for Africa, we test of the hypothesis that a greater exposure to the health-threatening risks of malaria epidemics increases the likelihood of civil violence (by reducing its opportunity cost). Following the epidemiological literature we identify peaks in malaria risk (and not malaria exposure) to account for the role of acquired and genetic immunities. In a dynamic (panel) perspective the risk of malaria is highest in periods that are unusually suitable for malaria out-brakes and in areas with low to intermediate malaria ecology (exposure). We investigate the link between malaria risk and civil violence in econometric specifications that incorporate these insights by allowing for non-linear and heterogeneous effects and by exploiting variation across grid cells (within country) and within cells variation in malaria risk measures using variation in malaria suitable months. For their nature malaria threats are limited in time and can be predicted using monthly panel data which also allows to control for all relevant cell specific determinants of civil conflicts that change from a year to the other (like e.g. cell specific income production in a given year). The findings suggest that malaria risk increases, in particular, the likelihood of unorganized (non battle related) civil violence. The results are robust to extensive checks including the use of alternative grid sizes and spatial econometrics techniques.

# Chapter 1

# Side-Effects of Immunities: the African Slave Trade

## 1.1 Introduction

The practice of slavery during colonialism compromised the long-term economic prosperity of the affected areas. It prejudiced the growth prospects of enslaved populations, as it was the case for Africans [Nunn, 2007], by engendering social distrust [Nunn and Wantchekon, 2009], population decline [Manning, 1990, McEvedy, Jones, et al., 1978], by fostering predatory political attitudes, ethnic stratification [Barry, 1998, Bates, 2008, Daaku, 1970, Inikori, 2003] and by discouraging state-building [Rodney, 1972, Whatley and Gillezeau, 2011][1]. It was argued also that slavery had been a burden for societies historically relying on it as a source of labor [Engerman and Sokoloff, 1997, Genovese, 1989, Ransom and Sutch, 2001, Wright, 1978], as new empirical evidence documents that a history of slavery fostered long-term poverty, political and economic inequality [Acemoglu, García-Jimeno, and Robinson, 2012, Dell, 2010, Nunn, 2008].

This paper aims to document the role played by diseases in determining why slavery was practiced in certain American regions and not in others, and why Africans, and certain Africans in particular, were transported and enslaved to the New World so numerously. Borrowing insights from historians, such as Curtin [1968], Wood [1974], Coelho and McGuire [1997], Kiple and King [2003] and McNeill [2010], I argue that the immunities and resistance of Africans from tropical regions to the disease environment that was

---

[1] Supporting a different view, Eltis and Jennings [1988] argued that the long-term economic impact of the slave trade was minor, while Fage [1969] sustained that the slave trade actually encouraged the consolidation of political states in Africa

prevalent in the American tropical and semi-tropical areas made them especially attractive for employment in these regions. In other words, Africans' resistances to malaria and to other tropical diseases increased the profitability of African slave labor, and of African slaves from more malarial countries in particular. The differential endowments of disease immunities and resistance of Africans, Europeans and Native Americans was a consequence of the geography of diseases prevailing at the eve of America's discovery. Until the sixteenth century geographical isolation protected the Americas from the burden of both human-contact civilization diseases (measles, smallpox...) and tropical diseases (such as malaria and yellow fever). Soon after discovery all these diseases started to be introduced into the New World, infecting a completely nonimmune population which thus began to suffer from ravaging epidemics. Human-contact civilization diseases could spread all over the Americas virtually relentlessly, while tropical diseases thrived in the warm and humid weather of the American tropics[2]. Among old world populations, while both Africans and Europeans had been exposed - even if with different frequencies - to the major civilization diseases, it was mainly Africans from tropical regions that had to face the most burdensome strains of tropical diseases since the origin of humankind. In particular, malaria represented such an evolutionary threat that, over the centuries, innate resistances (such as the sickle cell disease) were selected for in order to top up the protection granted by acquired immunities. Thus, the balance of disease immunities and resistances of African slaves, Native Americans slaves and Europeans servants increased comparative labor productivity of African slaves in tropical regions and contributed in raising colonizers' preference for African slave labor, fostering the demand (and the supply) of it.

There are several reasons why it is difficult to provide an empirical foundation to the African Slave Trade "disease hypothesis". In Table A.1, I show that a positive correlation exists between malaria prevalence and the late nineteenth century fraction of Africans - and of slaves - on population across New World regions. These correlations might be spurious for many reasons. First of all, the intensity of tropical disease incidence, malaria included, is a consequence of both geo-climatic exogenous characteristics and of living standards and technological progress of a region[3]. In order to obtain a purely exogenous measure of malaria incidence, researchers have been increasingly relying on indexes of geographical suitability to malaria, which predict the likelihood that a disease have to spread based on climatic and geographical factors. However, since malaria is likely to thrive in the same areas suitable for crops traditionally raised in plantations, in a cross-sectional setting it is virtually impossible to disentangle the role

---

[2] For insightful reviews see, among others, Crosby [2003] and Mann [2011]

[3] All vector diseases, such as malaria and yellow fever, require pools of mosquitoes for the pathogens to be able to spread and surive. The chances to eradicate a disease rest greatly on the feasibility of a reduction in the pool of available vectors (by the draining of marshes, the clearing of swamps) and on the possibility to protect man from mosquitoes (through bed nets, screened houses)

of soil suitabilities from the effect played by geographic suitability to malaria. As several authors proposed, the price differentials of African slaves with Native Americans' slaves [Schwartz, 1978] and with Europeans' servants [Coelho and McGuire, 1997] could reflect health component of productivity differences. However, Native American, European and African labor were heterogeneous along many different dimensions. For instance, Native Americans lacked some of the agricultural skills that Africans and Europeans mastered Wood [1974]. Moreover, since malaria spreads in hot and humid climates, it is possible that African labor had been preferred just because European servants had no familiarity with the weather of the American tropics. More importantly, we cannot dismiss the importance of racist beliefs in supporting the whole institutions of slavery, as myths of African inferiority surely acted as ideological foundations for the legitimization of savagely exploitative behaviors [Krieger, 1987].

In this article, I propose three empirical strategies in support of the the African Slave Trade 'disease hypothesis'. First, I document a strong positive cross-sectional correlation between geographic suitability to malaria and the recourse to slavery across 1790 United States counties. Given the paucity of historical information available at a country level and the numerous potential omitted factors driving cross-country (or cross-region) correlations, I focus on a within country analysis in order to get rid of unobserved commonalities in the colonization experience. In order to exploit only the exogenous component of malaria exposure, following Kiszewski, Mellinger, Spielman, Malaney, Sachs, and Sachs [2004], I employ an estimated index of malaria risk predicted on the bases of exogenous environmental and geographic characteristics. Additionally, I exploit a novel measure of malaria endemicity at the beginning of the twentieth century produced by Lysenko [1968]. The results document a strong positive correlation between the ratio of slaves over total population in the county and malaria risk. Importantly, the correlation is robust to the inclusion of measure of soil suitability for sugar, cotton, rice, tea and tobacco in the county. Despite a visually striking spatial correspondence between slave counties and malarial counties, I am unable to exclude that some unobservable characteristics, such as the fertility of the soil or local climatic features, might be correlated with the index of malaria risk and thus drive my results. Therefore, aiming at controlling for time-invariant unobservable characteristics of slave regions versus non slave regions that a cross-sectional regression may hide, I exploit a panel of 12 northern American states in the decades from 1640 to 1770. For identification, I exploit the time variation arising from the introduction of malaria falciparum into the US colonies, and the additional fact that malaria could thrive only in a subset of American colonies because of climatic characteristics that malaria requires. More precisely, I rely on historical evidence suggesting that the plasmodium of *malaria falciparum*, i.e. the more severe and often deadly form of malaria, was likely introduced in North America in the 1680s [Wood, 1974]. Thus, I compare the percentage of slaves in the US colonial states that were more suitable to malaria with states that were less suitable before and after the introduction of malaria falciparum, and show that the introduction of the most debilitating form of malaria sharply increased the number of blacks over total population *only* in regions with higher preconditions for malaria incidence.

Finally, to show that resistance to malaria increased the profitability of African labor, I document the existence of a malaria premium. In other words, since health and resistance to the local disease environment affected the productivity of slaves, I expect to see higher prices paid for healthier slaves. Given that acquired resistance to malaria are primarily acquired during early childhood and that genetic resistance to malaria are higher in regions historically most exposed to the disease, I proxy slave resistance to malaria with the level of malaria incidence in the country of origin of the slave. Using historical slave prices from the Louisiana slave market, I assemble a dataset of prices for over 3500 slaves, born in 21 different Africa countries. I show that slaves born in African countries with more malaria commanded higher prices. The results are robust to the inclusion of a full set of geographical, climatic and institutional controls. The fact that slaves originally from more malarial regions sold at higher prices suggests the existence of a premium for slaves with higher resistance to the malarial environment of southern plantations.

## 1.2 Conceptual Framework

**The Geography of Immunities and Resistance in 1500** With the discovery of the Americas, two previously isolated continental blocks came into contact and a formidable exchange of plants, animals and technologies took place. The long list of Old World diseases introduced into the New World represented one of the most dramatic sides of the "Columbian Exchange"[4] [Crosby, 2003, Mann, 2011, Nunn and Qian, 2010]. In order to understand the massive population movements that followed the discovery of the New World, it is useful to draw a map of the geography of diseases prevailing at the eve of the 'Columbian Exchange' by grouping infections into two main categories: *human-contact civilization diseases* and *tropical diseases.* Human-contact civilization diseases - such as smallpox, measles, and flu - were endemic in all the Eurasian block and in large parts of Africa. Their distribution in the Old World landmass was a strict consequence of their origins: almost all civilization diseases originated from the proximity between man and domestic animals (such as cattle, pigs, and chickens) and - since they

---

[4] Smallpox, measles, chickenpox, influenza, typhus, typhoid and parathyroid fever, diphtheria, cholera, bubonic plague, scarlet fever, whooping cough, malaria and many others.

were transmitted by human contact - could virtually spread across all latitudes just requiring sufficiently numerous human population to sustain themselves[5]. Differently, tropical diseases - such as malaria, yellow fever, hookworm and dengue fever - which relied on vectors for transmission, were largely confined to tropical, semi-tropical and temperate areas where the vectors (mosquitoes for malaria and yellow fever, snails for schistosomiasis) could find a suitable environment[6]. The geographical isolation of the Americas spared the continent from both types of infectious diseases. First, the New World had a relatively low number of animals available for domestication, thus less scope for the development of indigenous infections[7]. Secondly, the relative scarcity of diseases was a direct consequence of the way the continent was populated during the migration of humans out-of-Africa: small bands of man migrated to North America through the Bering Strait, so that no vector disease could complete the voyage in the cold weather of the Strait and very few human-contact diseases could sustain themselves in these small migrating bands [Diamond and Ford, 2000, Dobson and Carper, 1996, McNeill, 2010, Wolfe, Dunavan, and Diamond, 2007].

The geography of immunities follows directly, as a mirror image, the distribution of infectious diseases. Since, wherever infections are endemic, immunities and resistance are acquired in infancy, Old World population had developed forms of defensive responses to all civilization diseases. On the contrary, protective immunities to tropical diseases were widespread only in areas where these diseases were sufficiently frequent. Moreover, historically malaria have represented such a burden that, beside acquired immunities, a whole set of genetic traits were 'selected for' over the centuries in order to protect the most afflicted populations. Inherited innate resistance to malaria is a complex mix of genetic factors granting different types of protection: i) reducing the risk of developing parasitemia, ii) once parasitizes, reducing the risk of becoming ill with fever, iii) once infected with malaria, reducing the risk of developing severe malaria [Kwiatkowski, 2005]. The most well-known and studied genetic resistance to malaria are a set of blood cell abnormalities, which we summarize in Table 1 [Allison, 2006, Carter and Mendis, 2002][8].

---

[5] If a disease infects only human and lacks an animal reservoir, it will require a large population of susceptible individuals as human hosts.

[6] Note, however, that in the summer one variety of malaria, malaria vivax, reached even coastal northern European regions such as the Scotland and Finland.

[7] Llamas and alpacas, traditional herd animals in South America, tend to live in small groups high in the mountains, with limited contacts with human settlement.

[8] There is no clear-cut evidence of any genetic resistance to yellow fever and hookworms, despite a large amount of historical narratives and studies reporting substantially lower morbidity and mortality rates registered for Africans in the American colonies (Wood [1974] and Kiple and King [2003], Coelho and McGuire [2006] and McNeill [2010]). Moreover, being infected with yellow fever lead to a lifelong acquired immunity, more likely for people born and raised in areas exposed to frequent epidemics.

Despite the limited knowledge and understanding of infectious diseases prevalent at the time of the New World colonization, the asymmetries in the resistance to diseases among Native Americans, Europeans and Africans struck all observers and became immediately a matter of facts. European settlers were rapidly taking control of a vast sparsely populated land with an enormous potential for agriculture, and this spurred a substantial demand for labor. The labor pool which colonizers ended up exploiting, with different intensity depending on the area and the period, was basically constituted of Native American labor (both slave labor and wage labor)[9], African slave labor and European servants.

**Enslaving Native Americans** At first contact with Europeans, the immediate introduction of several Old World diseases caused massive epidemics involving indigenous people throughout the Americas[10]. Nevertheless, Natives represented for a long time the main source of labor exploited by colonists, somewhere in the form of slavery, somewhere of barter and somewhere else in form of wage labor, depending on the Indians reactions towards European's requests. Spanish and Portuguese shipped indigenous slaves from denser areas to already depopulated areas where profitable activities, such as mining or plantations, required labor. For instance, 40,000 Lucanians were moved from their homeland in the space of five years, Las Casas (1542) commented:

> *"most Spanish ships could voyage without compass or chart, merely by following for the distance between Lucanians Island and Espanola, which is sixtly or seventy leagues, the trace of those Indian corpses floating the sea, corpses that had been cast overboard by earlier ships*[11]*."*

The high mortality of Natives pushed Spaniards and Portuguese to impose restrictions to their enslavement, but also dissuaded slave-owners from investing in Natives' labors, as they argued

> *"so many die that even at the lowest price they are expensive*[12]*."*

As soon as the first African slaves commenced to be shipped to Brazil and central America, slaveholders were willing to pay higher prices for them. And not only African slaves sold at higher prices in all European colonies, but Africans were also paid higher wages than Natives where wage labor was in place Schwartz [1978], suggesting a possibly health-related differential in productivity on top of the differential in mortality rates.

---

[9] Especially in Latin America which was more densely inhabitedCrosby [2003].

[10] There is no consensus among historians and demographers over the size of the decline. See Livi-Bacci [2006] and Borah and Cook [1969] for two different positions in the debate.

[11] [Watts, 2000] in Carribean Slavery in the Atlantic World, page 138

[12] [Schwartz, 1978], page 71

**Indentured Servitude of Europeans** For several decades, European indentured servants[13] had been the major source of labor in many areas of the New World, before being replaced in tropical and semi-tropical areas with African slave labor. It is estimated that between half and two thirds of all white immigrants to the American colonies after 1630s came under indenture, and that up to 75% of Virginian settlers in the seventeenth century were servants [Galenson, 1981]. Even in the Caribbeans, European servants constituted the majority of field laborers until about 1660s, such that the first phase of sugar revolution was based largely on white servants' labor [Downes, 1987]. Differently from slaves, white servants were migrating to the New World voluntarily and for them the health environment was one of the key variables for deciding whether and where to migrate. Up to a certain extent, information on health conditions in the colonies reached the mainland country, despite various attempts of the colonial governments to hide news of endemic diseases and epidemics to potential settlers [Wood, 1974]. Moreover, it was not uncommon for settlers to move to safer areas after experiencing directly the effective health conditions at destination. For instance, after the introduction of malaria *falciparum*, South Carolina started to be considered "the great charnel house of the country" and found increasing difficulties in both attracting new Europeans and keeping the ones already settled[14].

**African Slave Labor** Spaniard settlers, surely erring by overstatement, were used to claim:

> *"if we did not hang a Negro, he would never die*[15]*."*

Indeed, European settlers rapidly reached the conviction that Africans were more resistant to tropical diseases than Europeans and indigenous people. Curtin [1968] provides some of the most suggestive statistics documenting a substantial differential in mortality rates between Africans and Eurpeans in the American tropics. Relying on data of British soldiers serving in the Caribbeans, he looks at mortality rate for Africans British soldiers and European British soldiers sharing similar diets and working conditions. According to his findings, European troops in Jamaica experienced an annual mortality rate of 130 out of 1000, while for African troops the rate was of 40 out of 1000. Such figures were no surprise for settlers and slaves, as it seems apparent from the lyrics of a song that African slaves in Jamaican plantations dedicated to the white man *buckra*:

---

[13] Under the contract of servitude called indentures, the emigrant agreed to work for a designated master during a fixed period of time in return for passage to a specified colony [Galenson, 1981].

[14] To the point that South Carolina's government adopted laws to forbid new settlers to migrate [Wood, 1974].

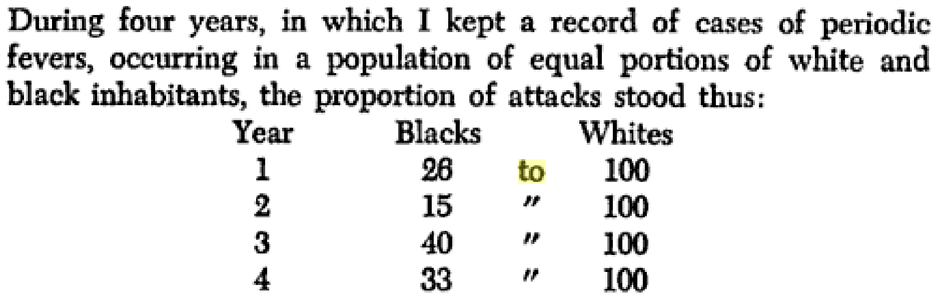
[15] Kiple [1986].

> *”New-come buckra, He get sik, He tak fever, He be die, He be die*[16]*.”*

Along a similar line, white planters all across the New World grew convinced that Africans slaves represented the most productive form of labor they could employ. In north American semi-tropical regions similar views were widespread:

> *”The old plantation was situated in rich lands, abounding in malaria, against which only the negro was proof*[17]*.”*

The health differential was so apparent to attract inquiry of the scientific community. Dr. Alfred Tebault, in 1856 on *American Journal of the Medical Sciences*, was reporting the results of his studies of malaria incidence on African and White Americans [Savitt, 2002]:



During four years, in which I kept a record of cases of periodic fevers, occurring in a population of equal portions of white and black inhabitants, the proportion of attacks stood thus:

| Year | Blacks | | Whites |
|---|---|---|---|
| 1 | 26 | to | 100 |
| 2 | 15 | ” | 100 |
| 3 | 40 | ” | 100 |
| 4 | 33 | ” | 100 |

Figure 1.1: Malaria Susceptibility: Blacks and Whites

Interestingly, Africans did not benefit an absolute health advantage, on the contrary, they tended on average to die earlier than Europeans, especially in cold weather areas[18]. Africans had an advantage specifically on warm weather diseases, this can be easily seen in the work of Rutman, Wetherell, and Rutman [1980], which shows a seasonality pattern in death for seventeenth century Virginia: African death peaked in the winter just as Europeans death, however European death records had a second autumnal peak, precisely when malaria is known to take its heaviest tool.

Slaveowners’ perceptions of differential susceptibility to diseases went even further than

---

[16] McNeill2010.

[17] [Mallard, 1892].

[18] Note, however, how it is virtually impossible to separate the role played by worse living conditions of Africans and Africans descendants from different susceptibilities to diseases when commenting on such figures. In other words, we can quite safely assume that living conditions of Africans were on average no worse than living conditions of Europeans (in terms of housing, food intake etc...). Therefore, when observing a lower incidence of diseases for Africans than Europeans, we can less fearfully attribute this figure to the difference in susceptibility than when observing data documenting a morbidity differential to the advantage of Europeans.

this, to the point that planters' could discriminate the health susceptibility even among Africans:

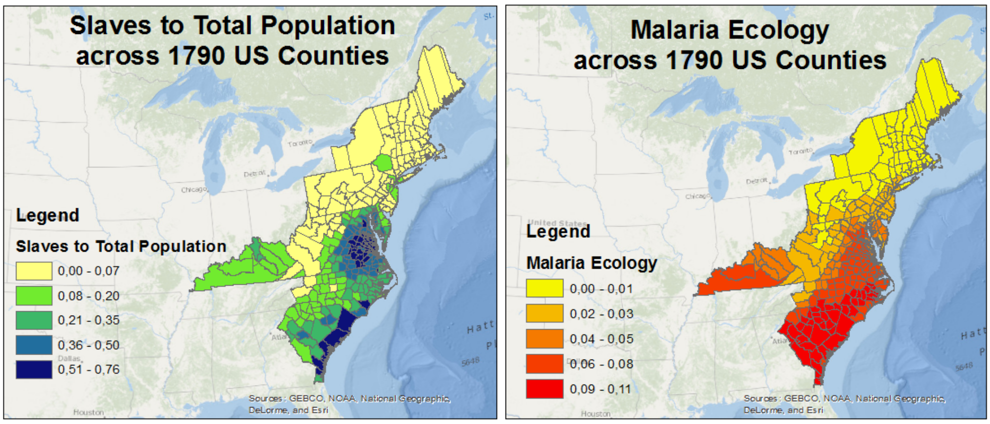
> *"This dislike [of Congos] was due to the West Central Africans' below-average height, the ill-health they experienced on lowland plantations, ... relatively little agricultural experience... planters preferred Africans from the Bight of Benin."*

## 1.3 Empirical Analysis

### 1.3.1 Malaria and Slavery across 1790 United States Counties

**Data and Descriptive Statistics** As it can be easily observed in Figure 1.2, already in 1790, slavery was not practiced homogeneously in all United States' counties. All Northern States had a very low number or no slaves registered in the Census, two states - Vermont and Massachusetts - had already abolished slavery and other three states - New Hampshire, Connecticut and Rhode Island - were in the process of becoming "free states". On a completely different track, in North Carolina and South Carolina about one fourth of the population was a slave, and up to one third in Virginia and Georgia. The picture was heterogeneous even within states, so that in South Carolina we see counties with a slave population reaching 70% of all registered population, and counties with less than 10% The distribution of African and African descendants - registered as non whites in the census - followed very closely the distribution of slavery[19].

FIGURE 1.2: Malaria and Slavery across 1790 US Counties



---

[19] The ratio of slaves over total population correlates at 0.99 with the ratio of blacks in the total population. Note that under nonwhite are included also mixed races

Malaria is a warm-weather disease. In order for malaria to spread, there needs to be a pool of mosquitoes to transmit the infection. The more the mosquito vectors of a region are used to feed on human blood, and the easier it is for the vectors to benefit of a warm and humid environment for breeding, the more malaria is harmful to man. Kiszewski, Mellinger, Spielman, Malaney, Sachs, and Sachs [2004] incorporated all these insights into a geographical suitability measure of malaria incidence. The measure that Kiszewski, Mellinger, Spielman, Malaney, Sachs, and Sachs [2004] produced is a stability index which, by taking into account mosquito habits and climatic features, predicts how common is the danger of being infected with malaria. In this paper, next to this measure, we exploit a new historical index of malaria endemicity at the beginning of the twentieth centuty, produced by Lysenko [1968] and digitalized by Hay S.I. [2004]. In this map, we have a disaggregated classification of the world based on the level of endemicity of malaria[20]. The fact that the measure predates large scale public health intervention for malaria eradication is partly reassuring insofar as it reduces the scope for endogeneity.

**Results** Tabel 1.4 reports regression estimates. Given that the malaria stability index in our sample goes from 0 to 0.10, going from counties where the index predicts no malaria to counties with the maximum malaria stability in the sample increases the ratio of slaves over total population by about 0.43, which is a substantial change, given that the average slave ratio is 0.22. Unsurprisingly, the pattern is confirmed with the other historical index of malaria incidence, when controlling for suitabilities to cotton, sugar, rice, tea and tobacco[21]. Finally, we find the same picture when looking at the ratio of blacks over total population (see Table 3.5).

### 1.3.2 The Introduction of Malaria Falciparum into the Colonies

**Identification Strategy** Cross-section results can be fundamentally flawed, insofar we cannot rule out the bias from omitted factors. In this section we propose an identification strategy that exploits the timing in the introduction of a dangerous strain of malaria - *plasmodium falciparum* - into the colonies. While *plasmodium vivax* was likely introduced into the North American colonies as early as the late sixteenth century Mann [2011], the introduction of *plasmodium falciparum*, the most virulent form of malaria, is dated around 1680, see Wood [1974] for South Carolina and Rutman and Rutman [1976] for Virginia. Importantly malaria, unlike human-contact diseases, requires specific climatic conditions to spread. For this reason, once the malaria *falciparum* was introduced

---

[20] The measure goes from 0, no transmission, to 5 holoendemic (transmission occurs all year long). The intermediate steps are epidemic, hypoendemic (very intermittent transmission), hyperendemic (intense, but with periods of no transmission), mesoendemic (regular seasonal transmission).

[21] Coffee is another crop traditionally associated with slave plantation, however, there is no variability in coffee suitability across our sample of counties.

into the colonies, it eradicated itself only in warm and humid states. Moreover, since the population was completely not immune to malaria *falciparum*: it first struck in the form of epidemics, as according to Waring [1975] was the case for South Carolina in 1684.

**Data and Descriptive Statistics** Since we do not have precise figures on the precise timing of malaria epidemics into US pre-federal states, we propose a sort of dif-in-dif exercise. We create a dummy variables taking value 0 before 1680 and 1 afterwords, then, we interact this variable with our index of malaria stability. What we search for is a differential effect of the change in blacks over total population before and after 1690 comparing states where malaria can thrive with states where the climate does not permit so. Population figures are taken from the pre-federal US census. The unit of observation is the decade, and we are able to assemble a panel going from 1640 to 1780.

**Results** Figure 1.3 graphically summarizes main results. We see that after 1690, the fraction of black over total population rapidly raised only in malaria-suitable states, and not in the others. Regression results confirm graphical findings, which are robust to the inclusion of time-varying soil suitability indexes[22].

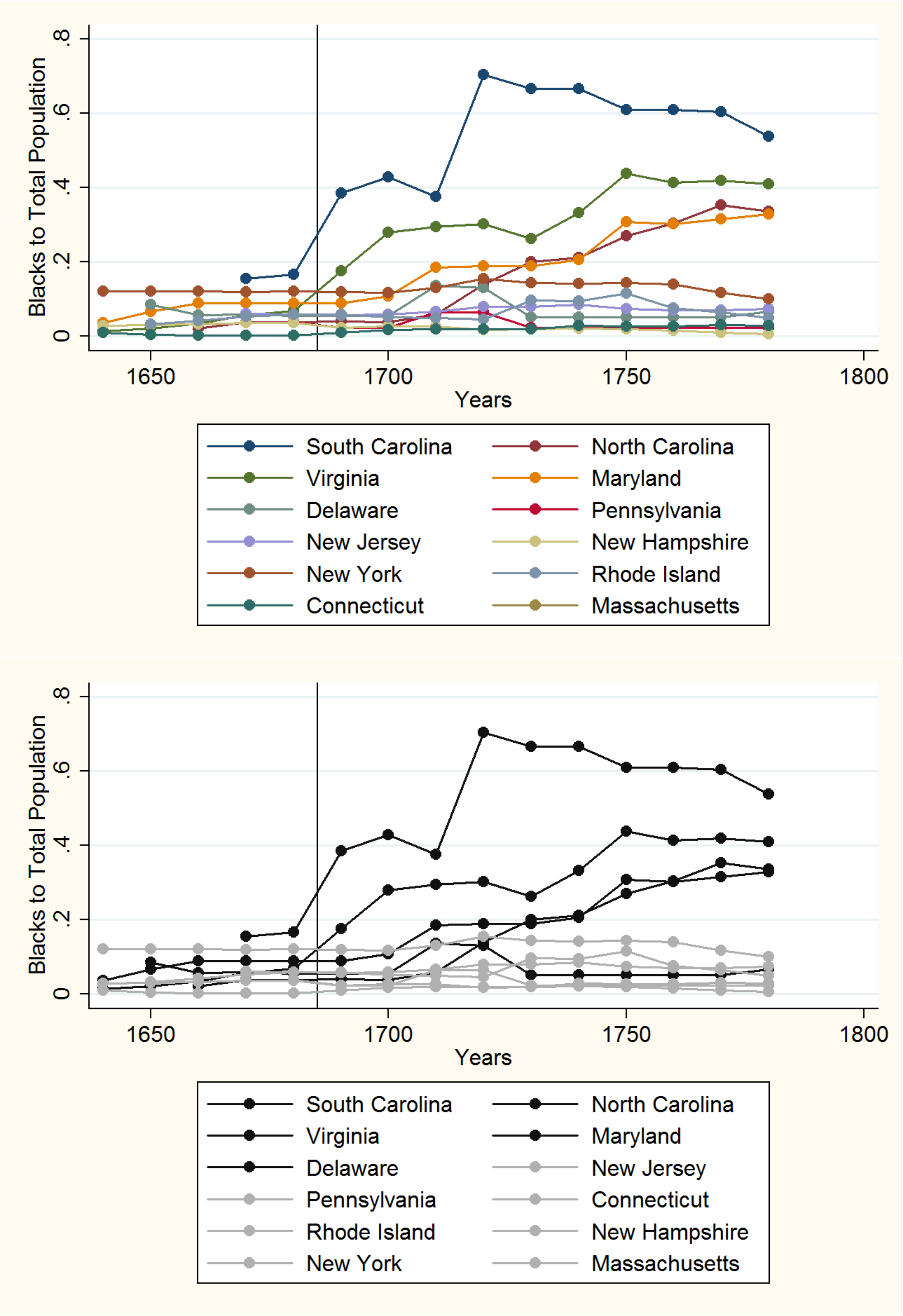
### 1.3.3 Malaria Resistance and Slave Prices

**Malaria Premium** Up to now, we are unable to exclude that African slaves represented the only available source of cheap labor for profit-seeking planters in diseased regions of the Americas. With our last strategy, we aim to provide evidence of slave-owners' preferences for slaves with higher resistance and immunities to the local disease environment. For doing so, we search in the price data a malaria premium. Since the market for slaves was a highly competitive market, as health and resistance to local diseases are known to affect the productivity of slaves, we expect to see higher prices paid for healthier slaves. Since malaria represented one of the main health threat in the slave regions, we expect to see a premium for slaves with higher innate and acquired resistance to malaria. In order to avoid comparisons between largely heterogeneous labor sources, we focus exclusively on slaves born in Africa. Importantly, we proxy the level of immunities to malaria of each slave with the malaria stability index in his/her country of birth. As epidemiology suggests, protective immunities are acquired in infancy and, moreover, in most malarial areas innate resistance are widespread.

---

[22] In other words, we interact the suitability indexes with each decade dummies, in order to allow for a time-varying effect of soil suitabilities on the faction of blacks over total population.

Figure 1.3: Blacks over Total Population in US Colonies



[†] The two graphs show the ratio of Africans to total population in the US pre-feder states from 1640 to 1780, in the panel below, the line is black for all states with a malaria stability index higher than the average and grey for all states with an index above the average.

FIGURE 1.4: Slaves' Advertisements

*Charleſtown, July 24th, 1769.*
TO BE SOLD,
On THURSDAY the third Day
of AUGUST next,
A CARGO
OF
NINETY-FOUR
PRIME, HEALTHY
NEGROES,
CONSISTING OF
Thirty-nine MEN, Fifteen BOYS,
Twenty-four WOMEN, and
Sixteen GIRLS.
JUST ARRIVED,
In the Brigantine DEMBIA, *Francis Bare*, Maſter, from SIERRA-LEON, by
DAVID & JOHN DEAS.

---

TO BE SOLD on board the
Ship *Bance Yland*, on tueſday the 6th
of *May* next, at *Aſhley-Ferry*; a choice
cargo of about 250 fine healthy
NEGROES,
juſt arrived from the
Windward & Rice Coaſt.
—The utmoſt care has
already been taken, and
ſhall be continued, to keep them free from
the leaſt danger of being infected with the
SMALL-POX, no boat having been on
board, and all other communication with
people from *Charles-Town* prevented.
*Auſtin, Laurens, & Appleby.*

*N. B.* Full one Half of the above Negroes have had the
SMALL-POX in their own Country.

**Data** We exploit a rich recently assembled database of slave transactions that took place in the Louisiana slave market between 1720 and 1820The database was designed and created by Gwendolyn Midlo Hall, see Hall [2005]. The database is an unbalanced panel of over 100,000slave transactions containing information such as age, sex, place of birth[23], race and price of the slave sold are available[24].

**Results** Table 1.6 and A.7 summarize main results. A one point increase in the malaria suitability index raises the price of the slave by 2%. As malaria suitability might be correlated with a whole set of geographical characteristics, I control for temperature, precipitation, elevation, ruggedness, soil suitability, distance to the coast and humidity. In Table A.8, I report results of regressions including potentially endogenous, but virtually relevant controls, such as proxy of agricultural and technological sophistication. The correlation is robust to the exclusion of each African country and of each of the four macro-region.

## 1.4 Tables

## 1.5 Conclusion

The paper "Side-Effects of Immunities: The African Slave Trade" advances a novel hypothesis about the **origins of slavery** in the tropical areas of the New World. Why slavery was practiced in certain American regions and not in others, and why Africans, and certain Africans in particular, were transported to the New World so numerously and enslaved is an economically relevant question. In fact, there is ample evidence that the American regions that relied on slavery in colonial times, as well as the African regions from which the slaves came, are poorer today (Engerman and Sokoloff [1997], Nunn [2008], Dell [2010]). Borrowing insights from historians, such as Curtin [1968], Wood [1974], Coelho and McGuire [1997], Kiple and King [2003] and McNeill [2010], I argue that the immunities of Africans from tropical regions to the disease environment that was prevalent in the American South made them especially attractive for employment in these regions. In other terms, Africans' resistance to malaria and other tropical diseases increased the profitability of African slave labor, and of African slaves from more malarial countries in particular. To verify the hypothesis, I exploit the time variation arising from the introduction of malaria falciparum into the US colonies, and the additional fact that malaria could thrive only in a subset of American colonies because of climatic characteristics that malaria requires. Thus, I compare the percentage of slaves in the US colonial states that were more suitable to malaria with states that were less suitable before and after the introduction of malaria falciparum. Finally, to show that resistance to malaria increased the profitability of African labor, I document the existence of a **malaria premium** using historical slave prices. Namely, I show that, among Africans, slaves born in African countries with more malaria commanded higher prices.

---

[23] While information on the race of the slaves (whether native, black, mulatto...) are available for a large fraction of our sample. We know the place of birth of about one tenth of the slaves in the sample.

[24] As standard in the literature that examined slave prices, we consider only slaves sold individually, and for which sale prices are available. A little less than one half of the slaves in the sample were sold as groups, and we are obliged to drop those from the sample.

Table 1.1: Common Polymorphisms That Affect Resistance to Malaria

| | **Type of Protection** | **Geographical Distribution** |
|---|---|---|
| **Thalassemias** | Approximately 50% reduction in the risk of malarial disease | High frequencies around Mediterranean sea shores, through most of Africa, Middle East, Central Asia, Arabian peninsula, Indian subcontinent, Southeast Asia, southern China, and Western Pacific Island (from Philippines to New Guinea and Melanesia) |
| **Sickle Cell Trait** | Approximately 90% protection against P. falciparum malarial mortality | In many parts of Africa, frequent at 30% |
| **G6PD** | Approximately 50% protection against severe P. falciparum malaria | same as Thalassemias |
| **Hemoglobin C** | Approximately 90% protection against P. falciparum malarial infection in the homozygote (30% in a heterozygous combination) | among certain West African population, frequent at 10%-20% |
| **Hemoglobin E** | May protect against P. vivax, and clear P. falciparum infection more rapidly | High frequencies in population across South-East Asia |
| **Ovalocytosis** | Reduced risk of P. vivax and P. falciparum infection | New Guinea (up tp 20%), Solomon Islands and Vanuatu |
| **RBC Duffy Negativity** | Complete refractoriness to P. vivax infection | More than 95% frequencies in West and Central Africa, at lower frequencies through the Arabian peninsula, across the Middle East and to the edges of Central Asia |

Table 1.2: Health Reports from South Carolina

| **Year** | **Source** | **Opinion on the Health of the Colonies** |
|---|---|---|
| **1674** | Joseph West | *Our people (God be praised) doe continue very well in health and the country seemes to be very healthfull and delightsome.* |
| **1671** | John Locke | *The rivers generally run through marshes which are not unhealthy.* |
| **1684** | Sir Peter Colleton, contingent of Scottish settlers | *We found the place so extraordinarily sicklies that sickness quickley seased many of our number and took away great many...* |
| **1682** | Samuel Wilson | *Who in this Country seated themselves near great marshes are subject to Agues.* |

TABLE 1.3: Malaria and Slavery across US Counties

| | **Dependent Variable: Colored to Total Population in 1790** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | PANEL A<br>Malaria Ecology - Full Sample | | | | | | | |
| Malaria Ecology | 4.280 | 5.941 | 7.143 | 6.037 | 5.726 | 5.548 | 6.191 | 5.241 |
| Conley s.e. 100 km | [0.571]*** | [1.062]*** | [1.205]*** | [1.073]*** | [1.059]*** | [0.976]*** | [1.029]*** | [1.016]*** |
| Cluster (State) s.e | (1.005)*** | (1.781)*** | (2.217)*** | (1.944)*** | (1.851)*** | (1.073)*** | (1.975)*** | (1.187)*** |
| *Crop Suitability Indexes* | | | | | | | | |
| Cotton Suitability | No | No | Yes | No | No | No | No | Yes |
| Sugar Suitability | No | No | No | Yes | No | No | No | Yes |
| Rice Suitability | No | No | No | No | Yes | No | No | Yes |
| Tea Suitability | No | No | No | No | No | Yes | No | Yes |
| Tobacco Suitability | No | No | No | No | No | No | Yes | Yes |
| State FE | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 285 | 285 | 285 | 285 | 285 | 285 | 285 | 285 |
| R-squared | 0.349 | 0.593 | 0.614 | 0.601 | 0.640 | 0.607 | 0.655 | 0.712 |
| | PANEL B<br>Malaria Ecology<br>Full Sample | | | | PANEL C<br>Malaria Endemicity<br>Full Sample | | PANEL D<br>Malaria Ecology<br>Only Slave States | |
| Malaria Ecology | 3.801 | 5.572 | 5.451 | 4.293 | | | 5.976 | 4.237 |
| Conley s.e. 100 km | [0.826]*** | [0.935]*** | [1.187]*** | [0.905]*** | | | [1.067]*** | [0.929]*** |
| Cluster (State) | (1.330)** | (1.139)*** | (1.313)*** | (1.475)** | | | (1.807)** | (1.500)** |
| Malaria Endemicity | | | | | 0.129 | 0.045 | | |
| Conley s.e. | | | | | [0.025]*** | [0.018]*** | | |
| Cluster (State) | | | | | (0.038)*** | (0.019)** | | |
| *Crop Suitability Indexes* | Yes | Yes | Yes | Yes | No | Yes | No | Yes |
| *Distances* | | | | | | | | |
| Sea Distance | Yes | No | No | Yes | No | Yes | No | Yes |
| River Distance | No | Yes | No | Yes | No | Yes | No | Yes |
| Distance Charleston | No | No | Yes | Yes | No | Yes | No | Yes |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 285 | 285 | 285 | 285 | 285 | 285 | 244 | 244 |
| R-squared | 0.795 | 0.725 | 0.713 | 0.800 | 0.602 | 0.772 | 0.499 | 0.760 |

*Notes:* Table reports OLS estimates. The unit of observation is the 1790 US county. The dependent variable is the ratio of "colored" people to total population. Malaria Ecology is an index measuring the force and stability of malaria transmission, Malaria Endemicity measures the level of malaria parasite rate at the beginning of the 20th century. Conley standard errors are reported in squared brackets, standard errors clustered at the state level are reported in parenthesis. ***, **, * indicate significance at 1-, 5-, and 10-%level, respectively.

TABLE 1.4: Malaria and Slavery across US Counties

| **Dependent Variable: Slaves to Total Population in 1790** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | PANEL A<br>Malaria Ecology - Full Sample | | | | | | | |
| Malaria Ecology | 4.289 | 5.672 | 6.820 | 5.768 | 5.461 | 5.289 | 5.907 | 4.997 |
| Conley s.e. | [0.550]**0 | [1.017]*** | [1.143]*** | [1.029]*** | [1.014]*** | [0.922]*** | [0.960]*** | [0.958]*** |
| Cluster (State) s.e | (0.960)*** | (1.781)*** | (2.166)*** | (1.940)** | (1.849)** | (1.095)*** | (1.959)*** | (1.162)*** |
| *Crop Suitability Indexes* | | | | | | | | |
| Cotton Suitability | No | No | Yes | No | No | No | No | Yes |
| Sugar Suitability | No | No | No | Yes | No | No | No | Yes |
| Rice Suitability | No | No | No | No | Yes | No | No | Yes |
| Tea Suitability | No | No | No | No | No | Yes | No | Yes |
| Tobacco Suitability | No | No | No | No | No | No | Yes | Yes |
| State FE | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 285 | 285 | 285 | 285 | 285 | 285 | 285 | 285 |
| R-squared | 0.366 | 0.601 | 0.621 | 0.609 | 0.648 | 0.615 | 0.658 | 0.716 |
| | PANEL B<br>Malaria Ecology<br>Full Sample | | | | PANEL C<br>Malaria Endemicity<br>Full Sample | | PANEL D<br>Malaria Ecology<br>Only Slave States | |
| Malaria Ecology | 3.678 | 5.301 | 5.252 | 4.193 | | | 5.704 | 4.145 |
| Conley s.e. | [0.803]*** | [0.884]*** | [1.126]*** | [0.881]*** | | | [1.022]*** | [0.909]*** |
| Cluster (State) | (1.322)** | (1.111)*** | (1.298)*** | (1.465)** | | | (1.809)** | (1.496)** |
| Malaria Endemicity | | | | | 0.126 | 0.050 | | |
| Conley s.e. | | | | | [0.022]*** | [0.017]*** | | |
| Cluster (State) | | | | | (0.034)*** | (0.017)** | | |
| *Crop Suitability Indexes* | Yes | Yes | Yes | Yes | No | Yes | No | Yes |
| *Distances* | | | | | | | | |
| Sea Distance | Yes | No | No | Yes | No | Yes | No | Yes |
| River Distance | No | Yes | No | Yes | No | Yes | No | Yes |
| Distance Charleston | No | No | Yes | Yes | No | Yes | No | Yes |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 285 | 285 | 285 | 285 | 285 | 285 | 244 | 244 |
| R-squared | 0.788 | 0.727 | 0.717 | 0.793 | 0.615 | 0.767 | 0.505 | 0.749 |

*Notes:* Table reports OLS estimates. The unit of observation is the 1790 US county. The dependent variable is the ratio of slaves to total population. Malaria Ecology is an index measuring the force and stability of malaria transmission, Malaria Endemicity measures the level of malaria parasite rate at the beginning of the 20th century. Conley standard errors are reported in squared brackets, standard errors clustered at the state level are reported in parenthesis. ***, **, * indicate significance at 1-, 5-, and 10-%level, respectively.

TABLE 1.5: Malaria and Share of Blacks across US States

| | **Blacks to Total Population** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Malaria Ecology x Post 1690 | 4.222 | 3.599 | 3.822 | 6.145 | 3.899 | 4.923 | 6.232 | 5.890 | 9.373 |
| Cluster (State) s.e. | (0.825)*** | (0.943)*** | (1.064)*** | (1.097)*** | (1.098)*** | (1.563)*** | (1.625)*** | (1.123)*** | (1.830)*** |
| Bootstrap s.e. p-value | 0.004 | 0.032 | 0.036 | 0.012 | 0.024 | 0.028 | 0.008 | 0.06 | 0.04 |
| Rice Suitability x Post 1700 | No | Yes | No | No | No | No | No | No | No |
| Rice Suitability x Year FE | No | No | Yes | No | No | No | No | Yes | Yes |
| Tea Suitability x Year FE | No | No | No | Yes | No | No | No | Yes | Yes |
| Tobacco Suitability x Year FE | No | No | No | No | Yes | No | No | Yes | Yes |
| Cotton Suitability x Year FE | No | No | No | No | No | Yes | No | Yes | Yes |
| Average Temperature x Year FE | No | No | No | No | No | No | Yes | Yes | Yes |
| Decade FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | No |
| Country Time Trend | No | No | No | No | No | No | No | No | Yes |
| Observations | 166 | 166 | 166 | 166 | 166 | 166 | 166 | 166 | 166 |
| R-squared | 0.586 | 0.611 | 0.683 | 0.745 | 0.639 | 0.702 | 0.712 | 0.915 | 0.939 |
| Number of States | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 |

*Notes:* Table reports panel OLS estimates. The unit of observation is the US state, the panel includes all decades from 1640 to 1770. The dependent variable is the ratio of "colored" people to total population. Malaria Ecology is an index measuring the force and stability of malaria transmission. The variable Malaria Post 1690 is an indicator variable equals 1 from 1690 onwards, and 0 otherwise. All regressions include decade fixed effects and state fixed effects, except for the last Column which includes state specific time trend. Controls variable x Year FE are cross-sectional variables interacted with a full set of decade fixed effects. Standard errors clustered at the state level are reported in parenthesis. Since we only have 12 clusters, we report p-values for the null hypothesis (Malaria Ecology x Post 1690) computed with wild bootstrap standard erros. ***, **, * indicate significance at 1-, 5-, and 10-%level, respectively.

Table 1.6: Malaria in the Country of Origin and Slave Price

| | **Ln(Slave Price at Sale)** | | | | | |
|---|---|---|---|---|---|---|
| Malaria Ecology | | 0.018*** | 0.037*** | 0.020*** | 0.022*** | 0.032** |
| | | (0.006) | (0.010) | (0.004) | (0.002 ) | (0.012) |
| *Wild Bootstrap P-value* | | 0.014 | 0.008 | 0.002 | 0.004 | 0.028 |
| Distance Coast | | | 0.221** | | | 0.172* |
| | | | (0.091) | | | (0.095) |
| Ruggedness | | | | -0.120*** | | -0.158 |
| | | | | (0.028) | | (0.172) |
| Land Suitability | | | | | -0.158*** | 0.111 |
| | | | | | (0.052) | (0.252) |
| Slave Age | 0.051*** | 0.052*** | 0.051*** | 0.052*** | 0.052*** | 0.051*** |
| | (0.003) | (0.003) | (0.003) | (0.004) | (0.004) | (0.004) |
| Slave Age Squared | -0.001*** | -0.001*** | -0.001*** | -0.001*** | -0.001*** | -0.001*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Male Slave | 0.186*** | 0.190*** | 0.190*** | 0.189*** | 0.189*** | 0.189*** |
| | (0.019) | (0.019) | (0.019) | (0.019) | (0.019) | (0.019) |
| Region FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Document Language FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Document Type FE | Yes | Yes | Yes | Yes | Yes | Yes |
| R-squared | 0.451 | 0.453 | 0.454 | 0.454 | 0.453 | 0.454 |
| Observations | 3675 | 3675 | 3675 | 3675 | 3675 | 3675 |

*Notes*: Table reports panel OLS estimates. The unit of observation is the individual slave. The dependent variable is the natural logarithm of the slave price at sale. Malaria Ecology is an index measuring the force and stability of malaria transmission. All regressions also control for age, age squared, sex, language of the sale price document fixed effects, type of document fixed effects, region fixed effects (Upper Guinea, Bight of Benin, Western Africa and Southeastern Africa) and year fixed effects. Standard errors are clustered at the country level (21 clusters). Since we only have 21 clusters, we report p-values for the null hypothesis (Malaria Ecology = 0) computed with wild bootstrap standard erros. ***, **, * indicate significance at 1-, 5-, and 10-%level, respectively.

# Chapter 2

# Caste Systems and Technology in Pre-Modern Societies

## 2.1 Introduction

Caste systems have long been considered an extremely inefficient type of economic and social institution, insofar as they constrain social mobility, fractionalize societies and foster discrimination. The long term negative detrimental effects of castes on modern economic growth can be hardly questioned. Nonetheless, the depth and persistence of this very ancient form of social fractionalization[1] is suggestive of the possibility that castes, under different historical constraints, may had represented an asset, instead of a cost. In fact, caste systems have not simply been a special form of social stratification, but have constituted a complex institutional set up with important effects over the economic organization of societies.

In this article, we explore the possibility that caste systems had been comparatively advantageous at an early phase of economic development. In particular, we test the hypothesis that castes had positively affected technological sophistication and labor specialization in the institutional, economical and technological context of pre-modern societies. The hypothesis builds on the idea that caste systems, by promoting strong ties of solidarity and cooperation within groups, might had facilitated and accelerated the process of labor specialization and technological sophistication. In fact, the strong reciprocal feelings of trust might have contributed to solve common problems of production and exchange in pre-modern societies, such as the danger of information disclosure,

---

[1] The origin of Indian caste systems has been dated variously between 3000 B.C. to approximately 1000 B.C. [Thapar, 2004]. It is more complicated to date the origin of castes in Western and Eastern Africa, because of the lack of written records. Tamari [1991] summarized evidence regarding the emergence of caste systems in western Africa, which appeared no later than 1300 AD according to some sources, and no later than 1500 according to other sources.

the enforcement of contracts through collective action and the cost of transferring skills. Moreover, anytime that economies of scale emerged, the endogamous group might had represented the natural basin where to recruit labor for extending production beyond the family unit. On top of this, workers in the same endogamous group might had benefited of technological spillovers from other kins. In this sense, castes might had facilitated the development of artisan clusters in pre-modern societies.

We use data from Murdock's Standard Cross Cultural Sample and *Ethnographic Atlas*, two ethnicity-level datasets which permit to compare pre-modern societies before the European colonization. To our knowledge, these are the only data sources allowing to map the worldwide distribution of caste systems. For comparing technological sophistication in pre-modern societies, we need to pull far back the frontier of technology, back to a time where the mere presence of specialized occupational practices represented, comparatively, a leap forward. The first measure of technological sophistication we use is based on the presence of three key types of artisanal skills, i.e. pottery, weaving and metalworking. Indeed, less than one fifth of societies in our sample disposed of all these three basic types of technologies. The second index of technological sophistication we employ, inspired by the work of Comin, Easterly, and Gong [2006], attempt to measure the level of technological sophistication along several dimensions, and namely: agriculture, artisan crafts, writing and land transportation. Finally, we also investigate the degree of labor specialization. For this reason, we explore to what degree occupational practices are performed as specialized arts or occupations.

We find a strong positive relationship between caste stratification and the level of technological sophistication and labor specialization in pre-industrial societies. The relationship remains economically meaningful and statistically significant even after controlling for various potentially relevant geographical and environmental characteristics of the land occupied by the society, as well as location controls such as distance from rivers and seas. Nevertheless, the correlation may be driven by omitted factor, and a reverse effect of technological sophistication on caste stratification cannot be ruled out.

In order to understand whether the effect of castes on technology is causal, we follow three strategies. First, we try to control for all major established determinants of pre-modern development. The correlation between caste stratification and pre-modern technological sophistication is robust to all relevant controls: to the timing of transition to agriculture, the migratory distance from Adis Abeba and the distance from the regional technological frontier (Ashraf and Galor [2011a] and Ashraf and Galor [2011c]). Moreover, controlling for observable society characteristics, such as population density and the level of political centralization, reduces the magnitude of the correlation only slightly. This evidence suggests that for controlling for observable characteristics affect our results only mildly.

The second strategy we employ rely on an exogenous source of variation in the likelihood of observing caste stratification. Following historian William McNeill [1976], who argued that castes represented an early social attempt to reduce the risk of contagion in highly diseased environments, we test whether castes prevailed in areas that, for bio-geographical reasons, had to deal with a higher burden of pathogens. We construct an index of geographical disease exposure exogenous to population size and development level and we find that it significantly predicts a higher probability of observing societies with castes, suggesting thus a tentatively causal relation. Instrumenting the variable castes with our index of disease exposure, we find a strong and causal effect running from caste stratification to pre-modern societies technological sophistication and labor specialization.

This paper contributes to multiple strands of the literature. First, it complements the evidence of recent studies documenting the crucial role of social institutions and cultural norms for economic outcomes [Greif, 1993, 1994, Guiso, Sapienza, and Zingales, 2008, Nunn and Wantchekon, 2009, Tabellini, 2007] and, along the same line, the impact of social fractionalization on growth and institutional quality [Alesina and Ferrara, 1999, 2004, Easterly and Levine, 1997, Fearon and Laitin, 2003, among many others]. In particular, our study enriches this literature substantiating the idea that the economic effects of institutions and norms, as the effects of social fractionalization, are not fixed but vary with the historical context. Second, this research add new elements to the literature exploring determinants of historical development [Ashraf and Galor, 2011a,b, Comin, Easterly, and Gong, 2006].

Moreover, the article tangentially contributes to the debate over the role of institutions versus geography in economic development [Acemoglu, Johnson, and Robinson, 2001, Gallup, Sachs, and Mellinger, 1998, Sachs, 2001, Sokoloff and Engerman, 2000, among the various contributions]. In fact, from one hand, we document that institutions mattered for economic development even in pre-modern societies, insofar as we find that castes spurred technology and labor specialization. On the other hand, we slightly depart from the institutional view since the effect of geography on institutions that we find pre-dates European colonization. In the same way, our findings slightly diverge from the standard geography view as we show that, in pre-modern times, diseases did not negatively affect technology but, through caste stratification, indirectly favored labor diversification and early technological sophistication. Along the same line, by highlighting an effect of disease exposure on the presence of caste systems, our analysis complements recent literature on the geographical origin of institutions, norms and social diversity, which traces the roots of their historical emergence in the bio-geographic environments that societies had to confront [Alesina, Giuliano, and Nunn, 2011a,b, Michalopoulos, 2008]. Finally, this paper, despite focusing on pre-modern times, indirectly adds new elements to the literature exploring the impact of a diseased environment over various economic, social and political outcomes [Acemoglu and Johnson, 2006, Bloom, Canning, and Sevilla, 2003, Cashdan, 2001, Cervellati, Sunde, and Valmori, 2011a,b, Fincher, Thornhill, Murray, and Schaller, 2008, Gallup, Sachs, and Mellinger, 1998, Lorentzen, McMillan, and Wacziarg, 2005, Sachs, 2003, Weil, 2007].

In Section II, we begin our analysis by laying out the historical and conceptual framework. In Section III, we turn to a description of the data and present OLS estimates. In Section IV we present the identification strategy for pinning down the causal relationship and report our IV estimates. Section V concludes.

## 2.2 Historical Background and Conceptual Framework

### 2.2.1 Historical Background

Caste-like stratification existed - and often still persist - in various geographical regions and within ethnic communities at different levels of economic development. The most studied and known example of caste system is the Indian one, by far the most structured caste system we know, being at the time an economic, religious, social and political institution. But castes - or better, system of social stratification with significant similarities to the Indian one - are wide-spread in western and eastern Africa and in other parts of Asia[2]. Sure enough, caste-like stratifications vary widely across regions, complicating the task of cross-cultural analysis. For instance, in the Indian system virtually any Hindu was recognized as part of an endogamous caste, or *jati*, while in Western Africa caste people formed a small minority of the population [Tamari, 1991][3]. In order to be able to compare castes cross-culturally, we proceed by providing a parsimonious operational definition. Following Berreman [1968], we define caste systems as "a hierarchy of endogamous divisions in which membership is hereditary and permanent". This definition stresses two main features of caste stratification, and namely: the hereditary status - one belongs to a caste only if he/she is born in that caste - and the rigid endogamic nature of the group - one belongs to a caste only by marrying within the caste, otherwise risking expulsion.

The definition we employ is a sort of lowest common denominator of caste-like institutions found in all the societies that we will compare. However, in no way this definition

---

[2] Anthropologists and sociologists have been long debating over whether it is appropriate to name "castes" the caste-like institutions found outside the Indian peninsula. In this context, we take a step back from the debate and circumscribe our analysis by providing a clear operational definition of what we consider a caste.

[3] "They have been estimated to form ten to twenty per cent of the Wolof population, and about five per cent of the Soninke population. They also seem to form about five per cent of the Bambara population. Among the Dogon, and a fortiori among the Dan, Minianka and other peoples who recognize only one or two castes, they form a yet smaller proportion of the population" [Tamari, 1991].

exhausts the religious meanings, the functions and roles that castes have been playing cross-culturally. Besides these common features, there are several characteristics that have been found across the majority of societies examined, but not in all of them. First of all, in the majority of cases castes are associated with a traditional hereditary occupation, which is transmitted from father to son (or mother to daughter) and is often forbidden to other members of the community. Moreover, caste stratification is very frequently supported by ideological-religious beliefs of purity-pollution. Caste people - or lower caste people - are considered impure and thus to merit segregation and disdain. The impure nature of caste people is sometimes attributed to the profession practiced (i.e. washers, sweepers...), or to the dietary habits of the community (i.e. eating porks, dead animals etc...). Along the same lines, several communities prescribe limits to the interactions among members of different castes, the strict prohibition of commensality being the most frequent one [Dumont, 1980, Hutton, 1946, Tamari, 1991].

Our notion of caste might, in some cases, not permit a clear-cut distinction with other kinds of social institutions or other type of social fractionalization, like kin groups, ethnic groups and social classes. Indeed, these various types of social divisions often overlap. Still, caste institutions maintain their specificities. For instance, we might say that a caste is a sort of kin group, since kin groups sometimes have very precisely set endogamic rules which may recall the endogamicity of castes. However, very often kin groups marry outside the group, in many cases the relationships among groups have no hierarchical nature and they are rarely associated to a traditional occupation. For what concerns ethnic stratification, different castes sometimes belong to different ethnic groups. Theories over the origin of caste systems argued that castes originated from different ethnic groups which happened to form economically-integrated but socially segregated communities [Hutton, 1946][4]. However, in the majority of cases, ethnic differences are not apparent and castes of a same community tend to speak the same language. Finally, caste stratification might coincide with class stratification. For instance, in India the possession of land was often prohibited to outcasts and lower castes. At the same time, being granted an exclusive right to perform a specific occupation often permitted caste people to reach a considerable wealth:

> “Sanitary workers veiled themselves in Lagos to avoid recognition, but

---

[4] Recent evidence from genetics supported the idea that Indian castes are genetically distinct and might had originated from different tribal groups, however, the debate is still ongoing. From one hand Mountain, Hebert, Bhattacharyya, Underhill, Ottolenghi, Gadgil, and Luca Cavalli-Sforza [1995] found no clear genetic separation along caste lines, on the other hand, Bamshad, Kivisild, Watkins, Dixon, Ricker, Rao, Naidu, Prasad, Reddy, Rasanayagam, et al. [2001] documented that higher caste Indians are closer to Europeans and lower caste Indians to Asians. Similar studies are not available for other caste systems located in Asia and Africa. Nonetheless, ethnographic evidence acquired by anthropologists raised similar hypothesis regarding the origins of caste systems in other areas. See, for instance Shack [1964] on castes from South-West Ethiopia, and Barth [1956] on a caste system in Northern Pakistan.

> everywhere they showed much corporate solidarity and some achieved considerable economic success ... The Dorze weavers of Addis Ababa suffered extreme insult and discrimination but earned perhaps three or four times the national average in 1970[5]"

### 2.2.2 Conceptual Framework

Castes have long been considered an obstacle to economic growth. Weber [1958] was among the first social scientists pointing to an alleged incompatibility between modern economic growth and the caste system. First of all, a division of labor based on hereditary status instead of contracts may discourage entrepreneurship no less than forestalls workers' ambitions [Berreman, 1968, Hutton, 1946]. Beside the distortionary effects on workers' incentives, Bose [1916] considered the hereditary passing of crafts and skills as an intrinsic limit to the introduction of innovation, since it made an innovation in method appear as a sin against the craftsman's ancestors. In a recent analysis, Santacreu-Vasut [2009] argued that social diversity generates information asymmetries which cause the adoption of less growth-enhancing managerial institutions. She explained the diverging path of productivity between Indian and Japanese cotton industries as the result of a higher social diversity, and of the less efficient institutions which it engendered. Even within the same Indian state, factories with more diverse labor force tended to recur to less profitable managerial set ups, with more workers supervision when the labor force is diverse.

Moreover, castes might prevent an efficient allocation of talents across professions by prescribing hereditary occupational status or by discriminating workers of lower castes. The effects of Indian caste-based discrimination are dramatically persistent, in a recent experiment Hoff, Pandey, and Team [2004] showed how low-caste students tend to perform significantly worse anytime salience is placed on their caste status.

Beside the direct effects on productivity and innovation, caste-based fractionalization might affect economic growth even through indirect channels, such as the quality of institutions a society decide to adopt and the level of public goods to provide. Banerjee, Iyer, and Somanathan [2005] find that social heterogeneity in Indian villages weakens the political relevance of the community and reduces the amount of public goods it is able to secure. More broadly, caste stratification might damage the quality of institutions just as ethnic fractionalization has been found to do [Alesina and Ferrara, 1999, 2004].

All this evidence suggests to consider castes as extremely costly institutions, among other things, in terms of efficiency. Still, caste systems are present in historically relatively prosperous societies, and in the majority of cases have lasted till our days. As

---

[5] Iliffe [1987].

pointed by Greif [2006], institutions tend to be persistent, because of the significant adjustment costs to be faced for transforming them. On top of this, institutions are for the most part embedded in a complex matrix of cultural and social norms, which are intrinsically slowly changing [North, 1990]. In this light, it becomes relevant to inquire whether what represents a truly inefficient institution in the context-specificity of a modern society, might have provided some advantages under different technological, social, economic and institutional circumstances. In other words, can castes had been a booster of growth at a certain point in history? In this article, we test whether this had been the case. More precisely, our hypothesis build on the idea that the previously documented detrimental effects of castes on growth might have prevailed only at a relatively late stage of technological sophistication. And, on the contrary, caste stratification had boosted technological sophistication and specialization at an earlier phase. Indeed, caste organization might have had the crucial function of "artificially" spur the division of labor, which since Adam Smith has been recognized as the first driver of growth. In this light, Bougle's remark on the economic consequences of the Indian castes is particularly suggestive:

> "..the caste which mends shoes refrains from making a pair of them... it takes three distinct craftsmen to make a bow and arrows..."[6].

The role of castes at an early stage of technological sophistication has not been explored in the economic literature. On the contrary, anthropologists and historians have long been arguing in this sense. Diop and Salemson [1987] claimed that labor specialization strictly followed clan and caste organization. Along the same line, Nesfield (1901)[7], one of the first indianologist studying the economic features of the caste system, was reading the Indian system as an institutional device for achieving labor specialization.

A considerable size of the literature on castes had been emphasizing the exploitative traits of the economic relations in caste societies. And indeed, the power relationship between social groups in caste societies cannot be underplayed. Nevertheless, limiting the analysis to the exploitative side of caste stratification does not allow to fully account for some relevant features of the economic relationships in caste societies. In this sense, it is worth quoting again Diop and Salemson [1987]:

> "For each caste: inconveniences and advantages, alienations and compensations balance each other out... the stability of the caste system is secured by the hereditary transmission of social functions that corresponds, to a certain extent, to a monopoly, disguised in religious interdiction, in order to eliminate competition".

---

[6] From Hutton [1946].

[7] From Hutton [1946].

It is necessary to qualify that Diop is specifically referring to caste institutions in Wester Africa and, still, is probably underplaying some of the dramatic social costs of being part of a despised group. Moreover, the same analysis would misrepresent the truly dismal conditions of many Indian out-castes. Still, the economic analysis of castes merits further scrutiny and cannot be simply dismissed as a particular form of slavery. In this light, the question we need to explore is why we expect castes to favor technological sophistication and labor specialization in pre-modern societies.

One possible explanation is that the caste - the kin-group or the clan - was the natural boundary where to recruit whenever arose the possibility of exploiting economies of scale in production, and whenever the production process required an amount of labor larger than what could be found within the family unit. The fact that each caste tended to be geographically localized in a specific area of the village could represent a favoring pre-condition for artisan clustering, which is a common feature of pre-modern crafts because of the positive organizational and technological externalities it engendered[8].

Another possible explanation is that the trust and cooperative nature of the relationships which were likely to prevail within the endogamous group could solve well-known problems of production in pre-modern societies. First of all, confining the artisan knowledge to the group could work as an early form of intellectual property protection. It is suggestive to recall how, in medieval Europe, craft guilds emphasized the "brotherly" nature of the association through elaborate symbolic liturgies and ceremonies, with the primary scope of promoting feelings of trust and cooperation among members. In a caste, such trustworthiness was granted by birth status within the endogamous group and this common background might had reduced fears of information disclosure. Epstein [1998] stresses the importance of intellectual property protection for European guilds, arguing that "deliberate inventions will not be forthcoming if the inventor cannot claim more than his proportional share of the gains". Of the possible solutions to the problem (patent rights, state support for research, etc.), only secret transmission of crafts to trusted members was likely to be an available option in pre-modern societies.

Another possible explanation why endogamous groups might had found it convenient to restrict production within the caste is the potentially high access cost to artisanal skills in pre-literate societies. Stressing again the comparison with medieval Europe might provide some further insights. Epstein [1998] argued that craft guilds emerged in order to provide transferable skills through apprenticeship and to share out the unattributed costs and benefits of training among its members. Again, the transfers of skills within a group of kin might be safer and less costly.

Finally, in a context where enforcement institutions may had been missing or lacking,

---

[8] Epstein [1998] pointed to the tendency of pre-modern craft guilds to cluster in the same neighborhood as a common feature of European medieval cities. It is sufficient to remember the district of the silk workers in Bologna, and the names of the roads in the ancient parts of the town which recall the craft professionals clustering along the road (goldsmiths, carpenters...).

specialization within the caste may had represented a tool for contract enforcement. Freitas [2006] argues that caste, as an information-sharing institution, may had facilitated collective action as a way to enforce contracts, with beneficial effects on trade.

## 2.3 Data and Empirical Estimations

### 2.3.1 Data Sources

Our analysis relies on information from two main data sources: the *Ethnographic Atlas* the Standard Cross Cultural Sample (SCCS), two ethnicity-level datasets constructed by George Peter Murdock providing information on societies as early in time as information is available [Murdock, 1967]. The *Ethnographic Atlas* includes 1,267 ethnic groups, primarily from observations recorded in the late 19th and early 20th century. The SCCS, a sub-sample of the Ethnographic Atlas, is composed by 186 societies chosen to maximize independence in cultural and historical origin. The scope of these datasets is to "photograph" societies in their pristine state, before that contact with Europeans radically modified their technologies, economies and institutions. With this aim, the earliest documenting material available from missionaries, voyagers, merchants and anthropologists was analyzed and coded.
One limitation of these data is that societies are observed at different points in time[9]. However, it is reassuring to note that the pin-pointing time is much more homogeneous within the same region. Moreover, since the dataset provides precise information regarding the year each society was observed, we try to guarantee robustness of results in two additional ways: i) we control for the pin-pointing date; ii) we provide estimates of the baseline specifications for homogeneous sub-periods. To our knowledge, these are the two only datasets available providing information about the presence of castes in pre-modern societies.

**Castes Systems** SCCS variable $v272$ and Ethnographic Atlas Variable $v98$ track the presence of caste stratification in each society, specifying which type of endogamous stratification is observed, and namely whether stratification takes the form of despised occupational groups, ethnic stratification or a complex type of stratification[10]. In the

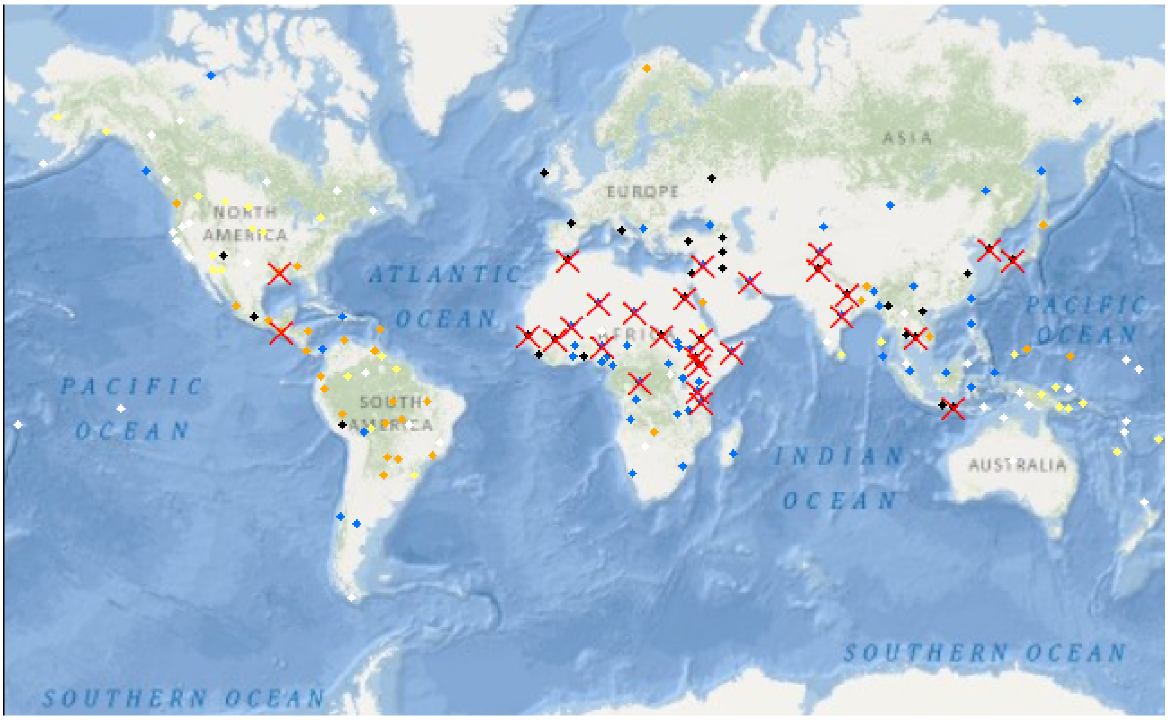
---

[9] In the Ethnographic Atlas, the pinpointing date for 3 societies dates before 1000 AD, for 42 societies between 1001 and 1800 AD, 32 after 1950 and all remaining between 1801 and 1950. For the SCCS, the pinpointing date for 3 of the 186 societies is before 1500 AD (Khmer, Hebrews and Romans), for 23 societies between 1500 and 1850, and the remaining societies' information is from late 19th and 20th century.

[10] Complex caste stratification include all multidimensional types of stratification, which can be at the time occupational, traditional, ethnic etc...

SCCS, out of 181 societies for which we have information, 18 societies had despised occupational groups, 3 societies presented ethnic endogamous stratification and 7 societies had a complex type of endogamous stratification. Endogamous stratification in the Ethnographic Atlas follows a similar pattern: of the 1079 societies of the Atlas for which we have information, about 10% had endogamy of despised occupational groups, 3% a complex type of caste stratification and a little more than 2% an ethnic endogamous stratification. Based on these variables, we created a dummy variable indicating whether a type of endogamous caste stratification is present ($CASTE_i$=1) or not ($CASTE_i = 0$) within the society[11]. As Figure 2.1 shows, castes have been mainly African and Euroasiatic institutions. The figure maps the distribution of societies in the SCCS. Each dot corresponds to a society. The color of each dot represents the level of technological sophistication of the society: white stands for no artisan, yellow stands for pottery only, green stands for weaving only, blue stands for metal-work and one other craft (either potters or weavers), black stands for societies with potters, weavers and metal-workers. The red cross around some of the societies in the sample indicates the presence of endogamous groups within the society.

FIGURE 2.1: Caste and Technology in the SCCS



**Technological Sophistication** In order to compare the degree of technological sophistication, we exploit two indexes aiming at capturing the level of sophistication in

---

[11] For robustness, we created a second dummy variable $CASTEnoethnic_i$ where we exclude ethnic stratification and replicated the analysis. Results remain qualitatively unchanged.

artisanal skills and the level of general technological sophistication.

The first index is intended to measure the degree of complexity and specialization in artisanal crafts and corresponds to SCCS variable $v153$. The index takes value one in societies where metalworking, loom-weaving and pottery are all absent while 5 is assigned to societies reported to have a variety of craft specialists, including at least smiths weavers and potters. Between these boundaries, the index is equal to 2 for societies where only pottery is practiced, 3 for societies which have specialized weavers (weavers only or weavers and potters) and 4 to societies which have metalworking but lack loom-weaving and/or pottery[12]. In the Ethnographic Atlas, a corresponding variable is missing but, by recombining information from other variables, it is possible to reconstruct a similar index of artisanal sophistication[13]. The measure of craft sophistication we are using might strike as a very raw and reductive indicator of technology. However, in order to make possible a comparison of technologies in pre-modern societies we need to pull back the frontier of technological sophistication. Indeed, it is worth emphasizing that less than 20% of all societies in the sample disposed of the three types of artisan practice.

The second index is meant to rank societies according to their overall level of technological sophistication in multiple fields, and namely: agriculture, artisanal crafts, writing and land transportation. The index is intended to mimic the technological index devised by Comin, Easterly, and Gong [2006], and was constructed aggregating information from several SCCS variables[14]. The All-Technologies index ranges from 4, the lowest value of technological sophistication, to 12, the highest rank, which is attributed to societies scoring the highest in all dimensions. Table 2.1 shows that the Circum-Mediterrean region and the EuroAsiatic region presented a higher level of both artisanal and general technological sophistication. Note, however, that in our analysis we will mainly focus on within-region variation of technological sophistication[15].

Thanks to the larger size of the Ethnographic Atlas, we are also able to explore the society level of labor specialization for various artisanal crafts. While previous indexes focus on the mere existence of a specific technological know-how, we are interested in

---

[12] The specific definition of crafts employed for devising SCCS variable $v153$ were defined as follows: i)Metal working: only such art as smelting, casting and forging, which involve the application of fire ; ii)Weaving: only the manufacture of true cloth on a lume or frame is indicated, not the manufacture of nets, baskets, mats or non-woven fabrics like barkloth or felt; iii)pottery: only the manufacture of earthenware utensils is indicated.

[13] See the Appendix for a detailed description of variables' construction and sources.

[14] The only technological area which is not covered by SCCS is military technology. See the Appendix for details regarding index construction.

[15] Results would not change if we were to lead the analysis by exploiting overall variation, and not just within region variation.

Table 2.1: Technological Sophistication

| | Afr | Med | Asia | Pac | N.Am | S.Am | Total |
|---|---|---|---|---|---|---|---|
| | | | | **Artisan Crafts** | | | |
| None | 3 | 1 | 4 | 14 | 13 | 4 | 39 |
| Pottery | 0 | 1 | 2 | 8 | 11 | 5 | 27 |
| Loom weaving | 1 | 2 | 4 | 2 | 5 | 17 | 31 |
| Metalwork | 21 | 8 | 15 | 5 | 2 | 5 | 56 |
| Smiths, weavers, potters | 3 | 16 | 9 | 2 | 2 | 1 | 33 |
| Total | 28 | 28 | 34 | 31 | 33 | 32 | 186 |
| | | | | **All Technologies** | | | |
| Index: 4-6 | 2 | 0 | 7 | 15 | 17 | 16 | 57 |
| Index: 7-9 | 24 | 4 | 11 | 12 | 15 | 11 | 77 |
| Index: 10-12 | 1 | 22 | 14 | 2 | 1 | 3 | 43 |
| Tolal | 27 | 26 | 32 | 29 | 33 | 30 | 177 |

*Notes*: The table reports the number of societies for each technological sophistication category. The upper panel summarizes the artisanal craft index, while the lower panel groups society into three level based on the All-Technology Index. Societies are grouped into six geographical region: "Afr." stands for Sub-Saharan Africa, "Med." for Circum-Mediterranean, "Asia" for East-Eurasia, "Pac." for Insular Pacific, "N.Am" for North America and "S.Am" for South America.

inquiring whether, given the presence of a specific technique, the occupation is exclusively practiced by specialized artisans or not. For several crafts, the dataset informs us on whether the occupation is performed as a craft specialization or not. We thus constructed a dichotomous variable $CRAFT_i$ which, given the existence of the technique, takes value 1 if the occupation is performed as a specialized craft and 0 if not. Table 2.2 has to be read as such: among the 435 societies which were known to work metal, only 17 did not have specialized metalworkers. On the other hand, among the 549 societies known to perform some kind of pottery, for the vast majority (475) pottery was not performed as a specialized craft[16].

### 2.3.2 Preliminary Evidence

We begin by estimating the relationship between the level of technological sophistication of the society and the presence of endogamous groups within the community. Given that the dependent variable $TECHN_i$ is ordinal, we proceed with two estimation strategies. First, we ignore the nature of the data and estimate an OLS model. The second strategy is to maintain the categorical feature of the variable and estimate an ordered logit model. As results remain qualitatively identical, we report results for the second strategy in the

---

[16] It is important to point that the Ethnographic Atlas is an unbalanced sample, i.e. some variables - regarding both the existence and the degree of specialization - are available for some societies and not for others. Therefore, descriptive statistics for each craft cannot be directly compared - we cannot say, for instance, that there are more societies which practice metalwork than weaving, because such a conclusion would heavily depend on the amount of missing data for each variable.

Table 2.2: Technological Sophistication

| | Afr | Med | Asia | Pac | N.Am | S.Am | Total |
|---|---|---|---|---|---|---|---|
| | | | | **Potters** | | | |
| Not Specialized | 178 | 32 | 40 | 38 | 113 | 74 | 475 |
| Specialized | 21 | 33 | 15 | 1 | 2 | 1 | 73 |
| Total | 199 | 65 | 55 | 39 | 115 | 75 | 548 |
| | | | | **Leather Workers** | | | |
| Not Specialized | 65 | 34 | 32 | 4 | 227 | 16 | 378 |
| Specialized | 16 | 41 | 9 | 0 | 1 | 1 | 68 |
| Total | 81 | 75 | 41 | 4 | 228 | 17 | 446 |
| | | | | **Weavers** | | | |
| Not Specialized | 62 | 36 | 52 | 30 | 88 | 54 | 322 |
| Specialized | 7 | 29 | 20 | 2 | 1 | 1 | 60 |
| Total | 69 | 65 | 72 | 32 | 89 | 55 | 382 |
| | | | | **Metal-Workers** | | | |
| Not Specialized | 8 | 1 | 0 | 0 | 2 | 6 | 17 |
| Specialized | 210 | 98 | 79 | 27 | 3 | 1 | 418 |
| Total | 218 | 99 | 79 | 27 | 5 | 7 | 435 |

*Notes*: The table reports the number of societies with specialized artisans. Societies are grouped into six geographical region: "Afr." stands for Sub-Saharan Africa, "Med." for Circum-Mediterranean, "Asia" for East-Eurasia, "Pac." for Insular Pacific, "N.Am" for North America and "S.Am" for South America.

Appendix. Our baseline estimating equation is:

$$TECHN_i = \beta_0 + \beta_1 CASTE_i + \beta_2 \mathbf{X_i} + \epsilon_\mathbf{i}$$

where $i$ indexes the society. The variable $TECHN_i$ denotes the level of technological sophistication. From Column (1) to Column (3), the dependent variable is the index of artisanal sophistication, while from Column (4) to Column (6) the dependent variable is the index of general technological sophistication. The dichotomous variable $CASTE_i$ takes value 1 if in the community is present at least one endogamous group, and 0 otherwise. From the second specification onwards, we add a vector of strictly exogenous control variables, **X**, which include geographical, climatic and location controls. Finally, we add region fixed effects in all specifications[17].

Regression results, reported in Table 2.3, highlight a positive and highly significant correlation between castes and the level of technological sophistication of the society, which changes modestly when including geographical controls, climatic controls and

---

[17] The geographical areas are: 1)Sub-Saharan Africa, 2) Circum-Mediterrenean (which includes Northern Africa and Western Europe), 3) East Eurasia (which include Eastern Europe and Asia), 4) Insular Pacific, 5) North America and 6) South America.

location controls. Note, from Column (1) and (4), that the variable $CASTES$, together with unobservables captured by the regional fixed effects, explains as much as around 40% of the variation in both technological sophistication in artisanal craft and general technological sophisticationWithout region fixed effects, the estimated coefficient of caste stratification on artisan craft is 1.57 (with a standard error of 0.170). The R-squared of the regression without regional fixed effects 0.17, implying that caste stratification alone explains up to 17% of the overall variation in craft sophistication. For the all-technology index, the coefficient is 2.55 (standard error 0.36) and part explained by caste stratification is up to 20%. The presence of castes is associated with a jump of almost one ladder in both indexes of sophistication, more precisely, a 0.8 increase in the craft index and an about 1 point increase in the technology index.

Column (1) of Table 2.4 shows how the estimated relation between castes and artisanal sophistication is almost half of the one estimated in the SCCS, however, reassuringly, it is highly statistically significant.
The larger size of the sample allows us to go a step further and explore patterns of labor specialization. In Column (2) to (6) we estimate a linear probability model for each main artisan craft. Having castes increases by around 20% the probability of having pottery and weaving performed by specialized artisans, while up to 40% the same probability for leather-working. We do not observe the same pattern for metal-working specialization and this is likely to be because metal-working tended to be practiced by specialists in the great majority of cases. However, if we were to run the same regressions on the probability of having metal-working practiced within the society, i.e. if we were interested in mere practice of metal work within the society, we would observe a strong and positive correlation.

**Robustness Checks** We performed a number of robustness checks. First, as previously mentioned, we replicated the baseline estimate with an ordered logit model. Since results are qualitatively unchanged we report estimations in Table B.4 of the Appendix. Second, we pursue two strategies to control for the possible bias generated by comparing societies pin-pointed at different points in time: we add the pin-pointing date as a control variable; alternatively, we reduce the sample keeping only societies observed respectively after 1850 and 1900. Again, results, reported in the Appendix, Table B.5, remain qualitatively unchanged, with estimated sizes of coefficients even larger in the smaller sub-samples than in the baseline.

## 2.4 Identifying a Causal Relationship

The regressions presented above document the existence of a previously unexplored strong correlation between castes and technological sophistication and specialization. This correlation cannot be interpreted causally, though. First, there might be some omitted factors which explain both technology and the presence of castes. For example, we might expect castes to prevail in historically more developed areas which are more technologically advanced to begin with. Second, the existence of a causal effect running from technological sophistication to castes cannot be excluded *a priori.* In fact, it might well be that castes emerged as a by-product of technological sophistication. Castes, just as class stratification, may be consequential to the degree of economic complexity which systematically follows technological innovations. Finally, we cannot exclude that early missionaries and voyagers, from whose diaries anthropologists draw information for coding the variables in the datasets, might had misunderstood some of the institutional features of societies they were visiting, not recognizing or misinterpreting endogamous stratifications[18]. If this is a true concern, the explanatory variable caste would be measured with error, exposing the OLS estimator to attenuation bias.

For untying this thigh knot we rely on three strategies. We first attempt to control for relevant observable society characteristics, mainly in order to exclude that the correlation is driven by well-known patterns of historical development. The second strategy is based on the introduction of an external source of variation in the explanatory variable - caste - which we claim exogenous to the main equation. With this aim, we exploit an early hypothesis put forward by several social scientists that has remained empirically untested, i.e. caste systems might have been an early evolutionary social device adopted by societies for reducing the transmission of diseases in areas facing a high burden of pathogens. Finally, with our last strategy, we try to exclude the existence of a direct effect running from technological sophistication to the presence of castes.

### 2.4.1 Controlling for Observables

The literature points at three main determinants of pre-modern development. Ashraf and Galor [2011a], inspired by the insightful perspective of Diamond and Ford [2000], document the effect of the Neolithic revolution - more precisely, of the timing of transition to agriculture - on comparative historical development, measured in terms of

---

[18] For instance, James Vaughan (1970) (in Essay included in the volume by Tuden, Plotnicov, and of Congress. Hebraic Section) [1970]) documented the presence of endogamous groups within the Margi population in Western Sudan, which was not coded as a society with endogamous stratification by the *Ethnographic Atlas*.

population density. Given the well-documented long lasting effects of early development [see Bockstette, Chanda, and Putterman, 2002, Comin, Easterly, and Gong, 2006, Olsson, Hibbs, et al., 2005, Putterman, 2008, among many others], the timing of transition to agriculture is very likely to affect technological sophistication and institutional features of societies in our sample and to act as a possible confounder. Therefore, we imputed the variable year since agricultural transition (from Putterman [2008][19]) to our data. Moreover, Ashraf and Galor [2011a] further show how the migratory distance from Adis Abeba is a strong predictor of historical development, measured again in terms of population density. The relationship they document is non-monotonic, for this reason we add a measure of Migratory Distance together with its quadratic form. Finally, as technologies can spread, following Ashraf and Galor [2011a] we add a measure of aerial distance to the regional technological frontier. Results, reported in Tabel 2.5 and 2.6, imply that none of these variables was driving the results, as coefficients even marginally increase after the introduction of the variables.

Another possible option is to control for other society characteristics, in order to compare societies at similar level of development. The SCCS offers plenty of potentially interesting proxy for development, i.e. a measure of population density, the level of political centralization of the society and an index meafsuring the intensity of agriculture. As regression results reported in B.6 of the Appendix show, including all these controls on top of Regressions in Tabel 2.5 and 2.6, reduces the magnitude of the caste coefficient by around 20% for crafts and by about 30% for all technologies. Again, the estimated coefficients remain significant at standard statistical levels. Another way to look at these results, it is to exploit the measurable bias from omitting observables characteristics for assessing the bias from unobservable ones, following Altonji, Elder, and Taber [2005]Detailed results and computations are reported in Table B.6, B.7 and B.8 of the Appendix.. By taking as baseline a regression of castes on technological sophistication with latitude as the only control estimated with region fixed effects, we find that, for unobsevables to be driving the results, the overall effects of unobservables on artisanal crafts should 29 times higher than the combined effects from all included controls, and namely all geographical and location controls, population density, agricultural intensity and political centralization, the Ln(Timing of Neolithic Transition), the distance to technological frontier and Migratory Distance and Migratory Distance squared. The same unobserved effect for general technological sophistication should be 3.3 times higher.

In any case, it is important to point that contemporaneous controls are likely to be endogenous to the level of technological sophistication, thus coefficient magnitudes should

---

[19] Note that the Putternam index is currently available only at country level and not disaggregated at the society level. This means that societies once localized in what is today the same country are imputed the same year for transition to agriculture. This is likely to attenuate the magnitude of the variable estimated coefficient.

be assessed with extreme care, as adding endogenous variable would bias the size of our coefficient of interest.

bose

### 2.4.2 Caste Systems and Disease Exposure

The impossibility to fully account for all possible confounders that may drive the relation between castes and technology, together with the necessity to exclude a reverse effect of technology on caste stratification, motivate us for pursuing a second alternative strategy. Relying on anthropological theories proposing castes as an evolutionary social device for reducing disease transmission, we exploit an index of disease exposure as an instrument for the presence of castes.

**Anthropological Background** The paternity of the idea that links castes to diseases should be accredited, in our knowledge, to historian William McNeill [1976]. In the well-known book "Plagues and Peoples", he argued:

> "the caste organization of Indian civilization may have partly been a response to the kind of epidemiological standoff that arose when intrusive Aryans, who had probably learned to live with some acute "civilized" diseases - e.g. perhaps smallpox - encountered various "forest folks" who had acquired tolerance for formidable local infections ... therefore " ... instead of digesting the various primitive communities that had occupied southern and eastern India... Indian civilization expanded incorporating ex-forest folks as castes".

More broadly speaking, the idea of the existence of a relationship between the disease environment and social norms, habits and practices that prevail in a given society is not new. Evolutionary anthropologist Alland [1970] claimed that "Social isolation of various subgroups is also important to disease ecology. A rigid caste system with minimal social and physical contact between groups may affect epidemic routes in community". Along the same line, medical anthropologist Cockburn [1971] was pointing that societies evolve "pragmatic means of sanitation that, whatever the intended religious or magical purpose, may in fact function to protect health and ward of diseases". More recently, evolutionary psychologists Curtis, Aunger, and Rabie [2004], Curtis, De Barra, and Aunger [2011] argued, very much in line with McNeill's ideas, that "The Law of Manu (one of the major Hindu holy books) prescribes... avoidance of polluted caste. These social distinctions may have their origins in biological avoidance strategies... through the emotional resources of disgust and contagion". Not referring precisely to castes but to ethnic fractionalization in general, Cashdan [2001] added that "parasites are like a wedge driving groups apart through their effective creation of anti-contagion behaviors". Finally, interesting complementary evidence is emerging from recent studies in genetics. Brahmajothi, Pitchappan, Kakkanaiah, Sashidhar, Rajaram, Ramu, Palanimurugan, Paramasivan, and Prabhakar [1991] and Siddiqui, Meisner, Tosh, Balakrishnan, Ghei, Fisher, Golding, Narayan, Sitaraman, Sengupta, et al. [2001] found that isolated caste and sub-caste populations of southern India differ in their HLA (Human leukocyte antigen) and other immune repertoire for diseases such as tuberculosis and leprosy. With this evidence in mind, some early historical descriptions of the treatment reserved to Indian low castes can be read under a new insightful perspective:

> "there were castes that had to carry bells just as medieval lepers... contact with them by sight or by the passage of a breath of air necessitated ceremonial purification[20]".

Given the previously mentioned similarities of Indian caste system and caste systems found in other geographical areas along dimensions such as the purity-pollution concept or the taboo regarding commensality, we test whether castes prevailed in environments burdened by higher disease exposure. Our hypothesis is that a highly diseased environment might had induced societies to create social separations in order to limit the spread of contagious diseases, which could otherwise had spread and thrived taking advantage of the large size of communities. The high burden of diseases might also had prevented different neighboring ethnic groups, adapted to different types of diseases, to merge by promoting norms of social distance. Even when the migrating groups had adapted to the new disease environment, the social norms imposing an endogamous stratification of society might had persisted[21].

**An Exogenous Measure of Disease Exposure** For testing whether caste systems prevailed in regions facing a heavier burden of diseases, we need a proxy for measuring

---

[20] Hutton [1946].

[21] The idea that diseases might affect long-term economic outcomes by shaping social institutions have recently received much attention by economists. Birchenall [2010], following anthropological evidence, find a significant and positive effect of disease exposure on the level of ethnic fractionalization in Sub-Saharan Africa. Birchenal claims that ethnic fractionalization slowed down the process of state centralization in Sub-Saharan Africa. Fogli and Veldkamp [2011], relying on recent findings in evolutionary psychology, argue that geographical areas dealing with higher disease hazards developed more collectivist cultures, with the implicit aim of promoting "safe" relations within the groups and limiting contacts with external potential source of contagion. Fogli and Veldkamp [2011] claim that the lack of individualism reduced the spread of innovations and was at the root of long run growth stagnation of countries with more collectivist values.

disease exposure. There are several methodological difficulties which complicate our search for a suitable instrument. First of all, the risk of endogeneity prevent us from using classical epidemiological measure of disease exposure, i.e. people at risk, number of new cases per year etc... The problem is that pathogens' effects are highly dependent on several social and economic variables related to the level of development, such as the quality of the diet, the sanitary habits and the type of household just to name a few. Second, several diseases, and namely those carried and transmitted by humans, tend to spread, it is therefore intrinsically difficult to obtain a measure of exogenous and time-invariant geographical disease exposure. Third, we would ideally need an index that captures disease exposure at least as early in time as societies were pin-pointed. Of course, we do not have information regarding the historical presence of diseases except for few diseases in a few number of countries.

In order to circumvent the methodological problems and in line with previous works [Cervellati, Sunde, and Valmori, 2011b, Murray and Schaller, 2010], we implement the following strategy: i) we rely on information retrieved from historical epidemiological Atlas from early 20th century (before the epidemiological revolution of the 40's); ii) among the diseases that we combine in the index, we rely mainly on vector diseases, i.e. diseases that are transmitted through a vector whose existence in a area strictly depends on exogenous climatic-environmental conditions; iii) we recode the index of intensity on a two point scale, in order to have it equal to 1 if the disease is present in the area where the society live, and 0 if not, for reducing the potential endogeneity of the variable disease intensity[22]. The SCCS provides an index of disease exposure for seven diseases, namely Leishmaniasis, Trypanosomes, Malaria, Schistosomes, Filariae, Spirochetes and Leprosy - note that 5 of the seven diseases are vector-transmitted[23].

It can be argued that the choice of employing vector diseases, despite guaranteeing exogeneity, might not serve the purpose of explaining the presence of castes, as vector diseases do not spread by human contact but are spread by an intermediate agent (mosquito, fly, etc...) which could even infect people living relatively far apart. However, there are two sets of reason why focusing mainly on vector-transmitted diseases can be an appropriate choice. First of all, epidemiological studies repeatedly shown that different diseases are "complementary shocks", in the sense that, for instance, having

---

[22] In fact, in a pre-modern context, human vector diseases such as tuberculosis or leprosy may prevail in highly populated societies, more than among hunters living spread in the forest. Therefore, the difference in disease intensity may be explained by the population density and not by exogenous climatic and environmental conditions.

[23] The index was made available by Low [1988], which collected the data for her study linking diseases to polygyny. In the SCCS, there is another index of disease exposure, made available by Ludvico Ludvico and Kurland [1995], who exploited the index for studies on human scarification practices.). This index is composed by 12 diseases. We exploited the second index for robustness checks: first, we constructed an alternative index for the seven main diseases of the baseline index (Leishmaniasis, Trypanosomes, Malaria, Schistosomes, Filariae, Spirochetes and Leprosy); second, we constructed an alternative index composed by all 12 diseases. Results remain qualitatively unchanged.

heavy malaria weakens the immunitary system and raise the risk of contracting other contagious diseases. Moreover, misconceptions regarding the way vector diseases are transmitted is vague, even today. For instance, interestingly for our analysis, Ramaiah, Kumar, and Ramu [1996] points that a share of the people interviewed in Tamil Nadu, India, (about 7%) still believed that filariasis can be contracted by the action of weaving.

Instrumental variable estimates are unbiased if the instrument, disease exposure, affects technological sophistication only through the presence of castes. In other words, disease exposure should not affect technological sophistication and artisanal specialization in pre-modern societies beyond the channel of caste stratification. In the economic literature the effects of diseases, through life expectancy, on economic performance of countries have been largely investigated. A large amount of contributions finds that increases in life expectancy cause increases in country income per capita [see Ashraf, Lester, and Weil, 2008, Bloom, Canning, and Sevilla, 2003, Gallup, Sachs, and Mellinger, 1998, Lorentzen, McMillan, and Wacziarg, 2005, Shastry and Weil, 2003, Weil, 2007, among others]. Acemoglu and Johnson [2006] reached an opposite conclusion, arguing that positive changes in mortality raise population growth but decrease income per capita. Cervellati and Sunde [2009] shows that the effects of higher life expectancy on income per capita tend to be negative before the demographic transition, but highly positive after the country transitions. Regarding the individual effects of health, diseases and high mortality rates have been found to negatively affect workers' productivity and days of work and to reduce investments in schooling [Ashraf, Fink, and Weil, 2010, Bleakley, 2006, Weil, 2007, just to mention a few].

On the contrary, the historical effects of diseases on development and, in particular, on technological sophistication have not been subject to empirical investigations. Acemoglu, Johnson, and Robinson [2001] argued that the impact of a diseased environment on the economic development of nations is not due to a direct effect of health conditions on income but passes through the channel of institutional quality settled at the time of colonization. On the other hand, historian McNeill [1976] argued that higher disease exposure historically represented a serious burden for societies because of the energy drain on individuals it engendered.

For our purposes, it is not obvious to see how lower productivity (days of work) and scarcer energies could had prevented societies from developing artisanal crafts like pottery, weaving and metalworking. A possible channel linking a diseased environment with lack of labor specialization runs through the potential population shocks which follows sudden epidemics and which could reduce demands and decrease specialization. However, the high level of technological sophistication and labor specialization historically reached in areas widely exposed to diseases, like the ancient civilizations of Southern

India, the Egyptians or the Ghana Empire, seems to contrast this possibility. Moreover, if a direct relation between geographical disease exposure and technology exists, according to the literature this should affect technological sophistication negatively. For this reason, since we expect disease to increase caste stratification and caste to foster technology, our estimate of the effect of castes on technological sophistication would be biased downward.

IV estimates In Table 2.7, we present instrumental variables estimates of the effect of castes on technological sophistication[24]. Since results do not change, for brevity, we report only the first sets of estimation results.

As shown in Column (1) and (2) of 2.7, the instrumental variable estimated effect of caste on artisan technological sophistication is about 5 time higher than the OLS one in the first and second specification. For what concern general technological sophistication, the IV coefficients is about 7 times larger. In specification (3) and (6), we propose a third specification where we include population density, this is in order to reassure that violations of the exclusion restrictions are not driving the results. However, since population density might be endogenous to the level of technological sophistication, the estimated coefficient cannot be directly compared with the baseline ones.
The wide difference in the size of OLS and IV estimated coefficients may result from the combination of several effects. The higher magnitude of the coefficients suggests that omitted variables might have weakened the estimated effects of castes on technological sophistication. Moreover, there might be attenuation biases caused by imperfections in the data source. On top of this, it is worth pointing that the OLS and IV estimators measure the effect of castes on technologies for two different sub-groups. OLS measures the effect on technologies for all societies which are recorded to have endogamous stratification. On the other hand, IV estimates the effect of castes on technological sophistication for those societies whose caste stratification, in our data, is related to the adaptive relationship with a highly diseased environment. We have no way to say whether all societies with endogamous stratification generated such institutions as a consequence of disease exposure.

It is crucial to note, as F Statistics Reported in Table 2.7 suggests, that we might be dealing with a weak instrument. Weak instruments generate problems both for what

---

[24] As we have a dichotomous endogenous variable, we proceed in two different ways. First, we estimate a linear first and second stage, which generates consistent second-stage estimates even with a dichotomous endogenous variable [Angrist and Pischke, 2008]. Another possible way is to estimate a non-linear first stage, and to instrument the endogenous variable with the fitted values of the non-linear first stage. The second procedure would enhance efficiency in the estimation as far as the non-linear model better approximates the CEF in the first stage [Angrist and Pischke, 2008].

concern the coefficient estimation, which risk to be biased in the same direction of the OLS' one, and the size of the confidence intervals for inference, that tend to be too narrow and lead to an underejection of the null[25]. In order to find support for our estimates, we undertake two strategies: i) we follow Moreira [2003] and compute confidence intervals which take into account the presence of a weak instrument, these are reported in square brackets in Table 2.7; iii) following suggestion by [Angrist and Pischke, 2008], we report reduced form estimates of the effect of diseases on technology, as the OLS estimator of the reduced form is always unbiased. Results are reported in the Appendix. Reassuringly, the 95% confidence regions based on Moreira's (2003) conditional likelihood ratio (CLR) approach are confined in the positive axis. As you can see from Column (1), (2), (4) and (5), the lower bound for the IV estimates is always higher than the OLS estimating coefficients, suggesting that OLS indeed bias the effect downwards. Moreover, the reduced form relation (reported in Table B.9 of the Appendix) between our disease index and technological sophistication confirm the existence a positive and significant effect of diseases on technology which, in our view, can possibly be attributed to no other channel other than the one of caste stratification.

To conclude, despite the fact that, given possible weak instrument issues, our IV coefficients may not be precisely estimated, weak-inference robust confidence intervals confirm a positive effect of caste stratification on technological sophistication.

### 2.4.3 Excluding Reverse Causality

Our main argument for excluding a reverse effect of technological sophistication over the likelihood of observing castes, it is to claim that castes are a pre-existing form of stratification. Therefore, our final attempt is to show that caste stratification is not the results of social complexity, which might derive from technological advancements. In order to do so, we exploit a well-established exogenous determinants of technological sophistication, i.e. the timing of transition to agriculture, and show how societies in areas that transitioned earlier to agriculture tend to have higher levels of technological sophistication, present a more structured division in social classes, but are not more likely to have caste stratification.

The Ethnographic Atlas and the SCCS report the presence and on the type of social classes reported within the society. The Atlas codes several type of social classes[26] but

---

[25] See Murray [2006a,b] for insightful reviews of best practices when dealing with weak instrument.

[26] Variables codes are: Absence of class stratification among freeman; Complex stratification into social classes correlated in large measure with extensive differentiation of occupational statuses; Dual stratification into a hereditary aristocracy and a lower class of ordinary commoners or freeman; Landlord versus landless class and Wealth distinction, based on possession and distribution

the coding offer no straightforward ordering. Our working hypothesis is that technological advances had generated new professional figures, and related new occupational positions, as well as major wealth distinctions between different professions. We create a variable $CLASS_i$ which takes value 1 for societies that present no sorts of class stratification (among freeman, as slavery is another matter); 2 for societies having a dual stratification into a hereditary aristocracy and a lower class of ordinary commoners or freeman, a landlord versus landless class; we assign value 3 to the index for societies with complex stratification into social classes and to societies which present wealth distinctions based on possession and distribution[27].

Results reported in Table 2.8 show, consistently with our hypothesis suggesting that castes are pre-existing forms of stratification, that the timing of Neolithic transition is positively correlated with both the technological sophistication index and the social classes index, however, there is no correlation with caste stratification.

## 2.5 Conclusion

We tested the hypothesis that caste systems positively affected technological sophistication and specialization at an early stage of development by “artificially” spurring division of labor in pre-modern societies. We used data from Murdock’s *Ethnographic Atlas* and Standard Cross Cultural Sample and found a robust and strong relationship between the presence of castes and two indexes of basic technological sophistication, the first index measuring artisan craft sophistication and the second general technological sophistication. Moreover, we found that castes increased the probability of having occupations performed by specialists. The correlations are robust to all major determinants of pre-modern development. In order to give a causal interpretation to the observed results, we exploit an exogenous variation in the presence of caste systems, i.e. the geographical disease exposure of the area in which the society is located. In fact, various historians and anthropologists have been suggesting that castes represented a social device for reducing contagion in highly diseased environments. We built an index of geographical disease exposure conceived to capture only the component of disease exposure related to exogenous bio-climatic characteristics of the geographical area. Exploiting disease exposure as an instrument for castes, we identified a positive and causal effect of caste systems on technological sophistication and labor specialization. Indeed, our IV estimates of the effect of caste on technology confirm the positive effect of caste systems

---

[27] Note that results do not rest on this coding. By considering wealth distinction as an intermediate type of stratification, we would obtain similar results

on technological sophistication and labor specialization in pre-modern societies.

Table 2.3: OLS for Castes and Technological Sophistication

| | **Technological Sophistication** | | | | | |
|---|---|---|---|---|---|---|
| | Artisanal Craft | | | All Technologies | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| CASTES | 0.788*** | 0.836*** | 0.873*** | 1.124** | 1.102** | 1.094** |
| | (0.241) | (0.263) | (0.270) | (0.481) | (0.469) | (0.486) |
| Latitude | | 0.013* | 0.012* | | 0.019* | 0.018* |
| | | (0.007) | (0.007) | | (0.010) | (0.010) |
| Mean Temperature | | 0.002 | 0.002 | | -0.005 | -0.006 |
| | | (0.010) | (0.010) | | (0.017) | (0.017) |
| Mean Rainfall | | 0.032 | 0.016 | | -0.098 | -0.096 |
| | | (0.044) | (0.047) | | (0.060) | (0.063) |
| Mean Elevation | | 0.000 | 0.000 | | 0.001** | 0.001** |
| | | (0.000) | (0.000) | | (0.000) | (0.000) |
| Agricultural Potential | | 0.010 | 0.015 | | 0.118** | 0.122*** |
| | | (0.029) | (0.029) | | (0.045) | (0.046) |
| Sea Distance | | | -0.406* | | | -0.122 |
| | | | (0.220) | | | (0.345) |
| River Distance | | | -0.215* | | | -0.139 |
| | | | (0.120) | | | (0.213) |
| Island Dummy | | | -0.343 | | | -0.488 |
| | | | (0.335) | | | (0.710) |
| Region FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 160 | 160 | 160 | 153 | 153 | 153 |
| R-squared | 0.419 | 0.440 | 0.458 | 0.408 | 0.464 | 0.469 |

*Notes*: The table reports OLS estimates. The unit of observation is the Standard Cross Cultural Sample society. The dependent variable in column (1), (2) and (3) is an index of artisanal sophistication which captures the type of specialized artisanal crafts performed in the society and ranges from 1 to 5. In column (4), (5) and (6), the dependent variable is an index of technological sophistication indicating the degree of sophistication of the society along multiple dimensions, i.e. crafts, agriculture, writing and land transportation. The explanatory variable of interest is the presence of castes, $CASTES$, which takes value 1 if at least one endogamous group is present and 0 otherwise. Geography control variables include: latitude, agricultural potential of the land occupied by the society, average temperature of the coldest month, average rainfall and mean elevation. Location controls are: distance of the society centroid from the closest river (in thousand Km), distance of the society centroid from the closest sea coast (in thousand Km) and a dummy variable for islands. Continent dummies for the six geographical areas in the sample, i.e. Sub-Saharan Africa, Circum-Mediterranean, East-Eurasia, Insular Pacific, North America and South America, are included in all specifications. Robust standard errors are reported in parentheses. *** stands for significant at the 1 percent level, ** at the 5 percent level, * at the 10 percent level.

Table 2.4: Castes and Technological Specialization

| | Technological Sophistication | Craft Specialization | | | |
|---|---|---|---|---|---|
| | | Pottery | Leather | Weaving | Metal |
| | (1) | (2) | (3) | (4) | (5) |
| CASTES | 0.420***<br>(0.146) | 0.220***<br>(0.034) | 0.449***<br>(0.058) | 0.186***<br>(0.068) | 0.020<br>(0.018) |
| Geography Controls | Yes | Yes | Yes | Yes | Yes |
| Location Controls | Yes | Yes | Yes | Yes | Yes |
| Continent FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 647 | 497 | 416 | 335 | 383 |
| R-squared | 0.511 | 0.301 | 0.504 | 0.254 | 0.390 |

*Notes*: The table reports OLS estimates. The unit of observation is the Ethnographic Atlas society. The dependent variable in column (1) is an index of artisanal sophistication which captures the types of artisanal crafts performed in the society and ranges from 1 to 5. In column (2), (3), (4) and (5), the dependent variable is a dichotomous variable indicating whether the existing craft is performed by specialized artisans or not. The explanatory variable of interest is the presence of castes, $CASTES$, which takes value 1 if at least one endogamous group is present and 0 otherwise. Geography control variables include: latitude, average agricultural suitability, average annual cumulated temperature above 0 degree, average precipitation and mean elevation, all measured within a 100 km radius of society centroid. Location controls are: distance of the society centroid from the closest river (in thousand Km), distance of the society centroid from the closest sea coast (in thousand Km) and a dummy variable for societies located on islands. Continent dummies for the six geographical areas in the sample, i.e. Sub-Saharan Africa, Circum-Mediterranean, East-Eurasia, Insular Pacific, North America and South America, are included in all specifications. Robust standard errors, clustered at the linguistic family level (the total number of linguistic family cluster is 68), are reported in parentheses. *** stands for significant at the 1 percent level, ** at the 5 percent level, * at the 10 percent level.

Table 2.5: Castes and Technological Sophistication

| | **Technological Sophistication** | | | | | |
|---|---|---|---|---|---|---|
| | Artisanal Craft | | | All Technologies | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| CASTES | 0.936***<br>(0.280) | 0.930***<br>(0.282) | 0.994***<br>(0.294) | 1.046**<br>(0.515) | 1.092**<br>(0.498) | 1.179**<br>(0.493) |
| Ln(Timing of Neolithic Trans.) | 0.497**<br>(0.196) | 0.177<br>(0.208) | 0.267<br>(0.226) | 0.950***<br>(0.326) | 0.343<br>(0.424) | 0.728*<br>(0.428) |
| Distance to Techn. Frontier | | -0.339***<br>(0.078) | -0.347***<br>(0.078) | | -0.545***<br>(0.152) | -0.619***<br>(0.159) |
| Migratory Distance | | | 0.119<br>(0.111) | | | 0.332**<br>(0.130) |
| Migratory Distance square | | | -0.004<br>(0.005) | | | -0.017***<br>(0.005) |
| Geography Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Location Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Continent FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 150 | 150 | 150 | 144 | 144 | 144 |
| R-squared | 0.447 | 0.498 | 0.503 | 0.475 | 0.528 | 0.551 |

*Notes*: The table reports OLS estimates. The unit of observation is the Standard Cross Cultural Sample society. The dependent variable in column (1), (2) and (3) is an index of artisanal sophistication which captures the type of specialized artisanal crafts performed in the society and ranges from 1 to 5. In column (4), (5) and (6), the dependent variable is an index of technological sophistication indicating the degree of sophistication of the society along multiple dimensions, i.e. crafts, agriculture, writing and land transportation. The explanatory variable of interest is the presence of castes, $CASTES$, which takes value 1 if at least one endogamous group is present and 0 otherwise. The variable Ln(Timing of Neolithic Transition) is the natural log of the number of years passed since the area where the society is located went through the transition to agriculture. Distance to technological frontier is the distance, in thousands of km, from the closest regional technological frontier. Migratory Distance and Migratory Distance squared are, respectively, the migratory distance (in thousand of km on a land path) and square migratory distance from Adis Abeba. Geography control variables include: latitude, agricultural potential of the land occupied by the society, average temperature of the coldest month, average rainfall and mean elevation. Location controls are: distance of the society centroid from the closest river (in thousand Km), distance of the society centroid from the closest sea coast (in thousand Km) and a dummy variable for islands. Continent dummies for the six geographical areas in the sample, i.e. Sub-Saharan Africa, Circum-Mediterranean, East-Eurasia, Insular Pacific, North America and South America, are included in all specifications. Robust standard errors are reported in parentheses. *** stands for significant at the 1 percent level, ** at the 5 percent level, * at the 10 percent level.

Table 2.6: Castes and Technology - Robustness

| | Technological Sophistication | Craft Specialization | | | |
|---|---|---|---|---|---|
| | | Pottery | Leather | Weaving | Metal |
| | (1) | (2) | (3) | (4) | (5) |
| CASTES | 0.629** | 0.200*** | 0.404*** | 0.168** | 0.023 |
| | (0.258) | (0.047) | (0.086) | (0.074) | (0.014) |
| Ln(Timing of Neolithic Trans.) | Yes | Yes | Yes | Yes | Yes |
| Distance to Techn. Frontier | Yes | Yes | Yes | Yes | Yes |
| Migratory Distance | Yes | Yes | Yes | Yes | Yes |
| Migratory Distance square | Yes | Yes | Yes | Yes | Yes |
| Geography Controls | Yes | Yes | Yes | Yes | Yes |
| Location Controls | Yes | Yes | Yes | Yes | Yes |
| Continent FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 643 | 493 | 413 | 332 | 379 |
| R-squared | 0.594 | 0.313 | 0.524 | 0.311 | 0.393 |

*Notes*: The table reports OLS estimates. The unit of observation is the Ethnographic Atlas society. The dependent variable in column (1) is an index of artisanal sophistication which captures the type of specialized artisanal crafts performed in the society and ranges from 1 to 5. In column (2), (3), (4) and (5), the dependent variable is a dichotomous variable indicating whether the existing craft is performed by specialized artisans or not. The explanatory variable of interest is the presence of castes, $CASTES$, which takes value 1 if at least one endogamous group is present and 0 otherwise. The variable Ln(Timing of Neolithic Transition) is the natural log of the number of years passed since the area where the society is located went through the transition to agriculture. Distance to technological frontier is the distance, in thousands of km, from the closest regional technological frontier. Migratory Distance and Migratory Distance squared are, respectively, the migratory distance (in thousand of km on a land path) and square migratory distance from Adis Abeba. Geography control variables include: latitude, average agricultural suitability, average annual cumulated temperature above 0 degree, average precipitation and mean elevation, all measured within a 100 km radius of society centroid. Location controls are: distance of the society centroid from the closest river (in thousand Km), distance of the society centroid from the closest sea coast (in thousand Km) and a dummy variable for society located on islands. Continent dummies for the six geographical areas in the sample, i.e. Sub-Saharan Africa, Circum-Mediterranean, East-Eurasia, Insular Pacific, North America and South America, are included in all specifications. Robust standard errors, clustered at the linguistic family level (the total number of linguistic family cluster is 68), are reported in parentheses. *** stands for significant at the 1 percent level, ** at the 5 percent level, * at the 10 percent level.

Table 2.7: IV for Castes and Technological Sophistication

| | **Technological Sophistication** | | | | | |
|---|---|---|---|---|---|---|
| | Artisanal Craft | | | All Technologies | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| CASTES | 4.036**<br>(1.752)<br>[ 1.47 , 15.88] | 4.162**<br>(1.836)<br>[1.50, 19.05] | 2.976*<br>(1.614)<br>[ 0.41, 24.13] | 6.880**<br>(3.035)<br>[ 2.65, 31.06] | 7.068**<br>(3.251)<br>[ 2.57, 40.56 ] | 4.850*<br>(2.644)<br>[ 0.91, 51.99] |
| Geography Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Location Controls | | Yes | Yes | | Yes | Yes |
| Population Density | | | Yes | | | Yes |
| Continent FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 159 | 159 | 159 | 152 | 152 | 152 |
| | **First Stage Statistics** | | | | | |
| Disease Index | 0.057**<br>(0.022) | 0.055**<br>(0.022) | 0.050**<br>(0.023) | 0.056**<br>(0.023) | 0.053**<br>(0.023) | 0.050**<br>(0.023) |
| Partial R-squared | 0.043 | 0.040 | 0.032 | 0.042 | 0.038 | 0.032 |
| F Stat | 6.64 | 6.02 | 4.86 | 6.17 | 5.51 | 4.61 |

*Notes*: IV estimates are reported. The unit of observation is the Standard Cross Cultural Sample society. In the upper panel we report the second stage estimated coefficients. The dependent variable in column (1), (2) and (3) is an index of artisanal sophistication which captures the type of specialized artisanal crafts performed in the society and ranges from 1 to 5. In column (4), (5) and (6), the dependent variable is an index of technological sophistication indicating the degree of sophistication of the society along multiple dimensions, i.e. crafts, agriculture, writing and land transportation. The explanatory variable of interest is the presence of castes, $CASTES$, which takes value 1 if in the society are present endogamous groups and 0 otherwise. The excluded instrument is an index of geographical disease exposure ranging from 0 to 7. Geography controls include: latitude, average temperature of the coldest month and average rainfall. Location controls are: distance of the society centroid from the closest river (in thousand Km), distance of the society centroid from the closest sea coast (in thousand Km) and a dummy variable for islands. Continent dummies for the six geographical areas in the sample, i.e. Sub-Saharan Africa, Circum-Mediterranean, East-Eurasia, Insular Pacific, North America and South America, are included in all specifications. Second stage coefficients are reported with robust standard errors in brackets, 95% confidence regions based on Moreira's (2003) conditional likelihood ratio (CLR) approach are reported in squared brackets. In the lower panel, the first stage coefficient estimates are reported. First stage statistics F statistics and Partial R-squared are also reported. *** stands for significant at the 1 percent level, ** at the 5 percent level, * at the 10 percent level.

Table 2.8: Neolithic Transition, Technology, Classes and Castes

| | **Dependent Variables** | | | | | |
|---|---|---|---|---|---|---|
| | Artisanal Craft | | Classes | | Castes | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Ln(Time of Neolithic Trans.) | 0.241*** | 0.221** | 0.179*** | 0.148*** | -0.011 | -0.009 |
| | (0.086) | (0.088) | (0.058) | (0.055) | (0.023) | (0.022) |
| Absolute Latitude | Yes | Yes | Yes | Yes | Yes | Yes |
| Geography Controls | | Yes | | Yes | | Yes |
| Location Controls | | Yes | | Yes | | Yes |
| Continent FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 684 | 684 | 1,003 | 1,003 | 997 | 997 |
| R-squared | 0.666 | 0.669 | 0.160 | 0.214 | 0.302 | 0.337 |

*Notes*: The table reports reduced form OLS estimates. The unit of observation is the Standard Cross Cultural Sample society. The dependent variable in column (1), (2) is an index of artisanal sophistication which captures the type of specialized artisanal crafts performed in the society and ranges from 1 to 5. In column (3) and (4) the dependent variable is an index capturing the presence of social classes within the society. In column (5) and (6) the dependent variable is a dummy variable that takes value 1 if in the society are present endogamous groups and 0 otherwise. Geography controls include: mean elevation and average rainfall. Location controls are: distance of the society centroid from the closest river (in thousand Km), distance of the society centroid from the closest sea coast (in thousand Km) and a dummy variable for islands. Absolute latitude and continent dummies for the six geographical areas in the sample, i.e. Sub-Saharan Africa, Circum-Mediterranean, East-Eurasia, Insular Pacific, North America and South America, are included in all specifications. Robust standard errors are reported in parentheses. *** stands for significant at the 1 percent level, ** at the 5 percent level, * at the 10 percent level.

# Chapter 3

# Malaria as Determinant of Modern Ethno-linguistic Diversity

## 3.1 Introduction

A vast and consolidated stream of economic literature has been documenting the economic consequences of ethno-linguistic diversity [among many others] [Alesina and Ferrara, 1999, 2004, Easterly and Levine, 1997, Fearon and Laitin, 2003]. More recent contributions, and in particular Michalopoulos [2008] and Ashraf and Galor [2013], make the point that ethno-linguistic diversity is in itself a consequence of more fundamental features of the geographical environment, showing that geo-physical characteristics, such as the heterogeneity in land suitability and the migratory distance from the cradle of humankind in East Africa, are associated with higher contemporary ethnic diversity today.

In this paper we document that an additional important geographical feature - geographic suitability to malaria - explains part of the currently observed ethno-linguistic diversity across and within modern countries. We hypothesize that in highly malarial areas the necessity to adapt and preserve immunities specific to the local disease environment increased isolation, reduced mobility and population admixture with neighboring groups, with the consequence of producing and preserving a larger number of different ethno-linguistic groups. From one hand, as we know from historical narratives, the danger associated with moving into diseased environments has been historically an obstacle to trade and commerce. Ramen [2002] mentions the practice of “silent trade” - i.e. trade at a distance between caravans from North Africa and population at the south of the

Sahara - devised to prevent the traders from exchanging dangerous germs on top of the goods. It is well known that malaria represented one of the most lethal diseases in history and, for this reason, it likely constituted a major obstacle to trade and, more generally, to population movements. As a matter or fact, the Boer trekkers faced high morbidity and mortality while adventuring into northern tropical areas, having in the end to resettle in more temperate regions [Becker, 1985]. Malaria represented a barrier between places where malaria was endemic and places where it was absent, but also between highly malarial areas and equally malarial neighboring areas. Diamond and Ford [2000] argued that "tropical Africans were combating malaria with more than just antibodies... by living in relatively small communities, spread out over vast areas...[to] limit the level of malaria transmission". On top of this, it is important to mention that malaria exists in various strains and that a whole set of location-specific strains were discovered, against which only strain-specific resistance can offer some form of protection. Curtin [1968, 1998] showed that African troops had higher mortality rates in foreign African countries than in their home-countries, possibly a consequence of different strains of malaria encountered in foreign African countries. Historian William McNeill [1976] argued that in the Indian sub-continent the heterogeneity in epidemiological endowments between intrusive Aryans and local "forest folks" prevented the local primitive communities from being "digested" into the invaders' civilization, remaining semi-autonomous and separated. Malaria, for instance, protected for centuries the isolation of Tharu people of Nepal [Brower and Johnston, 2007]. After centuries of residence in malaria-infested regions, Tharu people developed significant genetic resistance to malaria, to the point that they faced an about sevenfold lower malaria incidence than synpatric non-Tharu people [Modiano, Morpurgo, Terrenato, Novelletto, Di Rienzo, Colombo, Purpura, Mariani, Santachiara-Benerecetti, Brega, et al., 1991]. Malaria immunities allowed Tharu people to live undisturbed from neighboring powerful civilization, as they were the only group able to survive in those infested lands. Strict endogamy practiced by the Tharu confined these traits to this indigenous group and preserved this location-specific advantage.

In this work, we provide empirical evidence of the relationship linking malaria incidence with the historical and contemporaneous number of ethno-linguistic groups observed across and within countries. In order to do so we exploit the framework devised by Michalopoulos [2008] and conduct the analysis at a disaggregated level, by superimposing over modern countries a grid of artificial squares, measuring 1x1 degree of size. By doing so, we are able to concentrate on within-country variation of malaria incidence and diversity, as cross-country variation in diversity could hide key historical state-formation processes as unobservable confounders. One particularly challenging obstacle faced by the literature exploring the consequences of disease exposure lies in the search for exogenous measures of disease incidence, as it is well known that disease prevalence is highly influenced by living standards (diets, housing condition, public health, agricultural practices) and by patterns of population density and isolation. For these reasons, our attempt is to exploit only the exogenous component of historical malaria incidence, exploiting two reasonably exogenous disaggregated indexes. The first index we use is the malaria stability index devised by Kiszewski, Mellinger, Spielman, Malaney, Sachs, and Sachs [2004], which aims at measuring the force and stability of malaria transmission, based on the biological characteristics of the regionally dominant vector mosquitoes, such as their propensity to feed on humans and their daily survival, and how these features interact with the climatic environment. The second index is a recently available historical measure of malaria endemicity measured at the beginning of the twentieth century, produced by Lysenko [1968] and recently digitalized by [Hay S.I., 2004]. In our view, the index has to be preferred to modern measures of clinical malaria incidence, insofar as it measures the degree of malaria endemicity in a time where the main massive malaria eradication campaigns had still to be conceived and realized.

We depart from previous work exploring the relationship between disease and diversity along several dimensions[1]. First, we concentrate on the role played by malaria alone, based on the fact that malaria has likely been the most destructive disease in human history. Malaria is known as the “strongest known selective pressure in the recent history of the human genome” Kwiatkowski [2005], and just as much as it affected the evolution of the human genome, we argue it molded the social structure and habits of societies residing in the affected areas. We do not exclude that malaria acted in combination with other relevant human diseases, however, we maintain that malaria was the major driving force, even if its effects may have been worsened by the interaction with other harmful co-existing diseases. Secondly, we highlight that the relation between malaria incidence and diversity exists at different points in history, today and before colonization. Exploiting the fact that malaria was not present in the New World before colonization, we show that the relationship between geographic suitability to malaria and ethno-linguistic diversity is absent in pre-colonial Americas. This exercise, which we perform as a sort of “placebo test”, mitigates the concern that our index of geographic suitability to malaria is capturing the effect on diversity of climatic and geographic characteristics spuriously related with malaria suitability. In order to test this, we exploit Murdock (1959) data on the distribution of ethnic groups in pre-colonial North and South America, recently digitized by Chiovelli (2014).

Finally, the main contribution of our exercise to the literature exploring the fundamental determinants of ethno-linguistic diversity is to propose and explicitly test one of the

---

[1] A previous empirical work by [Cashdan, 2001] highlighted a correlation between ethnic diversity and pathogens load, she concentrated on six diseases: leishmanias (three species), trypanosomes (two species), malaria (four species), schistosomes (three species), filariae (two species), spirochetes (two species and one genus), and leprosy. Birchenall [2010] exploits the same of set of diseases and a similar framework to investigate the long term effect of diseases on development

possible channels that links malaria and ethno-linguistic diversity. Throughout history, patterns of marriage have been one of the most frequent form of migration and mixing with neighboring communities. In this paper, we hypothesize that in highly malarial environment the need to exploit and preserve location-specific immunities reduced the practice of marrying away from home and from former kinship ties, thus increasing the propensity to arrange endogamous marriages. We test this hypothesis using contemporary data on marriage patterns in Africa, from the Demographic and Health Survey. We exploit several waves of the survey conducted in 1815 cluster across 22 African countries, whenever information on both the ethic identity of the wife and the husband was available. Since information on ethnicity varies along waves and countries - as for some waves very detailed information on ethnic identity is provided whereas for other waves more aggregated ethnic families are reported - we associate all reported ethnic identities to the various branches of the Ethnologue tree. In this way, we are able to compute the propensity to marry within the ethnic family at various level of ethnic family aggregation. We show that, at intermediate levels of ethnic disaggregation, individuals living in areas with a higher suitability to malaria have higher propensity to marry somebody from the same group. Results hold when controlling for individual level of education, age and whether the individual is a urban or rural dweller.

## 3.2 Malaria Exposure and Ethno-Linguistic Diversity

### 3.2.1 Data and Empirical Strategy

**Empirical Strategy** We follow the framework devised by Michalopoulos [2008] and create a grid of cells of 1x1 degree of size, corresponding to about 110 km at the equator. By doing so, we create a set of squared "artificial countries" which we exploit as unit of observation. This approach has two main advantages: from one hand, the small size of cells allow us to measure more accurately geographical and environmental features, variables such as elevation, average precipitation and temperature are more accurately measured this way than in the aggregated form of a standard cross-country regressions; ii) from another hand, by looking at small artificial cells we can address an additional source of variability, i.e. within country variability. In fact, following this framework, we can easily account for common country effects and get rid the potential confounding factors which act at the country level, such as institutional or cultural features. This approach has also a set of shortcomings, the most compelling one being that many variables for which we would like to control for, such as GDP or average human capital, are not available at such a finer spatial resolution. However, it is also true that all these controls would be endogenous to our variable of interests and could not be included into the regression without fears of biased estimates.

We estimate the relationship between the natural logarithm of the number of linguistic groups in the 1x1 degree cell and the incidence of malaria. Our baseline specification follows:

$$Ln(NumberofGroups)_{i,c} = \beta_0 + \beta_1 Malaria_{i,c} + \beta_2 \mathbf{X_{i,c}} + \mu_{\mathbf{c}} + \epsilon_{\mathbf{i,c}}$$

where $i$ indicates the cell, and $c$ the country. $\mathbf{X_{i,c}}$ includes a vast set of climatic, geographical and location controls, $\mu_c$ stands for country fixed effects. Based on our conceptual framework, we hypothesize further that the effect of malaria on linguistic diversity could be higher in places with a high heterogeneity in the stock of acquired and innate immunities. We expect this heterogeneity to be higher in places with a higher variation in elevation since, because of its reliance on mosquitoes for transmission, malaria cannot be transmitted in highlands even whenever present in the neighboring lowlands. Because of this precise reason, we might find neighboring areas where people living in the lowlands developed strong immunities to malaria and people in the highland did not. This heterogeneity in immunity endowment could in principle represent a barrier to the admixing of these groups. We test this prediction by looking at the interaction between malaria incidence with a measure of the standard deviation of the elevation in the cell $Mal_{i,c} * STDElev_{i,c}$:

$$Ln(NofGroups)_{i,c} = \beta_0 + \beta_1 Mal_{i,c} + \beta_2 STDElev_{i,c} + \beta_3 Mal_{i,c} * STDElev_{i,c} + \beta_2 \mathbf{X_{i,c}} + \mu_{\mathbf{c}} + \epsilon_{\mathbf{i,c}}$$

**The Data** The World Language Mapping System database offers, to our knowledge, the most comprehensive mapping of the world's known living languages[2]. We compute the number of languages spoken in all cells of our dataset, excluding languages spoken by less than 10000 people overall. Figure 1 shows how the number of languages per cell varies across the world. Note that in about half of the cells composing our sample only one language is spoken, while in one fourth of the cell 3 or more languages are currently in use.

In order to measure malaria exposure, we rely on the work of Kiszewski, Mellinger, Spielman, Malaney, Sachs, and Sachs [2004], which constructed their index as follows: they associated to each country a dominant vector of *Anopheles* mosquitoes (for countries with different dominant mosquitoes, mosquitoes were association to sub-regions). A monthly index of stability was then computed as a parametric function of the share

---

[2] Note that we exploit a definition of ethnicity based on language, this choice follows both theoretical and empirical consideration.

of blood meal taken by the mosquito, the daily survival rate and the extrinsic incubation period. Once this regional monthly index (constructed for about 260 regions in the world) was created, in order to obtain a finer data resolution, a mimimum lagged threshold of precipitation (10 mm) was exploited as a pre-condition for malaria transmission. The yearly aggregation of such an index is the malaria stability index which we exploit in our analysis.

The second measure of malaria incidence that we use is a historical one. It was produced by Lysenko [1968] and recently digitized by [Hay S.I., 2004], it aims at measuring the level of malaria endemicity at the beginning of the 20th century. Endemicity is defined as the parasite rate (PR) in the 2-10 year age cohort [3]. The index takes value 0 wherever malaria is absent, 1 for epidemic areas, 2 where malaria is hypoendemic, 3 for mesoendemic areas, 4 for hyperendemic and 5 for holoendemic areas. As previously mentioned, the level of malaria endemicity could be endogenous to several factors such as agricultural activity and population density, however, insofar as it is measured at the begininning of the 20th it precedes the timing of the massive malaria eradication campaings which took place after the IIWW. It is reasonable to consider than before the era of massive eradication campaing, the incidence of malaria was more tightly related to climatic factors than it is today.

As a robustness, we exploit what we consider three additional proxies of historical malaria incidence: three blood related genetic variants which have been shown to be strictly associated with historical malaria incidence. HbS stands for sickle haemoglobin allele frequency in 2010, G6PD for allele frequency for G6PD deficiency in 2010, and Duffy for Duffy negative phenotype again measured in 2010.

---

[3] Hypoendemic with PR lower than 0.1; mesoendemic with PR between 0.11-0.5; hyperendemic for 0.51-0.75 for the holoendemic class (PR higher than 0.75), the PR refers to the 1-year age group

### 3.2.2 Results

Table 3.1 reports baseline results. An increase in malaria incidence is associated with a sizable and robust increase in the number of languages spoken in a cell, see Column 1-5 in Table 3.1. In terms of magnitude, going from a cell with no malaria to a cell which was historically holoendemic increases the average number of language spoken in the cell by almost one fourth. We obtain a similar estimate when we look at the results of the regression with the malaria ecology index of stability in transmission, in this case going from a place with 0 malaria stability to a place with an average stability of 34, increases the number of groups by 0.23%, see Column 6-10 in Table 3.1. Results are confirmed when looking at the effect of genetic immunities, Table 3.2 and when looking at the number of pre-colonial ethnic group as reported by Murdock (1959), see Table 3.4. Consistent with our predictions, the correlation between the indexes of malaria and the number of ethno-linguistc groups is not present when looking at the distribution of ethnic groups of pre-colonial Americas, since the disease was not present in the New World before colonization. Finally, there is an extra effect of malaria incidence on ethnic diversity in areas with a larger than average heterogeneity in elevation.

Table 3.1: Malaria Incidence and Diversity

| | **Ln(Number of Languages)** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Malaria Endemicity** | 0.200*** | 0.162*** | 0.090*** | 0.052*** | 0.051*** | | | | | |
| | (0.02) | (0.02) | (0.02) | (0.01) | (0.01) | | | | | |
| **Malaria Ecology** | | | | | | 0.045*** | 0.039*** | 0.023*** | 0.007* | 0.008** |
| | | | | | | (0.01) | (0.01) | (0.01) | (0.00) | (0.00) |
| Average Temperature | | 0.006* | 0.004 | -0.006 | -0.007 | | 0.006 | 0.007 | -0.004 | -0.005 |
| | | (0.00) | (0.01) | (0.01) | (0.01) | | (0.00) | (0.01) | (0.01) | (0.01) |
| Average Precipitation | | 0.002*** | 0.004*** | 0.002*** | 0.002*** | | 0.002*** | 0.004*** | 0.002*** | 0.002*** |
| | | (0.00) | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) | (0.00) |
| Mean Suitability | | -0.074 | 0.027 | 0.083 | 0.067 | | 0.098 | 0.120 | 0.110 | 0.086 |
| | | (0.08) | (0.08) | (0.08) | (0.08) | | (0.09) | (0.08) | (0.07) | (0.08) |
| Variation Suitability | | 0.784** | 0.433* | 0.651*** | 0.626*** | | 0.611* | 0.315 | 0.636*** | 0.600** |
| | | (0.33) | (0.23) | (0.23) | (0.22) | | (0.33) | (0.22) | (0.24) | (0.23) |
| Mean Elevation | | 0.000*** | 0.000 | -0.000 | -0.000 | | 0.000** | 0.000 | -0.000 | -0.000 |
| | | (0.00) | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) | (0.00) |
| Variation Elevation | | 0.000*** | 0.000*** | 0.000*** | 0.000*** | | 0.000*** | 0.001*** | 0.000*** | 0.000*** |
| | | (0.00) | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) | (0.00) |
| Ruggedness | | 0.000 | 0.000 | 0.000 | 0.000 | | 0.000 | 0.000 | 0.000 | 0.000 |
| | | (0.00) | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) | (0.00) |
| Total Water | | | 0.026* | 0.007 | 0.006 | | | 0.033** | 0.007 | 0.005 |
| | | | (0.01) | (0.01) | (0.01) | | | (0.01) | (0.01) | (0.01) |
| Total Area | | | 0.014** | 0.023*** | 0.023*** | | | 0.015** | 0.024*** | 0.023*** |
| | | | (0.01) | (0.00) | (0.00) | | | (0.01) | (0.00) | (0.00) |
| Number of Country | | | 0.200*** | 0.210*** | 0.209*** | | | 0.196*** | 0.211*** | 0.210*** |
| | | | (0.04) | (0.03) | (0.03) | | | (0.04) | (0.03) | (0.03) |
| Within Country | | | -0.021 | -0.003 | -0.003 | | | -0.038 | -0.002 | -0.004 |
| | | | (0.06) | (0.05) | (0.05) | | | (0.06) | (0.05) | (0.05) |
| Ln(Migratory Distance) | | | -0.153*** | -0.191 | -0.193 | | | -0.146*** | -0.171 | -0.174 |
| | | | (0.06) | (0.12) | (0.12) | | | (0.05) | (0.13) | (0.13) |
| Ln(Distance Coast) | | | 0.087*** | 0.039** | 0.041** | | | 0.085*** | 0.039** | 0.043** |
| | | | (0.02) | (0.02) | (0.02) | | | (0.02) | (0.02) | (0.02) |
| Ln(Distance Border) | | | -0.069*** | -0.037** | -0.037** | | | -0.064*** | -0.038** | -0.039** |
| | | | (0.02) | (0.02) | (0.02) | | | (0.02) | (0.01) | (0.02) |
| Ln(Distance River) | | | 0.017 | -0.009 | -0.009 | | | 0.012 | -0.009 | -0.009 |
| | | | (0.02) | (0.01) | (0.01) | | | (0.02) | (0.01) | (0.01) |
| Ln(Distance Capital) | | | 0.017 | 0.042* | 0.044* | | | 0.017 | 0.038* | 0.042* |
| | | | (0.02) | (0.02) | (0.02) | | | (0.02) | (0.02) | (0.02) |
| Absolute Latitude | | | 0.001 | -0.002 | -0.002 | | | 0.002 | -0.004 | -0.004 |
| | | | (0.01) | (0.01) | (0.01) | | | (0.01) | (0.01) | (0.01) |
| Night Lights | | | | | -0.004 | | | | | -0.005 |
| | | | | | (0.00) | | | | | (0.00) |
| Ln(Population Density) | | | | | 0.011 | | | | | 0.015* |
| | | | | | (0.01) | | | | | (0.01) |
| Country FE | | | | Yes | Yes | | | | Yes | Yes |
| N | 9450 | 9450 | 9450 | 9450 | 9450 | 9450 | 9450 | 9450 | 9450 | 9450 |
| R-Squared | 0.187 | 0.289 | 0.391 | 0.549 | 0.549 | 0.163 | 0.295 | 0.393 | 0.546 | 0.547 |

*Notes* Table reports OLS estimates. The unit of observation is a grid square of 1x1 degree of size. Standard Errors (in parenthesis) are clustered at the country level. The dependent variable is the natural logarithm of the number of languages per cell, constructed based on the World Language Mapping System database. Explanatory variables measure the average values (and the standard deviation, in the case of elevation and suitability) of the variable within the 1x1 degree cell. See the Appendix for further details on data construction. ***, **, * indicate significance at 1-, 5-, and 10-% level, respectively.

Table 3.2: Malaria Immunities and Diversity

| | Ln(Number of Languages) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Hbs** | 7.785*** | -0.254 | -0.292 | | | | | | |
| | (1.26) | (0.99) | (0.99) | | | | | | |
| **G6PD** | | | | 3.750*** | 1.675** | 1.633** | | | |
| | | | | (1.10) | (0.77) | (0.77) | | | |
| **Duffy** | | | | | | | 0.484*** | 0.618*** | 0.590*** |
| | | | | | | | (0.13) | (0.18) | (0.19) |
| Average Temperature | | -0.006 | -0.007 | | -0.007 | -0.008 | | 0.000 | -0.001 |
| | | (0.01) | (0.01) | | (0.01) | (0.01) | | (0.01) | (0.01) |
| Average Precipitation | | 0.003*** | 0.003*** | | 0.003*** | 0.003*** | | 0.003*** | 0.003*** |
| | | (0.00) | (0.00) | | (0.00) | (0.00) | | (0.00) | (0.00) |
| Mean Suitability | | 0.109 | 0.087 | | 0.178* | 0.137 | | 0.148 | 0.116 |
| | | (0.07) | (0.08) | | (0.10) | (0.12) | | (0.11) | (0.12) |
| Suitability | | 0.665*** | 0.633*** | | 0.924*** | 0.875*** | | 0.926*** | 0.888*** |
| | | (0.24) | (0.23) | | (0.31) | (0.31) | | (0.34) | (0.33) |
| Mean Elevation | | -0.000 | -0.000 | | -0.000 | -0.000* | | -0.000 | -0.000 |
| | | (0.00) | (0.00) | | (0.00) | (0.00) | | (0.00) | (0.00) |
| Variation Elevation | | 0.000*** | 0.000*** | | 0.000*** | 0.000*** | | 0.000*** | 0.000*** |
| | | (0.00) | (0.00) | | (0.00) | (0.00) | | (0.00) | (0.00) |
| Ruggedness | | 0.000 | 0.000 | | 0.000*** | 0.000*** | | 0.000*** | 0.000*** |
| | | (0.00) | (0.00) | | (0.00) | (0.00) | | (0.00) | (0.00) |
| Total Water | | 0.004 | 0.003 | | 0.003 | 0.002 | | 0.004 | 0.003 |
| | | (0.01) | (0.01) | | (0.02) | (0.02) | | (0.02) | (0.02) |
| Total Area | | 0.025*** | 0.024*** | | 0.025*** | 0.023*** | | 0.028*** | 0.026*** |
| | | (0.00) | (0.00) | | (0.01) | (0.01) | | (0.00) | (0.01) |
| Number of Country | | 0.219*** | 0.217*** | | 0.166*** | 0.161*** | | 0.165*** | 0.162*** |
| | | (0.03) | (0.03) | | (0.04) | (0.04) | | (0.04) | (0.04) |
| Within Country | | 0.007 | 0.007 | | 0.025 | 0.019 | | 0.017 | 0.013 |
| | | (0.05) | (0.05) | | (0.05) | (0.06) | | (0.05) | (0.05) |
| Ln(Migratory Distance) | | -0.184 | -0.187 | | -0.200 | -0.200 | | -0.127 | -0.126 |
| | | (0.12) | (0.12) | | (0.14) | (0.14) | | (0.15) | (0.15) |
| Ln(Distance Coast) | | 0.041** | 0.045** | | 0.069*** | 0.077*** | | 0.067*** | 0.074*** |
| | | (0.02) | (0.02) | | (0.02) | (0.02) | | (0.02) | (0.02) |
| Ln(Distance Border) | | -0.039** | -0.040** | | -0.050*** | -0.051*** | | -0.048*** | -0.049** |
| | | (0.02) | (0.02) | | (0.02) | (0.02) | | (0.02) | (0.02) |
| Ln(Distance River) | | -0.010 | -0.010 | | -0.023 | -0.022 | | -0.026 | -0.025 |
| | | (0.01) | (0.01) | | (0.02) | (0.02) | | (0.02) | (0.02) |
| Ln(Distance Adis Ababa) | | 0.039* | 0.042* | | 0.062* | 0.073** | | 0.061** | 0.070** |
| | | (0.02) | (0.02) | | (0.03) | (0.03) | | (0.03) | (0.03) |
| Absolute Latitude | | -0.005 | -0.006 | | -0.004 | -0.004 | | -0.001 | -0.001 |
| | | (0.01) | (0.01) | | (0.01) | (0.01) | | (0.01) | (0.01) |
| Night Lights | | | -0.005 | | | 0.001 | | | -0.001 |
| | | | (0.00) | | | (0.01) | | | (0.01) |
| Ln(Population) | | | 0.014 | | | 0.019 | | | 0.016 |
| | | | (0.01) | | | (0.01) | | | (0.01) |
| Country FE | | Yes | Yes | | Yes | Yes | | Yes | Yes |
| N | 9368 | 9368 | 9368 | 5599 | 5599 | 5599 | 5643 | 5643 | 5643 |
| R-Squared | 0.093 | 0.546 | 0.546 | 0.071 | 0.532 | 0.533 | 0.068 | 0.530 | 0.531 |

*Notes* Table reports OLS estimates. The unit of observation is a grid square of 1x1 degree of size. Standard Errors (in parenthesis) are clustered at the country level. The dependent variable is the natural logarithm of the number of languages per cell, constructed based on the World Language Mapping System database. Explanatory variables of interests are measures of genetic resistance to malaria. HbS stands for sickle haemoglobin allele frequency in 2010, G6PD for allele frequency for G6PD deficiency in 2010, and Duffy for Duffy negative phenotype in 2010. ***, **, * indicate significance at 1-, 5-, and 10-% level, respectively.

Table 3.3: Malaria Incidence, Variation in Elevation and Diversity

| | **Ln(Number of Languages)** | | | |
|---|---|---|---|---|
| **Malaria Endemicity** | 0.016 | | 0.017 | |
| | (0.01) | | (0.01) | |
| **Malaria Endemicity x Variation Elevation** | 0.000*** | | 0.000*** | |
| | (0.00) | | (0.00) | |
| **Malaria Ecology** | | 0.004 | | 0.005 |
| | | (0.00) | | (0.00) |
| **Malaria Ecology x Variation Elevation** | | 0.000** | | 0.000* |
| | | (0.00) | | (0.00) |
| **Variation Elevation** | 0.000 | 0.000*** | 0.000 | 0.000*** |
| | (0.00) | (0.00) | (0.00) | (0.00) |
| Average Temperature | -0.006 | -0.006 | -0.006 | -0.006 |
| | (0.01) | (0.01) | (0.01) | (0.01) |
| Average Precipitation | 0.002*** | 0.002*** | 0.002*** | 0.002*** |
| | (0.00) | (0.00) | (0.00) | (0.00) |
| Mean Suitability | 0.051 | 0.077 | 0.051 | 0.078 |
| | (0.08) | (0.08) | (0.08) | (0.08) |
| Variation Suitability | 0.667*** | 0.607*** | 0.667*** | 0.602*** |
| | (0.21) | (0.23) | (0.21) | (0.23) |
| Mean Elevation | -0.000 | -0.000 | -0.000 | -0.000 |
| | (0.00) | (0.00) | (0.00) | (0.00) |
| Total Water | 0.004 | 0.007 | 0.002 | 0.006 |
| | (0.01) | (0.01) | (0.01) | (0.01) |
| Total Area | 0.022*** | 0.023*** | 0.022*** | 0.023*** |
| | (0.00) | (0.00) | (0.00) | (0.00) |
| Number of Country | 0.208*** | 0.210*** | 0.209*** | 0.210*** |
| | (0.03) | (0.03) | (0.03) | (0.03) |
| Within Country | -0.007 | -0.000 | -0.005 | 0.001 |
| | (0.05) | (0.05) | (0.05) | (0.05) |
| Ruggedness | 0.000 | 0.000 | 0.000 | 0.000 |
| | (0.00) | (0.00) | (0.00) | (0.00) |
| Ln(Migratory Distance) | -0.215* | -0.180 | -0.213* | -0.179 |
| | (0.12) | (0.13) | (0.12) | (0.13) |
| Ln(Distance Coast) | 0.039** | 0.043** | 0.038** | 0.043** |
| | (0.02) | (0.02) | (0.02) | (0.02) |
| Ln(Distance Border) | -0.037** | -0.040** | -0.036** | -0.039** |
| | (0.02) | (0.02) | (0.02) | (0.02) |
| Ln(Distance River) | -0.010 | -0.009 | -0.010 | -0.009 |
| | (0.01) | (0.01) | (0.01) | (0.01) |
| Ln(Distance Capital) | 0.046** | 0.041* | 0.047** | 0.041* |
| | (0.02) | (0.02) | (0.02) | (0.02) |
| Absolute Latitude | -0.003 | -0.004 | -0.003 | -0.004 |
| | (0.01) | (0.01) | (0.01) | (0.01) |
| Night Lights | -0.004 | -0.005 | -0.004 | -0.005 |
| | (0.00) | (0.00) | (0.00) | (0.00) |
| Ln(Population) | 0.011 | 0.016* | 0.011 | 0.016* |
| | (0.01) | (0.01) | (0.01) | (0.01) |
| Country FE | Yes | Yes | Yes | Yes |
| N | 9450 | 9450 | 9450 | 9450 |
| R-Squared | 0.552 | 0.548 | 0.553 | 0.548 |

*Notes* Table reports OLS estimates. The unit of observation is a grid square of 1x1 degree of size. Standard Errors (in parenthesis) are clustered at the country level. The dependent variable is the natural logarithm of the number of languages per cell, constructed based on the World Language Mapping System database. The variables Malaria Endemicity x Variation Elevation, Malaria Ecology x Variation Elevation are interaction terms. ***, **, * indicate significance at 1-, 5-, and 10-% level, respectively.

TABLE 3.4: Malaria and Diversity in Pre-Colonial Africa

| | **Ln(Number of Languages: Pre-Colonial Africa)** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Malaria Endemicity | 0.115*** | 0.061*** | 0.043** | 0.055** | 0.049** | | | | | |
| | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | | | | | |
| Malaria Ecology | | | | | | 0.017*** | 0.017*** | 0.010** | 0.006** | 0.006* |
| | | | | | | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| Average Temperature | | 0.023** | 0.002 | 0.002 | 0.001 | | 0.002 | -0.008 | 0.002 | 0.000 |
| | | (0.01) | (0.01) | (0.02) | (0.02) | | (0.01) | (0.02) | (0.02) | (0.02) |
| Average Precipitation | | 0.001 | 0.001 | 0.001 | 0.001 | | 0.001* | 0.001 | 0.001 | 0.001 |
| | | (0.00) | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) | (0.00) |
| Mean Suitability | | 0.213** | 0.223** | 0.019 | 0.013 | | 0.144 | 0.198* | 0.042 | 0.031 |
| | | (0.09) | (0.10) | (0.13) | (0.13) | | (0.11) | (0.11) | (0.13) | (0.13) |
| Variation Suitability | | 1.171*** | 1.062*** | 1.032*** | 0.970*** | | 0.892*** | 0.916*** | 0.978*** | 0.909*** |
| | | (0.39) | (0.34) | (0.28) | (0.28) | | (0.33) | (0.31) | (0.28) | (0.27) |
| Mean Elevation | | 0.000* | -0.000* | -0.000 | -0.000 | | 0.000** | -0.000 | 0.000 | 0.000 |
| | | (0.00) | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) | (0.00) |
| Variation Elevation | | 0.001** | 0.001*** | 0.001*** | 0.001*** | | 0.001*** | 0.001*** | 0.001*** | 0.001*** |
| | | (0.00) | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) | (0.00) |
| Ruggedness | | -0.000** | -0.000 | -0.000 | -0.000 | | -0.000** | -0.000* | -0.000 | -0.000 |
| | | (0.00) | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) | (0.00) |
| Total Water | | | 0.042*** | 0.041*** | 0.036*** | | | 0.051*** | 0.042*** | 0.036*** |
| | | | (0.01) | (0.01) | (0.01) | | | (0.01) | (0.01) | (0.01) |
| areaMURD1000 | | | 0.031*** | 0.040*** | 0.043*** | | | 0.032*** | 0.042*** | 0.045*** |
| | | | (0.01) | (0.01) | (0.01) | | | (0.01) | (0.01) | (0.01) |
| Ln(Migratory Distance) | | | 0.031 | -0.100 | -0.113 | | | 0.044 | -0.060 | -0.078 |
| | | | (0.05) | (0.13) | (0.14) | | | (0.04) | (0.15) | (0.15) |
| Ln(Distance Coast) | | | 0.096*** | 0.032 | 0.044 | | | 0.082*** | 0.025 | 0.038 |
| | | | (0.03) | (0.02) | (0.03) | | | (0.02) | (0.02) | (0.03) |
| Ln(Distance River) | | | -0.025 | -0.033* | -0.031* | | | -0.020 | -0.031* | -0.029* |
| | | | (0.02) | (0.02) | (0.02) | | | (0.02) | (0.02) | (0.02) |
| Ln(Distance Adis Ababa) | | | -0.068*** | -0.034 | -0.048* | | | -0.057** | -0.035 | -0.050* |
| | | | (0.03) | (0.03) | (0.03) | | | (0.03) | (0.03) | (0.03) |
| Absolute Latitude | | | 0.000 | -0.005 | -0.004 | | | -0.001 | -0.005 | -0.004 |
| | | | (0.01) | (0.01) | (0.01) | | | (0.00) | (0.01) | (0.01) |
| Number of Country | | | | | -0.017 | | | | | -0.018 |
| | | | | | (0.05) | | | | | (0.05) |
| Within Country | | | | | -0.034 | | | | | -0.038 |
| | | | | | (0.06) | | | | | (0.06) |
| Ln(Distance Border) | | | | | -0.030* | | | | | -0.032* |
| | | | | | (0.02) | | | | | (0.02) |
| Country FE | | | | Yes | Yes | | | | Yes | Yes |
| N | 1972 | 1972 | 1972 | 1972 | 1972 | 1972 | 1972 | 1972 | 1972 | 1972 |
| R-Squared | 0.103 | 0.166 | 0.223 | 0.322 | 0.326 | 0.091 | 0.186 | 0.227 | 0.318 | 0.323 |

*Notes* Table reports OLS estimates. The unit of observation is a grid square of 1x1 degree of sizeThe dependent variable is the natural logarithm of the number of languages per cell, constructed based on the Murdock map of Africa. Standard Errors (in parenthesis) are clustered at the country level. ***, **, * indicate significance at 1-, 5-, and 10-% level, respectively.

TABLE 3.5: Malaria and Diversity in Pre-Colonial America

| | **Ln(Number of Languages: North and South America)** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Malaria Endemicity | 0.051 | 0.005 | -0.015 | -0.010 | -0.012 | | | | | |
| | (0.03) | (0.03) | (0.03) | (0.03) | (0.02) | | | | | |
| Malaria Ecology | | | | | | 0.034* | 0.022* | 0.014 | 0.012 | 0.007 |
| | | | | | | (0.02) | (0.01) | (0.01) | (0.01) | (0.01) |
| Average Temperature | | 0.008** | 0.013*** | 0.014*** | 0.014*** | | 0.005 | 0.013*** | 0.014*** | 0.014*** |
| | | (0.00) | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) | (0.00) |
| Average Precipitation | | 0.001*** | 0.001*** | 0.001*** | 0.001*** | | 0.001** | 0.001*** | 0.001*** | 0.001*** |
| | | (0.00) | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) | (0.00) |
| Mean Suitability | | 0.117** | 0.057 | 0.045 | 0.040 | | 0.152** | 0.079* | 0.060 | 0.051 |
| | | (0.04) | (0.05) | (0.06) | (0.05) | | (0.05) | (0.04) | (0.06) | (0.05) |
| Variation Suitability | | 0.875*** | 0.522*** | 0.382*** | 0.399*** | | 0.856*** | 0.531*** | 0.401*** | 0.415*** |
| | | (0.16) | (0.13) | (0.11) | (0.11) | | (0.16) | (0.13) | (0.12) | (0.11) |
| Mean Elevation | | 0.000 | 0.000 | 0.000 | 0.000 | | 0.000 | 0.000 | 0.000 | 0.000 |
| | | (0.00) | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) | (0.00) |
| Variation Elevation | | -0.000 | 0.000** | 0.000 | 0.000 | | 0.000 | 0.000*** | 0.000 | 0.000 |
| | | (0.00) | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) | (0.00) |
| Ruggedness | | 0.000** | 0.000** | 0.000*** | 0.000*** | | 0.000** | 0.000* | 0.000*** | 0.000*** |
| | | (0.00) | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) | (0.00) |
| Total Water | | | 0.001 | 0.008 | 0.005 | | | -0.000 | 0.007 | 0.004 |
| | | | (0.01) | (0.01) | (0.01) | | | (0.01) | (0.01) | (0.01) |
| areaCHIOVELLI1000 | | | 0.080*** | 0.083*** | 0.085*** | | | 0.081*** | 0.083*** | 0.085*** |
| | | | (0.01) | (0.01) | (0.01) | | | (0.01) | (0.01) | (0.01) |
| Ln(Migratory Distance) | | | -0.721*** | -1.595*** | -1.629*** | | | -0.704*** | -1.553** | -1.619** |
| | | | (0.12) | (0.55) | (0.49) | | | (0.11) | (0.62) | (0.58) |
| Ln(Distance Coast) | | | -0.058*** | -0.066*** | -0.041** | | | -0.065*** | -0.070*** | -0.045** |
| | | | (0.02) | (0.02) | (0.02) | | | (0.02) | (0.02) | (0.02) |
| Ln(Distance River) | | | -0.024* | -0.006 | -0.006 | | | -0.023* | -0.006 | -0.006 |
| | | | (0.01) | (0.01) | (0.01) | | | (0.01) | (0.01) | (0.01) |
| Ln(Distance Adis Ababa) | | | 0.031* | 0.005 | -0.007 | | | 0.030* | 0.006 | -0.006 |
| | | | (0.02) | (0.03) | (0.03) | | | (0.02) | (0.03) | (0.02) |
| Absolute Latitude | | | 0.004* | 0.007** | 0.007** | | | 0.006* | 0.009** | 0.008** |
| | | | (0.00) | (0.00) | (0.00) | | | (0.00) | (0.00) | (0.00) |
| Number of Country | | | | | 0.212** | | | | | 0.207*** |
| | | | | | (0.08) | | | | | (0.07) |
| Within Country | | | | | 0.241*** | | | | | 0.235*** |
| | | | | | (0.08) | | | | | (0.08) |
| Ln(Distance Border) | | | | | -0.036** | | | | | -0.035** |
| | | | | | (0.02) | | | | | (0.02) |
| Country FE | | | | Yes | Yes | | | | Yes | Yes |
| N | 2516 | 2516 | 2516 | 2516 | 2516 | 2516 | 2516 | 2516 | 2516 | 2516 |
| R-Squared | 0.020 | 0.089 | 0.181 | 0.211 | 0.216 | 0.020 | 0.093 | 0.181 | 0.211 | 0.215 |

*Notes* Table reports OLS estimates. The unit of observation is a grid square of 1x1 degree of sizeThe dependent variable is the natural logarithm of the number of Ethnic Groups per cell, constructed based on the Murdock map of South America. Standard Errors (in parenthesis) are clustered at the country level.

## 3.3 Looking for the Channel: Malaria Exposure and endogamous Marriages

In the previous section we establish the following empirical regularity: areas that are more suitable for malaria present higher levels of ethno-linguistic diversity. In this section we want to explore one of the possibile channels linking malaria to contemporaneous levels of ethno-linguistic diversity. The fact the malaria exposure increases ethno-lingustic heterogeneity is consistent with the idea of ethno-linguistic diversity being affected by human interactions. Along with fertility and mortality rates, human behaviors - such as migration and ethnic mixing - may alter the distribution of ethnic groups in a given area. In a partial equilibrium world, endogamous ethnic marriage is a crucial factor for the persistence of a given ethnic group in a given location. Moreover, historical narratives (McNeill, 1975; Browner and Johnston, 2007) and genetic studies (Kwaiatkowski, 2005) point out that heterogeneity in epidemiological endowments is conducive to the development of location-specific immunities. These two facts coupled together imply that ethnic groups historically exposed to malaria may have a comparative advantage in terms of differential mortality with respect to groups that were not exposed to malaria, or to the same indigenous strains. Our hypothesis is that ethnic endogamicity may be an optimal individual behavioral response to malaria exposure. If marriage decisions are driven by the observed differential mortality among ethnic groups, marrying within the group could maximize the survival probability of the couple and of their descendants. This response is optimal in the sense that it reduces the risk of losing the ethnic-specific immunities of the couple.

**Data** In order to test this hypothesis we use contemporary georeferenced data on marriage patterns in Africa provided by the Demographic and Health Survey (DHS). DHS surveys are constructed to provide detailed information on a sub-national representative sample of women of age ranging from 15 to 49. We exploit several waves of the survey conducted in 1815 cluster across 22 African countries. In most of the waves, data on married couples are provided. In this exercise we use those waves for which information on both the ethnic identity of the wife and the husband is available. The exercise of identification of endogamous marriages is complicated by the fact that the definition of ethnicity is not homogenous and varies along waves and countries. More precisely, for the same country we might have a wave where it is specified the ethnicity $x$ of individual $W_i$, while in the following wave individual $W_j$ is reported to belong to ethnicity $x1$, which actually represents a sub-family of ethnicity $x$. In fact, in some cases the ethno-lingustic level reported is a dialect, while in other cases we have only official national languages. This is problematic to the extent to which the variability of ethno-linguistic diversity across waves and across country is just artificially generated by the different levels of details and precision that characterized the process of DHS data collection. To mitigate this issue, our strategy is to associate all ethnic groups reported in the DHS to the various branches of the Ethnologue trees [Gordon and Grimes, 2005]. Following this procedure, we are able to assign each ethno-linguistic entry in the DHS to the respective ethno-linguistc family in the Ethnologue. For every spouse in the sample, we are therefore able to see whether he/she formed a couple with an individual belonging to her ethnic group at various level of ethno-linguistic family aggregation.

From the DHS data, we extract individual - both for female and male - socio-economic (religion, urban residence, education) and democraphic (age) characteristics. Moreover, we are able to proxy for the ethnic group population in the region. Moreover, we exploit the spatial disaggregation of the DHS data in order to construct our geographical controls. Notice that, in order to preserve respondents confindenciality, DHS uses random displacement in the georeferencing process of different clusters in each country. Urban clusters are displaced with a minimum of 0 and a maximum of 2 kilometers of error. Rural clusters are instead displaced with an error of a minimum of 0 and a maximum of 5 kilometers of positional error with a further 1% of the rural clusters displaced a minimum of 0 and a maximum of 10 kilometers (DHS website) [4]. In order to minimize this displacement issue, we construct in ArcGIS a buffer of 10 km radius around the coordinates of each cluster. In such a way, we are able to use the area inside this circle to construct cluster-level measure of malaria suitability, elevation, suitability of agriculture, temperature and precipitation. Since the area of the buffer contains the "true" location of each cluster, we think this strategy represents a reasonable solution to the random displacement adopted in DHS data.

**Empirical Strategy** Our unit of observation is the individual respondent in cluster $c$ in country $z$ for each avalaible wave $t$. In order to test whether there is a behavioral endogamous respone to malaria exposure, we are going to exploit the within-country across cluster variability using the following specification:

$$Endogamy_{i,c,z,t} = \beta_0 + \beta_1 Mal_{c,z} + \beta_2 \mathbf{X}_{c,z} + \mathbf{Z}_{i,c,z} + \mu_z + \mu_t + \epsilon_{i,c,z,t}$$

---

[4] Notice that there is no issue in the assignment of cluster near to country borders. In fact, "displacement is restricted so that the points stay within the country and within the DHS survey region (DHS website). Morevoer, "In surveys released since 2009 the displacement is restricted to the country's second administrative level where possible." (DHS website)

where $Endogamy_{i,c,z,t}$ is a dummy variable taking value one if the indivual is married with somebody belonging to the same ethnic group in cluster $c$ in country $z$ in wave $t$. $Mal_{c,z}$ is a measure of malaria suitability in cluster $c$ in country $z$. $X_{c,z}$ is a vector of time invariant geographical controls (elevation (level and standard deviation), suitability of agriculture, temperature, precipitation. $Z_{i,c,z}$ is a vector of individual characteristics (urban resident, education, age). $\mu_z$ is a country dummy, $\mu_e$ is a female ethnic dummy, and $\mu_t$ is a wave dummy.

### 3.3.1 Results

Results are presented in different tables for levels of ethnolinguist familiy aggregation of Ethnologue Level 2 to 7. Results for each Ethnologue Level are reported in Table 3.6 to 3.11. Our baseline specification is the one exploiting within-country across cluster variation in the relationship between endogamous marriages and malaria exposure. Results show a robust correlation between endogamous marriages and malaria exposure when we look at at intermediate levels of ethnic disaggregation. Within a country, individuals living in areas with a higher suitability to malaria have higher propensity to marry somebody of the same group. The result holds when controlling for individual level of education, age and whether the individual is a urban or rural dwellers.

TABLE 3.6: Malaria and Endogamy, Ethnologue Level 2

| | **Endogamy Ethnologue Level 2** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Malaria Endemicity Buf 10km | 0.003 | 0.005* | 0.000 | -0.002 | | | | |
| | (0.00) | (0.00) | (0.00) | (0.00) | | | | |
| Malaria Ecology Buf 10km | | | | | 0.001 | 0.002** | 0.002 | 0.001 |
| | | | | | (0.00) | (0.00) | (0.00) | (0.00) |
| Suitability Buf 10km | | -0.067** | -0.056* | -0.062** | | -0.076** | -0.062** | -0.064** |
| | | (0.03) | (0.03) | (0.03) | | (0.03) | (0.03) | (0.03) |
| Temperature Buf 10km | | -0.006** | -0.007** | -0.005*** | | -0.008** | -0.009** | -0.006** |
| | | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) |
| Elevation std Buf 10km | | -0.000 | -0.000 | 0.000 | | -0.000 | -0.000 | -0.000 |
| | | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) |
| Elevation Buf 10km | | -0.000 | -0.000 | -0.000 | | -0.000 | -0.000 | -0.000 |
| | | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) |
| Precipitation Buf 10km | | -0.000 | -0.000 | -0.000 | | -0.000 | 0.000 | -0.000 |
| | | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) |
| Urban Residence Female | | | -0.043*** | -0.041*** | | | -0.041*** | -0.039*** |
| | | | (0.01) | (0.01) | | | (0.01) | (0.01) |
| Age Female | | | 0.000 | 0.000 | | | 0.000 | 0.000 |
| | | | (0.00) | (0.00) | | | (0.00) | (0.00) |
| Highest Education Female | | | -0.005 | -0.007 | | | -0.005 | -0.007 |
| | | | (0.00) | (0.00) | | | (0.00) | (0.00) |
| Age Male | | | -0.000 | -0.000 | | | -0.000 | -0.000 |
| | | | (0.00) | (0.00) | | | (0.00) | (0.00) |
| Highest Education Male | | | -0.004 | -0.004 | | | -0.004 | -0.004 |
| | | | (0.00) | (0.00) | | | (0.00) | (0.00) |
| People in the Group Tree2 | | | 0.000** | 0.000*** | | | 0.000** | 0.000*** |
| | | | (0.00) | (0.00) | | | (0.00) | (0.00) |
| Religion FE | | | Yes | Yes | | | Yes | Yes |
| Ethnicity FE | | | | Yes | | | | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 76888 | 76871 | 73887 | 73887 | 76888 | 76871 | 73887 | 73887 |
| R-Squared | 0.037 | 0.041 | 0.064 | 0.077 | 0.038 | 0.042 | 0.064 | 0.077 |

*Notes* Table reports OLS estimates. The unit of observation is the individual respondent to the Demographic Health Survey. Standard Errors (in parenthesis) are clustered at the country level. The dependent variable is a dummy variable taking value 1 if the individual is married with somebody of the same ethnic group family. The ethnic group family is defined as for Ethnologue Level Tree number 2. The geographic variables were constructed drawing a 10 km radius around the coordinates of the DHS cluster to which the individual respondent belong to. ***, **, * indicate significance at 1-, 5-, and 10-% level, respectively.

TABLE 3.7: Malaria and Endogamy, Ethnologue Level 3

| | **Endogamy Ethnologue Level 3** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Malaria Endemicity Buf 10km | 0.002 | 0.003 | -0.001 | -0.004 | | | | |
| | (0.00) | (0.00) | (0.00) | (0.00) | | | | |
| Malaria Ecology Buf 10km | | | | | 0.002*** | 0.003*** | 0.002* | 0.001 |
| | | | | | (0.00) | (0.00) | (0.00) | (0.00) |
| Suitability Buf 10km | | -0.034 | -0.022 | -0.029 | | -0.045 | -0.030 | -0.032 |
| | | (0.03) | (0.03) | (0.03) | | (0.03) | (0.03) | (0.02) |
| Temperature Buf 10km | | -0.006** | -0.007** | -0.005** | | -0.010** | -0.010** | -0.006*** |
| | | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) |
| Elevation std Buf 10km | | 0.000*** | 0.000** | 0.000*** | | 0.000*** | 0.000** | 0.000*** |
| | | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) |
| Elevation Buf 10km | | -0.000** | -0.000*** | -0.000*** | | -0.000** | -0.000*** | -0.000*** |
| | | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) |
| Precipitation Buf 10km | | -0.000 | -0.000 | -0.000 | | -0.000 | 0.000 | -0.000 |
| | | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) |
| Urban Residence Female | | | -0.043*** | -0.040*** | | | -0.040*** | -0.038*** |
| | | | (0.01) | (0.01) | | | (0.01) | (0.01) |
| Age Female | | | 0.000 | 0.000 | | | 0.000 | 0.000 |
| | | | (0.00) | (0.00) | | | (0.00) | (0.00) |
| Highest Education Female | | | -0.006 | -0.011*** | | | -0.006 | -0.011*** |
| | | | (0.00) | (0.00) | | | (0.00) | (0.00) |
| Age Male | | | -0.000 | -0.000 | | | -0.000 | -0.000 |
| | | | (0.00) | (0.00) | | | (0.00) | (0.00) |
| Highest Education Male | | | -0.004 | -0.006** | | | -0.003 | -0.006** |
| | | | (0.00) | (0.00) | | | (0.00) | (0.00) |
| People in the Group Tree3 | | | 0.000** | 0.000** | | | 0.000** | 0.000** |
| | | | (0.00) | (0.00) | | | (0.00) | (0.00) |
| Religion FE | | | Yes | Yes | | | Yes | Yes |
| Ethnicity FE | | | | Yes | | | | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 75049 | 75032 | 72050 | 72050 | 75049 | 75032 | 72050 | 72050 |
| R-Squared | 0.044 | 0.046 | 0.070 | 0.091 | 0.045 | 0.048 | 0.071 | 0.091 |

*Notes* Table reports OLS estimates. The unit of observation is the individual respondent to the Demographic Health Survey. Standard Errors (in parenthesis) are clustered at the country level. The dependent variable is a dummy variable taking value 1 if the individual is married with somebody of the same ethnic group family. The ethnic group family is defined as for Ethnologue Level Tree number 3. The geographic variables were constructed drawing a 10 km radius around the coordinates of the DHS cluster to which the individual respondent belong to. ***, **, * indicate significance at 1-, 5-, and 10-% level, respectively.

TABLE 3.8: Malaria and Endogamy, Ethnologue Level 4

| | **Endogamy Ethnologue Level 4** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Malaria Endemicity Buf 10km | 0.003 | 0.005* | 0.001 | -0.004 | | | | |
| | (0.00) | (0.00) | (0.00) | (0.00) | | | | |
| Malaria Ecology Buf 10km | | | | | 0.002* | 0.003** | 0.002 | 0.001 |
| | | | | | (0.00) | (0.00) | (0.00) | (0.00) |
| Suitability Buf 10km | | -0.036 | -0.021 | -0.022 | | -0.044 | -0.027 | -0.025 |
| | | (0.03) | (0.04) | (0.03) | | (0.03) | (0.03) | (0.03) |
| Temperature Buf 10km | | -0.007** | -0.008** | -0.002 | | -0.010** | -0.010** | -0.004* |
| | | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) |
| Elevation std Buf 10km | | 0.000*** | 0.000 | 0.000* | | 0.000** | 0.000 | 0.000* |
| | | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) |
| Elevation Buf 10km | | -0.000** | -0.000*** | -0.000** | | -0.000** | -0.000*** | -0.000** |
| | | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) |
| Precipitation Buf 10km | | -0.000 | -0.000 | -0.000 | | -0.000 | 0.000 | -0.000 |
| | | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) |
| Urban Residence Female | | | -0.041*** | -0.038*** | | | -0.038*** | -0.035*** |
| | | | (0.01) | (0.01) | | | (0.01) | (0.01) |
| Age Female | | | 0.000 | -0.000 | | | 0.000 | -0.000 |
| | | | (0.00) | (0.00) | | | (0.00) | (0.00) |
| Highest Education Female | | | -0.006 | -0.014*** | | | -0.006 | -0.014*** |
| | | | (0.00) | (0.00) | | | (0.00) | (0.00) |
| Age Male | | | -0.000 | -0.000 | | | -0.000 | -0.000 |
| | | | (0.00) | (0.00) | | | (0.00) | (0.00) |
| Highest Education Male | | | -0.005 | -0.009** | | | -0.005 | -0.009** |
| | | | (0.00) | (0.00) | | | (0.00) | (0.00) |
| People in the Group Tree4 | | | 0.000** | 0.000* | | | 0.000** | 0.000* |
| | | | (0.00) | (0.00) | | | (0.00) | (0.00) |
| Religion FE | | | Yes | Yes | | | Yes | Yes |
| Ethnicity FE | | | | Yes | | | | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 73125 | 73108 | 70163 | 70163 | 73125 | 73108 | 70163 | 70163 |
| R-Squared | 0.044 | 0.046 | 0.069 | 0.092 | 0.045 | 0.048 | 0.070 | 0.092 |

*Notes* Table reports OLS estimates. The unit of observation is the individual respondent to the Demographic Health Survey. Standard Errors (in parenthesis) are clustered at the country level. The dependent variable is a dummy variable taking value 1 if the individual is married with somebody of the same ethnic group family. The ethnic group family is defined as for Ethnologue Level Tree number 4. The geographic variables were constructed drawing a 10 km radius around the coordinates of the DHS cluster to which the individual respondent belong to. ***, **, * indicate significance at 1-, 5-, and 10-% level, respectively.

TABLE 3.9: Malaria and Endogamy, Ethnologue Level 5

| | **Endogamy Ethnologue Level 5** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Malaria Endemicity Buf 10km | 0.000 | 0.003 | -0.002 | -0.003 | | | | |
| | (0.00) | (0.00) | (0.00) | (0.00) | | | | |
| Malaria Ecology Buf 10km | | | | | 0.003*** | 0.004*** | 0.004*** | 0.003** |
| | | | | | (0.00) | (0.00) | (0.00) | (0.00) |
| Suitability Buf 10km | | -0.007 | 0.009 | 0.008 | | -0.018 | -0.001 | -0.000 |
| | | (0.03) | (0.03) | (0.03) | | (0.03) | (0.03) | (0.03) |
| Temperature Buf 10km | | -0.005 | -0.007** | -0.003** | | -0.010** | -0.013*** | -0.007*** |
| | | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) |
| Elevation std Buf 10km | | 0.000 | 0.000 | 0.000 | | -0.000 | 0.000 | -0.000 |
| | | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) |
| Elevation Buf 10km | | -0.000 | -0.000** | -0.000 | | -0.000 | -0.000*** | -0.000* |
| | | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) |
| Precipitation Buf 10km | | -0.000 | -0.000 | -0.000 | | -0.000 | 0.000 | -0.000 |
| | | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) |
| Urban Residence Female | | | -0.036*** | -0.029*** | | | -0.030*** | -0.025*** |
| | | | (0.01) | (0.01) | | | (0.01) | (0.00) |
| Age Female | | | -0.000 | -0.000 | | | -0.000 | -0.000 |
| | | | (0.00) | (0.00) | | | (0.00) | (0.00) |
| Highest Education Female | | | -0.011** | -0.018*** | | | -0.011** | -0.018*** |
| | | | (0.01) | (0.00) | | | (0.00) | (0.00) |
| Age Male | | | 0.000 | 0.000 | | | 0.000 | 0.000 |
| | | | (0.00) | (0.00) | | | (0.00) | (0.00) |
| Highest Education Male | | | -0.005 | -0.008** | | | -0.004 | -0.007** |
| | | | (0.00) | (0.00) | | | (0.00) | (0.00) |
| People in the Group Tree5 | | | 0.000** | 0.000*** | | | 0.000*** | 0.000*** |
| | | | (0.00) | (0.00) | | | (0.00) | (0.00) |
| Religion FE | | | Yes | Yes | | | Yes | Yes |
| Ethnicity FE | | | | Yes | | | | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 66107 | 66090 | 63164 | 63164 | 66107 | 66090 | 63164 | 63164 |
| R-Squared | 0.055 | 0.057 | 0.076 | 0.103 | 0.058 | 0.060 | 0.079 | 0.104 |

*Notes* Table reports OLS estimates. The unit of observation is the individual respondent to the Demographic Health Survey. Standard Errors (in parenthesis) are clustered at the country level. The dependent variable is a dummy variable taking value 1 if the individual is married with somebody of the same ethnic group family. The ethnic group family is defined as for Ethnologue Level Tree number 5. The geographic variables were constructed drawing a 10 km radius around the coordinates of the DHS cluster to which the individual respondent belong to. ***, **, * indicate significance at 1-, 5-, and 10-% level, respectively.

TABLE 3.10: Malaria and Endogamy, Ethnologue Level 6

| | **Endogamy Ethnologue Level 6** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Malaria Endemicity Buf 10km | 0.004 | 0.006 | 0.000 | -0.001 | | | | |
| | (0.00) | (0.00) | (0.00) | (0.00) | | | | |
| Malaria Ecology Buf 10km | | | | | 0.004*** | 0.005*** | 0.003** | 0.002 |
| | | | | | (0.00) | (0.00) | (0.00) | (0.00) |
| Suitability Buf 10km | | 0.001 | -0.002 | 0.012 | | -0.003 | -0.006 | 0.006 |
| | | (0.03) | (0.03) | (0.04) | | (0.02) | (0.03) | (0.03) |
| Temperature Buf 10km | | 0.000 | -0.005 | -0.003 | | -0.007 | -0.009** | -0.006* |
| | | (0.00) | (0.00) | (0.00) | | (0.01) | (0.00) | (0.00) |
| Elevation std Buf 10km | | 0.000 | 0.000 | -0.000 | | 0.000 | -0.000 | -0.000 |
| | | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) |
| Elevation Buf 10km | | 0.000 | -0.000* | -0.000 | | -0.000 | -0.000* | -0.000 |
| | | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) |
| Precipitation Buf 10km | | -0.001 | -0.000 | -0.001** | | -0.000 | -0.000 | -0.000** |
| | | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) |
| Urban Residence Female | | | -0.040*** | -0.034*** | | | -0.036*** | -0.031*** |
| | | | (0.01) | (0.01) | | | (0.01) | (0.01) |
| Age Female | | | -0.000 | -0.000 | | | -0.000 | -0.000 |
| | | | (0.00) | (0.00) | | | (0.00) | (0.00) |
| Highest Education Female | | | -0.020*** | -0.022*** | | | -0.020*** | -0.022*** |
| | | | (0.01) | (0.00) | | | (0.00) | (0.00) |
| Age Male | | | 0.000 | 0.000 | | | 0.000 | 0.000 |
| | | | (0.00) | (0.00) | | | (0.00) | (0.00) |
| Highest Education Male | | | -0.010*** | -0.012*** | | | -0.010** | -0.012** |
| | | | (0.00) | (0.00) | | | (0.00) | (0.00) |
| People in the Group Tree6 | | | 0.000** | 0.000** | | | 0.000** | 0.000** |
| | | | (0.00) | (0.00) | | | (0.00) | (0.00) |
| Religion FE | | | Yes | Yes | | | Yes | Yes |
| Ethnicity FE | | | | Yes | | | | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 58790 | 58773 | 55886 | 55886 | 58790 | 58773 | 55886 | 55886 |
| r2 | 0.078 | 0.082 | 0.114 | 0.141 | 0.083 | 0.086 | 0.116 | 0.142 |

*Notes* Table reports OLS estimates. The unit of observation is the individual respondent to the Demographic Health Survey. Standard Errors (in parenthesis) are clustered at the country level. The dependent variable is a dummy variable taking value 1 if the individual is married with somebody of the same ethnic group family, and 0 otherwise. The ethnic group family is defined as for Ethnologue Level Tree number 6. The geographic variables were constructed drawing a 10 km radius around the coordinates of the DHS cluster to which the individual respondent belong to. ***, **, * indicate significance at 1-, 5-, and 10-% level, respectively.

Table 3.11: Malaria and Endogamy, Ethnologue Level 7

| | **Endogamy Ethnologue Level 7** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Malaria Endemicity Buf 10km | 0.001 | 0.003 | -0.002 | -0.001 | | | | |
| | (0.01) | (0.01) | (0.00) | (0.00) | | | | |
| Malaria Ecology Buf 10km | | | | | 0.003* | 0.003** | 0.002 | 0.002 |
| | | | | | (0.00) | (0.00) | (0.00) | (0.00) |
| Suitability Buf 10km | | -0.010 | -0.016 | -0.003 | | -0.017 | -0.021 | -0.010 |
| | | (0.02) | (0.03) | (0.04) | | (0.02) | (0.03) | (0.04) |
| Temperature Buf 10km | | -0.000 | -0.005 | -0.001 | | -0.006 | -0.009** | -0.004 |
| | | (0.01) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) |
| Elevation std Buf 10km | | 0.000 | 0.000 | -0.000 | | 0.000 | 0.000 | -0.000 |
| | | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) |
| Elevation Buf 10km | | 0.000 | -0.000 | 0.000 | | -0.000 | -0.000 | -0.000 |
| | | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) |
| Precipitation Buf 10km | | -0.000* | -0.000 | -0.000 | | -0.000 | -0.000 | -0.000 |
| | | (0.00) | (0.00) | (0.00) | | (0.00) | (0.00) | (0.00) |
| Urban Residence Female | | | -0.041*** | -0.035*** | | | -0.038*** | -0.033*** |
| | | | (0.01) | (0.01) | | | (0.01) | (0.01) |
| Age Female | | | -0.000 | -0.000 | | | -0.000 | -0.000 |
| | | | (0.00) | (0.00) | | | (0.00) | (0.00) |
| Highest Education Female | | | -0.015** | -0.018*** | | | -0.015** | -0.018*** |
| | | | (0.01) | (0.01) | | | (0.01) | (0.01) |
| Age Male | | | 0.000 | 0.000 | | | 0.000 | 0.000 |
| | | | (0.00) | (0.00) | | | (0.00) | (0.00) |
| Highest Education Male | | | -0.009* | -0.011** | | | -0.009* | -0.011** |
| | | | (0.00) | (0.00) | | | (0.00) | (0.00) |
| People in the Group Tree7 | | | 0.000** | 0.000* | | | 0.000** | 0.000* |
| | | | (0.00) | (0.00) | | | (0.00) | (0.00) |
| Religion FE | | | Yes | Yes | | | Yes | Yes |
| Ethnicity FE | | | | Yes | | | | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| N | 51647 | 51630 | 49325 | 49325 | 51647 | 51630 | 49325 | 49325 |
| R-squared | 0.064 | 0.066 | 0.092 | 0.113 | 0.066 | 0.068 | 0.093 | 0.114 |

*Notes* Table reports OLS estimates. The unit of observation is the individual respondent to the Demographic Health Survey. Standard Errors (in parenthesis) are clustered at the country level. The dependent variable is a dummy variable taking value 1 if the individual is married with somebody of the same ethnic group family. The ethnic group family is defined as for Ethnologue Level Tree number 7. The geographic variables were constructed drawing a 10 km radius around the coordinates of the DHS cluster to which the individual respondent belong to. ***, **, * indicate significance at 1-, 5-, and 10-% level, respectively.

## 3.4 Conclusion

In this paper we hypothesize that malaria is a historical determinant of contemporary ethno-linguistic diversity. In order to test this hypothesis, we show that higher historical malaria incidence and higher geographic suitability to malaria are both associated to higher linguistic diversity. We propose a channel linking malaria incidence and diversity, we conjecture that malaria increased the propensity to marry within the groups. We test this hypothesis by looking at marriage patterns within 22 African countries, we show that geographic suitability to malaria is associated with a higher rate of endogamous marriages.

# Chapter 4

# Malaria Risk and Civil Violence: A Disaggregated Analysis for Africa

## 4.1 Introduction

This research provides an empirical test of the hypothesis that a greater exposure to the health-threatening risks of malaria epidemics increases the likelihood of civil violence in Africa. This hypothesis is motivated by the observation that poor health, and the exposure to life threatening pathogens, potentially reduces the opportunity costs for violent behavior through several channels. Recent contributions to the literature suggest that poorer health is associated with riskier behavior.[1] Individuals who face high mortality rates also tend to be less future oriented.[2] Bad health might sensibly deteriorate living conditions and thus be conducive to violence, similar to reductions in economic well-being that have been argued to be conducive to violence.[3] Moreover, a deterioration of

---

[1] High extrinsic mortality rates related with extreme events like the Chernobyl accident in 1986 substantially increased death rates related to smoking, drinking and car speeding, the practice of unsafe sex and the spread of HIV, see, e.g., Lorentzen/etal:2008. Adda2013 show that individuals in poorer health are more likely to start smoking. Experimental findings by Lammers and van Wijnbergen (2008) show that HIV positive South African students are less risk-averse and display larger discount rates. Oster:2012 finds that behavioral responses to HIV prevention efforts are smaller in area highly exposed to malaria.

[2] The link between mortality and future orientation has been formalized by Becker and Mulligan (1997). In theory, a high discount rate increases defection in repeated strategic interactions (both at the individual and group level) thereby reducing the scope for cooperative equilibria involving a peaceful balancing of conflicts of interests. See, for instance, Dutta (1995). That the exposure to health hazards affects future orientation and behavior by increasing people's short-sightedness has been documented in social psychology, see Aspinwall:2005.

[3] See, e.g., Bellows/Miguel:2009 for a survey.

health conditions may reduce the utility from adopting low risk (or risk-avoiding) behaviors, thereby increasing the willingness to take risks and resort to violence.[4] Finally, wore health conditions and lower fighting power among potential target groups might increase incentives for violent acts.

The empirical analysis exploits disaggregated and geo-referenced panel data that superimpose a grid of equally-sized cells on the territory of interest. The unit of observation is a grid cell (in grids of 1×1 and 2.5×2.5 degrees of size), whose borders are conceptually unrelated to conflicts as they are based exclusively on geodesic principles (latitude and longitude) and do not coincide with borders of national or regional entities. The use of spatially (cell-level) and temporally (yearly and monthly) disaggregated data allows exploiting more disaggregate information and higher variability (both in terms of violent events data, geo-covariates and weather conditions) than an analysis at the country-level as in most of the literature. The identification exploits not only within-country, but even within cell variation over time. Another advantage of using disaggregate data is that it allows testing the hypothesis of a link between health conditions and risky or violent behaviour, which is not confined to organized civil conflicts and civil wars, by using data on (unorganized) civil violence from the Armed Conflict Location & Event Data Project (ACLED).

To identify areas and periods in which malaria represents a serious health threat we follow the epidemiological literature, which has documented that malaria malaria threat or malaria risk is conceptually different from malaria incidence. The risk of malaria is higher when the conditions are suitable for the spread and infection through the vector (the *Anopheles* mosquito) *and* when the human population has low malaria resistance in terms of both acquired and genetic immunities. More specifically, this has two implications. Firstly, in a cross-sectional perspective the malaria risk is highest in areas (cells) with intermediate levels of malaria endemicity. The reason is that in areas with high malaria endemicity individuals develop acquired immunities (by surviving malaria infections) and genetic immunities (such as different types of blood disorders that protect against malaria infections). These immunities effectively reduce the risk of contracting the disease. Secondly, from a dynamic (panel) perspective the risk of malaria is highest in periods that are unusually suitable for malaria out-brakes and in areas with low to intermediate malaria incidence. The reason is that the adult population in such geographic areas (cells) does not build up an adequate stock of acquired and genetic immunities when the exposure is low over long periods and high only in infrequent (rare)

---

[4] In a standard expected utility perspective, a reduction in the utility associated with the safest alternative (e.g., abstaining from violent activities) increases the optimal level of risk taking. Individuals with lower levels of utility will invest less in safety-seeking activities since they have less to loose in expected terms. Following earlier experimental findings, Goudie et al. (2014) formalize this argument and provide specific evidence for this channel using data for seat belt wearing and automobile accidents.

periods that are unusually suitable for malaria out-brakes. We investigate the link between malaria risk and civil violence in econometric specifications that incorporate these insights by allowing for non-linear effects and exploiting variation across geographic cells and variation within cells over time.

As a first step of the analysis, we conduct estimations that exploit cross-cell variability in malaria incidence. Depending on cell size (in grids of 1×1 and 2.5×2.5 degrees of size), the analysis exploits between 440 and 2,600 observations. The malaria exposure is proxied by an index devised by Kiszewski et. al. (2004) , which measures the cell-specific malaria ecology. This measure effectively captures the likelihood of being bitten by a mosquito, depending on the degree of anthropophily of the mosquitoes specific to the region, and on their degree of dangerousness based on climatic and geographical features.[5] In line with the predictions of epidemiologists, the estimation results document the existence of a reverse U-shaped effect of malaria on civil violence: Conflict is higher in grid cells in which malaria incidence is intermediate and the malaria threat is at the maximum. The findings are compatible with the view that violence is higher in grid cells with a sufficient geo-climatic suitability for malaria, but a relatively low level of immunity in the population.

In a second step, the analysis is extended to exploiting within-cell variability over time. The estimation is conducted using partially interacted models that allow for the possibility that months suitable for malaria have heterogenous effects on civil conflict depending on the level of long-term malaria exposure (and therefore indirectly depending on the level of acquired immunities) in a given cell. Periods that are unusually suitable for malaria are identified following the approach by Tanser et. al. (2003) and using high definition data on weather conditions to predict months that are suitable for the mosquito vectors to become active and infectious. The resulting index has been shown to predict the probability of malaria outbreaks at the cell-month level with high accuracy.[6] The short-term (epidemic) nature of malaria risk allows exploiting within-cell variability both at yearly and monthly frequencies, thereby accounting for cell specific and cell-year specific fixed effects.[7] The analysis is based on panel data with about

---

[5] We also use alternative information from a historical measure of malaria endemicity at the beginning of the twentieth century, produced by Lysenko1968 and recently digitalized by Hay et. al. (2004). This index is supposed to measure malaria endemicity in a time in which the massive malaria eradication campaigns had still to be conceived and realized. Furthermore, this measure is informative about the historical presence and intensity of the Plasmodium parasites in the blood of surviving children, which is indirectly informative about their genetic immunities.

[6] A recent application by kudamatsu2012weather links malaria outbreaks to child mortality. As discussed in Section 4.2, the measure exploits weather information about precipitation and temperature and requires four necessary conditions that need to be met in the past three to twelve months for plasmodium to be transmittable.

[7] Differently from other possible triggers of civil conflicts that may involve delays in changes to living conditions, like income shocks, the actual risk of epidemics is limited in time and can be predicted with high accuracy exploiting information on the weather conditions in preceding months.

35,000 cell-level observations at yearly frequency and about 430,000 cell-level observations at monthly frequency. The results show that periods that are unusually suitable for malaria increase the likelihood of civil violence only in areas with low to intermediate malaria ecology. The results are very similar when exploiting data with yearly or monthly variability. The results further suggest that malaria risk mainly affects the likelihood of non-organized (non-battle-related) civil violence in the form of riots, protests and violence involving civilians. The main findings are robust to accounting for alternative weather shocks (in terms of temperature, precipitation, and the inclusion of an evapotranspiration index) and their interactions with intermediate malaria suitability. The patterns also hold when accounting for spatial autocorrelation in the errors and the co-variates using spatial econometric techniques.

The paper contributes to the empirical literature that investigates the determinants of civil conflicts and civil violence. Most of the literature has focused attention to countries as unit of analysis. The paper that is closest in spirit to the current analysis is by Cervellati, Sunde and Valmori (2014), which links the exposure to a harsher disease environment, in terms of multi-host vector-transmitted pathogens, to civil conflicts using cross-country data. Their results complement the existing evidence on the role of geographic, socio-economic and ethnic (or genetic) diversity and the role of short term triggers, like income and weather shocks, in explaining civil conflicts across countries.[8] This paper also contributes to a smaller, but growing literature, that studies the determinants of civil conflicts using disaggregate data at sub-national levels for Africa. The closest analysis in terms of data and methodology is by harari2013conflict who investigate the role of weather shocks that favor agricultural production, on conflict.[9] Their analysis uses variation in the Standard Precipitation and Evapotranspiration Index (SPEI) during the (cell-specific) growing season.[10] Our results complement this literature by focusing on health shocks, rather than income shocks.[11] In this paper we also extend the scope of the analysis by exploiting within-cell variation over time. The year-to-year variation in the data allows us to account for all time invariant cell specific

---

[8] The cross-country literature identifies income and poverty, weak or non-democratic institutions, political instability, geographic features such as rough terrain, ethnic fragmentation and polarization, genetic diversity, and population pressure that have been identified, among others, the works by Fearon/Laitin:2003, Collier/Hoeffler:2004, Montalvo/Reynal-Querol:2005a, Collier/Rohner:2008, Collier/Hoeffler/Rohner:2009, Esteban/Mayoral/Ray:2012, Arbatli/Ashraf/Galor:2013. Investigations of the short term triggers of conflicts that exploit within country year-to-year variation include Miguel/Satyanath/Sergenti:2004, Ciccone:2011, Besley and Persson (2009, 2011), and Couttenier/-Soubeyran:2014.

[9] The works by Rohner/Thoenig/Zilibotti:2013 and Amodio/Chiovelli:2014 also exploit disaggregate panel data to investigate the determinants of civil conflict in Congo and South Africa, respectively.

[10] Similarly to harari2013conflict, we use data for all African countries and account for country-specific characteristics by including country fixed effects.

[11] The results therefore complement the findings by Harari and La Ferrara (2013) that found a role for weather shocks related to agricultural production for civil conflicts involving battles but little, or no, effect on civil violence. In turn, our results suggest that temporary changes in malaria risk are highly related to non organized, non-battle related, civil violence.

characteristics. The month-to-month variation further allows us to account for time varying cell specific characteristics by including cell-year fixed effects. By highlighting the potentially relevant, but so far neglected, role of weather-related short-term changes in malaria risk for civil violence, the paper also contributes to the ongoing debate on the role of weather for civil conflicts, see, e.g., Hsiang et al. (2013). The results have relevant policy implications in view of the increasing warnings on the possible role of climate change for an expansion of malaria exposure to areas inhabited by populations that have not developed immunity to malaria and are therefore particularly at risk.

## 4.2 Empirical Analysis

### 4.2.1 Background and Data

To test the hypothesis that health hazards may facilitate violent activities we use disaggregated data for Africa and focus on Malaria for numerous reasons. First, malaria is highly relevant for human health.[12] Second, being transmitted only through vectors (the female *Anopheles* mosquito) the risk of contracting malaria in a particular area, and the occurrence of epidemics, is closely related to climatic conditions; malaria breeds where there is an abundance of humidity and rain. This implies as additional advantage for the analysis that malaria risk can be mapped at a high geographical resolution.

The measure of malaria incidence used is the Malaria Ecology Index, an ecology-based spatial index of stability and force of malaria transmission devised by . This index is constructed by exploiting the interaction of climate with the biological characteristics of the *Anopheles* mosquito dominant in the respective region. More precisely, constructed a map of the distribution of dominant *Anopheles* mosquitoes. This map consists of about 260 distinct regions. Based on the distribution of dominant *Anopheles* mosquitos, they devised an index of transmission force and stability as a function of several characteristics of the mosquito: i) share of blood meals taken by the mosquito, ii) the daily survival rate, and iii) the extrinsic incubation period. In order to obtain a finer data resolution, a minimum threshold of precipitation (10 mm) was imposed as necessary precondition for subsequent malaria transmission. The resulting monthly index was finally aggregated into a cross-sectional yearly index of malaria incidence. The final index ranges from 0 to 34.

---

[12] Malaria is a serious (and sometimes fatal) disease that is endemic in 109 countries around the world. About 3.3 billion people (corresponding to half of the world population) is susceptible to the disease. In 2010, 216 million clinical cases of malaria were recorded worldwide, about 90% of malaria-related deaths occur south of the Sahara in Africa.

The possibility of malaria transmission is closely related to weather conditions, as growth and survival of the *Plasmodium* parasites that cause the disease and of the vectors (the *Anopheles* mosquito) depend heavily on temperature and precipitation. More importantly, abnormal levels of temperature and precipitation can give rise to particularly harmful conditions. The 20th Report of the WHO Expert Committee on Malaria called for the need of simple indicators relying on meteorological information to be used to predict and prevent malaria epidemics.[13] Tanser et. al. (2003) provide such a weather-based index for conditions particularly suitable for malaria for Africa. The resulting binary index, labelled "Malaria Suitable Month" (or simply MSM) in the following, is constructed as follows. The binary index for grid cell $g$ for month $t$, $MSM_{g,t}$, takes value 1 if, and only if, all of the following four conditions are satisfied:

- Average monthly rainfall during the past 3 months ($t-2$, $t-1$, $t$) is at least 60mm/m$^2$.
- Rainfall in at least one of the past 3 months is at least 80 mm/m$^2$.
- No month in the past 12 months has an average temperature below $5^oC$.
- The average temperature in the past 3 months exceeds $19.5^oC$+SD (monthly temperature in the past 12 months).

Thus, the variable Malaria-Suitable Month (MSM) takes value 1 if malaria can be transmitted in the given cell and month, and 0 otherwise. For use with yearly data frequency, the variable Malaria Suitable Months is constructed as the yearly aggregate of monthly suitability conditions, and thus ranges from 0 to 12. For constructing the variable Malaria Suitable Months Demeaned (MSM Demeaned), we compute the yearly average of malaria-suitable months over the sample and subtract if from the yearly MSM variable. Weather data used for the construction of MSM, and for the covariates included in the regressions, are taken from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim dataset, which offers the advantage of being not based on gauge data but on data re-analysis.[14]

The data for civil violence is from the ACLED Dataset (version 3, 2012). The data cover all African countries over the period 1997-2012 and contain event types classified into conflicts, battles, riots and protests, and actors classified into governments, rebels,

---

[13] The report states "An increasing number of malaria epidemics have been recently documented throughout the world, particularly in Africa. Areas become epidemic when conditions that normally limit transmission change radically as a result of abnormally heavy rains, long periods of increased humidity and temperature.".

[14] Re-analysis involves model simulations of past events that include the incorporation of historical observations taken form various sources, such as weather stations, satellites, and sensors. These are the best available weather data and have previously been used by kudamatsu2012weather and harari2013conflict.

militias, ethnic groups, active political organizations, and civilians. In the cross-sectional regression we construct a variable measuring the fraction of years during which at least one conflict event was registered. In the panel analysis, the indicator of conflict is a dummy equal to one if at least one conflict event of any type occurred in a given cell in a given month or year, respectively. We also consider a breakdown of conflict events into sub-categories to explore the existence of different determinants of organized battle-related events and non-battle related civil violence in the form of riots, protests and violence involving civilians.

We construct an extensive set of covariates to be used as control variables (and for counterfactual exercises) at the grid-cell level. Under the heading “Geographic Controls”, we include absolute latitude, mean elevation, average terrain ruggedness, total cell area, total area of the cell occupied by water, average precipitation and average temperature. The controls summarized as “Location and Distances” include the natural logarithm of the distance to the country capital, to the coast, to the country border, to the closest river and to Addis Ababa. The group of variables labelled “Natural Resources” comprise the average land suitability for agriculture, the presence of diamond mines, and the presence of petrol fields. The ”Ethnic Diversity” variable measures the number of ethnic groups in the cell. In some specifications we also control for the average population density in the cell and average night light intensity. See Table D.2 in the Appendix for a description of data definitions, construction and sources.

## 4.3 Empirical Results: Summary

### 4.3.1 Cross-Cell Results

Table 4.1 reports OLS estimates of the cross-sectional relation between Malaria Ecology and the fraction of years with a at least one conflict event in the period 1998-2011. In this period, a violent event has been registered in about 40% of the cells. All regressions include country fixed effects to control for all characteristics that are common for cells belonging to the same country. The linear effect, reported in Column (1), (2) and (3), is positive but becomes insignificant in the most extensive specifications. In light of the findings from the epidemiological literature reported in the Introduction, the linear specification might not be appropriate if malaria risk is higher for intermediate malaria exposure in terms of ecology. Consistent with this prediction, the results for a quadratic specification in Columns (4), (5) and (6) reveal a hump-shaped relationship between Malaria Ecology and the frequency of conflict events. Once including strictly exogenous controls, Column (5), the estimated linear and quadratic coefficients become both significant at the 1 percent level. Depending on the specifications, the effect is maximum for levels of malaria ecology ranging from 12 to 18.

Table 4.2 shows that the same hump-shaped profile also emerges from pooled OLS estimates that exploit observations at the year level. The specifications include country and year fixed effects. In Columns (3), (4) and (5) we add time-varying climatic controls in order to account for the effect of weather-related income shocks on the probability of observing a conflict event in the year. Column (4) and (5) include a full set of country×year fixed effects.

### 4.3.2 Panel Results based on within cell variation overtime

**Yearly Data.** We next move to exploiting exogenous changes in weather conditions to identify short-term variations in malaria risk. Based on the epidemiological literature, we hypothesize malaria threat to be higher in cells with a positive, but low to intermediate, average malaria exposure. To test this hypothesis, we allow for the possibility that unusual changes in malaria suitable months (MSM) have heterogenous effects in cells with low-intermediate malaria exposure and in cells with high malaria exposure. Table 4.3 reports the effect of having one more (less) malaria-suitable month in cells with intermediate malaria ecology, on the likelihood of violence. The variable $IntermediateME \times MSMDem.$ represents the interaction term between the Intermediate Malaria Ecology and $Mal.SuitableMonthsDem$, the deviation from average malaria-suitable months in a cell. As before, areas with intermediate malaria exposure in terms of ecology reveal a higher conflict likelihood. The main effect of weather suitability is weak. The interaction term, however, is significant at the 1 percent level, consistent with the hypothesis. The estimated overall effect of an extra month of malaria transmission is positive, significant, and sizable in magnitude. The interaction term remains highly significant when also accounting for time varying weather conditions in terms of precipitation, temperature and evapotranspiration. The results are also robust to the inclusion of country×year fixed effects.

Table 4.4 exploits within-cell variation overtime by including cell fixed effects in the yearly panel data. This specification accounts for the possible confounder for the pooled OLS specifications that unobservable (or unmeasurable) characteristics of the specific cells are correlated with our measures of malaria risks. In Table 4.4, we report estimates of specifications with cell and year fixed effects at the yearly panel to account for observable and unobservable time-invariant cell-specific characteristics. Since Malaria Ecology is time-invariant, its main effect is subsumed in the cell fixed effects. The results confirm that an increase in the predicted number of months of malaria suitability increases conflict in cells with low to intermediate Malaria Ecology. The overall effect of an extra malaria month is positive in these cells.

The results in Table 4.5 investigate the role of changes in malaria risk for different types of civil violence. Columns (1)-(3) report results for organized forms of civil conflicts related to battle events, whereas Columns (4)-(6) report results for more unorganized forms of civil violence in terms of events related to riots, civil protests, or violence involving civilians. The results reveal that the effect of increased malaria risk is particularly relevant for the unorganized forms of violence, such as riots and violent actions against civilians but only weakly related to battles events. This supports the hypothesis stated in the Introduction that health-related variation may change the opportunity costs (or incentives) for risky and violent behaviors at the individual level, leading to spontaneous outbreaks of violence (and less so to outbreaks of violence that require considerable planning, preparation, and infrastructure, such as battles).

**Monthly Data.** The nature of malaria risk implies that the risk of malaria epidemic is limited in terms of time. By its nature, the Malaria Suitable Month index is designed to predict malaria threats at the month level. Table 4.6, reports the results of the analysis when replicated for exploiting exogenous variation at the monthly level. The variable Malaria Suitable Month indicates whether the climatic preconditions for malaria transmission were met in a particular month in a particular cell. Again, the goal of the analysis is to identify the effect of a malaria-suitable month in cells where the average malaria incidence is intermediate. The results closely resemble those at yearly levels presented before. In particular, the interaction effect is positive, and sizable in magnitude, as is the overall effect of a malaria suitable-month in these cells, see Columns (1)-(3). The specification in Column (4) controls for the average monthly precipitation, temperature and evapotranspiration up to twelve months before the month of measurement. One interesting feature of the month-to-month variation is that it allows to account for time varying cell specificities by including cell×year fixed effects. This effectively allows to control for all relevant determinants of civil conflicts that change from a year to the other (including, e.g. changes in yearly income production) in a given cell. The results, reported in Columns (6) to (8), confirm the results and reveal coefficient estimates that are quantitatively even larger.

**Robustness.** Table D.1 includes alternative interactions between weather shocks, in the form of precipitation, temperature and the Standard Precipitation and Evapotranspiration Index (SPEI) to investigate if the effect in low-intermediate malaria ecology cells is related to other weather shocks than measured by the malaria suitable months indicator. The results reveal no significant alternative interactions. Finally, Table 4.7 controls for spatial autocorrelation in the error term, in the explanatory variables and in the covariates by means of spatial autocorrelation models. The main results remain qualitatively and quantitatively unaffected.

Table 4.1: Non Linear Effect of Malaria Exposure - Cross Section

| Dependent Variable | Fraction of Years with a Conflict Event - ACLED | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Malaria Ecology | 0.012** | 0.007 | 0.008 | 0.026** | 0.031*** | 0.029*** |
| | [0.005] | [0.006] | [0.006] | [0.011] | [0.011] | [0.010] |
| Malaria Ecology square | | | | -0.001 | -0.001*** | -0.001*** |
| | | | | [0.000] | [0.000] | [0.000] |
| Geographic Controls | No | Yes | Yes | No | Yes | Yes |
| Location and Distances | No | Yes | Yes | No | Yes | Yes |
| Natural Resources | No | Yes | Yes | No | Yes | Yes |
| Ethnic Diversity | No | Yes | Yes | No | Yes | Yes |
| Population and Night Lights | No | No | Yes | No | No | Yes |
| Weather Time Varying | No | No | No | No | No | No |
| Observations | 440 | 440 | 440 | 440 | 440 | 440 |
| R-squared | 0.520 | 0.672 | 0.714 | 0.524 | 0.682 | 0.722 |
| Number of Cells | 440 | 440 | 440 | 440 | 440 | 440 |

The dependent variable is a dummy variable taking value 1 if in the cell was registered at least one conflict event. The dependent variable is the fraction of years with conflicts in the period 1998-2011 (ACLED dataset). Malaria Ecology is an index measuring the force and degree of stability in malaria transmission. The ”Geographic Controls” include absolute latitude, mean elevation, average terrain ruggedness, total cell area, total area of the cell occupied by water, average precipitation and average temperature. The “Location and Distances” controls includes the natural logarithm of the distance to the country capital, to the coast, to the country border, to the closest river and to Adis Ababa. The “natural resources” controls include the average land suitability for agriculture, the presence of diamond mines and the presence of petrol fields. The ”Ethnic-Diversity” controls for the number of ethnic groups in the cell (GREG). The “Weather Time-Varying” controls include the average temperature, the average precipitation and the effective rainfall (the Standard Precipitation and Evapotranspiration Index -SPEI) registered in the year. Country-Year fixed effects are a set of country specific year fixed effects. OLS estimates. The unit of observation is a 2.5 x 2.5 degree cell. Standard errors are clustered at the country level. ***, **, * indicate significance at 1-, 5-, and 10-% level, respectively.

Table 4.2: Non Linear Effect of Malaria Exposure - Pooled OLS

| Dependent Variable | Any Conflict Event - ACLED | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Malaria Ecology | 0.031*** | 0.029*** | 0.028*** | 0.029** | 0.027** |
| | [0.011] | [0.010] | [0.010] | [0.011] | [0.010] |
| Malaria Ecology square | -0.001*** | -0.001*** | -0.001*** | -0.001*** | -0.001** |
| | [0.000] | [0.000] | [0.000] | [0.000] | [0.000] |
| Geographic Controls | Yes | Yes | Yes | Yes | Yes |
| Location and Distances | Yes | Yes | Yes | Yes | Yes |
| Natural Resources | Yes | Yes | Yes | Yes | Yes |
| Ethnic Diversity | Yes | Yes | Yes | Yes | Yes |
| Population and Night Lights | No | Yes | Yes | No | Yes |
| Weather Time-Varying | No | No | Yes | Yes | Yes |
| Country FE | Yes | Yes | Yes | No | No |
| Year FE | Yes | Yes | Yes | No | No |
| Country-Year FE | No | No | No | Yes | Yes |
| Observations | 6,118 | 6,118 | 6,118 | 6,118 | 6,118 |
| R-squared | 0.356 | 0.377 | 0.380 | 0.447 | 0.468 |
| Number of Cell | 437 | 437 | 437 | 437 | 437 |

The dependent variable is a dummy variable taking value 1 if in the cell was registered at least one conflict event - ACLED database. Malaria Ecology is an index measuring the force and degree of stability in malaria transmission. The "Geographic Controls" include absolute latitude, mean elevation, average terrain ruggedness, total cell area, total area of the cell occupied by water, average precipitation and average temperature. The "Location and Distances" controls includes the natural logarithm of the distance to the country capital, to the coast, to the country border, to the closest river and to Adis Ababa. The "natural resources" controls include the average land suitability for agriculture, the presence of diamond mines and the presence of petrol fields. The "Ethnic-Diversity" controls for the number of ethnic groups in the cell (GREG). The "Weather Time-Varying" controls include the average temperature, the average precipitation and the effective rainfall (the Standard Precipitation and Evapotranspiration Index -SPEI) registered in the year. Country-Year fixed effects are a set of country specific year fixed effects. OLS estimates. The unit of observation is a 2.5 x 2.5 degree cell. Standard errors are clustered at the country level. ***, **, * indicate significance at 1-, 5-, and 10-% level, respectively.

Table 4.3: The Effect of Malaria Suitable Months - Pooled OLS Yearly Data

| Dependent Variable | Any Conflict Event | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Intermediate ME | 0.137*** | 0.137*** | 0.095*** | 0.069*** | 0.038* | 0.038* | 0.038* |
| | [0.035] | [0.035] | [0.026] | [0.022] | [0.021] | [0.021] | [0.021] |
| Mal. Suitable Months Dem. | -0.004 | -0.018** | -0.012** | -0.011** | -0.011** | -0.009** | -0.008* |
| | [0.005] | [0.007] | [0.005] | [0.005] | [0.005] | [0.004] | [0.004] |
| Inter. ME × MSM Dem. | | 0.022*** | 0.019*** | 0.017*** | 0.018*** | 0.015*** | 0.014*** |
| | | [0.006] | [0.005] | [0.005] | [0.005] | [0.005] | [0.005] |
| Geographic Controls | No | No | No | Yes | Yes | Yes | Yes |
| Location and Distances | No | No | No | Yes | Yes | Yes | Yes |
| Natural Resources | No | No | No | Yes | Yes | Yes | Yes |
| Ethnic Diversity | No | No | No | Yes | Yes | Yes | Yes |
| Population and Night Lights | No | No | No | No | Yes | Yes | Yes |
| Weather Time-Varying | No | No | Yes | Yes | Yes | Yes | Yes |
| Weather Time-Varying Lag | No | No | No | No | No | Yes | Yes |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes | No |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | No |
| Country-Year FE | No | No | No | No | No | No | Yes |
| Observations | 35,784 | 35,784 | 35,784 | 35,784 | 35,784 | 35,784 | 35,784 |
| R-squared | 0.190 | 0.191 | 0.217 | 0.234 | 0.252 | 0.253 | 0.310 |
| Number of Cells | 2,556 | 2,556 | 2,556 | 2,556 | 2,556 | 2,556 | 2,556 |

The dependent variable is a dummy variable taking value 1 if in the cell was registered at least one conflict event (ACLED dataset). Intermediate Malaria Ecology is a dummy variable taking value 1 for cells with an average malaria ecology index larger than 0 and lower than 15. Malaria-Suitable Shock is an index which measures how the number of months predicted as suitable for malaria in the current year differ from the average months of predicted malaria suitability over the whole sample. The "Geographic Controls" include absolute latitude, mean elevation, average terrain ruggedness, total cell area, total area of the cell occupied by water, average precipitation and average temperature. The "Location and Distances" controls includes the natural logarithm of the distance to the country capital, to the coast, to the country border, to the closest river and to Adis Ababa. The "natural resources" controls include the average land suitability for agriculture, the presence of diamond mines and the presence of petrol fields. The "Ethnic-Diversity" controls for the number of ethnic groups in the cell (GREG). The "Weather Time-Varying" controls include the average temperature, the average precipitation and the effective rainfall (the Standard Precipitation and Evapotranspiration Index -SPEI) registered in the year. Country-Year fixed effects are a set of country specific year fixed effects. OLS estimates. The unit of observation is a 1 x 1 degree cell. Standard errors are clustered at the country level. ***, **, * indicate significance at 1-, 5-, and 10-% level, respectively.

TABLE 4.4: The Effect of Malaria Suitable Months - Within Cell Variation Yearly Data

| Dependent Variable | Any Conflict Event - ACLED Yearly Data | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Malaria Suitable Months - MSM | -0.016** | -0.011** | -0.009** | -0.009** | -0.009** |
| | [0.006] | [0.005] | [0.004] | [0.004] | [0.004] |
| Intermediate ME × MSM | 0.021*** | 0.019*** | 0.017*** | 0.017*** | 0.017*** |
| | [0.006] | [0.005] | [0.006] | [0.006] | [0.005] |
| Weather Time-Varying | No | No | Yes | Yes | Yes |
| Weather Time-Varying Lag | No | No | No | Yes | Yes |
| Any Conflict Event Lag | No | No | No | No | Yes |
| Cell FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 35,784 | 35,784 | 35,784 | 35,784 | 35,784 |
| R-squared | 0.007 | 0.008 | 0.011 | 0.011 | 0.020 |
| Number of Cells | 2,556 | 2,556 | 2,556 | 2,556 | 2,556 |

The dependent variable is a dummy variable taking value 1 if in the cell was registered at least one conflict event (ACLED dataset). Intermediate Malaria Ecology is a dummy variable taking value 1 for cells with an average malaria ecology index larger than 0 and lower than 15. Malaria-Suitable Months is an index which predicts the number of malaria suitable months in the current year. The "Weather Time-Varying" controls include the average temperature, the average precipitation and the effective rainfall (the Standard Precipitation and Evapotranspiration Index -SPEI) registered in the year. The unit of observation is a 1 x 1 degree cell. Panel data from 1998 to 2011 at yearly frequencies. ***, **, * indicate significance at 1-, 5-, and 10-% level, respectively.

Table 4.5: Different Types of Conflicts

| Dependent Variable | Civil Conflicts Battles | | | Civil Violence Riots - Protests - Violence Civilians | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Malaria Suitable Months | 0.004 | 0.004 | 0.004 | -0.015*** | -0.015*** | -0.016*** |
| | [0.002] | [0.002] | [0.002] | [0.005] | [0.005] | [0.005] |
| Int.ME × MSM | -0.001 | -0.001 | -0.002 | 0.020*** | 0.018*** | 0.020*** |
| | [0.003] | [0.004] | [0.003] | [0.006] | [0.005] | [0.005] |
| Malaria Suitable Months Lag | | | 0.000 | | | 0.001 |
| | | | [0.003] | | | [0.006] |
| Int.ME × MSM | | | 0.004 | | | -0.009 |
| | | | [0.005] | | | [0.008] |
| Weather Time-Varying | No | Yes | Yes | No | Yes | Yes |
| Weather Time-Varying Lag | No | Yes | Yes | No | Yes | Yes |
| Weather Time-Varying Lag 2 | No | Yes | Yes | No | Yes | Yes |
| Conflict Lag | No | No | Yes | No | No | Yes |
| Cell Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Year Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 35,784 | 33,228 | 33,228 | 35,784 | 33,228 | 33,228 |
| R-squared | 0.004 | 0.006 | 0.007 | 0.006 | 0.007 | 0.008 |
| Number of Cell | 2,556 | 2,556 | 2,556 | 2,556 | 2,556 | 2,556 |

The Dependent variable in From Column (1) to (3) is a dummy variable taking value 1 if in the cell was registered at least one organized conflict event, i.e. battles with either government or rebel groups involved (ACLED dataset). In columns (4) to (6), the dependent variable is a dummy variable taking value 1 if in the cell was registered at least one (unorganized) violent event including riots, protests and violence against civilian. Intermediate Malaria Ecology is a dummy variable taking value 1 for cells with an average malaria ecology index larger than 0 and lower than 15. Malaria-Suitable Months is an index which predicts the number of malaria suitable months in the current year. The variable (b) x (a) is the interaction term between the time-invariant variable of Intermediate Malaria Ecology and the time-varying variable of Malaria-Suitable Months. The Climate controls" comprise average temperature, precipitation and effective rainfall (SPEI index) registered in the year. Standard errors are clustered at the country level. OLS estimates. The unit of observation is a 1 x 1 degree cell. The panel data covers years from 1998 to 2011. ***, **, * indicate significance at 1-, 5-, and 10-% level, respectively.

Table 4.6: Malaria Risk and Conflict Events: within cell variation Monthly Data

| | Any Conflict Event - ACLED Monthly | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Malaria Suitable Month | -0.009*** | -0.005*** | -0.004** | -0.004** | -0.003* | -0.009* | -0.009* | -0.009* |
| | [0.002] | [0.002] | [0.002] | [0.002] | [0.002] | [0.005] | [0.005] | [0.005] |
| ME Interm.× MSM | 0.009*** | 0.009*** | 0.008*** | 0.008*** | 0.007*** | 0.020*** | 0.020*** | 0.020*** |
| | [0.002] | [0.002] | [0.002] | [0.002] | [0.002] | [0.006] | [0.006] | [0.006] |
| Weather | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Weather Lag 1 Month | No | No | Yes | No | No | No | No | No |
| Weather Lags 1-12 Month | No | No | No | Yes | Yes | Yes | Yes | Yes |
| Conflict Lag | No | No | No | No | Yes | No | No | No |
| Cell Fixed Effects | Yes | Yes | Yes | Yes | Yes | No | No | No |
| Year Fixed Effects | Yes | Yes | Yes | Yes | Yes | No | No | No |
| Cell×Year Fixed Effects | No | No | No | No | No | Yes | Yes | Yes |
| Observations | 427,201 | 427,201 | 427,201 | 427,201 | 427,201 | 427,201 | 427,201 | 427,201 |
| R-squared | 0.035 | 0.035 | 0.035 | 0.036 | 0.079 | 0.142 | 0.142 | 0.142 |
| Number of Cell | 2,556 | 2,556 | 2,556 | 2,556 | 2,556 | 2,556 | 2,556 | 2,556 |

The dependent variable is a dummy variable taking value 1 if in the cell was registered at least one conflict event (ACLED dataset) in the month. Malaria-Suitable is an index which predicts whether the months is suitable for malaria to be transmitted. The "Weather" controls include average temperature, precipitation and effective rainfall (SPEI index) registered in the current month, the "Weather Lag 1 month" include the same controls in the previous months, while "Weather Lags 1-12 Month " include average temperature, precipitation and effective rainfall (SPEI index) in each of the 12 months proceeding the current month. Standard errors are clustered at the country level. The table reports OLS estimates. The unit of observation is a 1 x 1 degree cell. The panel includes monthly data from January 1998 to December 2011. ***, **, * indicate significance at 1-, 5-, and 10-% level, respectively.

Table 4.7: Within cell yearly variation, robustness: spatial econometrics

| | Any Conflict Event - ACLED Yearly | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Malaria Suitable Month | -0.009*** | -0.004 | -0.003 | -0.004 |
| | [0.003] | [0.003] | [0.004] | [0.004] |
| Mal. Suitable Months× Inter. ME | 0.015*** | 0.011*** | 0.01** | 0.01** |
| | [0.004] | [0.004] | [0.004] | [0.004] |
| W×Malaria Suitable Months | | -0.018*** | -0.011* | -0.011** |
| | | [0.006] | [0.006] | [0.006] |
| W×MSM×Inter. Mal. Ecology | | 0.014 * | 0,006 | 0,006 |
| | | [0,008] | [0,006] | [0,006] |
| W×Weather Time-Varying | No | Yes | Yes | Yes |
| W×Weather Time-Varying Lag | No | Yes | Yes | Yes |
| W×Conflict Lag | No | No | Yes | Yes |
| Weather Time Varying | Yes | Yes | Yes | Yes |
| Weather Time Varying Lagged | Yes | Yes | Yes | Yes |
| Unorganized Violence Lag | No | No | No | Yes |
| Cell Fixed Effects | Yes | Yes | Yes | Yes |
| Year Fixed Effects | Yes | Yes | Yes | Yes |
| Observations | 3,5784 | 3,5784 | 3,5784 | 3,5784 |
| Number of Cell | 2,556 | 2,556 | 2,556 | 2,556 |

The dependent variable is a dummy variable taking value 1 if in the cell was registered at least one conflict event (ACLED dataset). Intermediate Malaria Ecology is a dummy variable taking value 1 for cells with an average malaria ecology index larger than 0 and lower than 15. Malaria-Suitable Months is an index which predicts the number of malaria suitable months in the current year. The “Weather Time-Varying” controls include the average temperature, the average precipitation and the effective rainfall (the Standard Precipitation and Evapotranspiration Index -SPEI) registered in the year. The “W x Weather Time-Varying” include the time-varying controls’ averages in the eight contiguous cells, while “W x Weather Time-Varying Lag” include the time-varying lagged controls’ averages in the eight contiguous cells. The “W x Conflict Lag” is the dependent variable spatial lag in the eight contiguous cells. In Column (1) and (2), we allow for spatial correlation in the eight contiguous cells’ errors. The unit of observation is a 1 x 1 degree cell. Maximum Likelihood Estimates. Panel data from 1998 to 2011 at yearly frequencies. ***, **, * indicate significance at 1-, 5-, and 10-% level, respectively.

# Appendix A

# Side-Effects of Immunities: Additional Results

TABLE A.1: Malaria and Slavery: Cross-Regional Analysis

| | Colored to Total Population | | Slaves to Total Population | |
|---|---|---|---|---|
| Malaria Ecology | 0.077***<br>(0.024) | | 0.027**<br>(0.013) | |
| Malaria Endemicity | | 0.086***<br>(0.020) | | 0.044***<br>(0.012) |
| Cotton Suitability | Yes | Yes | Yes | Yes |
| Rice Suitability | Yes | Yes | Yes | Yes |
| Sugar Suitability | Yes | Yes | Yes | Yes |
| Tea Suitability | Yes | Yes | Yes | Yes |
| Tobacco Suitability | Yes | Yes | Yes | Yes |
| Distance Coast | Yes | Yes | Yes | Yes |
| Country FE | Yes | Yes | Yes | Yes |
| Observations | 73 | 73 | 73 | 73 |
| R-Squared | 0.807 | 0.830 | 0.688 | 0.721 |

*Note:* Table shows OLS estimates. The unit of observation is the region: 20 states in 1872 Brazil, 13 provinces in 1872 Cuba and 40 states in the 1860 United States. Data are taken from Bergad [2007]. The dependent variable in Column 1 and 2 is the ration of blacks over total population, in Column 3 and 4 is the ratio of slaves over total population. Malaria Ecology is an index measuring the force and stability of malaria transmission, Malaria Endemicity measures the level of malaria parasite rate at the beginning of the 20th century. Robust standard errors in parenthesis. ***, **, * indicate significance at 1-, 5-, and 10-% level, respectively.

Table A.2: Summary Statistics of Cross-County Analysis

| **Variable** | **Mean** | **Std. Dev.** | **Min** | **Max** |
|---|---|---|---|---|
| Blacks to Total Population | 0.239 | 0.209 | 0.002 | 0.772 |
| Slaves to Total Population | 0.224 | 0.205 | 0 | 0.762 |
| Malaria Ecology | 0.043 | 0.029 | 0 | 0.108 |
| Cotton Suitability | 2469.092 | 1746.858 | 0 | 7486.781 |
| Rice Suitability | 1071.552 | 1422.706 | 0 | 5802.241 |
| Sugar Suitability | 140.035 | 508.068 | 0 | 2874.037 |
| Tea Suitability | 2142.92 | 2162.941 | 0 | 7170.96 |
| Tobacco Suitability | 4000.996 | 1380.764 | 0.305 | 7261.313 |
| Distance Sea | 118427.216 | 136128.241 | 0 | 713286.313 |
| Distance River | 11391.349 | 6168.794 | 745.679 | 35114.699 |
| Distance Charleston | 630095.203 | 362693.139 | 708.659 | 1689229.25 |
| N | 285 | | | |

Table A.3: Malaria and Share of Blacks across US States: Robustness

| | **Blacks to Total Population** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Full Sample | Full Sample | No S.Carolina | No N.Carolina | No Virginia | No Delaware | Only 1680+1690 | Balanced Sample |
| Mal. Eco. x Post 1690 | | 4.026*** | 3.679** | 4.710*** | 3.655*** | 4.609*** | 1.847** | 6.104*** |
| | | (0.761) | (1.270) | (0.896) | (0.851) | (0.780) | (0.808) | (1.331) |
| Mal. Endem. x Post 1690 | 0.144*** | | | | | | | |
| | (0.034) | | | | | | | |
| Bootstrap s.e. p-value | 0.000 | 0.004 | 0.014 | 0.028 | 0.032 | 0.000 | 0.136 | 0.006 |
| Whites | No | Yes | No | No | No | No | No | No |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 166 | 166 | 154 | 153 | 151 | 152 | 24 | 109 |
| R-squared | 0.543 | 0.662 | 0.503 | 0.613 | 0.536 | 0.662 | 0.600 | 0.699 |
| Number of States | 12 | 12 | 11 | 11 | 11 | 11 | 12 | 8 |

*Notes:* Table reports panel OLS estimates. The unit of observation is the US state, the panel includes all decades from 1640 to 1770. The dependent variable is the ratio of "colored" people to total population. Malaria Ecology is an index measuring the force and stability of malaria transmission. Malaria Endemicity measures the level of malaria parasite rate at the beginning of the 20th century. The variable Malaria Post 1690 is an indicator variable equals 1 from 1690 onwards, and 0 otherwise. All regressions include decade fixed effects and state fixed effects. The variable Whites measures the size of the white population. As a robustness, we perform baseline estimates on sub-samples: we exclude one Southern state at the time, we include only 1680 and 1690, we exclude all states for which we miss observations for some decades. Standard errors clustered at the state level are reported in parenthesis. Since we only have 12 clusters, we report p-values for the null hypothesis (Malaria Ecology x Post 1690) computed with wild bootstrap standard erros. ***, **, * indicate significance at 1-, 5-, and 10-%level, respectively.

TABLE A.4: Malaria and Blacks across US States

| | **Ln Blacks** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Malaria Ecology x Post 1690 | 35.661***<br>(5.970) | 28.749***<br>(6.494) | 31.806***<br>(8.099) | 55.232***<br>(11.981) | 22.510*<br>(11.272) | 60.897***<br>(12.732) | 12.687<br>(22.776) | 92.857**<br>(29.901) | 108.147***<br>(27.446) |
| Bootstrap s.e. p-value | 0.012 | 0.004 | 0.004 | 0.072 | 0.36 | 0.036 | 0.64 | 0.2 | 0.128 |
| Rice Suitability x Post 1700 | No | Yes | No | No | No | No | No | No | No |
| Rice Suitability x Year FE | No | No | Yes | No | No | No | No | Yes | Yes |
| Tea Suitability x Year FE | No | No | No | Yes | No | No | No | Yes | Yes |
| Tobacco Suitability x Year FE | No | No | No | No | Yes | No | No | Yes | Yes |
| Cotton Suitability x Year FE | No | No | No | No | No | Yes | No | Yes | Yes |
| Average Temperature x Year FE | No | No | No | No | No | No | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| State FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | No |
| Country Time Trend | No | No | No | No | No | No | No | No | Yes |
| Observations | 166 | 166 | 166 | 166 | 166 | 166 | 166 | 166 | 166 |
| R-squared | 0.924 | 0.930 | 0.948 | 0.956 | 0.946 | 0.950 | 0.955 | 0.985 | 0.991 |
| Number of States | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 |

*Notes:* Table reports panel OLS estimates. The unit of observation is the US state, the panel includes all decades from 1640 to 1770. The dependent variable is the natural logarithm of "colored" population. Malaria Endemicity measures the level of malaria parasite rate at the beginning of the 20th century. The variable Malaria Post 1690 is an indicator variable equals 1 from 1690 onwards, and 0 otherwise. All regressions include decade fixed effects and state fixed effects, except for the last Column which includes state specific time trend. Controls variable x Year FE are cross-sectional variables interacted with a full set of decade fixed effects. Standard errors clustered at the state level are reported in parenthesis. Since we only have 12 clusters, we report p-values for the null hypothesis (Malaria Ecology x Post 1690 = 0) computed with wild bootstrap standard erros. ***, **, * indicate significance at 1-, 5-, and 10-%level, respectively.

TABLE A.5: Summary statistics - Malaria and Slave Prices

| **Variable** | **Mean** | **Std. Dev.** | **Min.** | **Max.** | **N** |
|---|---|---|---|---|---|
| Slave Age | 29.509 | 12.493 | 1.8 | 80 | 3675 |
| Slave Age Squared | 1026.83 | 903.420 | 3.24 | 6400 | 3675 |
| Male Slave | 0.689 | 0.463 | 0 | 1 | 3675 |
| Document Year | 1801.943 | 11.727 | 1741 | 1820 | 3675 |
| Ln(Slave Price) | 6.168 | 0.647 | 2.398 | 8.102 | 3675 |
| Malaria Ecology | 19.062 | 3.528 | 5.197 | 30.935 | 3675 |
| Distance Coast | 0.379 | 0.184 | 0.045 | 1.081 | 3675 |
| Ruggedness | 0.345 | 0.234 | 0.141 | 1.194 | 3675 |
| Land Suitability | 0.389 | 0.147 | 0.112 | 0.633 | 3675 |
| Average Temperature | 25.877 | 1.429 | 21.035 | 28.717 | 3675 |
| Absolute Latitude | 7.461 | 5.195 | 0.623 | 19 | 3675 |
| Average Precipitation | 121.291 | 39.316 | 25.701 | 222.689 | 3675 |
| Average Relative Humidity | 66.222 | 12.006 | 33.67 | 80.565 | 3675 |
| Mean Elevation | 339.957 | 127.817 | 31.821 | 1044.682 | 3675 |
| TseTse Fly Suitability | 0.555 | 0.207 | 0.046 | 0.831 | 3675 |
| Distance Atlantic Markets | 4.759 | 0.929 | 3.705 | 10.595 | 3675 |
| Distance to Rivers | 0.105 | 0.025 | 0.061 | 0.159 | 3675 |
| Historical Croplands Cover | 0.034 | 0.024 | 0.003 | 0.097 | 3675 |
| Transition to Agriculture | 3045.116 | 270.743 | 1250 | 3500 | 3675 |
| Ln(Population in 1400) | 13.177 | 1.029 | 11.383 | 15.592 | 3675 |
| State Antiquity | 260.189 | 196.844 | 0 | 637.5 | 3670 |

Table A.6: Cross-correlation table - Malaria and Slave Prices

| Variables | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) | (15) | (16) | (17) | (18) | (19) | (20) | (21) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (1) Slave Age | 1.000 | | | | | | | | | | | | | | | | | | | | |
| (2) Slave Age Sq. | 0.971 | 1.000 | | | | | | | | | | | | | | | | | | | |
| (3) Male Slave | 0.086 | 0.059 | 1.000 | | | | | | | | | | | | | | | | | | |
| (4) Document Year | 0.083 | 0.081 | -0.000 | 1.000 | | | | | | | | | | | | | | | | | |
| (5) Ln(Slave Price) | -0.385 | -0.462 | 0.130 | 0.246 | 1.000 | | | | | | | | | | | | | | | | |
| (6) Malaria Ecology | 0.107 | 0.102 | -0.045 | -0.131 | -0.066 | 1.000 | | | | | | | | | | | | | | | |
| (7) Distance Coast | 0.056 | 0.041 | 0.054 | 0.056 | -0.013 | -0.515 | 1.000 | | | | | | | | | | | | | | |
| (8) Ruggedness | 0.096 | 0.093 | -0.005 | -0.130 | -0.066 | 0.250 | -0.462 | 1.000 | | | | | | | | | | | | | |
| (9) Land Suitability | 0.124 | 0.116 | -0.033 | -0.127 | -0.076 | 0.580 | -0.550 | 0.847 | 1.000 | | | | | | | | | | | | |
| (10) Av. Temperature | 0.126 | 0.121 | 0.025 | -0.147 | -0.079 | 0.602 | 0.018 | -0.096 | -0.035 | 1.000 | | | | | | | | | | | |
| (11) Abs. Latitude | 0.157 | 0.152 | 0.035 | -0.199 | -0.100 | 0.443 | -0.147 | 0.465 | 0.326 | 0.767 | 1.000 | | | | | | | | | | |
| (12) Av. Precipitation | -0.081 | -0.077 | -0.047 | 0.089 | 0.059 | -0.105 | -0.384 | 0.399 | 0.402 | -0.722 | -0.515 | 1.000 | | | | | | | | | |
| (13) Av. Rel. Humidity | -0.162 | -0.155 | -0.049 | 0.187 | 0.107 | -0.364 | -0.131 | -0.237 | -0.103 | -0.858 | -0.923 | 0.698 | 1.000 | | | | | | | | |
| (14) Mean Elevation | 0.017 | 0.007 | 0.006 | 0.047 | -0.001 | -0.506 | 0.295 | 0.388 | 0.280 | -0.764 | -0.394 | 0.516 | 0.407 | 1.000 | | | | | | | |
| (15) TseTse Fly Suit. | -0.079 | -0.075 | -0.045 | 0.070 | 0.061 | -0.008 | -0.339 | 0.273 | 0.407 | -0.700 | -0.592 | 0.830 | 0.733 | 0.495 | 1.000 | | | | | | |
| (16) Distance Atlantic | -0.068 | -0.073 | -0.016 | 0.152 | 0.056 | -0.419 | 0.314 | -0.562 | -0.402 | -0.443 | -0.567 | -0.058 | 0.554 | 0.190 | -0.024 | 1.000 | | | | | |
| (17) Distance Rivers | -0.024 | -0.022 | -0.030 | -0.020 | 0.033 | 0.121 | -0.452 | 0.430 | 0.568 | -0.483 | -0.265 | 0.628 | 0.440 | 0.397 | 0.846 | -0.214 | 1.000 | | | | |
| (18) Croplands Cover | 0.031 | 0.041 | -0.002 | -0.113 | -0.037 | 0.536 | -0.505 | 0.167 | 0.102 | 0.715 | 0.710 | -0.419 | -0.611 | -0.761 | -0.555 | -0.411 | -0.373 | 1.000 | | | |
| (19) Transition Agri. | -0.008 | -0.002 | -0.011 | -0.043 | 0.013 | 0.391 | -0.288 | 0.232 | 0.305 | 0.076 | -0.040 | 0.331 | 0.034 | -0.073 | 0.575 | -0.663 | 0.535 | -0.027 | 1.000 | | |
| (20) Ln(Pop. in 1400) | 0.153 | 0.137 | 0.000 | -0.072 | -0.106 | 0.360 | 0.265 | 0.102 | 0.160 | 0.499 | 0.458 | -0.398 | -0.542 | -0.105 | -0.593 | 0.001 | -0.667 | 0.319 | -0.365 | 1.000 | |
| (21) State Antiquity | 0.088 | 0.085 | 0.023 | -0.065 | -0.087 | 0.205 | 0.300 | -0.156 | -0.222 | 0.659 | 0.489 | -0.691 | -0.674 | -0.390 | -0.830 | -0.079 | -0.872 | 0.495 | -0.347 | 0.834 | 1.000 |

TABLE A.7: Malaria in the Country of Origin and Slave Price

| | **Ln(Slave Price at Sale)** | | | | | | |
|---|---|---|---|---|---|---|---|
| Malaria Ecology | 0.034** | 0.032** | 0.031** | 0.031** | 0.040*** | 0.035** | 0.059** |
| | (0.013) | (0.012) | (0.012) | (0.012) | (0.014) | (0.015) | (0.027) |
| *Wild Bootstrap P-value* | 0.016 | 0.016 | 0.024 | 0.024 | 0.016 | 0.048 | 0.10 |
| Average Temperature | -0.043 | | | | | | -0.023 |
| | (0.053) | | | | | | (0.176) |
| Absolute Latitude | | 0.002 | | | | | 0.032*** |
| | | (0.011) | | | | | (0.011) |
| Average Precipitation | | | 0.000 | | | | -0.001 |
| | | | (0.000) | | | | (0.002) |
| Average Relative Humidity | | | | 0.000 | | | 0.024* |
| | | | | (0.003) | | | (0.013) |
| Mean Elevation | | | | | 0.001 | | 0.001** |
| | | | | | (0.000) | | (0.000) |
| TseTse Fly Suitability | | | | | | -0.056 | -0.446 |
| | | | | | | (0.119) | (0.295) |
| Distance Coast | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Ruggedness | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Land Suitability | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Slave Age and Age Squared | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Male Slave | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Region FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Document Language FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Document Type FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| R-squared | 0.454 | 0.454 | 0.454 | 0.454 | 0.455 | 0.454 | 0.455 |
| Observations | 3675 | 3675 | 3675 | 3675 | 3675 | 3675 | 3675 |

*Notes*: Table reports panel OLS estimates. The unit of observation is the individual slave. The dependent variable is the natural logarithm of the slave price at sale. Malaria Ecology is an index measuring the force and stability of malaria transmission. All regressions also control for age, age squared, sex, language of the sale price document fixed effects, type of document fixed effects, region fixed effects (Upper Guinea, Bight of Benin, Western Africa and Southeastern Africa) and year fixed effects. Standard errors are clustered at the country level (21 clusters). Since we only have 21 clusters, we report p-values for the null hypothesis (Malaria Ecology = 0) computed with wild bootstrap standard erros. ***, **, * indicate significance at 1-, 5-, and 10-%level, respectively.

TABLE A.8: Malaria in the Country of Origin and Slave Price

| | **Ln(Slave Price at Sale)** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Distances | | | Agriculture and Civilization | | | | |
| Malaria Ecology | 0.032**<br>(0.013) | 0.029**<br>(0.011) | 0.032**<br>(0.014) | 0.032**<br>(0.012) | 0.033**<br>(0.013) | 0.038**<br>(0.015) | 0.031**<br>(0.012) | 0.023**<br>(0.009) |
| *Wild Bootstrap P-value* | 0.016 | 0.016 | 0.048 | 0.016 | 0.024 | 0.068 | 0.016 | 0.04 |
| Distance Atlantic Markets | 0.011<br>(0.076) | | | | | | | |
| Distance to Rivers | | 1.351<br>(0.987) | | | | | | |
| Longitude | | | -0.000<br>(0.006) | | | | | |
| Historical Croplands Cover | | | | 0.973<br>(1.064) | | | | |
| % Fertile soil | | | | | -0.001<br>(0.001) | | | |
| Transition to Agriculture$_a$ | | | | | | -0.059<br>(0.056) | | |
| Ln(Population in 1400) | | | | | | | -0.005<br>(0.026) | |
| State Antiquity$_a$ | | | | | | | | -0.118<br>(0.078) |
| Distance Coast | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Ruggedness | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Land Suitability | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Slave Age and Age Squared | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Male Slave | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Region FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Document Language FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Document Type FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| R-squared | 0.454 | 0.454 | 0.454 | 0.454 | 0.454 | 0.454 | 0.454 | 0.454 |
| Observations | 3675 | 3675 | 3675 | 3675 | 3675 | 3675 | 3675 | 3670 |

*Notes*: Table reports panel OLS estimates. The unit of observation is the individual slave. The dependent variable is the natural logarithm of the slave price at sale. Malaria Ecology is an index measuring the force and stability of malaria transmission. All regressions also control for age, age squared, sex, language of the sale price document fixed effects, type of document fixed effects, region fixed effects (Upper Guinea, Bight of Benin, Western Africa and Southeastern Africa) and year fixed effects. Standard errors are clustered at the country level (21 clusters and 19 for specification 8). Since we only have 21 clusters, we report p-values for the null hypothesis (Malaria Ecology = 0) computed with wild bootstrap standard erros. ***, **, * indicate significance at 1-, 5-, and 10-%level, respectively. a) Coefficients and standard errors multiplied by 1000.

Table A.9: Malaria in the Country of Origin and Slave Price

| | **Ln(Slave Price at Sale)** | | | | | |
|---|---|---|---|---|---|---|
| Malaria Endemicity | 0.040* | 0.003 | 0.080*** | 0.083*** | 0.039*** | 0.074 |
| | (0.020) | (0.055) | (0.014) | (0.015) | (0.013) | (0.065) |
| *Wild Bootstrap P-value* | 0.276 | 1.000 | 0.000 | 0.000 | 0.112 | 0.288 |
| Distance Coast | | -0.084 | | | | -0.099 |
| | | (0.103) | | | | (0.183) |
| Ruggedness | | | -0.172*** | | | -0.601** |
| | | | (0.034) | | | (0.233) |
| Land Suitability | | | | -0.188*** | | 0.270 |
| | | | | (0.043) | | (0.337) |
| Mean Elevation | | | | | -0.000*** | 0.000 |
| | | | | | (0.000) | (0.001) |
| Slave Age and Age Squared | Yes | Yes | Yes | Yes | Yes | Yes |
| Male Slave | Yes | Yes | Yes | Yes | Yes | Yes |
| Region FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Document Language FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Document Type FE | Yes | Yes | Yes | Yes | Yes | Yes |
| R-squared | 0.451 | 0.451 | 0.452 | 0.452 | 0.452 | 0.453 |
| Observations | 3675 | 3675 | 3675 | 3675 | 3675 | 3675 |

*Notes*: Table reports panel OLS estimates. The unit of observation is the individual slave. The dependent variable is the natural logarithm of the slave price at sale. Malaria Ecology is an index measuring the force and stability of malaria transmission. All regressions also control for age, age squared, sex, language of the sale price document fixed effects, type of document fixed effects, region fixed effects (Upper Guinea, Bight of Benin, Western Africa and Southeastern Africa) and year fixed effects. Standard errors are clustered at the country level (21 clusters). Since we only have 21 clusters, we report p-values for the null hypothesis (Malaria Ecology = 0) computed with wild bootstrap standard erros. ***, **, * indicate significance at 1-, 5-, and 10-%level, respectively.

TABLE A.10: Malaria and Slave Price: Excluding one Country at the Time

| | **Ln(Slave Price at Sale)** | | | | | | |
|---|---|---|---|---|---|---|---|
| | NO ANGOLA | NO BENIN | NO BURKINA | NO CAMEROON | NO CAF | NO CONGO | NO CDR |
| Malaria Ecology 0.018** | 0.016***<br>(0.006) | 0.018***<br>(0.006) | 0.018***<br>(0.006) | 0.018***<br>(0.006) | 0.018***<br>(0.006) | 0.018***<br>(0.006) | 0.018***<br>(0.006) |
| Observations | 3672 | 3594 | 3675 | 3675 | 3675 | 3674 | 3675 |
| R-squared | 0.452 | 0.455 | 0.453 | 0.453 | 0.453 | 0.453 | 0.453 |
| | NO COTE | NO GABON | NO GHANA | NO GUINEA | NO GBISSAU | NO LIBERIA | NO MALI |
| Malaria Ecology | 0.018***<br>(0.006) | 0.018***<br>(0.006) | 0.018***<br>(0.006) | 0.018***<br>(0.005) | 0.018***<br>(0.006) | 0.018***<br>(0.006) | 0.031***<br>(0.007) |
| Observations | 3675 | 3665 | 3661 | 3266 | 3675 | 3674 | 3525 |
| R-squared | 0.453 | 0.453 | 0.452 | 0.456 | 0.453 | 0.453 | 0.459 |
| | NO MOZAMBIQUE | NO NIGERIA | NO SENEGAL | NO SLEONE | NO TANZANIA | NO TOGO | NO ZIMBABWE |
| Malaria Ecology | 0.019***<br>(0.006) | 0.015**<br>(0.007) | 0.017**<br>(0.007) | 0.018***<br>(0.006) | 0.018***<br>(0.006) | 0.018***<br>(0.006) | 0.018***<br>(0.006) |
| Observations | 3637 | 3215 | 3406 | 3675 | 3675 | 3525 | 3674 |
| R-squared | 0.455 | 0.444 | 0.455 | 0.454 | 0.454 | 0.459 | 0.454 |

*Notes*: Table reports panel OLS estimates. The unit of observation is the individual slave. The dependent variable is the natural logarithm of the slave price at sale. Malaria Ecology is an index measuring the force and stability of malaria transmission. We exclude one country at the time. All regressions also control for age, age squared, sex, language of the sale price document fixed effects, type of document fixed effects, region fixed effects (Upper Guinea, Bight of Benin, Western Africa and Southeastern Africa) and year fixed effects. Standard errors are clustered at the country level (21 clusters). Since we only have 21 clusters, we report p-values for the null hypothesis (Malaria Ecology = 0) computed with wild bootstrap standard erros. ***, **, * indicate significance at 1-, 5-, and 10-%level, respectively.

Table A.11: Malaria and Slave Price in African Regions

| | **Ln(Slave Price at Sale)** | | | | |
|---|---|---|---|---|---|
| | Baseline Sample (1) | Excluding Upper Guinea (2) | Excluding Bight of Benin (3) | Excluding Central Africa (4) | Excluding Southeastern Africa (5) |
| Malaria Ecology | 0.018*** | 0.025*** | 0.012** | 0.018** | 0.019** |
| | (0.006) | (0.007) | (0.005) | (0.007) | (0.006) |
| *Wild Bootstrap P-value* | 0.012 | 0.012 | 0.132 | 0.06 | 0.012 |
| Slave Age and Age Squared | Yes | Yes | Yes | Yes | Yes |
| Male Slave | Yes | Yes | Yes | Yes | Yes |
| Region FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Document Language FE | Yes | Yes | Yes | Yes | Yes |
| Document Type FE | Yes | Yes | Yes | Yes | Yes |
| R-squared | 0.453 | 0.454 | 0.464 | 0.453 | 0.453 |
| Observations | 3675 | 2220 | 2751 | 2421 | 3633 |
| | Only Upper Guinea (1) | Only Bight of Benin (2) | Only Central Africa (3) | Only Southeastern Africa (4) | Upper Guinea and B. of Benin (5) |
| Malaria Ecology | 0.013* | 0.032** | 0.016** | -0.042 | 0.019** |
| | (0.006) | (0.009) | (0.004) | (0.024) | (0.007) |
| *Wild Bootstrap P-value* | 0.240 | 0.036 | 0.112 | 0.292 | 0.06 |
| Slave Age and Age Squared | Yes | Yes | Yes | Yes | Yes |
| Male Slave | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| Document Language FE | Yes | Yes | Yes | Yes | Yes |
| Document Type FE | Yes | Yes | Yes | Yes | Yes |
| R-squared | 0.480 | 0.470 | 0.459 | 0.886 | 0.454 |
| Observations | 1455 | 924 | 1254 | 42 | 2379 |

*Notes*: Table reports panel OLS estimates. The unit of observation is the individual slave. The dependent variable is the natural logarithm of the slave price at sale. Malaria Ecology is an index measuring the force and stability of malaria transmission. In the panel above, we exclude one African region at the time. In the panel below, we report the baseline specification for each African region. All regressions also control for age, age squared, sex, language of the sale price document fixed effects, type of document fixed effects, region fixed effects (Upper Guinea, Bight of Benin, Western Africa and Southeastern Africa) and year fixed effects. Standard errors are clustered at the country level (21 clusters). Since we only have 21 clusters, we report p-values for the null hypothesis (Malaria Ecology = 0) computed with wild bootstrap standard erros. ***, **, * indicate significance at 1-, 5-, and 10-%level, respectively.

TABLE A.12: Malaria and Slave Price - Controlling for Slave Heights

| | **Ln(Slave Price at Sale)** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Malaria Ecology | 0.044**<br>(0.012) | 0.047**<br>(0.011) | 0.027**<br>(0.008) | 0.031***<br>(0.005) | 0.025*<br>(0.009) | 0.029**<br>(0.007) | 0.023**<br>(0.007) | 0.024**<br>(0.006) |
| *Wild Bootstrap P-value* | 0.188 | 0.100 | 0.100 | 0.100 | 0.100 | 0.100 | 0.156 | .048 |
| Ethnic Group Mean Height | | -0.006*<br>(0.003) | | -0.009<br>(0.006) | | -0.012<br>(0.007) | | -0.010<br>(0.004) |
| Region FE | Yes | Yes | No | No | No | No | No | No |
| Distance Coast | No | No | Yes | Yes | No | No | No | No |
| Transition to Agriculture | No | No | No | No | Yes | Yes | No | No |
| Ln(Population in 1400) | No | No | No | No | No | No | Yes | Yes |
| Slave Age and Age Squared | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Male Slave | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Document Language FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Document Type FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| R-squared | 0.537 | 0.538 | 0.535 | 0.536 | 0.535 | 0.535 | 0.535 | 0.536 |
| Observations | 841 | 841 | 841 | 841 | 841 | 841 | 841 | 841 |

*Notes*: Table reports panel OLS estimates. The unit of observation is the individual slave. The dependent variable is the natural logarithm of the slave price at sale. Malaria Ecology is an index measuring the force and stability of malaria transmission. We include the variable slave heights. All regressions also control for age, age squared, sex, language of the sale price document fixed effects, type of document fixed effects, region fixed effects (Upper Guinea, Bight of Benin, Western Africa and Southeastern Africa) and year fixed effects. Standard errors are clustered at the country level (21 clusters). Since we only have 21 clusters, we report p-values for the null hypothesis (Malaria Ecology = 0) computed with wild bootstrap standard erros. ***, **, * indicate significance at 1-, 5-, and 10-%level, respectively.

# Appendix B

# Caste and Technology: Additional Results

## B.1 Appendix

### B.1.1 Data Sources

The analysis is performed relying on two main sources of data: the Standard Cross Cultural Sample and the Ethnographic Atlas. When necessary, additional data were joined to the existing dataset through a mapping system.

#### B.1.1.1 Standard Cross Cultural Sample

**Technological Sophistication: Craft Specialization** SCCS variable $v153$, captures the degree of complexity and specialization in technological crafts. The variable ranges from 1 to 5: 5 is attributed to societies reported to have a variety of craft specialists, including at least smiths, weavers, and potters; 4 to societies reported to have specialized metalworkers or smiths, but to lack loom-weaving and or pottery; 3 wherever loom weaving is practiced but metalwork in absent or unreported; 1 to societies where pottery is made but metalworking and loomweaving are absent or unreported; and 0 to societies where metalworking, loom-weaving and pottery are all absent or unreported. *Source: Murdock, George P., and Caterina Provost. 1971. ETHNOLOGY 12:379-392.* **Technological Sophistication: All Technologies**. Index of technological sophistication of the society along several dimensions: agriculture, artisanal crafts, writing and land transportation. Composed aggregating four index constructed through the recoding of information available in the SCCS. Agriculture index (SCCS variable $v3$ was recoded on a three point scale): 1 for societies where agriculture provide no food contribution; 2 for societies where agriculture contributes up to 50% of the food supply; 3 to societies relying primarily on agriculture. Craft index (constructed based on SCCS variable $v248$ and $v251$): 1 whenever neither metalwork nor pottery is present; 2 to societies which practice pottery (and not metal); 3 to societies which practice both pottery and metal. The presence/absence of pottery and metal was retrieved indirectly through information on the sexual division of labor; i.e. metal/pottery were considered absent if the activity was absent or not reported. Writing index (constructed based on SCCS variable $v149$): 1 if no writing and records are reported; 2 if non-written records and mnemonic devices are reported; 3 if true writing is reported. Land transportation index (constructed based on SCCS variable $v154$): 1 to societies where no technologies for land transportation is reported; 2 for societies which exploits packed/draft animals; 3 for societies with vehicles. The four indexes were aggregated into a single index, thus ranging from 0 to 12. *Source: variable* $v3$, $v149$ *and* $v154$, *Subsistence Economy and Supportive Practices George P. Murdock and Diana O. Morrow. 1970. ETHNOLOGY 9:302-330. Variable* $v248$ *and* $v251$, *Source: Ethnographic Atlas. World Cultures revision by J. Patrick Gray, 1998.*

**Castes** The variable CASTE takes value 1 if it is reported at least one type of caste stratification, and 0 if not. It is based on information retrieved from SCCS variable $v272$, which classify societies in 4 categories based on the forms of endogamous stratification they present, and namely 1) Caste distinction absent or insignificant; 2) Ethnic stratification; 3) Despised Occupational Groups; 4) Complex caste stratification. The baseline variable CASTE include all three types of caste stratification (2, 3 and 5). The variable CASTE1, include only complex caste stratification, and endogamous occupational groups. *Source: Ethnographic Atlas. World Cultures revision by J. Patrick Gray, 1998.* **Latitude**. Latitude of society centroid, measured in degrees. SCCS variable *latitude*. *Source: Standard Cross Cultural Sample (Murdock, George Peter and Douglas R. White. (1969)).* **Mean Temperature**. Mean Temperature of the coldest month. SCCS variable $v188$, *Source: Walter, H., and H. Leith (1964) Klimadiagramm-Weltatlas, Jena: Gustav Fischer.* **Mean Rainfall**. Average Annual Rainfall, coded from 0 to 8 based on annual mm of rainfall. SCCS variable $v929$. Source: World Meteorological Organization (1971) Climatological Norms (CLINO) for Climate and Climate Ship Stations for the period 1931-1960. Geneva. **Agricultural Potential**. Index of agricultural potential of the land: sum of Land Slope, Soils and Climate Scales. SCCS variable $v929$. *Standard Cross Cultural Sample (Murdock, George Peter and Douglas R. White. (1969)).* **Sea Distance** Distance of society centroid to the closest sea coast, measured in thousands of Km. Computed through a geographical mapping system, based on shorelines maps. *Source: Wessel, P. and Smith, W.H.F., 1996. A global, self-consistent, hierarchical, high-resolution shoreline database. J. Geophys. Res., 101(B4): 8741743. http://www.soest.hawaii. edu/wessel/gshhs/* **River Distance**. Distance of society centroid to the closest river, measured in thousands of Km. Computed through a geographical mapping system, based on shorelines maps. *Source: World Water Bodies Dataset. http://www.arcgis.com/home/item.html?id=e750071279bf450cbd510454a8 0f2e63* **Island Dummy**. Dichotomous variables computed measuring the size of the land mass where the society centroid is located. All land masses with an area smaller than Australia were considered islands. *Source: Wessel, P. and Smith, W.H.F., 1996. A global, self-consistent, hierarchical, high-resolution shoreline database. J. Geophys. Res., 101(B4): 8741743. Available at: http://www.soest.hawaii.edu/wessel/gshhs/*

**Ln(Timing of Transition to Agriculture)** Natural logarithm of the timing to the transition to agriculture, from Putterman [2008]. Since data are available only at country a level, each society was matched to the modern country based on geographical location. For some societies, timing of Neolithic transition of the modern country is not available, and namely: Bahamas, Eritrea, Fiji, French Polynesia, Kiribati, Marshall Islands, Micronesia, New Caledonia, Palau, Samoa, Solomon Islands and Vanuatu.

**Distance to Technological Frontier** Aerial distance (Haversine formula), in thousands of km, to the regional technological frontier in 1500 as in Ashraf and Galor [2011a], and namely London and Paris in Europe, Fez and Cairo in Africa, Constantinople and Peking in Asia, and Tenochtitlan and Cuzco in the Americas.

**Migratory Distances** Migratory distance, on a land path, from Adis Ababa. Computed following Ashraf and Galor [2011b]. The distance of the society centroid from Adis Abeba is computed using the Haversine formula. In order to replicate the most likely migration pattern followed by early men, we calculated the distance from Adis Ababa of the path that connect several obligatory intermediate points, and namely: Cairo, Istambul, Phnom Phen, Anadyr and Prince Rupert.

**Agricultural Intensity** Index capturing the fraction of food provided through agriculture, SCCS variable $v3$ *Source: Subsistence Economy and Supportive Practices, George P. Murdock and Diana O. Morrow. 1970. ETHNOLOGY 9:302-330.*

**Population Density** Approximate number of people per square male, coded on a 7 point scale. SCCS variable $v64$ *Source: George P. Murdock and Suzanne F. Wilson. 1972. ETHNOLOGY 11: 254-295.*

**Political Integration** Number of political jurisdictions above the local. SCCS variable $v157$ *Source: Murdock, George P., and Caterina Provost. 1971. ETHNOLOGY 12:379-392.*

**Disease Index** Constructed recombining SCCS variables $v1253 - v1259$. Originally, variables coded on a three point scale: "Absent or not recorded", "Present, no indication of severity", "Present and serious, widespread, or endemic". In the main specification, in order to ensure exogeneity of the index, we recoded the variables on a two point scale with "Present, no indication of severity" and "Present and serious, widespread, or endemic" being equal to 1, and zero otherwise. The diseases include: Leishmaniasis, Trypanosomes, Malaria, Schistosomes, Filariae, Spirochetes and Leprosy. *Source: Low [1988].*

#### B.1.1.2 Ethnographic Atlas

**Technological Sophistication: Artisanal Crafts** The information regarding the presence of the craft was retrieved, indirectly, from variable $V44$, $V45$, $V47$. In order to replicate SCCS variables $v153$, the same index was constructed: the artisanal craft index is equal 1 for societies where no craft is practiced, to 2 for societies where only pottery is practiced, 3 for societies which have weavers (weavers only or weavers and potters) and 4 to societies which have metalworking and eventually one or both other craft specialization (either weaving and/or pottery), 5 to societies with all 3 artisanal crafts. *Source: Ethnographic Atlas. World Cultures revision by J. Patrick Gray, 1998. Available at: http://eclectic.ss.uci.edu/rwhite/ worldcul/ EthnographicAtlasWCRevisedByWorldCultures.sav*

**Craft Specialization** Constructed recoding Ethnographic Variables $V55$, $V56$, $V58$, respectively for metal, pottery and weaving. The variable of CRAFT specialization was set equal to one for each observation of $V55, V56, V58$ having value 3 or 4, and 0 otherwise. For each craft, only societies where the craft was practiced (independently of having it performed as a specialized profession or not) are considered. The information regarding the presence of the craft was retrieved from variable $V44, V45, V47$, which describe the type of sexual division of labor of the society and are indirectly informative regarding the mere presence of the craft. Observations of variables $V44, V45, V47$ having value of 9 were considered to imply that the craft was not practiced. *Source: Ethnographic Atlas. World Cultures revision by J. Patrick Gray, 1998. Available at: as above.*

**Castes** The variable CASTE takes value 1 if it is reported at least one type of caste stratification, and 0 if not. It is based on information retrieved from Ethnographic Atlas variable $V68$, which classify societies in 4 categories based on the forms of endogamous stratification they present, and namely 1) Caste distinction absent or insignificant; 2) Ethnic stratification; 3) Despised Occupational Groups; 4) Complex caste stratification.

The baseline variable CASTE include all three types of caste stratification (2, 3 and 5). The variable CASTE1, include only complex caste stratification, and endogamous occupational groups. *Source: ETHNOGRAPHIC ATLAS. World Cultures revision by J. Patrick Gray, 1998. Available at: as above.*

**Latitude** Latitude of society centroid, measured in degrees. Ethnographic Atlas variable $V106$. *Source: Ethnographic Atlas. World Cultures revision by J. Patrick Gray, 1998. Available at: as above.*

**Mean Temperature** Mean global accumulated temperature (T¿0) in an area defined by a 100 km radius drawn from society centroid. Computed through a geographical mapping system. *Source: Global accumulated temperatures (Tmean¿ 0 C), Land and Plant Nutrition Management Service - AGLL - FAO-UN.*

**Mean Rainfall** Average Annual Precipitation, coded on a scale based on annual mm of rainfall, in an area defined by a 100 km radius drawn from society centroid. Computed through a geographical mapping system. *Source: Climate Research Unit, Univ. of East Anglia. New, M.G., M. Hulme and P.D. Jones, 1999: Representing 20th century space-time climate variability. I: Development of a 1961-1990 mean monthly terrestrial climatology. J. Climate. 12, 829-856. Available at: http://atlas.sage.wisc.edu/*

**Agricultural Suitability** Average suitability index, ranging from 0 to 1, in an area defined by a 100 km radius drawn from society centroid. Computed through a geographical mapping system. *Source: C. Ramankutty, N., J.A. Foley , J. Norman, and K. McSweeney. The global distribution of cultivable lands:current patterns and sensitivity to possible climate change. Submitted to Global Ecology and Biogeography, March 2001. Available at: http://atlas.sage.wisc.edu/*

**Mean Elevation** Mean elevation, in meters, in an area defined by a 100 km radius drawn from society centroid. Computed through a geographical mapping system. National Oceanic and Atmospheric Administration (NOAA) and U.S. National Geophysical Data Center, TerrainBase, release 1.0 (CD-ROM), Boulder, Colo. Available at: http://atlas.sage.wisc.edu/

**Sea Distance** Distance of society centroid to the closest sea coast, measured in thousands of Km. Computed through a geographical mapping system, based on shorelines maps. *Source: Wessel, P. and Smith, W.H.F., 1996. A global, self-consistent, hierarchical, high-resolution shoreline database. J. Geophys. Res., 101(B4): 87411743. http://www.soest.hawaii.edu/wessel/gshhs/*

**River Distance** Distance of society centroid to the closest river, measured in thousands of Km. Computed through a geographical mapping system, based on shorelines maps. *Source: World Water Bodies Dataset.*

**Island Dummy** Dichotomous variables computed measuring the size of the land mass where the society centroid is located. All land masses with an area smaller than Australia were considered islands. *Source: Wessel, P. and Smith, W.H.F., 1996. A global, self-consistent, hierarchical, high-resolution shoreline database. J. Geophys. Res., 101(B4): 8741743. http://www.soest.hawaii.edu/ wessel/gshhs/*

**Ln(Timing of Transition to Agriculture)** Natural logarithm of the timing to the transition to agriculture, from Putterman [2008]. Since data are available only at country a level, each society was matched to the modern country based on geographical location. For some societies, timing of Neolithic transition of the modern country is not available, and namely: Bahamas, Eritrea, Fiji, French Polynesia, Kiribati, Marshall Islands, Micronesia, New Caledonia, Palau, Samoa, Solomon Islands and Vanuatu.

**Distance to Technological Frontier** Aerial distance (Haversine formula), in thousands of km, to the regional technological frontier in 1500 as in Ashraf and Galor [2011a], and namely London and Paris in Europe, Fez and Cairo in Africa, Constantinople and Peking in Asia, and Tenochtitlan and Cuzco in the Americas.

**Migratory Distances** Migratory distance, on a land path, from Adis Ababa. Computed following Ashraf and Galor [2011b]. The distance of the society centroid from Adis Abeba is computed using the Haversine formula. In order to replicate the most likely migration pattern followed by early men, we calculated the distance from Adis Ababa of the path that connect several obligatory intermediate points, and namely: Cairo, Istambul, Phnom Phen, Anadyr and Prince Rupert.

**Agricultural Intensity** Index capturing the fraction of food provided through agriculture, Ethnographic Atlas variable $V28$ *Source: Ethnographic Atls. World Cultures revision by J. Patrick Gray, 1998. Available at: as above.*

**Political Centralization** Number of political jurisdictions above the local. Ethnographic Atlas variable $V33$ *Source: Ethnographic Atlas. World Cultures revision by J. Patrick Gray, 1998. Available at: as above.*

**Social Classes** We recode Ethnographic Atlas variable $V66$ into an index which takes values: 1) for societies with no social classes; 2) societies with a dual stratification into a hereditary aristocracy and commoners, and landlord versus landless societies, and 3) societies presenting wealth distinction based on possession and distribution, and complex stratification correlated with differentiation of occupational statuses.

### B.1.2 Descriptive Statistics

TABLE B.1: Societies with Caste Stratification

| | Afr | Med | Asia | Pac | N.Am | S.Am | Total |
|---|---|---|---|---|---|---|---|
| | | **Standard Cross Cultural Sample** | | | | | |
| Caste | 4 | 13 | 8 | 1 | 1 | 1 | 28 |
| No Caste | 24 | 12 | 24 | 30 | 32 | 31 | 153 |
| | | **Ethnographic Atlas** | | | | | |
| Caste | 57 | 70 | 31 | 2 | 2 | 1 | 163 |
| No Caste | 286 | 47 | 68 | 148 | 270 | 97 | 916 |

*Notes*: The table reports the number of societies in each geographical area which present caste-like endogamous stratification. Societies are grouped into six geographical region: "Afr." stands for Sub-Saharan Africa, "Med." for Circum-Mediterranean, "Asia" for East-Eurasia, "Pac." for Insular Pacific, "N.Am" for North America and "S.Am" for South America.

TABLE B.2: SCCS Summary Statistics

| Variable | Mean | Std. Dev. | Min. | Max. | N |
|---|---|---|---|---|---|
| Technology: Artisan Crafts | 3.091 | 1.413 | 1 | 5 | 186 |
| Technology: All technologies | 7.774 | 2.232 | 4 | 12 | 177 |
| Castes | 0.155 | 0.363 | 0 | 1 | 181 |
| Latitude | 14.656 | 25.162 | -55.5 | 68.7 | 186 |
| Mean Temperature | 12.75 | 14.544 | -28 | 28 | 180 |
| Mean Rainfall | 6.205 | 2.408 | 0 | 8 | 171 |
| Mean Elevation | 448.615 | 673.341 | 0 | 3822 | 182 |
| Agricultural Potential | 16.726 | 3.466 | 4 | 23 | 186 |
| Sea Distance | 0.367 | 0.415 | 0 | 1.579 | |
| River Distance | 0.231 | 0.564 | 0 | 5.068 | 186 |
| Island | 0.22 | 0.416 | 0 | 1 | 186 |
| Ln(Timing of Neolithic Trans.) | 8.237 | 0.554 | 5.991 | 9.259 | 171 |
| Distance to Techn. Frontier | 3.278 | 1.878 | 0.018 | 10.19 | 186 |
| Migratory Distance | 12.335 | 7.729 | 0.254 | 28.065 | 186 |
| Agricultural Intensity | 4.366 | 1.902 | 1 | 6 | 186 |
| Population Density | 3.761 | 1.977 | 1 | 7 | 184 |
| Political Centralization | 2.962 | 1.183 | 1 | 5 | 186 |

### B.1.3 Preliminary Evidence

#### B.1.3.1 Robustness Checks

### B.1.4 Identifying a Causal Relationship

Table B.3: Ethnographic Atlas Summary Statistics

| **Variable** | **Mean** | **Std. Dev.** | **Min.** | **Max.** | **N** |
|---|---|---|---|---|---|
| Technology: Artisan Crafts | 1.147 | 1.511 | 0 | 4 | 743 |
| Specialized Potters | 0.133 | 0.34 | 0 | 1 | 548 |
| Specialized Leather Workers | 0.152 | 0.36 | 0 | 1 | 446 |
| Specialized Weavers | 0.157 | 0.364 | 0 | 1 | 382 |
| Specialized Metal Workers | 0.961 | 0.194 | 0 | 1 | 435 |
| Castes | 0.151 | 0.358 | 0 | 1 | 1079 |
| Latitude | 14.522 | 22.16 | -55 | 78 | 1267 |
| Mean Elevation | 704.515 | 637.652 | 0 | 4732.333 | 1224 |
| Mean Rainfall | 2.488 | 2.168 | 0 | 14 | 1196 |
| Mean Elevation | 7223.961 | 2753.316 | 122 | 10910.563 | 1184 |
| Average Land Suitability | 0.359 | 0.267 | 0 | 0.997 | 1157 |
| Sea Distance | 0.202 | 0.596 | 0 | 7.334 | 1267 |
| River Distance | 0.451 | 0.426 | 0 | 1.848 | 1267 |
| Island | 0.173 | 0.378 | 0 | 1 | 1267 |
| Ln(Timing of Neolithic Trans.) | 8.148 | 0.473 | 5.991 | 9.259 | 1205 |
| Distance to Techn. Frontier | 3.3 | 1.651 | 0 | 10.213 | 1267 |
| Migratory Distance | 10.145 | 7.244 | 0.11 | 28.039 | 1267 |
| Agricultural Intensity | 3.036 | 1.756 | 0 | 6 | 1267 |
| Political Centralization | 1.701 | 1.153 | 0 | 5 | 1267 |
| Social Classes | 1.582 | 0.624 | 1 | 3 | 1085 |
| Disease Index | 3.398 | 2.154 | 0 | 7 | 186 |

Table B.4: Ordered Logit for Castes and Technological Sophistication

| | **Technological Sophistication** | | | |
|---|---|---|---|---|
| | Artisanal Craft | | | |
| | (1) | (2) | (3) | (4) |
| CASTES | 0.932** | 1.037** | 1.136** | 1.591** |
| | (0.457) | (0.509) | (0.537) | (0.650) |
| | **Change in Predicted Probability** | | | |
| No artisan craft | -0.066 | -0.280 | -0.246 | -0.077 |
| At least Pottery | -0.077 | -0.200 | -0.192 | -0.123 |
| At least Weaving | -0.080 | -0.082 | -0.088 | -0.159 |
| At least Metalwork | 0.114 | 0.341 | 0.327 | 0.175 |
| Pottery, Weaving, Metalwork | 0.109 | 0.220 | 0.199 | 0.184 |
| Geography Controls | | Yes | Yes | Yes |
| Location Controls | | | Yes | Yes |
| Ln(Timing of Neolithic Trans.) | | | | Yes |
| Distance to Techn. Frontier | | | | Yes |
| Migratory Distance | | | | Yes |
| Migratory Distance square | | | | Yes |
| Continent FE | Yes | Yes | Yes | Yes |
| Observations | 150 | 150 | 150 | 150 |

*Notes*: The table reports ordered logit estimates. The upper part of the panel reports the coefficient estimates, the lower part of the panel reports changes in the predicted probability of having a certain degree of technological sophistication having castes, compared not to have them. The unit of observation is the Standard Cross Cultural Sample society. The dependent variable in column (1), (2) and (3) is an index of artisanal sophistication which captures the type of specialized artisanal crafts performed in the society and ranges from 1 to 5. The explanatory variable of interest is the presence of castes, $CASTES$, which takes value 1 if at least one endogamous group is present and 0 otherwise. Geography control variables include: latitude, agricultural potential of the land occupied by the society, average temperature of the coldest month, average rainfall and mean elevation. Location controls are: distance of the society centroid from the closest river (in thousand Km), distance of the society centroid from the closest sea coast (in thousand Km) and a dummy variable for islands. Continent dummies for the six geographical areas in the sample, i.e. Sub-Saharan Africa, Circum-Mediterranean, East-Eurasia, Insular Pacific, North America and South America, are included in all specifications. Robust standard errors are reported in parentheses. *** stands for significant at the 1 percent level, ** at the 5 percent level, * at the 10 percent level.

Table B.5: OLS for Castes and Technological Sophistication

| | **Technological Sophistication** | | | | | |
|---|---|---|---|---|---|---|
| | Artisanal Craft | | | All Technologies | | |
| | After 1500 | After 1850 | After 1900 | After 1500 | After 1850 | After 1900 |
| CASTES | 0.920***<br>(0.280) | 1.055***<br>(0.299) | 1.078***<br>(0.326) | 1.154**<br>(0.507) | 1.381**<br>(0.564) | 1.373**<br>(0.555) |
| Geography Controls | | Yes | Yes | | Yes | Yes |
| Location Controls | | | Yes | | | Yes |
| Continent FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 158 | 137 | 94 | 151 | 132 | 89 |
| R-squared | 0.448 | 0.431 | 0.478 | 0.459 | 0.443 | 0.486 |

*Notes*: The table reports OLS estimates. The unit of observation is the Standard Cross Cultural Sample society. The dependent variable in column (1), (2) and (3) is an index of artisanal sophistication which captures the type of specialized artisanal crafts performed in the society and ranges from 1 to 5. In column (4), (5) and (6), the dependent variable is an index of technological sophistication indicating the degree of sophistication of the society along multiple dimensions, i.e. crafts, agriculture, writing and land transportation. The explanatory variable of interest is the presence of castes, $CASTES$, which takes value 1 if at least one endogamous group is present and 0 otherwise. In Column (1) and (4) we include the pinpointing date as a control, and we estimate the baseline equation over the whole sample. In Column (2), (5) we restrict the sample to societies observed after 1850 and in Column (3), (6) to societies observed after 1900. Geography control variables include: latitude, agricultural potential of the land occupied by the society, average temperature of the coldest month, average rainfall and mean elevation. Location controls are: distance of the society centroid from the closest river (in thousand Km), distance of the society centroid from the closest sea coast (in thousand Km) and a dummy variable for islands. Continent dummies for the six geographical areas in the sample, i.e. Sub-Saharan Africa, Circum-Mediterranean, East-Eurasia, Insular Pacific, North America and South America, are included in all specifications. Robust standard errors are reported in parentheses. *** stands for significant at the 1 percent level, ** at the 5 percent level, * at the 10 percent level.

Table B.6: Castes and Technological Specialization

| | **Technological Sophistication** | | | | | |
|---|---|---|---|---|---|---|
| | Artisanal Craft | | | All Technologies | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| CASTES | 0.849*** | 0.807*** | 0.779*** | 0.927** | 0.864** | 0.811** |
| | (0.271) | (0.254) | (0.243) | (0.406) | (0.366) | (0.360) |
| Agricultural Intensity | 0.308*** | 0.210*** | 0.173*** | 0.649*** | 0.519*** | 0.461*** |
| | (0.050) | (0.060) | (0.062) | (0.050) | (0.064) | (0.065) |
| Population Density | | 0.221*** | 0.132** | | 0.311*** | 0.191*** |
| | | (0.063) | (0.060) | | (0.083) | (0.071) |
| Political Centralization | | | 0.270*** | | | 0.432*** |
| | | | (0.082) | | | (0.104) |
| Ln(Timing of Neolithic Trans.) | Yes | Yes | Yes | Yes | Yes | Yes |
| Distance to Techn. Frontier | Yes | Yes | Yes | Yes | Yes | Yes |
| Migratory Distance | Yes | Yes | Yes | Yes | Yes | Yes |
| Migratory Distance square | Yes | Yes | Yes | Yes | Yes | Yes |
| Geography Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Location Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Continent FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 149 | 149 | 149 | 143 | 143 | 143 |
| R-squared | 0.613 | 0.653 | 0.675 | 0.748 | 0.780 | 0.803 |

*Notes*: The table reports OLS estimates. The unit of observation is the Standard Cross Cultural Sample society. The dependent variable in column (1), (2) and (3) is an index of artisanal sophistication which captures the type of specialized artisanal crafts performed in the society and ranges from 1 to 5. In column (4), (5) and (6), the dependent variable is an index of technological sophistication indicating the degree of sophistication of the society along multiple dimensions, i.e. crafts, agriculture, writing and land transportation. The explanatory variable of interest is the presence of castes, $CASTES$, which takes value 1 if at least one endogamous group is present and 0 otherwise. Agricultural Intensity is an index which capture what fraction of food is provided to agriculture. Population density measures the approximate number of people per square male, coded on a 7 point scale. Political Centralization captures the number of political jurisdictions above the local. The variable Ln(Timing of Neolithic Transition) is the natural log of the number of years elapsed since the area where the society is located went through the transition to agriculture. Distance to technological frontier is the distance, in thousands of km, from the closest regional technological frontier. Migratory Distance and Migratory Distance squared are, respectively, the migratory distance (in thousand of km on a land path) and square migratory distance from Adis Abeba. Geography control variables include: latitude, agricultural potential of the land occupied by the society, average temperature of the coldest month, average rainfall and mean elevation. Location controls are: distance of the society centroid from the closest river (in thousand Km), distance of the society centroid from the closest sea coast (in thousand Km) and a dummy variable for islands. Continent dummies for the six geographical areas in the sample, i.e. Sub-Saharan Africa, Circum-Mediterranean, East-Eurasia, Insular Pacific, North America and South America, are included in all specifications. Robust standard errors are reported in parentheses. *** stands for significant at the 1 percent level, ** at the 5 percent level, * at the 10 percent level.

Table B.7: Assessing Bias From Unobservables

| | **Technological Sophistication** | | | | | |
|---|---|---|---|---|---|---|
| | Artisanal Craft | | | All Technologies | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Size of the Bias | | 5.6 | 28.8 | | 2.1 | 3.4 |
| CASTES | 0.806***<br>(0.249) | 0.683***<br>(0.230) | 0.779***<br>(0.243) | 1.050**<br>(0.483) | 0.715**<br>(0.341) | 0.811**<br>(0.360) |
| Agricultural Intensity | | Yes | Yes | | Yes | Yes |
| Population Density | | Yes | Yes | | Yes | Yes |
| Political Centralization | | Yes | Yes | | Yes | Yes |
| Ln(Timing of Neolithic Trans.) | | | Yes | | | Yes |
| Distance to Techn. Frontier | | | Yes | | | Yes |
| Migratory Distance | | | Yes | | | Yes |
| Migratory Distance square | | | Yes | | | Yes |
| Geography Controls | Only Latitude | All | All | Only Latitude | All | All |
| Location Controls | | All | All | | All | All |
| Continent FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 149 | 149 | 149 | 143 | 143 | 143 |
| R-squared | 0.613 | 0.653 | 0.675 | 0.748 | 0.780 | 0.803 |

*Notes*: The upper panel of the table report the computed size of the bias from unobservables inferred from selection on observables. The lower panel of the table reports OLS estimates of the relative regressions. Regression in Column (1) and (3) are our "restricted regressions" and only include latitude and region fixed effects. Regressions in Column (2) and (4) are the first set of "unrestricted regressions" ($R_u$) and include country fixed effects, all geographical controls, all location controls, population density, agricultural intensity and political centralization. Regressions in Column (3) and (6) are the second set of "restricted regressions" ($R_r$) and include all controls from the first restricted regression, plus the Ln(Timing of Neolithic Transition), the distance to technological frontier, Migratory Distance and Migratory Distance squared. The bias from unobservables is calculated as $R_r/(R_u - R_r)$.

Table B.8: Castes and Technological Specialization

| | Technological Sophistication | Craft Specialization | | | |
|---|---|---|---|---|---|
| | | Pottery | Leather | Weaving | Metal |
| | (1) | (2) | (3) | (4) | (5) |
| CASTES | 0.380** | 0.162*** | 0.354*** | 0.111* | 0.019 |
| | (0.156) | (0.050) | (0.102) | (0.056) | (0.016) |
| Agricultural Intensity | Yes | Yes | Yes | Yes | Yes |
| Political Centralization | Yes | Yes | Yes | Yes | Yes |
| Ln(Timing of Neolithic Trans.) | Yes | Yes | Yes | Yes | Yes |
| Distance to Techn. Frontier | Yes | Yes | Yes | Yes | Yes |
| Migratory Distance | Yes | Yes | Yes | Yes | Yes |
| Migratory Distance square | Yes | Yes | Yes | Yes | Yes |
| Geography Controls | Yes | Yes | Yes | Yes | Yes |
| Location Controls | Yes | Yes | Yes | Yes | Yes |
| Continent FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 632 | 480 | 407 | 319 | 366 |
| R-squared | 0.652 | 0.351 | 0.551 | 0.378 | 0.362 |

*Notes*: The table reports OLS estimates. The unit of observation is the Ethnographic Atlas society. The dependent variable in column (1) is an index of artisanal sophistication which captures the type of specialized artisanal crafts performed in the society and ranges from 1 to 5. In column (2), (3), (4) and (5), the dependent variable is a dichotomous variable indicating whether the existing craft is performed by specialized artisans or not. The explanatory variable of interest is the presence of castes, $CASTES$, which takes value 1 if at least one endogamous group is present and 0 otherwise. Geography control variables include: latitude, average agricultural suitability, average annual cumulated temperature above 0 degree, average precipitation and mean elevation, all measured within a 100 km radius of society centroid. Location controls are: distance of the society centroid from the closest river (in thousand Km), distance of the society centroid from the closest sea coast (in thousand Km) and a dummy variable for society located on islands. Continent dummies for the six geographical areas in the sample, i.e. Sub-Saharan Africa, Circum-Mediterranean, East-Eurasia, Insular Pacific, North America and South America, are included in all specifications. Robust standard errors, clustered at the linguistic family level (the total number of linguistic family cluster is 68), are reported in parentheses. *** stands for significant at the 1 percent level, ** at the 5 percent level, * at the 10 percent level.

TABLE B.9: OLS for Castes and Technological Sophistication

| | Technological Sophistication | | | | | |
|---|---|---|---|---|---|---|
| | Artisanal Craft | | | All Technologies | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Disease Index | 0.229***<br>(0.076) | 0.227***<br>(0.077) | 0.148**<br>(0.064) | 0.366***<br>(0.111) | 0.363***<br>(0.112) | 0.248***<br>(0.091) |
| Geography Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Location Controls | | Yes | Yes | | Yes | Yes |
| Population Density | | | Yes | | | Yes |
| Continent FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 159 | 159 | 159 | 143 | 143 | 143 |
| R-squared | 0.433 | 0.446 | 0.574 | 0.413 | 0.414 | 0.622 |

*Notes*: The table reports reduced form OLS estimates. The unit of observation is the Standard Cross Cultural Sample society. The dependent variable in column (1), (2) and (3) is an index of artisanal sophistication which captures the type of specialized artisanal crafts performed in the society and ranges from 1 to 5. In column (4), (5) and (6), the dependent variable is an index of technological sophistication indicating the degree of sophistication of the society along multiple dimensions, i.e. crafts, agriculture, writing and land transportation. The explanatory variable of interest is an index of geographical disease exposure ranging from 0 to 7. Geography controls include: latitude, average temperature of the coldest month and average rainfall. Location controls are: distance of the society centroid from the closest river (in thousand Km), distance of the society centroid from the closest sea coast (in thousand Km) and a dummy variable for islands. Continent dummies for the six geographical areas in the sample, i.e. Sub-Saharan Africa, Circum-Mediterranean, East-Eurasia, Insular Pacific, North America and South America, are included in all specifications. Robust standard errors are reported in parentheses. *** stands for significant at the 1 percent level, ** at the 5 percent level, * at the 10 percent level.

# Appendix C

# Malaria and Diversity: Additional Results

Table C.1: Summary statistics

| Variable | Mean | Std. Dev. | Min. | Max. | N |
|---|---|---|---|---|---|
| Ln(Number of Language) | 0.533 | 0.65 | 0 | 3.932 | 12762 |
| Ln(Number of Language, Africa) | 0.735 | 0.554 | 0 | 2.565 | 2662 |
| Ln(Number of Language, Americas) | 0.427 | 0.465 | 0 | 2.197 | 4637 |
| Malaria Endemicity | 1.31 | 1.477 | 0 | 5 | 15683 |
| Malaria Ecology | 1.809 | 4.943 | 0 | 34.728 | 16522 |
| Average Temperature | 8.081 | 15.203 | -30.331 | 30.386 | 17263 |
| Average Precipitation | 60.697 | 59.003 | 0 | 640.889 | 17272 |
| Mean Suitability | 0.27 | 0.308 | 0 | 0.999 | 16498 |
| Suitability | 0.032 | 0.047 | 0 | 0.409 | 16498 |
| Mean Elevation | 624.722 | 792.77 | -720.643 | 5725.512 | 17213 |
| Variation Elevation | 137.56 | 163.493 | 0 | 1868.89 | 17213 |
| Ruggedness | 85098.69 | 112044.254 | 30.148 | 1016771.313 | 17272 |
| Total Water | 0.513 | 0.877 | 0 | 9.146 | 17272 |
| Total Area | 8.072 | 3.31 | 1.001 | 12.308 | 12762 |
| Number of Country | 1.148 | 0.402 | 1 | 5 | 17272 |
| Within Country | 0.869 | 0.337 | 0 | 1 | 17272 |
| Ln(Migratory Distance) | 9.15 | 0.79 | 4.358 | 10.247 | 17263 |
| Ln(Distance Coast) | 11.746 | 2.055 | 0.357 | 14.546 | 17272 |
| Ln(Distance Border) | 11.035 | 1.83 | -1.409 | 13.854 | 17272 |
| Ln(Distance River) | 12.653 | 1.456 | 3.875 | 15.878 | 17272 |
| Ln(Distance Adis Ababa) | 14.004 | 1.231 | 8.247 | 17.001 | 17170 |
| Absolute Latitude | 40.81 | 22.502 | 0.25 | 83.417 | 17272 |
| Night Lights | 3.604 | 3.169 | 0 | 49.196 | 16385 |
| Ln(Population) | 2.534 | 1.959 | -2.197 | 9.037 | 10297 |

Figure C.1: Languages

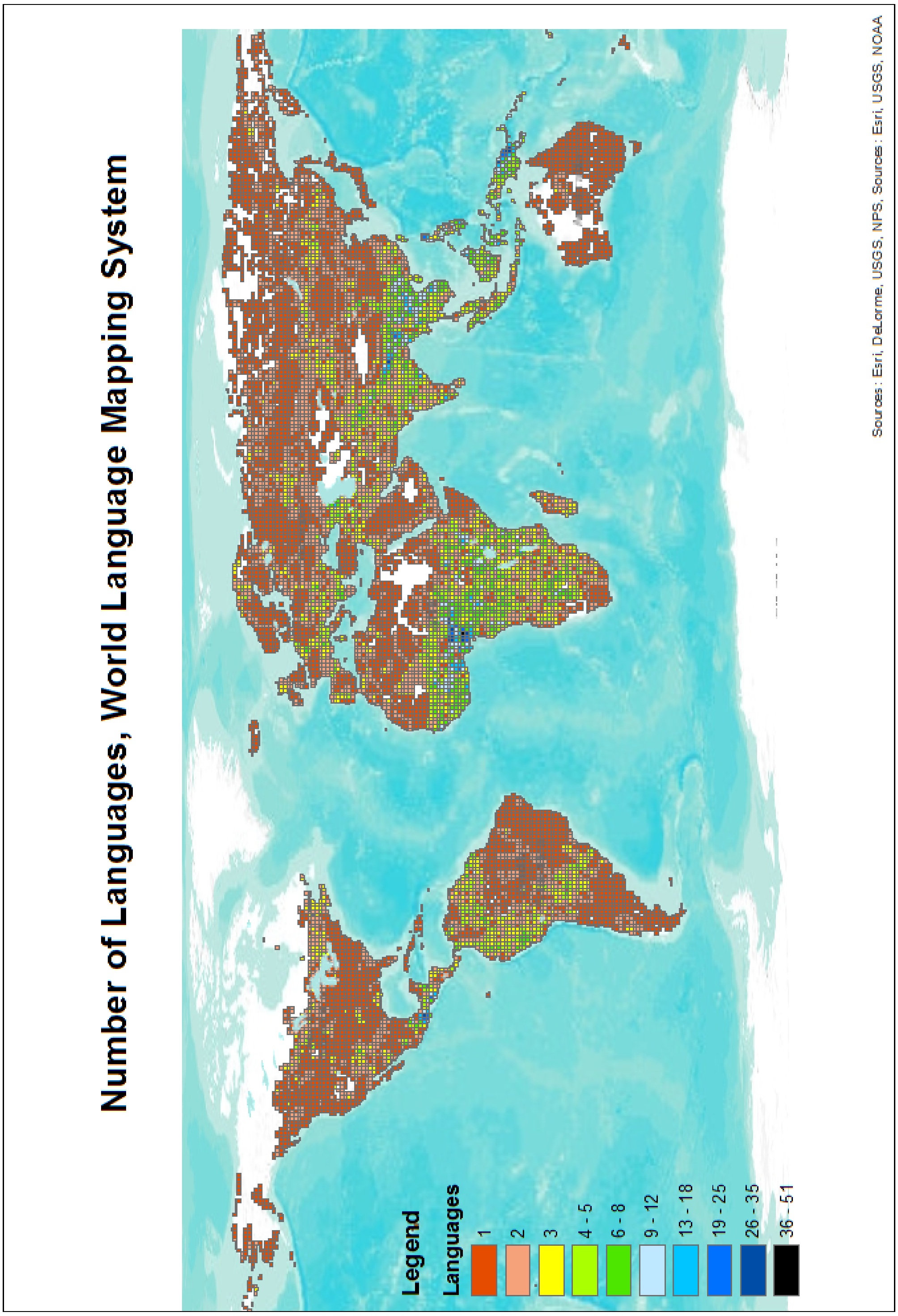


Figure C.2: Malaria Ecology

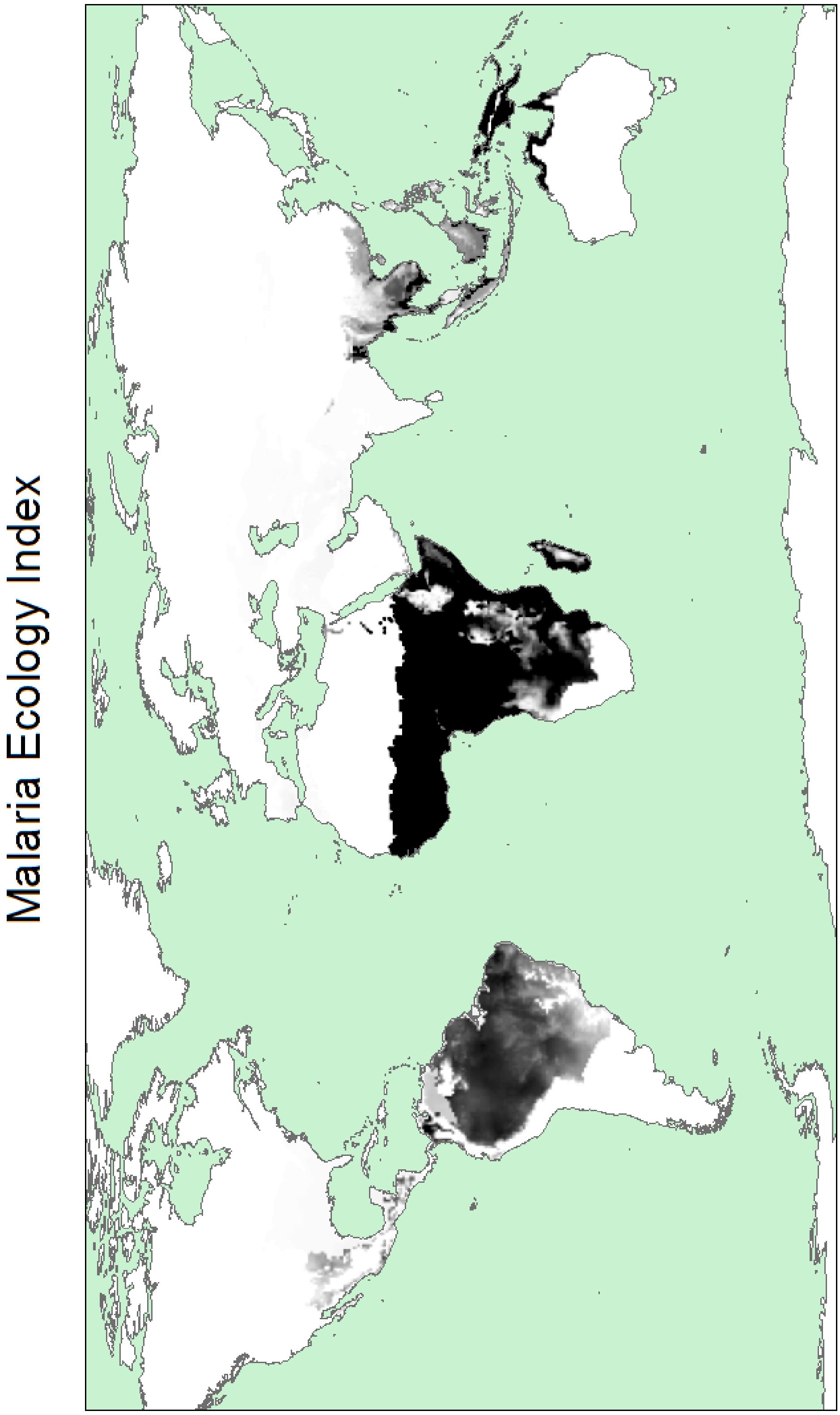


Figure C.3: Malaria Endemicity

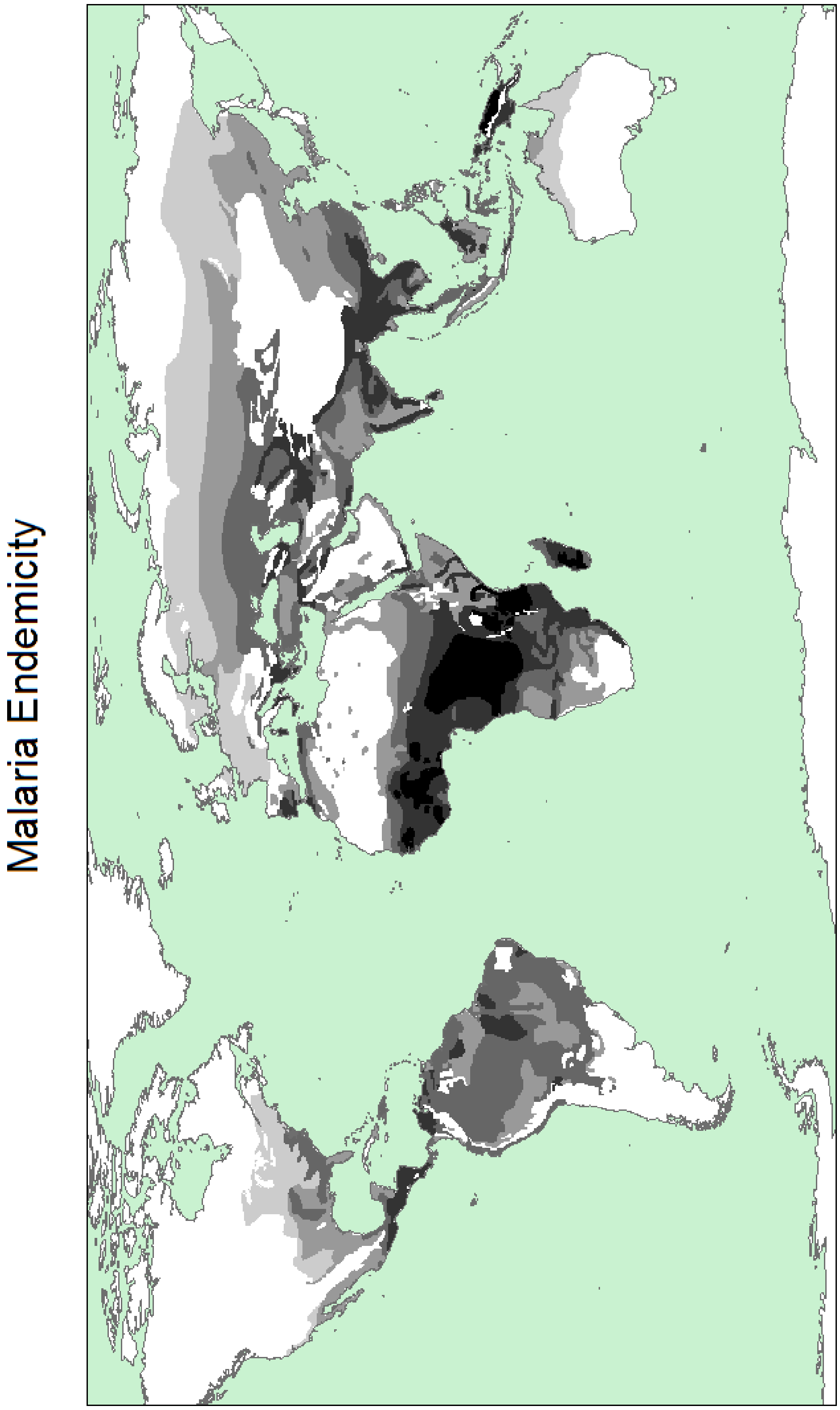
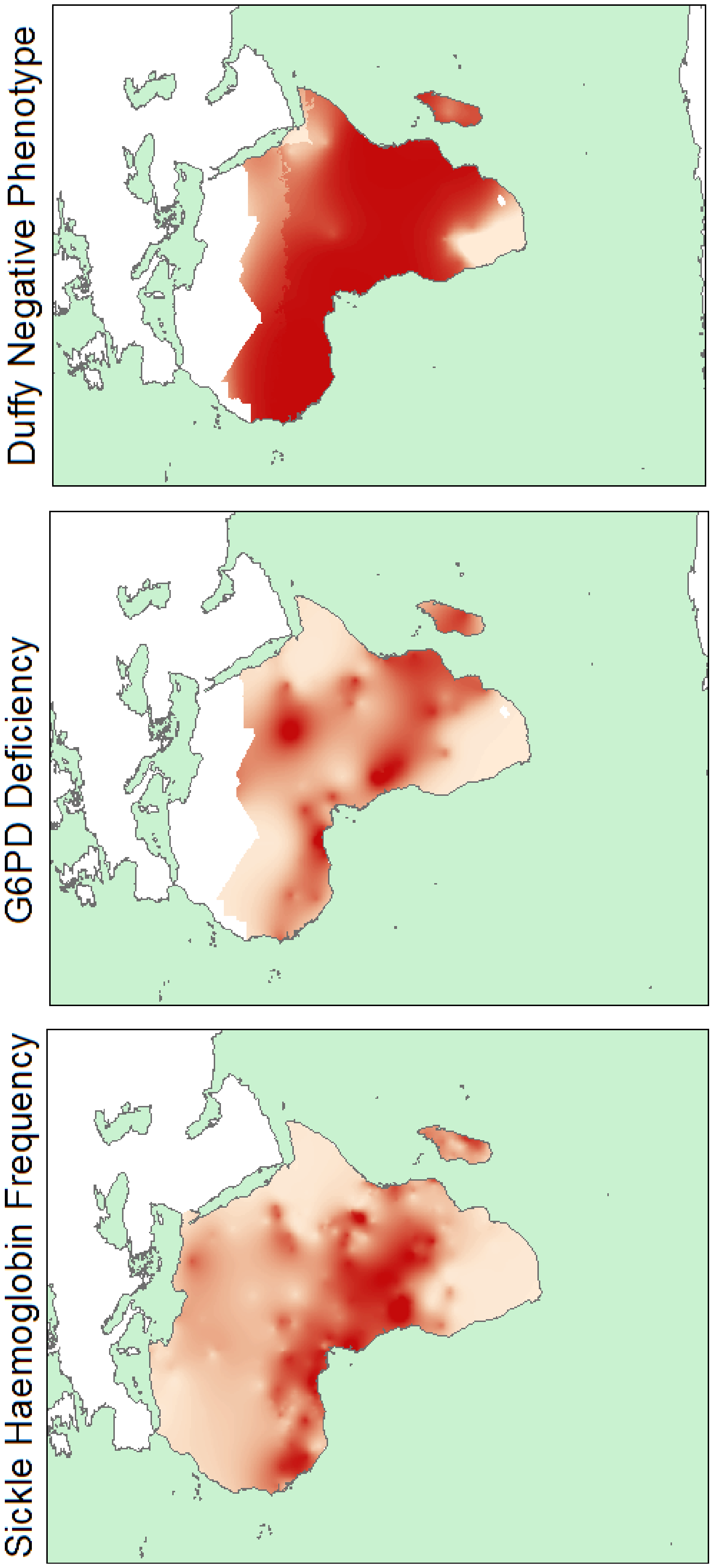


Figure C.4: HbS, G6PD and Duffy



**Data Construction and Sources: Malaria and Diversity**

**Number of Languages, World Language Mapping System**
Number of languages in the 1x1 degree cell. Source: constructed with a (ArcGIS) spatial join between a 1x1 degree grid and the World Language Mapping System language shapefile. We keep only languages that are spoken by more than 1000 people overall. We exclude all intersections that measures less than 10 km squared.

**Number of Ethnic Group, Murdock (1959)**
Number of ethnic groups in the 1x1 degree cell. Source: constructed with a (ArcGIS) spatial join between a 1x1 degree grid and the digitalized map of Murdock Africa map, from Nunn (2009). We exclude all intersections that measures less than 10 km squared.

**Number of Ethnic Group, Murdock (1959)**
Number of ethnic groups in the 1x1 degree cell. Source: constructed with a (ArcGIS) spatial join between a 1x1 degree grid and the digitalized map of Murdock map for North and South America, digitalize by Chiovelli (2013). We exclude all intersections that measures less than 10 km squared.

**Malaria Ecology**
Average Malaria Ecology Index in the 1x1 degree grid cell. Source: average Malaria Ecology is constructed as the 1x1 degree cell average of the Malaria Ecology index from Kiszewski (2004) across grids, computed using ArcGIS with data in EASE GLOBAL GRID projection.

**Malaria Endemicity**
Average Historical Malaria Endemicity in the 1x1 degree grid cell. Source: average Historical Malaria Endemicity is constructed as the 1x1 degree cell average of the Malaria Endemicity level, devised by Lysenko (1968) and digitalized by Hay 2004, computed using ArcGIS with data in EASE GLOBAL GRID projection.

**Average Temperature**
Mean annual 1x1 degree cell temperature (baseline period 1961-1990). Source: average temperature is constructed as the 1x1 degree cell of the mean annual temperature across raster grids, computed using ArcGIS with data in EASE GLOBAL GRID projection, from FAO/IIASA, 2011-2012. Global Agro-ecological Zones (GAEZ v3.0). FAO Rome, Italy and IIASA, Laxenburg, Austria.

**Average Precipitation**
Average 1x1 degree cell monthly precipitation mm/month (baseline period 1961-1990). Source: average monthly precipitation is constructed as the 1x1 degree cell average of the mean monthly precipitation across 10 minute grids, computed using ArcGIS with data in EASE GLOBAL GRID projection, with CRU CL 2.0 data from New (2002).

**Land Suitability**
Average land suitability in the 1x1 degree cell . Source: average Land Suitability is constructed as the 1x1 degree cell average of the land suitability index from Ramankutty (2002) across 0.5 degree raster grids, computed using ArcGIS with data in EASE GLOBAL GRID projection.

**Variation in Land Suitability**
Standard deviation of land suitability in the 1x1 degree cell. Source: standard deviation for Land Suitability is constructed as the 1x1 degree cell standard deviation of the land suitability index from Ramankutty (2002) across 0.5 degree raster grids, computed using ArcGIS with data in EASE GLOBAL GRID projection.

**Mean Elevation**

**Data Construction and Sources: Malaria and Diversity**

Average 1x1 degree cell elevation. Source: mean elevation is constructed as the 1x1 degree cell average of elevation across grids, computed using ArcGIS with data in Africa EASE GLOBAL GRID projection, with data from National Oceanic and Atmospheric Administration (NOAA) and U.S. National Geophysical Data Center, TerrainBase, release 1.0 (CD-ROM), Boulder, Colo.

**Variation in Elevation**

Standard deviation of elevation in the 1x1 degree cell. Source: constructed as the 1x1 degree cell standard deviation of elevation across grids, computed using ArcGIS with data in Africa EASE GLOBAL GRID projection, with data from National Oceanic and Atmospheric Administration (NOAA) and U.S. National Geophysical Data Center, TerrainBase, release 1.0 (CD-ROM), Boulder, Colo.

**Ruggedness**

Average 1x1 degree cell ruggedness (Terrain Ruggedness Index, 100 m). Source: mean ruggedness is constructed as the 1x1 degree cell average of elevation across grids, computed using ArcGIS with data in EASE GLOBAL GRID projection, with data from Terrain Ruggedness Index originally devised by Riley, DeGloria, and Elliot (1999), obtained through http://diegopuga.org/data/rugged/grid.

**Total Water Area**

Total area occupied by water within the 1x1 degree cell. Source: constructed with ArcGIS by intersecting the 1x1 degree cell grid and the Digital Chart of the World inwater shapefile, by intersecting the 1x1 degree cell grid and the Digital Chart of the World oceans and sea shapefile. We sum up total in-cell water area and the areas of the cell occupied by seas and oceans, areas computed with data in EASE GLOBAL GRID projection.

**Total Area**

Total area of the 1x1 degree cell. Source: constructed with ArcGIS by intersecting the 1x1 degree cell grid and the World Language Mapping System shapefile from the Digital Chart of the World. We exclude cell parts not covered by World Language Mapping System data (and by the Africa Murdock Map, and the Murdock map for North and South America), areas computed with data in EASE GLOBAL GRID projection.

**Number of Countries**

Total number of countries in the 1x1 degree cell. Source: constructed with ArcGIS by intersecting the 1x1 degree cell grid and the country boundaries shapefile from the Digital Chart of the World, areas computed with data in EASE GLOBAL GRID projection.

**Within Country**

Dummy variable taking value one if the 1x1 degree cell belong to one single country, 0 otherwise. Source: constructed with ArcGIS by intersecting the 1x1 degree cell grid and the country boundaries shapefile from the Digital Chart of the World.

**Ln Migratory Distance**

Migratory distance, on a land path, from Adis Ababa. Source: computed following Ashraf and Galor [2011b]. The distance of the centroid of 1x1 degree cell from from Adis Abeba is computed using the Haversine formula. In order to replicate the most likely migration pattern followed by early men, we calculated the distance from Adis Ababa of the path that connect several obligatory intermediate points, and namely: Cairo, Istambul, Phnom Phen, Anadyr and Prince Rupert.

**Ln Distance Coast**

Distance to closest coast. Source: constructed with ArcGIS using the digital Chart of the World coastline shapefile.

**Data Construction and Sources: Malaria and Diversity**

**Ln Distance Border**
Distance to closest border. Source: constructed with ArcGIS using the digital Chart of the World boundaries shapefile.

**Ln Distance Capital**
Distance to the capital of the country where lies the centroid of the 1x1 degree cell. Source: constructed with ArcGIS using the digital World Capital shapefile.

**Ln Distance River**
Distance to closest river. Source: constructed with ArcGIS using Major Rivers World selected p3w shapefile, retrived from www.naturalearth.com.

**Absolute Latitude**
Absolute latitudinal distance from the equator in decimal degrees of the 1x1 degree cell. Source: computed using ArcGIS with data in WSG1984.

**Night Lights**
Average population in 1x1 degree cell. Source: constructed as the 1x1 degree cell average of population across raster grids, computed using ArcGIS with data in EASE GLOBAL GRID projection, with data from the Center for International Earth Science Information Network - CIESIN - Columbia University, United Nations Food and Agriculture Programme - FAO, and Centro Internacional de Agricultura Tropical - CIAT. 2005. Gridded Population of the World, Version 3 (GPWv3): Population Count Grid. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://sedac.ciesin.columbia.edu/data/set/gpw-v3-population-count.

**Ln Population Density**
Average night lights intesity in 1x1 degree cell. Source: constructed as the 1x1 degree cell average of night lights intesity grids, computed using ArcGIS with data in EASE GLOBAL GRID projection, with data from NOOA National Geophysical Data Centre for the year 2000.

**Data Construction and Sources: Malaria Exposure and and endogamous Marriages**

**Intra-Ethnic Marriages**
Variable equals to one for a marriage between two individuals belonging to the same ethnic group. Zero otherwise. Source: Demographic and Health Survey (DHS). We create this variable starting from ethnic variables ($v131$ for female and $mv131$ for male for most of the waves) in the DHS. We validate the ethnic groups using Ethnologue.

**Malaria Ecology**
Average Malaria Ecology Index in the 10km radius around the coordinates of each cluster. Source: Malaria Ecology index from Kiszewski (2004).

**Malaria Endemicity**
Average Historical Malaria Endemicity in the 10km radius around the coordinates of each cluster. Source: devised by Lysenko (1968) and digitalized by Hay 2004.

**Average Temperature**
Mean annual temperature in the 10km radius around the coordinates of each cluster (baseline period 1961-1990). Source: FAO/IIASA, 2011-2012. Global Agro-ecological Zones (GAEZ v3.0). FAO Rome, Italy and IIASA, Laxenburg, Austria.

**Average Precipitation**
Average monthly precipitation mm/month in the 10km radius around the coordinates of each cluster (baseline period 1961-1990). Source: CRU CL 2.0 data from New (2002).

**Land Suitability**
Average land suitability in the 10km radius around the coordinates of each cluster. Source: Ramankutty (2002).

**Mean Elevation**
Average elevation in the 10km radius around the coordinates of each cluster. Source: National Oceanic and Atmospheric Administration (NOAA) and U.S. National Geophysical Data Center, TerrainBase, release 1.0 (CD-ROM), Boulder, Colo.

**Variation in Elevation**
Standard deviation of elevation in the 10km radius around the coordinates of each cluster. Source: National Oceanic and Atmospheric Administration (NOAA) and U.S. National Geophysical Data Center, TerrainBase, release 1.0 (CD-ROM), Boulder, Colo.

**Urban Residence Female**
Dummy variable taking value one the individual is living in urban areas. Zero otherwise. Source: Demographic and Health Survey (DHS) ($v025$ for female and $mv025$ for male).

**Education**
Highest year of education of the respondent individual. Source: Demographic and Health Survey (DHS) ($v133$ for female and $mv133$ for male).

**Age**
Age of the respondent individual. Source: Demographic and Health Survey (DHS) ($vv012$ for female and $mv012$ for male).

**Population in the Relevant Ethnic Aggregation**
Absolute number of DHS survey respondent belonging to any given ethnolinguistic family in the region. Source: Demographic and Health Survey (DHS) ($v012$ for female and $mv012$ for male).

# Appendix D

# Malaria Risk and Civil Violence: Additional Results

TABLE D.1: Robustness: Alternative Interactions Yearly Data

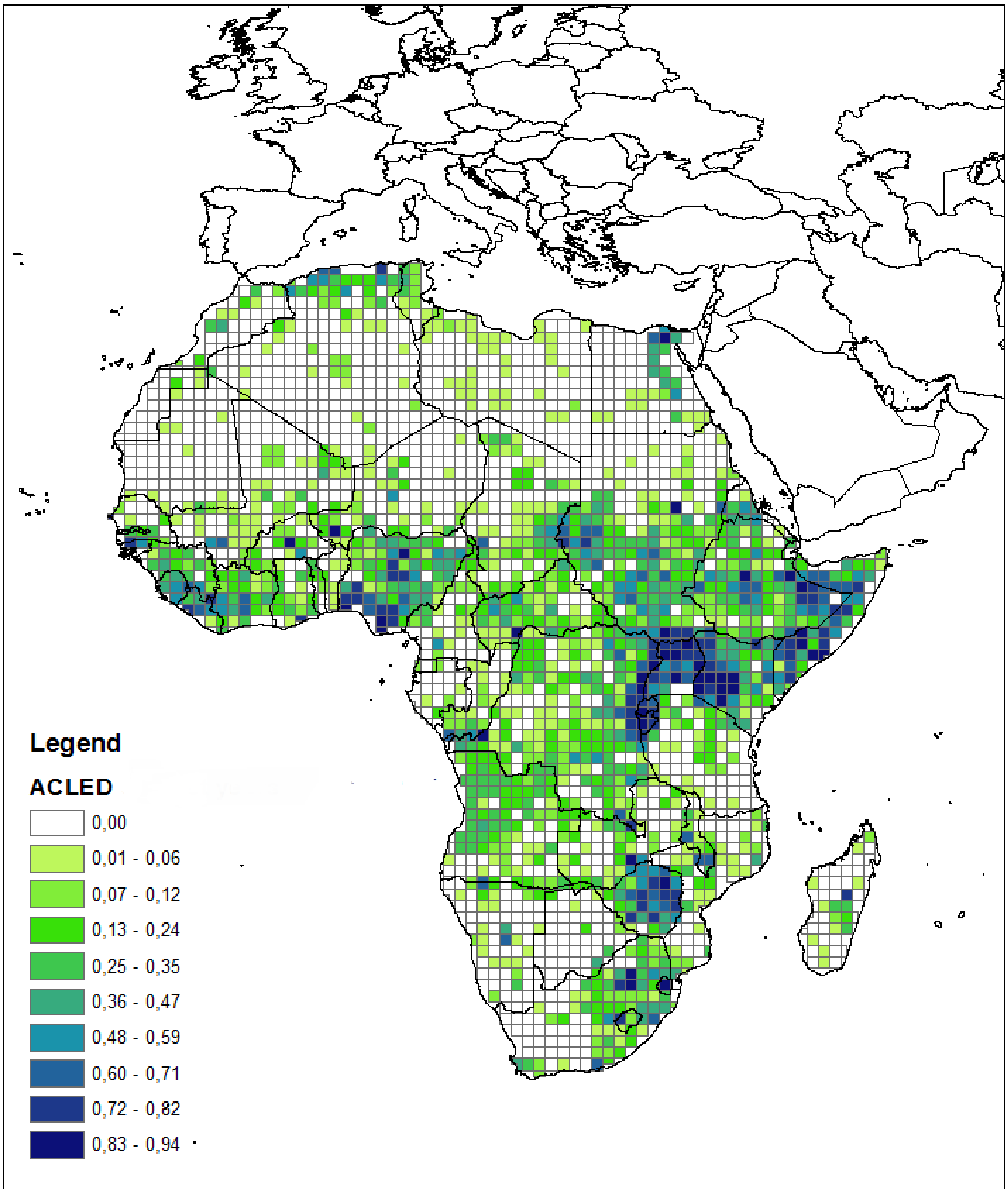
| | Any Conflict Event - ACLED | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Pooled OLS | | | Cell Fix Effects OLS | | | Monthly Cell Fix Effects OLS | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (4) | (5) | (6) |
| Inter. ME | 0.038* | 0.039* | 0.042** | | | | | | |
| | [0.021] | [0.021] | [0.021] | | | | | | |
| MSM D. | -0.008* | -0.008* | -0.009* | | | | | | |
| | [0.004] | [0.005] | [0.005] | | | | | | |
| Int. ME×MSM D. | 0.014** | 0.015** | 0.015** | | | | | | |
| | [0.005] | [0.005] | [0.006] | | | | | | |
| Malaria SM | | | | -0.008* | -0.009* | -0.010** | -0.003 | -0.003* | -0.003* |
| | | | | [0.004] | [0.005] | [0.005] | [0.002] | [0.002] | [0.002] |
| Int. ME×MSM | | | | 0.015*** | 0.016*** | 0.017*** | 0.007*** | 0.007*** | 0.007*** |
| | | | | [0.005] | [0.005] | [0.006] | [0.002] | [0.002] | [0.002] |
| Int. ME ×: | ×Prec.D | ×Tem.D. | ×SPEI D. | ×Prec. | ×Temp. | ×SPEI | ×Prec. | ×Temp. | ×SPEI |
| | 0.000 | -0.015 | 0.013 | 0.000 | -0.010 | -0.010 | 0.000 | -0.000 | -0.001 |
| | [0.001] | [0.023] | [0.014] | [0.001] | [0.020] | [0.011] | [0.000] | [0.001] | [0.001] |
| Geographic C. | Yes | Yes | Yes | No | No | No | No | No | No |
| Location/Dist. | Yes | Yes | Yes | No | No | No | No | No | No |
| Natural Res. | Yes | Yes | Yes | No | No | No | No | No | No |
| Ethnic Div. | Yes | Yes | Yes | No | No | No | No | No | No |
| Pop. and Lights | Yes | Yes | Yes | No | No | No | No | No | No |
| Weather | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Weather Lag | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Conflict Lag | No | No | No | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 35,784 | 35,784 | 35,784 | 35,784 | 35,784 | 35,784 | 427,201 | 427,201 | 427,201 |
| R-squared | 0.253 | 0.253 | 0.253 | 0.020 | 0.020 | 0.020 | 0.079 | 0.079 | 0.079 |
| Number of Cells | 2,556 | 2,556 | 2,556 | 2,556 | 2,556 | 2,556 | 2,556 | 2,556 | 2,556 |

The dependent variable is a dummy variable taking value 1 if in the cell was registered at least one conflict event (ACLED dataset). Intermediate Malaria Ecology is a dummy variable taking value 1 for cells with an average malaria ecology index larger than 0 and lower than 15. Malaria-Suitable Shock is an index which measures how the number of months predicted as suitable for malaria in the current year differ from the average months of predicted malaria suitability over the whole sample. The "Geographic Controls" include absolute latitude, mean elevation, average terrain ruggedness, total cell area, total area of the cell occupied by water, average precipitation and average temperature. The "Location and Distances" controls includes the natural logarithm of the distance to the country capital, to the coast, to the country border, to the closest river and to Adis Ababa. The "natural resources" controls include the average land suitability for agriculture, the presence of diamond mines and the presence of petrol fields. The "Ethnic-Diversity" controls for the number of ethnic groups in the cell (GREG). The "Weather Time-Varying" controls include the average temperature, the average precipitation and the effective rainfall (the Standard Precipitation and Evapotranspiration Index -SPEI) registered in the year. Country-Year fixed effects are a set of country specific year fixed effects. Standard errors are clustered at the country level. OLS estimates. The unit of observation is a 1 x 1 degree cell. ***, **, * indicate significance at 1-, 5-, and 10-% level, respectively.

Table D.2: Data Sources and Description of Main Variables of Interest

| **Variable Description and Data Sources** |
|---|
| **Malaria Exposure:** |
| *Malaria Ecology.* Index measuring the force and stability of malaria transmission based on biologic characteristics of diverse vector mosquitoes interact with climate. Data source: Kiszewski2004. |
| *Intermediate Malaria Ecology.* Based on the Malaria Ecology index, we constructed a dummy variable - Malaria Intermediate - which takes value one in cells with low to intermediate force and stability of transmission. This variables takes value 0 in cells where malaria cannot be transmitted, either because *Anopheles* vectors are not present or because climatic conditions for transmission are absent, and in cells with with very high Malaria Ecology (larger than 15). Data source: Kiszewski2004. |
| **Malaria Risk predicted from Weather Variables:** |
| *Malaria Suitable Month.* Index predicting whether the month is suitable for malaria transmission, based on an algorithm requiring determined threshold of temperature and precipitation in up to 12 months before. The algorithm was computed with temperature and precipitation data from ECMWF ERA-Interim dataset ECMWF ERA-Interim dataset dee2011era. |
| *Malaria Suitable Months.* Index predicting the number of malaria-suitable months for the year, based on the Malaria Suitable Month described above. |
| *Malaria Suitable Months Demeaned.* Index predicting the difference between the malaria-suitable months for the year and the sample average, based on the Malaria Suitable Month described above. |
| **Measures of Civil Violence:** |
| *Baseline: Any Conflict.* Dummy variable taking value one if in the cell there was at least once conflict event within the year (within the month, for the monthly panel analysis). Conflict Events include the categories: Battle-Government regains territory, Battle-No change of territory, Battle-Rebels overtake territory, Headquarters or base established, Non-violent activity by a conflict actor, Non-violent transfer of territory, Riots/Protests and Violence against civilians. Source: ACLED Version 2 (1997-2011), ACLED - Armed Conflict Location and Event Data Project. |
| *Organized Events.* Dummy variable taking value one if in the cell there was at least once organized conflict event within the year (within the month, for the monthly panel analysis). Organized conflict Events include the categories: Battle-Government regains territory, Battle-No change of territory, Battle-Rebels overtake territory, Headquarters or base established, Non-violent activity by a conflict actor, Non-violent transfer of territory. Source: ACLED Version 2 (1997-2011) - Armed Conflict Location and Event Data Project. |
| *Unorganized Events.* Dummy variable taking value one if in the cell there was at least once unorganized conflict event within the year (within the month, for the monthly panel analysis). Organized conflict Events include the categories: Riots/Protests and Violence against civilians. Source: ACLED Version 2 (1997-2011), ACLED - Armed Conflict Location and Event Data Project. |
| **Cell Specific Controls:** |
| *Average Temperature* Mean annual cell temperature (baseline period 1961-1990). Source: FAO/IIASA, 2011-2012. Global Agro-ecological Zones (GAEZ v3.0). FAO Rome, Italy and IIASA, Laxenburg, Austria. |
| *Average Precipitation.* Average cell monthly precipitation mm/month (baseline period 1961-1990). Source: CRU CL 2.0 data from New2002. |
| *Land Suitability.* Average land suitability in the cell. Source: Ramankutty2002. |
| *Mean Elevation.* Average cell elevation. Source: National Oceanic and Atmospheric Administration (NOAA) and U.S. National Geophysical Data Center, TerrainBase, release 1.0 (CD-ROM), Boulder, Colo. |
| *Total Water Area.* A Total area occupied by water in the cell (seas, oceans, lakes and rivers). Source: constructed with Digital Chart of the World inwater shapefile nd the Digital Chart of the World oceans and sea shapefile. |
| *Cell Area.* Natural logarithm of the cell area. |
| *Ln Distance Coast* Natural logarithm of the average cell distance to closest coast. Source: constructed with coastline shapefile from Global Self-consistent Hierarchical High-resolution Geography Version 2.2.2 January 1, 2013. |
| *Ln Distance Border* Natural logarithm of the average cell distance to country capital. Source: constructed with the World Capital shapefile. |
| *Ln Distance Capital* Natural logarithm of the average cell distance to closest river. Source: constructed with ArcGIS using $Major_{R}ivers_{W}orld_{s}elected_{p}3w$ shapefile (from www.naturalearth.com). |
| *Ln Distance Adis Ababa* Natural logarithm of the geodesic distance to Adis Ababa. |
| *Absolute Latitude* Absolute latitudinal distance of the centroid of the cell. |
| *Population* Average population in the cell in year 1995. Source: constructed as the mean population across 2.5 arc-minutes grid. Data from he Center for International Earth Science Information Network - CIESIN - Columbia University, United Nations Food and Agriculture Programme - FAO, and Centro Internacional de Agricultura Tropical - CIAT. 2005. Gridded Population of the World, Version 3 (GPWv3): Population Count Grid. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://sedac.ciesin.columbia.edu/data/set/gpw-v3-population-count. |
| *Night Lights* Average night lights intesity in the cell. Source: constructed with data from NOOA National Geophysical Data Centre for the year 2000. |
| **Weather Time-Varying Controls:** |
| *Average Temperature.* Average monthly/yearly temperature in the cell. Source: data from ECMWF ERA-Interim dataset ECMWF ERA-Interim dataset dee2011era. |
| *Average Precipitation.* Total monthly/average yearly precipitation in the cell. Source: data from ECMWF ERA-Interim dataset ECMWF ERA-Interim dataset dee2011era. |
| *SPEI.* Average standardised precipitation-evapotranspiration index in the cell, normalized at 4 months. Source: SPEIbase v2.2 from http://digital.csic.es/handle/10261/48169. |

Figure D.1: Fractions of Years with at least one conflict event - ACLED



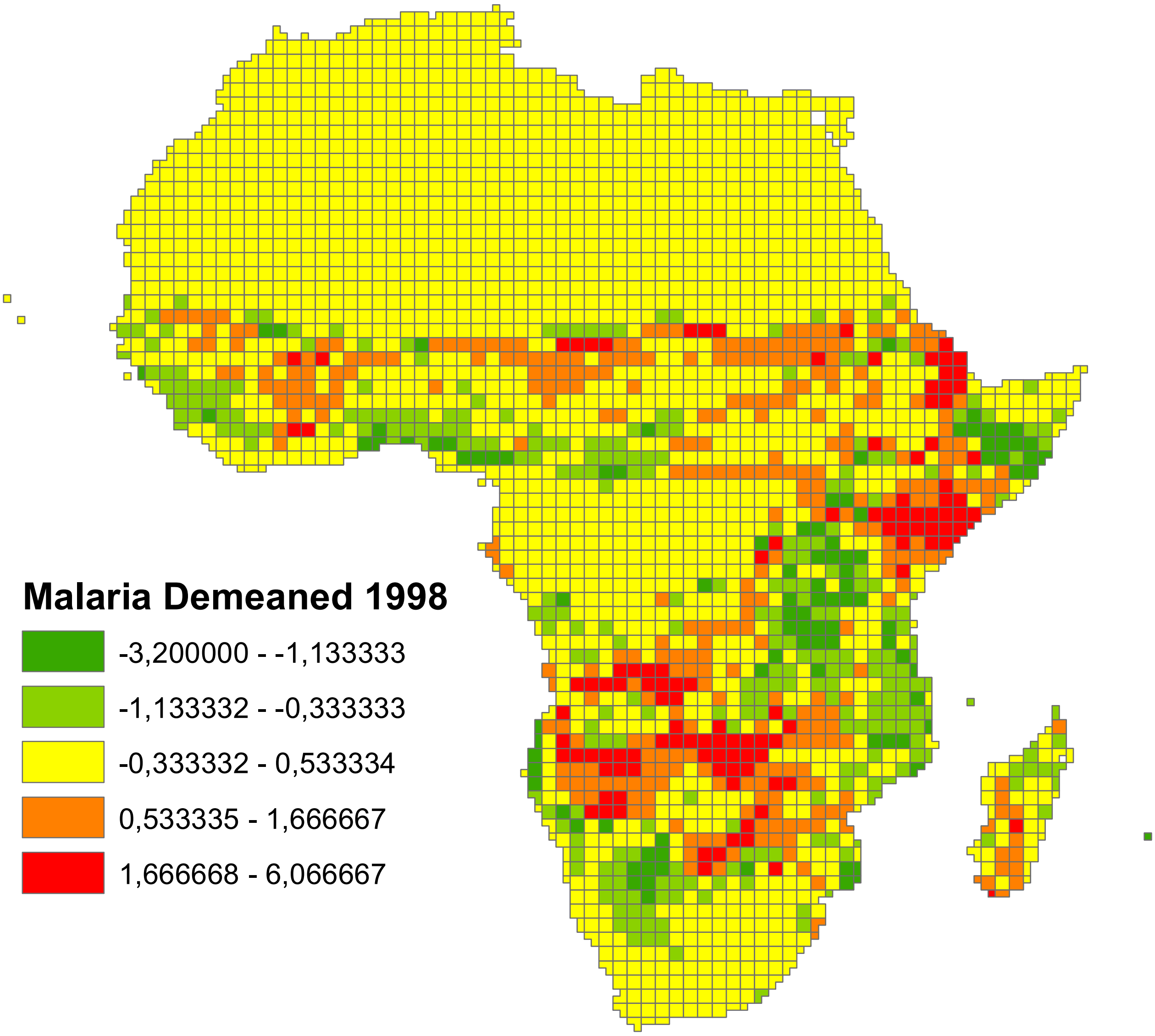
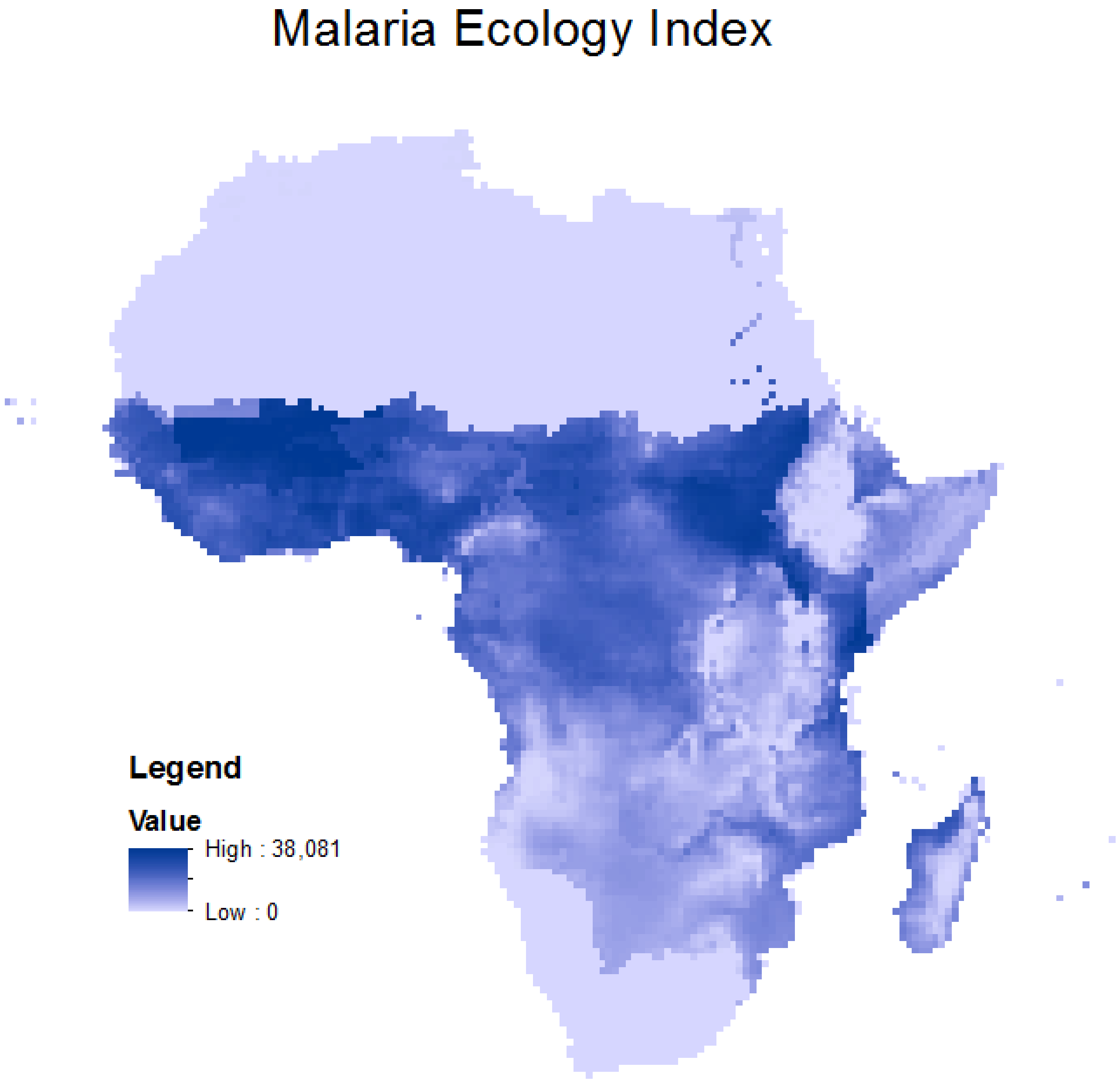


FIGURE D.2: Malaria Suitable Months Demeaned 1998

Figure D.3: Malaria Ecology in Africa



# Bibliography

Acemoglu, D., C. García-Jimeno, and J. Robinson (2012): "Finding Eldorado: Slavery and Long-run Development in Colombia," *Journal of Comparative Economics.*

Acemoglu, D., and S. Johnson (2006): "Disease and development: the effect of life expectancy on economic growth," .

Acemoglu, D., S. Johnson, and J. A. Robinson (2001): "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, 91(5), 1369–1401.

Aiyar, S., C. Dalgaard, and O. Moav (2008): "Technological progress and regress in pre-industrial times," *Journal of Economic Growth*, 13(2), 125–144.

Alesina, A., and E. Ferrara (1999): "Participation in heterogeneous communities," .

——— (2004): "Ethnic diversity and economic performance," .

Alesina, A., P. Giuliano, and N. Nunn (2011a): "Fertility and the Plough," *American Economic Review*, 101(3), 499–503.

Alesina, A., P. Giuliano, and N. Nunn (2011b): "On the origins of gender roles: Women and the plough," .

Alland, A. (1970): *Adaptation in cultural evolution: an approach to medical anthropology.* Columbia University Press New York.

Allison, A. (2006): "Genetic factors in resistance to malaria," *Annals of the New York Academy of Sciences*, 91(3), 710–729.

Altonji, J., T. Elder, and C. Taber (2005): "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools," *Journal of Political Economy*, 113(1), 151–184.

Angrist, J., and J. Pischke (2008): *Mostly harmless econometrics: An empiricist's companion.* Princeton Univ Pr.

Ashraf, N., G. Fink, and D. Weil (2010): "Evaluating the effects of large scale health interventions in developing countries: The zambian malaria initiative," .

Ashraf, Q., and O. Galor (2011a): "Dynamics and Stagnation in the Malthusian Epoch," *The American Economic Review*, 101(5), 2003–2041.

——— (2011b): "The" Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic Development," .

——— (2011c): "The" Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic Development," .

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

Ashraf, Q. H., A. Lester, and D. N. Weil (2008): "When Does Improving Health Raise GDP?," (14449).

Bamshad, M., T. Kivisild, W. Watkins, M. Dixon, C. Ricker, B. Rao, J. Naidu, B. Prasad, P. Reddy, A. Rasanayagam, et al. (2001): "Genetic evidence on the origins of Indian caste populations," *Genome research*, 11(6), 994–1004.

Banerjee, A., L. Iyer, and R. Somanathan (2005): "History, social divisions, and public goods in rural India," *Journal of the European Economic Association*, 3(2-3), 639–647.

Barry, B. (1998): *Senegambia and the Atlantic slave trade.* Cambridge University Press.

Barth, F. (1956): "Ecologic relationships of ethnic groups in Swat, North Pakistan," *American Anthropologist*, 58(6), 1079–1089.

Bates, R. H. (2008): *When things fell apart.* Cambridge University Press Cambridge.

Becker, G. S., and C. B. Mulligan (1997): "The Endogenous Determination of Time Preference," *Quarterly Journal of Economics*, 112(3), 729–758.

Becker, P. (1985): *The pathfinders: the saga of exploration in southern Africa.* Viking.

Bergad, L. W. (2007): *The comparative histories of slavery in Brazil, Cuba, and the United States.* Cambridge University Press.

Berreman, G. (1968): "Stratification, Pluralism and Interaction: a comparative analysis of caste," in *Ciba Foundation Symposium-Caste and Race: Comparative Approaches*, pp. 45–73. Wiley Online Library.

Besley, T., and T. Persson (2009): "Repression or Civil War?," *American Economic Review*, 99(4), 292–297.

——— (2011): "The Logic of Political Violence," *Quarterly Journal of Economics*, 126(3), 1411–1445.

Birchenall, J. (2010): "Disease and diversity in Africa's Long-term Economic Development," .

Bleakley, H. (2006): "Malaria In The Americas: A Retrospective Analysis Of Childhood Exposure," (003185).

Bloom, D., D. Canning, and J. Sevilla (2003): *The demographic dividend: a new perspective on the economic consequences of population change*, vol. 5. Rand Corp.

Bockstette, V., A. Chanda, and L. Putterman (2002): "States and markets: The advantage of an early start," *Journal of Economic Growth*, 7(4), 347–369.

Borah, W., and S. F. Cook (1969): "Conquest and population: a demographic approach to Mexican history," *Proceedings of the American Philosophical Society*, 113(2), 177–183.

Bose, P. (1916): *The Illusions of New India.* W. Newman.

Brahmajothi, V., R. Pitchappan, V. Kakkanaiah, M. Sashidhar, K. Rajaram, S. Ramu, K. Palanimurugan, C. Paramasivan, and R. Prabhakar (1991): "Association of pulmonary tuberculosis and HLA in South India," *Tubercle*, 72(2), 123–132.

Brower, B., and B. R. Johnston (2007): *Disappearing peoples?: indigenous groups and ethnic minorities in South and Central Asia.* Left Coast Press.

Carter, R., and K. N. Mendis (2002): "Evolutionary and historical aspects of the burden of malaria," *Clinical Microbiology Reviews*, 15(4), 564–594.

Cashdan, E. (2001): "Ethnic diversity and its environmental determinants: effects of climate, pathogens, and habitat diversity," *American Anthropologist*, 103(4), 968–991.

Cervellati, M., and U. Sunde (2009): "Life expectancy and economic growth: the role of the demographic transition," *Journal of Economic Growth*, pp. 1–35.

Cervellati, M., U. Sunde, and S. Valmori (2011a): "Disease Environment and Civil Conflicts," (5614).

——— (2011b): "The Distribution of Infectious Diseases and Extrinsic Mortality Across Countries," .

——— (2014): "Pathogens, Weather Shocks and Civil Conflicts," Discussion paper, Mimeo, University of Bologna.

Cockburn, T. (1971): "Infectious diseases in ancient populations," *Current Anthropology*, 12(1), 45–62.

Coelho, P., and R. McGuire (1997): "African and European Bound Labor in the British New World: The Biological Consequences of Economic Choices," *The Journal of economic history*, 57(01), 83–115.

Coelho, P. R., and R. A. McGuire (2006): "Racial differences in disease susceptibilities: Intestinal worm infections in the early twentieth-century American South," *Social history of medicine*, 19(3), 461–482.

Comin, D., W. Easterly, and E. Gong (2006): "Was the Wealth of Nations Determined in 1000 bc?," .

Crosby, A. W. (2003): *The Columbian exchange: biological and cultural consequences of 1492*, vol. 2. Greenwood Publishing Group.

Curtin, P. (1968): "Epidemiology and the slave trade," *Political Science Quarterly*, pp. 190–216.

Curtin, P. D. (1998): *Disease and empire: The health of European Troops in the Conquest of Africa*. Cambridge University Press.

Curtis, V., R. Aunger, and T. Rabie (2004): "Evidence that disgust evolved to protect from risk of disease," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271(Suppl 4), S131–S133.

Curtis, V., M. De Barra, and R. Aunger (2011): "Disgust as an adaptive system for disease avoidance behaviour," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1563), 389.

Daaku, K. Y. (1970): *Trade and politics on the Gold Coast, 1600-1720: a study of the African reaction to European trade*. Clarendon Press.

Dell, M. (2010): "The persistent effects of Peru's mining Mita," *Econometrica*, 78(6), 1863–1903.

Diamond, J., and L. Ford (2000): "Guns, germs, and steel: the fates of human societies," *Perspectives in Biology and Medicine*, 43(4), 609.

Diop, C., and H. Salemson (1987): *Precolonial Black Africa: a comparative study of the political and social systems of Europe and Black Africa, from antiquity to the formation of modern states*. L. Hill.

DOBSON, A. P., AND E. R. CARPER (1996): “Infectious diseases and human population history,” *Bioscience*, 46(2), 115–126.

DOWNES, A. (1987): “The economics of transition to the black labor system in Barbados, 1630-1680,” *The Journal of Interdisciplinary History*, 18(2), 225–247.

DUMONT, L. (1980): *Homo hierarchicus: the caste system and its implications*. University of Chicago Press.

DUTTA, P. K. (1995): “Collusion, Discounting, and Dynamic Games,” *Journal of Economic Theory*, 66, 289–306.

EASTERLY, W., AND R. LEVINE (1997): “Africa’s growth tragedy: policies and ethnic divisions,” *The Quarterly Journal of Economics*, 112(4), 1203.

ELTIS, D., AND L. C. JENNINGS (1988): “Trade between Western Africa and the Atlantic world in the pre-colonial era,” *The American Historical Review*, 93(4), 936–959.

ENGERMAN, S., AND K. SOKOLOFF (1997): *Factor Endowments, Institutions, and Differential Growth Paths among New World Economies*, vol. 89. Stanford University Press.

EPSTEIN, S. (1998): “Craft guilds, apprenticeship, and technological change in preindustrial Europe,” *Journal of Economic History*, 58, 684–713.

FAGE, J. D. (1969): *A history of West Africa*. Cambridge Univ Press.

FEARON, J., AND D. LAITIN (2003): “Ethnicity, insurgency, and civil war,” *American Political Science Review*, 97(1), 75–90.

FINCHER, C., R. THORNHILL, D. MURRAY, AND M. SCHALLER (2008): “Pathogen prevalence predicts human cross-cultural variability in individualism/collectivism,” *Proceedings of the Royal Society B: Biological Sciences*, 275(1640), 1279.

FOGLI, A., AND L. VELDKAMP (2011): “Germs, social networks and growth,” .

FREITAS, K. (2006): “The Indian caste system as a means of contract enforcement,” mimeo Northwestern University. Paper presented at the NEUDC conference 2006, Cornell University.

GALENSON, D. W. (1981): *White servitude in colonial America: An economic analysis*. Cambridge University Press Cambridge.

GALLUP, J., J. SACHS, AND A. MELLINGER (1998): “Geography and economic development,” .

GENOVESE, E. D. (1989): *The political economy of slavery: Studies in the economy and society of the slave South.* Wesleyan University Press.

GORDON, R. G., AND B. F. GRIMES (2005): *Ethnologue: Languages of the world*, vol. 15. SIL international Dallas, TX.

GOUDIE, R. J., S. MUKHERJEE, J. E. DE NEVE, A. J. OSWALD, AND S. WU (2014): "Happiness as a Driver of Risk Avoiding Behavior," *Economica, forthcoming.*

GREIF, A. (1993): "Contract enforceability and economic institutions in early trade: The Maghribi traders' coalition," *American economic review*, 83, 525–525.

——— (1994): "Cultural beliefs and the organization of society: A historical and theoretical reflection on collectivist and individualist societies," *Journal of political economy*, pp. 912–950.

——— (2006): *Institutions and the path to the modern economy: Lessons from medieval trade.* Cambridge Univ Pr.

GUISO, L., P. SAPIENZA, AND L. ZINGALES (2006): "Does Culture Affect Economic Outcomes?," *Journal of Economic Perspectives*, 20(2), 23–48.

GUISO, L., P. SAPIENZA, AND L. ZINGALES (2008): "Long term persistence," .

HALL, G. M. (2005): *Slavery and African ethnicities in the Americas: restoring the links.* University of North Carolina Press.

HAY S.I., GUERRA C.A., T. A. N. A. . S. R. (2004): "The global distribution and population at risk of malaria: past, present and future," *Lancet Infectious Diseases*, 4(6), 327–336.

HOFF, K., P. PANDEY, AND W. B. D. R. G. I. C. TEAM (2004): *Belief systems and durable inequalities: an experimental investigation of Indian caste*, vol. 3351. World Bank, Development Research Group, Investment Climate.

HSIANG, S. M., M. BURKE, AND E. MIGUEL (2013): "Quantifying the Influence of Climate on Human Conflict," *Science*, DOI: 10.1126/science.1235367(published online August 1, 2013).

HUTTON, J. (1946): *Caste in India.* University Press.

ILIFFE, J. (1987): *The African poor: A history*, vol. 58. Cambridge Univ Pr.

INIKORI, J. E. (2003): "The Struggle against the Trans-Atlantic Slave Trade," *fighting the slave Trade: West African strategies*, pp. 170–198.

Kiple, K., and V. King (2003): *Another Dimension to the Black Diaspora: diet, disease and racism.* Cambridge University Press.

Kiple, K. F. (1986): "A Survey of Recent Literature on the Biological Past of the Black," *Social Science History*, 10(4), 343–367.

Kiszewski, A., A. Mellinger, A. Spielman, P. Malaney, S. E. Sachs, and J. Sachs (2004): "A global index representing the stability of malaria transmission," *American Journal of Tropical Medicine and Hygiene*, 70(5), 486–498.

Krieger, N. (1987): "Shades of Differnce: Theoretical Underpinnings of the Medical Controversy on Black/White Differences in the United States, 1830-1870," *International Journal of Health Services*, 17(2), 259–278.

Kwiatkowski, D. P. (2005): "How malaria has affected the human genome and what human genetics can teach us about malaria," *The American Journal of Human Genetics*, 77(2), 171–192.

Lammers, J., and S. van Wijnbergen (2008): "HIV/AIDS, Risk Aversion and Intertemporal Choice," *Tinbergen Institute Discussion Paper*, 2007-098/1.

Livi-Bacci, M. (2006): "The depopulation of Hispanic America after the conquest," *Population and Development Review*, 32(2), 199–232.

Lorentzen, P., J. McMillan, and R. Wacziarg (2005): "Death and development," .

Low, B. (1988): "Pathogen stress and polygyny in humans," *Human reproductive behaviour: A Darwinian perspective*, pp. 115–128.

Ludvico, L., and J. Kurland (1995): "Symbolic or not-so-symbolic wounds: The behavioral ecology of human scarification," *Ethology and Sociobiology*, 16(2), 155–172.

Lysenko, A.J., S. I. (1968): "Geography of malaria. Amedico-geographic profile of an ancient disease," *Medicinskaja Geografija*, p. 25146.

Mallard, R. Q. (1892): *Plantation life before emancipation.* Negro History Press.

Mann, C. C. (2011): *1493: Uncovering the New World Columbus Created.* Alfred a Knopf Incorporated.

Manning, P. (1990): *Slavery and African Life: occidental, oriental, and African slave trades*, vol. 67.

McEvedy, C., R. Jones, et al. (1978): *Atlas of world population history.* Penguin Books Ltd, Harmondsworth, Middlesex, England.

McNeill, J. (2010): "Mosquito Empires: Ecology And War In The Greater Caribbean, 1620-1914 (New Approaches To The Americas) Author: JR McNe," .

McNeill, W. (1976): "Plagues and Peoples. 1976," *Garden City, NY: Anchor P.*

Michalopoulos, S. (2008): "The Origins of Ethnolinguistic Diversity: Theory and Evidence," (11531).

——— (2012): "The origins of ethnolinguistic diversity," *The American Economic Review*, 102(4), 1508–1539.

Modiano, G., G. Morpurgo, L. Terrenato, A. Novelletto, A. Di Rienzo, B. Colombo, M. Purpura, M. Mariani, S. Santachiara-Benerecetti, A. Brega, et al. (1991): "Protection against malaria morbidity: near-fixation of the $\alpha$-thalassemia gene in a Nepalese population," *American journal of human genetics*, 48(2), 390.

Moreira, M. (2003): "A conditional likelihood ratio test for structural models," *Econometrica*, 71(4), 1027–1048.

Mountain, J., J. Hebert, S. Bhattacharyya, P. Underhill, C. Ottolenghi, M. Gadgil, and L. Luca Cavalli-Sforza (1995): "Demographic history of India and mtDNA-sequence diversity," *American journal of human genetics*, 56, 979–992.

Murdock, G. (1967): "Ethnographic atlas: A summary," *Ethnology*, 6(2), 109–236.

Murray, D., and M. Schaller (2010): "Historical prevalence of infectious diseases within 230 geopolitical regions: A tool for investigating origins of culture," *Journal of Cross-Cultural Psychology*, 41(1), 99–108.

Murray, M. (2006a): "Avoiding invalid instruments and coping with weak instruments," *The Journal of Economic Perspectives*, 20(4), 111–132.

Murray, M. (2006b): "The bad, the weak, and the ugly: Avoiding the pitfalls of instrumental variables estimation," .

North, D. (1990): *Institutions, institutional change, and economic performance.* Cambridge Univ Pr.

Nunn, N. (2007): "The long-term effects of Africa's slave trades," .

——— (2008): "Slavery, Inequality, and Economic Development in the Americas," *Institutions and economic performance.*

NUNN, N., AND N. QIAN (2010): "The Columbian exchange: A history of disease, food, and ideas," *The Journal of Economic Perspectives*, 24(2), 163–188.

NUNN, N., AND L. WANTCHEKON (2009): "The slave trade and the origins of mistrust in Africa," .

OLSSON, O., D. HIBBS, ET AL. (2005): "Biogeography and long-run economic development," *European Economic Review*, 49(4), 909–938.

ORGANIZATION, W. H. (2000): "WHO expert committee on malaria: twentieth report," .

PUTTERMAN, L. (2008): "Agriculture, diffusion and development: Ripple effects of the neolithic revolution," *Economica*, 75(300), 729–748.

RAMAIAH, K., K. KUMAR, AND K. RAMU (1996): "Knowledge and beliefs about transmission, prevention and control of lymphatic filariasis in rural areas of South India," *Tropical Medicine & International Health*, 1(4), 433–438.

RAMEN, F. (2002): *Sleeping Sickness and Other Parasitic Tropical Diseases*. The Rosen Publishing Group.

RANSOM, R. L., AND R. SUTCH (2001): *One kind of freedom: The economic consequences of emancipation*. Cambridge University Press.

RODNEY, W. (1972): "How europe underdeveloped africa," *Beyond borders: Thinking critically about global issues*, pp. 107–125.

RUTMAN, D. B., AND A. H. RUTMAN (1976): "Of agues and fevers: malaria in the early Chesapeake.," *The William and Mary Quarterly*, 33, 31.

RUTMAN, D. B., C. WETHERELL, AND A. H. RUTMAN (1980): "Rhythms of life: Black and white seasonality in the early Chesapeake," *Journal of Interdisciplinary History*, pp. 29–53.

SACHS, J. (2001): "Tropical underdevelopment," .

——— (2003): "Institutions don't rule: direct effects of geography on per capita income," .

SANTACREU-VASUT, E. (2009): "Diversity and Institutions: Theory and Evidence from the Spread of Industrialization," .

SAVITT, T. L. (2002): *Medicine and slavery: The diseases and health care of blacks in antebellum Virginia*. University of Illinois Press.

Schwartz, S. B. (1978): "Indian labor and New World plantations: European demands and Indian responses in northeastern Brazil," *The American Historical Review*, pp. 43–79.

Shack, W. (1964): "54. Notes on Occupational Castes Among the Gurage of South-West Ethiopia," *Man*, 64(1), 50–52.

Shastry, G., and D. Weil (2003): "HOW MUCH OF CROSS-COUNTRY INCOME VARIATION IS EXPLAINED BY HEALTH?," *Journal of the European Economic Association*, 1(2-3), 387–396.

Siddiqui, M., S. Meisner, K. Tosh, K. Balakrishnan, S. Ghei, S. Fisher, M. Golding, N. Narayan, T. Sitaraman, U. Sengupta, et al. (2001): "A major susceptibility locus for leprosy in India maps to chromosome 10 p 13," *Nature genetics*, 27(4), 439–441.

Sokoloff, K., and S. Engerman (2000): "History lessons: Institutions, factors endowments, and paths of development in the new world," *The Journal of Economic Perspectives*, 14(3), 217–232.

Tabellini, G. (2007): "Culture and institutions," *CEPR Discussion Paper No. DP6589.*

Tamari, T. (1991): "The development of caste systems in West Africa," *Journal of African History*, 32(2), 221–50.

Tanser, F. C., B. Sharp, and D. Le Sueur (2003): "Potential effect of climate change on malaria transmission in Africa," *The Lancet*, 362(9398), 1792–1798.

Thapar, R. (2004): *Early India: From the origins to AD 1300.* Univ of California Pr.

Tuden, A., L. Plotnicov, and T. L. K. C. L. of Congress. Hebraic Section) (1970): *Social stratification in Africa.* The Free Press New York.

Waring, J. I. (1975): "Colonial medicine in Georgia and South Carolina.," *The Georgia historical quarterly*, 59, 141.

Watts, D. (2000): "Early Hispanic New World Agriculture, 1492 to 1509," *Caribbean slavery in the Atlantic world: a student reader.*

Weber, M. (1958): "The Religion of India: The Sociology of Hinduism and Buddhism, trans," *HH Gerth and Don Martindale (Glencoe, Ill.: The Free Press, 1958)*, p. 207.

Weil, D. N. (2007): "Accounting for The Effect of Health on Economic Growth," *The Quarterly Journal of Economics*, 122(3), 1265–1306.

Whatley, W. C., and R. Gillezeau (2011): "The fundamental impact of the slave trade on African economies," *Economic Evolution and Revolution in Historical Time*, pp. 86–110.

Wolfe, N. D., C. P. Dunavan, and J. Diamond (2007): "Origins of major human infectious diseases," *Nature*, 447(7142), 279–283.

Wood, P. (1974): "Black Majority: Negroes in Colonial South Carolina from 1670 through the Stono Rebellion," *Norton Library, New York.*

Wright, G. (1978): *The political economy of the cotton South: households, markets, and wealth in the nineteenth century.* Norton New York.