AN ANALYSIS OF GLOBAL POPULATION MOVEMENTS FROM 1500 TO 2000

MEHMET BURAK AKREN

BOĞAZİÇİ UNIVERSITY

2024

AN ANALYSIS OF GLOBAL POPULATION MOVEMENTS FROM 1500 TO 2000

Thesis submitted to the

Institute for Graduate Studies in Social Sciences

in partial fulfillment of the requirements for the degree of

Master of Arts

in

Economics

by

Mehmet Burak Akren

Boğaziçi University

2024

An Analysis of Global Population Movements from 1500 to 2000

The thesis of Mehmet Burak Akren

has been approved by:

Assoc. Prof. Kamil Kıvanç Karaman _____
(Thesis Advisor)

Assoc. Prof. Sabri Arhan Ertan _____
(Thesis Co-Advisor)

Assoc. Prof. Murat Koyuncu _____

Assist. Prof. Selcen Çakır _____

Assoc. Prof. Cemal Eren Arbatlı _____
(External Member)

June 2024

DECLARATION OF ORIGINALITY

I, Mehmet Burak Akren, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produces or published by another person or institution;
- this thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature ……

Date …….

ABSTRACT

An Analysis of Global Population Movements from 1500 to 2000

This thesis explores the main reasons behind human mobility since 1500, focusing on the different regional effects of various economic, political, social, and geographic factors. The initial cross-country analysis highlights the significant impact of geographic elements and historical influences on migration. The bilateral data structure in the second analysis allowed us to examine the effects in both source and destination countries at the same time, along with bilateral factors. Three main conclusions emerged from this study. First, migration is influenced by both source and destination countries, requiring an origin-destination framework for accurate historical migration assessment. Second, analyzing deep-rooted factors at the regional level provides more accurate results, as shown by the differences between the Eurasia and Sub-Saharan Africa regions. A global analysis might lead to biased results. Lastly, historical migrations were shaped by complex economic, political, and geographic influences in unique ways. Although often used as natural experiments in economic analysis, colonial migrations are affected by a complex network of factors and need to be considered carefully.

ÖZET

1500'den 2000'e Küresel Nüfus Hareketlerinin Analizi

Bu tez, 1500’den itibaren gerçekleşen insan hareketliliğinin başlıca nedenlerini, çeşitli ekonomik, politik, sosyal ve coğrafi faktörlerin farklı bölgesel etkilerine odaklanarak araştırmaktadır. İlk yapılan çapraz-ülke analizi, coğrafi unsurların ve tarihsel etkilerin göç üzerindeki önemli etkisini vurgulamaktadır. İkinci analizdeki ikili veri yapısı, kaynak ve varış ülkelerindeki etkileri ile beraber, ikili faktörleri incelememize olanak tanımıştır. Bu çalışmadan üç ana sonuç ortaya çıkmıştır. Birincisi, göç hem kaynak hem de varış ülkelerindeki faktörlerden etkilenir ve tarihi göçleri doğru bir şekilde değerlendirmek için bir kaynak-varış veri yapısı gerekmektedir. İkincisi, derin etkili faktörlerin bölgesel düzeyde analiz edilmesi, Avrasya ve Sahra Altı Afrika bölgeleri arasındaki farklılıklar gibi daha doğru sonuçlar sağlar. Küresel bir analiz hatalı olabilecek sonuçlara yol açabilir. Son olarak, tarihi göçler, karmaşık ekonomik, politik ve coğrafi etkilerle kendine özgü yollarla şekillenmiştir. Ekonomik analizlerde genellikle doğal deneyler olarak kullanılsalar da kolonyal göçleri karmaşık bir değişken ağı tarafından etkilenir ve dikkatle ele alınmalıdır.

ACKNOWLEDGMENTS

I want to thank my co-advisor, Arhan Ertan, for his invaluable motivation and guidance throughout this process. Completing this thesis would have been much more challenging without his insightful comments. I also extend my deepest gratitude to my other co-advisor, Kıvanç Karaman, for allowing me to work with him and always being ready to support me with his ideas. I would also like to express my gratitude to Professors Eren Arbatlı, Murat Koyuncu, and Selcen Çakır for their valuable support. Additionally, I would like to thank the Economics Department and the International Trade Department of Boğaziçi University.

I am deeply grateful to everyone who supported me throughout my master's program. This was a unique and important period in my life, and their support meant a lot to me. Specifically, I want to thank my dear friend Hakan for his never-ending support.

Finally, of course, I want to thank my dear family, Bülent, Nilgün and Şevval, for always being there to support me. I was not able to achieve this without them.

To all the wonderful people, striving to live and be good…

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1

CHAPTER 2: LITERATURE REVIEW 6

2.1 Persistence literature 7

2.2 Spatial models 11

2.3 Origin destination framework and The Gravity Model 12

CHAPTER 3: BACKGROUND INFORMATION 14

CHAPTER 4: DATA AND VARIABLES 18

4.1 World Migration Matrix 18

4.2 Dependent variables 20

4.3 Independent variables 23

CHAPTER 5: METHODOLOGY AND ANALYSES 32

5.1 Cross-sectional unilateral analyses with regional shares 32

5.2 Bilateral analyses 46

CHAPTER 6: CONCLUSION 59

APPENDIX A: FIGURES OF CHAPTER 4 62

APPENDIX B: TABLES OF CHAPTER 4 66

APPENDIX C: FIGURES OF CHAPTER 5 68

APPENDIX D: TABLES OF CHAPTER 5 72

REFERENCES 90

LIST OF FIGURES

Figure 1. A small portion of WMM …….. 62

Figure 2. Mean of EuroShare across regions ……..62

Figure 3. Mean of AsiaShare across regions ……..63

Figure 4. Mean of SSAshare across regions ……...63

Figure 5. Mean of DS by regions for Latin America and Caribbean …….64

Figure 6. Mean of DSv2 by regions for Latin America and Caribbean …….....64

Figure 7. Mean of DS by regions for North America ……65

Figure 8. Mean of DSv2 by regions for North America ..……..65

Figure 9. Moran's I plot for EuroShare ……..……68

Figure 10. Moran's I plot for SSAshare …….68

Figure 11. Moran's I plot for AsiaShare ……69

Figure 12. S_lnpd1500 and DSv2 change for Eurasia to Non-Eurasia ……..69

Figure 13. D_MalariaEcology and DSv2 change for Eurasia to Non-Eurasia ……...70

Figure 14. D_lnpd1500 and DSv2 change for Europe to Non-Eurasia ……..70

Figure 15. D_Precious Metals and DSv2 change for Europe to Non-Eurasia ……..70

Figure 16. Absolute Latitude Difference and DSv2 change for Sub-Saharan Africa to new lands ……...…….……71

Figure 17. D_Percentage of Arable Lands and DSv2 change for Sub-Saharan Africa to new lands ……...……...…….……..71

LIST OF TABLES

Table 1. Number of Countries for Regions ……..…… 66

Table 2. Descriptive Statistics for Bilateral Data Structure ……...66

Table 3. Descriptive Statistics for Unilateral Independent Variables ……....67

Table 4. Descriptive Statistics for Regional Share Variables …….....72

Table 5. Baseline Regression Results for EuroShare …….....72

Table 6. Regression Results for EuroShare with More Controls …….....73

Table 7. Baseline Regression Results for SSAshare …….....74

Table 8. Regression Results for SSAshare with More Controls …….....75

Table 9. Moran's I and Lagrange Multiplier Tests …….....6

Table 10. Spatial ML Results for EuroShare …….....77

Table 11. Spatial ML Results for SSAshare ……..…….....77

Table 12. Alternative Estimation Method for Spatial Analyses : EuroShare ……78

Table 13. Alternative Estimation Method for Spatial Analyses : SSAshare ……..78

Table 14. Only Non-Eurasia for EuroShare ……...……...……...……..79

Table 15. Only New Lands for EuroShare ……..…….....79

Table 16. Only New Lands for SSAshare ..…….....79

Table 17. Alternative Independent Variables for EuroShare …….....80

Table 18. Alternative Independent Variables for SSAshare …….....80

Table 19. Additional Control Variables for EuroShare …….....80

Table 20. Additional Control Variables for SSAshare …….....81

Table 21. Spatial Analyses with Colonial Dummies for EuroShare ……..81

Table 22. Spatial Analyses with Colonial Dummies for SSAshare ……...81

Table 23. Comparison with WMM 2000 for EuroShare ……..…….82

Table 24. Comparison with WMM 2000 for SSAshare …….....82

Table 25. AsiaShare ……...……...……..……...82

Table 26. From Eurasia to Non-Eurasia Baseline Results …….....83

Table 27. From Eurasia to Non-Eurasia Detailed Results …….....83

Table 28. Regional Bilateral Analyses ……..…….....84

Table 29. Bilateral Analyses for Europe …….……...84

Table 30. Bilateral Analyses for SSA ……..……...85

Table 31. Bilateral Analyses for Asia ……..……..…85

Table 32. Bilateral Results with DS from WMM 1960 Version ...…….....86

Table 33. .Bilateral Results with DS from WMM 2000 Version ……...….....86

Table 34. Bilateral Results with DSv2 from WMM 2000 Version …….....87

Table 35. Different Regional Settings: Source is Eurasia and Destination is "New Lands" ……..……...…….....87

Table 36. Different Regional Settings: Source is Europe and Destination is "Non-Europe" ……..……...…….....88

Table 37. Different Regional Settings: Source Asia and Destination is "Non-Asia" ... 88

Table 38. Different Regional Settings: Source Asia and Destination is "New Lands" ..89

CHAPTER 1

INTRODUCTION

Human mobility can occur for various reasons, like geographical factors, economic opportunities, resource availability, as well as coercion, the need to escape from poverty; and it leads to several outcomes, such as wars and economic transformations. Perhaps most notably, many cultures have converged, different ideas have spread, and new political and economic systems have emerged due to human mobility. Additionally, the movement of people has also spread diseases, causing widespread despair and suffering.

Developments in marine technology around the 1400s significantly increased mobility and activity in the oceans. Several societies used newer technologies to explore previously unknown areas. These movements significantly impacted both new and old locations (Manning & Trimmer, 2020). The history of European colonialism in the Americas clearly illustrates this impact. The displacement of indigenous communities, the introduction of new political regimes, and new ways of production influenced the direction of development and still affect the region today (Acemoglu et al., 2001). Also, the internal and international labor migrations that occurred over the years following the Industrial Revolution have significant effects today (Haas et al., 2019).

This thesis will conduct a macro-level, aggregated analysis of the factors that influenced human mobility in source and destination countries over an extended period. In the literature, these factors are called “deep-rooted factors” that significantly impact past and contemporary events (Spolaore & Wacziarg, 2013). For instance, geographic factors are one of the most significant factors since the dawn of the history of homo

sapiens and its ancestors (Gallup et al., 1999; Hoerder, 2002; Sachs, 2012). In addition, numerous economic, historical, and cultural factors also influenced human history (Banerjee & Iyer, 2005; Nunn, 2008; Voigtländer & Voth, 2012).

Regarding the several factors that influence migration, the study's initial focus is on geographical elements. Existing literature widely acknowledges their profound impact on the course of history and the economic development of countries (Diamond, 1999; Olsson & Hibbs, 2004; Sachs, 2003). Furthermore, historical factors play a significant role in shaping migration patterns and historical events. One extensively discussed factor in the literature on long-run effects is the characteristics of the population in both the source and destination regions. The existing composition of the people in these areas had significant impacts on the interactions of arriving populations. Also, various other factors, such as local population features, mortality rates, state history, historical characteristics, and technological levels, influenced people's migration decisions throughout the studied period (Acemoglu et al., 2001; Easterly & Levine, 2016). When analyzing migrations between various regions through the lens of push-pull factors, particularly from Eurasia to areas beyond, the study identified both positive and negative influences of several variables. Among these, significant adverse impacts were observed from diseases, climate differences, and population density in the destination areas.

Conversely, the positive effects were exhibited from the proportion of arable land, valuable metals in the destination region, colonial relationships, and the economic development level of the source country. When examining the influence of these factors separately for both the European and Asian regions, it is evident that a significant portion of these variables are impactful. Additionally, when looking at migrations from

Sub-Saharan Africa to regions discovered during the colonial period, the direction of some variables' effects is reversed. During the period under study, people enslaved from African regions were often moved to areas with higher disease prevalence and to tropical plantations. However, the positive impact of variables such as arable land and valuable metals like gold and silver confirms that the destinations were economically prioritized areas.

The primary data used in this research as the dependent variable is obtained from the World Migration Matrix introduced by Putterman and Weil (2010), which covers 1500 to 2000. It presents a detailed breakdown of the ancestral composition of the populations in nearly every country worldwide, determining 1500 CE as a starting point. The main goal of the WMM is to indicate the countries in which the ancestors of today's residents lived in 1500 CE. Utilizing this matrix may provide valuable information about migration rates and patterns throughout this timeframe. Using this dataset, different dependent variables have been created for two analyses. The details of these will be explained in the following sections. This study focuses on mobility between 1500 and 2000, offering an overview of the historical events that have greatly influenced population movements. The reason for selecting 1500 to 2000 is that the WMM used in this study is designed to encompass these years, and the chosen period guarantees the inclusion of a wide array of significant historical events that have influenced the course of population movements on a global scale (Hoerder, 2002; Lucassen & Lucassen, 2009; Manning & Trimmer, 2020). On top of that, the updated 1500-1960 matrix is used to focus on the colonial era, which is from Chanda et al. (2014). Using this secondary matrix will eliminate the effects of modern migrations that occurred after the 1960s, thus highlighting the impacts of colonial migrations.

In the first part of the study, cross-country analyses were conducted with a focus on regionally clustered ancestry rates, followed by regression analyses. Considering the extensive historical period under examination, it is highly plausible that spatial factors played a significant role in shaping historical migrations, much like their influence on contemporary migrations as documented by recent studies (Bu et al., 2022; Jeanty et al., 2010; LeSage & Pace, 2010). Consequently, spatial models were also employed in the cross-country analyses on top of the OLS models.

Then, in the second part of the study, several analyses were conducted using the origin-destination format to examine migration dynamics in depth. Also, the Gravity model is used, a framework initially applied to migration patterns and later adapted to international trade analysis (Sen & Smith, 2012; Head & Mayer, 2014; Beine et al., 2016). The baseline model suggests that population sizes and distances between origin and destination influence migration. It expanded to include many variables, aligning with neoclassical economic theories emphasizing push and pull factors. Recently, the availability of detailed migration data has renewed interest in gravity models for migration analysis (Beine et al., 2011; Maurel & Tuccio, 2016; Ramos, 2016). Under the origin-destination framework, The Poisson Pseudo Maximum Likelihood (PPML) method has been used in the analyses. This method was praised in the recent literature primarily for its effectiveness in handling bilateral datasets where the dependent variable contains a significant proportion of zeros called a sparse matrix. Furthermore, it has been noted that biased estimates of elasticities are produced by parameters from log-linearized models estimated using Ordinary Least Squares in the presence of heteroskedasticity (Silva & Tenreyro, 2006; Santos Silva & Tenreyro, 2022). The details of the methodology will be explained in the following chapters.

The main hypothesis of the thesis is that migration is affected by conditions in both the origin and destination countries, so it is important to examine both to understand historical movements accurately. Second, studying these factors regionally, such as in Eurasia versus Sub-Saharan Africa, provides more precise results than a global overview, which could be misleading. Lastly, historical migrations were shaped by economic, political, and geographical influences, each affected differently by historical events.

The rest of this study is structured as follows. The next section reviews relevant literature, starting with studies on persistence, deep-rooted analyses, spatial models, and the gravity model literature, examining their connection to this research. Then, a brief historical background is presented, followed by the data and variables of interest, which outlines the preparation of the dependent variables and vital independent variables central to the research. Next is the methodology and analysis section, which presents the study's structure and findings. Finally, the conclusion chapter provides a summary and highlights further research opportunities.

CHAPTER 2

LITERATURE REVIEW

This study primarily intersects with three bodies of literature. First, it relates to the persistence literature, which mainly explores the deep historical roots of economic development and contemporary economic results. Given that the data spans a long period and encompasses the entire world, it is well-suited for exploration within this field. Second, spatial models will be used in the first part of the study, which looks at different ancestry shares of countries. Applying these methods to study historical migrations contributes to this field of research. Also, prepared ancestry share variables could be helpful for further studies in the literature. Third, it aligns with the studies that applied the origin-destination framework and Gravity Model to study migration. Some of these studies employed the Poisson Pseudo Maximum Likelihood (PPML) method to analyze factors unique to both source and destination regions and the bilateral factors. Additionally, although there are analyses of historical migrations at a more local level within the existing literature, this study is the primary one that has approached the subject from a broader, more comprehensive perspective.

2.1 Persistence literature

The impact of historical events on the performance of contemporary economies is a significant subject of inquiry in the social sciences. However, the relationship between economics and the study of historical events has experienced significant fluctuations over the centuries. History and economics were studied closely during the 18th and 19th centuries, with numerous prominent thinkers and economists, such as Karl Marx and Max Weber, emphasizing the historical underpinnings of economic phenomena (Marx, 2000; Weber, 2009). However, from the early 1900s onwards, the two fields diverged significantly.

The influential work by North and Greif in the 1980s and 1990s effectively brought historical analysis back into focus within economics. Their research emphasized the importance of integrating history into economic thought (North, 1990; Greif, 1993, 1994, 1998). Following that, the late 1990s saw a substantial change in economic history analysis, as evidenced by three innovative research series. Hypotheses that colonial policies and institutions have a long-term impact on contemporary economic and political performances are developed (Acemoglu et al., 2002). Meanwhile, how colonial legal frameworks impact modern financial and contractual systems is highlighted (La Porta et al., 1998). Similarly, Engerman & Sokoloff (2002, 2005) and Glaeser et al. (2004) pointed out how the contemporary economic disparity was affected by distinct colonial practices and natural endowments. This interest arises from a desire to comprehend why societies follow different trajectories and why some prosper and progress while others stagnate or decline. Abramitzky (2015) labels this academic attitude as “Big Think” and explains that it involves using detailed data and modern econometric methods to answer important questions in economic history.

Today, it is widely acknowledged that a significant portion of current economic development can be traced back to historical events such as the agricultural revolution, colonial rule, forced labor, older legislative structures, global conflicts, and revolutions (Banerjee & Iyer, 2005; Nunn, 2008; S. O. Becker & Woessmann, 2009; Ashraf & Galor, 2011; Michalopoulos & Papaioannou, 2016). The term "persistence" is central to this analysis. It emphasizes the lasting influence of historical events, policies, and conditions on economic outcomes.

In the field of persistence research, various topics are examined; still, they can be grouped into broad themes. The initial branch searches for the effects of events from more ancient periods in history on the present day, especially on contemporary economic development and social consequences. For instance, recent studies extend the foundational work of Olsson & Hibbs (2004) on the persistent effects of the agricultural revolution. Putterman (2008) investigates the correlation between adopting early agricultural practices and the economic performance of nations today. Vollrath (2011) specifically addresses the impact of agricultural productivity on the evolution of economic structures. Olsson & Paik (2016) link the agricultural revolution's early adoption to long-lasting collectivist norms and the formation of autocratic societies with social inequalities.

Not only the Agricultural Revolution but also many other developments from ancient times have been examined in various studies. For example, Bockstette et al.(2002) and Borcan et al. (2018, 2021) argue that the length of a state's history is a crucial predictor of a country's social and economic development. Also, expanding upon Comin et al. (2010) on the importance of technology rates, recent research further explores how technological advancements shape economic development and societal

norms. For example, Alesina et al. (2013) reveals how historical differences in plow usage have lasting impacts on contemporary gender roles. This branch of persistence literature is closely related to historian Fernand Braudel's concept of *longue durée* emphasizes the influence of structural and systemic factors, such as access to navigable rivers, proximity to trade routes, and environmental conditions, on economic patterns over extended periods (Braudel & Wallerstein, 2009). Expanding on Braudel's insights, persistence literature scholars have extensively explored geographical variables' impact on economic development. Engerman & Sokoloff (2002) examined how the availability of natural resources, the climate, and biodiversity can dictate economic success. Hall & Jones (1999) similarly examined the role of these natural factors in shaping economic landscapes. Olsson & Hibbs (2004) show that countries with richer environments adopted agriculture earlier and achieved higher GDPs. Carstensen & Gundlach (2006) explored the connection between geography and economic performance, suggesting that natural conditions significantly affect economic growth. Fenske (2011) merges 19th-century African ethnic group boundaries with data on 18 ecological zones to create an ecological diversity index for each group. Michalopoulos (2012) studied the intricate ways geographical conditions influence linguistic diversity. Alsan (2015) further emphasized the complex relationship between nature and economic progress.

The second part of the persistence literature investigates the effects of deep-rooted variables on various other events. Studies such as Voigtländer & Voth (2012) on the Black Death reveal how major plagues reshaped Europe's labor markets and wealth distribution. Acemoglu et al. (2011) show how the French Revolution led to significant changes in property rights. Greif (2006) underscores the economic benefits of ancient trade networks while Becker et al. (2011) highlighting the lasting positive effects of

early investments in education on Western Europe's economic prosperity. (Buggle et al., 2023) examines the factors that affect the migration of Jews facing persecution.

Despite the substantial research on "persistent effects," there remains a gap in understanding the complex interactions between geographical, economic, cultural, and historical factors that shape historical migration patterns. While the economics literature has extensively studied the economic, political, and social factors driving contemporary migration, less research has analyzed the "deep-rooted" factors behind historical and contemporary population movements; migration is often treated as an explanatory factor for other outcomes (Spolaore & Wacziarg, 2013). This thesis aims to fill the research gap by investigating the human and geographical features from history that may have influenced population movements over time. By exploring the interaction between these factors, the study seeks to understand how social, economic, and environmental characteristics shaped migration patterns. Of course, given that this is a very macro perspective, several details might be overlooked. However, the main motivation is to see if we can identify patterns that have emerged as a result of these deep-rooted effects despite these challenges. This is closely aligned with the core motivation behind all studies on persistence effects.

2.2 Spatial models

Spatial interaction models are important tools in geographical and regional economic studies, analyzing the flow of goods, people, or information between different locations (Anselin, 1988; Anselin & Griffith, 1988; Elhorst, 2014). A general spatial regression model, which is also known as the Spatial Autoregressive Model (SAR), for cross-sectional data is :

$$y = \rho Wy + \beta X + \varepsilon \quad (1)$$

In Equation (1), β represents the coefficient linked to exogenous factors X that is different from any lagged dependent variables. The effect of the spatially lagged dependent variable y is represented by the coefficient ρ. The disturbance vector ε has a constant variance and a covariance of zero, and it is practical to assume that ε has a normal distribution centered on zero for the sake of generalization. The N-by-N spatial distance weight W used to make this lag while creating the connections between an origin and destination country. This study uses inverse-distance weighting with latitude and longitude variables to produce this distance weight matrix. For example, this matrix produces the spatial lag vector Wy when paired with the vector y. The conceptual framework put out by Griffith & Jones (1980) is formally encapsulated by this spatial lag vector, which successfully captures the dynamics of spatial dependency they highlighted Anselin (1988).

This study looks at a broad historical period; it is conceivable that spatial factors, just as they influence contemporary migrations (Bu et al., 2022; Jeanty et al., 2010; LeSage & Pace, 2010), could have also affected historical migrations. Therefore, spatial

models will be utilized in the initial cross-country section of the analyses. It is believed that using these methods in the context of historical migration will contribute to the field.

2.3 Origin destination framework and The Gravity Models

Migration has been a part of human history for a long time; nevertheless, developing theories to explain migration is a relatively new area. In the 1880s, Ravenstein is one of the forerunners in the study of migration. He developed "Laws of Migration" based on statistical data on migration (Tobler, 1995) . These laws were established from observed trends, such as the typical inclination of migrants to migrate across shorter distances. These “laws” did not provide an extensive framework for comprehending migration, as expected. The push-pull model argues that migration is influenced by both "push" factors that force people to leave their homes, like poverty or conflict, and "pull" factors that draw them to a new location (Dorigo & Tobler, 1983; Parkins, 2010). It is said that neo-classical migration theories are rooted in push-pull theories, which regard migration as a macro-level result of income and other opportunity inequalities between origin and destination areas (Lucassen & Lucassen, 2014). In addition, applied applications, such as regression, usually have a similar fashion with push and pull analysis. Typically, migration studies naturally consider positive and negative factors when formulated. Similarly, in this study, factors that increase or decrease the chance of migration have been considered together.

The Gravity Model suggests that migration is related to distance and driven by population sizes at both the origin and destination. This model predicts that areas with higher population densities are more likely to experience migration. It aligns with the

neoclassical framework, emphasizing macro-level push and pull factors and holding similar assumptions about observations (Wilson, 1971; Anderson & van Wincoop, 2004, 2003; de Haas, 2014).

The Gravity Model has been used to explain various migration patterns, including cross-border migrations and internal shifts inside a country's borders. The basis of the model is the bilateral preparation of the data. In this setting, each variable appears to be both an origin and a destination, and all relationships between origin-destination pairs are the observations. This is implied by getting n^2 observations for n different countries. Most gravity models use this kind of data called the Origin-Destination type modeling. It is claimed that the structure can be compared to a closed-form system resembling a general equilibrium (Sen & Smith, 2012; Beine et al., 2011; Ramos, 2016).

The Origin-Destination Framework has consistently been useful in explaining much of the variation in international trade flows. Additional variables such as common borders, shared languages, and trade agreements have been added to the model to further analyze factors influencing trade. Several economists have offered theoretical arguments for the model (Anderson, 1979; Deardorff, 1998; De Benedictis & Taglioni, 2011; Head & Mayer, 2014). Recently, with the increased availability of dyadic migration data, there has been renewed interest in using gravity models for studying migration (Beine et al., 2011; Ramos, 2016). The term "gravity model" has since been applied more generally to linear or nonlinear regressions that relate migration flows to relevant predictor variables (Beine et al., 2016).

CHAPTER 3

BACKGROUND INFORMATION

During the colonial era, several European nations launched ambitious voyages, established extensive slave trade networks, traded valuable mines, made conquests, and engaged in various political, economic, and religious activities. These actions caused major demographic shifts across continents, permanently affecting communities, economies, and political landscapes.

One of the harshest features of this era was the high mortality rates of indigenous people due to illnesses coming from the old world and the displacement of these people from their homelands, followed by their coerced labor in mines. Along with indigenous people, millions of Africans were removed from their home regions and transported over the Atlantic to work in plantations, mines, and other areas. Also, many people willingly migrated to new locations during this period, while some of these migrations involved compulsion. By the 18th century, the Industrial Revolution introduced new job opportunities and boosted economic growth. Consequently, millions moved in search of better economic prospects, often relocating to rapidly expanding new industrial centers (Blackburn, 1997; Thornton, 1998; Hoerder, 2002; Haas et al., 2019). Therefore, there are three main categories of people: those who migrated voluntarily, those who were enslaved, and those who relocated or died because of illnesses.

- Indigenous people

The first step of colonial expansion was the discovery of valuable metals like gold and silver in the Americas. Explorers from Spain and Portugal were the initiators of the

colonization of Latin America. For instance, Francisco Pizarro overthrew the Inca Empire in Peru, while Hernán Cortés led the invasion of the Aztec Empire in Mexico. These conquests involved using military force and pacts with different native populations. After these conquests, colonies were founded throughout Latin America (Moch, 2003; McKeown, 2004; Lucassen & Lucassen, 2014). Shortly after their initial contact with the region, the Spaniards created colonial governments intending to control the natural resources of Latin America. Thus, they forcibly relocated indigenous people from their original lands and employed them in mining activities. The repercussions of colonial encounters for indigenous communities were frequently disastrous, and forced migration resulted in significant population losses and the disappearance of inhabitants' traditional ways of living. Furthermore, and more crucially, the transmission of diseases between European, African, and American populations was catastrophic. It is estimated that the number of Indigenous American settlements had fallen to just one-tenth of their original population. While European and African people were also exposed to several new diseases, they did not suffer from catastrophic population reductions. In the following phases of colonization, many indigenous people were employed on plantations, although the primary labor force mainly comprised enslaved Africans. Additionally, individuals were involved in mining and various other occupations within the plantation setting. In South America, the changes in the labor market caused by the disastrous diseases and subsequent developments in market structures led to alterations in labor systems. This facilitated the transition from compulsory labor arrangements to a free labor market and created opportunities for individuals to relocate freely (Zwart & Zanden, 2018).

- Enslavement

During this era, the Atlantic slave trade, which began in the early 16th century and continued into the 19th century, led to the enslavement of millions of Africans (Rawley & Behrendt, 2005; Eltis & Richardson, 2013). Over the centuries, these individuals were forcibly brought to the New World to work in mines and on plantations, growing cash crops like sugar, tobacco, cotton, and coffee. This demand for labor led European countries to establish vast agricultural plantations in South America, the Caribbean, and North America (Zwart & Zanden, 2018). Although slavery was present in early African communities, it was not a central aspect of their societies, as noted by (Lovejoy, 2011). The slave trade disrupted the social structures of both the African continent and the colonies, causing long-lasting demographic changes (Thomas, 1997; Thornton, 1998).

Furthermore, the transatlantic slave trade caused major population changes. In some regions, the population became mostly African. This can be seen today in the diverse cultural environment and complex racial dynamics in many parts of the Americas. The long-term effects of the slave trade show the lasting impact of this dark period in history (Eltis, 1999; Lovejoy, 2011).

- Voluntarily migration

European emigration to the colonies often occurred during colonization because of new opportunities for work, arable lands, precious metals in the region, and other individual reasons, such as escaping from religious pressures. Also, colonial European powers actively recruited miners, engineers, administrators, and bureaucratic staff from their home countries to supervise, lead, and monitor the sophisticated mining processes. This influx of European immigrants changed the demographics of the colonies. (Canny,

1994) Moreover, internal migration was further stimulated as residents from other regions of Latin America relocated there in pursuit of employment opportunities (Brown, 2012). Settlers from Europe usually came from densely populated places to less densely populated areas, while merchants usually located in the previously dense areas in new regions (Manning & Trimmer, 2020). By the 19th century, economic prospects, poverty, and land scarcity were significant drivers of large-scale global migrations. Although the slave trade was abolished, Africans continued to be moved, though it was small in numbers compared to the millions from Europe and Asia. The late 19th century also marked the Age of Mass Migration, when many people from different areas of Eurasia settled in the new regions (Haas et al., 2019).

CHAPTER 4

DATA AND VARIABLES

4.1 World Migration Matrix

World Migration Matrix (WMM or ancestral matrix afterward) was introduced by Putterman and Weil (2010), and it presents a detailed breakdown of the ancestral composition of the populations of 172 countries worldwide[1]. The main goal of the WMM is to indicate the ratios of countries in which the ancestors of today's residents lived in 1500 CE. Additionally, a newly developed matrix, created specifically for the period between 1500 and 1960 by the authors, has been used as the primary data source for this study (Chanda et al., 2014) . The reason for using this matrix is that it covers the period known as the colonial era and discards the effects and potential bias of contemporary migrations after 60s. Nevertheless, the analyses conducted using the 1500-2000 matrix have been included in the appendix as robustness checks.

The conceptual point emphasized by the authors during this matrix and their subsequent works is that focusing solely on countries as the unit of analysis in studying the long-term impact of factors such as geography and culture on the economy may lead to flawed results. With the development of this matrix, authors propose an "ancestry adjustment" in future studies. The use of this matrix in this fashion, as a robustness check or as an adjustment for ancestry in the preliminary results, has been frequent in the long-run effects literature.

[1] Only 163 countries are suitable for our study.

Additionally, the authors justified the choice of 1500 CE as the starting point by highlighting the great population movements after 1500 CE. The selected time frame encompasses numerous pivotal historical events that have significantly impacted global population movements; in history literature, the year 1500 is also considered a good starting point for periodization (Hoerder, 2002; Lucassen & Lucassen, 2009; Manning & Trimmer, 2020).

While preparing this matrix, they have utilized diverse primary and secondary sources, which the authors have explained in detail. Additionally, as mentioned earlier, this dataset has been frequently utilized as an adjusting factor in numerous high-level research projects and has received considerable acclaim.

On top of these, two findings are available to test the performance of the matrix, one derived from the authors and the other from me. Initially, as the authors claimed in the article, there is a strong correlation of 0.89 between their measurement of the European fraction and the European share measure from Acemoglu et al. (2002). To further assess the matrix's performance, I developed a new "Euro share" metric, which represents each country's proportion of European origin by aggregating the shares of properties from European countries. Compared with the "Euro share" variable developed by Easterly and Levine (2016) this new metric revealed a strong correlation of 0.70.

It is a 172 x 172 matrix in which country names are used as both the row and column titles; a small portion of this matrix is depicted in Figure 1 (Appendix A) . The numbers in each cell are estimates of the percentage of ancestors from the nation listed in the row heading who were residing in the country in the column heading in 1500. For clarity, if we focus on Turkey as a row and find a 0.072 value under the Ukraine column, it suggests that 7,2 % of contemporary Turkey's inhabitants have ancestral roots in areas

corresponding to modern-day Ukraine as of 1500 CE. This data exhibits migratory movements from Ukraine to Turkey over the subsequent centuries. As mentioned above, there are 163 countries available for use. These countries are divided into 9 different regions and conduct regional analyses accordingly later. The number of countries by region can be seen in Table 1 (Appendix B) . It should be acknowledged that the WMM does not only reflect migration but also encompasses various influences such as historical accidents, wars, genocides, natural disasters, and diseases, leading to different mortality and fertility rates across groups. Simplifying the data to just migration between points may be misleading. Despite these complexities, it is crucial to identify deep-rooted factors shaping ancestral compositions. This thesis aimed to show that certain factors remain significant despite various influences. Through controlled and regional analyses, we demonstrated that these key factors consistently impact migration patterns and population compositions.

4.2 Dependent variables: EuroShare, AsiaShare, and SSAshare

For the first part of the data, three variables are created using the World Migration Model (WMM), and the structure of the data is cross-country. *EuroShare* measures the proportion of individuals of European descent in the population. Similarly, *AsiaShare* measures the proportion of people of Asian descent, and *SSAshare* measures the proportion of individuals of Sub-Saharan descent. Each variable provides insights into the migrations from these specific regions to other countries across the world. Below, the mean of each dependent variable can be seen across the regions.

DS and DSv2: Two new dependent variables are created for the second part of the analysis, and the data structure has been changed to dyadic. WMM for 1960 is used to create these variables for primary analyses. The relationship between all countries in the matrix was prepared with a bilateral structure for this thesis, also called the origin-destination structure (LeSage & Pace, 2010). This led to a substantial increase in observations and explanatory power. The initial step to this was to convert the WMM into a "long" format. This transformation allows for bilaterally analyzing data, with each country represented 163 times as both destination and source.

For instance, we can observe population movement from Greece to Afghanistan and Afghanistan to Greece separately. This format provides a comprehensive data set for examining all migration relationships, consisting of 26.569 observations, equal to 163 times 163, as expected. After this step, the “country” columns were renamed 'Source Country' and 'Destination Country'. Subsequently, the numbers in the appropriate places on the matrix began to denote the relationship between these two countries. The results equal to the matrix values were named *'Descendent Share (DS).*'

However, analyzing only the *'Descendent Share'* might not give us the whole picture and could lead to skewed results because it needs to consider how big or small countries are. For example, if Angola is 0.03 of Brazil and Iraq is 0.03 of Turkey, “Angola to Brazil” and “Iraq to Turkey” would have the same in the *Descendent Share*. However, we know that these countries have different populations and sizes. So, we need to adjust the numbers to take those differences into account. To address this issue, we needed a variable weighted by population sizes. In creating this second variable, the first step was multiplying each value in the matrix by the year 1960 population figures of the destination countries. This hypothetical calculation suggests that, for example, the

ancestors of 17 million people living in modern-day Afghanistan were also residing there in the year 1500. After this step, the sum of these populations is given as a column base to the "All Descendants." For instance, this gives us the number of people around the world today who are descendants of those who lived in Afghanistan in the past. Following this procedure, the calculation of "DS version 2" or "DSv2" for each pair of country interactions (such as ALB-AFG, AFG-ALB, etc.) involves dividing the multiplication of each country's year 1960 population with the corresponding matrix values by the 'All Descendants' figure, which has been calculated based on the destination country. This computation was carried out across the entire matrix.
Here is the equation that represents this calculation:

$$DS^{ij}_{version2} = \frac{Pop^{i}_{1960} \times Matrix\ Value^{ij}}{All\ Descendants^{j}}$$

In this equation:

- Let $DS^{ij}_{version2}$ represent the DSversion2 (DSv2) value for a pair of countries i and j, where i is the destination country and j is the source country.
- Pop^{i}_{1960} denotes the population of the destination country i in the year 1960.
- $MatrixValue^{ij}$ corresponds to the matrix value for the country pair i and j.
- $All\ Descendents^{j}$ is the aggregate number of descendants for the source country j, as calculated from the matrix.

After completing this matrix-level operation, we convert the data into the long format. *DSv2* improves upon the original *DS* by considering how big or small countries' populations are. The original *DS* might tell how many people from one country live in

another, but *DSv2* goes a step further by looking at these numbers to see the total number of people who lived in the source country in the year 1960. Most of the analyses, *DSv2* are multiplied with 100 first then used in the analyses for editorial purposes. This way, it gives a clearer picture that considers the size of each country, which is especially helpful when comparing larger countries to smaller ones. Additionally, measuring the share that a specific country receives from the global distribution of its citizens proves to be more beneficial for our global analysis. Nonetheless, the analyses made with the *DS* variable will also be used for robustness checks; also, *DS* and *DSv2* are created with the original matrix and used for the same analyses as mentioned above.Since the fundamental dynamic of the study is migration to the New World countries discovered during the colonial period, the proportion of people who migrated from other continents to these countries is significant. Several figures in the Appendix A show the breakdowns for *DS* and *DSv2* for the Americas.

4.3 Independent variables

4.3.1 Historical factors

One of the factors extensively discussed in the long-run effects literature is the characteristics of the population in the destination regions, especially considering the colonial era in which European colonists migrated worldwide. The existing composition of the people in those areas significantly influenced the settlement decisions of the arriving populations and, indirectly, the formation of the relationship networks that facilitated the emergence of institutions (Acemoglu et al., 2001, 2002). The local population features, mortality rates, urbanization levels, historical background of

governance, and technological advancements were influential factors that significantly shaped people's decisions not only during the colonial era but throughout the entire period under study.

1- Population density

Population Density at 1500 CE is a commonly used indicator prepared by McEvedy and Jones (1975). The rationale behind the population density is that it can be argued that places with a denser population throughout history will be more advanced in development. Population Density at 1500 CE is widely used in the economics literature (Glaeser et al., 2004; Ashraf & Galor, 2011; Easterly & Levine, 2016). To noted that, *Urbanization Rate* is also an indicator of the population in 1500 CE living in an urban area with a population minimum of 5,000. It is claimed to be representative of the economic development of a society. This indicator was first developed and used by Acemoglu et al. (2002). Following this article, *Urbanization Rate* is often incorporated in the persistence literature. However, several studies criticize it by stating that it does not have sufficient indicator power. Chanda et al. (2014) claims that urbanization data for 1500 is only available for a minimal sample set and only for 41 countries. Also Ashraf and Galor (2011) argues that this variable is only a proxy for technological advancement in society. Therefore, it is not used in this study.

2- State history

The state is one of the most critical factors for economic development. Countries' recent economic performance can be traced back to important decisions or events in history, which led to the foundation of several institutions (North, 1990; Acemoglu et al., 2005; Besley & Persson, 2009). A state-related indicator can be quite helpful in determining the motivators of population movements, considering the degree of the

state's impact on its citizens. "*State History at 1500 variable*", prepared by Borcan et al. (2018, 2021), is added to my dataset, which covers a sample of 152 countries in my dataset. This variable exclusively indicates the level of development up to the year 1500. It has been used as an alternative variable in robustness checks in place of population density at 1500.

3- Colonial relationship and European colonial origin indicators

Colonial history is an important factor for the analyses. It is essential to use judgment to determine whether countries should be considered to have experienced colonization. There are also ongoing discussions about what conditions a nation should be classified as a colony in the literature, too. As a result, various colonial dummies have been used in several economic research studies (Acemoglu et al., 2001; Banerjee & Iyer, 2005; Ertan et al., 2016). The main colony dummy is taken from Olsson (2009), and its name is *colony*. On top of that, due to the bilateral structure of our data, bilateral variables can also be utilized; therefore, the "*Colonial Dependency Ever*" variable, taken from (Conte et al., 2022), and prepared by Head et al., (2010), is used. Also, European colonial origin indicators are taken from Nunn & Puga (2012). These dummies indicate which specific European country colonized each nation. These two variables have been checked, and some inconsistencies between them have been resolved. Due to the potential for creating endogeneity, estimations have also been conducted in some baseline analyses without considering colonies.

4- Slave exports and slave dummy

Slavery throughout the Colonial Era is one the most tragic and vital concepts to discuss, which is a far extent of this study. In the economics literature, it is claimed that forced population movement due to the enslavement of the people in Africa had

persistent effects on source and destination countries Nunn (2008). The *slave exports* variable is taken from Nunn & Puga (2012). As they stated, this variable is based on Nunn (2008)'s estimates of the total number of slaves traded in Africa's four slave trades between 1400 and 1900. Additionally, using the slave exports variable, a *slave* dummy is created to use in regressions as an alternative. This variable takes a value of 1 if *slave exports* is higher than 0; otherwise, it takes a value of 0, which measures whether there are any enslaved people or not.

5- Precious metals

The search for valuable minerals has been widely acknowledged as one of the significant driving forces behind European navigations and settlements during the Colonial Era. Also, several "gold rushes" occurred after the exploration of substantial gold mines and other valuable sources. For instance, towards the end of the 19th century, several regions, such as Australia, the United States, New Zealand, Brazil, South Africa, China, and Canada, became highly popular destinations for immigrants seeking employment in mines. So, adding precious metal dummies, which demonstrates whether there is a valuable natural source within the country border, would be beneficial for analyzing population movement. The Precious Metals dummy taken by Easterly and Levine (2016) analyzes these motivations.

4.3.2 Geographical variables

Geographical factors, including latitude and climate, strongly correlate with development, even though the precise interpretation of this correlation continues to be a subject of debate (Acemoglu et al., 2005; Ashraf & Galor, 2011; Spolaore & Wacziarg, 2013). Therefore, when assessing geographic factors, it is crucial to exercise caution regarding their potential channels of influence and their endogenous relationship with other variables.

1- Latitude

Latitude is claimed as a highly explanatory variable that has been frequently utilized in the "persistence literature" (Ashraf & Galor, 2013; Spolaore & Wacziarg, 2013; Ertan et al., 2016; Michalopoulos & Papaioannou, 2016). This variable measures the distance from the equator, and its absolute version is more frequently used in the literature due to its enhanced descriptive power. Regions at similar latitudes usually have similar temperatures, making it possible to grow the same kinds of crops and create a familiar climate. Also, several latitude regions were more likely to cultivate plantation crops, which was highly critical in the colonial era (Engerman & Sokoloff, 2002; Easterly & Levine, 2003). Consequently, it is anticipated that this factor will significantly influence population movements. Additionally, to utilize the origin-destination setup of the data, the differences in absolute latitudes are calculated as a bilateral variable, which is called *abslatdiff*. The aim here is that this difference could approximately serve as a proxy for climate differences.

2- Malaria ecology and EDE

The literature has extensively discussed the influence of diseases on population movements and economic development. Therefore, diseases in the countries are essential for understanding the driving forces behind human mobility. A recent variable, *malaria ecology,* which is taken from Erhan and Putterman (2016), is used in this study. The biological characteristics of mosquitoes involved in malaria transmission are evaluated by the malaria ecology index, which was created by Kiszewski et al. (2004).

The Early Disease Environment (*EDE*) indicator, which was created by Auer (2013), is used as an alternative to *malaria ecology*. It has a broader scope. To predict the "logarithm of the annualized probability of death for European males within the age cohort of soldiers" this measure utilizes data on non-combat soldier mortality rates during the colonial era, supplemented by various climate-related variables.

3- Percentage of arable lands

arable is taken from Ashraf and Galor (2013). This variable is important for the research because agriculture has played a crucial role in production throughout history. Regions with more arable land have likely attracted more people due to their suitability for farming and agricultural activities, which are essential for sustaining large populations and attracting new people. Thus, including arable land as a variable in the analyses provides valuable insights into the historical factors that influenced human mobility and settlement patterns.

4- Distance and contiguity

Bilateral distance is added to the data as a geographic control variable because it is a natural factor affecting population movements. To capture this relation, bilateral *dist* and *lndist* that are taken from (Conte et al., 2022), which is known as the *CEPII Gravity*

Database, are included in the data. Distance is expected to have an impact on migration but including it in regressions will be important for the accurate functioning of other variables. Additionally, the contiguity dummy variable, which measures whether countries are neighboring and taken from (Conte et al., 2022), will also be used in the regressions. Also, *Distance to the UK* and *Distance to China* variables, taken from Spolaore & Wacziarg (2013), are also used in several estimations.

5- Biogeography

Biogeography is a variable created by Hibbs and Olsson (2004). It shows the geographic dispersion of farming crops that can be stored for long periods and animals that can be domesticated. This variable is critical because it indicates how suitable different regions are for agriculture and animal husbandry throughout history. Regions with higher *Biogeography* values have better environments for farming and raising animals, making them more attractive for human settlement and development. Similar to the arable land percentage, storing crops and domesticating animals has been crucial for sustaining large populations and attracting migrations. This variable is used as a control on top of the arable because the variable's sample size is very small compared to the *arable*.

6- Elevation

Elevation is an important factor in this because it directly affects temperature, living conditions, and social structures. Higher elevations tend to have cooler temperatures, which can influence the types of crops that can be grown and the kinds of animals that can be raised, which is pretty similar to the *arable* and *biogeography* variables. These environmental conditions, in turn, affect the suitability of a region for human habitation. Elevation also has a relation with the availability of water and the

types of diseases that are prevalent in an area. This variable is taken from Ashraf and Galor (2013). This variable will be used as a geographical control variable and the correlations with the *arable* and *biogeography* will be considered carefully while doing the estimations.

7- Distance to waterways

The variable represents the average distance, measured in thousands of kilometers to the nearest ice-free coast or navigable river accessible to ships. This information is important because being close to waters has always been crucial for trade, transportation, and colonial activities. Regions near the water could easily engage in commerce and transport resources and people, all of which make them more attractive for settlement and migration. The variable is taken from Ashraf and Galor (2013) and will be used as a geographical control variable.

8- Island, small and landlocked regions:

They are the conventional landlocked, small, and island dummy variables. They are important for this study because they represent the isolation of countries to some extent, and isolation is critical for economic development, migration, and many other historical cases, such as indigenous mortality (Easterly & Levine, 2016).

9- Temperature and precipitation:

Temperature measures the spatial average of the intertemporal monthly mean temperatures across the grid cells within a country. Additionally, to utilize the origin-destination setup of the data, the differences in absolute temperatures are calculated as a bilateral variable, which is called *abstempdiff*. *Precipitation* measures the spatial average of the intertemporal monthly precipitation across the grid cells within a country. These variables are taken from Ashraf and Galor (2013).

10- Ruggedness:

Ruggedness, taken from Nunn & Puga (2012), is used to analyze how geographic features influence economic and social outcomes, particularly in historical contexts. That's why it is used as a additional geographic control in this study.

11- Suitavg:

Suitavg is taken from Ashraf and Galor (2013). It is used to evaluate whether a piece of land is suitable for farming, considering several ecological factors of climate suitability.

12- Settler mortality:

Settler Mortality was developed by (Acemoglu et al., 2001) using the data on what the Europeans would have experienced if they had attempted to settle in the same area 150 years earlier. It is one of the most famous and most disputed variables in the literature, due to disputes and the limited sample set (available only for 64 countries), this variable will only be used as an additional control for several estimations.

In Appendix B, descriptive statistics for unilateral and bilateral versions can be seen for variables.

CHAPTER 5

METHODOLOGY AND ANALYSES

5.1 Cross-sectional unilateral analyses with regional shares

5.1.1 Descriptive statistics

The first part of the study includes cross-sectional analyses. The dependent variables are created for this thesis using DS variables from WMM. As explained before, three variables were created for this analysis: *EuroShare*, *AsiaShare*, and *SSAshare,* each representing the total proportion of people from these specific regions in a country. In other words, dependent variable *EuroShare* measures the proportion of individuals of European descent in the population of the country and these variables serve as a proxy for migration and can be used in future studies alongside Euroshare.

The Table 2 in the Appendix D presents the basic descriptive statistics for these dependent variables, these regions are set according to the literature (Ashraf & Galor, 2011).

The analyses with "European Share" variable are similar to the analyses in Acemoglu et al. (2002) and Easterly and Levine (2016). On top of these, this study uniquely adds analysis for Asian and Sub-Saharan African ancestry composition analysis for all countries. In other words, it helps us see how our deep-rooted variables affect the migrations not only from Europe, but also from Asia, and Sub-Saharan Africa to other parts of the world.

5.1.2 Regression Analyses

The initial analyses will be conducted using Ordinary Least Squares (OLS).

Euroshare:

Firstly, dependent variable *EuroShare* measures the proportion of individuals of European descent in the population of the country. Following three variables are backbone of this study. The initial variable *"Log Population Density at 1500*" is a proxy for the early economic development and settlements. In these estimations it will offer insights into how historical development have influenced current population conditions. Secondly, *Malaria ecology*, representing the ecological suitability for malaria, serves as a proxy for the historical disease environment. It is highly critical for the settlement procedures during the colonial era. Finally, the *"Percentage of Arable Land"* variable indicates the proportion of land that can be cultivated, agricultural production is highly critical during the period for European migrants; therefore, this variable is included in the estimations.

"*Average Elevation"* controls for geographic attributes that could influence economic activities, particularly agriculture potential, disease environment and accessibility of the region. *"Average Distance to Waterways"* is also added as a geographical control. Regions near the water could easily engage in commerce and transport resources and people. Finally, it should be considered that only non-European countries have been included in the analysis because otherwise, *"European Share"* in the European countries will lead to a significant bias in the estimations.

The first estimation results are displayed in Table 5 (Appendix D). The negative and statistically significant coefficients of *"Log Population Density at 1500*"

(*ln_pd1500)* in all three models indicate that higher population density in 1500 is associated with a lower proportion of individuals of European descent today. This suggests that areas with higher initial population densities were less attractive to European settlers, possibly due to the presence of established local populations and infrastructure. *Malaria ecology*, the second variable, has negative coefficients, significant at the 1% and 5% levels, show that higher ecological suitability for malaria correlates with a lower *EuroShare*. This finding aligns with historical evidence that Europeans tended to avoid areas with high malaria risk, which would have posed significant health challenges during the colonial era. The third variable, *"Percentage of Arable Land" (arable)* have positive and statistically significant coefficients suggest that a higher percentage of arable land is associated with a higher proportion of individuals of European descent. This underscores the importance of agricultural suitability in attracting European settlers, as fertile land would have been crucial for sustaining colonial economies and supporting migrant populations.

The three estimations are presented in the table. Column 1 includes the basic variables without any additional controls, achieving an R-squared of 0.399, indicating that about 39.9% of the variance in EuroShare is explained by the model. Column 2 adds geographical controls, such as average elevation, landlock and island dummies, and average distance to the nearest ice-free coast, improving the R-squared to 0.485. Column 3 further includes colonial dummies, which indicate whether a country was colonized by Great Britain, France, Spain, Portugal, or other European powers, , improving the R-squared to 0.669.

Then, more control variables in the literature added to the estimation and controlled. First, *Absolute Latitude* is added, frequently used in the literature as an

climate proxy, Also, it can be used to assess the similarity between European and destination country in aspect of agriculture. The *Presence of Gold and Silver* is a dummy variable, indicating the presence of gold or silver, highlighting the role of natural resource endowments in shaping the attractiveness for colonization of the state and migration flows. *Biogeography* encompasses the effects of the historical environmental and ecological conditions that have historically influenced agricultural productivity. However, it should be noted that adding this variable decreases the sample size. Finally, *Settler Mortality* is also controlled. The estimation results are displayed in Table 6 (Appendix D).

In Column 2, absolute latitude is added to the variables in Model 1. The results remain largely unchanged, with *ln_pd1500, malaria_ecology, and arable* retaining their significance and coefficients. The inclusion of *absolute latitude*, which is insignificant, does not significantly alter the model's explanatory power, as reflected in the R-squared value of 0.486. Column 3 adds the presence of gold and silver to the previous variables. While the coefficients for *ln_pd1500, malaria ecology, and arable* remain significant, the *presence of gold and silver* is not significant. This addition slightly improves the R-squared to 0.505, indicating a minor increase in explanatory power. *Biogeography* is included in Column 4, along with the previous variables. The results show that *ln_pd1500, malaria ecology, and arable* are still significant, while *biogeography* is not. The R-squared value increases to 0.665, suggesting that biogeographical factors, though not significant individually, contribute to explaining the variation in *EuroShare* when included in the model. *Settler Mortality* is added in Column 5. The results indicate that *ln_pd1500, malaria_ecology, and arable* remain significant, while settler mortality is not. This model achieves the highest R-squared value of 0.727, indicating that 72.7% of

the variation in *EuroShare* is explained by the model. This significant increase demonstrates the importance of considering settler mortality in understanding the distribution of populations of European descent. Column 6 includes region fixed effects, further controlling for unobserved regional characteristics that might influence settlement patterns. The coefficients for *ln_pd1500, malaria ecology, and arable* remain significant, while *absolute latitude, presence of gold and silver, biogeography,* and *settler mortality* are not. The R-squared value increases to 0.765, highlighting the added explanatory power of controlling for regional fixed effects. Finally, Column 7 includes colonial dummies, indicating whether a country was colonized by major European powers. The results show that *ln_pd1500, malaria ecology, and arable* are still significant, while the other variables remain insignificant. The R-squared value is 0.794, the highest among all models. The significant role of early economic development, disease environment, agricultural suitability, and colonial history is evident. The inclusion of geographical and regional controls further enhances the models' explanatory power, providing a nuanced and robust understanding of the factors influencing the distribution of populations of European descent in non-European countries.

SSAshare:

The dependent variable, *SSAshare*, quantifies the proportion of individuals of Sub-Saharan African descent within the population. Same independent variables are controlled for the baseline estimations. The estimation results are displayed in Table 7 (Appendix D). Column 1 includes the basic set of independent variables without any additional controls. The results show that *ln_pd1500* has a negative and statistically significant coefficient at the 5% level, suggesting that higher population density in 1500

is associated with a lower proportion of individuals of Sub-Saharan African descent. *malaria ecology* has a negative but not significant coefficient, while *arable* has a positive but not significant coefficient. The constant term is positive and significant at the 10% level. In Column 2, geographical controls are added, including average elevation, landlock and island dummies, and average distance to the nearest ice-free coast. The results show an increase in the explanatory power of the model, with an R-squared value of 0.289. The coefficient for *ln pd1500* becomes more negative and remains significant at the 5% level, indicating a stronger relationship. *malaria ecology* becomes significant at the 10% level, suggesting that higher malaria suitability negatively affects SSAshare. The coefficient for *arable* becomes positive and significant at the 10% level, indicating that a higher percentage of arable land is associated with a higher proportion of individuals of Sub-Saharan African descent. Column 3 includes both geographical controls and colonial dummies, indicating whether a country was colonized by major European powers such as Great Britain, France, Spain, Portugal, or other European powers. The R-squared value increases to 0.324, indicating that 32.4% of the variation in SSAshare is explained by this model. The coefficient for *ln_pd1500* remains negative and significant, although at the 10% level. *malaria_ecology* remains negative but not significant. The coefficient for *arable* remains positive and significant at the 10% level. The constant term becomes negative and insignificant. The results highlight the importance of historical and geographical factors in explaining the proportion of individuals of Sub-Saharan African descent in non-Sub-Saharan African countries. Higher population density in 1500 consistently shows a negative relationship with SSAshare, indicating that more developed and densely populated areas in 1500 are associated with a lower proportion of Sub-Saharan African descent today. The

ecological suitability for malaria has a negative impact on SSAshare, particularly when geographical controls are included. The percentage of arable land shows a positive relationship with SSAshare, suggesting that agricultural suitability attracts settlement. Including geographical and colonial controls improves the explanatory power of the models, providing a more comprehensive understanding of the factors influencing the distribution of populations of Sub-Saharan African descent in non-Sub-Saharan African countries.

Then, more control variables in the literature added to the estimation and controlled. First, *absolute latitude* and The *Presence of Gold and Silver* are added like the previous estimation with *EuroShare*, then *Slave* dummy is controlled. The estimation results are displayed in Table 8 (Appendix D). Column 1 includes *ln_pd1500*, *malaria ecology*, and *arable*, with geographical controls such as average elevation, landlock and island dummies, and average distance to the nearest ice-free coast. The results show that *ln_pd1500* is negative and significant at the 5% level, indicating that higher population density in 1500 is associated with a lower SSAshare. *Malaria ecology* is negative and significant at the 10% level, suggesting higher malaria suitability negatively impacts SSAshare. *Arable* is positive and significant at the 10% level, indicating that more arable land is associated with higher SSAshare. Column 2 adds *absolute latitude*, which is negative and significant at the 5% level, indicating that higher latitude (further from the equator) is associated with a lower SSAshare. The coefficients for *ln_pd1500, malaria ecology, and arable* remain significant and similar in magnitude. The R-squared value increases to 0.330. In Column 3, the presence of gold and silver is added, but it is not significant. The other variables retain their significance and similar coefficients. The R-squared value is 0.290. Column 4 introduces the slave variable, which is negative and

significant at the 10% level, indicating that countries that have sent slaves have a lower SSAshare. The coefficients for *ln_pd1500, malaria_ecology, and arable* remain significant. The R-squared value is 0.336. Column 5 adds colonial dummies, indicating whether a country was colonized by major European powers. The results show that *ln_pd1500* remains negative and significant, while *malaria ecology* and *arable* are no longer significant. The *slave* variable remains negative and significant. Column 6 includes region fixed effects to control for unobserved regional characteristics. In this estimation, only *Percentage of Arable Land* stays significant and positive.

The changes in significance value of the *malaria ecology* variable for SSAshare may have raised doubts about whether using an alternative variable might be more efficient. As a robustness check, the *EDE* variable was used, can be seen at the end of the chapter and Appendix D. When this variable was employed, a positive relationship between the EDE variable and SSAshare is observed. This may imply that more Africans migrated to more disease-prone areas, suggesting that historical circumstances led them to these regions. However, conversely, both malaria ecology and EDE have a negative impact on EuroShare. This indicates that historical European migrations occurred less frequently in disease-prone environments. Another conclusion that can be drawn from this is the benefit of using regional categories when analyzing historical migrations, as a global control would not have revealed the positive effect on Africans observed in this study.

5.1.3 Spatial ML analyses

Tobler (1970) emphasizes that no region exists in isolation, and population movements further enhance the relationship between regions. Considering the expected spatial spillover effects of migration and other variables under consideration in this study, it is better to utilize a spatial econometric approach for empirical analyses.

Below the general framework for the spatial analyses can be seen. This complete model tends to be too complex, so many studies use simpler versions (Elhorst, 2014).

General Framework

$$y = \rho W y + \alpha \iota + X \beta + W X \theta + \varepsilon$$
$$\mu = \lambda W \mu + \varepsilon \qquad (1)$$

Here W stands for the weighting matrix. In spatial econometrics, weights are important for quantifying the relationships between different units. To create this weighting matrix, inverse-weighting is used in all these models which assigns weights based on the inverse of the distance between locations.

The process of formulating and testing different models by assuming certain coefficients to be zero is referred to as "filtering." by Anselin (1988). This method entails continuously evaluating and adjusting models by applying specific coefficient constraints to determine and select the most suitable model specifications. The spatial econometric model is typically structured in six fundamental forms. The first type is endogenous interaction effects, where an area's outcome depends on its neighbors' outcomes (independent variable), and this is analyzed using the spatial autoregressive

model (SAR). The second type involves interaction effects within the error term, analyzed with the spatial error model (SEM). The Spatial Lag of X (SLX) model captures exogenous interaction effects, where an area's outcome is influenced by the independent variables of neighboring areas. This model is used to assess how changes in explanatory variables in neighboring regions affect the dependent variable in a given region. The next type is exogenous interaction effects, where an area's outcome is influenced by both the dependent and independent variables of neighboring areas, and this is analyzed using the spatial Durbin model (SDM) (Elhorst, 2014). Spatial Autocorrelation Coefficients (SAC) model captures both endogenous interaction effects and spatial autocorrelation in the error terms. This model is used to analyze how the dependent variable of neighboring areas and the spatially correlated error terms affect the dependent variable in a given region. Finally, Spatial Durbin Error Model (SDEM) is an extension of the SLX model and the spatial error model (SEM). It includes both exogenous interaction effects and spatial autocorrelation in the error terms. This model helps to analyze how the independent variables of neighboring areas and the spatially correlated error terms influence the dependent variable.

The Spatial Durbin Model (SDM) is increasingly popular in recent studies and can be adapted into the Spatial Autoregressive Model (SAR) and the Spatial Error Model (SEM) under certain conditions, making it a more general form of spatial econometric model (Elhorst, 2014). Estimation will be carried out using all these models; however, according to Moran's I and interpretations in the literature, the results of the SDM, SAR, and SDEM models are more significant for this study. Here are the cross-section formulations of these models:

Spatial autoregressive model (SAR)

$$y = \rho W y + X \beta + \varepsilon \quad (2)$$

Spatial error model (SEM)

$$y = X\beta + \mu$$

$$\mu = \lambda W \mu + \varepsilon \quad (3)$$

Spatial Durbin Model (SDM)

$$y = \rho W y + \alpha \iota + X \beta + W X \theta + \varepsilon \quad (4)$$

Spatial Lag of X (SLX)

$$y = X\beta + WX\theta + \varepsilon \quad (5)$$

Spatial Durbin Error Model (SDEM)

$$y = X\beta + WX\theta + \mu$$

$$\mu = \lambda W \mu + \varepsilon \quad (6)$$

Spatial Autocorrelation Coefficients (SAC) Model

$$y = \rho W y + X\beta + \mu$$

$$\mu = \lambda W \mu + \varepsilon \quad (7)$$

Moran's I test for spatial autocorrelation:

To better understand the spatial structure of the variables, first "Moran's I" test for spatial autocorrelation is made, which measures how much correlation exists within regions.

Below is the mathematical formula for Moran's I (Moran, 1950; Anselin, 1995):

$$I=N/W_{total}*(\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}(x_i - \bar{x})(x_j-\bar{x}))/\sum_{i=1}^{N}(x_i -\bar{x})^2) \qquad (8)$$

Here, x is the standardized interested variable, W is a row-standardized spatial weight matrix, and N denotes the total number of observations. Moran's I ranges from -1 to 1. Positive values indicate positive spatial autocorrelation, while negative values indicate negative spatial autocorrelation or dispersion. The null hypothesis assumes complete randomness.

Additionally, Lagrange Multiplies test results are also shown in the tables, which are developed by Anselin (1995). Thanks to these tests, it can be identified that whether a spatial effect is caused by errors or dependent variables. On top of that, The Robust Lagrange Multiplier (LM) test is a statistical method used to identify the presence of spatial dependence in econometric models. Developed by Anselin (1995), this test helps determine whether spatial lag or spatial error models are more appropriate for the data. Also, plots for Moran's I are also shown. The points in the positive-positive and negative-negative ,which are also called as high-high and low-low, are sign of spatial autocorrelation. As can be seen from Table 9 (Appendix D) and Figure 9, 10 and 11 (Appendix C) , the variables *EuroShare*, *AsiaShare*, and *SSAshare* exhibit significant spatial autocorrelation, though it's important to note that it is comparatively low for *SSAshare* and *AsiaShare*. Also, spatial analyses should be made for the independent variables because they have significant and positive Moran I result. In the literature, Maximum Likelihood and Generalized Spatial Two Stage Least Squares methods are popular for spatial analyses (Elhorst, 2014), which will be used as a robustness check.

- *Spatial analyses for EuroShare*

Table 10 (Appendix D) presents the results from six spatial econometric models (SEM, SAR, SLX, SAC, SDM, SDEM) estimations for the "*EuroShare*" variable. It shows that "Ln Population Density at 1500" and "Malaria Ecology" have a consistently negative and statistically significant effect on EuroShare across all models, while the "Percentage of Arable Lands" has a positive and significant effect. The constant term is positive and significant in each model. All models include geographical controls, and the mean value of EuroShare is 0.321 across 105 observations, highlighting robust findings across different spatial econometric approaches.

- *Spatial analyses for SSAshare*

Table 11 (Appendix D) presents the results from six different spatial econometric models analyzing the dependent variable "SSAshare," which represents the share in Sub-Saharan Africa. The independent variables included in these models are "Ln Population Density at 1500," "Malaria Ecology," and "Percentage of Arable Lands." The coefficients for "Ln Population Density at 1500" are generally negative and statistically significant in the SAR, SLX, SAC, SDM, and SDEM models, suggesting that regions with higher population density in 1500 tend to have a lower SSAshare today. "Malaria Ecology" shows a negative coefficient across all models, but these are not statistically significant, indicating no strong evidence of its impact on SSAshare. The "Percentage of Arable Lands" is positively associated with SSAshare, with statistically significant coefficients in the SAR and SAC models, implying that regions with more arable land have a higher SSAshare.

The OLS results in the table indicate significant relationships between historical population density, malaria ecology, percentage of arable lands, and EuroShare, but these estimates can be biased and inefficient without considering spatial interactions. By integrating spatial components, these advanced models provide a more accurate and comprehensive understanding of how geographical and spatial factors influence the dependent variable, leading to more robust and insightful conclusions. Along with these controls, it was shown that the effects of the main three variables remain strong, consistent in direction across different estimations, and significant.

5.1.4 Robustness checks

- Alternative estimation method (the Generalized Spatial Two-Stage Least-Squares Estimator) for spatial analyses: According to the Table 12 and 13 (Appendix D) Results stay similar for most of the estimations.
- Different regional settings for OLS and Spatial Analyses: When Only Non-Eurasia countries are taken for EuroShare, results stay similar for all estimations which can be seen in Table 14 (Appendix D). When only “New Lands” are taken for EuroShare, "Ln Population Density at 1500" and “Malaria Ecology” stay similar for most estimations, but “Percentage of Arable Lands” is not significant, Table 15 (Appendix D). Finally, when only “New Lands” sample used for SSAshare, “malaria ecology” and EDE are significant and positive, also “Percentage of Arable Lands” is significant and positive for most of the estimations, while "Ln Population Density at 1500" is negative and significant Table 16 (Appendix D).

- Estimations with alternative independent variables: Similar results. Table 17 and Table 18 (Appendix D).
- Estimations with Additional Controls: Results are similar. Can be seen in Table 19 and Table 20 (Appendix D). And spatial estimations with Colony Dummies yields similar results. Can be seen in Table 21 and Table 22 (Appendix D).
- Estimations with WMM (year 2000): Results are similar. Can be seen in Table 23 and Table 24 (Appendix D).
- Estimations with ”AsiaShare”: Only malaria ecology is negative for several estimations. Can be seen in Table 25 (Appendix D).

5.2 Bilateral analyses

5.2.1 The origin-destination format and the Gravity Model

Conducting bilateral analysis is the backbone of the study and broadens the scope of the study, allowing for more detailed and comprehensive analysis. It enables us to examine both source and bilateral variables, providing a more in-depth understanding of migration flows between regions. Additionally, with this extensive dataset, more detailed regional analyses can be performed. This approach not only enriches the analysis but also provides a platform for exploring various factors that influence migration at both the source and destination levels. Building on the gravity model literature in both trade (Frankel & Romer, 1999; Anderson & van Wincoop, 2003) and migration (Grogger & Hanson, 2011; Beine et al., 2011, 2016) a pseudo-gravity model is established for analyzing bilateral migration:

$$DSv2_{od}=\alpha+\beta_j*X_o+\beta_k*X_d+\beta_m*\gamma_{od}+\varepsilon \quad (9)$$

$DSv2_{od}$ (appears in the text as DSv2 for simplicity) represents the calculated bilateral immigration rate, which reflects the hypothetical rate of immigrants from origin country 'o' to destination country 'd,' considering all descendants globally. X_o represents the origin-specific unilateral variables, which will be included in the estimation, β_j with representing their results individually. Similarly, X_d represents the destination-specific unilateral variables, and β_k represents their results. Finally, γ_{od} denotes the bilateral variables between these two countries. The variables added will vary across estimations and robustness analyses, but the equation will remain the same.

5.2.2 Methodology

In the literature, the Poisson Pseudo Maximum Likelihood (PPML) method, an estimation technique used with the gravity model, has been gaining increasing prominence recently. This method is used primarily to handle datasets where the dependent variable contains a significant proportion of zeros, indicating a sparse matrix (Beine et al., 2016; Silva & Tenreyro, 2006; 2022). Unlike Ordinary Least Squares (OLS), which require logging the dependent variable under the gravity model and hence excludes zero entries, the non-linear nature of PPML allows for the inclusion of these values. Such inclusion is crucial, as omitting zero entries can distort the data and potentially overlook key insights (Nejad & Young, 2016; Arif, 2022).

However, there are approaches like the one used by Michalopoulos (2012) which apply the log (x + ”Delta”) transformation in conjunction with OLS, a method often

referred to as the 'popular fix.' This “delta,” which varies depending on the dataset, is designed to incorporate zero values. Despite its widespread use, there is considerable debate and a growing consensus that such methods can lead to biased outcomes (Bellégo et al., 2022; Cohn et al., 2022). Nonetheless, results are calculated using the log (x+1) transformation and OLS. Secondly, Silva and Tenreyro (2006) noted that parameters from log-linearized models estimated using OLS produce biased estimates of elasticities in heteroskedasticity. Consequently, they recommend the PPML method as a corrective measure. Since then, the adoption of PPML has grown significantly in the literature (Beine et al., 2011; Kaiser et al., 2014; Alesina et al., 2016; Mitnik & Grusky, 2020). Additionally, as Silva and Tenreyro (2022) explained, the PPML is employed to estimate models where the dependent variable is not a count, closely aligning with practices in the trade literature (Beine & Parsons, 2015; Foubert & Ruyssen, 2024).

5.2.3 Main results

Preliminary figures shows that there is a relationship between critical variables and the independent variable. Figures can be seen in the Appendix C. In the bilateral analysis, several clusterings have been performed for both source and destination regions. For example, this includes examining migrations from Eurasia to regions outside Eurasia. The aim here is to uncover the different factors influencing migrations between regions through these groupings. The study has been conducted with historically and geographically logical clustering.

Eurasia to Non-Eurasia migrations

Baseline results:

The Table 26 (Appendix D) presents the results of regressions analyzing the factors influencing migration flows from Eurasia to non-Eurasian regions. The analysis uses two different estimation methods: Ordinary Least Squares (OLS) and Poisson Pseudo Maximum Likelihood (PPML). The dependent variable here is $DSv2_{od}$. The independent variables considered in the models include bilateral distance, destination population density in 1500, destination malaria ecology, and the percentage of arable lands at the destination. The regressions also include geographical controls such as average elevation, landlock and island dummies, and average distance to the nearest ice-free coast in some models.

The coefficients for *"Ln Distance"* are negative and statistically significant across all models, indicating that greater distances between regions reduce migration flows. This effect is more pronounced in the PPML models, where the coefficients are larger (-2.514 and -2.810) compared to the OLS models (-0.331 and -0.322). Similarly, *"the destination population density in 1500"* has a consistently negative and significant effect on migration flows in all models, with a stronger effect observed in the PPML models. The variable for *"destination malaria ecology"* also shows a negative and significant impact, with a much larger effect in the PPML models. Conversely, the *"percentage of arable lands at the destination"* has a positive and significant effect on migration flows, with a greater magnitude in the PPML models. The constant terms are positive and significant in all models, but they are much larger in the PPML models, reflecting a different scale of the dependent variable. These findings are aligned with the

results of unilateral analyses. We could argue that historical factors such as distance, population density in 1500, malaria ecology, and arable land significantly influence migration flows from Eurasia to non-Eurasian regions, with greater distances, malaria and higher historical population densities generally reducing migration, while more arable land increases it. The stronger effects observed in the PPML models highlight the importance of accounting for distributional characteristics in migration data.

Detailed results:

The Table 27 (Appendix D) presents the results of regressions analyzing the factors influencing migration flows from Eurasia to non-Eurasian regions using both Ordinary Least Squares (OLS) and Poisson Pseudo Maximum Likelihood (PPML) methods. The dependent variable is $DSv2_{od}$. The independent variables include bilateral distance, destination population density in 1500, destination malaria ecology, destination percentage of arable lands, absolute temperature difference, colonial dependency, source population density in 1500, and source percentage of arable lands. The analysis incorporates geographical controls and colonial dummies, with some models also including region fixed effects.

The coefficient for *"Ln Distance"* is negative and statistically significant across all models, indicating that greater distances between regions reduce migration flows. This effect is more pronounced in the PPML models, with coefficients ranging from -1.972 to -2.538, compared to the OLS models, where the coefficients range from -0.259 to -0.408. Similarly, the *"Destination Ln Population Density at 1500"* variable has a consistently negative and significant effect on migration flows in all models, suggesting that higher historical population density in destination regions decreases their

attractiveness for migrants. The effect is stronger in the PPML models, with coefficients ranging from -0.708 to -0.832, compared to the OLS models, with coefficients ranging from -0.237 to -0.264.

The *"Destination Malaria Ecology"* variable also shows a negative and significant impact on migration flows in most models, indicating that regions with higher malaria prevalence are less attractive to migrants. This effect is particularly notable in the PPML models, with coefficients ranging from -0.230 to -0.239. The *"Destination Percentage of Arable Lands"* has a positive and significant effect on migration flows in both OLS and PPML models, suggesting that regions with more arable land are more attractive to migrants. The *"Absolute Temperature Difference"* has a negative and significant effect across all models, while *"Colonial Dependency"* shows a positive and significant effect, indicating that former colonies have higher migration flows. The *"Source Ln Population Density at 1500"* has a positive and significant effect in all models, while the *"Source Percentage of Arable Lands"* has a negative but mostly insignificant effect.

The results indicate that greater distances, and higher historical population densities in destination regions reduce migration flows, while more arable land and colonial dependency increase these flows. The stronger effects observed in the PPML models highlight the importance of accounting for distributional characteristics in migration data. Additionally, regions with higher malaria prevalence and greater temperature differences are less attractive to migrants, underscoring the role of environmental factors in shaping migration patterns.

Regional migration analyses

Baseline results:

Next table presents the results of regressions analyzing the factors influencing migration flows from different source regions (Europe, Sub-Saharan Africa (SSA), and Asia) to Non-Eurasian regions. The analysis uses both Ordinary Least Squares (OLS) and Poisson Pseudo Maximum Likelihood (PPML) methods. The independent variable in these regressions is DSv2, and the regressions consider the effects of bilateral distance, destination population density in 1500, destination malaria ecology, and the percentage of arable lands in the destination region. Geographical controls such as average elevation, landlock and island dummies, and average distance to the nearest ice-free coast are included in all models. Results can be seen in The Table 28 (Appendix D).

For migrations from Europe to Non-Eurasian regions, the *"Ln Distance"* variable has a negative and statistically significant effect on migration flows, with the effect being more pronounced in the PPML model (-4.335) compared to the OLS model (-1.344). The *"Destination Ln Population Density at 1500"* also has a negative and significant effect, indicating that higher historical population density in the destination regions decreases their attractiveness for migrants. This effect is stronger in the PPML model (-1.112) than in the OLS model (-0.557). The *"Destination Malaria Ecology"* has a negative and significant effect in both models, particularly in the PPML model (-0.317). The *"Destination Percentage of Arable Lands"* has a positive and significant impact on migration flows, with coefficients of 0.039 in the OLS model and 0.110 in the PPML model. The constant term is large and significant in both models.

For migrations from Sub-Saharan Africa (SSA) to Non-SSA & Non-Eurasia (labeled as in the rest of the study “New Lands”) regions, "Ln Distance" similarly shows a negative and significant effect on migration flows, with a stronger effect in the PPML model (-3.272) compared to the OLS model (-1.887). The "Destination Ln Population Density at 1500" has a negative and significant effect in the OLS model (-0.280) and a smaller but still negative effect in the PPML model (-0.194). Interestingly, *"Destination Malaria Ecology"* has a positive and significant effect in both models, indicating that regions with higher malaria prevalence took more migrants from SSA. The *"Destination Percentage of Arable Lands"* also positively and significantly affects migration flows, with higher coefficients in the PPML model (0.159) compared to the OLS model (0.096). The constant terms in both models are positive and significant.

For migrations from Asia to Non-Eurasian regions, the results show that *"Ln Distance"* has a negative and significant effect only in the PPML model (-1.531). The *"Destination Ln Population Density at 1500"* has a negative and significant effect in both models, particularly in the PPML model (-1.381). The *"Destination Malaria Ecology"* has a negative and significant effect in the PPML model (-0.337) but is not significant in the OLS model. The *"Destination Percentage of Arable Lands"* positively and significantly affects migration flows in both models, with a much higher coefficient in the PPML model (0.299) compared to the OLS model (0.00334). The constant term is significant in the PPML model (3.138) but not in the OLS model.

In summary, the table shows that greater distances generally reduce migration flows across all regions, with stronger effects observed in the PPML models. Higher historical population densities at the destination tend to decrease their attractiveness, while more arable land increases it. Malaria ecology shows varying effects depending on

the region, and colonial dependencies appear to have a significant positive impact on migration flows, particularly in the context of SSA. The PPML models consistently reveal stronger and more nuanced effects, underscoring their importance in analyzing migration data.

Detailed results:

Following tables present the results of multiple regressions analyzing the factors influencing migration flows from different regions to Non-Eurasian regions or New Lands, using both Ordinary Least Squares (OLS) and Poisson Pseudo Maximum Likelihood (PPML) methods. The dependent variable in these regressions is DSv2. The tables consider several variables: bilateral distance (lndist), destination population density in 1500 (d_ln_pd1500), destination malaria ecology (d_malaria_ecology), destination percentage of arable lands (d_arable), absolute temperature difference (abstempdiff), colonial dependency (col_dep_ever), source population density in 1500 (s_ln_pd1500), and source percentage of arable lands (s_arable). Geographical controls are included in all models, with some models also including colonial dummies.

Europe

In the Table 29 (Appendix D) sample is European sources to Non-Eurasian regions. Bilateral distance (lndist) consistently shows a negative and significant effect across all models, indicating that greater distances reduce migration flows. The destination population density in 1500 (d_ln_pd1500) has a negative and significant effect in most models, suggesting that higher historical population density in destination regions decreases their attractiveness for migrants. The destination malaria ecology

(d_malaria_ecology) variable is generally negative and significant, indicating that regions with higher malaria prevalence are less attractive to migrants. An exception is found in column 7, where the coefficient is positive, suggesting a possible region-specific effect. The percentage of arable lands at the destination (d_arable) shows a positive and significant effect across all models, implying that regions with more arable land are more attractive to migrants. The absolute temperature difference (abstempdiff) also has a negative and significant effect in most models, indicating that larger temperature differences reduce migration flows. Colonial dependency (col_dep_ever) is positive and significant in most models, highlighting that former colonies tend to have higher migration flows. The source population density in 1500 (s_ln_pd1500) and source percentage of arable lands (s_arable) have mixed results, with significance in some models but not in others, indicating less consistent effects.

Sub-Saharan Africa

In The Table 30 (Appendix D) sample is SSA regions to *New Lands*. Bilateral distance (lndist) remains negative and significant across all models, reinforcing the idea that greater distances are a barrier to migration. Destination population density in 1500 (d_ln_pd1500) is again negative and significant in most models, indicating a consistent trend that regions with higher historical populations are less attractive destinations. Destination malaria ecology (d_malaria_ecology) shows mixed results; while it is generally positive and significant, it is not significant in some models. The destination percentage of arable lands (d_arable) remains positively significant, emphasizing the importance of arable land in attracting migrants. Absolute temperature difference (abstempdiff) continues to show a negative and significant effect in most models,

confirming that temperature differences are a deterrent to migration. The source population density in 1500 (s_ln_pd1500) and source percentage of arable lands (s_arable) show significant results in some models, indicating that source region characteristics also play a role, though their impact is less consistent compared to the destination variables.

Asia

In The Table 31 (Appendix D) the sample is Asian regions to non-eurasian regions. The patterns largely align with the European tables. Bilateral distance (lndist) is negative and significant, highlighting the consistent barrier posed by distance to migration. Destination population density in 1500 (d_ln_pd1500) remains negative and significant, reinforcing the finding that historically populous regions are less attractive as destinations. Destination malaria ecology (d_malaria_ecology) is generally negative and significant, further indicating the deterrent effect of malaria prevalence on migration. The destination percentage of arable lands (d_arable) consistently shows a positive and significant effect, reiterating the attractiveness of regions with more arable land. Absolute temperature difference (abstempdiff) continues to be a significant deterrent in most models. The significance of source region characteristics, such as population density in 1500 (s_ln_pd1500) and percentage of arable lands (s_arable), varies across models, indicating their less consistent but still notable role in migration dynamics. The constant terms are significant across all models, indicating the presence of unobserved factors influencing migration flows.

In summary, it can be claimed that migration is influenced by both source and destination countries, making an origin-destination framework essential for analyzing historical migrations, just as in contemporary studies. Estimations with cross-country and bilateral relationships show that while destination factors are influential, source and bilateral factors are also significant. Also, regional analysis provides more accurate results for deep-rooted migration factors. The results show that several variable signs may differ between Eurasia and Sub-Saharan Africa regions which demonstrates that global analysis might obscure regional differences or miss significant results. Thus, as explained above, a regional approach offers targeted insights and more robust results. Finally, historical migrations are shaped by intertwined economic, political, and geographical factors. Therefore, colonial migrations, often used as natural experiments in economic analysis, are influenced by a complex network of factors that must be acknowledged.

5.2.4 Robustness checks

Different independent variable

Results with *DS* (1960), results with *DS* and *DSv2* from WMM (2000) shows similar results, can be seen in the Table 32, Table 33, and Table 34 (Appendix D).

Different regional settings and additional controls

- Only New Lands for Eurasian Sources: Main variables mostly significant even controlled for colonial dependency and presence of precious metals. Can be seen in the Table 35 (Appendix D).

- Only Non-Europe for European Sources: Arable and lnpd1500 are significant. Colonial dependency and presence of precious metals have positive significant effects. Can be seen in the Table 36 (Appendix D).
- Only Non-Asian for Asian Sources: Arable is positive but ln pd1500 is negative. Can be seen in the Table 37 (Appendix D).
- Only New Lands for Asian Sources: Like previous results but Malaria Ecology is positive. Can be seen in the Table 38 (Appendix D).

CHAPTER 6

CONCLUSION

When considering these results, three main conclusions emerge. First, due to its nature, migration is influenced by both source and destination countries. Therefore, analyzing historical migrations through an origin-destination framework is beneficial, just as in contemporary migration studies. As estimations with cross-country and bilateral relationships have shown, while destination-based outcomes are influential, source and bilateral factors also play a significant role. In fact, in bilateral structures, it has been discovered that some source factors, previously undetected, can also be impactful. Therefore, focusing solely on destination-based studies could lead to inaccurate conclusions.

Second, when examining deep-rooted factors that influence migration, regional analysis provides more accurate results, as demonstrated by my results where the signs of various variables changed when comparing Eurasia and Sub-Saharan Africa. A global analysis might obscure regional differences or fail to capture significant results. Additionally, approaching migration from a regional perspective not only provides targeted insights but also enhances the utility of spatial analysis.

Thirdly, historical migrations highlight the intertwined influences of economic, political, and geographical factors, which shape historical narratives in different ways. The concept of the state and economic interests has been highly influential in migration, but nature, which humans have always had to contend with, also plays a significant role in shaping historical events. While colonial migrations have often been used as a natural

experiment in economic analysis, it is important to remember that a complex network of factors shapes this phenomenon.

Also, it should be acknowledged that The WMM does not solely reflect the impact of migration. Instead, it encapsulates a bundle of influences, including not only the deliberate decisions of groups of people but also the unintended consequences of historical accidents, and cultural differences such as, wars, genocides, natural disasters, and diseases have disproportionately affected different groups, leading to variations in mortality rates across ancestral groups. Additionally, cultural differences have resulted in varying fertility rates among these groups. These factors contribute to the ancestral compositions we observe today. Describing the data merely in terms of migration between source and destination would be misleading and simplistic. Despite these complexities, it is essential to investigate whether it can be identified some deep-rooted factors that have systematically shaped ancestral compositions. The aim of this thesis was to demonstrate, by this data in various controlled and regional analyses, that certain factors remain critically important despite the multitude of influences. By employing different controls and conducting regional analyses, we were able to show that some factors consistently exert a significant impact. This finding underscores the importance of these key factors in shaping migration patterns and their lasting effects on population compositions.

As a final note for further research, a key upgrade to the Gravity Model involves looking beyond direct origin-destination connections to consider the impact of nearby points. Originally, gravity model assumes that each pair of origin and destination was independent, an assumption LeSage and Pace (2010) critically called "heroic." Therefore, combining these two methods will be beneficial for further analyses. Another

developmental point will be to periodize the analysis, dividing *DS* and its versions into 100-year periods with more data to be collected. This will enrich the analysis and potentially eliminate endogeneity problems.

APPENDIX A

FIGURES OF CHAPTER 4

wbcode	wbname	afg	alb	dza	ago	arg	arm	aus	aut
AFG	Afghanistar	0,87	0	0	0	0	0	0	0
ALB	Albania	0	0,963	0	0	0	0	0	0
DZA	Algeria	0	0	1	0	0	0	0	0
AGO	Angola	0	0	0	0,98	0	0	0	0
ARG	Argentina	0	0	0	0,005459	0,05	0	0	0
ARM	Armenia	0	0	0	0	0	0,975	0	0
AUS	Australia	0	0	0	0	0	0,001	0,02	0,002
AUT	Austria	0	0	0	0	0	0	0	0,952
AZE	Azerbaijan	0	0	0	0	0	0,015	0	0
BHR	Bahrain	0	0	0	0	0	0	0	0
BGD	Bangladesh	0	0	0	0	0	0	0	0
BLR	Belarus	0	0	0	0	0	0	0	0
BEL	Belgium	0	0	0,006	0	0	0	0	0
BLZ	Belize	0	0	0	0,026096	0	0	0	0
BEN	Benin	0	0	0	0	0	0	0	0
BTN	Bhutan	0	0	0	0	0	0	0	0
BOL	Bolivia	0	0	0	0,002477	0	0	0	0
BIH	Bosnia and	0	0	0	0	0	0	0	0
BWA	Botswana	0	0	0	0	0	0	0	0
BRA	Brazil	0	0	0	0,066072	0	0	0	0

Figure 1. A small portion of WMM

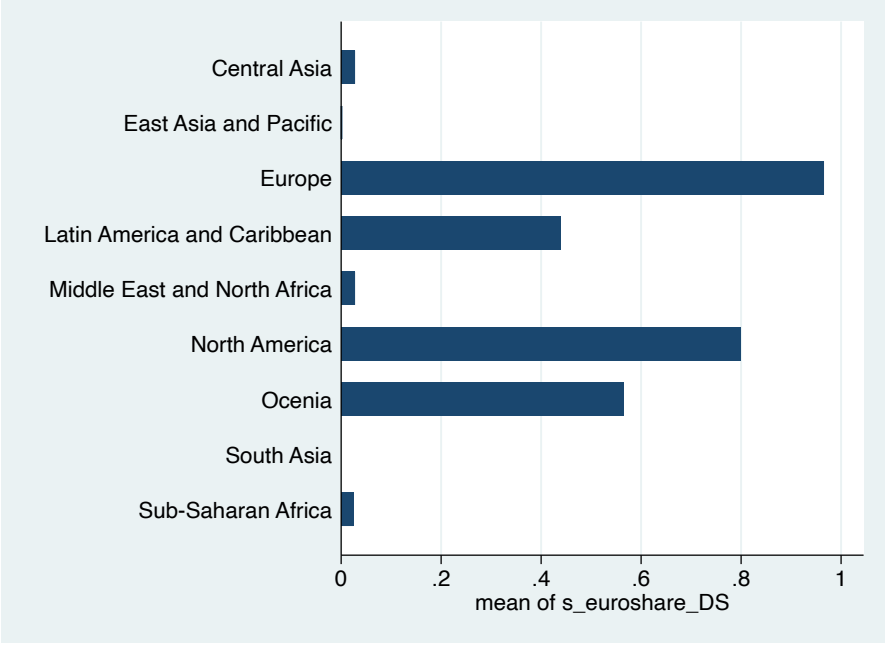


Figure 2. Mean of *EuroShare* across regions

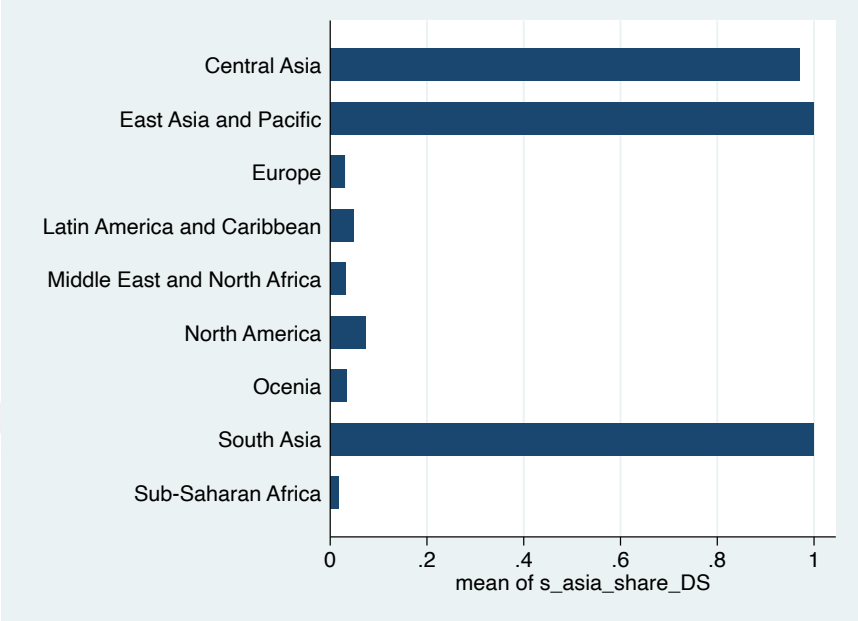


Figure 3. Mean of *AsiaShare* across regions

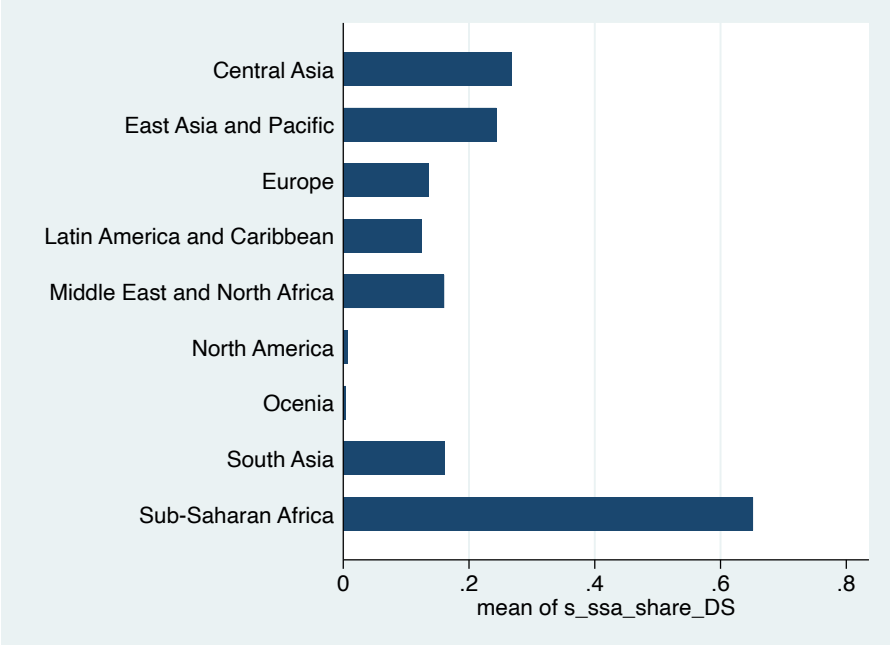


Figure 4. Mean of *SSAshare* across regions

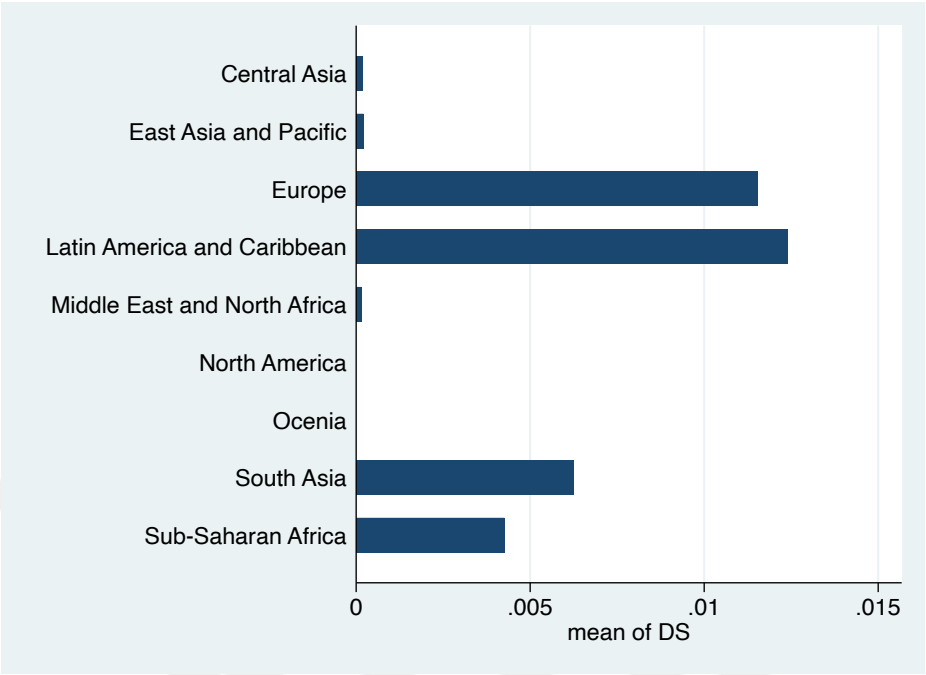


Figure 5. Mean of *DS* by regions for Latin America and Caribbean

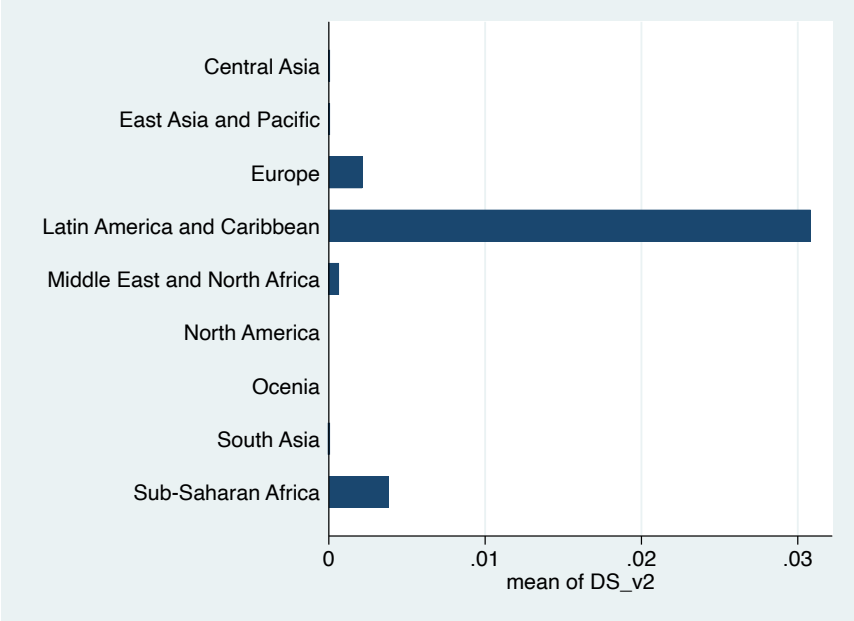


Figure 6. Mean of *DSv2* by regions for Latin America and Caribbean

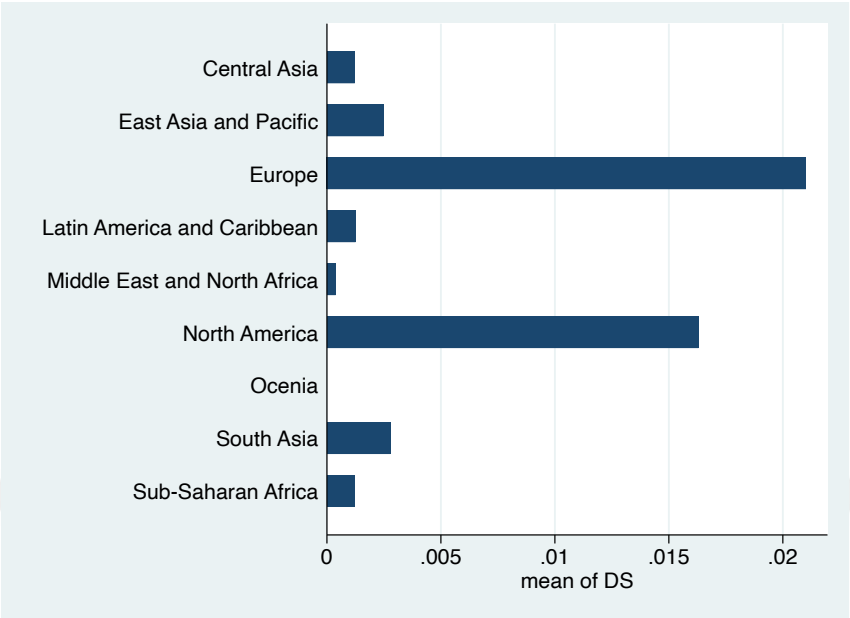


Figure 7. Mean of *DS* by regions for North America

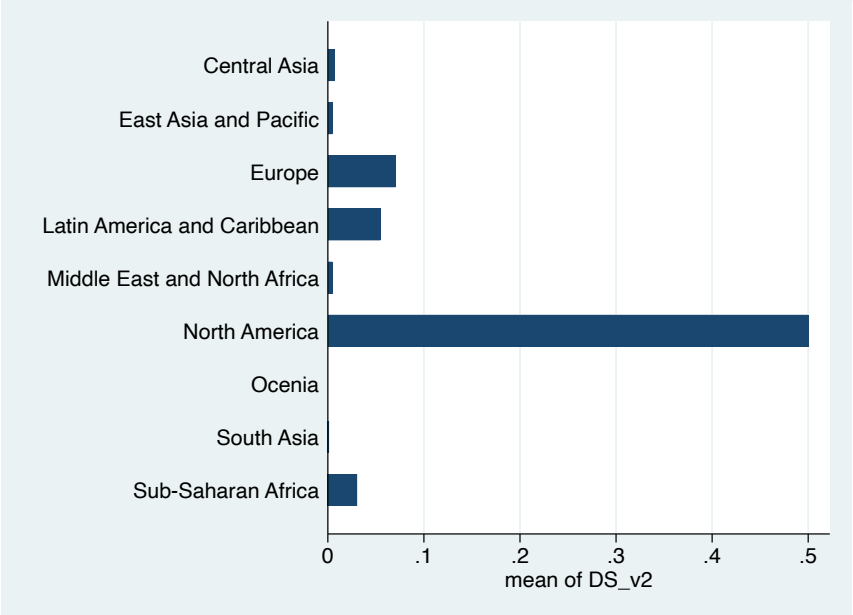


Figure 8. Mean of *DS_v2* by regions for North America

APPENDIX B

TABLES OF CHAPTER 4

Table 1. Number of Countries for Regions

	Frequency
Central Asia	8
East Asia and Pacific	17
Europe	37
Latin America and Caribbean	25
Middle East and North Africa	19
North America	2
Oceania	3
South Asia	7
Sub-Saharan Africa	45
Total	163

Table 2. Descriptive Statistics for Bilateral Data Structure

Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
DSv2	26569	.006	.067	0	1
DS	26569	.006	.068	0	1
d ln pd1500	25265	.906	1.491	-3.817	4.135
d malaria ecology	17930	5.609	8.173	0	32.203
d arable	26080	15.436	13.743	.12	62.1
abslatdiff	26569	27.825	20.534	0	106.287
abstempdiff	26244	9.085	7.325	0	36.569
col dep ever	26248	.01	.101	0	1
d gold silver	18419	.319	.466	0	1
s ln pd1500	25265	.906	1.491	-3.817	4.135
s arable	26080	15.436	13.743	.12	62.1
d island	26080	.144	.351	0	1
d landlock	26080	.237	.426	0	1
d elevavg	26406	.539	.488	.024	2.674
d dist coast	26569	.349	.418	.002	2.206
d rugged	26569	1.336	1.217	.016	6.74
d temp	26406	18.311	8.252	-7.929	28.639
d precip	26406	89.347	60.836	2.911	259.952
d climate	18093	2.586	1.026	1	4
d statehist1500	24450	.176	.185	0	.76

Table 3. Descriptive Statistics for Unilateral Independent Variables:

Variable	Obs	Mean	Std. Dev.	Min	Max
ln pd1500	155	.906	1.496	-3.817	4.135
malaria ecology	110	5.609	8.211	0	32.203
arable	160	15.436	13.786	.12	62.1
elevavg	162	.539	.489	.024	2.674
small	160	.25	.434	0	1
landlock	160	.237	.427	0	1
island	160	.144	.352	0	1
statehist1500	150	.176	.185	0	.76
suitavg	151	.378	.25	0	.96
EDE	110	.294	.899	-2.184	2.605

APPENDIX C

FIGURES OF CHAPTER 5

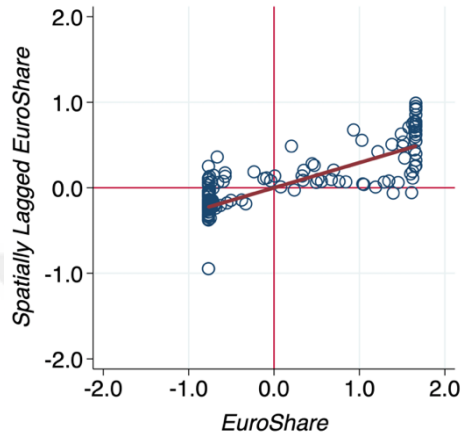


Figure 9. Moran's I Plots for EuroShare:

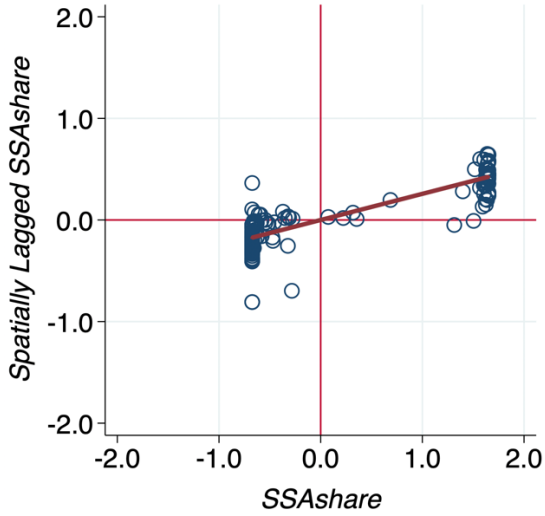


Figure 10. Moran's I Plots for SSAshare:

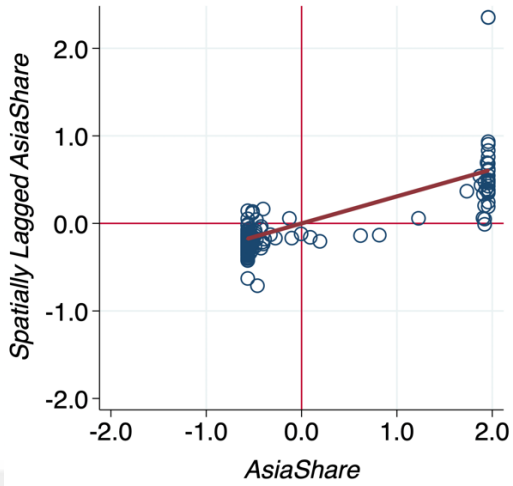


Figure 11. Moran's I Plots for AsiaShare

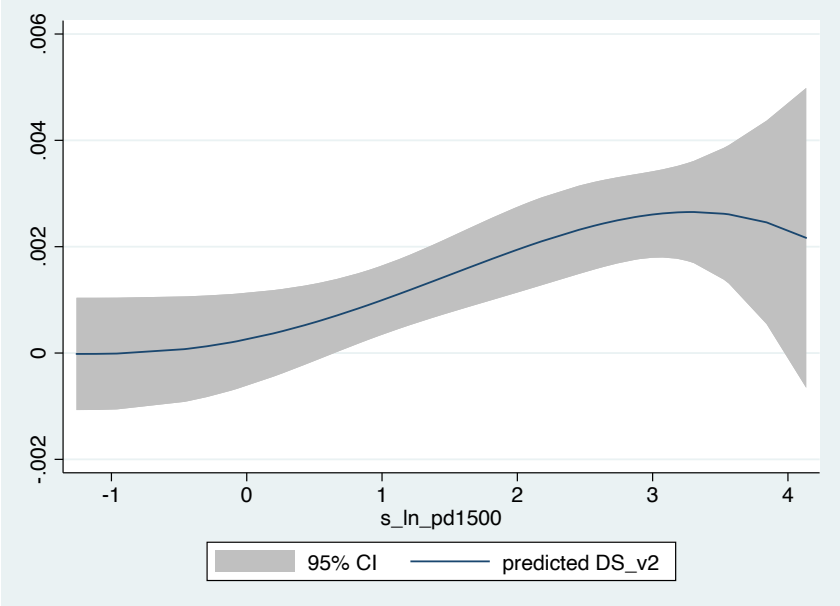


Figure 12. Source Log Population Density at 1500 and DSv2 Change for Eurasia to Non-Eurasia

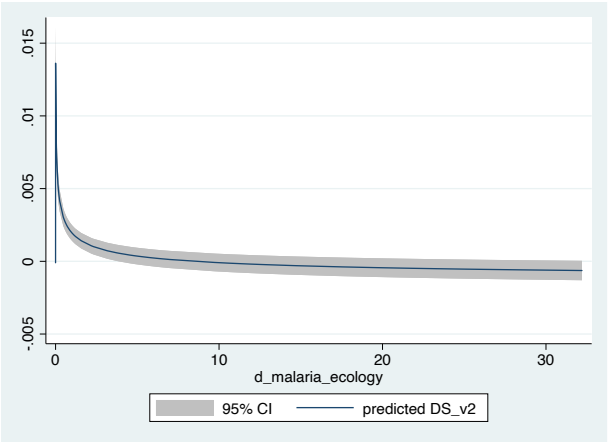


Figure 13. Destination Malaria Ecology and DSv2 Change for Eurasia to Non-Eurasia

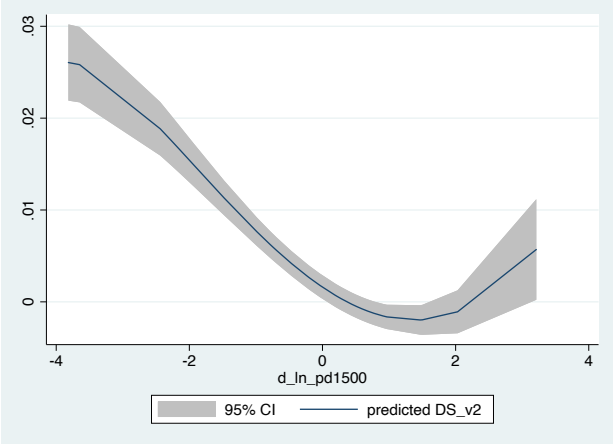


Figure 14. Destination Log Population Density at 1500 and DSv2 Change for Europe to Non-Eurasia

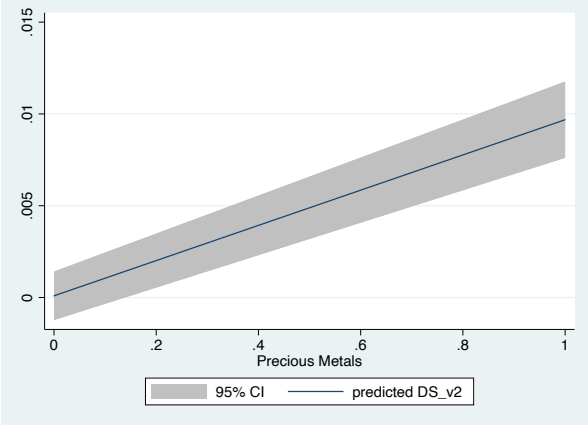


Figure 15. Destination Precious Metals and DSv2 Change for Europe to Non-Eurasia

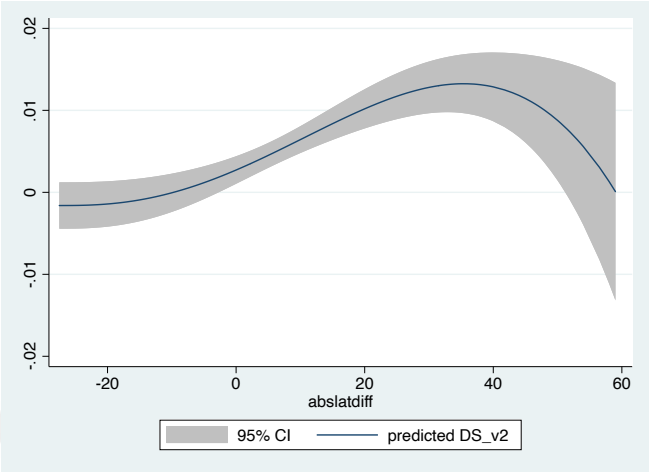


Figure 16. Absolute Latitude Difference and DSv2 Change for Sub-Saharan Africa to New Lands

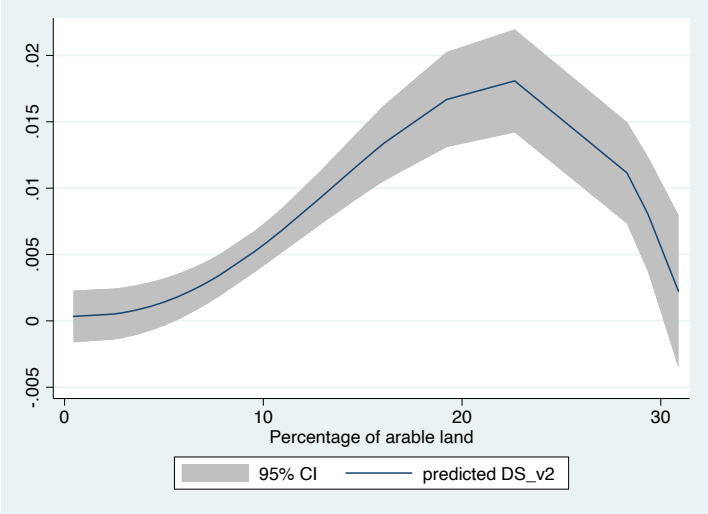


Figure 17. Destination Percentage of Arable Lands and DSv2 Change for Sub-Saharan Africa to New Lands

APPENDIX D

TABLES OF CHAPTER 5

Table 4. Descriptive Statistics for Regional Share Variables

Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
EuroShare	163	.321	.416	0	1
AsiaShare	163	.215	.395	0	1
SSAshare	163	.297	.433	0	1

Table 5. Baseline Regression Results for EuroShare

Variables	OLS / EuroShare		
	(1)	(2)	(3)
Ln Population Density at 1500	-0.115***	-0.118***	-0.097***
	(0.020)	(0.019)	(0.022)
Malaria Ecology	-0.008***	-0.008***	-0.004**
	(0.002)	(0.002)	(0.002)
Percentage of Arable Lands	0.005***	0.005***	0.005***
	(0.002)	(0.002)	(0.002)
Constant	0.178***	0.275***	0.121**
	(0.034)	(0.057)	(0.059)
Mean	0.321	0.321	0.321
Observations	105	105	105
R-squared	0.399	0.485	0.669
Controls for:			
Geographical Controls	No	Yes	Yes
Colonial Dummies	No	No	Yes

Notes: Geographical Controls are "Average Elevation", landlock and island dummies, and "Average distance to nearest ice-free coast". Colonial dummies are whether colonized by Great Britain, France, Spain, Portugal or Other European Powers. *, **, or *** indicates significance at the 10%, 5% and 1% levels, respectively.

Table 6. Regression Results for EuroShare with More Controls

Variables	OLS / EuroShare						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ln Population Density at 1500	-0.118***	-0.117***	-0.117***	-0.108***	-0.153***	-0.041**	-0.048**
	(0.0192)	(0.0189)	(0.019)	(0.020)	(0.018)	(0.020)	(0.022)
Malaria Ecology	-0.008***	-0.008***	-0.007***	-0.012***	-0.009**	-0.006**	-0.005*
	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)	(0.002)	(0.002)
Percentage of Arable Lands	0.005***	0.005***	0.005***	0.003	0.004***	0.002	0.003**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Absolute Latitude		0.001					
		(0.002)					
Presence of Gold and Silver			0.062				
			(0.051)				
Biogeography				-0.001			
				(0.001)			
Settler Mortality					-0.003		
					(0.033)		
Constant	0.275***	0.260***	0.245***	0.371***	0.351**	0.098**	0.058
	(0.057)	(0.076)	(0.068)	(0.090)	(0.141)	(0.050)	(0.050)
Mean	0.321	0.321	0.321	0.321	0.321	0.321	0.321
Observations	105	105	105	105	105	105	105
R-squared	0.485	0.486	0.505	0.570	0.665	0.727	0.794
Controls for:							
Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region Fixed Effects	No	No	No	No	No	Yes	Yes
Colonial Dummies	No	No	No	No	No	No	Yes

Notes: Geographical Controls are "Average Elevation", landlock and island dummies, and "Average distance to nearest ice-free coast". Colonial dummies are whether colonized by Great Britain, France, Spain, Portugal or Other European Powers. Countries divided to 9 regions taken from world bank. *, **, or *** indicates significance at the 10%, 5% and 1% levels, respectively.

Table 7. Baseline Regression Results for SSAshare

Variables	OLS / SSAshare		
	(1)	(2)	(3)
Ln Population Density at 1500	-0.031**	-0.039**	-0.037*
	(0.014)	(0.017)	(0.020)
Malaria Ecology	-0.008	-0.014*	-0.013
	(0.007)	(0.007)	(0.008)
Percentage of Arable Lands	0.004	0.004*	0.003*
	(0.003)	(0.002)	(0.002)
Constant	0.041*	0.027	-0.004
	(0.022)	(0.036)	(0.053)
Mean	0.215	0.215	0.215
Observations	67	67	67
R-squared	0.076	0.289	0.324
Controls for:			
Geographical Controls	No	Yes	Yes
Colonial Dummies	No	No	Yes

Notes: Geographical Controls are "Average Elevation", landlock and island dummies, and "Average distance to nearest ice-free coast". Colonial dummies are whether colonized by Great Britain, France, Spain, Portugal or Other European Powers. *, **, or *** indicates significance at the 10%, 5% and 1% levels, respectively.

Table 8. Regression Results for EuroShare with More Controls

Variables	OLS / SSAshare					
	(1)	(2)	(3)	(4)	(5)	(6)
Ln Population Density at 1500	-0.039**	-0.039**	-0.039**	-0.051**	-0.037*	-0.014
	(0.017)	(0.016)	(0.016)	(0.020)	(0.020)	(0.015)
Malaria Ecology	-0.014*	-0.021**	-0.013*	-0.014*	-0.013	-0.001
	(0.007)	(0.008)	(0.007)	(0.007)	(0.008)	(0.005)
Percentage of Arable Lands	0.004*	0.005*	0.004*	0.004*	0.003*	0.004**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Absolute Latitude		-0.003**				
		(0.001)				
Presence of Gold and Silver			0.021			
			(0.049)			
Slave				-0.097*		
				(0.056)		
Constant	0.027	0.096**	0.023	0.057	-0.004	-0.073
	(0.036)	(0.044)	(0.041)	(0.040)	(0.053)	(0.055)
Mean	0.215	0.215	0.215	0.215	0.215	0.215
Observations	67	67	67	67	67	67
R-squared	0.289	0.330	0.290	0.336	0.324	0.820
Controls for:						
Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes
Region Fixed Effects	No	No	No	No	No	Yes
Colonial Dummies	No	No	No	No	Yes	Yes

Notes: Geographical Controls are "Average Elevation", landlock and island dummies, and "Average distance to nearest ice-free coast". Colonial dummies are whether colonized by Great Britain, France, Spain, Portugal or Other European Powers. *, **, or *** indicates significance at the 10%, 5% and 1% levels, respectively.

Table 9. Moran's I and Lagrange Multiplier Tests

Diagnostics for SSAshare

Test	Statistic	p-value
Spatial Error:		
Moran's I	10.046	0.000
Lagrange Multiplier	40.255	0.000
Robust LM	0.236	0.627
Spatial Lag:		
Lagrange Multiplier	52.854	0.000
Robust LM	12.835	0.000

Diagnostics for EuroShare

Test	Statistic	p-value
Spatial Error:		
Moran's I	8.077	0.000
Lagrange Multiplier	29.513	0.000
Robust LM	3.404	0.065
Spatial Lag:		
Lagrange Multiplier	30.991	0.000
Robust LM	4.882	0.027

Diagnostics for AsiaShare

Test	Statistic	p-value
Spatial Error:		
Moran's I	3.572	0.000
Lagrange Multiplier	3.833	0.050
Robust LM	0.025	0.874
Spatial Lag:		
Lagrange Multiplier	3.956	0.047
Robust LM	0.148	0.701

Table 10. Spatial ML Results for EuroShare

Variables	EuroShare			EuroShare		
	SEM	SAR	SLX	SAC	SDM	SDEM
	(1)	(2)	(3)	(4)	(5)	(6)
Ln Population Density at 1500	-0.061***	-0.078***	-0.045***	-0.060***	-0.045***	-0.046***
	(0.018)	(0.016)	(0.017)	(0.017)	(0.017)	(0.017)
Malaria Ecology	-0.011***	-0.011***	-0.014***	-0.011***	-0.014***	-0.014***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Percentage of Arable Lands	0.004***	0.004**	0.004***	0.004***	0.004**	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	0.473***	0.258***	0.641***	0.470***	0.639***	0.645***
	(0.088)	(0.068)	(0.099)	(0.097)	(0.099)	(0.092)
Mean	0.321	0.321	0.321	0.321	0.321	0.321
Observations	105	105	105	105	105	105
Controls for:						
Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Geographical Controls are "Average Elevation", landlock and island dummies, and "Average distance to nearest ice-free coast". *, **, or *** indicates significance at the 10%, 5% and 1% levels, respectively.

Table 11. Spatial ML Results for SSAshare

Variables	SSAshare			SSAshare		
	SEM	SAR	SLX	SAC	SDM	SDEM
	(1)	(2)	(3)	(4)	(5)	(6)
Ln Population Density at 1500	0.00114	-0.0221*	-0.0265*	-0.0219*	-0.0277**	-0.0254*
	(0.0177)	(0.0127)	(0.0137)	(0.0130)	(0.0136)	(0.0136)
Malaria Ecology	-0.00255	-0.00637	-0.00484	-0.00617	-0.00564	-0.00460
	(0.00910)	(0.00779)	(0.00822)	(0.00784)	(0.00817)	(0.00814)
Percentage of Arable Lands	0.00141	0.00355**	0.00241	0.00353**	0.00233	0.00272*
	(0.00162)	(0.00159)	(0.00155)	(0.00159)	(0.00153)	(0.00156)
Constant	0.0598	-0.0603	-0.155**	-0.0603	-0.176**	-0.163**
	(0.0669)	(0.0399)	(0.0664)	(0.0411)	(0.0702)	(0.0677)
Mean	0.215	0.215	0.215	0.215	0.215	0.215
Observations	67	67	67	67	67	67
Controls for:						
Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Geographical Controls are "Average Elevation", landlock and island dummies, and "Average distance to nearest ice-free coast". *, **, or *** indicates significance at the 10%, 5% and 1% levels, respectively.

Table 12. Alternative Estimation Method for Spatial Analyses : EuroShare

VARIABLES	(1) SAC / EuroShare	(2) SDEM / EuroShare	(3) SDM / EuroShare	(4) GNS / EuroShare
ln_pd1500	-0.0783***	-0.0535***	-0.0506***	-0.0562***
	(0.0191)	(0.0175)	(0.0173)	(0.0172)
malaria_ecology	-0.0107***	-0.0121***	-0.0117***	-0.0116***
	(0.00288)	(0.00405)	(0.00453)	(0.00403)
arable	0.00412***	0.00457***	0.00477***	0.00481***
	(0.00145)	(0.00119)	(0.00127)	(0.00122)
Constant	0.239***	0.705***	0.681***	0.665***
	(0.0876)	(0.129)	(0.145)	(0.122)
Observations	105	105	105	105
Controls	YES	YES	YES	YES
Pseudo R-squared	0.589	0.719	0.724	0.724

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 13. Alternative Estimation Method for Spatial Analyses : SSAshare

VARIABLES	(1) SAC / SSAshare	(2) SDEM / SSAshare	(3) SDM / SSAshare	(4) GNS / SSAshare
ln_pd1500	-0.0131	-0.0263**	-0.0230**	-0.0256**
	(0.00898)	(0.0103)	(0.0109)	(0.0114)
malaria_ecology	-0.00274	-0.00481	-0.00254	-0.00310
	(0.00488)	(0.00564)	(0.00583)	(0.00657)
arable	0.00327*	0.00245	0.00266	0.00240
	(0.00179)	(0.00160)	(0.00174)	(0.00169)
Constant	-0.107**	-0.156**	-0.0957	-0.105
	(0.0508)	(0.0629)	(0.0761)	(0.0813)
Observations	67	67	67	67
Controls	YES	YES	YES	YES
Pseudo R-squared	0.252	0.612	0.304	0.456

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 14. Only Non-Eurasia for EuroShare

VARIABLES	(1) SAC / EuroShare	(2) SDEM / EuroShare	(3) SDM / EuroShare	(4) OLS / EuroShare
ln_pd1500	-0.0971***	-0.0861***	-0.0850***	-0.0836***
	(0.0246)	(0.0305)	(0.0305)	(0.0222)
malaria_ecology	-0.0137***	-0.0115**	-0.0114**	-0.0186***
	(0.00466)	(0.00511)	(0.00516)	(0.00313)
arable	0.00496**	0.00546**	0.00556**	0.00479**
	(0.00235)	(0.00233)	(0.00233)	(0.00211)
Constant	0.346***	0.652***	0.656***	0.472***
	(0.120)	(0.140)	(0.145)	(0.0762)
Observations	68	68	68	68
R-squared				0.634
Controls	YES	YES	YES	YES
Pseudo R-squared	0.638	0.739	0.739	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 15. Only New Lands for EuroShare

VARIABLES	(1) SAC / EuroShare	(2) SDEM / EuroShare	(3) SDM / EuroShare	(4) OLS / EuroShare
ln_pd1500	-0.111*	0.0878*	0.135**	-0.133***
	(0.0624)	(0.0480)	(0.0540)	(0.0369)
malaria_ecology	-0.0456**	-0.108***	-0.100***	-0.0410
	(0.0223)	(0.0109)	(0.0129)	(0.0244)
arable	0.00182	0.00503	0.00408	0.00143
	(0.00525)	(0.00319)	(0.00387)	(0.00606)
Constant	0.603***	1.462***	1.544***	0.508***
	(0.232)	(0.141)	(0.172)	(0.121)
Observations	28	28	28	28
R-squared	0.571	0.705	0.746	0.566
Controls	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 16. Only New Lands for EuroShare

VARIABLES	(1) SAC / SSAshare	(2) SDEM / SSAshare	(3) SDM / SSAshare	(4) OLS / SSAshare
ln_pd1500	-0.0148	-0.0608	-0.0828*	0.0310
	(0.0346)	(0.0454)	(0.0479)	(0.0289)
malaria_ecology	-0.0182	0.0195*	0.0151	-0.0277
	(0.0167)	(0.0101)	(0.0114)	(0.0235)
arable	0.00344	0.00509*	0.00390	0.00359
	(0.00434)	(0.00291)	(0.00343)	(0.00720)
Constant	-0.0611	-0.543***	-0.593***	0.125
	(0.0992)	(0.133)	(0.152)	(0.105)
Observations	28	28	28	28
R-squared	0.582	0.691	0.677	0.496
Controls	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 17. Alternative Independent Variables for EuroShare

VARIABLES	(1) SAC / EuroShare	(2) SDEM / EuroShare	(3) SDM / EuroShare	(4) OLS / EuroShare
statehist1500	-0.281**	-0.307**	-0.304**	-0.291**
	(0.128)	(0.127)	(0.128)	(0.117)
EDE	-0.121***	-0.106***	-0.106***	-0.166***
	(0.0261)	(0.0275)	(0.0277)	(0.0262)
suitavg	-0.00618	0.0238	0.0259	0.0788
	(0.0791)	(0.0761)	(0.0762)	(0.0944)
Constant	0.483***	0.503***	0.495***	0.306***
	(0.0950)	(0.0870)	(0.0860)	(0.0599)
Observations	100	100	100	100
R-squared	0.484	0.689	0.689	0.522
Controls	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 18. Alternative Independent Variables for SSAshare

VARIABLES	(1) SAC / SSAshare	(2) SDEM / SSAshare	(3) SDM / SSAshare	(4) OLS / SSAshare
statehist1500	0.0503	0.139	0.130	0.0820
	(0.114)	(0.116)	(0.116)	(0.104)
EDE	0.0315	0.00524	0.00616	0.0412*
	(0.0237)	(0.0236)	(0.0235)	(0.0207)
suitavg	0.0412	-0.0126	0.000457	0.0492
	(0.0645)	(0.0608)	(0.0611)	(0.0897)
Constant	-0.000698	-0.128*	-0.126*	0.0794
	(0.0624)	(0.0659)	(0.0685)	(0.0492)
Observations	64	64	64	64
R-squared	0.376	0.556	0.565	0.360
Controls	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 19. Additional Control Variables for EuroShare

VARIABLES	(1) SAC / EuroShare	(2) SDEM / EuroShare	(3) SDM / EuroShare	(4) OLS / EuroShare
ln_pd1500	-0.0573***	-0.0538***	-0.0518***	-0.0560***
	(0.0172)	(0.0162)	(0.0163)	(0.0188)
malaria_ecology	-0.00729**	-0.00678*	-0.00694*	-0.0123***
	(0.00355)	(0.00398)	(0.00405)	(0.00286)
arable	0.00344**	0.00363**	0.00366**	0.00341**
	(0.00149)	(0.00142)	(0.00143)	(0.00139)
Constant	0.662***	0.966***	0.969***	0.618***
	(0.128)	(0.127)	(0.131)	(0.138)
Observations	96	96	96	96
R-squared	0.667	0.796	0.796	0.684
Controls	YES	YES	YES	YES

Controls are average elevation, landlock, eurasia and island dummies, temperature, precipitation, ruggedness, gold_silver dummy, and average distance to the nearest ice-free coast.
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 20. Additional Control Variables for SSAshare

VARIABLES	(1) SAC / SSAshare	(2) SDEM / SSAshare	(3) SDM / SSAshare	(4) OLS / SSAshare
ln_pd1500	-0.0173	0.0107	-0.00217	-0.00658
	(0.0177)	(0.0155)	(0.0179)	(0.0175)
malaria_ecology	0.00333	0.00673	0.00595	-0.0104
	(0.00948)	(0.00853)	(0.0104)	(0.00970)
arable	0.00312**	0.00455***	0.00306**	0.00324*
	(0.00157)	(0.00106)	(0.00135)	(0.00190)
Constant	-0.139	-0.279***	-0.259**	-0.119
	(0.112)	(0.0846)	(0.110)	(0.118)
Observations	59	59	59	59
R-squared				0.537
Controls	YES	YES	YES	YES
Pseudo R-squared	0.601	0.714	0.649	

Controls are average elevation, landlock, eurasia and island dummies, temperature, precipitation, ruggedness, gold silver dummy, and average distance to the nearest ice-free coast.
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 21. Spatial Analyses with Colonial Dummies for EuroShare

VARIABLES	(1) SAC / EuroShare	(2) SDEM / EuroShare	(3) SDM / EuroShare	(4) OLS / EuroShare
ln_pd1500	-0.0736***	-0.0633***	-0.0611***	-0.0799***
	(0.0161)	(0.0160)	(0.0159)	(0.0217)
malaria_ecology	-0.00971***	-0.0109***	-0.0110***	-0.00832**
	(0.00332)	(0.00346)	(0.00359)	(0.00325)
arable	0.00474***	0.00514***	0.00530***	0.00480***
	(0.00146)	(0.00132)	(0.00135)	(0.00151)
Constant	0.344***	0.573***	0.595***	0.232***
	(0.0972)	(0.0981)	(0.104)	(0.0846)
Observations	105	105	105	105
R-squared				0.686
Controls	YES	YES	YES	YES
Colony Dummies	YES	YES	YES	YES
Pseudo R-squared	0.685	0.771	0.771	

Controls are average elevation, landlock, island and eurasia dummies, and average distance to the nearest ice-free coast.
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 22. Spatial Analyses with Colonial Dummies for SSAshare

VARIABLES	(1) SAC / SSAshare	(2) SDEM / SSAshare	(3) SDM / SSAshare	(4) OLS / SSAshare
ln_pd1500	-0.00271	-0.0304***	-0.0341***	0.0186
	(0.0136)	(0.0116)	(0.0119)	(0.0170)
malaria_ecology	-0.00658	-0.00719	-0.00578	-0.0165
	(0.00687)	(0.00595)	(0.00594)	(0.0114)
arable	0.00206*	0.00376***	0.00336***	0.00258
	(0.00120)	(0.00108)	(0.00111)	(0.00170)
Constant	0.0813	-0.235***	-0.238***	0.290***
	(0.0770)	(0.0711)	(0.0749)	(0.105)
Observations	67	67	67	67
R-squared				0.549
Controls	YES	YES	YES	YES
Colony Dummies	YES	YES	YES	YES
Pseudo R-squared	0.00942	0.837	0.839	

Controls are average elevation, landlock, island and eurasia dummies, and average distance to the nearest ice-free coast.
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 23. Comparison with WMM 2000 for EuroShare

VARIABLES	(1) SAC / EuroShare	(2) SAC / EuroShare2000	(3) OLS / EuroShare	(4) OLS / EuroShare2000
ln_pd1500	-0.0608***	-0.0594***	-0.0719***	-0.0703***
	(0.0176)	(0.0172)	(0.0197)	(0.0197)
malaria_ecology	-0.0111***	-0.0109***	-0.0157***	-0.0158***
	(0.00356)	(0.00349)	(0.00245)	(0.00242)
arable	0.00403***	0.00388**	0.00397**	0.00381**
	(0.00156)	(0.00153)	(0.00151)	(0.00148)
Constant	0.470***	0.466***	0.382***	0.382***
	(0.0976)	(0.0957)	(0.0555)	(0.0557)
Observations	105	105	105	105
R-squared			0.581	0.582
Controls	YES	YES	YES	YES
Pseudo R-squared	0.541	0.538		

Controls are average elevation, landlock, eurasia, and island dummies,
and average distance to the nearest ice-free coast.
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 24. Comparison with WMM 2000 for SSAshare

VARIABLES	(1) SAC / SSAshare	(2) SAC / SSAshare2000	(3) OLS / SSAshare	(4) OLS / SSAshare2000	(5) SAC / SSAshare	(6) SAC / SSAshare2000	(7) OLS / SSAshare	(8) OLS / SSAshare2000
ln_pd1500	-0.0219*	-0.0201	-0.0390**	-0.0382**	-0.0257**	-0.0235*	-0.0459***	-0.0450***
	(0.0130)	(0.0126)	(0.0167)	(0.0163)	(0.0131)	(0.0127)	(0.0171)	(0.0166)
malaria_ecology	-0.00617	-0.00523	-0.0138*	-0.0137*				
	(0.00784)	(0.00769)	(0.00762)	(0.00743)				
arable	0.00353**	0.00333**	0.00415*	0.00404*	0.00332**	0.00314**	0.00384	0.00373
	(0.00159)	(0.00151)	(0.00236)	(0.00228)	(0.00157)	(0.00150)	(0.00241)	(0.00233)
EDE					0.0326	0.0317	0.0471*	0.0468*
					(0.0239)	(0.0229)	(0.0245)	(0.0240)
Constant	-0.0603	-0.0601	0.0271	0.0292	-0.0733*	-0.0718*	0.000865	0.00325
	(0.0411)	(0.0411)	(0.0367)	(0.0359)	(0.0397)	(0.0396)	(0.0425)	(0.0414)
Observations	67	67	67	67	67	67	67	67
R-squared			0.289	0.293			0.294	0.298
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Pseudo R-squared	0.453	0.441			0.438	0.427		

Controls are average elevation, landlock, eurasia, and island dummies,
and average distance to the nearest ice-free coast.
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 25. AsiaShare

VARIABLES	(1) SAC / AsiaShare	(2) SDEM / AsiaShare	(3) SDM / AsiaShare	(4) OLS / AsiaShare	(5) OLS / AsiaShare
ln_pd1500	-0.00802	-0.0124	-0.0112	-0.00561	-0.00561
	(0.00869)	(0.0117)	(0.0118)	(0.00826)	(0.00826)
malaria_ecology	-0.000824	-0.000411	-0.000717	-0.00185*	-0.00185*
	(0.00146)	(0.00205)	(0.00205)	(0.00101)	(0.00101)
arable	0.000104	7.11e-05	-5.59e-05	-0.000188	-0.000188
	(0.000930)	(0.000996)	(0.000975)	(0.000550)	(0.000550)
Constant	0.0439	0.0182	0.0416	0.0726	0.0726
	(0.0386)	(0.0672)	(0.0618)	(0.0456)	(0.0456)
Observations	82	82	82	82	82
R-squared				0.164	0.164
Controls	YES	YES	YES	YES	YES
Colony Dummies	NO	NO	NO	NO	YES
Pseudo R-squared	0.0656	0.123	0.121		

Controls are average elevation, landlock, island dummies, and average distance to the nearest ice-free coast.
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 26. From Eurasia to Non-Eurasia Baseline Results

Variables	OLS		PPML	
	(1)	(2)	(4)	(5)
Ln Distance	-0.331***	-0.322***	-2.514***	-2.810***
	(0.095)	(0.0955)	(0.443)	(0.486)
Destination Ln Population Density at 1500	-0.269***	-0.254***	-1.256***	-0.741***
	(0.054)	(0.0519)	(0.154)	(0.164)
Destination Malaria Ecology	-0.006***	-0.007**	-0.228***	-0.306***
	(0.002)	(0.003)	(0.056)	(0.057)
Destination Percentage of Arable Lands	0.019***	0.020***	0.158***	0.153***
	(0.004)	(0.004)	(0.034)	(0.019)
Constant	2.999***	2.865***	17.760***	22.040***
	(0.861)	(0.859)	(3.622)	(4.441)
Mean	0.122	0.122	0.122	0.122
Observations	5,916	5,916	5,916	5,916
R-Squared	0.027	0.033	---	---
Pseudo R-Squared	---	---	0.473	0.542
Controls for:				
Geographical Controls	No	Yes	No	Yes

Notes: Geographical Controls are "Average Elevation", landlock and island dummies, and "Average distance to nearest ice-free coast". *, **, or *** indicates significance at the 10%, 5% and 1% levels, respectively.

Table 27. From Eurasia to Non-Eurasia Detailed Results

Variables	OLS				PPML			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Ln Distance	-0.313***	-0.318***	-0.259**	-0.408***	-1.972***	-1.899***	-2.489***	-2.538***
	(0.0996)	(0.108)	(0.101)	(0.123)	(0.605)	(0.605)	(0.773)	(0.799)
Destination Ln Population Density at 1500	-0.250***	-0.264***	-0.237***	-0.108***	-0.708***	-0.685***	-0.832***	-0.748***
	(0.0509)	(0.0555)	(0.0492)	(0.0285)	(0.169)	(0.173)	(0.261)	(0.260)
Destination Malaria Ecology	-0.00627***	-0.00648**	-0.0110***	0.0150***	-0.239***	-0.230***	-0.564*	-0.0488
	(0.00235)	(0.00259)	(0.00418)	(0.00442)	(0.0778)	(0.0732)	(0.321)	(0.0809)
Destination Percentage of Arable Lands	0.0205***	0.0216***	0.0229***	0.00733***	0.101***	0.0947***	0.112***	0.0702***
	(0.00452)	(0.00491)	(0.00508)	(0.00230)	(0.0205)	(0.0213)	(0.0263)	(0.0189)
Absolute Temperature Difference	-0.0166***	-0.0166***	-0.0167***	-0.0213***	-0.118***	-0.124***	-0.101***	-0.0940***
	(0.00402)	(0.00404)	(0.00408)	(0.00498)	(0.0198)	(0.0230)	(0.0210)	(0.0222)
Colonial Dependency	2.272*	2.209*	2.223*	2.113*	2.201***	1.946**	1.740***	1.708***
	-1.161	-1.164	-1.165	-1.168	(0.783)	(0.786)	(0.618)	(0.625)
Source Ln Population Density at 1500		0.0440**	0.0445**	0.0411**		0.454***	0.435***	0.435***
		(0.0184)	(0.0185)	(0.0180)		(0.146)	(0.149)	(0.155)
Source Percentage of Arable Lands		-0.000691	-0.000621	-0.000389		-0.0225	-0.0213	-0.0220*
		(0.00149)	(0.00148)	(0.00142)		(0.0144)	(0.0131)	(0.0131)
Constant	2.946***	2.937***	2.666***	4.157***		14.52***	14.69**	16.99**
	(0.929)	-1.004	(0.963)	-1.228		-5.526	-6.300	-7.870
Observations	5,916	5,372	5,372	5,372	5,916	5,372	5,214	5,214
R-squared	0.051	0.052	0.053	0.096	-	-	-	-
Pseudo R-Squared	-	-	-	-	0.600	0.620	0.648	0.664
Controls for:								
Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region Fixed Effects	No	No	No	Yes	No	No	No	Yes
Colonial Dummies	No	No	Yes	Yes	No	No	Yes	Yes

Notes: Geographical Controls are "Average Elevation", landlock and island dummies, and "Average distance to nearest ice-free coast". Colonial dummies are whether colonized by Great Britain, France, Spain, Portugal or Other European Powers. *, **, or *** indicates significance at the 10%, 5% and 1% levels, respectively.

Table 28. Regional Bilateral Analyses

Variables	Europe		SSA		Asia	
	OLS	PPML	OLS	PPML	OLS	PPML
	(1)	(2)	(4)	(5)	(4)	(5)
Ln Distance	-1.344***	-4.335***	-1.887***	-3.272***	-0.017	-1.531
	(0.420)	(0.735)	(0.427)	(0.389)	(0.0158)	(1.477)
Destination Ln Population Density at 1500	-0.557***	-1.112***	-0.280***	-0.194*	-0.035***	-1.381***
	(0.133)	(0.208)	(0.108)	(0.113)	(0.013)	(0.438)
Destination Malaria Ecology	-0.029***	-0.317**	0.135***	0.230***	0.001	-0.337***
	(0.009)	(0.148)	(0.0406)	(0.0658)	(0.001)	(0.0743)
Destination Percentage of Arable Lands	0.039***	0.110***	0.096***	0.159***	0.00334**	0.299***
	(0.010)	(0.0147)	(0.023)	(0.0169)	(0.00135)	(0.0409)
Constant	12.020***	37.080***	15.890***	25.260***	0.130	3.138
	(3.760)	(6.464)	(3.635)	(3.415)	(0.145)	(13.57)
Observations	2.516	2.516	1.260	1.260	2.108	2.108
R-Squared	0.065	---	0.084	---	0.031	---
Pseudo R-Squared	---	0.615	---	0.405	---	0.593
Controls for:						
Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Geographical Controls are "Average Elevation", landlock and island dummies, and "Average distance to nearest ice-free coast". *, **, or *** indicates significance at the 10%, 5% and 1% levels, respectively.

Table 29. Bilateral Analyses for Europe

VARIABLES	(1) OLS / DSv2	(2) PPML / DSv2	(3) OLS / DSv2	(4) PPML / DSv2	(5) OLS / DSv2	(6) PPML / DSv2	(7) OLS / DSv2	(8) PPML / DSv2
lndist	-1.155***	-2.967**	-1.379***	-2.677**	-1.108**	-4.720***	-1.465***	-7.456***
	(0.393)	(1.271)	(0.471)	(1.113)	(0.487)	(1.276)	(0.542)	(0.977)
d_ln_pd1500	-0.425***	-0.525	-0.433***	-0.536*	-0.321***	-0.502	-0.135***	-0.106
	(0.101)	(0.366)	(0.104)	(0.318)	(0.0835)	(0.450)	(0.0450)	(0.336)
d_malaria_ecology	-0.0246***	-0.229**	-0.0290***	-0.220**	-0.0376***	-1.625***	0.0182***	-1.221***
	(0.00786)	(0.115)	(0.00931)	(0.0987)	(0.0116)	(0.369)	(0.00622)	(0.422)
d_arable	0.0356***	0.0466**	0.0367***	0.0575***	0.0398***	0.00731	0.0103**	0.0926***
	(0.00847)	(0.0210)	(0.00891)	(0.0186)	(0.00921)	(0.0423)	(0.00440)	(0.0260)
abstempdiff	-0.0801***	-0.183***	-0.109***	-0.173***	-0.114***	-0.146***	-0.0770***	-0.188***
	(0.0196)	(0.0566)	(0.0293)	(0.0476)	(0.0303)	(0.0429)	(0.0234)	(0.0422)
col_dep_ever	2.048*	1.866**	2.164*	1.932***	2.180*	1.472***	2.042*	1.283***
	(1.176)	(0.764)	(1.182)	(0.747)	(1.183)	(0.458)	(1.193)	(0.380)
s_ln_pd1500			-0.236***	0.115	-0.233***	0.0291	-0.177**	-0.0737
			(0.0883)	(0.164)	(0.0883)	(0.145)	(0.0799)	(0.154)
s_arable			-0.00634*	-0.0327*	-0.00667*	-0.0246**	-0.00578*	-0.0145
			(0.00376)	(0.0168)	(0.00382)	(0.0120)	(0.00342)	(0.0115)
Constant	11.64***	27.32**	14.84***	24.96**	12.90***	43.34***	15.58***	59.90***
	(3.811)	(10.88)	(4.906)	(9.712)	(4.960)	(13.92)	(5.389)	(8.846)
Observations	2,516	2,516	2,244	2,244	2,244	2,178	2,244	2,178
R-squared	0.092	--	0.101	--	0.104	--	0.175	--
Pseudo R-squared	--	0.679	--	0.693	--	0.755	--	0.784
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Regional Dummies	NO	NO	NO	NO	NO	NO	YES	YES
Colonial Dummies	NO	NO	NO	NO	YES	YES	YES	YES

Controls are average elevation, landlock and island dummies, and average distance to the nearest ice-free coast for destinations.
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 30. Bilateral Analyses for SSA

VARIABLES	(1) OLS / DSv2	(2) PPML / DSv2	(3) OLS / DSv2	(4) PPML / DSv2	(5) OLS / DSv2	(6) PPML / DSv2
lndist	-1.913***	-3.184***	-1.751***	-2.687***	-1.740***	-2.909***
	(0.459)	(0.451)	(0.450)	(0.524)	(0.502)	(0.451)
d_ln_pd1500	-0.265***	-0.436***	-0.273***	-0.519***	0.0823*	0.305**
	(0.0860)	(0.150)	(0.0884)	(0.146)	(0.0431)	(0.125)
d_malaria_ecology	0.144***	-0.0946	0.147***	-0.0457	0.0535*	-1.370***
	(0.0532)	(0.210)	(0.0543)	(0.181)	(0.0323)	(0.478)
d_arable	0.0963***	0.156***	0.0983***	0.152***	0.113***	0.150***
	(0.0245)	(0.0154)	(0.0249)	(0.0149)	(0.0307)	(0.0161)
abstempdiff	0.0128	-0.101***	0.0156	-0.100***	0.0338	-0.0849**
	(0.0339)	(0.0334)	(0.0343)	(0.0322)	(0.0239)	(0.0401)
s_ln_pd1500			0.302***	0.431***	0.308***	
			(0.0893)	(0.132)	(0.0872)	
s_arable			-0.0238**	-0.0300	-0.0246**	
			(0.0112)	(0.0204)	(0.0113)	
Constant	16.13***	23.94***	14.71***	19.36***	15.14***	26.74***
	(3.911)	(3.977)	(3.819)	(4.568)	(4.197)	(4.152)
Observations	1,260	1,260	1,232	1,232	1,232	1,260
R-squared	0.084	---	0.090	---	0.113	---
Pseudo R-squared	---	0.402	---	0.416	---	0.449
Controls	YES	YES	YES	YES	YES	YES
Colonial Dummies	NO	NO	NO	NO	YES	YES

Controls are average elevation, landlock and island dummies, and average distance to the nearest ice-free coast for destinations.
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 31. Bilateral Analyses for Asia

VARIABLES	(1) OLS / DSv2	(2) PPML / DSv2	(3) OLS / DSv2	(4) PPML / DSv2	(5) OLS / DSv2	(6) PPML / DSv2
lndist	-0.0195	-1.785*	-0.0250	-2.138**	-0.00139	-1.789*
	(0.0207)	(0.939)	(0.0211)	(0.961)	(0.0219)	(1.013)
d_ln_pd1500	-0.0347***	-2.250***	-0.0349**	-2.099***	-0.0270**	-1.739***
	(0.0131)	(0.447)	(0.0142)	(0.493)	(0.0109)	(0.634)
d_malaria_ecology	4.97e-06	-0.197***	-0.000103	-0.212***	-0.00154**	-1.432
	(0.000341)	(0.0625)	(0.000337)	(0.0780)	(0.000655)	(1.055)
d_arable	0.00335**	0.309***	0.00337**	0.298***	0.00337**	0.256***
	(0.00136)	(0.0458)	(0.00148)	(0.0478)	(0.00148)	(0.0442)
abstempdiff	-0.000354	0.0121	-0.000651	0.0106	-0.000563	0.0155
	(0.000913)	(0.0420)	(0.00102)	(0.0518)	(0.00102)	(0.0413)
s_ln_pd1500			0.00336	0.234	0.00223	
			(0.00379)	(0.181)	(0.00383)	
s_arable			-0.000467	-0.0225	-0.000374	
			(0.000309)	(0.0314)	(0.000284)	
Constant	0.152	1.463	0.211	5.120	0.0497	8.961
	(0.201)	(8.846)	(0.206)	(9.090)	(0.213)	(10.53)
Observations	2,108	2,108	1,904	1,904	1,904	1,457
R-squared	0.031		0.031		0.033	
Controls	YES	YES	YES	YES	YES	YES
Pseudo R-squared		0.590		0.590		0.576
Colonial Dummies					YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 32. Bilateral Results with DS from WMM 1960 version

Variables	Eurasia		Europe		SSA		Asia	
	OLS	PPML	OLS	PPML	OLS	PPML	OLS	PPML
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ln Distance	-0.311***	-1.848***	-0.621	-2.481**	-0.893***	-1.643***	0.005	0.305
	(0.115)	(0.446)	(0.553)	(1.009)	(0.119)	(0.219)	(0.021)	(0.815)
Destination Ln Population Density at 1500	-0.122**	-0.211	-0.274**	-0.416**	0.069***	0.215*	-0.017	-0.482*
	(0.049)	(0.139)	(0.113)	(0.200)	(0.022)	(0.126)	(0.020)	(0.275)
Destination Malaria Ecology	-0.030***	-0.392***	-0.063***	-0.504***	-0.045***	-0.0335	-0.006*	-0.202**
	(0.008	(0.100)	(0.021)	(0.164)	(0.015)	(0.0504)	(0.003)	(0.095)
Destination Percentage of Arable Lands	0.005	-0.013	0.011	-0.025	0.008	0.0112	-0.001	-0.259
	(0.004)	(0.029)	(0.010)	(0.040)	(0.005)	(0.0135)	(0.001)	(0.194)
Constant	3.560***	17.540***	7.001	23.950**	8.460***	13.710***	0.166	-2.165
	(1.158)	(4.169)	(5.139)	(9.389)	(1.106)	(2.028)	(0.182)	(6.795)
Observations	5.916	5.916	2.516	2.516	1.260	1.260	1.260	1.260
R-Squared	0.010	---	0.019	---	0.105	---	0.084	---
Pseudo R-Squared	---	0.188	---	0.213	---	0.231	---	0.405
Controls for:								
Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Geographical Controls are "Average Elevation", landlock and island dummies, and "Average distance to nearest ice-free coast". *, **, or *** indicates significance at the 10%, 5% and 1% levels, respectively.

Table 33. Bilateral Results with DS from WMM 2000 version

VARIABLES	(1) Europe / OLS	(2) Europe / PPML	(3) SSA / OLS	(4) SSA / PPML	(5) Asia / OLS	(6) Asia / PPML
lndist	-0.00603	-2.464**	-0.00885***	-1.648***	2.87e-05	0.353
	(0.00542)	(1.001)	(0.00117)	(0.217)	(0.000213)	(0.816)
d_ln_pd1500	-0.00267**	-0.412**	0.000699***	0.224*	-0.000227	-0.486*
	(0.00111)	(0.198)	(0.000218)	(0.126)	(0.000204)	(0.269)
d_malaria_ecology	-0.000626***	-0.500***	-0.000445***	-0.0335	-6.46e-05**	-0.197**
	(0.000215)	(0.163)	(0.000151)	(0.0502)	(3.25e-05)	(0.0950)
d_arable	0.000108	-0.0252	7.77e-05	0.0107	-5.07e-06	-0.249
	(9.81e-05)	(0.0406)	(5.45e-05)	(0.0133)	(1.21e-05)	(0.181)
Constant	0.0685	19.20**	0.0839***	9.161***	0.00182	-7.435
	(0.0504)	(9.329)	(0.0109)	(2.014)	(0.00184)	(6.847)
Observations	2,516	2,516	1,260	1,260	2,108	2,108
R-squared	0.019		0.106		0.004	
Controls	YES	YES	YES	YES	YES	YES
Pseudo R-squared		0.148		0.105		0.180

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 34. Bilateral Results with DSv2 from WMM 2000 version

VARIABLES	(1) Europe / OLS	(2) Europe / PPML	(3) SSA / OLS	(4) SSA / PPML	(5) Asia / OLS	(6) Asia / PPML
lndist	-0.0147***	-3.916***	-0.0161***	-3.224***	-0.000142	-1.450
	(0.00433)	(0.616)	(0.00344)	(0.352)	(0.000140)	(1.414)
d_ln_pd1500	-0.00645***	-1.073***	-0.00158*	-0.0162	-0.000295***	-1.072***
	(0.00139)	(0.197)	(0.000816)	(0.119)	(0.000105)	(0.334)
d_malaria_ecology	-0.000322***	-0.296***	0.00107***	0.216***	1.72e-08	-0.337***
	(9.48e-05)	(0.103)	(0.000337)	(0.0640)	(2.34e-06)	(0.0758)
d_arable	0.000450***	0.112***	0.000693***	0.138***	2.82e-05***	0.273***
	(9.77e-05)	(0.0125)	(0.000174)	(0.0166)	(1.09e-05)	(0.0400)
Constant	0.131***	28.84***	0.139***	20.82***	0.00105	-0.983
	(0.0388)	(5.467)	(0.0298)	(3.096)	(0.00127)	(12.71)
Observations	2,516	2,516	1,260	1,260	2,108	2,108
R-squared	0.074		0.076		0.035	
Controls	YES	YES	YES	YES	YES	YES
Pseudo R-squared		0.411		0.178		0.357

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 35. Different Regional Settings: Source is Eurasia and Destination is "New Lands"

VARIABLES	(1) PPML / DSv2	(2) OLS / DSv2	(3) PPML / DS	(4) OLS / DS
lndist	-2.761***	-1.481***	-0.153	-0.00248
	(0.666)	(0.462)	(0.753)	(0.00352)
d_ln_pd1500	-0.704***	-0.356***	-0.273**	-0.00192*
	(0.167)	(0.0804)	(0.128)	(0.00114)
d_malaria_ecology	0.222	0.0230	-0.295*	-0.00100*
	(0.223)	(0.0140)	(0.152)	(0.000538)
d_arable	0.132***	0.0596***	-0.0246	-8.04e-05
	(0.0188)	(0.0144)	(0.0263)	(0.000130)
col_dep_ever	1.820**	4.635*	4.894***	0.289***
	(0.778)	(2.582)	(0.337)	(0.0401)
d_gold_silver	1.468***	0.0805	0.413	0.000779
	(0.471)	(0.0516)	(0.334)	(0.00180)
Constant	19.99***	12.95***	-3.955	0.0292
	(6.064)	(4.128)	(6.991)	(0.0329)
Observations	2,436	2,436	2,436	2,436
R-squared		0.092		0.426
Controls	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 36. Different Regional Settings: Source is Europe and Destination is "Non-Europe"

VARIABLES	(1) PPML / DSv2	(2) OLS / DSv2	(3) PPML / DS	(4) OLS / DS
lndist	-1.686***	-0.163*	0.851**	0.00284**
	(0.575)	(0.0843)	(0.428)	(0.00131)
d_ln_pd1500	-1.605***	-0.364***	-0.404***	-0.00256***
	(0.366)	(0.0864)	(0.128)	(0.000730)
d_malaria_ecology	0.0199	-0.000310	-0.262***	-0.000206***
	(0.0400)	(0.00188)	(0.0979)	(7.44e-05)
d_arable	0.138***	0.0295***	0.00122	0.000111*
	(0.0312)	(0.00749)	(0.0199)	(6.52e-05)
col_dep_ever	1.960**	1.761*	4.629***	0.108***
	(0.841)	(0.924)	(0.273)	(0.0202)
d_gold_silver	2.731***	0.222***	0.346	0.00136
	(0.835)	(0.0596)	(0.301)	(0.00152)
Constant	7.468*	1.190*	-13.30***	-0.0207*
	(4.035)	(0.673)	(3.755)	(0.0107)
Observations	3,552	3,552	3,552	3,552
R-squared		0.056		0.190
Controls	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 37. Different Regional Settings: Source Asia and Destination is "Non-Asia"

VARIABLES	(1) PPML / DSv2	(2) OLS / DSv2	(3) PPML / DS	(4) OLS / DS
lndist	-2.877***	-0.0909*	-0.510	-0.000452
	(0.457)	(0.0501)	(0.680)	(0.000455)
d_ln_pd1500	-3.847**	-0.0370***	-0.278*	-0.000272*
	(1.535)	(0.0125)	(0.166)	(0.000150)
d_malaria_ecology	-0.598	-0.000337	-0.126*	-4.19e-05
	(0.739)	(0.000828)	(0.0707)	(2.59e-05)
d_arable	0.410***	0.00403***	-0.0455	2.02e-06
	(0.108)	(0.00141)	(0.0436)	(9.16e-06)
d_acdummy	-5.184	-0.0218	-0.854	-7.26e-05
	(4.335)	(0.0142)	(0.551)	(0.000167)
d_gold_silver	-4.717**	0.0187	0.450	0.000210
	(2.383)	(0.0138)	(0.845)	(0.000642)
Constant	14.75***	0.835*	0.757	0.00603
	(3.593)	(0.486)	(5.413)	(0.00411)
Observations	2,387	2,387	2,387	2,387
R-squared		0.032		0.004
Controls	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 38. Different Regional Settings: Source Asia and Destination is “New Lands”

VARIABLES	(1) PPML / DSv2	(2) OLS / DSv2	(3) PPML / DS	(4) OLS / DS
lndist	-1.600	-0.153	0.461	0.00143
	(1.569)	(0.0977)	(0.717)	(0.00213)
d_ln_pd1500	-1.347***	-0.0352***	0.101	-0.000648
	(0.464)	(0.0136)	(0.537)	(0.000664)
d_malaria_ecology	0.421**	0.00101	-0.0121	9.12e-05
	(0.213)	(0.00413)	(0.130)	(0.000486)
d_arable	0.298***	0.00895**	-0.110	-0.000102
	(0.0565)	(0.00387)	(0.0783)	(7.39e-05)
d_gold_silver	4.152***	0.0124	0.715	-0.000112
	(1.258)	(0.00871)	(1.411)	(0.00139)
Constant	-0.411	1.308	-7.716	-0.00972
	(15.18)	(0.896)	(7.732)	(0.0188)
Observations	806	868	806	868
R-squared		0.052		0.004
Controls	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

REFERENCES

Abramitzky, R. (2015). Economics and the modern economic historian (Working Paper 21636). National Bureau of Economic Research. https://doi.org/10.3386/w21636

Acemoglu, D., Cantoni, D., Johnson, S., & Robinson, J. A. (2011). The Consequences of radical reform: The French Revolution. American Economic Review, 101(7), 3286–3307. https://doi.org/10.1257/aer.101.7.3286

Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The Colonial origins of comparative development: An empirical investigation. American Economic Review, 91(5), 1369–1401. https://doi.org/10.1257/aer.91.5.1369

Acemoglu, D., Johnson, S., & Robinson, J. A. (2002). Reversal of fortune: geography and institutions in the making of the modern world income distribution. The Quarterly Journal of Economics, 117(4), 1231–1294.

Alesina, A., Giuliano, P., & Nunn, N. (2013). On the origins of gender roles: women and the plough. The Quarterly Journal of Economics, 128(2), 469–530. https://doi.org/10.1093/qje/qjt005

Alesina, A., Harnoss, J., & Rapoport, H. (2016). Birthplace diversity and economic prosperity. Journal of Economic Growth, 21(2), 101–138. https://doi.org/10.1007/s10887-016-9127-6

Alsan, M. (2015). The Effect of the TseTse fly on African development. American Economic Review, 105(1), 382–410. https://doi.org/10.1257/aer.20130604

Anderson, J. E. (1979). A theoretical foundation for the gravity equation. The American Economic Review, 69(1), 106–116.

Anderson, J. E., & van Wincoop, E. (2003). Gravity with gravitas: a solution to the border puzzle. American Economic Review, 93(1), 170–192. https://doi.org/10.1257/000282803321455214

Anderson, J. E., & van Wincoop, E. (2004). Trade costs. Journal of Economic Literature, 42(3), 691–751. https://doi.org/10.1257/0022051042177649

Anselin, L. (1988). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. Geographical Analysis, 20(1), 1–17. https://doi.org/10.1111/j.1538-4632.1988.tb00159.x

Anselin, L. (1995). Local indicators of spatial association—LISA. Geographical Analysis, 27(2), 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Anselin, L., & Griffith, D. A. (1988). Do spatial effects really matter in regression analysis? Papers in Regional Science, 65(1), 11–34. https://doi.org/10.1111/j.1435-5597.1988.tb01155.x

Arif, I. (2022). Educational attainment, corruption, and migration: An empirical analysis from a gravity model. Economic Modelling, 110, 105802. https://doi.org/10.1016/j.econmod.2022.105802

Ashraf, Q., & Galor, O. (2011). Dynamics and stagnation in the Malthusian Epoch. The American Economic Review, 101(5), 2003–2041.

Ashraf, Q., & Galor, O. (2013). The "Out of Africa" hypothesis, human genetic diversity, and comparative economic development. American Economic Review, 103(1), 1–46. https://doi.org/10.1257/aer.103.1.1

Auer, R. A. (2013). Geography, institutions, and the making of comparative development. Journal of Economic Growth, 18(2), 179–215. https://doi.org/10.1007/s10887-013-9087-z

Banerjee, A., & Iyer, L. (2005). History, institutions, and economic performance: the legacy of colonial land tenure systems in India. American Economic Review, 95(4), 1190–1213. https://doi.org/10.1257/0002828054825574

Becker, A., Enke, B., & Falk, A. (2020). Ancient origins of the global variation in economic preferences. AEA Papers and Proceedings, 110, 319–323. https://doi.org/10.1257/pandp.20201071

Becker, S. O., Hornung, E., & Woessmann, L. (2011). Education and catch-up in the Industrial Revolution. American Economic Journal: Macroeconomics, 3(3), 92–126. https://doi.org/10.1257/mac.3.3.92

Becker, S. O., & Woessmann, L. (2009). Was Weber wrong? A human capital theory of protestant economic history. The Quarterly Journal of Economics, 124(2), 531–596. https://doi.org/10.1162/qjec.2009.124.2.531

Beine, M., Bertoli, S., & Fernández-Huertas Moraga, J. (2016). A practitioners' guide to gravity models of international migration. The World Economy, 39(4), 496–512. https://doi.org/10.1111/twec.12265

Beine, M., Docquier, F., & Özden, Ç. (2011). Diasporas. Journal of Development Economics, 95(1), 30–41. https://doi.org/10.1016/j.jdeveco.2009.11.004

Beine, M., & Parsons, C. (2015). Climatic factors as determinants of international migration. The Scandinavian Journal of Economics, 117(2), 723–767. https://doi.org/10.1111/sjoe.12098

Bellégo, C., Benatia, D., & Pape, L. (2022). Dealing with logs and zeros in regression models (arXiv:2203.11820). arXiv. https://doi.org/10.48550/arXiv.2203.11820

Besley, T., & Persson, T. (2009). The origins of state capacity: property rights, taxation, and politics. American Economic Review, 99(4), 1218–1244. https://doi.org/10.1257/aer.99.4.1218

Blackburn, R. (1997). The making of new world slavery: from the baroque to the modern, 1492-1800. Verso.

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

Borcan, O., Olsson, O., & Putterman, L. (2018). State history and economic development: Evidence from six millennia. Journal of Economic Growth, 23(1), 1–40. https://doi.org/10.1007/s10887-017-9152-0

Borcan, O., Olsson, O., & Putterman, L. (2021). Transition to agriculture and first state presence: A global analysis. Explorations in Economic History, 82, 101404. https://doi.org/10.1016/j.eeh.2021.101404

Borjas, G. J. (1989). Economic theory and international migration. The International Migration Review, 23(3), 457–485. https://doi.org/10.2307/2546424

Braudel, F., & Wallerstein, I. (2009). History and the social sciences: The Longue Durée. Review (Fernand Braudel Center), 32(2), 171–203.

Brown, K. W. (2012). A history of mining in Latin America: From the colonial era to the present. UNM Press.

Bu, Y., Wang, E., Qiu, Y., & Möst, D. (2022). Impact assessment of population migration on energy consumption and carbon emissions in China: A spatial econometric investigation. Environmental Impact Assessment Review, 93, 106744. https://doi.org/10.1016/j.eiar.2022.106744

Buggle, J., Mayer, T., Sakalli, S. O., & Thoenig, M. (2023). The refugee's dilemma: evidence from Jewish migration out of Nazi Germany. The Quarterly Journal of Economics, 138(2), 1273–1345. https://doi.org/10.1093/qje/qjad001

Carstensen, K., & Gundlach, E. (2006). The primacy of institutions reconsidered: Direct income effects of malaria prevalence. The World Bank Economic Review, 20(3), 309–339. https://doi.org/10.1093/wber/lhl001

Chanda, A., Cook, C. J., & Putterman, L. (2014). Persistence of fortune: Accounting for population movements, there was no post-Columbian reversal. American Economic Journal: Macroeconomics, 6(3), 1–28. https://doi.org/10.1257/mac.6.3.1

Cohn, J. B., Liu, Z., & Wardlaw, M. I. (2022). Count (and count-like) data in finance. Journal of Financial Economics, 146(2), 529–551. https://doi.org/10.1016/j.jfineco.2022.08.004

Comin, D., Easterly, W., & Gong, E. (2010). Was the wealth of nations determined in 1000 BC? American Economic Journal: Macroeconomics, 2(3), 65–97. https://doi.org/10.1257/mac.2.3.65

Conte, M., Cotterlaz, P., & Mayer, T. (2022). The CEPII gravity database.

De Benedictis, L., & Taglioni, D. (2011). The Gravity Model in international trade. In L. De Benedictis & L. Salvatici (Eds.), The trade impact of European Union preferential policies: an analysis through Gravity Models (pp. 55–89). Springer. https://doi.org/10.1007/978-3-642-16564-1_4

de Haas, H. (2014). Migration theory: Quo vadis? https://ora.ox.ac.uk/objects/uuid:45aacf94-8f24-4294-9c74-cbc8453fcbfb

Deardorff, A. (1998). Determinants of bilateral trade: Does gravity work in a neoclassical world? In the Regionalization of the world economy (pp. 7–32). University of Chicago Press. https://www.nber.org/books-and-chapters/regionalization-world-economy/determinants-bilateral-trade-does-gravity-work-neoclassical-world

Diamond, J. M. (1999). Guns, germs, and steel: The fates of human societies. W.W. Norton & Company.

Dorigo, G., & Tobler, W. (1983). Push-pull migration laws. Annals of the Association of American Geographers, 73(1), 1–17. https://doi.org/10.1111/j.1467-8306.1983.tb01392.x

Easterly, W., & Levine, R. (2003). Tropics, germs, and crops: How endowments influence economic development. Journal of Monetary Economics, 50(1), 3–39. https://doi.org/10.1016/S0304-3932(02)00200-3

Easterly, W., & Levine, R. (2016). The European origins of economic development. Journal of Economic Growth, 21(3), 225–257. https://doi.org/10.1007/s10887-016-9130-y

Elhorst, J. P. (2014). Spatial panel data models. In J. P. Elhorst (Ed.), Spatial Econometrics: From Cross-Sectional Data to Spatial Panels (pp. 37–93). Springer. https://doi.org/10.1007/978-3-642-40340-8_3

Eltis, D. (1999). The rise of African slavery in the Americas. Cambridge University Press. https://doi.org/10.1017/CBO9780511583667

Eltis, D., & Richardson, D. (2013). Routes to slavery: direction, ethnicity and mortality in the transatlantic slave trade. Routledge.

Engerman, S. L., & Sokoloff, K. L. (2002). Factor endowments, inequality, and paths of development among new world economics (Working Paper 9259). National Bureau of Economic Research. https://doi.org/10.3386/w9259

Engerman, S. L., & Sokoloff, K. L. (2005). Colonialism, inequality, and long-run paths of development (Working Paper 11057). National Bureau of Economic Research. https://doi.org/10.3386/w11057

Ertan, A., Fiszbein, M., & Putterman, L. (2016). Who was colonized and when? A cross-country analysis of determinants. European Economic Review, 83, 165–184. https://doi.org/10.1016/j.euroecorev.2015.10.012

Europeans on the move: Studies on European migration 1500–1800. (1994). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198204190.001.0001

Fenske, J. (2011). Land tenure and investment incentives: Evidence from West Africa. Journal of Development Economics, 95(2), 137–156. https://doi.org/10.1016/j.jdeveco.2010.05.001

Foubert, K., & Ruyssen, I. (2024). Global migration and the role of terrorist attacks. Journal of Economic Behavior & Organization, 220, 507–530. https://doi.org/10.1016/j.jebo.2024.02.022

Frankel, J. A., & Romer, D. H. (1999). Does trade cause growth? American Economic Review, 89(3), 379–399. https://doi.org/10.1257/aer.89.3.379

Gallup, J. L., Sachs, J. D., & Mellinger, A. D. (1999). Geography and economic development. International Regional Science Review, 22(2), 179–232. https://doi.org/10.1177/016001799761012334

Glaeser, E. L., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2004). Do institutions cause growth? Journal of Economic Growth, 9(3), 271–303. https://doi.org/10.1023/B:JOEG.0000038933.16398.ed

Greif, A. (1993). Contract enforceability and economic institutions in early trade: the Maghribi traders' coalition. The American Economic Review, 83(3), 525–548.

Greif, A. (1994). Cultural beliefs and the organization of society: A historical and theoretical reflection on collectivist and individualist societies. Journal of Political Economy, 102(5), 912–950. https://doi.org/10.1086/261959

Greif, A. (1998). Historical and comparative institutional analysis. The American Economic Review, 88(2), 80–84.

Greif, A. (2006). Institutions and the path to the modern economy: Lessons from medieval trade. Cambridge University Press.

Griffith, D. A., & Jones, K. G. (1980). Explorations into the relationship between spatial structure and spatial interaction. Environment and planning a: economy and space, 12(2), 187–201. https://doi.org/10.1068/a120187

Grogger, J., & Hanson, G. H. (2011). Income maximization and the selection and sorting of international migrants. Journal of Development Economics, 95(1), 42–57. https://doi.org/10.1016/j.jdeveco.2010.06.003

Haas, H. de, Castles, S., & Miller, M. J. (2019). The Age of migration: International population movements in the modern world. Bloomsbury Publishing.

Hall, R. E., & Jones, C. I. (1999). Why do some countries produce so much more output per worker than others? The Quarterly Journal of Economics, 114(1), 83–116.

Head, K., & Mayer, T. (2014). Chapter 3 - Gravity equations: Workhorse, toolkit, and cookbook. In G. Gopinath, E. Helpman, & K. Rogoff (Eds.), Handbook of International Economics (Vol. 4, pp. 131–195). Elsevier. https://doi.org/10.1016/B978-0-444-54314-1.00003-3

Head, K., Mayer, T., & Ries, J. (2010). The erosion of colonial trade linkages after independence. Journal of International Economics, 81(1), 1–14. https://doi.org/10.1016/j.jinteco.2010.01.002

Hoerder, D. (2002). Cultures in Contact: World migrations in the second millennium. Duke University Press.

Jeanty, P. W., Partridge, M., & Irwin, E. (2010). Estimation of a spatial simultaneous equation model of population migration and housing price dynamics. Regional Science and Urban Economics, 40(5), 343–352. https://doi.org/10.1016/j.regsciurbeco.2010.01.002

Kaiser, U., Mendez, S. J., Rønde, T., & Ullrich, H. (2014). Regulation of pharmaceutical prices: Evidence from a reference price reform in Denmark. Journal of Health Economics, 36, 174–187. https://doi.org/10.1016/j.jhealeco.2014.04.003

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

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. W. (1998). Law and finance. Journal of Political Economy, 106(6), 1113–1155. https://doi.org/10.1086/250042

LeSage, J. P., & Pace, R. K. (2010). Spatial econometric models. In M. M. Fischer & A. Getis (Eds.), Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications (pp. 355–376). Springer. https://doi.org/10.1007/978-3-642-03647-7_18

Lovejoy, P. E. (2011). Transformations in slavery: A history of slavery in Africa. Cambridge University Press.

Lucassen, J., & Lucassen, L. (2009). The mobility transition revisited, 1500–1900: What the case of Europe can offer to global history. Journal of Global History, 4(3), 347–377. https://doi.org/10.1017/S174002280999012X

Lucassen, J., & Lucassen, L. (2014). Globalising migration history: The Eurasian experience (16th-21st Centuries). BRILL. http://ebookcentral.proquest.com/lib/bogazici-ebooks/detail.action?docID=1694719

Manning, P., & Trimmer, T. (2020). Migration in world history (3rd ed.). Routledge. https://doi.org/10.4324/9781351256681

Marx, K. (2000). Karl Marx: Selected writings. Oxford University Press.

Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., & Taylor, J. E. (1993). Theories of international migration: A review and appraisal. Population and Development Review, 19(3), 431–466. https://doi.org/10.2307/2938462

Maurel, M., & Tuccio, M. (2016). Climate instability, urbanisation and international migration. The Journal of Development Studies, 52(5), 735–752. https://doi.org/10.1080/00220388.2015.1121240

McKeown, A. (2004). Global migration, 1846-1940. Journal of World History, 15(2), 155–189.

Michalopoulos, S. (2012). The origins of ethnolinguistic diversity. American Economic Review, 102(4), 1508–1539. https://doi.org/10.1257/aer.102.4.1508

Michalopoulos, S., & Papaioannou, E. (2016). The long-run effects of the scramble for Africa. American Economic Review, 106(7), 1802–1848. https://doi.org/10.1257/aer.20131311

Mitnik, P. A., & Grusky, D. B. (2020). The intergenerational elasticity of what? The case for redefining the workhorse measure of economic mobility. Sociological Methodology, 50(1), 47–95. https://doi.org/10.1177/0081175019886613

Moch, L. P. (2003). Moving Europeans, second edition: Migration in Western Europe since 1650. Indiana University Press.

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. Biometrika, 37(1/2), 17–23. https://doi.org/10.2307/2332142

Nejad, M. N., & Young, A. T. (2016). Want freedom, will travel: Emigrant self-selection according to institutional quality. European Journal of Political Economy, 45, 71–84. https://doi.org/10.1016/j.ejpoleco.2016.06.002

North, D. C. (1990). Institutions, institutional change and economic performance. Cambridge University Press.

Nunn, N. (2008). The long-term effects of Africa's slave trades. The Quarterly Journal of Economics, 123(1), 139–176. https://doi.org/10.1162/qjec.2008.123.1.139

Nunn, N., & Puga, D. (2012). Ruggedness: The blessing of bad geography in Africa. The Review of Economics and Statistics, 94(1), 20–36. https://doi.org/10.1162/REST_a_00161

Olsson, O. (2009). On the democratic legacy of colonialism. Journal of Comparative Economics, 37(4), 534–551. https://doi.org/10.1016/j.jce.2009.08.004

Olsson, O., & Hibbs, J. (2004). Geography, biogeography and why some countries are rich and others poor [Report]. https://gupea.ub.gu.se/handle/2077/2780

Olsson, O., & Paik, C. (2016). Long-run cultural divergence: Evidence from the Neolithic Revolution. Journal of Development Economics, 122, 197–213. https://doi.org/10.1016/j.jdeveco.2016.05.003

Parkins, N. C. (2010). Push and pull factors of migration. American Review of Political Economy, 8(2), Article 2. https://doi.org/10.38024/arpe.119

Putterman, L. (2008). Agriculture, diffusion and development: Ripple effects of the Neolithic Revolution. Economica, 75(300), 729–748. https://doi.org/10.1111/j.1468-0335.2007.00652.x

Putterman, L., & Weil, D. N. (2010). Post-1500 population flows and the long run determinants of economic growth and inequality. The Quarterly Journal of Economics, 125(4), 1627–1682. https://doi.org/10.1162/qjec.2010.125.4.1627

Ramos, R. (2016). Gravity models: A tool for migration analysis. IZA World of Labor. https://doi.org/10.15185/izawol.239

Rawley, J. A., & Behrendt, S. D. (2005). The transatlantic slave trade: A history. U of Nebraska Press.

Robinson, J., Acemoglu, D., & Johnson, S. (2005). Institutions as a fundamental cause of long-run growth. Handbook of Economic Growth, 1A, 386–472.

Sachs, J. D. (2003). Institutions don't rule: Direct effects of geography on per capita income (Working Paper 9490). National Bureau of Economic Research. https://doi.org/10.3386/w9490

Sachs, J. D. (2012). Government, geography, and growth: the true drivers of economic development. Foreign Affairs, 91(5), 142–150.

Santos Silva, J. M. C., & Tenreyro, S. (2022). The log of gravity at 15. Portuguese Economic Journal, 21(3), 423–437. https://doi.org/10.1007/s10258-021-00203-w

Sen, A., & Smith, T. E. (2012). Gravity Models of spatial interaction behavior. Springer Science & Business Media.

Silva, J. M. C. S., & Tenreyro, S. (2006). The log of gravity. The Review of Economics and Statistics, 88(4), 641–658. https://doi.org/10.1162/rest.88.4.641

Spolaore, E., & Wacziarg, R. (2013). How deep are the roots of economic development? Journal of Economic Literature, 51(2), 325–369. https://doi.org/10.1257/jel.51.2.325

Thomas, H. (1997). The slave trade: the story of the Atlantic slave trade: 1440 - 1870. Simon and Schuster.

Thornton, J. K. (1998). Africa and Africans in the making of the Atlantic world, 1400-1800. Cambridge University Press.

Tobler, W. (1995). Migration: Ravenstein, Thornthwaite, and beyond. Urban Geography, 16(4), 327–343. https://doi.org/10.2747/0272-3638.16.4.327

Voigtländer, N., & Voth, H.-J. (2012). Persecution perpetuated: The medieval origins of anti-semitic violence in Nazi Germany. The Quarterly Journal of Economics, 127(3), 1339–1392. https://doi.org/10.1093/qje/qjs019

Vollrath, D. (2011). The agricultural basis of comparative development. Journal of Economic Growth, 16(4), 343–370. https://doi.org/10.1007/s10887-011-9074-1

Weber, M. (2009). From Max Weber: Essays in sociology. Routledge. https://doi.org/10.4324/9780203452196

Wilson, A. G. (1971). A family of spatial interaction models, and associated developments. Environment and Planning a: Economy and space, 3(1), 1–32. https://doi.org/10.1068/a030001

Zwart, P. de, & Zanden, J. L. van. (2018). The origins of globalization: World trade in the making of the global economy, 1500-1800. Cambridge University Press.