



Munich Personal RePEc Archive

# **Population Diversity, Division of Labor and Comparative Development**

Depetris-Chauvin, Emilio and Özak, Ömer

Departamento de Economía, Pontificia Universidad Católica de  
Chile, Department of Economics, Southern Methodist University

4 April 2016

Online at <https://mpra.ub.uni-muenchen.de/70503/>

MPRA Paper No. 70503, posted 13 Apr 2016 13:57 UTC

# Population Diversity, Division of Labor and Comparative Development\*

Emilio Depetris-Chauvin<sup>†</sup> and Ömer Özak<sup>‡</sup>

April 4, 2016

## Abstract

This research explores the emergence and prevalence of economic specialization and trade in pre-modern societies. It advances the hypothesis, and establishes empirically that population diversity had a positive causal effect on economic specialization and trade. Based on a novel ethnic level dataset combining geocoded ethnographic, linguistic and genetic data, this research exploits the exogenous variation in population diversity generated by the “Out-of-Africa” migration of anatomically modern humans to causally establish that higher levels of population diversity were conducive to economic specialization and the emergence of trade-related institutions that, in turn, translated into pre-modern era differences in comparative development. Additionally, this research provides suggestive evidence that regions historically inhabited by pre-modern societies with high levels of economic specialization have higher levels of contemporary occupational heterogeneity, economic complexity and development.

*Keywords: Economic Specialization, Division of Labor, Trade, Comparative Development, Economic Development, Population Diversity, Population Heterogeneity, Genetic Diversity, Linguistic Diversity, Diversity, Persistence, Out of Africa, Serial Founder Effect*

*JEL Classification: D74, F10, N10, O10, O11, O12, O40, Z10*

---

\*We wish to thank Klaus Desmet, Oded Galor, Pete Klenow, Stelios Michalopoulos, Dan Millimet, and David Weil, as well as participants of the 2015 Meeting of the Latin American and Caribbean Economic Association (LACEA), and seminar participants at Brown University, Clark University, University of Connecticut, Banco de la República de Colombia, Universidad de los Andes, Universidad del Rosario, and Universidad Nacional de Colombia, for useful comments and discussions. Additionally, we thank Anthon Eff for sharing the EA and SCCS datasets, and James Fenske for sharing his data for the mapping of ethnic groups to their historical homelands. An earlier version of this paper circulated under the title “Population Diversity, Division of Labor and the Emergence of Trade and State”.

<sup>†</sup>Departamento de Economía, Pontificia Universidad Católica de Chile. E-mail: edepetris@uc.cl

<sup>‡</sup>Department of Economics, Southern Methodist University. E-mail: ozak@smu.edu

“The greatest improvement in the productive powers of labour, and the greater part of the skill, dexterity, and judgment with which it is any where directed, or applied, seem to have been the effects of the division of labour.”

Adam Smith (1776)

# 1 Introduction

At least since Adam Smith, division of labor and economic specialization have been considered fundamental to comparative economic development. Their importance is attributed to their essential role in increasing trade, productivity, innovation and economic growth, as well as to their positive effects on institutions. As occupational specialization has been prevalent since pre-modern times, with a complex division of labor that often involved specialization by communities and regions (Nolan and Lenski, 1999), it is not surprising that variations in the existence and extent of trade in the pre-colonial era may partially explain observed differences in contemporary economic outcomes across countries and regions (Berg, 1991; Findlay and O’Rourke, 2007; Greif, 1993; Putterman and Weil, 2010; Smith, 1776).<sup>1</sup> Interestingly, little, if anything, is known about the deep-rooted historical determinants of the division of labor and economic specialization.

This research explores the emergence and prevalence of economic specialization and trade in pre-modern societies. It advances the hypothesis, and establishes empirically that population diversity had a positive causal effect on economic specialization and trade. Based on a novel ethnic level dataset combining geocoded ethnographic, linguistic and genetic data, this research exploits the exogenous variation in population diversity generated by the “Out-of-Africa” migration of anatomically modern humans to causally establish that higher levels of population diversity were conducive to economic specialization and the emergence of trade-related institutions that, in turn, translated into pre-modern era differences in comparative development. Additionally, this research provides suggestive evidence that regions historically inhabited by pre-modern societies with high levels of economic specialization have higher levels of contemporary occupational heterogeneity, economic complexity and development.

In particular, this research proposes the hypothesis that higher levels of population diversity during the pre-modern era were conducive to economic specialization and trade. A

---

<sup>1</sup> Additionally, experience with trade and trade enhancing institutions during the pre-modern era has been associated with interethnic tolerance (Jha, 2013), religious adherence (Michalopoulos et al., 2012), Western European economic growth (Acemoglu et al., 2005), and European city growth (De Long and Shleifer, 1993).

diverse population implied larger variations in preferences and abilities across individuals. This in turn increased the complementarities between preferences, abilities and the environment, fostering the emergence of trade due to the potential gains of increased specialization. Thus, the theory predicts that during the pre-modern era economic specialization and trade should emerge and be more prevalent among diverse populations. Consequently, and as a result of the aforementioned gains from specialization, pre-modern era societies with more diverse populations also benefited from higher levels of prosperity in pre-modern times as measured by their level of societal complexity, technological specialization, population density, and the scope of their centralized institutions.

To empirically test this hypothesis, this research constructs a novel dataset of ethnic level measures of pre-modern economic specialization and population diversity. By performing the analysis at the *ethnic level* in pre-modern societies, the analysis sidesteps potential pitfalls from the aggregation of data to the country level as well as from the effects of migrations and population replacements in the post-1500CE era. Moreover, the *intra-ethnic* analysis performed in this research overcomes the potential confounding effects of *country-level inter-ethnic* measures of diversity, which have been widely exploited in the previous literature.

The research constructs novel measures of the number of economic activities in which specialization existed in the pre-modern era for over 1100 ethnicities. Additionally, the analysis constructs proxies of historical population diversity by exploiting data on genetic and linguistic diversity for around 300 ethnic groups. An essential feature of these diversity measures is that a great deal of their variation was largely generated exogenously by a fundamental historical process, namely the serial founder effect (SFE) behind the dispersion of anatomically modern humans out of East Africa more than 60,000 years ago (Ramachandran et al., 2005). In particular, the SFE implies that successive divisions of an original population into various subpopulations generates a loss of diversity in cultural traits, preferences, knowledge, abilities, and other intergenerationally transmitted characteristics, similar to the loss in genetic and linguistic diversity, especially in an era when knowledge and culture, among others, were passed orally between generations.<sup>2</sup> Additionally, given the African origin of modern humans, the SFE implies that diversity decreases along migratory routes from East Africa. Importantly, it has been established that genetic and linguistic diversity decrease with the migratory distance from East Africa (Atkinson, 2011; Manica et al., 2007; Ramachandran et al., 2005). Given the especially strong empirical and causal relation between migratory distances from East Africa and genetic diversity, this measure plays a central role in the

---

<sup>2</sup>This is not the first research suggesting the existence of SFE for non-genetic traits. E.g. the diversity of phenotypes, phonemes, handaxes, and arrow heads has been shown to decrease along migratory paths (Atkinson, 2011; Hamilton and Buchanan, 2009; Lycett, 2008; Lycett and von Cramon-Taubadel, 2008).

analysis. Still, it is important to note that similar results would be obtained if, instead, one were to use linguistic diversity. This suggests these measures capture general aspects of population diversity above and beyond genes.

The research establishes the robust positive causal effect of population diversity on the emergence and prevalence of economic specialization and trade in pre-modern societies in various steps. First, using ordinary least squares and a restricted sample of 116 ethnicities for which ethnic and genetic data exist, the empirical analysis documents the robust positive statistically and economically significant relation between diversity and economic specialization. Clearly, these statistical associations do not necessarily imply causality and could arise from omitted confounders, such as heterogeneity in environmental factors, or as a result of reverse causality from either propensity to trade or the emergence of trade-related institutions on population compositions.

In order to overcome these potential concerns, the research follows various strategies. First, it accounts for the confounding effect of a large set of geographical and climatic controls, such as absolute latitude, average elevation, terrain ruggedness, accessibility to navigable water, average temperature, etc. Second, it establishes that the main results are not driven by other competing hypotheses on the emergence of trade such as variation in agricultural suitability, ecological diversity, and spatial and intertemporal temperature volatility. Third, it follows an instrumental variable approach by exploiting the Out-of-Africa hypothesis, which posits that migratory distance from East Africa is strongly negatively related to diversity at the ethnic group level. By exploiting these three strategies jointly, the research establishes the positive causal effect of population diversity on the emergence and prevalence of economic specialization for the restricted sample of 116 ethnic groups.

In a second stage of the analysis, the research exploits the predictions of the Out-of-Africa hypothesis and the variations in the distance to East Africa in order to generate predicted population diversity measures (Ashraf and Galor, 2013b). This allows the analysis to be performed on a sample of more than 900 ethnicities. By increasing the sample size, the analysis overcomes potential concerns regarding geographical coverage and representativeness of the restricted sample. Furthermore, by increasing the sample size the effect of population diversity can be estimated more precisely. Reassuringly, and in line with the proposed hypothesis, the estimated causal effect of population diversity on the emergence and prevalence of economic specialization and trade is statistically and economically significant. In particular, a one standard deviation increase in population diversity generates about half a standard deviation increase in economic specialization. Moreover, the research establishes the positive complementary effect of diversity in population and environment on the emergence and prevalence of economic specialization. These results are robust to accounting for

other historical confounding processes such as the number of years since (a) the Neolithic revolution and (b) first settlement. Moreover, the analysis establishes that the positive effect of diversity on specialization is robust to the distance to technological frontiers and to the presence of centralized institutions. Additionally, the analysis establishes that population diversity has a positive causal effect on trade and trade related institutions like money and credit.

In a third stage of the analysis, the research explores the positive effect of population diversity and the mediating effect of specialization on pre-industrial economic development. In order to overcome potential endogeneity concerns due to reverse causality, the analysis exploits the heteroskedastic structure of the residuals to identify the exogenous variation in economic specialization (Lewbel, 2012). In particular, it establishes that population diversity had a positive causal effect on pre-industrial economic development. Additionally, it shows that this effect is largely mediated by its effect on economic specialization. Thus, suggesting that economic specialization is a crucial mechanism behind the relation between economic development and population diversity.

Finally, the research establishes the persistent effect of economic specialization on economic development. In particular, it shows that regions historically inhabited by pre-modern societies with higher levels of economic specialization have higher levels of contemporary development, economic complexity, and occupational heterogeneity. This suggests a novel channel through which deep historical factors affect contemporary economic development (Spolaore and Wacziarg, 2013).

This research is the first attempt to identify the deep-rooted historical factors behind economic specialization and the emergence of trade, as well as their effect on comparative economic development. Moreover, it is the first to identify the positive causal effect of (i) population diversity on economic specialization and the emergence of trade, and (ii) pre-modern economic specialization on economic development. In doing so, this research contributes to three strands of literature.

First, this research contributes to the literature on the effects of diversity on economic development, which has previously been explored using various measures of genetic, ethnic, cultural, and religious diversity (Alesina et al., 2003; Alesina, Harnoss and Rapoport, 2013; Ashraf and Galor, 2013a,b; Desmet et al., 2012, 2015; Easterly and Levine, 1997). Although economic theory suggests that higher diversity should be beneficial for productivity and, thus, development, due to larger skill complementarities, the empirical evidence on the benefits of diversity is strikingly absent. Notably, the existing empirical evidence at the country level suggests that population diversity adversely affects social cohesiveness, trust and development.

Second, this research contributes to the emerging literature that uses genetic diversity to understand the deep-rooted determinants of modern comparative development and of diversity in general. In particular, Arbatli et al. (2013) have argued that genetic diversity provides a “deeper” and better measure of diversity at the country level. Similarly, using cross-country data it has been shown that a sizable variation in income (Ashraf and Galor, 2013b), prevalence of civil conflict (Arbatli et al., 2013), mistrust and cultural fragmentation (Ashraf and Galor, 2013a) can be attributed to variations in genetic diversity.<sup>3</sup> By additionally using linguistic diversity data, this is the first paper to exploit non-genetic historical measures of intra-ethnic diversity. In particular, the results of this research are robust to the measure of historical population diversity used, which sheds new light on the role of genetic diversity and diversity in general as a source of variation in comparative development.

Finally, this research contributes to the literature on the deep-rooted historical sources of contemporary economic development (Alesina, Giuliano and Nunn, 2013; Ashraf and Galor, 2013b; Spolaore and Wacziarg, 2013) by providing a novel channel through which historical conditions determined thousands of years ago still have an effect today.

The remainder of the paper is organized as follows. Section 2 presents ethnographic evidence in favor of the hypothesis. Section 3 describes the data. Section 4 presents the empirical analysis on the impact of population heterogeneity on economic specialization and trade, as well as the effect of economic specialization on pre-industrial development. Section 5 analyzes the persistent effect of pre-industrial economic specialization on modern economic development. Section 6 concludes.

## 2 Ethnographic Narratives on Population Diversity, Division of Labor, and Trade

This section presents ethnographic evidence in support of the hypothesis that higher levels of population diversity during the pre-modern era were conducive to economic specialization of labor and trade.

An illustrative example of the link between diversity and division of labor and trade is given by the Konso people of South-Western Ethiopia and the Aché people of Eastern Paraguay. These two ethnic groups are located on both extremes of the sample distribution of genetic diversity, separated by more than five standard deviations from each other. Due to their proximity to the Ethiopian rift valley, Konso’s genetic diversity is among the highest

---

<sup>3</sup>Genetic distance between populations has been also studied and linked to differences in income (Spolaore and Wacziarg, 2009a), level of trust and bilateral trade (Guiso et al., 2009), and the propensity to engage in conflict (Spolaore and Wacziarg, 2009b).

in the world; while the Aché is the less diverse group in the sample of societies analyzed in this research. For thousands of years, both groups inhabited remote locations with little influence from outsiders (Hallpike, 1972, 2008; Hill and Hurtado, 1996). The ecological environment for both societies was hard and not particularly rich. More specifically, the Konso historically lived in a rocky high elevation (Freeman and Pankhurst, 2003), whereas the Aché inhabited a flat tropical forest (Hill and Hurtado, 1996). The difference in diversity between these two groups maps into differences in their economic specialization of labor. In particular, according to the Ethnographic Atlas, the Konso have labor specialization in 5 activities, whereas the Aché have none. Moreover, when it comes to trade activities, the two groups were very dissimilar as well. Markets were ancient in Konso society and held daily at different locations (Hallpike, 1968), with artisans selling wares, farmers selling grains, butter, and honey, as well as butchers selling raw meat. Contrarily, there was no trade either between the Aché and outsiders nor within the Aché people in pre-modern times (Hill and Hurtado, 1996).

The role of trade on early state formation has been well discussed in the historical, anthropological, political science, and economics literature (Bates, 1983; Braudel, 1972; Fenske, 2014; Gluckman, 1941; Polanyi, 1957; Service, 1978). In fact, a strand of literature in archeology proposes an adaptationist model of specialization, exchange, and state formation, which hypothesizes that centralization of power develops in regions where high resource diversity facilitated regional trade. In this sense, this adaptationist theory proposes that centralized government emerged to maintain peace and mediate diverse interests of different specialists across regions (Sanders, 1965). Additionally, full-time specialists are found in almost every study of early states (Claessen and Skalník, 1978). Despite of the strong link between economic specialization of labor, trade and statehood already documented in previous works, the direction of causality is hard to identify. Nonetheless, examples of highly centralized societies without division of labor and not engaged in trade are virtually absent in the anthropological, archeological and historical literature on pre-modern societies. On the contrary, several examples of stateless pre-modern societies engaging in trade activities and having a noticeable division of labor suggest that statehood was not a necessary precondition for trade and specialization (see e.g. Table 10 below). In particular, examine the case of the Konso of Ethiopia, discussed above, who have a high degree of specialization without any level of jurisdictional hierarchy above the local level. Similarly, consider the cases of the Karen in Myanmar and the Guajiros at the Colombia-Venezuela border. The Karen people are a culturally and linguistically diverse and historically stateless society that have traditionally traded cotton, forest products, and domestic animals to neighboring Burmese and Hmong people -another stateless society- in exchange for rice, pottery, and salt (Hinton, 1979). The



Guajiros, mostly a pastoralist society nowadays, were an egalitarian society that historically based their economy on gathering, hunting, horticulture, and fishing activities depending of the location (Perrin, 1996). According to early European explorers, around the 15th century there were several indigenous groups living in the homeland of the Guajiros, but all those groups were probably part of the same society receiving different names depending of the different economic activities they practiced (Perrin, 1996). Trade was historically important among the Guajiros who commonly held weekly markets (Perrin, 1996).

### 3 Data

This section introduces measures of economic specialization, trade, pre-industrial economic development, historical population diversity, and geographical controls at the ethnic level required by the empirical strategy. In particular, it explains the sources and construction of the various measures used in the analysis.

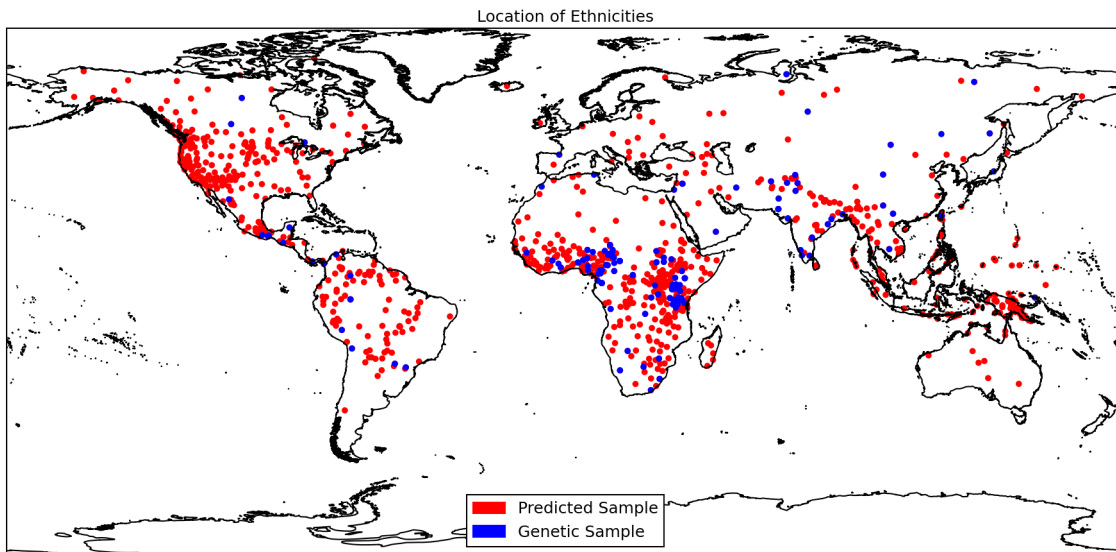


Figure 1: Location of Ethnicities employed in the Analysis (Full and Restricted Samples)

#### 3.1 Dependent Variables: Economic Specialization, Trade and Pre-Industrial Development

The analysis employs the two main sources for ethnic level data currently available, namely the Ethnographic Atlas (EA) and the Standard Cross-Cultural Sample (SCCS). Both datasets have been widely used in anthropology and economics for the study on pre-industrial societies and the long-term effects of pre-industrial culture and institutions (Alesina, Giuliano

and Nunn, 2013; Fenske, 2014; Gennaioli and Rainer, 2007; Michalopoulos and Papaioannou, 2013). The Ethnographic Atlas (Murdock, 1967) includes information on 115 characteristics for 1267 ethnicities around the globe. On the other hand, the Standard Cross-Cultural Sample (Murdock and White, 1969) expands the set of characteristics to over 2000 for a subsample of 180 independent ethnicities. By combining both datasets the analysis overcomes the restriction in terms of thematic coverage of the EA and ethnic/geographic coverage of the SCCS.<sup>4</sup> Figure 1 depicts the location of the full sample of ethnicities used in the main analysis. Additionally, it highlights the ethnicities for which genetic data is available and those for which it is predicted as explained below.

In order to analyze the impact of population diversity on trade and pre-industrial comparative development, the analysis employs various trade related measures from the SCCS and, additionally, constructs a novel measure of economic specialization at the ethnic level using data from both the EA and SCCS. In particular, the analysis employs the following measures from the SCCS: the importance of trade (v819), inter-community trade as food source (v1), money (media of exchange) and credit (v17), credit source (v18), writing and records (v149), technological specialization (v153), complexity (v158.1), and population density (v64). Additionally, the analysis employs measures of the mean size of local communities (v31), the level of statehood (v33) and class stratification from the EA as additional measures of pre-industrial comparative development. As the trade variables from the SCCS are only available for a small subset of ethnicities, especially once the availability of genetic information is taken into account, the main analysis of the impact of population diversity on trade uses a novel measure of economic specialization as the dependent variable.

In particular, since the EA does not have any direct measures of trade, the analysis uses the data available in order to construct various measures of economic specialization. In particular, both data sets include variables on the existence of “age or occupational specialization” for metal working (v55), weaving (v56), leather working (v57), pottery making (v58), boat building (v59), house construction (v60), gathering (v61), hunting (v62), fishing (v63), animal husbandry (v64), and agriculture (v65). For each of these activities, the EA and SCCS assess if the ethnic group had “craft, industrial or age specialization” or if the “activity was absent or no specialization occurred”. These variables allow the identification of ethnicities in which specialization existed in the pre-modern era. On the other hand, these variables do not allow for the differentiation of ethnicities where no specialization occurred from those in which the activity was absent, thus confounding the lack of specialization with

---

<sup>4</sup>The main reason behind the construction of the SCCS was to overcome Galton’s independence problem, i.e., the difficulties of drawing inferences from cross-cultural data due to spatial auto-correlation. The sample of ethnicities in the SCCS were chosen so as to minimize this problem.

the lack of the activity. In order to overcome this problem, the analysis uses additional information from variables v44-v54 in order to assess, for the same activities, whether the activity was “absent or unimportant” or “present”.

Based on this information, the analysis constructs three measures of specialization. The first measure of the level of specialization in ethnicity  $e$ ,  $s_e^1$  counts the number of specialized activities, i.e.  $s_e^1 = \sum_a s_{ea}$ , where  $s_{ea}$  equals 1 if the activity was present and specialized in ethnicity  $e$  and zero otherwise. The second measure of the level of specialization in ethnicity  $e$ ,  $s_e^2$  is the share of activities present that were specialized, i.e.  $s_e^2 = s_e^1/n_e$ , where  $s_e^1$  is the first measure and  $n_e$  is the number of activities available in ethnicity  $e$ . Finally, the third measure of the level of specialization is  $s_e^3 = \sum_a \tilde{s}_{ea}$ , where  $\tilde{s}_{ea}$  equals 0 if the activity  $a$  is not present, 1 if it is present but it is not specialized, and 3 if the activity is present and specialized in ethnicity  $e$ .<sup>5</sup> The main dependent variable in the analysis is the number of specialized activities in an ethnicity,  $s_e^1$ , but the results remain qualitatively unchanged when using the other measures. Reassuringly, these new measures correlate strongly among themselves and with the trade measures from the SCCS (Table A.3).

### 3.2 Independent variables: Population Diversity

This research constructs a novel dataset on georeferenced population diversity at the ethnicity level using two proxies, namely genetic and linguistic diversity. It is important to note that both measures capture intra-ethnic population diversity as opposed to inter-ethnic diversity, which has been widely used in the existing literature that analyzes cross-country differences in population diversity.

The analysis constructs a novel dataset on georeferenced genetic diversity at the ethnicity level using the most comprehensive genomic data set on human micro-satellite variation to date (Pemberton et al., 2013). In particular, Pemberton et al. (2013) combine eight previous population-genetic data sets and analyze them following a standardized procedure, which ensures all the data is produced following a uniform method, ensuring comparability across populations and samples. This data set contains information on 645 common single-nucleotide protein (SNP) loci for 5435 individuals from 267 independent ethnicities. There are two main advantages of using this data. First, it is based on predominantly indigenous populations (Pemberton et al., 2013), which ensures the population inhabited the same location for a prolonged period of time and lowers a potential concern generated by a possible admixture of populations. Second, the SNP’s included in the analysis are “neutral” to selec-

---

<sup>5</sup>The analysis assigns a higher value to specialization in order to differentiate the effect of specialization from technological development. Reassuringly, using a value of 2 for specialization does not alter the main results.

tion, i.e. they are not involved in processes that encode proteins and thus are not subject to natural selection (Kimura, 1983).

Based on this data, this research constructs for each ethnicity a measure of genetic diversity based on what population geneticists call the expected heterozygosity within a population. In particular, the genetic diversity or expected heterozygosity of a population measures the average probability that two randomly chosen individuals in the population do not share the same allele of a gene, i.e. that they do not have the same variant form of the gene.<sup>6</sup> In order to ensure comparability across populations, the analysis constrains the construction of the genetic diversity to the set of 619 common SNP loci for which information exists for all ethnic groups.<sup>7</sup>

Finally, out of the 267 ethnicities this research is able to match a subset of 149 ethnicities to the Ethnographic Atlas (EA). This maps the genetic diversity data to the EA, and thus, to all the cultural, institutional and geographic data contained in the EA or to other data sets to which the EA can be mapped. In particular, and as discussed below, ethnicities can be mapped to the geographical characteristics of their historical homelands.

In order to expand the sample, the analysis generates predicted levels of genetic diversity for the full sample of 1265 ethnicities available in the EA. In particular, according to the "Out-of-Africa" theory of the geographic origin and early migration of anatomically modern humans (Cann et al., 1987; Pemberton et al., 2013; Ramachandran et al., 2005), genetic diversity decreases with the distance from East Africa due to the serial founder effect. Thus, the analysis exploits the variations in the pre-historical migratory distance to East Africa (Addis Ababa) in order to generate the predicted genetic diversity for the full sample of ethnicities available in the EA.

Finally, the research uses linguistic diversity as an alternative proxy of population diversity. In particular, the analysis employs measures of consonant inventories, vowel quality inventories, and the number of genders (Dryer, 2013) as measures of intra-ethnic (language) diversity.

---

<sup>6</sup>The literature on diversity has measured this population attribute using various characteristics like religion, language, ethnicity, or genetics. Diversity within a population is usually defined as the probability that two random individuals in a population do not share the same characteristic. For example, religious, linguistic or ethnic diversity/fractionalization estimate the probability that two random individuals in a population do not share the same religion, speak the same language or have the same ethnic background. Similarly, genetic diversity or expected heterozygosity measure the expected genetic similarity between any two individuals in a population. It is important to note that *all* these measures capture diversity and do not measure any innate superiority of a certain type of characteristic over another. For example, a population in which there exists only one religion, language, ethnicity, or blood type, will be less diverse than one in which there are many, but the measures of diversity do not and cannot be used to identify if one *specific* religion, language, ethnicity or blood type is better than others.

<sup>7</sup>The genetic diversity on the full set of 645 loci is almost perfectly correlated with the measure used in the paper for the 267 original ethnicities in Pemberton et al. (2013). Their correlation is 0.99 ( $p < 0.01$ ).

### 3.3 Geographical Controls

An ethnicity’s pattern of economic activities, opportunities to trade, as well as its genetic and linguistic diversity may be confounded with the geographical characteristics of the ethnicity’s homeland. Thus, the analysis accounts for a large set of geographical controls in order to attenuate any concerns about omitted variable bias. In particular, using the mapping between geographic information systems (GIS) geometries of ethnic homelands and the EA and SCCS generated by Fenske (2014), the analysis constructs for each ethnicity a large set of geographical characteristics of its homeland. Tables A.1-A.2 show the list of all variables and their summary statistics for the various samples used in the paper.

## 4 Empirical Analysis

This section analyzes the effect of population diversity, as measured by intra-ethnic genetic and linguistic diversity, on economic specialization of labor, trade, and pre-industrial development. In order to economize space and ease the presentation, the main analysis focuses on genetic diversity, although robustness to the diversity measure are included in various parts of the text.

### 4.1 Population Diversity and Economic Specialization (Ordinary Least Squares Analysis)

This subsection explores the statistical relationship between population diversity and economic specialization at the ethnicity level. It focuses on 116 ethnic groups for which both genetic and ethnographic data to construct the proposed measure of economic specialization is available. Figure 2 shows for these 116 ethnicities the distribution of population diversity for groups above and below the mean economic specialization. Clearly, more specialized groups also have higher population diversity.

In order to analyze this relation more systematically, the following baseline econometric specification is adopted and estimated by ordinary least squares (OLS):

$$Specialization_i = \alpha + \beta PD_i + G_i' \Gamma + X_i' \Delta + \epsilon_i \quad (1)$$

where  $Specialization_i$  is the measure of economic specialization of the ethnic group  $i$  introduced in section 3.1. The variable  $PD_i$  is a measure of population diversity as reflected by the expected heterozygosity of ethnic group  $i$ . The vector  $G_i'$  denotes a set of basic geographic controls whereas the vector  $X_i'$  includes a set of additional potential confounders that are

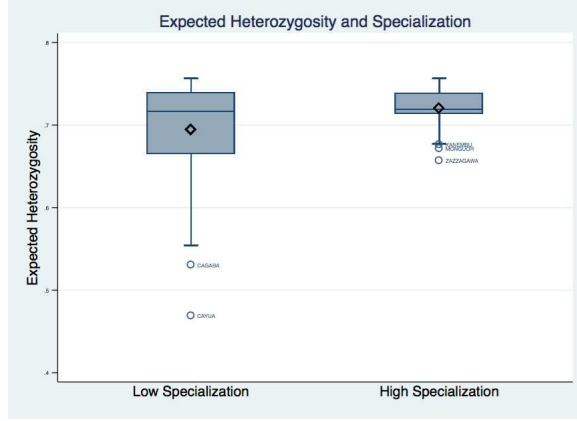


Figure 2: Population Diversity and Economic Specialization

discussed below in detail. Finally,  $\epsilon_i$  is an error term that is allowed to be heteroskedastic. The proposed hypothesis in this paper implies  $\beta > 0$ .<sup>8</sup>

Table 1 analyzes the association between economic specialization and population diversity accounting for a basic set of geographic characteristics of ethnicities' homelands using OLS. In particular, column 1 shows the unconditional relationship between population diversity and economic specialization. The estimated coefficient is statistically significant at the 1 percent level and is consistent with an economically significant effect of population diversity. In particular, a one standard deviation increase in population diversity is associated with a 0.27 standard deviation increase in economic specialization.

A potential concern is that population diversity might be capturing the effect of absolute latitude. In particular, technologies and institutions have historically spread more easily across similar latitudes, where climate and the duration of days were not drastically different. Furthermore, the positive high correlation between absolute latitude and development, which has been widely documented in the economic growth and development literature (Spolaore and Wacziarg, 2013), might confound the effect of population diversity. In order to address this potential concern column 2 accounts for the effect of absolute latitude. Reassuringly, although absolute latitude enters positively (albeit statistically insignificantly) in this specification, the effect of population diversity remains highly statistically significant and increases by 10 percent. This increase in the point estimate for  $\beta$  accurately reflects the fact that there is a strong negative relationship between absolute latitude and diversity (Michalopoulos, 2012).

Column 3 accounts for the total area of the ethnic homeland, since all else equal, larger

<sup>8</sup>In order to ease the interpretation of the results and compare them across the different specifications presented in this paper, all tables report standardized coefficients. The standard coefficients report the number of standard deviation changes in the dependent variable for a one-standard deviation change in the independent variable.

Table 1: Population Diversity and Economic Specialization

|                      | Economic Specialization |                   |                   |                   |                   |                   |                   |
|----------------------|-------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
|                      | (1)                     | (2)               | (3)               | (4)               | (5)               | (6)               | (7)               |
| Population Diversity | 0.27***<br>(0.05)       | 0.31***<br>(0.05) | 0.27***<br>(0.05) | 0.27***<br>(0.05) | 0.25***<br>(0.05) | 0.27***<br>(0.05) | 0.36***<br>(0.09) |
| Absolute Latitude    |                         | 0.15<br>(0.09)    |                   |                   |                   |                   | 0.80***<br>(0.30) |
| Area                 |                         |                   | 0.01<br>(0.04)    |                   |                   |                   | 0.01<br>(0.05)    |
| Elevation (Avg.)     |                         |                   |                   | -0.03<br>(0.11)   |                   |                   | 0.31*<br>(0.16)   |
| Precipitation (Avg.) |                         |                   |                   |                   | -0.08<br>(0.09)   |                   | 0.13<br>(0.16)    |
| Temperature (Avg.)   |                         |                   |                   |                   |                   | 0.04<br>(0.08)    | 0.73***<br>(0.25) |
| Adjusted- $R^2$      | 0.07                    | 0.08              | 0.06              | 0.06              | 0.06              | 0.06              | 0.15              |
| Observations         | 116                     | 116               | 116               | 116               | 116               | 116               | 116               |

Notes: This table establishes the positive statistically and economically significant correlation between economic specialization and population diversity as measured by expected heterozygosity after accounting for a set of basic geographical controls. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

areas may contain a more diverse populations by construction. In particular, cultural assimilation may be more difficult in large territories, thus, contributing to cultural diversity. Additionally, total area may confound the effect of market potential, which is a potential driver of economic specialization.<sup>9</sup> Reassuringly, the inclusion of this control does not affect the estimated effect of population diversity. Column 4 accounts for the effect of mean elevation, which has been shown to negatively correlate with ethnolinguistic heterogeneity at the country level (Michalopoulos, 2012). Reassuringly, the point estimate remains virtually unaltered.

Another potential concern is that population diversity correlates with precipitation and temperature. In particular, it has been shown that both species and cultural diversity are positively correlated with precipitation and net primary productivity, which in turn depends on temperature (Moore et al., 2002; Nettle, 1998). Furthermore, precipitation and temperature might directly affect economic activities and specialization. Thus, omission of precipitation and temperature might bias the results. Columns 5 and 6 address this potential concern by accounting for average precipitation and average temperature, respectively.

<sup>9</sup>It is worth noting that total area is determined by ethnic homeland borders, which can be arguably endogenous to both heterogeneity and economic specialization or trade.

As shown in the table, the estimated coefficients on both these controls are negative and not statistically nor economically significant. On the other hand, the effect of population diversity remains positive statistically and economically significant.

Finally, column 7 accounts for the joint effect of all these basic geographic controls. The statistical relationship between population diversity and economic specialization is statistically significant at the 1 percent level and implies an economically significant effect of population diversity. In particular, an increase of one standard deviation in population diversity increases economic specialization by more than one-third of its standard deviation.

While these results support the proposed hypothesis, the estimated effect of population diversity might be biased due to omitted variables. In order to address this potential concern and to account for other possible sources of economic specialization, Table 2 adds a further set of controls to the analysis. In order to compare with the previous results, column 1 includes all the controls in Table 1.

A potential concern is that higher genetic diversity may be a result of a hostile disease environment. For example, Birchenall (2014) argues that pathogen stress influenced pre-colonial ethnic diversity. Furthermore, a “bad” disease environment can also negatively affect economic activities. Thus, column 2 considers the potential confounding effect of the disease environment by accounting for the ecology of malaria (Kiszewski et al., 2004). As expected, malaria ecology negatively correlates with economic specialization. Given the positive correlation between the disease environment and population diversity, the inclusion of malaria ecology increases the size and statistical significance of the point estimate for population diversity.

Column 3 accounts for the diversity of the ecological environment, which could potentially affect specialization directly (Fenske, 2014) and be correlated with linguistic and cultural diversity (Michalopoulos, 2012; Moore et al., 2002). Reassuringly, although ecological diversity correlates strongly with economic specialization, the point estimate for population diversity is virtually unaltered.<sup>10</sup>

Columns 4 and 5 account for the potentially confounding effects of agricultural and caloric suitability. In particular, Michalopoulos (2012) shows that variation in soil quality correlates with inter-ethnic linguistic diversity, which could foster trade. Moreover, variation in soil quality could potentially be conducive to trade directly. On the other hand, Galor and Özak (2014, 2015) show that pre-industrial population (density) levels are highly correlated with their Caloric Suitability Index (CSI).<sup>11</sup> Since population (density) potentially affects

---

<sup>10</sup>A measure of ecological diversity is constructed following Fenske (2014) -a Herfindahl index constructed from the shares of each ethnic homeland’s area occupied by each ecological type (White, 1983).

<sup>11</sup>The Caloric Suitability Index (CSI) measures for each cell of 10 kms  $\times$  10 kms in the world, the average number of calories that could be potentially produced given the climatic conditions in that cell and the crops



Table 2: Population Diversity and Economic Specialization

|   | Economic Specialization |                    |                   |                   |                   |                    |                   |                   |                    |
|---|-------------------------|--------------------|-------------------|-------------------|-------------------|--------------------|-------------------|-------------------|--------------------|
|   | (1)                     | (2)                | (3)               | (4)               | (5)               | (6)                | (7)               | (8)               | (9)                |
| Population Diversity                    | 0.36***<br>(0.09)       | 0.41***<br>(0.10)  | 0.36***<br>(0.10) | 0.37***<br>(0.09) | 0.34***<br>(0.11) | 0.33***<br>(0.10)  | 0.31***<br>(0.09) | 0.40***<br>(0.11) | 0.31***<br>(0.10)  |
| Malaria Ecology                         |                         | -0.36***<br>(0.12) |                   |                   |                   |                    |                   |                   | -0.41***<br>(0.12) |
| Ecological Diversity                    |                         |                    | 0.26***<br>(0.10) |                   |                   |                    |                   |                   | 0.20*<br>(0.11)    |
| Agricultural<br>Suitability (avg.)      |                         |                    |                   | 0.00<br>(0.13)    |                   |                    |                   |                   | 0.13<br>(0.10)     |
| Agricultural<br>Suitability (std.)      |                         |                    |                   | 0.22*<br>(0.13)   |                   |                    |                   |                   | 0.32**<br>(0.14)   |
| Caloric Suitability<br>(Pre-1500 ,avg.) |                         |                    |                   |                   | -0.24*<br>(0.14)  |                    |                   |                   | -0.34**<br>(0.14)  |
| Caloric Suitability<br>(Pre-1500 ,std.) |                         |                    |                   |                   | 0.30**<br>(0.11)  |                    |                   |                   | 0.07<br>(0.14)     |
| Temperature (Spatial<br>Corr., Avg.)    |                         |                    |                   |                   |                   | 0.01<br>(0.09)     |                   |                   | -0.05<br>(0.08)    |
| Temperature<br>(Volatility, Avg)        |                         |                    |                   |                   |                   | -0.58***<br>(0.19) |                   |                   | -0.11<br>(0.20)    |
| Pct. Area within<br>100kms of Sea       |                         |                    |                   |                   |                   |                    | 0.02<br>(0.10)    |                   | -0.16<br>(0.11)    |
| Coast Length                            |                         |                    |                   |                   |                   |                    | 0.49**<br>(0.22)  |                   | 0.60***<br>(0.20)  |
| Ruggedness (Avg.)                       |                         |                    |                   |                   |                   |                    |                   | -0.22<br>(0.22)   | 0.07<br>(0.18)     |
| Pre-Industrial<br>Mobility (avg.)       |                         |                    |                   |                   |                   |                    |                   | 0.81*<br>(0.41)   | 1.06**<br>(0.46)   |
| Pre-Industrial<br>Mobility (std.)       |                         |                    |                   |                   |                   |                    |                   | -0.04<br>(0.12)   | -0.36**<br>(0.16)  |
| Main Controls                           | Yes                     | Yes                | Yes               | Yes               | Yes               | Yes                | Yes               | Yes               | Yes                |
| Altonji et al                           |                         | -8.86              | 303.72            | -242.74           | 13.92             | 10.19              | 6.36              | -11.69            | 6.34               |
| $\delta$                                |                         | 0.83               | 1.26              | 0.89              | 1.03              | 1.01               | 1.02              | 0.84              | 1.18               |
| $\beta$ -Oster                          |                         | 0.62               | 0.36              | 0.38              | 0.12              | 0.04               | 0.03              | 0.76              | 0.30               |
| $R^2$                                   | 0.20                    | 0.27               | 0.26              | 0.23              | 0.24              | 0.24               | 0.26              | 0.23              | 0.50               |
| Adjusted- $R^2$                         | 0.15                    | 0.22               | 0.21              | 0.17              | 0.18              | 0.18               | 0.20              | 0.17              | 0.40               |
| Observations                            | 116                     | 116                | 116               | 116               | 116               | 116                | 116               | 116               | 116                |

Notes: This table establishes the positive statistically and economically significant correlation between economic specialization and population diversity as measured by expected heterozygosity after accounting for the set of basic geographical controls of Table 1 and an extended set of confounders. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

market size and thus specialization, including the mean and the standard deviation of the CSI accounts for this potential confounding channel. Reassuringly, the qualitative results remain unaltered.

Column 6 controls for the confounding effects of both the spatial correlation and the intertemporal volatility of temperature. In particular, Dean et al. (1985) argue that trade alliances among communities were common in regions with high spatial variability in climate. In addition, pre-modern societies could have mitigated the negative impact of climatic variation by extending the set of subsistence activities. Additionally, Ahlerup and Olsson (2012) show that temperature variation predicts ethnic diversity. Accounting for these potential confounders does not alter the results.

Columns 7 and 8 account for a potential concern that ethnicities' isolation and access to the sea might jointly affect their genetic diversity and their economic specialization. In particular, proximity and access to the sea may ease contact with other societies, thus increasing genetic diversity and facilitating trade. Similarly, isolated ethnicities may be forced to specialize and also be less diverse. Reassuringly, accounting for the fraction of the ethnic homeland located within 100 kilometers from the sea as well as the length of the ethnic homeland's coastline (Column 7), and for the average ruggedness of the terrain, the average and the standard deviation of the pre-industrial mobility index developed by Özak (2010, 2012) does not alter the qualitative results.

Finally, column 9 accounts for the joint effect of all the previous confounders. The estimated effect of population diversity on economic specialization remains positive statistically and economically significant. In particular, a one standard deviation increase in population diversity increases, on average, economic specialization by one-third of its standard deviation.

The point estimates reported so far may still be biased due to unobservable factors that correlate with both population diversity and economic specialization. In order to assess the effects of this potential bias on the results, Table 2 reports statistics for selection on unobservables (Altonji et al., 2005; Bellows and Miguel, 2009; Oster, 2014). To construct these statistics the specification in column 1, which only controls for the basic geographic variables discussed in Table 1, is taken as the baseline. Both the Altonji et al's (Altonji et al., 2005; Bellows and Miguel, 2009) and  $\delta$  (Oster, 2014) statistics measure how strongly correlated any unobservables would have to be in order to account for the full size of the coefficient on population diversity. As can be seen, in all columns Altonji et al's statistic is larger (in absolute value) than 1, while the  $\delta$  statistic, which penalizes additionally for changes in the  $R^2$ , is larger than 1 once all the controls are included, suggesting that omitted variable bias is not driving the results. Moreover, the bias corrected  $\beta$ -Oster statistic is always positive, suggesting that *even* under omitted variable bias, the effect of population

diversity on economic specialization is positive and economically significant. In particular, the estimates of column 9 suggest that the true effect of population diversity belongs to the interval  $[0.30, 0.31]$ , i.e. that a one standard deviation increase in genetic diversity generates almost one-third of a standard deviation increase in economic specialization.

## 4.2 Population Diversity and Distance to Addis Ababa

This section establishes the negative statistically and economically significant causal effect of the migratory distance from East Africa on population diversity as measured by genetic and linguistic diversity. In particular, the “Out-of-Africa” theory of the geographic origin and early migration of anatomically modern humans posits that the process leading to the peopling of planet Earth by anatomically modern humans started with their migration out of East Africa more than sixty thousand years ago (Cann et al., 1987; Pemberton et al., 2013; Ramachandran et al., 2005). This process consisted of a series of discrete successive migrations, in which new settlements were established by smaller subgroups from an originally larger population. Since the population of a new settlement was not necessarily representative of the original population, the sampling process from subsequently smaller populations led to a loss of population diversity, i.e., the serial founder effect. Therefore, the Out-of-Africa theory predicts that population diversity decreases along the different migratory routes that humans followed out of East Africa.<sup>12</sup>

The analysis estimates the pre-industrial migratory distance to East Africa by finding the minimal travel times to Addis Ababa using the Human Mobility Index with Seafaring - HMISea (Özak, 2010, 2012). HMISea estimates the time (in weeks) required to walk across each square kilometer of land, accounting for the topographic, climatic, terrain conditions, and human biological abilities, as well as the time required to cross major seas with pre-industrial technologies. Figure 3 shows the potential migratory routes out of East Africa to the historical ethnic homelands that minimize the travel time according to HMISea. To overcome the potential concern of endogeneity of the actual historical patterns of migration, the analysis employs the HMISea travel time to the ethnic homeland as an instrument for population diversity.

Table 3 explores the relationship between migratory distance to Addis Ababa and genetic diversity (as measured by expected heterozygosity) for 144 ethnic groups for which geo-coded genetic and ethnographic data is available.<sup>13</sup> Two facts stand out from the results in Table 3: (i) migratory distance to Addis Ababa alone explains 72 percent of the variation in

---

<sup>12</sup>This prediction has been supported empirically using data from various population samples (Ashraf and Galor, 2013b; Atkinson, 2011; Cann et al., 1987; Pemberton et al., 2013; Ramachandran et al., 2005).

<sup>13</sup>Similar results are obtained in the full sample of 267 ethnicities for which genetic data alone is available.

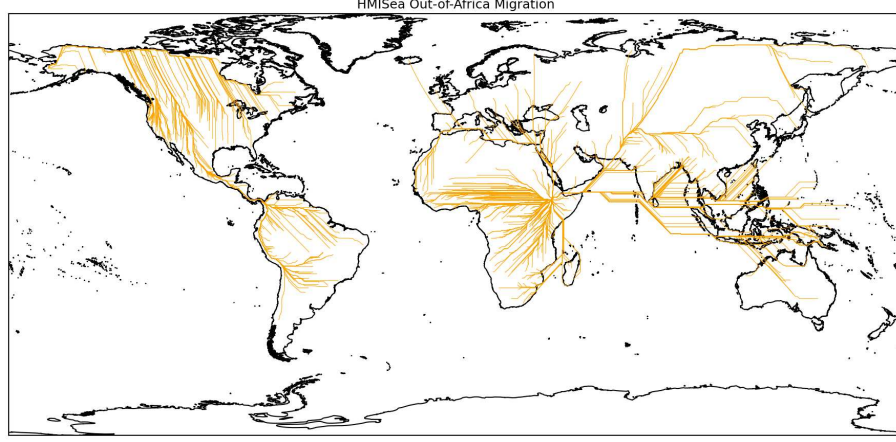


Figure 3: Potential Migratory Routes Out of East Africa

population diversity (column 1); and (ii) accounting for the potential confounding effects of all the controls included in Tables 1 and 2, both individually and jointly, affects remarkably little the point estimates for pre-industrial migratory distance to Addis Ababa. Furthermore, as shown in column 8, these results hold also for the restricted sample of 116 ethnic groups from previous section. Figures 4(a) and 4(b) depict respectively the unconditional and conditional strong negative relationship between population diversity and the pre-industrial migratory distance to Addis Ababa.

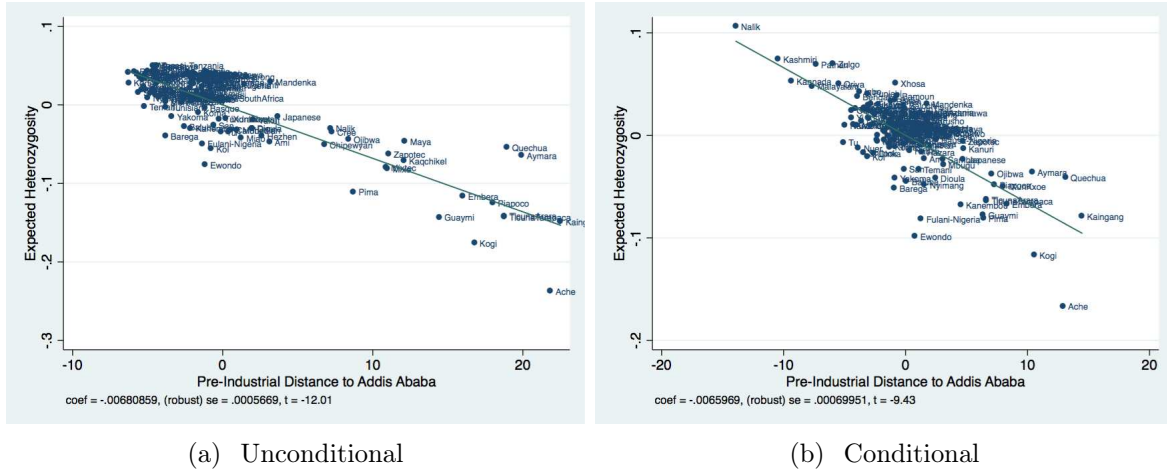


Figure 4: Pre-Industrial Distance to Addis-Ababa and Population Diversity

The importance of effect of the distance to Addis Ababa on genetic diversity is further confirmed by the semi-partial  $R^2$ .<sup>14</sup> In particular, the distance to Addis Ababa has the

<sup>14</sup>Results not shown, but can be obtained from authors.

Table 3: Population Diversity and Distance to Addis-Ababa

|   | Population Diversity (Expected Heterozygosity) |                    |                    |                    |                    |                    |                    |                    |                     |
|---|--|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|---------------------|
|   | Full Sample                                    |                    |                    |                    |                    |                    |                    |                    | Specia-<br>lization |
|   | (1)  | (2)                | (3)                | (4)                | (5)                | (6)                | (7)                | (8)                | (9)                 |
| Pre-Industrial Distance to Addis Ababa      | -0.85***<br>(0.07)                             | -0.80***<br>(0.10) | -0.80***<br>(0.09) | -0.81***<br>(0.09) | -0.80***<br>(0.10) | -0.79***<br>(0.10) | -0.80***<br>(0.10) | -0.82***<br>(0.09) | -0.85***<br>(0.09)  |
| Malaria Ecology                             |  |                    | 0.17***<br>(0.05)  |                    |                    |                    |                    | 0.20***<br>(0.07)  | 0.16**<br>(0.08)    |
| Agricultural Suitability (avg.)             |  |                    |                    | -0.00<br>(0.06)    |                    |                    |                    | -0.01<br>(0.07)    | -0.01<br>(0.07)     |
| Agricultural Suitability (std.)             |  |                    |                    | 0.08<br>(0.06)     |                    |                    |                    | 0.13**<br>(0.06)   | 0.14*<br>(0.07)     |
| Caloric Suitability Index (Pre-1500CE)      |  |                    |                    |                    | 0.02<br>(0.07)     |                    |                    | 0.07<br>(0.07)     | 0.09<br>(0.09)      |
| Caloric Suitability (Pre-1500 ,std.)        |  |                    |                    |                    | -0.08<br>(0.05)    |                    |                    | -0.13**<br>(0.06)  | -0.13*<br>(0.07)    |
| Pct. Area within 100kms of Sea Coast Length |  |                    |                    |                    |                    | -0.00<br>(0.06)    |                    | 0.13**<br>(0.06)   | 0.14**<br>(0.07)    |
|   |  |                    |                    |                    |                    | 0.03<br>(0.06)     |                    | 0.04<br>(0.05)     | 0.01<br>(0.07)      |
| Ruggedness (Avg.)                           |  |                    |                    |                    |                    |                    | -0.03<br>(0.12)    | -0.19<br>(0.13)    | -0.19<br>(0.16)     |
| Pre-Industrial Mobility (avg.)              |  |                    |                    |                    |                    |                    | 0.05<br>(0.20)     | 0.12<br>(0.23)     | 0.13<br>(0.25)      |
| Pre-Industrial Mobility (std.)              |  |                    |                    |                    |                    |                    | -0.13<br>(0.08)    | -0.07<br>(0.11)    | -0.07<br>(0.12)     |
| Main Controls                               | No   | Yes                | Yes                | Yes                | Yes                | Yes                | Yes                | Yes                | Yes                 |
| Adjusted- $R^2$                             | 0.72   | 0.72               | 0.74               | 0.72               | 0.72               | 0.72               | 0.73               | 0.75               | 0.73                |
| $R^2$                                       | 0.72   | 0.74               | 0.75               | 0.74               | 0.74               | 0.74               | 0.75               | 0.78               | 0.76                |
| Observations                                | 144  | 144                | 144                | 144                | 144                | 144                | 144                | 144                | 116                 |

Notes: This table establishes the negative statistically and economically significant relation between expected heterozygosity and the distance to Addis Ababa after accounting for the set of basic geographical controls of Table 1 and an extended set of confounders and measures of isolation. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

largest semi-partial  $R^2$  in the analysis, e.g. in column 8 it is 0.3, which is 15 times larger than the semi-partial of malaria ecology, which is the variable with the second largest value. This suggests that the variation that is uniquely related to the distance to Addis Ababa, explains 30% of the total variation in genetic diversity, while the variation that is specific to the each of other variables explains less than 2% of the total variation in genetic diversity.

The strong predictive power of the pre-industrial distance to Addis Ababa on genetic

Table 4: Linguistic Diversity and Distance to Addis-Ababa

|   | Linguistic Diversity |                    |                    |                    |                    |                    |                    |                    |                    |
|---|----------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
|   | Full Sample          |                    |                    |                    |                    |                    |                    |                    | Specialization     |
|   | (1)                  | (2)                | (3)                | (4)                | (5)                | (6)                | (7)                | (8)                | (9)                |
| Panel A: Linguistic Diversity (Consonant Inventory)     |                      |                    |                    |                    |                    |                    |                    |                    |                    |
| Pre-Industrial Distance to Addis Ababa                  | -0.42***<br>(0.05)   | -0.36***<br>(0.05) | -0.35***<br>(0.05) | -0.35***<br>(0.05) | -0.36***<br>(0.05) | -0.36***<br>(0.05) | -0.38***<br>(0.06) | -0.37***<br>(0.06) | -0.37***<br>(0.07) |
| Adjusted- $R^2$   | 0.17                 | 0.26               | 0.26               | 0.26               | 0.26               | 0.26               | 0.26               | 0.27               | 0.27               |
| $R^2$   | 0.17                 | 0.27               | 0.27               | 0.28               | 0.28               | 0.28               | 0.28               | 0.31               | 0.31               |
| Observations  | 299                  | 299                | 299                | 299                | 299                | 299                | 299                | 299                | 254                |
| Panel B: Linguistic Diversity (Vowel Quality Inventory) |                      |                    |                    |                    |                    |                    |                    |                    |                    |
| Pre-Industrial Distance to Addis Ababa                  | -0.29***<br>(0.06)   | -0.32***<br>(0.07) | -0.27***<br>(0.07) | -0.32***<br>(0.07) | -0.34***<br>(0.07) | -0.34***<br>(0.07) | -0.36***<br>(0.07) | -0.31***<br>(0.07) | -0.33***<br>(0.08) |
| Adjusted- $R^2$   | 0.08                 | 0.14               | 0.17               | 0.14               | 0.15               | 0.16               | 0.15               | 0.19               | 0.22               |
| $R^2$   | 0.09                 | 0.16               | 0.19               | 0.16               | 0.17               | 0.19               | 0.18               | 0.23               | 0.27               |
| Observations  | 301                  | 301                | 301                | 301                | 301                | 301                | 301                | 301                | 255                |
| Panel C: Linguistic Diversity (Number of Genders)       |                      |                    |                    |                    |                    |                    |                    |                    |                    |
| Pre-Industrial Distance to Addis Ababa                  | -0.25***<br>(0.06)   | -0.25***<br>(0.06) | -0.23***<br>(0.06) | -0.26***<br>(0.06) | -0.27***<br>(0.06) | -0.25***<br>(0.06) | -0.31***<br>(0.08) | -0.30***<br>(0.09) | -0.33***<br>(0.08) |
| Adjusted- $R^2$   | 0.06                 | 0.17               | 0.17               | 0.18               | 0.18               | 0.16               | 0.19               | 0.20               | 0.23               |
| $R^2$   | 0.06                 | 0.20               | 0.21               | 0.22               | 0.23               | 0.20               | 0.23               | 0.28               | 0.32               |
| Observations  | 152                  | 152                | 152                | 152                | 152                | 152                | 152                | 152                | 130                |
| Main Controls   | No                   | Yes                | Yes                | Yes                | Yes                | Yes                | Yes                | Yes                | Yes                |
| Additional Controls                                     | -                    | -                  | Malaria            | Agr. Suit.         | CSI                | Sea                | Mobility           | All                | All                |

Notes: This table establishes the negative statistically and economically significant relation between measures of linguistic diversity and the distance to Addis Ababa after accounting for the set of basic geographical controls of Table 1 and an extended set of confounders and measures of isolation. Each column includes the same set of controls as the same column in Table 3. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

diversity, and the stability of the estimated effect of distance to Addis Ababa to the inclusion of various potential confounders, suggests that this distance is a valid instrument for diversity, giving, in particular, credence to the validity of the exclusion restriction. Nonetheless, the analysis below provides additional checks on the validity of this instrument by accounting for the effect of other historical determinants of development.

Finally, the negative relation between the pre-industrial distance to Addis Ababa and population diversity is further confirmed in Table 4, which shows the relation between this distance and 3 measures of linguistic diversity. While the explanatory power of the pre-industrial distance to Addis Ababa is lower for linguistic diversity than for genetic diversity,

the results still support the validity of pre-industrial distance to Addis Ababa as an instrument for population diversity.

### 4.3 Population Diversity and Economic Specialization (Instrumental Variable Analysis)

This section establishes the positive causal effect of population diversity on economic specialization by exploiting an instrumental variable strategy based on the migratory distance to East Africa. As shown in the previous section, the migratory distance to East Africa is a valid instrument for population diversity, since (i) it is the main predictor of population diversity, due to the serial founder effect and the Out-of-Africa theory, and (ii) it only affects economic outcomes through its effect on diversity.<sup>15</sup>

Table 5 presents the results of this instrumental variables (IV) analysis, in which population diversity as proxied by genetic diversity is instrumented by the migratory distance to East Africa for the set of 116 ethnicities for which genetic, ethnographic and geographic data exists. In order to facilitate comparison with the OLS results, column 1 replicates the analysis of column 5 in Table 1 by accounting for the effect of the set of basic geographic controls. Columns 2 through 10 use this IV strategy to establish the positive causal effect of population diversity on economic specialization, accounting for the set of controls of Table 2. The estimated effect is 22-55% larger than in the OLS analysis, and ranges between 0.44 and 0.56, implying an economically significant effect of population diversity on economic specialization. In particular, after accounting for all the confounders analyzed in table 2, a one standard deviation increase in population diversity causes about half a standard deviation increase in economic specialization.

These results are not subject to a weak instrument problem, since the Kleibergen-Paap F-statistics for the first stage, reported at the bottom of the table, are all larger than the critical values suggested by Stock-Yogo. Additionally, the results are robust to the measure of economic specialization used (see section 3.1). In particular, Table A.9 shows that employing the alternative measures of economic specialization generates qualitatively identical results and imply a positive causal effect of population diversity on economic specialization.

Finally, Table 6 establishes the robustness of the results to the measure of population diversity by replicating the analysis of Table 5 using different proxies of population diversity. In particular, proxying population diversity with linguistic diversity as measured by consonant inventory, vowel quality inventory and the number of genders generates qualitatively similar results.

---

<sup>15</sup>Section 4.6 presents additional evidence in support of the exclusion restriction.

Table 5: Population Diversity and Economic Specialization (IV)

|   | Economic Specialization |                   |                    |                   |                   |                   |                    |                   |                   |                    |
|---|-------------------------|-------------------|--------------------|-------------------|-------------------|-------------------|--------------------|-------------------|-------------------|--------------------|
|   | OLS                     | IV                |                    |                   |                   |                   |                    |                   |                   |                    |
|   | (1)                     | (2)               | (3)                | (4)               | (5)               | (6)               | (7)                | (8)               | (9)               | (10)               |
| Population Diversity                      | 0.36***<br>(0.09)       | 0.51***<br>(0.12) | 0.53***<br>(0.12)  | 0.54***<br>(0.12) | 0.54***<br>(0.13) | 0.45***<br>(0.14) | 0.49***<br>(0.13)  | 0.44***<br>(0.11) | 0.56***<br>(0.14) | 0.46***<br>(0.14)  |
| Malaria Ecology                           |                         |                   | -0.38***<br>(0.11) |                   |                   |                   |                    |                   |                   | -0.44***<br>(0.11) |
| Ecological Diversity                      |                         |                   |                    | 0.26***<br>(0.10) |                   |                   |                    |                   |                   | 0.19*<br>(0.10)    |
| Agricultural<br>Suitability (avg.)        |                         |                   |                    |                   | -0.01<br>(0.14)   |                   |                    |                   |                   | 0.08<br>(0.10)     |
| Agricultural<br>Suitability (std.)        |                         |                   |                    |                   | 0.22*<br>(0.12)   |                   |                    |                   |                   | 0.31**<br>(0.12)   |
| Caloric Suitability<br>Index (Pre-1500CE) |                         |                   |                    |                   |                   | -0.21<br>(0.14)   |                    |                   |                   | -0.28**<br>(0.13)  |
| Caloric Suitability<br>(Pre-1500 ,std.)   |                         |                   |                    |                   |                   | 0.29***<br>(0.11) |                    |                   |                   | 0.09<br>(0.12)     |
| Temperature (Spatial<br>Corr., Avg.)      |                         |                   |                    |                   |                   |                   | 0.01<br>(0.10)     |                   |                   | -0.03<br>(0.07)    |
| Temperature<br>(Volatility, Avg)          |                         |                   |                    |                   |                   |                   | -0.53***<br>(0.19) |                   |                   | -0.03<br>(0.18)    |
| Pct. Area within<br>100kms of Sea         |                         |                   |                    |                   |                   |                   |                    | 0.01<br>(0.10)    |                   | -0.17<br>(0.11)    |
| Coast Length                              |                         |                   |                    |                   |                   |                   |                    | 0.46**<br>(0.21)  |                   | 0.57***<br>(0.17)  |
| Ruggedness (Avg.)                         |                         |                   |                    |                   |                   |                   |                    |                   | -0.28<br>(0.22)   | 0.02<br>(0.18)     |
| Pre-Industrial<br>Mobility (avg.)         |                         |                   |                    |                   |                   |                   |                    |                   | 0.91**<br>(0.40)  | 1.23***<br>(0.44)  |
| Pre-Industrial<br>Mobility (std.)         |                         |                   |                    |                   |                   |                   |                    |                   | -0.03<br>(0.12)   | -0.37**<br>(0.15)  |
| Main Controls                             | Yes                     | Yes               | Yes                | Yes               | Yes               | Yes               | Yes                | Yes               | Yes               | Yes                |
| First-stage F-statistic                   |                         | 56.99             | 59.31              | 59.04             | 65.63             | 52.61             | 55.27              | 53.29             | 63.44             | 81.54              |
| Adjusted- $R^2$                           | 0.15                    | 0.14              | 0.21               | 0.19              | 0.15              | 0.17              | 0.16               | 0.19              | 0.15              | 0.39               |
| Observations                              | 116                     | 116               | 116                | 116               | 116               | 116               | 116                | 116               | 116               | 116                |

Notes: This table establishes the positive statistically and economically significant causal effect of population diversity on economic specialization, by instrumenting population diversity with the distance to Addis Ababa (see section 4.2). These results are robust to accounting for the set of basic geographical controls of Table 1 and an extended set of confounders. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.



Table 6: Population Diversity and Economic Specialization (IV)  
Linguistic Diversity

|   | Economic Specialization |                   |                   |                   |                   |                   |                   |                   |                   |                   |
|---|-------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
|   | OLS                     | IV                |                   |                   |                   |                   |                   |                   |                   |                   |
|   | (1)                     | (2)               | (3)               | (4)               | (5)               | (6)               | (7)               | (8)               | (9)               | (10)              |
| Panel A: Linguistic Diversity (Consonant Inventory)     |                         |                   |                   |                   |                   |                   |                   |                   |                   |                   |
| Linguistic Diversity                                    | 0.20***<br>(0.06)       | 1.13***<br>(0.20) | 1.15***<br>(0.22) | 1.12***<br>(0.20) | 1.13***<br>(0.21) | 1.18***<br>(0.21) | 1.12***<br>(0.21) | 1.13***<br>(0.20) | 1.15***<br>(0.22) | 1.22***<br>(0.28) |
| First-stage F-statistic                                 |                         | 46.23             | 44.68             | 45.37             | 43.58             | 44.29             | 43.27             | 44.19             | 36.73             | 27.68             |
| Adjusted- $R^2$   | 0.06                    | -0.67             | -0.71             | -0.61             | -0.68             | -0.73             | -0.67             | -0.68             | -0.72             | -0.75             |
| Observations  | 254                     | 254               | 254               | 254               | 254               | 254               | 254               | 254               | 254               | 254               |
| Panel B: Linguistic Diversity (Vowel Quality Inventory) |                         |                   |                   |                   |                   |                   |                   |                   |                   |                   |
| Linguistic Diversity                                    | 0.39***<br>(0.06)       | 1.15***<br>(0.24) | 1.38***<br>(0.34) | 1.16***<br>(0.24) | 1.14***<br>(0.25) | 1.13***<br>(0.22) | 1.02***<br>(0.21) | 1.06***<br>(0.21) | 1.05***<br>(0.22) | 1.29***<br>(0.33) |
| First-stage F-statistic                                 |                         | 22.60             | 16.32             | 24.31             | 21.43             | 26.88             | 27.78             | 30.08             | 26.28             | 18.78             |
| Adjusted- $R^2$   | 0.17                    | -0.39             | -0.69             | -0.39             | -0.37             | -0.35             | -0.19             | -0.24             | -0.24             | -0.52             |
| Observations  | 255                     | 255               | 255               | 255               | 255               | 255               | 255               | 255               | 255               | 255               |
| Panel C: Linguistic Diversity (Number of Genders)       |                         |                   |                   |                   |                   |                   |                   |                   |                   |                   |
| Linguistic Diversity                                    | 0.13<br>(0.08)          | 1.19***<br>(0.28) | 1.30***<br>(0.34) | 1.16***<br>(0.28) | 1.18***<br>(0.29) | 1.21***<br>(0.29) | 1.09***<br>(0.26) | 1.16***<br>(0.28) | 0.91***<br>(0.22) | 0.87***<br>(0.31) |
| First-stage F-statistic                                 |                         | 29.97             | 24.50             | 29.02             | 29.16             | 28.00             | 30.19             | 28.64             | 33.65             | 17.21             |
| Adjusted- $R^2$   | -0.01                   | -1.03             | -1.25             | -0.96             | -0.96             | -1.07             | -0.85             | -1.00             | -0.52             | -0.43             |
| Observations  | 130                     | 130               | 130               | 130               | 130               | 130               | 130               | 130               | 130               | 130               |
| Main Controls   | No                      | Yes               | Yes               | Yes               | Yes               | Yes               | Yes               | Yes               | Yes               |                   |
| Additional Controls                                     | -                       | -                 | Malaria           | Agr. Suit.        | CSI               | Sea               | Mobility          | All               | All               |                   |

Notes: This table establishes the positive statistically and economically significant causal effect of population diversity on economic specialization, by instrumenting population diversity with the distance to Addis Ababa (see section 4.2). These results are robust to accounting for the set of basic geographical controls of Table 1 and an extended set of confounders. Each column includes the same set of controls as the same column in Table 5. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

## 4.4 Predicted Population Diversity and Economic Specialization

This section provides additional support for the positive causal effect of population diversity on economic specialization. In particular, a potential concern with the previous analysis is that it is based on a sample of ethnicities, for which both genetic and specialization data

is available, which could be a source of potential bias. In order to address this potential concern, this section follows Ashraf and Galor (2013b) and uses a measure of population diversity as predicted by the pre-industrial migratory distance to Addis Ababa. In particular, based on the estimated relation between the migratory distance to Addis Ababa and population diversity in the subsample of ethnicities analyzed in section 4.2, the analysis predicts population diversity for all ethnicities in the Ethnographic Atlas. This strategy expands the sample of ethnicities for which diversity and specialization data is available to 934. Moreover, it allows the analysis to be performed on additional ethnographic data on trade. Finally, as in the case of the previous IV approach, the estimated effect of predicted population diversity can be given a causal interpretation, since by construction it captures only the exogenous variation in diversity generated by the serial founder effect and the Out-of-Africa theory.

The baseline regression specification in this section is given by

$$Specialization_i = \alpha + \beta \widehat{PD}_i + G'_i \Gamma + X'_i \Delta + \epsilon_i \quad (2)$$

where the only difference with respect to equation (1) is the inclusion of  $\widehat{PD}_i$ , which is the predicted population diversity implied by the relation between migratory distance to Addis Ababa and population diversity accounting for all additional controls. Since this analysis exploits a generated regressor, standard errors are computed following the bootstrapping procedure discussed in Ashraf and Galor (2013b).<sup>16</sup>

Based on this extended sample, the analysis replicates in columns 1 to 10 of Table 7 the main econometric specifications of Tables 1, 2, and 5. Reassuringly, the positive causal effect of population diversity on economic specialization remains statistically and economically significant. Furthermore, the point estimates are remarkably stable across specifications, supporting the view that the effect of predicted population diversity is not biased by omitted factors. Moreover, the size of the estimated effect of population diversity on economic specialization in this expanded sample lies between the OLS and IV estimates of the reduced sample (see Tables 2 and 5).

Column 11 establishes that the positive causal effect of population diversity on economic specialization is robust to accounting for unobserved time-invariant continent-specific attributes. Indeed, if anything, the inclusion of continental fixed effects increases the estimated effect of diversity. In particular, the estimates in columns 10 and 11 imply that a standard deviation increase in predicted population diversity increases economic specializa-

---

<sup>16</sup>In particular, a random sample of 144 ethnicities with both genetic and migratory distance data is drawn with replacement out of the original sample. Then the specification of column 8 of Table 3 of section 4.2 is re-estimated. Using these new estimates population diversity is predicted again and equation (2) is re-estimated. This procedure is repeated 1001 times and the distribution of the bootstrapped coefficients is used to compute the standard errors.

Table 7: Predicted Population Diversity and Economic Specialization

|                                      | Economic Specialization |                   |                    |                   |                   |                   |                    |                   |                   |                    |                    |
|--------------------------------------|-------------------------|-------------------|--------------------|-------------------|-------------------|-------------------|--------------------|-------------------|-------------------|--------------------|--------------------|
|                                      | (1)                     | (2)               | (3)                | (4)               | (5)               | (6)               | (7)                | (8)               | (9)               | (10)               | (11)               |
| Predicted Population Diversity       | 0.44***<br>(0.02)       | 0.42***<br>(0.03) | 0.49***<br>(0.04)  | 0.42***<br>(0.03) | 0.41***<br>(0.03) | 0.46***<br>(0.04) | 0.40***<br>(0.03)  | 0.42***<br>(0.05) | 0.42***<br>(0.04) | 0.53***<br>(0.07)  | 0.60***<br>(0.21)  |
| Malaria Ecology                      |                         |                   | -0.18***<br>(0.05) |                   |                   |                   |                    |                   |                   | -0.18***<br>(0.07) | -0.14**<br>(0.07)  |
| Ecological Diversity                 |                         |                   |                    | 0.15***<br>(0.04) |                   |                   |                    |                   |                   | 0.10***<br>(0.04)  | 0.09***<br>(0.05)  |
| Agricultural Suitability (avg.)      |                         |                   |                    |                   | -0.05<br>(0.05)   |                   |                    |                   |                   | -0.14***<br>(0.07) | -0.17***<br>(0.08) |
| Agricultural Suitability (std.)      |                         |                   |                    |                   | 0.09**<br>(0.05)  |                   |                    |                   |                   | 0.03<br>(0.06)     | -0.02<br>(0.07)    |
| Caloric Suitability (Pre-1500 ,avg.) |                         |                   |                    |                   |                   | 0.08**<br>(0.05)  |                    |                   |                   | 0.09***<br>(0.06)  | 0.09***<br>(0.07)  |
| Caloric Suitability (Pre-1500 ,std.) |                         |                   |                    |                   |                   | 0.12***<br>(0.05) |                    |                   |                   | 0.12***<br>(0.06)  | 0.12***<br>(0.07)  |
| Temperature (Spatial Corr., Avg.)    |                         |                   |                    |                   |                   |                   | -0.01<br>(0.05)    |                   |                   | -0.04<br>(0.06)    | -0.08<br>(0.08)    |
| Temperature (Volatility, Avg)        |                         |                   |                    |                   |                   |                   | -0.20***<br>(0.10) |                   |                   | 0.01<br>(0.15)     | 0.04<br>(0.14)     |
| Pct. Area within 100kms of Sea       |                         |                   |                    |                   |                   |                   |                    | 0.00<br>(0.05)    |                   | -0.14***<br>(0.07) | -0.11**<br>(0.07)  |
| Coast Length                         |                         |                   |                    |                   |                   |                   |                    | 0.00<br>(0.13)    |                   | 0.01<br>(0.12)     | -0.02<br>(0.12)    |
| Ruggedness (Avg.)                    |                         |                   |                    |                   |                   |                   |                    |                   | 0.14**<br>(0.11)  | 0.16**<br>(0.16)   | 0.18**<br>(0.18)   |
| Pre-Industrial Mobility (avg.)       |                         |                   |                    |                   |                   |                   |                    |                   | 0.20**<br>(0.16)  | 0.46***<br>(0.19)  | 0.20*<br>(0.21)    |
| Pre-Industrial Mobility (std.)       |                         |                   |                    |                   |                   |                   |                    |                   | -0.05<br>(0.09)   | -0.22***<br>(0.12) | -0.13*<br>(0.14)   |
| Main Controls                        | No                      | Yes               | Yes                | Yes               | Yes               | Yes               | Yes                | Yes               | Yes               | Yes                | Yes                |
| Continental FE                       | No                      | No                | No                 | No                | No                | No                | No                 | No                | No                | No                 | Yes                |
| Adjusted- $R^2$                      | 0.19                    | 0.22              | 0.23               | 0.24              | 0.22              | 0.24              | 0.22               | 0.21              | 0.23              | 0.29               | 0.32               |
| Observations                         | 934                     | 934               | 934                | 934               | 934               | 934               | 934                | 934               | 934               | 934                | 934                |

Notes: This table establishes the positive statistically and economically significant causal effect of population diversity as predicted by the distance to Addis Ababa (see section 4.2) on economic specialization. These results are robust to accounting for the set of basic geographical controls of Table 1 and an extended set of confounders. Standardized coefficients. Bootstrap standard error estimates in parenthesis; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

tion by more than a half of a standard deviation.<sup>17</sup>

<sup>17</sup>Table A.12 shows the point estimates of the reduced form economic specialization-distance to Addis Ababa for all the specifications in Table 7. The point estimates for pre-industrial distance to Addis Ababa are remarkably stable and strongly statistically significant.

## 4.5 Heterogenous Effects of Population Diversity on Specialization

This section explores whether, as suggested by the theory, diverse populations enjoy complementarities with diverse geographical and ecological endowments. In particular, the effect of population diversity on economic specialization might be higher in locations with diverse geography, given that diverse preferences or abilities could potentially allow diverse endowments and ecologies to be exploited better and, thus, generate higher levels of economic specialization.

Table 8: Heterogeneous Effects of Predicted Population Diversity on Economic Specialization

|   | Economic Specialization |                   |                   |                   |                   |                   |
|---|-------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
|   | (1)                     | (2)               | (3)               | (4)               | (5)               | (6)               |
| Predicted Population Diversity                                      | 0.42***<br>(0.03)       | 0.33***<br>(0.04) | 0.34***<br>(0.04) | 0.13**<br>(0.08)  | 0.28***<br>(0.04) | 0.28***<br>(0.05) |
| Predicted Population Diversity<br>× Ecological Diversity            |                         | 1.07***<br>(0.43) |                   |                   |                   |                   |
| Predicted Population Diversity<br>× Agricultural Suitability (std.) |                         |                   | 1.91***<br>(0.66) |                   |                   |                   |
| Predicted Population Diversity<br>× Temperature (Volatility, avg.)  |                         |                   |                   | 2.24***<br>(0.74) |                   |                   |
| Predicted Population Diversity<br>× Ruggedness (Avg.)               |                         |                   |                   |                   | 2.77***<br>(0.63) |                   |
| Predicted Population Diversity<br>× Pre-Industrial Mobility (std.)  |                         |                   |                   |                   |                   | 1.54***<br>(0.60) |
| Main Controls   | Yes                     | Yes               | Yes               | Yes               | Yes               | Yes               |
| Adjusted- $R^2$   | 0.21                    | 0.23              | 0.23              | 0.24              | 0.26              | 0.23              |
| Observations  | 934                     | 934               | 934               | 934               | 934               | 934               |

Notes: This table establishes the positive statistically and economically significant causal effect of population diversity as predicted by the distance to Addis Ababa (see section 4.2) on economic specialization. Additionally, it establishes the heterogeneity of the effect and the complementarity between population diversity and variations in environmental and geographical factors. These results are robust to accounting for the set of basic geographical controls of Table 1 and an extended set of confounders. Standardized coefficients. Bootstrap standard error estimates in parenthesis; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

Table 8 analyzes the potential heterogenous effects of population diversity on economic specialization. In particular, it shows the main effect of population diversity and its interaction with ecological diversity, the standard deviation of agricultural suitability, temperature volatility, the standard deviation of ruggedness of the terrain, and the standard deviation of pre-industrial mobility.<sup>18</sup> As can be seen there, all main effects and interactions are positive

<sup>18</sup>The estimated coefficients are again reported as standardized betas, which simplifies the comparison

and highly statistically and economically significant. The estimates imply that the more diverse a population and the more diverse the geography in which it lives, the higher the level of economic specialization.

As suggested by the theory diverse populations enjoy complementarities with positive effects on economic specialization of living in diverse geographical areas. This result provides a link between the seemingly contradictory theories based on the composition of the population (Ashraf and Galor, 2013a,b) and those based on geographical factors (Galor and Özak, 2014, 2015). In particular, it provides an explanation as to why economies with similar populations or environments might have different economic outcomes.

## 4.6 Population Diversity, Economic Specialization, and Historical Confounders

This section establishes that the positive causal effect of population diversity on economic specialization is robust to accounting for other historical sources of pre-industrial development. Thus, overcoming the potential concern that population diversity is capturing the effect of factors like the transition to agriculture or the history of settlement or the existence of centralized institutions on economic specialization. Moreover, it overcomes the potential concern that the established causal effect of population diversity is capturing its effect on pre-industrial development, with the latter potentially being the actual source of economic specialization.

Table 9 analyzes the robustness of the results to accounting for additional potential historical sources of pre-industrial development and economic specialization, which were generated by migratory or diffusion processes. Thus, it additionally explores the validity of the exclusion restriction. Column 1 replicates the results of column 7 in Table 7 and serves as a baseline point of comparison. Column 2 includes an indicator of the duration of human settlements since prehistoric times, “origtime”, which estimates the date since the first uninterrupted settlement by anatomically modern humans (Ahlerup and Olsson, 2012). Clearly, this measure should be highly correlated with migratory distance to Addis Ababa and population diversity, since the closer a location is to Addis Ababa, the earlier it could have been populated by anatomically modern humans. Thus, the omission of origtime may bias the estimated effect of population diversity documented above, if a longer history of uninterrupted settlement facilitated the division of labor via, for example, a greater chance

---

of the main effects across tables. Of course, this makes the interpretation of the interactions difficult, but given that both main effects and interactions are positive, the qualitative nature of the effects is directly observable from the table.

Table 9: Predicted Population Diversity, Economic Specialization  
and Other Historical Confounders

|                                  | Economic Specialization |                   |                   |                   |                    |                    |                    |
|----------------------------------|-------------------------|-------------------|-------------------|-------------------|--------------------|--------------------|--------------------|
|                                  | (1)                     | (2)               | (3)               | (4)               | (5)                | (6)                | (7)                |
| Predicted Population Diversity   | 0.42***<br>(0.03)       | 0.44***<br>(0.06) | 0.37***<br>(0.02) | 0.43***<br>(0.03) | 0.48***<br>(0.03)  | 0.48***<br>(0.03)  | 0.48***<br>(0.03)  |
| Origtime                         |                         | -0.02<br>(0.07)   |                   |                   |                    |                    |                    |
| Years Since Neolithic Revolution |                         |                   | 0.30***<br>(0.04) |                   |                    |                    |                    |
| Distance Neolithic Frontier      |                         |                   |                   | -0.02<br>(0.03)   |                    |                    |                    |
| Distance Frontier (1CE)          |                         |                   |                   |                   | -0.21***<br>(0.03) |                    |                    |
| Distance Frontier (1000CE)       |                         |                   |                   |                   |                    | -0.21***<br>(0.03) |                    |
| Distance Frontier (1500CE)       |                         |                   |                   |                   |                    |                    | -0.22***<br>(0.03) |
| Main Controls                    | Yes                     | Yes               | Yes               | Yes               | Yes                | Yes                | Yes                |
| Adjusted- $R^2$                  | 0.22                    | 0.22              | 0.30              | 0.22              | 0.25               | 0.25               | 0.25               |
| Observations                     | 936                     | 927               | 926               | 934               | 934                | 934                | 934                |

Notes: This table establishes the positive statistically and economically significant causal effect of population diversity as predicted by the distance to Addis Ababa (see section 4.2) on economic specialization after accounting for other potential historical sources of specialization and development. Standardized coefficients. Bootstrap standard error estimates in parenthesis; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

for the emergence of social stratification or a dominant elite.<sup>19</sup> Reassuringly, the results in column 2 reveal that the inclusion of origtime has a negligible impact on the estimated effect of predicted population diversity. The effect of population diversity on economic specialization remains positive, strongly statistically and economically significant: a standard deviation increase in the proposed measure of population diversity explains one fourth of the standard deviation of economic specialization. This suggests the previous results were not picking up the potential effect of a longer settlement duration on the division of labor in pre-modern societies.

Columns 3 and 4 analyze the potential confounding effect of the long-lasting influence of the Neolithic Revolution by accounting for the timing of the Neolithic Revolution (Putterman, 2008) and the pre-industrial distance to the closest Neolithic frontier. As argued by Diamond (1997), an earlier transition from hunting and gathering practices to agriculture

<sup>19</sup>In fact, Ahlerup and Olsson (2012) show that the historical duration of human settlements is a strong predictor of ethnolinguistic fractionalization.

provided an initial advantage to some societies, which later translated into a persistent technological superiority. Moreover, it has been suggested that an earlier transition to agriculture allowed the creation of an economic surplus and the emergence of economic specialization (Boix, 2015). Additionally, country-level precolonial development has been positively associated with the time since the Neolithic Revolution (Ashraf and Galor, 2011). In line with these findings, column 3 shows that the time elapsed since the Neolithic Revolution positively affects economic specialization. On the other hand, the pre-industrial distance to the closest Neolithic frontier does not have an effect on economic specialization. Reassuringly, the estimated effect of predicted population diversity on economic specialization remains positive strongly statistically and economically significant suggesting that the omission of the Neolithic transition-timing was not spuriously driving the main results.<sup>20</sup>

Additionally, columns 5-7 analyze the potential confounding effect of the distance from the closest technological frontier in the years 1, 1000 and 1500CE. In particular, if technology diffuses from a technological frontier, one can expect ethnicities close to the frontier to acquire more technologies and develop economically, all of which might potentially be conducive to economic specialization. Indeed, the estimated effect of the distance from the frontier on economic specialization is negative statistically and economically significant, in line with this prediction. Still, the estimated positive causal effect of predicted population diversity on economic specialization remains statistically and economically significant.

Finally, Table 10 explores whether the existence of centralized institutions affects the causal effect of population diversity on economic specialization. In particular, the table replicates the analysis for the sample of ethnicities with and without a centralized state. As can be seen in Table 10 the positive causal effect of population diversity on economic specialization is not mediated by the existence of a state. Moreover, the fact that population diversity generates economic specialization in the sample of ethnicities that do not have a state, implies that the existence of a state is not a necessary precondition for the emergence of economic specialization and trade.

## 4.7 Predicted Population Diversity and Other Measures of Trade

This section analyzes the empirical relationship between predicted population diversity and a broader set of pre-industrial trade-related measures from the Standard Cross-Cultural Sample (SCCS). In particular, it establishes the positive effect of population diversity on the importance of trade for subsistence, the existence of inter-community trade as a food source, the existence of money as a medium of exchange, the existence of credit specialists,

---

<sup>20</sup>Alternatively, accounting for the degree of subsistence dependence on agriculture, as measured in the Ethnographic Atlas (v5), does not alter the results either.

Table 10: Predicted Population Diversity and Economic Specialization  
Ethnicities with and without Centralized States

|                                | Economic Specialization |                   |                   |                       |                   |                   |
|--------------------------------|-------------------------|-------------------|-------------------|-----------------------|-------------------|-------------------|
|                                | No Centralized State    |                   |                   | Any Centralized State |                   |                   |
|                                | (1)                     | (2)               | (3)               | (4)                   | (5)               | (6)               |
| Predicted Population Diversity | 0.46***<br>(0.05)       | 0.39***<br>(0.05) | 0.50***<br>(0.09) | 0.30***<br>(0.03)     | 0.36***<br>(0.04) | 0.40***<br>(0.06) |
| Main Controls                  | No                      | Yes               | Yes               | No                    | Yes               | Yes               |
| Additonal Controls             | No                      | No                | Yes               | No                    | No                | Yes               |
| Adjusted- $R^2$                | 0.21                    | 0.24              | 0.30              | 0.09                  | 0.14              | 0.24              |
| Observations                   | 433                     | 433               | 433               | 479                   | 479               | 479               |

Notes: This table establishes that the positive statistically and economically significant causal effect of population diversity as predicted by the distance to Addis Ababa (see section 4.2) on economic specialization is not mediated by the existence of a (pre-industrial) State. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

and the existence of writing and records. Reassuringly, these measures of trade correlate strongly and positively with the measure of economic specialization (see Table A.3). Table 11 presents the point estimates for the regression specifications given by:

$$Y_i = \alpha + \beta \widehat{PD}_i + G_i' \Gamma + X_i' \Delta + \epsilon_i, \quad (3)$$

where the only difference with respect to equation (2) is the dependent variable  $Y_i$  which represents different measures of pre-industrial trade-related variables or the sociocultural complexity of the ethnic group. Sample size varies between 153 and 168 ethnic groups depending on the availability of the outcome variable. Column 1 confirms that the previous results on the positive effect of predicted population diversity on economic specialization remains statistically and economically significant when using the SCCS sample.

Column 2 in Table 11 establishes the positive statistically and economically significant effect of population diversity on the importance of trade for subsistence -measured as percent importance in contribution to subsistence- (Barry, 1982). The estimated effect implies that one standard deviation increase in predicted population diversity increases the importance of trade by one-fourth of a standard deviation.

Column 3 provides additional evidence for the positive effect of population diversity on trade as captured by the extent the local food supply depends on trade between communities of an ethnicity. This trade measure ranges from 1 (no trade) to 7 (food imports contributes



to more than 50 percent of food supply). The results suggest that ethnic groups with a higher level of population diversity tend to trade more among its communities.

Table 11: Predicted Population Diversity, Economic Specialization and Trade in the SCCS

|                                | Pre-Industrial Measures of Trade |                          |  |                   |                   |                           |
|--------------------------------|----------------------------------|--------------------------|--|-------------------|-------------------|---------------------------|
|                                | Economic<br>Special-<br>ization  | Impor-<br>tance<br>Trade | Intercom-<br>of munity<br>Trade<br>as Food<br>Source | Money             | Credit            | Writing<br>and<br>Records |
|                                | (1)                              | (2)                      | (3)  | (4)               | (5)               | (6)                       |
| Predicted Population Diversity | 0.37***<br>(0.06)                | 0.25***<br>(0.07)        | 0.42***<br>(0.08)                                    | 0.44***<br>(0.08) | 0.19***<br>(0.07) | 0.19***<br>(0.06)         |
| Main Controls                  | Yes                              | Yes                      | Yes  | Yes               | Yes               | Yes                       |
| Additonal Controls             | Yes                              | Yes                      | Yes  | Yes               | Yes               | Yes                       |
| Adjusted- $R^2$                | 0.32                             | 0.04                     | 0.19   | 0.23              | 0.19              | 0.26                      |
| Observations                   | 168                              | 168                      | 165  | 165               | 153               | 168                       |

Notes: This table establishes the positive statistically and economically significant causal effect of population diversity as predicted by the distance to Addis Ababa (see section 4.2) on various measures of trade and trade-related institutions and technologies. These results are robust to accounting for the set of basic geographical controls of Table 1 and an extended set of confounders. Standardized coefficients. Bootstrap standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

Columns 4-6 analyze the effect of population diversity on trade related institutions and technologies. In particular, it establishes the positive effect of population diversity on the existence of money, the existence of credit institutions, and the existence of writing and records. The estimated effect is economically significant and implies that a one standard deviation increase in predicted population diversity increases the likelihood of the existence of (a) money by 0.4 standard deviations (column 4), (b) credit institutions by 0.2 standard deviations (column 5), and (c) the existence of writing and records by 0.2 standard deviations (column 6).

These results support the proposed theory and imply a positive effect of population diversity on economic specialization, trade, and trade-related institutions and technologies.

## 4.8 Predicted Population Diversity, Economic Specialization and Pre-industrial Economic Development

This section analyzes the effects of economic specialization on pre-industrial economic development. In particular, Table 12 explores the potentially beneficial effects of economic

Table 12: Predicted Population Diversity, Specialization and Pre-Industrial Development

|                                   | Pre-Industrial Development             |                   |                       |  |                    |                              |
|-----------------------------------|--|-------------------|-----------------------|--|--------------------|------------------------------|
|                                   | Technologi-<br>cal Special-<br>ization | Complexity        | Population<br>Density | Mean Size of<br>Local Com-<br>munities | Statehood<br>Level | Class<br>Stratifi-<br>cation |
|                                   | (1)                                    | (2)               | (3)                   | (4)                                    | (5)                | (6)                          |
| Panel A: Effect of Specialization |  |                   |                       |  |                    |                              |
| Economic Specialization           | 0.45***<br>(0.08)                      | 0.52***<br>(0.07) | 0.39***<br>(0.07)     | 0.38***<br>(0.04)                      | 0.43***<br>(0.03)  | 0.22***<br>(0.03)            |
| Adjusted- $R^2$                   | 0.50                                   | 0.57              | 0.51                  | 0.46                                   | 0.48               | 0.32                         |
| Panel B: Mediation (OLS)          |  |                   |                       |  |                    |                              |
| Economic Specialization           | 0.45***<br>(0.08)                      | 0.52***<br>(0.07) | 0.40***<br>(0.07)     | 0.40***<br>(0.04)                      | 0.43***<br>(0.03)  | 0.21***<br>(0.03)            |
| Predicted Population Diversity    | -0.16<br>(0.35)                        | -0.32<br>(0.25)   | -0.33<br>(0.33)       | -0.30*<br>(0.18)                       | 0.15<br>(0.12)     | 0.38**<br>(0.16)             |
| Adjusted- $R^2$                   | 0.49                                   | 0.57              | 0.51                  | 0.46                                   | 0.48               | 0.32                         |
| Panel C: Mediation (IV)           |  |                   |                       |  |                    |                              |
| Economic Specialization           | 0.38***<br>(0.09)                      | 0.42***<br>(0.09) | 0.34***<br>(0.08)     | 0.39***<br>(0.04)                      | 0.42***<br>(0.04)  | 0.24***<br>(0.03)            |
| Predicted Population Diversity    | -0.14<br>(0.32)                        | -0.30<br>(0.23)   | -0.32<br>(0.30)       | -0.29*<br>(0.17)                       | 0.15<br>(0.12)     | 0.36**<br>(0.15)             |
| Breusch-Pagan F-stat              | 22.63                                  | 22.63             | 20.48                 | 32.61                                  | 51.23              | 48.84                        |
| Breusch-Pagan p-value             | 0.00                                   | 0.00              | 0.00                  | 0.00                                   | 0.00               | 0.00                         |
| First-stage F-statistic           | 26.21                                  | 26.21             | 27.43                 | 47.88                                  | 54.85              | 51.89                        |
| Hansen's J-statistic              | 34.80                                  | 25.19             | 23.47                 | 32.34                                  | 28.17              | 34.83                        |
| J-stat p-value                    | 0.04                                   | 0.29              | 0.38                  | 0.07                                   | 0.17               | 0.04                         |
| Adjusted- $R^2$                   | 0.49                                   | 0.57              | 0.51                  | 0.46                                   | 0.48               | 0.32                         |
| Main Controls                     | Yes                                    | Yes               | Yes                   | Yes                                    | Yes                | Yes                          |
| Additional Controls               | Yes                                    | Yes               | Yes                   | Yes                                    | Yes                | Yes                          |
| Continental FE                    | Yes                                    | Yes               | Yes                   | Yes                                    | Yes                | Yes                          |
| Observations                      | 168                                    | 168               | 166                   | 509                                    | 912                | 879                          |

Notes: Notes: This table establishes the positive effect of economic specialization on pre-industrial development (Panel A). Panels B and C establish that economic specialization mediates the effect of population diversity on pre-industrial development. Panel C exploits an instrumental variable approach to establish the causal effect of economic specialization. These results account for the set of basic geographical controls of Table 1, an extended set of confounders and continental fixed effects. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

specialization on technological specialization (column 1), socio-economic complexity (column 2), population density (column 3), mean size of local communities (column 4), levels of statehood (column 5), and the existence of class stratification (column 6).

Panel A shows the positive association between economic specialization and these measures of pre-industrial development. In particular, the estimated coefficients imply that a one

standard deviation increase in economic specialization is associated with about 0.4 standard deviations increase in pre-industrial development.

Panel B suggests that, with the exception of class stratification, any potential effect of population diversity on pre-industrial development is mediated by its effect on economic specialization. In particular, predicted population diversity is not statistically significantly associated with pre-industrial development with the exception of its association with class stratification.

Although it is reassuring that economic specialization has a positive association with these measures of pre-industrial development, clearly, these associations cannot be given a causal interpretation due to endogeneity concerns. Particularly, the potential reverse causality from pre-industrial development to economic specialization is a major concern. Moreover, it is difficult to find an instrument based on theoretical arguments, which affects economic specialization *without* having a potential direct effect on pre-industrial development.

In order to overcome this issue, this research employs atheoretical instrumental variables based on the heteroskedastic structure of the residuals of the regression of economic specialization on all the additional control variables (Lewbel, 2012). In particular, consider the regression of a variable  $Y_1$  on an endogenous variable  $Y_2$  and a set of exogenous variables  $X$ . Lewbel (2012) establishes that if there exists a set  $Z \subseteq X$  of exogenous variables such that  $Z$  has at least two elements, then the set of variables  $(Z - E(Z))e_2$ , where  $E(Z)$  are the expected values of  $Z$  and  $e_2$  is the residual of the regression of  $Y_2$  on  $X$ , are valid instruments for  $Y_2$  in the regression of  $Y_1$  on  $Y_2$  and  $X$ , as long as  $e_2$  is not homoskedastic. Setting  $X = Z$  to be the set of all controls in the analysis (including continental fixed effects),<sup>21</sup> Panel C of Table 12 establishes that economic specialization has a positive economically and statistically significant effect on pre-industrial development. Moreover, the Breusch-Pagan test suggests the presence of heteroskedasticity, ensuring the conditions for identification are satisfied. Furthermore, Hansen’s J test for overidentification restrictions suggests that the instruments are valid. Interestingly, the estimated causal effect is quite similar across measures of pre-industrial development and imply that a one standard deviation increase in economic specialization increases pre-industrial development by 0.4 standard deviations.

---

<sup>21</sup>The analysis excludes the measures of isolation and mobility, since they tend to violate the exclusion restriction and invalidate some of the analysis.

## 5 Persistent Effects of Pre-Industrial Economic Specialization on Economic Development

This section explores whether historical levels of economic specialization have an effect on contemporary development. In particular, as established in the previous sections, pre-modern economic specialization is positively associated with trade and trade facilitating institutions, and had a positive effect on the emergence of pre-modern states, pre-industrial development, economic complexity and technology. Thus, if these institutions or technologies persist across time, it is conceivable that pre-modern economic specialization might have a persistent effect on economic development. On the other hand, pre-modern economic specialization may have fostered the emergence of certain cultural traits or the accumulation of a diverse set of production-specific knowledge due to learning by doing processes, all of which might still affect contemporary development.

Table 13 explores the potential persistent effect of pre-modern economic specialization on contemporary ethnic-level development. In particular, it establishes the positive statistically and economically significant association between pre-modern levels of economic specialization and contemporary development as measured by the intensity of night-time lights (Henderson et al., 2012; Michalopoulos and Papaioannou, 2013). Columns 1-3 show that after accounting for the effect of geography and continental fixed effects, ethnic groups with higher levels of pre-modern economic specialization have higher (log)-light intensity per area of their homeland, and thus higher levels of contemporary economic development. Additionally, column 4 shows that accounting for the effects of the transition to agriculture and the history of settlement does not alter the positive statistically and economically significant association between pre-modern economic specialization and contemporary economic development.

Columns 5 and 6 exclude the New World from the analysis, since light intensity of ethnic homelands in the Americas might be capturing the effects of population replacement and migration after 1500CE. Reassuringly, the estimated positive effect of economic specialization on contemporary economic development is even larger. Thus, the Old World sample suggests that a one-standard deviation increase in pre-modern economic development generates 0.2 standard deviations increase in log-light intensity.

Clearly, the positive correlation between pre-modern economic specialization and economic development cannot be given a fully causal interpretation, since the analysis may be subject to omitted variables bias. In order to delve further into the potential positive and persistent effect of specialization on development, Table 14 accounts for the full set of geographical controls, the transition to agriculture, the history of settlement and for regional/sub-continental fixed effects. Columns 1-2 and 5-6 establish that even after account-

Table 13: Pre-industrial Economic Specialization and Contemporary Economic Development

|                         | Log(Average Light Density + 0.01) |                   |                   |                   |                   |                   |
|-------------------------|-----------------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
|                         | Whole World                       |                   |                   |                   | Old World         |                   |
|                         | (1)                               | (2)               | (3)               | (4)               | (5)               | (6)               |
| Economic Specialization | 0.07**<br>(0.03)                  | 0.19***<br>(0.03) | 0.16***<br>(0.03) | 0.12***<br>(0.03) | 0.20***<br>(0.04) | 0.16***<br>(0.04) |
| Main Controls           | Yes                               | No                | Yes               | Yes               | Yes               | Yes               |
| Continental FE          | No                                | Yes               | Yes               | Yes               | Yes               | Yes               |
| Neolithic + Origtime    | No                                | No                | No                | Yes               | No                | Yes               |
| Adjusted- $R^2$         | 0.26                              | 0.35              | 0.40              | 0.41              | 0.31              | 0.33              |
| Observations            | 900                               | 900               | 900               | 900               | 565               | 565               |

Notes: This table establishes the persistent positive statistically and economically significant effect of pre-modern economic specialization on economic development. These results account for the main set of geographical controls in Table 7, continental fixed effects, the timing of transition to the Neolithic and the number of years of continuous settlement. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

ing for this expanded set of confounders, pre-modern economic specialization has a positive statistically and economically significant effect on contemporary development. Additionally, columns 3-4 and 7-8 use the heteroskedastic structure of the residuals in the regression of economic specialization on all the controls to generate instruments to identify the causal effect of pre-modern economic specialization on development (Lewbel, 2012).<sup>22</sup> Instrumenting economic specialization increases its estimated effect on development. Unfortunately, and although the first-stage  $F$ -statistic shows that the instruments are strong, Hansen’s over-identification test rejects the hypothesis that the instruments satisfy the exclusion restriction. Thus, the estimated effect might still be biased. Although these results cannot fully determine its causal nature, they do suggest that pre-modern economic specialization has a positive effect on contemporary development.

In order to further analyze potential channels through which pre-modern economic specialization might affect contemporary development, the analysis explores the effect of pre-modern economic specialization on contemporary occupational heterogeneity, i.e. the number of distinct economic occupations performed by members of an ethnicity as reported in the Afro-barometer. Columns 1-2 in Table 15 establish that pre-modern economic specialization has a positive statistically and economically significant effect on contemporary

<sup>22</sup>See section 4.8 for a presentation of the idea behind this instrumental variable approach.

Table 14: Pre-industrial Economic Specialization and Contemporary Economic Development

|                         | Log(Average Light Intensity + 0.01) |                  |                   |                   |                  |                  |                   |                  |
|-------------------------|-------------------------------------|------------------|-------------------|-------------------|------------------|------------------|-------------------|------------------|
|                         | Whole World                         |                  |                   |                   | Old World        |                  |                   |                  |
|                         | OLS                                 |                  | IV                |                   | OLS              |                  | IV                |                  |
|                         | (1)                                 | (2)              | (3)               | (4)               | (5)              | (6)              | (7)               | (8)              |
| Economic Specialization | 0.06**<br>(0.03)                    | 0.06**<br>(0.03) | 0.10***<br>(0.03) | 0.10***<br>(0.03) | 0.09**<br>(0.03) | 0.09**<br>(0.03) | 0.12***<br>(0.05) | 0.10**<br>(0.04) |
| Regional FE             | Yes                                 | Yes              | Yes               | Yes               | Yes              | Yes              | Yes               | Yes              |
| Main Controls           | Yes                                 | Yes              | Yes               | Yes               | Yes              | Yes              | Yes               | Yes              |
| All Additional Controls | Yes                                 | Yes              | Yes               | Yes               | Yes              | Yes              | Yes               | Yes              |
| Neolithic + Origtime    | No                                  | Yes              | No                | Yes               | No               | Yes              | No                | Yes              |
| First-stage F-statistic |                                     |                  | 75.85             | 76.73             |                  |                  | 31.06             | 30.31            |
| Hansen's J-statistic    |                                     |                  | 51.12             | 53.88             |                  |                  | 44.83             | 49.14            |
| J-stat p-value          |                                     |                  | 0.04              | 0.04              |                  |                  | 0.07              | 0.04             |
| Adjusted- $R^2$         | 0.25                                | 0.25             | 0.23              | 0.23              | 0.23             | 0.23             | 0.20              | 0.20             |
| Observations            | 898                                 | 898              | 898               | 898               | 563              | 563              | 563               | 563              |

Notes: This table establishes the persistent positive statistically and economically significant effect of pre-modern economic specialization on economic development. These results account for the main set of geographical controls in Table 7, regional fixed effects, the timing of transition to the Neolithic and number of years of continuous settlement. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

occupational heterogeneity after accounting for the full set of geographical controls, the transition to agriculture, the history of settlement and for regional/sub-continental fixed effects. Columns 3-4 follow Lewbel (2012) and instrument pre-modern economic specialization without affecting the qualitative results. Moreover, the F-statistic for the first stage suggests that the instruments are strongly correlated with specialization, while Hansen's over-identification tests suggests that the instruments are valid. Columns 5-8 replicate the analysis, but weigh each ethnicity according to the number of individuals of the ethnicity surveyed in the Afro-barometer. Reassuringly, the results are qualitatively similar. In particular, the estimates suggest that a one-standard deviation increase in pre-modern economic specialization increases contemporary occupational heterogeneity by about 0.2 standard deviations. Given the positive correlation between contemporary occupational heterogeneity and economic development, this result suggests a novel channel through which pre-modern economic specialization affects comparative development.

One potential mechanism that could explain the persistence of occupational heterogeneity

Table 15: Pre-colonial Economic Specialization, State Centralization  
and Contemporary Occupational Heterogeneity

|                         | Contemporary Occupational Heterogeneity |                  |                   |                  |                  |                  |                  |                  |
|-------------------------|---|------------------|-------------------|------------------|------------------|------------------|------------------|------------------|
|                         | Unweighted                              |                  |                   |                  | Weighted         |                  |                  |                  |
|                         | OLS                                     |                  | IV                |                  | OLS              |                  | IV               |                  |
|                         | (1)                                     | (2)              | (3)               | (4)              | (5)              | (6)              | (7)              | (8)              |
| Economic Specialization | 0.21***<br>(0.08)                       | 0.17**<br>(0.08) | 0.26***<br>(0.07) | 0.17**<br>(0.07) | 0.20**<br>(0.09) | 0.19**<br>(0.09) | 0.22**<br>(0.09) | 0.19**<br>(0.08) |
| Main Controls           | Yes                                     | Yes              | Yes               | Yes              | Yes              | Yes              | Yes              | Yes              |
| All Additional Controls | Yes                                     | Yes              | Yes               | Yes              | Yes              | Yes              | Yes              | Yes              |
| Regional FE             | No                                      | Yes              | No                | Yes              | No               | Yes              | No               | Yes              |
| First-stage F-statistic |   |                  | 31.03             | 364.59           |                  |                  | 31.87            | 15210.71         |
| Hansen's J-statistic    |   |                  | 17.81             | 20.70            |                  |                  | 24.36            | 26.65            |
| J-stat p-value          |   |                  | 0.40              | 0.48             |                  |                  | 0.11             | 0.18             |
| Adjusted- $R^2$         | 0.12                                    | 0.13             | 0.12              | 0.13             | 0.16             | 0.22             | 0.16             | 0.22             |
| Observations            | 102                                     | 102              | 102               | 102              | 102              | 102              | 102              | 102              |

Notes: This table establishes the persistent positive statistically and economically significant effect of pre-modern economic specialization on contemporary occupational heterogeneity. These results account for the full set of geographical controls in Table 7 and continental fixed effects. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

is learning by doing. In particular, societies that had higher levels of pre-modern economic specialization might have accumulated a more diverse set of production-specific human capital. This would allow them to produce a larger set of goods and thus have a more complex economic system. In particular, if the production of one type of good requires experience in the production of a related good, production processes will generate spillovers across sectors and products (Hausmann and Hidalgo, 2011; Hausmann et al., 2014; Hidalgo et al., 2007). Thus, societies with higher levels of pre-modern economic specialization would potentially have higher levels of contemporary economic complexity and produce a more diverse set of products.

Table 16 explores this prediction using country-level data. In particular, for each country the analysis constructs a pre-modern economic specialization measure, based on the population weighted average of pre-modern economic specialization across ethnicities located in the country. The table shows that pre-modern economic specialization has an economically and statistically significant association with the Economic Complexity Index (Hausmann et al., 2014), which measures the level of productive diversification in the country. In particular, a high value of the Economic Complexity Index suggests that a country produces complex

Table 16: Pre-modern Economic Specialization and Contemporary Economic Complexity

|                                    | Economic Complexity Index (2010) |                   |                   |                   |                   |                  |                  |                  |
|------------------------------------|----------------------------------|-------------------|-------------------|-------------------|-------------------|------------------|------------------|------------------|
|                                    | Main Measure                     |                   |                   |                   |                   |                  | Share            | Cardinal         |
|                                    | (1)                              | (2)               | (3)               | (4)               | (5)               | (6)              | (7)              | (8)              |
| Pre-modern Economic Specialization | 0.17**<br>(0.08)                 | 0.20***<br>(0.07) | 0.21***<br>(0.07) | 0.20***<br>(0.07) | 0.19***<br>(0.07) | 0.16**<br>(0.07) | 0.17**<br>(0.07) | 0.14**<br>(0.06) |
| Continental FE                     | Yes                              | Yes               | Yes               | Yes               | Yes               | Yes              | Yes              | Yes              |
| Geographical Controls              | No                               | Yes               | Yes               | Yes               | Yes               | Yes              | Yes              | Yes              |
| Caloric Suitability Controls       | No                               | No                | Yes               | Yes               | Yes               | Yes              | Yes              | Yes              |
| Geographical Diversity Controls    | No                               | No                | No                | Yes               | Yes               | Yes              | Yes              | Yes              |
| Disease Controls                   | No                               | No                | No                | No                | Yes               | Yes              | Yes              | Yes              |
| Years Since Neolithic Transition   | No                               | No                | No                | No                | No                | Yes              | Yes              | Yes              |
| Adjusted- $R^2$                    | 0.49                             | 0.59              | 0.61              | 0.64              | 0.66              | 0.67             | 0.68             | 0.67             |
| Observations                       | 95                               | 95                | 95                | 95                | 95                | 95               | 95               | 95               |

Notes: This table established the positive statistically and economically significant association between levels of pre-modern economic specialization and contemporary economic complexity at the country-level. Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

goods that few other countries produce. Similarly, Table C.1 in the Appendix shows that countries with high levels of pre-modern economic specialization tend to export a larger number of goods, tend to export more goods than they import, and have a larger share of global GDP.

## 6 Concluding Remarks

This research is the first attempt to identify the deep-rooted historical factors behind economic specialization and the emergence of trade, as well as their effect on comparative economic development. Moreover, it is the first to identify the positive causal effect of population diversity on economic specialization and the emergence of trade. In particular, by exploiting the exogenous decrease in population diversity due to the statistical sampling process generated by the serial founder effect, the analysis implements an instrumental variable approach to establish the positive statistically and economically significant causal effect of population diversity on pre-modern economic specialization. The analysis introduces a novel dataset combining geocoded ethnographic, linguistic and genetic data at the ethnicity level and a novel trade measure, which is based on the degree of economic specialization among eleven different pre-industrial economic activities.



Additionally, the analysis finds suggestive evidence of a persistent effect of pre-modern economic specialization on contemporary economic development. In particular, it establishes that the light intensity of an ethnic homeland increases with its exposure to higher levels of pre-modern economic specialization. Moreover, the analysis establishes the positive effect of pre-modern economic specialization on contemporary occupational heterogeneity. Suggesting that past economic specialization predicts contemporary levels of economic specialization, and with them contemporary economic development. Additionally, the research shows that countries with higher levels of pre-modern economic specialization tend to have more complex and diversified economic structures. This is a novel channel through which past economic development might still have an effect on comparative development. Further exploration of this channel and its potential effects is still an open issue.

## References

- Acemoglu, D., Johnson, S. and Robinson, J. (2005). Institutions as a fundamental cause of long-run growth, *Handbook of economic growth* **1**: 385–472.
- Ahlerup, P. and Olsson, O. (2012). The roots of ethnic diversity, *Journal of Economic Growth* **17**(2): 71–102.
- Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S. and Wacziarg, R. T. (2003). Fractionalization, *Journal of Economic Growth* **8**(2): 155–194.
- Alesina, A., Giuliano, P. and Nunn, N. (2013). On the origins of gender roles: Women and the plough, *The Quarterly Journal of Economics* **128**(2): 469–530.
- Alesina, A., Harnoss, J. and Rapoport, H. (2013). Birthplace diversity and economic prosperity, *National Bureau of Economic Research* .
- Altonji, J. G., Elder, T. E. and Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools, *Journal of Political Economy* **113**(1): 151–184.
- Arbatli, E., Ashraf, Q. and Galor, O. (2013). The nature of civil conflict, *Working Paper, Brown University, Department of Economics* .
- Ashraf, Q. and Galor, O. (2011). Dynamics and stagnation in the malthusian epoch, *The American Economic Review* **101**(5): 2003–2041.

- Ashraf, Q. and Galor, O. (2013a). Genetic diversity and the origins of cultural fragmentation, *The American Economic Review* **103**(3): 528–533.
- Ashraf, Q. and Galor, O. (2013b). The out of africa hypothesis, human genetic diversity, and comparative economic development, *The American Economic Review* **103**(1): 1–46.
- Atkinson, Q. D. (2011). Phonemic diversity supports a serial founder effect model of language expansion from africa, *Science* **332**(6027): 346–349.
- Barry, H. (1982). Cross-cultural codes on contributions by women to subsistence, *Ethnology: An international journal of cultural and social anthropology* **21**(2): 165–188.
- Bates, R. H. (1983). *Essays on the political economy of rural Africa*, Vol. no. 38 of *African studies series*, Cambridge University Press, Cambridge.
- Bellows, J. and Miguel, E. (2009). War and local collective action in sierra leone, *Journal of Public Economics* **93**(11): 1144–1157.
- Berg, M. (1991). *Markets and manufacture in early industrial Europe*, Routledge.
- Birchenall, J. A. (2014). Disease and diversity in africa’s long-term economic development, *Technical report*, University of California Santa Barbara.
- Boix, C. (2015). *Political order and inequality: their foundations and their consequences for human welfare*, Cambridge studies in comparative politics, Cambridge University Press.
- Braudel, F. (1972). *The Mediterranean and the mediterranean world in the age of Philip II*, Harper & Row, New York.
- Cann, R. L., Stoneking, M. and Wilson, A. C. (1987). Mitochondrial dna and human evolution, *Nature* **325**(6099): 31–36.
- Claessen, H. J. M. and Skalník, P. (1978). *The Early state*, New Babylon, studies in the social sciences ; 32, Mouton, The Hague.
- Conley, T. (1999). Gmm estimation with cross sectional dependence, *Journal of Econometrics* **92**(1): 1–45.
- De Long, J. B. and Shleifer, A. (1993). Princes and merchants: European city growth before the industrial revolution, *Journal of Law and Economics* **36**.

- Dean, J. S., Euler, R. C., Gumerman, G. J., Plog, F., Hevly, R. H. and Karlstrom, T. N. (1985). Human behavior, demography, and paleoenvironment on the colorado plateaus, *American Antiquity* pp. 537–554.
- Desmet, K., Ortuño-Ortín, I. and Wacziarg, R. (2012). The political economy of linguistic cleavages, *Journal of development Economics* **97**(2): 322–338.
- Desmet, K., Ortuño-Ortín, I. and Wacziarg, R. (2015). Culture, ethnicity and diversity, *Technical report*, National Bureau of Economic Research.
- Diamond, J. M. (1997). *Guns, germs, and steel: the fates of human societies*, 1st ed edn, W.W. Norton & Co., New York.
- Dryer, Matthew S & Haspelmath, M. e. (2013). *The World Atlas of of Language Structures Online.*, Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Easterly, W. and Levine, R. (1997). Africa’s growth tragedy: policies and ethnic divisions, *The Quarterly Journal of Economics* pp. 1203–1250.
- Fenske, J. (2014). Ecology, trade, and states in pre-colonial africa, *Journal of the European Economic Association* **12**(3): 612–640.
- Findlay, R. and O’Rourke, K. H. (2007). *Power and plenty: trade, war, and the world economy in the second millennium*, Princeton University Press, Princeton, N.J.
- Freeman, D. and Pankhurst, A. (2003). *Peripheral people: the excluded minorities of Ethiopia*, Red Sea Press, Lawrenceville, NJ.
- Galor, O. and Özak, Ö. (2014). The agricultural origins of time preference, *NBER Working Paper Series* .
- Galor, O. and Özak, Ö. (2015). Land productivity and economic development: Caloric suitability vs. agricultural suitability, *SMU Working Paper Series* .
- Gennaioli, N. and Rainer, I. (2007). The modern impact of precolonial centralization in africa, *Journal of Economic Growth* **12**(3): 185–234.
- Gluckman, M. (1941). *Economy of the central Barotse plain*, Vol. no. 7, Rhodes-Livingstone Institute, Livingstone, Northern Rhodesia.
- Greif, A. (1993). Contract enforceability and economic institutions in early trade: The maghribi traders’ coalition, *The American Economic Review* **83**(3): 525–548.

- Guiso, L., Sapienza, P. and Zingales, L. (2009). Cultural biases in economic exchange?, *Quarterly Journal of Economics* **124**(3): 1095–1131.
- Hallpike, C. R. (1968). The status of craftsmen among the konso of south-west ethiopia, *Africa* **38**(03): 258–269.
- Hallpike, C. R. (1972). *The Konso of Ethiopia: a study of the values of a Cushitic people*, Clarendon Press, Oxford.
- Hallpike, C. R. (2008). *The Konso of Ethiopia: a study of the values of a Cushitic people*, rev. ed edn, AuthorHouse, Central Milton Keynes.
- Hamilton, M. J. and Buchanan, B. (2009). The accumulation of stochastic copying errors causes drift in culturally transmitted technologies: quantifying clovis evolutionary dynamics, *Journal of Anthropological Archaeology* **28**(1): 55–69.
- Haspelmath, M. and Bibiko, H.-J. (2005). *The world atlas of language structures*, Oxford linguistics, Oxford University Press, New York.
- Hausmann, R. and Hidalgo, C. A. (2011). The network structure of economic output, *Journal of Economic Growth* **16**(4): 309–342.
- Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M., Simoes, A. and Yildirim, M. A. (2014). *The atlas of economic complexity: Mapping paths to prosperity*, Mit Press.
- Henderson, J. V., Storeygard, A. and Weil, D. N. (2012). Measuring economic growth from outer space, *The American Economic Review* **102**(2): 994–1028.
- Hidalgo, C. A., Klinger, B., Barabási, A.-L. and Hausmann, R. (2007). The product space conditions the development of nations, *Science* **317**(5837): 482–487.
- Hill, K. and Hurtado, A. M. (1996). *Aché life history: the ecology and demography of a foraging people*, Aldine de Gruyter, New York.
- Hinton, P. (1979). The karen, millennialism, and the politics of accommodation to lowland states, in C. F. Keyes (ed.), *Ethnic Adaptation and Identity: The Karen on the Thai Frontier with Burma*, Institute for the Study of Human Issues, pp. 81–98.
- Jha, S. (2013). Trade, institutions, and ethnic tolerance: Evidence from south asia, *American Political Science Review* **107**(04): 806–832.
- Kimura, M. (1983). *The neutral theory of molecular evolution*, Cambridge University Press, Cambridge.

- Kiszewski, A., Mellinger, A., Spielman, A., Malaney, P., Sachs, S. E. and Sachs, J. (2004). A global index representing the stability of malaria transmission, *The American Journal of Tropical Medicine and Hygiene* **70**(5): 486–498.
- Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models, *Journal of Business & Economic Statistics* **30**(1): 67–80.
- Lycett, S. J. (2008). Acheulean variation and selection: does handaxe symmetry fit neutral expectations?, *Journal of Archaeological Science* **35**(9): 2640–2648.
- Lycett, S. J. and von Cramon-Taubadel, N. (2008). Acheulean variability and hominin dispersals: a model-bound approach, *Journal of Archaeological Science* **35**(3): 553–562.
- Manica, A., Amos, W., Balloux, F. and Hanihara, T. (2007). The effect of ancient population bottlenecks on human phenotypic variation, *Nature* **448**(7151): 346–348.
- Michalopoulos, S. (2012). The origins of ethnolinguistic diversity, *The American Economic Review* **102**(4): 1508.
- Michalopoulos, S., Naghavi, A. and Prarolo, G. (2012). Trade and geography in the origins and spread of islam, *Technical report*, National Bureau of Economic Research.
- Michalopoulos, S. and Papaioannou, E. (2013). Pre-colonial ethnic institutions and contemporary african development, *Econometrica* **81**(1): 113–152.
- Moore, J. L., Manne, L., Brooks, T., Burgess, N. D., Davies, R., Rahbek, C., Williams, P. and Balmford, A. (2002). The distribution of cultural and biological diversity in africa, *Proceedings of the Royal Society of London B: Biological Sciences* **269**(1501): 1645–1653.
- Murdock, G. P. (1967). Ethnographic atlas: a summary, *Ethnology* pp. 109–236.
- Murdock, G. P. and White, D. R. (1969). Standard cross-cultural sample, *Ethnology* pp. 329–369.
- Nettle, D. (1998). Explaining global patterns of language diversity, *Journal of anthropological archaeology* **17**(4): 354–374.
- Nolan, P. and Lenski, G. E. (1999). *Human societies: an introduction to macrosociology*, 8th edn, McGraw Hill College, New York.
- Oster, E. (2014). Unobservable selection and coefficient stability: Theory and validation.
- Özak, Ö. (2010). The voyage of homo-economicus: Some economic measures of distance.

- Özak, Ö. (2012). Distance to the technological frontier and economic development, *Southern Methodist University Working Paper Series* .
- Pemberton, T. J., DeGiorgio, M. and Rosenberg, N. A. (2013). Population structure in a comprehensive genomic data set on human microsatellite variation, *G3: Genes/ Genomes/ Genetics* **3**(5): 891–907.
- Perrin, M. (1996). *Guajiro*, The Gale Group, Inc.
- Polanyi, K. (1957). *The great transformation*, Beacon Press, Boston.
- Putterman, L. (2008). Agriculture, diffusion and development: ripple effects of the neolithic revolution, *Economica* **75**(300): 729–748.
- Putterman, L. and 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.
- Ramachandran, S., Deshpande, O., Roseman, C. C., Rosenberg, N. A., Feldman, M. W. and Cavalli-Sforza, L. L. (2005). Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in africa, *Proceedings of the National Academy of Sciences of the United States of America* **102**(44): 15942–15947.
- Sanders, W. T. (1965). *The cultural ecology of the Teotihuacán Valley: a preliminary report of the results of the Teotihuacán Valley Project*, Dept. of Sociology & Anthropology, Pennsylvania State University.
- Service, E. R. (1978). Classical and modern theories on the origins of government, in R. Cohen and E. R. Service (eds), *Origins of the State: The Anthropology of Political Evolution*, ISHI, Philadelphia, PA, pp. 21–34.
- Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*, White-stone, Dublin.
- Spolaore, E. and Wacziarg, R. (2009a). The diffusion of development, *Quarterly Journal of Economics* **124**(2): 469–529.
- Spolaore, E. and Wacziarg, R. (2009b). War and relatedness, *Technical report*, National Bureau of Economic Research.
- Spolaore, E. and Wacziarg, R. (2013). How deep are the roots of economic development?, *Journal of Economic Literature* **51**(2): 325–69.

White, F. (1983). *The vegetation of Africa, a descriptive memoir to accompany the UNESCO/AETFAT/UNSO vegetation map of Africa (3 Plates, Northwestern Africa, Northeastern Africa, and Southern Africa, 1: 5,000,000)*, Unesco, Paris.

# Appendix

## A Trade: Additional Results and Supporting Material

### A.1 Summary Statistics

Table A.1: Summary Statistics on Base Sample

|  | Mean    | Std       | Min    | Max     | N   |
|--|---------|-----------|--------|---------|-----|
| Economic Specialization                | 1.34    | (1.41)    | 0.00   | 7.00    | 116 |
| Economic Specialization (Share)        | 0.20    | (0.19)    | 0.00   | 0.80    | 116 |
| Economic Specialization (Dev)          | 9.02    | (3.96)    | 2.00   | 25.00   | 116 |
| Population Diversity                   | 0.70    | (0.05)    | 0.47   | 0.76    | 116 |
| Absolute Latitude                      | 15.95   | (15.22)   | 0.04   | 68.67   | 116 |
| Area                                   | 0.18    | (0.85)    | 0.00   | 8.97    | 116 |
| Elevation (Avg.)                       | 823.71  | (727.51)  | 27.79  | 3581.35 | 116 |
| Precipitation (Avg.)                   | 91.00   | (57.54)   | 11.77  | 334.73  | 116 |
| Temperature (Avg.)                     | 20.69   | (8.43)    | -13.44 | 28.27   | 116 |
| Malaria Ecology                        | 7.88    | (9.07)    | 0.00   | 29.36   | 116 |
| Ecological Diversity                   | 0.26    | (0.22)    | 0.00   | 0.67    | 116 |
| Agricultural Suitability (avg.)        | 0.76    | (0.33)    | 0.00   | 1.00    | 116 |
| Agricultural Suitability (std.)        | 0.08    | (0.11)    | 0.00   | 0.45    | 116 |
| Caloric Suitability Index (Pre-1500CE) | 2699.11 | (1040.20) | 0.00   | 5030.97 | 116 |
| Caloric Suitability (Pre-1500 ,std.)   | 418.27  | (360.47)  | 0.00   | 1520.41 | 116 |
| Temperature (Spatial Corr., Avg.)      | 0.93    | (0.17)    | 0.00   | 1.00    | 116 |
| Temperature (Volatility, Avg)          | 0.84    | (0.48)    | 0.27   | 2.87    | 116 |
| Pct. Area within 100 kms of Sea        | 0.19    | (0.33)    | 0.00   | 1.00    | 116 |
| Coast Length                           | 0.49    | (2.16)    | 0.00   | 19.65   | 116 |
| Ruggedness (Avg.)                      | 110.62  | (149.48)  | 1.27   | 1076.01 | 116 |
| Pre-Industrial Mobility (avg.)         | 0.27    | (0.06)    | 0.07   | 0.37    | 116 |
| Pre-Industrial Mobility (std.)         | 0.05    | (0.04)    | 0.01   | 0.25    | 116 |



Table A.2: Summary Statistics on Full Sample

|  | Mean    | Std       | Min    | Max     | N   |
|--|---------|-----------|--------|---------|-----|
| Economic Specialization                | 0.85    | (1.20)    | 0.00   | 7.00    | 934 |
| Economic Specialization (Share)        | 0.13    | (0.17)    | 0.00   | 1.00    | 934 |
| Economic Specialization (Dev)          | 7.74    | (3.59)    | 1.00   | 25.00   | 934 |
| Predicted Population Diversity         | 0.68    | (0.05)    | 0.54   | 0.76    | 934 |
| Absolute Latitude                      | 20.77   | (16.59)   | 0.02   | 71.22   | 934 |
| Area                                   | 0.07    | (0.37)    | 0.00   | 8.97    | 934 |
| Elevation (Avg.)                       | 755.14  | (676.82)  | 1.06   | 4417.96 | 934 |
| Precipitation (Avg.)                   | 105.83  | (71.13)   | 0.00   | 499.24  | 934 |
| Temperature (Avg.)                     | 19.09   | (8.60)    | -15.31 | 29.58   | 934 |
| Malaria Ecology                        | 5.58    | (8.05)    | 0.00   | 33.95   | 934 |
| Ecological Diversity                   | 0.19    | (0.21)    | 0.00   | 0.82    | 934 |
| Agricultural Suitability (avg.)        | 0.76    | (0.34)    | 0.00   | 1.00    | 934 |
| Agricultural Suitability (std.)        | 0.07    | (0.10)    | 0.00   | 0.47    | 934 |
| Caloric Suitability Index (Pre-1500CE) | 2673.34 | (1282.61) | 0.00   | 6955.56 | 934 |
| Caloric Suitability (Pre-1500 ,std.)   | 362.60  | (333.18)  | 0.00   | 2436.89 | 934 |
| Temperature (Spatial Corr., Avg.)      | 0.86    | (0.28)    | 0.00   | 1.00    | 934 |
| Temperature (Volatility, Avg)          | 0.98    | (0.57)    | 0.00   | 3.08    | 934 |
| Pct. Area within 100 kms of Sea        | 0.30    | (0.41)    | 0.00   | 1.00    | 934 |
| Coast Length                           | 0.34    | (2.97)    | 0.00   | 81.92   | 934 |
| Ruggedness (Avg.)                      | 137.45  | (160.05)  | 0.05   | 1137.67 | 934 |
| Pre-Industrial Mobility (avg.)         | 0.27    | (0.07)    | 0.06   | 0.47    | 934 |
| Pre-Industrial Mobility (std.)         | 0.06    | (0.05)    | 0.00   | 0.27    | 934 |

## A.2 Economic Specialization, Trade and Distance to Addis Ababa

### A.2.1 Correlation between Economic Specialization and Trade Measures

Table A.3: Correlation of Economic Specialization and Trade Measures

| Economic Specialization and Trade Measures |                                 |   |                                       |                        |   |         |                  |                    |   |            |
|--|---------------------------------|---|---------------------------------------|------------------------|---|---------|------------------|--------------------|---|------------|
|  | Economic<br>Specializa-<br>tion | Economic<br>Specializa-<br>tion (Share) | Economic<br>Specializa-<br>tion (Dev) | Importance<br>of Trade | Intercommu-<br>nity Trade as<br>Food Source | Money   | Credit<br>Source | Writing<br>Records | and<br>Technological<br>Specializa-<br>tion | Complexity |
| Economic Special-<br>ization               | 1.00                            |   |                                       |                        |   |         |                  |                    |   |            |
| Economic Special-<br>ization (Share)       | 0.94***                         | 1.00                                    |                                       |                        |   |         |                  |                    |   |            |
| Economic Special-<br>ization (Dev)         | 0.87***                         | 0.74***                                 | 1.00                                  |                        |   |         |                  |                    |   |            |
| Importance of<br>Trade                     | 0.20***                         | 0.19**                                  | 0.13*                                 | 1.00                   |   |         |                  |                    |   |            |
| Intercommunity<br>Trade as Food<br>Source  | 0.32***                         | 0.32***                                 | 0.25***                               | 0.77***                | 1.00  |         |                  |                    |   |            |
| Money                                      | 0.35***                         | 0.38***                                 | 0.30***                               | 0.29***                | 0.43***                                     | 1.00    |                  |                    |   |            |
| Credit Source                              | 0.25***                         | 0.21***                                 | 0.26***                               | 0.28***                | 0.31***                                     | 0.35*** | 1.00             |                    |   |            |
| Writing and<br>Records                     | 0.55***                         | 0.55***                                 | 0.47***                               | 0.24***                | 0.31***                                     | 0.38*** | 0.38***          | 1.00               |   |            |
| Technological Spe-<br>cialization          | 0.60***                         | 0.58***                                 | 0.67***                               | 0.18**                 | 0.28***                                     | 0.37*** | 0.26***          | 0.49***            | 1.00  |            |
| Complexity                                 | 0.61***                         | 0.59***                                 | 0.63***                               | 0.27***                | 0.36***                                     | 0.55*** | 0.39***          | 0.70***            | 0.73***                                     | 1.00       |

## A.2.2 Robustness to Clustering and Spatial-Autocorrelation

Table A.4: Expected Heterozygosity and Economic Specialization

|                      | Economic Specialization |         |         |         |        |         |        |
|----------------------|-------------------------|---------|---------|---------|--------|---------|--------|
|                      | (1)                     | (2)     | (3)     | (4)     | (5)    | (6)     | (7)    |
| Expected             | 0.27***                 | 0.31*** | 0.27*** | 0.27*** | 0.25** | 0.27*** | 0.36** |
| Heterozygosity       | (0.05)                  | (0.09)  | (0.09)  | (0.09)  | (0.10) | (0.09)  | (0.14) |
| Absolute Latitude    |                         | 0.15    |         |         |        |         | 0.80   |
|                      |                         | (0.13)  |         |         |        |         | (0.47) |
| Area                 |                         |         | 0.01    |         |        |         | 0.01   |
|                      |                         |         | (0.05)  |         |        |         | (0.06) |
| Elevation (Avg.)     |                         |         |         | -0.03   |        |         | 0.31*  |
|                      |                         |         |         | (0.07)  |        |         | (0.17) |
| Precipitation (Avg.) |                         |         |         |         | -0.08  |         | 0.13   |
|                      |                         |         |         |         | (0.07) |         | (0.21) |
| Temperature (Avg.)   |                         |         |         |         |        | 0.04    | 0.73*  |
|                      |                         |         |         |         |        | (0.08)  | (0.37) |
| Adjusted- $R^2$      | 0.07                    | 0.08    | 0.06    | 0.06    | 0.06   | 0.06    | 0.15   |
| Observations         | 116                     | 116     | 116     | 116     | 116    | 116     | 116    |

Notes: Standardized coefficients. Heteroskedasticity robust standard error estimates clustered at the language phylum level are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

Table A.5: Population Diversity and Economic Specialization  
(Robustness to Clustering and Spatial Auto-Correlation)

|                      | Economic Specialization                           |   |   |   |   |   |   |   |   |
|----------------------|---|---|---|---|---|---|---|---|---|
|                      | (1)   | (2)   | (3)   | (4)   | (5)   | (6)   | (7)   | (8)   | (9)   |
| Population Diversity | 0.36***<br>(0.09)<br>([0.14])<br>[0.11]<br>{0.10} | 0.41***<br>(0.10)<br>([0.14])<br>[0.10]<br>{0.10} | 0.36***<br>(0.10)<br>([0.15])<br>[0.11]<br>{0.09} | 0.37***<br>(0.09)<br>([0.13])<br>[0.10]<br>{0.10} | 0.34***<br>(0.11)<br>([0.16])<br>[0.12]<br>{0.10} | 0.33***<br>(0.10)<br>([0.15])<br>[0.11]<br>{0.10} | 0.31***<br>(0.09)<br>([0.13])<br>[0.10]<br>{0.10} | 0.40***<br>(0.11)<br>([0.16])<br>[0.12]<br>{0.10} | 0.31***<br>(0.10)<br>([0.12])<br>[0.09]<br>{0.09} |
| Main Controls        | Yes   | Yes   | Yes   | Yes   | Yes   | Yes   | Yes   | Yes   | Yes   |
| Additional Controls  | No  | Yes   | Yes   | Yes   | Yes   | Yes   | Yes   | Yes   | Yes   |
| Adjusted- $R^2$      | 0.15  | 0.22  | 0.21  | 0.17  | 0.18  | 0.18  | 0.20  | 0.17  | 0.40  |
| Observations         | 116   | 116   | 116   | 116   | 116   | 116   | 116   | 116   | 116   |

Notes: This Table establishes the robustness of the OLS results to clustering by language phylum and spatial auto-correlation. The additional controls in each column are the ones of the same column in Table 2. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses, clustered at the language phylum in parenthesis and squared brackets, spatial auto-correlation corrected standard errors (Conley, 1999) in squared brackets and Cliff-Ord ML in curly brackets. \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

Table A.6: Population Diversity and Distance to Addis-Ababa  
(Robustness to Clustering and Spatial Auto-Correlation)

|  | Population Diversity                               |  |  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|--|--|--|
|  | Full Sample  |  |  |  |  |  |  |  | Specialization                                     |
|  | (1)  | (2)  | (3)  | (4)  | (5)  | (6)  | (7)  | (8)  | (9)  |
| Pre-Industrial<br>Distance<br>to Addis Ababa | -0.85***<br>(0.07)<br>([0.09])<br>[0.08]<br>{0.04} | -0.80***<br>(0.10)<br>([0.12])<br>[0.10]<br>{0.05} | -0.80***<br>(0.09)<br>([0.11])<br>[0.10]<br>{0.05} | -0.81***<br>(0.09)<br>([0.11])<br>[0.09]<br>{0.05} | -0.80***<br>(0.10)<br>([0.13])<br>[0.11]<br>{0.06} | -0.79***<br>(0.10)<br>([0.12])<br>[0.11]<br>{0.06} | -0.80***<br>(0.10)<br>([0.12])<br>[0.10]<br>{0.06} | -0.82***<br>(0.09)<br>([0.10])<br>[0.08]<br>{0.06} | -0.85***<br>(0.09)<br>([0.10])<br>[0.08]<br>{0.10} |
| Main Controls                                | No   | Yes  | Yes  | Yes  | Yes  | Yes  | Yes  | Yes  | Yes  |
| Additional Controls                          | No   | No   | Yes  | Yes  | Yes  | Yes  | Yes  | Yes  | Yes  |
| Adjusted- $R^2$                              | 0.72   | 0.72   | 0.74   | 0.72   | 0.72   | 0.72   | 0.73   | 0.75   | 0.73   |
| $R^2$  | 0.72   | 0.74   | 0.75   | 0.74   | 0.74   | 0.74   | 0.75   | 0.78   | 0.76   |
| Observations                                 | 144  | 144  | 144  | 144  | 144  | 144  | 144  | 144  | 116  |

Notes: This Table establishes the robustness of the negative effect of the migratory distance on population diversity to clustering by language phylum and spatial auto-correlation. The additional controls in each column are the ones of the same column in Table 3. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses, clustered at the language phylum in parenthesis and squared brackets, spatial auto-correlation corrected standard errors (Conley, 1999) in squared brackets and Cliff-Ord ML in curly brackets. \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

Table A.7: Population Diversity and Economic Specialization (Reduced Form)  
(Robustness to Clustering and Spatial Auto-Correlation)

|                         | Economic Specialization |          |          |          |          |          |          |          |          |
|-------------------------|-------------------------|----------|----------|----------|----------|----------|----------|----------|----------|
|                         | (1)                     | (2)      | (3)      | (4)      | (5)      | (6)      | (7)      | (8)      | (9)      |
| Pre-Industrial          | -0.41***                | -0.43*** | -0.44*** | -0.45*** | -0.37*** | -0.40*** | -0.36*** | -0.46*** | -0.39*** |
| Distance to Addis Ababa | (0.10)                  | (0.10)   | (0.10)   | (0.09)   | (0.12)   | (0.10)   | (0.09)   | (0.11)   | (0.13)   |
|                         | ([0.12])                | ([0.13]) | ([0.12]) | ([0.11]) | ([0.15]) | ([0.13]) | ([0.11]) | ([0.15]) | ([0.15]) |
|                         | [0.09]                  | [0.10]   | [0.10]   | [0.09]   | [0.11]   | [0.10]   | [0.09]   | [0.10]   | [0.10]   |
|                         | {0.10}                  | {0.10}   | {0.10}   | {0.10}   | {0.11}   | {0.10}   | {0.10}   | {0.10}   | {0.10}   |
| Main Controls           | Yes                     | Yes      | Yes      | Yes      | Yes      | Yes      | Yes      | Yes      | Yes      |
| Additional Controls     | No                      | Yes      | Yes      | Yes      | Yes      | Yes      | Yes      | Yes      | Yes      |
| Adjusted- $R^2$         | 0.17                    | 0.23     | 0.25     | 0.21     | 0.19     | 0.21     | 0.22     | 0.19     | 0.42     |
| Observations            | 116                     | 116      | 116      | 116      | 116      | 116      | 116      | 116      | 116      |

Notes: This Table establishes the robustness of the OLS results to clustering by language phylum and spatial auto-correlation. The additional controls in each column are the ones of the same column in Table A.11. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses, clustered at the language phylum in parenthesis and squared brackets, spatial auto-correlation corrected standard errors (Conley, 1999) in squared brackets and Cliff-Ord ML in curly brackets. \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

Table A.8: Population Diversity and Economic Specialization (Reduced Form)  
(Robustness to Clustering and Spatial Auto-Correlation)

|                         | Economic Specialization |          |          |          |          |          |          |          |          |          |          |
|-------------------------|-------------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
|                         | (1)                     | (2)      | (3)      | (4)      | (5)      | (6)      | (7)      | (8)      | (9)      | (10)     | (11)     |
| Pre-Industrial          | -0.45***                | -0.42*** | -0.45*** | -0.42*** | -0.41*** | -0.45*** | -0.40*** | -0.42*** | -0.42*** | -0.47*** | -0.36*** |
| Distance to Addis Ababa | (0.03)                  | (0.03)   | (0.03)   | (0.03)   | (0.03)   | (0.03)   | (0.03)   | (0.03)   | (0.03)   | (0.04)   | (0.10)   |
|                         | ([0.08])                | ([0.08]) | ([0.09]) | ([0.07]) | ([0.07]) | ([0.09]) | ([0.07]) | ([0.08]) | ([0.08]) | ([0.09]) | ([0.20]) |
|                         | [0.07]                  | [0.06]   | [0.06]   | [0.06]   | [0.06]   | [0.07]   | [0.06]   | [0.06]   | [0.06]   | [0.06]   | [0.18]   |
|                         | {0.03}                  | {0.03}   | {0.03}   | {0.03}   | {0.03}   | {0.03}   | {0.03}   | {0.03}   | {0.03}   | {0.03}   | {0.09}   |
| Main Controls           | No                      | Yes      | Yes      | Yes      | Yes      | Yes      | Yes      | Yes      | Yes      | Yes      | Yes      |
| Additional Controls     | Yes                     | Yes      | Yes      | Yes      | Yes      | Yes      | Yes      | Yes      | Yes      | Yes      | Yes      |
| Adjusted- $R^2$         | 0.20                    | 0.22     | 0.23     | 0.24     | 0.24     | 0.24     | 0.24     | 0.22     | 0.23     | 0.29     | 0.32     |
| Observations            | 932                     | 932      | 932      | 932      | 932      | 932      | 932      | 932      | 932      | 932      | 932      |

Notes: This Table establishes the robustness of the OLS results to clustering by language phylum and spatial auto-correlation. The additional controls in each column are the ones of the same column in Table A.11. Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses, clustered at the language phylum in parenthesis and squared brackets, spatial auto-correlation corrected standard errors (Conley, 1999) in squared brackets and Cliff-Ord ML in curly brackets. \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

### A.2.3 Robustness to Measure of Economic Specialization

Table A.9: Population Diversity and Economic Specialization:  
Robustness to Specialization Measure

|                         | Economic Specialization Measures |                   |                   |                   |                  |                  |
|-------------------------|----------------------------------|-------------------|-------------------|-------------------|------------------|------------------|
|                         | Main                             |                   | Share             |                   | Dev              |                  |
|                         | OLS                              | IV                | OLS               | IV                | OLS              | IV               |
|                         | (1)                              | (2)               | (3)               | (4)               | (5)              | (6)              |
| Population Diversity    | 0.27***<br>(0.05)                | 0.46***<br>(0.14) | 0.33***<br>(0.05) | 0.37***<br>(0.12) | 0.13**<br>(0.06) | 0.31**<br>(0.14) |
| Main Controls           | No                               | Yes               | No                | Yes               | No               | Yes              |
| All Additional Controls | No                               | Yes               | No                | Yes               | No               | Yes              |
| First-stage F-statistic |                                  | 81.54             |                   | 81.54             |                  | 81.54            |
| $R^2$                   | 0.08                             | 0.49              | 0.11              | 0.49              | 0.02             | 0.46             |
| Adjusted- $R^2$         | 0.07                             | 0.39              | 0.10              | 0.39              | 0.01             | 0.35             |
| Observations            | 116                              | 116               | 116               | 116               | 116              | 116              |

Notes: Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

### A.2.4 Robustness to Continental Fixed Effects

Table A.10: Predicted Population Diversity and Economic Specialization  
(Robustness to Continental Fixed Effects)

|   | Economic Specialization |                   |                    |                   |                   |                   |                    |                   |                   |                    |
|---|-------------------------|-------------------|--------------------|-------------------|-------------------|-------------------|--------------------|-------------------|-------------------|--------------------|
|   | (1)                     | (2)               | (3)                | (4)               | (5)               | (6)               | (7)                | (8)               | (9)               | (10)               |
| Predicted Population Diversity            | 0.73***<br>(0.15)       | 0.57***<br>(0.15) | 0.58***<br>(0.15)  | 0.51***<br>(0.15) | 0.51***<br>(0.16) | 0.53***<br>(0.16) | 0.59***<br>(0.15)  | 0.55***<br>(0.15) | 0.57***<br>(0.15) | 0.58***<br>(0.17)  |
| Malaria Ecology                           |                         |                   | -0.13***<br>(0.05) |                   |                   |                   |                    |                   |                   | -0.09*<br>(0.05)   |
| Ecological Diversity                      |                         |                   |                    | 0.12***<br>(0.03) |                   |                   |                    |                   |                   | 0.09***<br>(0.03)  |
| Agricultural<br>Suitability (avg.)        |                         |                   |                    |                   | -0.07*<br>(0.04)  |                   |                    |                   |                   | -0.16***<br>(0.04) |
| Agricultural<br>Suitability (std.)        |                         |                   |                    |                   | 0.04<br>(0.05)    |                   |                    |                   |                   | 0.01<br>(0.05)     |
| Caloric Suitability<br>Index (Pre-1500CE) |                         |                   |                    |                   |                   | 0.06*<br>(0.03)   |                    |                   |                   | 0.11***<br>(0.03)  |
| Caloric Suitability<br>(Pre-1500 ,std.)   |                         |                   |                    |                   |                   | 0.08*<br>(0.04)   |                    |                   |                   | 0.07<br>(0.05)     |
| Temperature (Spatial<br>Corr., Avg.)      |                         |                   |                    |                   |                   |                   | 0.01<br>(0.03)     |                   |                   | 0.03<br>(0.04)     |
| Temperature<br>(Volatility, Avg)          |                         |                   |                    |                   |                   |                   | -0.25***<br>(0.06) |                   |                   | -0.14*<br>(0.08)   |
| Pct. Area within<br>100kms of Sea         |                         |                   |                    |                   |                   |                   |                    | 0.05<br>(0.03)    |                   | -0.02<br>(0.04)    |
| Coast Length                              |                         |                   |                    |                   |                   |                   |                    | 0.01<br>(0.03)    |                   | 0.01<br>(0.03)     |
| Ruggedness (Avg.)                         |                         |                   |                    |                   |                   |                   |                    |                   | 0.10<br>(0.06)    | 0.08<br>(0.07)     |
| Pre-Industrial<br>Mobility (avg.)         |                         |                   |                    |                   |                   |                   |                    |                   | 0.15<br>(0.10)    | 0.28**<br>(0.11)   |
| Pre-Industrial<br>Mobility (std.)         |                         |                   |                    |                   |                   |                   |                    |                   | -0.04<br>(0.08)   | -0.14*<br>(0.08)   |
| Continental FE                            | Yes                     | Yes               | Yes                | Yes               | Yes               | Yes               | Yes                | Yes               | Yes               | Yes                |
| Main Controls                             | No                      | Yes               | Yes                | Yes               | Yes               | Yes               | Yes                | Yes               | Yes               | Yes                |
| Adjusted- $R^2$                           | 0.26                    | 0.29              | 0.29               | 0.30              | 0.29              | 0.29              | 0.29               | 0.29              | 0.29              | 0.32               |
| Observations                              | 934                     | 934               | 934                | 934               | 934               | 934               | 934                | 934               | 934               | 934                |

Notes: Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.



## A.2.5 Reduced Form Analysis: Distance to Addis Ababa and Economic Specialization

Table A.11: Population Diversity and Economic Specialization (Reduced Form)

|                         | Economic Specialization |          |          |          |          |          |          |          |          |
|-------------------------|-------------------------|----------|----------|----------|----------|----------|----------|----------|----------|
|                         | (1)                     | (2)      | (3)      | (4)      | (5)      | (6)      | (7)      | (8)      | (9)      |
| Pre-Industrial          | -0.41***                | -0.43*** | -0.44*** | -0.45*** | -0.37*** | -0.40*** | -0.36*** | -0.46*** | -0.39*** |
| Distance to Addis Ababa | (0.10)                  | (0.10)   | (0.10)   | (0.09)   | (0.12)   | (0.10)   | (0.09)   | (0.11)   | (0.13)   |
| Malaria Ecology         |                         | -0.31*** |          |          |          |          |          |          | -0.34*** |
|                         |                         | (0.12)   |          |          |          |          |          |          | (0.11)   |
| Ecological Diversity    |                         |          | 0.30***  |          |          |          |          |          | 0.23**   |
|                         |                         |          | (0.10)   |          |          |          |          |          | (0.10)   |
| Agricultural            |                         |          |          | 0.00     |          |          |          |          | 0.06     |
| Suitability (avg.)      |                         |          |          | (0.13)   |          |          |          |          | (0.10)   |
| Agricultural            |                         |          |          | 0.28**   |          |          |          |          | 0.36**   |
| Suitability (std.)      |                         |          |          | (0.13)   |          |          |          |          | (0.14)   |
| Caloric Suitability     |                         |          |          |          | -0.18    |          |          |          | -0.23    |
| (Pre-1500 ,avg.)        |                         |          |          |          | (0.15)   |          |          |          | (0.15)   |
| Caloric Suitability     |                         |          |          |          | 0.25**   |          |          |          | -0.01    |
| (Pre-1500 ,std.)        |                         |          |          |          | (0.12)   |          |          |          | (0.15)   |
| Temperature (Spatial    |                         |          |          |          |          | 0.04     |          |          | -0.01    |
| Corr., Avg.)            |                         |          |          |          |          | (0.10)   |          |          | (0.07)   |
| Temperature             |                         |          |          |          |          | -0.64*** |          |          | -0.21    |
| (Volatility, Avg)       |                         |          |          |          |          | (0.20)   |          |          | (0.20)   |
| Pct. Area within        |                         |          |          |          |          |          | 0.02     |          | -0.12    |
| 100kms of Sea           |                         |          |          |          |          |          | (0.10)   |          | (0.11)   |
| Coast Length            |                         |          |          |          |          |          | 0.47**   |          | 0.59***  |
|                         |                         |          |          |          |          |          | (0.21)   |          | (0.18)   |
| Ruggedness (Avg.)       |                         |          |          |          |          |          |          | -0.32    | -0.05    |
|                         |                         |          |          |          |          |          |          | (0.23)   | (0.19)   |
| Pre-Industrial          |                         |          |          |          |          |          |          | 0.93**   | 1.15**   |
| Mobility (avg.)         |                         |          |          |          |          |          |          | (0.42)   | (0.49)   |
| Pre-Industrial          |                         |          |          |          |          |          |          | -0.08    | -0.36**  |
| Mobility (std.)         |                         |          |          |          |          |          |          | (0.13)   | (0.16)   |
| Main Controls           | Yes                     | Yes      | Yes      | Yes      | Yes      | Yes      | Yes      | Yes      | Yes      |
| Adjusted- $R^2$         | 0.17                    | 0.23     | 0.25     | 0.21     | 0.19     | 0.21     | 0.22     | 0.19     | 0.42     |
| Observations            | 116                     | 116      | 116      | 116      | 116      | 116      | 116      | 116      | 116      |

Notes: Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

Table A.12: Distance to Addis Ababa and Economic Specialization

|   | Economic Specialization |                    |                    |                    |                    |                    |                    |                    |                    |                    |                    |
|---|-------------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
|   | (1)                     | (2)                | (3)                | (4)                | (5)                | (6)                | (7)                | (8)                | (9)                | (10)               | (11)               |
| Pre-Industrial Dist. to Addis Ababa         | -0.45***<br>(0.03)      | -0.42***<br>(0.03) | -0.45***<br>(0.03) | -0.42***<br>(0.03) | -0.41***<br>(0.03) | -0.45***<br>(0.03) | -0.40***<br>(0.03) | -0.42***<br>(0.03) | -0.42***<br>(0.03) | -0.47***<br>(0.04) | -0.36***<br>(0.10) |
| Malaria Ecology                             |                         |                    | -0.10**<br>(0.04)  |                    |                    |                    |                    |                    |                    | -0.07<br>(0.05)    | -0.06<br>(0.05)    |
| Ecological Diversity                        |                         |                    |                    | 0.15***<br>(0.03)  |                    |                    |                    |                    |                    | 0.10***<br>(0.03)  | 0.09***<br>(0.03)  |
| Agricultural Suitability (avg.)             |                         |                    |                    |                    | -0.05<br>(0.04)    |                    |                    |                    |                    | -0.17***<br>(0.04) | -0.14***<br>(0.04) |
| Agricultural Suitability (std.)             |                         |                    |                    |                    | 0.13***<br>(0.04)  |                    |                    |                    |                    | 0.09**<br>(0.04)   | 0.07<br>(0.04)     |
| Caloric Suitability Index (Pre-1500CE)      |                         |                    |                    |                    |                    | 0.09***<br>(0.03)  |                    |                    |                    | 0.15***<br>(0.03)  | 0.11***<br>(0.03)  |
| Caloric Suitability (Pre-1500 ,std.)        |                         |                    |                    |                    |                    | 0.08*<br>(0.05)    |                    |                    |                    | 0.06<br>(0.05)     | 0.07<br>(0.05)     |
| Temperature (Spatial Corr., Avg.)           |                         |                    |                    |                    |                    |                    | 0.02<br>(0.02)     |                    |                    | 0.02<br>(0.04)     | 0.03<br>(0.04)     |
| Temperature (Volatility, Avg)               |                         |                    |                    |                    |                    |                    | -0.32***<br>(0.06) |                    |                    | -0.23***<br>(0.07) | -0.14*<br>(0.08)   |
| Pct. Area within 100kms of Sea Coast Length |                         |                    |                    |                    |                    |                    |                    | 0.02<br>(0.03)     |                    | -0.05<br>(0.04)    | 0.01<br>(0.04)     |
|   |                         |                    |                    |                    |                    |                    |                    | 0.04<br>(0.04)     |                    | 0.04<br>(0.04)     | 0.04<br>(0.03)     |
| Ruggedness (Avg.)                           |                         |                    |                    |                    |                    |                    |                    |                    | 0.12*<br>(0.07)    | 0.05<br>(0.07)     | 0.05<br>(0.07)     |
| Pre-Industrial Mobility (avg.)              |                         |                    |                    |                    |                    |                    |                    |                    | 0.27***<br>(0.10)  | 0.41***<br>(0.11)  | 0.28**<br>(0.11)   |
| Pre-Industrial Mobility (std.)              |                         |                    |                    |                    |                    |                    |                    |                    | -0.12<br>(0.08)    | -0.21***<br>(0.08) | -0.14*<br>(0.08)   |
| Main Controls                               | No                      | Yes                | Yes                | Yes                | Yes                | Yes                | Yes                | Yes                | Yes                | Yes                | Yes                |
| Continental FE                              | No                      | No                 | No                 | No                 | No                 | No                 | No                 | No                 | No                 | No                 | Yes                |
| Adjusted- $R^2$                             | 0.20                    | 0.22               | 0.23               | 0.24               | 0.24               | 0.24               | 0.24               | 0.22               | 0.23               | 0.29               | 0.32               |
| Observations                                | 934                     | 934                | 934                | 934                | 934                | 934                | 934                | 934                | 934                | 934                | 934                |

Notes: Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

Table A.13: Distance to Addis Ababa and Economic Specialization

|  | Economic Specialization |                    |                    |                    |                    |                    |                    |                    |                    |                    |
|--|-------------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
|  | (1)                     | (2)                | (3)                | (4)                | (5)                | (6)                | (7)                | (8)                | (9)                | (10)               |
| Pre-Industrial Distance to Addis Ababa | -0.48***<br>(0.10)      | -0.37***<br>(0.10) | -0.33***<br>(0.10) | -0.35***<br>(0.10) | -0.36***<br>(0.10) | -0.36***<br>(0.11) | -0.40***<br>(0.10) | -0.39***<br>(0.10) | -0.36***<br>(0.11) | -0.36***<br>(0.10) |
| Malaria Ecology                        |                         |                    | -0.09*<br>(0.05)   |                    |                    |                    |                    |                    |                    | -0.06<br>(0.05)    |
| Ecological Diversity                   |                         |                    |                    | 0.13***<br>(0.03)  |                    |                    |                    |                    |                    | 0.09***<br>(0.03)  |
| Agricultural Suitability (avg.)        |                         |                    |                    |                    | -0.05<br>(0.04)    |                    |                    |                    |                    | -0.14***<br>(0.04) |
| Agricultural Suitability (std.)        |                         |                    |                    |                    | 0.10**<br>(0.04)   |                    |                    |                    |                    | 0.07<br>(0.04)     |
| Caloric Suitability Index (Pre-1500CE) |                         |                    |                    |                    |                    | 0.05*<br>(0.03)    |                    |                    |                    | 0.11***<br>(0.03)  |
| Caloric Suitability (Pre-1500 ,std.)   |                         |                    |                    |                    |                    | 0.09**<br>(0.04)   |                    |                    |                    | 0.07<br>(0.05)     |
| Temperature (Spatial Corr., Avg.)      |                         |                    |                    |                    |                    |                    | 0.00<br>(0.03)     |                    |                    | 0.03<br>(0.04)     |
| Temperature (Volatility, Avg)          |                         |                    |                    |                    |                    |                    | -0.25***<br>(0.06) |                    |                    | -0.14*<br>(0.08)   |
| Pct. Area within 100kms of Sea         |                         |                    |                    |                    |                    |                    |                    | 0.06**<br>(0.03)   |                    | 0.01<br>(0.04)     |
| Coast Length                           |                         |                    |                    |                    |                    |                    |                    | 0.04<br>(0.03)     |                    | 0.04<br>(0.03)     |
| Ruggedness (Avg.)                      |                         |                    |                    |                    |                    |                    |                    |                    | 0.10<br>(0.06)     | 0.05<br>(0.07)     |
| Pre-Industrial Mobility (avg.)         |                         |                    |                    |                    |                    |                    |                    |                    | 0.16<br>(0.11)     | 0.28**<br>(0.11)   |
| Pre-Industrial Mobility (std.)         |                         |                    |                    |                    |                    |                    |                    |                    | -0.06<br>(0.08)    | -0.14*<br>(0.08)   |
| Continental FE                         | Yes                     | Yes                | Yes                | Yes                | Yes                | Yes                | Yes                | Yes                | Yes                | Yes                |
| Main Controls                          | No                      | Yes                | Yes                | Yes                | Yes                | Yes                | Yes                | Yes                | Yes                | Yes                |
| Adjusted- $R^2$                        | 0.26                    | 0.28               | 0.29               | 0.30               | 0.29               | 0.29               | 0.29               | 0.29               | 0.29               | 0.32               |
| Observations                           | 934                     | 934                | 934                | 934                | 934                | 934                | 934                | 934                | 934                | 934                |

Notes: Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

Table A.14: Heterogeneous Effects of Distance to Addis Ababa on Economic Specialization

|   | Economic Specialization |                    |                    |                    |                    |                    |
|---|-------------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
|   | (1)                     | (2)                | (3)                | (4)                | (5)                | (6)                |
| Pre-Industrial Distance to Addis Ababa                                      | -0.07***<br>(0.00)      | -0.05***<br>(0.01) | -0.06***<br>(0.00) | -0.02**<br>(0.01)  | -0.05***<br>(0.00) | -0.05***<br>(0.01) |
| Ecological Diversity  |                         | 1.55***<br>(0.32)  |                    |                    |                    |                    |
| Pre-Industrial Distance to Addis Ababa<br>× Ecological Diversity            |                         | -0.07***<br>(0.02) |                    |                    |                    |                    |
| Agricultural Suitability (std.)   |                         |                    | 3.84***<br>(0.91)  |                    |                    |                    |
| Pre-Industrial Distance to Addis Ababa<br>× Agricultural Suitability (std.) |                         |                    | -0.24***<br>(0.07) |                    |                    |                    |
| Temperature (Volatility, Avg)   |                         |                    |                    | -0.01<br>(0.20)    |                    |                    |
| Pre-Industrial Distance to Addis Ababa<br>× Temperature (Volatility, Avg)   |                         |                    |                    | -0.05***<br>(0.01) |                    |                    |
| Ruggedness (Avg.)   |                         |                    |                    |                    | 0.00***<br>(0.00)  |                    |
| Pre-Industrial Distance to Addis Ababa<br>× Ruggedness (Avg.)               |                         |                    |                    |                    | -0.00***<br>(0.00) |                    |
| Pre-Industrial Mobility (std.)  |                         |                    |                    |                    |                    | 6.93***<br>(1.99)  |
| Pre-Industrial Distance to Addis Ababa<br>× Pre-Industrial Mobility (std.)  |                         |                    |                    |                    |                    | -0.44***<br>(0.14) |
| Main Controls   | Yes                     | Yes                | Yes                | Yes                | Yes                | Yes                |
| Adjusted- $R^2$   | 0.22                    | 0.25               | 0.25               | 0.25               | 0.26               | 0.24               |
| Observations  | 934                     | 934                | 934                | 934                | 934                | 934                |

Notes: Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

Table A.15: Distance to Addis Ababa, Pre-Industrial Development, and Economic Specialization

|  | Economic Specialization |                    |                    |                    |                    |                    |
|--|-------------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
|  | Full Sample             |                    |                    |                    | Community Sample   | Size               |
|  | (1)                     | (2)                | (3)                | (4)                | (5)                | (6)                |
| Pre-Industrial Distance to Addis Ababa | -0.42***<br>(0.03)      | -0.44***<br>(0.06) | -0.35***<br>(0.02) | -0.43***<br>(0.03) | -0.41***<br>(0.03) | -0.28***<br>(0.03) |
| Origtime                               |                         | -0.03<br>(0.07)    |                    |                    |                    |                    |
| Years Since Neolithic Revolution       |                         |                    | 0.27***<br>(0.04)  |                    |                    |                    |
| Population Density (1500CE)            |                         |                    |                    | 0.11***<br>(0.04)  |                    |                    |
| Main Controls                          | Yes                     | Yes                | Yes                | Yes                | Yes                | Yes                |
| Community Size FE                      | No                      | No                 | No                 | No                 | No                 | Yes                |
| Adjusted- $R^2$                        | 0.22                    | 0.22               | 0.29               | 0.24               | 0.22               | 0.47               |
| Observations                           | 938                     | 927                | 926                | 913                | 512                | 512                |

Notes: Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

Table A.16: Distance to Addis Ababa, Pre-Industrial Development, and Economic Specialization

|  | Economic Specialization |                    |                    |                    |                       |                    |
|--|-------------------------|--------------------|--------------------|--------------------|-----------------------|--------------------|
|  | Full Sample             |                    |                    |                    | Community Size Sample |                    |
|  | (1)                     | (2)                | (3)                | (4)                | (5)                   | (6)                |
| Pre-Industrial Distance to Addis Ababa | -0.42***<br>(0.03)      | -0.44***<br>(0.06) | -0.35***<br>(0.02) | -0.43***<br>(0.03) | -0.41***<br>(0.03)    | -0.27***<br>(0.03) |
| Origtime                               |                         | -0.03<br>(0.07)    |                    |                    |                       |                    |
| Years Since Neolithic Revolution       |                         |                    | 0.27***<br>(0.04)  |                    |                       |                    |
| Population Density (1500CE)            |                         |                    |                    | 0.11***<br>(0.04)  |                       |                    |
| Mean Size of Local Communities         |                         |                    |                    |                    |                       | 0.45***<br>(0.05)  |
| Main Controls                          | Yes                     | Yes                | Yes                | Yes                | Yes                   | Yes                |
| Adjusted- $R^2$                        | 0.22                    | 0.22               | 0.29               | 0.24               | 0.22                  | 0.39               |
| Observations                           | 938                     | 927                | 926                | 913                | 512                   | 512                |

Notes: Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

Table A.17: Distance to Addis Ababa, Pre-Industrial Development, and Economic Specialization

|   | Pre-Industrial Measures of Trade |                        |   |                    |                   |                        |                                      |                    |
|---|----------------------------------|------------------------|---|--------------------|-------------------|------------------------|--------------------------------------|--------------------|
|   | Economic<br>Specializa-<br>tion  | Importance<br>of Trade | Intercommu-<br>nity Trade as<br>Food Source | Money              | Credit<br>Source  | Writing and<br>Records | Technological<br>Specializa-<br>tion | Complexity         |
|   | (1)                              | (2)                    | (3)   | (4)                | (5)               | (6)                    | (7)                                  | (8)                |
| Pre-Industrial<br>Distance to Addis Ababa | -0.38***<br>(0.05)               | -0.23***<br>(0.06)     | -0.39***<br>(0.07)                          | -0.45***<br>(0.06) | -0.18**<br>(0.08) | -0.21***<br>(0.06)     | -0.29***<br>(0.05)                   | -0.36***<br>(0.06) |
| Main Controls                             | Yes                              | Yes                    | Yes   | Yes                | Yes               | Yes                    | Yes                                  | Yes                |
| Additonal Controls                        | Yes                              | Yes                    | Yes   | Yes                | Yes               | Yes                    | Yes                                  | Yes                |
| Adjusted- $R^2$                           | 0.34                             | 0.04                   | 0.18  | 0.25               | 0.14              | 0.27                   | 0.28                                 | 0.31               |
| Observations                              | 168                              | 168                    | 165   | 165                | 153               | 168                    | 168                                  | 168                |

Notes: Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

## B Linguistic Diversity and Economic Specialization

This section explores the relation between economic specialization and another proxy of population diversity, as measured by linguistic diversity. In this paper linguistic diversity refers to a language's diversity in terms of number of genders, consonant inventory, and vowel quality inventory (Haspelmath and Bibiko, 2005) and not to the number of languages in a location, i.e., in captures diversity *within* a population and *not across* populations.

Table B.1: Linguistic Diversity and Distance to Addis-Ababa

|  | Linguistic Diversity (Consonant Inventory) |                    |                    |                    |                    |                    |                    |                    |                    |
|--|--|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
|  | Full Sample                                |                    |                    |                    |                    |                    |                    |                    | Specialization     |
|  | (1)  | (2)                | (3)                | (4)                | (5)                | (6)                | (7)                | (8)                | (9)                |
| Pre-Industrial Distance to Addis Ababa | -0.42***<br>(0.05)                         | -0.36***<br>(0.05) | -0.35***<br>(0.05) | -0.35***<br>(0.05) | -0.36***<br>(0.05) | -0.36***<br>(0.05) | -0.38***<br>(0.06) | -0.37***<br>(0.06) | -0.37***<br>(0.07) |
| Malaria Ecology                        |  |                    | 0.04<br>(0.06)     |                    |                    |                    |                    | -0.00<br>(0.07)    | -0.04<br>(0.08)    |
| Agricultural Suitability (avg.)        |  |                    |                    | -0.05<br>(0.06)    |                    |                    |                    | -0.05<br>(0.08)    | -0.04<br>(0.08)    |
| Agricultural Suitability (std.)        |  |                    |                    | 0.05<br>(0.05)     |                    |                    |                    | 0.09<br>(0.06)     | 0.06<br>(0.06)     |
| Caloric Suitability (Pre-1500 ,avg.)   |  |                    |                    |                    | 0.01<br>(0.06)     |                    |                    | 0.03<br>(0.07)     | 0.03<br>(0.08)     |
| Caloric Suitability (Pre-1500 ,std.)   |  |                    |                    |                    | -0.11*<br>(0.06)   |                    |                    | -0.13*<br>(0.07)   | -0.11<br>(0.07)    |
| Pct. Area within 100kms of Sea         |  |                    |                    |                    |                    | -0.07<br>(0.06)    |                    | -0.08<br>(0.06)    | -0.10<br>(0.07)    |
| Coast Length                           |  |                    |                    |                    |                    | 0.11<br>(0.09)     |                    | 0.13<br>(0.08)     | 0.14<br>(0.09)     |
| Ruggedness (Avg.)                      |  |                    |                    |                    |                    |                    | 0.10<br>(0.12)     | 0.10<br>(0.13)     | 0.07<br>(0.14)     |
| Pre-Industrial Mobility (avg.)         |  |                    |                    |                    |                    |                    | 0.30<br>(0.22)     | 0.30<br>(0.25)     | 0.26<br>(0.27)     |
| Pre-Industrial Mobility (std.)         |  |                    |                    |                    |                    |                    | -0.17<br>(0.11)    | -0.14<br>(0.12)    | -0.06<br>(0.13)    |
| Main Controls                          | No   | Yes                | Yes                | Yes                | Yes                | Yes                | Yes                | Yes                | Yes                |
| Adjusted- $R^2$                        | 0.17                                       | 0.26               | 0.26               | 0.26               | 0.26               | 0.26               | 0.26               | 0.27               | 0.27               |
| $R^2$                                  | 0.17                                       | 0.27               | 0.27               | 0.28               | 0.28               | 0.28               | 0.28               | 0.31               | 0.31               |
| Observations                           | 299  | 299                | 299                | 299                | 299                | 299                | 299                | 299                | 254                |

Notes: Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.



Table B.2: Linguistic Diversity and Distance to Addis-Ababa

|                         | Linguistic Diversity (Vowel Quality Inventory) |          |          |          |          |          |          |          |                |
|-------------------------|--|----------|----------|----------|----------|----------|----------|----------|----------------|
|                         | Full Sample                                    |          |          |          |          |          |          |          | Specialization |
|                         | (1)  | (2)      | (3)      | (4)      | (5)      | (6)      | (7)      | (8)      | (9)            |
| Pre-Industrial          | -0.29***                                       | -0.32*** | -0.27*** | -0.32*** | -0.34*** | -0.34*** | -0.36*** | -0.31*** | -0.33***       |
| Distance to Addis Ababa | (0.06)   | (0.07)   | (0.07)   | (0.07)   | (0.07)   | (0.07)   | (0.07)   | (0.07)   | (0.08)         |
| Malaria Ecology         |  |          | 0.20***  |          |          |          |          | 0.21***  | 0.23***        |
|                         |  |          | (0.06)   |          |          |          |          | (0.07)   | (0.07)         |
| Agricultural            |  |          |          | 0.01     |          |          |          | -0.16**  | -0.19**        |
| Suitability (avg.)      |  |          |          | (0.06)   |          |          |          | (0.08)   | (0.08)         |
| Agricultural            |  |          |          | 0.05     |          |          |          | 0.03     | 0.03           |
| Suitability (std.)      |  |          |          | (0.06)   |          |          |          | (0.06)   | (0.07)         |
| Caloric Suitability     |  |          |          |          | 0.14**   |          |          | 0.20**   | 0.25***        |
| (Pre-1500 ,avg.)        |  |          |          |          | (0.07)   |          |          | (0.08)   | (0.08)         |
| Caloric Suitability     |  |          |          |          | -0.08    |          |          | -0.04    | -0.04          |
| (Pre-1500 ,std.)        |  |          |          |          | (0.07)   |          |          | (0.07)   | (0.07)         |
| Pct. Area within        |  |          |          |          |          | -0.19*** |          | -0.09    | -0.12          |
| 100kms of Sea           |  |          |          |          |          | (0.06)   |          | (0.07)   | (0.07)         |
| Coast Length            |  |          |          |          |          | 0.07     |          | 0.10     | 0.12           |
|                         |  |          |          |          |          | (0.09)   |          | (0.10)   | (0.11)         |
| Ruggedness (Avg.)       |  |          |          |          |          |          | -0.17    | -0.22*   | -0.19          |
|                         |  |          |          |          |          |          | (0.12)   | (0.12)   | (0.13)         |
| Pre-Industrial          |  |          |          |          |          |          | 0.36     | 0.35     | 0.28           |
| Mobility (avg.)         |  |          |          |          |          |          | (0.26)   | (0.26)   | (0.27)         |
| Pre-Industrial          |  |          |          |          |          |          | -0.16    | -0.04    | -0.02          |
| Mobility (std.)         |  |          |          |          |          |          | (0.12)   | (0.12)   | (0.13)         |
| Main Controls           | No   | Yes      | Yes      | Yes      | Yes      | Yes      | Yes      | Yes      | Yes            |
| Adjusted- $R^2$         | 0.08   | 0.14     | 0.17     | 0.14     | 0.15     | 0.16     | 0.15     | 0.19     | 0.22           |
| $R^2$                   | 0.09   | 0.16     | 0.19     | 0.16     | 0.17     | 0.19     | 0.18     | 0.23     | 0.27           |
| Observations            | 301  | 301      | 301      | 301      | 301      | 301      | 301      | 301      | 255            |

Notes: Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

Table B.3: Linguistic Diversity and Distance to Addis-Ababa

|  | Linguistic Diversity (Number of Genders) |                    |                    |                    |                    |                    |                    |                    |                    |
|--|--|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
|  | Full Sample                              |                    |                    |                    |                    |                    |                    |                    | Specialization     |
|  | (1)                                      | (2)                | (3)                | (4)                | (5)                | (6)                | (7)                | (8)                | (9)                |
| Pre-Industrial Distance to Addis Ababa | -0.25***<br>(0.06)                       | -0.25***<br>(0.06) | -0.23***<br>(0.06) | -0.26***<br>(0.06) | -0.27***<br>(0.06) | -0.25***<br>(0.06) | -0.31***<br>(0.08) | -0.30***<br>(0.09) | -0.33***<br>(0.08) |
| Malaria Ecology                        |  |                    | 0.11<br>(0.12)     |                    |                    |                    |                    | 0.05<br>(0.15)     | 0.04<br>(0.16)     |
| Agricultural Suitability (avg.)        |  |                    |                    | 0.18**<br>(0.07)   |                    |                    |                    | 0.14<br>(0.09)     | 0.17*<br>(0.09)    |
| Agricultural Suitability (std.)        |  |                    |                    | -0.08<br>(0.07)    |                    |                    |                    | -0.11<br>(0.09)    | -0.08<br>(0.09)    |
| Caloric Suitability (Pre-1500 ,avg.)   |  |                    |                    |                    | 0.21**<br>(0.10)   |                    |                    | 0.18<br>(0.12)     | 0.13<br>(0.10)     |
| Caloric Suitability (Pre-1500 ,std.)   |  |                    |                    |                    | -0.15<br>(0.11)    |                    |                    | -0.13<br>(0.12)    | -0.08<br>(0.11)    |
| Pct. Area within 100kms of Sea         |  |                    |                    |                    |                    | -0.05<br>(0.09)    |                    | -0.00<br>(0.09)    | -0.01<br>(0.10)    |
| Coast Length                           |  |                    |                    |                    |                    | 0.10<br>(0.13)     |                    | 0.18<br>(0.13)     | 0.23<br>(0.16)     |
| Ruggedness (Avg.)                      |  |                    |                    |                    |                    |                    | 0.23<br>(0.15)     | 0.33**<br>(0.13)   | 0.38***<br>(0.14)  |
| Pre-Industrial Mobility (avg.)         |  |                    |                    |                    |                    |                    | -0.06<br>(0.26)    | -0.47<br>(0.33)    | -0.58<br>(0.35)    |
| Pre-Industrial Mobility (std.)         |  |                    |                    |                    |                    |                    | -0.31*<br>(0.16)   | -0.19<br>(0.20)    | -0.20<br>(0.19)    |
| Main Controls                          | No                                       | Yes                | Yes                | Yes                | Yes                | Yes                | Yes                | Yes                | Yes                |
| Adjusted- $R^2$                        | 0.06                                     | 0.17               | 0.17               | 0.18               | 0.18               | 0.16               | 0.19               | 0.20               | 0.23               |
| $R^2$                                  | 0.06                                     | 0.20               | 0.21               | 0.22               | 0.23               | 0.20               | 0.23               | 0.28               | 0.32               |
| Observations                           | 152                                      | 152                | 152                | 152                | 152                | 152                | 152                | 152                | 130                |

Notes: Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

Table B.4: Population Diversity and Economic Specialization (IV)  
Linguistic Diversity (Consonant Inventories)

|   | Economic Specialization |                   |                   |                   |                   |                   |                   |                   |                   |                   |
|---|-------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
|   | OLS                     | IV                |                   |                   |                   |                   |                   |                   |                   |                   |
|   | (1)                     | (2)               | (3)               | (4)               | (5)               | (6)               | (7)               | (8)               | (9)               | (10)              |
| Linguistic Diversity                    | 0.20***<br>(0.06)       | 1.13***<br>(0.20) | 1.15***<br>(0.22) | 1.12***<br>(0.20) | 1.13***<br>(0.21) | 1.18***<br>(0.21) | 1.12***<br>(0.21) | 1.13***<br>(0.20) | 1.15***<br>(0.22) | 1.22***<br>(0.28) |
| Malaria Ecology                         |                         |                   | -0.03<br>(0.10)   |                   |                   |                   |                   |                   |                   | 0.01<br>(0.11)    |
| Ecological Diversity                    |                         |                   |                   | 0.26***<br>(0.09) |                   |                   |                   |                   |                   | 0.22**<br>(0.11)  |
| Agricultural<br>Suitability (avg.)      |                         |                   |                   |                   | -0.00<br>(0.10)   |                   |                   |                   |                   | -0.19<br>(0.12)   |
| Agricultural<br>Suitability (std.)      |                         |                   |                   |                   | 0.12<br>(0.11)    |                   |                   |                   |                   | -0.03<br>(0.12)   |
| Caloric Suitability<br>(Pre-1500 ,avg.) |                         |                   |                   |                   |                   | 0.07<br>(0.11)    |                   |                   |                   | 0.09<br>(0.12)    |
| Caloric Suitability<br>(Pre-1500 ,std.) |                         |                   |                   |                   |                   | 0.21*<br>(0.12)   |                   |                   |                   | 0.23<br>(0.15)    |
| Temperature (Spatial<br>Corr., Avg.)    |                         |                   |                   |                   |                   |                   | -0.11<br>(0.09)   |                   |                   | -0.27**<br>(0.13) |
| Temperature<br>(Volatility, Avg)        |                         |                   |                   |                   |                   |                   | 0.22<br>(0.23)    |                   |                   | 0.31<br>(0.28)    |
| Pct. Area within<br>100kms of Sea       |                         |                   |                   |                   |                   |                   |                   | -0.05<br>(0.09)   |                   | -0.14<br>(0.11)   |
| Coast Length                            |                         |                   |                   |                   |                   |                   |                   | 0.14<br>(0.21)    |                   | 0.15<br>(0.19)    |
| Ruggedness (Avg.)                       |                         |                   |                   |                   |                   |                   |                   |                   | -0.16<br>(0.18)   | -0.17<br>(0.19)   |
| Pre-Industrial<br>Mobility (avg.)       |                         |                   |                   |                   |                   |                   |                   |                   | -0.00<br>(0.31)   | 0.42<br>(0.35)    |
| Pre-Industrial<br>Mobility (std.)       |                         |                   |                   |                   |                   |                   |                   |                   | 0.03<br>(0.17)    | -0.17<br>(0.20)   |
| Main Controls                           | Yes                     | Yes               | Yes               | Yes               | Yes               | Yes               | Yes               | Yes               | Yes               | Yes               |
| First-stage F-statistic                 |                         | 46.23             | 44.68             | 45.37             | 43.58             | 44.29             | 43.27             | 44.19             | 36.73             | 27.68             |
| Adjusted- $R^2$                         | 0.06                    | -0.67             | -0.71             | -0.61             | -0.68             | -0.73             | -0.67             | -0.68             | -0.72             | -0.75             |
| Observations                            | 254                     | 254               | 254               | 254               | 254               | 254               | 254               | 254               | 254               | 254               |

Notes: Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

Table B.5: Population Diversity and Economic Specialization (IV)  
Linguistic Diversity (Vowel Quality Inventories)

|   | Economic Specialization |                   |                   |                   |                   |                   |                    |                   |                   |                   |
|---|-------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|--------------------|-------------------|-------------------|-------------------|
|   | OLS                     | IV                |                   |                   |                   |                   |                    |                   |                   |                   |
|   | (1)                     | (2)               | (3)               | (4)               | (5)               | (6)               | (7)                | (8)               | (9)               | (10)              |
| Linguistic Diversity                    | 0.39***<br>(0.06)       | 1.15***<br>(0.24) | 1.38***<br>(0.34) | 1.16***<br>(0.24) | 1.14***<br>(0.25) | 1.13***<br>(0.22) | 1.02***<br>(0.21)  | 1.06***<br>(0.21) | 1.05***<br>(0.22) | 1.29***<br>(0.33) |
| Malaria Ecology                         |                         |                   | -0.36**<br>(0.14) |                   |                   |                   |                    |                   |                   | -0.34**<br>(0.17) |
| Ecological Diversity                    |                         |                   |                   | -0.07<br>(0.10)   |                   |                   |                    |                   |                   | -0.19<br>(0.12)   |
| Agricultural<br>Suitability (avg.)      |                         |                   |                   |                   | -0.04<br>(0.09)   |                   |                    |                   |                   | 0.01<br>(0.15)    |
| Agricultural<br>Suitability (std.)      |                         |                   |                   |                   | 0.08<br>(0.09)    |                   |                    |                   |                   | 0.10<br>(0.11)    |
| Caloric Suitability<br>(Pre-1500 ,avg.) |                         |                   |                   |                   |                   | -0.10<br>(0.08)   |                    |                   |                   | -0.11<br>(0.14)   |
| Caloric Suitability<br>(Pre-1500 ,std.) |                         |                   |                   |                   |                   | 0.19*<br>(0.10)   |                    |                   |                   | 0.20*<br>(0.11)   |
| Temperature (Spatial<br>Corr., Avg.)    |                         |                   |                   |                   |                   |                   | -0.06<br>(0.08)    |                   |                   | -0.05<br>(0.12)   |
| Temperature<br>(Volatility, Avg)        |                         |                   |                   |                   |                   |                   | -0.42***<br>(0.16) |                   |                   | -0.27<br>(0.23)   |
| Pct. Area within<br>100kms of Sea       |                         |                   |                   |                   |                   |                   |                    | 0.11<br>(0.09)    |                   | 0.00<br>(0.14)    |
| Coast Length                            |                         |                   |                   |                   |                   |                   |                    | 0.19**<br>(0.09)  |                   | 0.15<br>(0.09)    |
| Ruggedness (Avg.)                       |                         |                   |                   |                   |                   |                   |                    |                   | 0.06<br>(0.15)    | -0.06<br>(0.20)   |
| Pre-Industrial<br>Mobility (avg.)       |                         |                   |                   |                   |                   |                   |                    |                   | -0.04<br>(0.29)   | 0.42<br>(0.36)    |
| Pre-Industrial<br>Mobility (std.)       |                         |                   |                   |                   |                   |                   |                    |                   | 0.13<br>(0.17)    | -0.11<br>(0.22)   |
| Main Controls                           | Yes                     | Yes               | Yes               | Yes               | Yes               | Yes               | Yes                | Yes               | Yes               | Yes               |
| First-stage F-statistic                 |                         | 22.60             | 16.32             | 24.31             | 21.43             | 26.88             | 27.78              | 30.08             | 26.28             | 18.78             |
| Adjusted- $R^2$                         | 0.17                    | -0.39             | -0.69             | -0.39             | -0.37             | -0.35             | -0.19              | -0.24             | -0.24             | -0.52             |
| Observations                            | 255                     | 255               | 255               | 255               | 255               | 255               | 255                | 255               | 255               | 255               |

Notes: Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

Table B.6: Population Diversity and Economic Specialization (IV)  
Linguistic Diversity (Number of Genders)

|   | Economic Specialization |                   |                   |                   |                   |                   |                   |                   |                   |                    |
|---|-------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|--------------------|
|   | OLS                     | IV                |                   |                   |                   |                   |                   |                   |                   |                    |
|   | (1)                     | (2)               | (3)               | (4)               | (5)               | (6)               | (7)               | (8)               | (9)               | (10)               |
| Linguistic Diversity                    | 0.13<br>(0.08)          | 1.19***<br>(0.28) | 1.30***<br>(0.34) | 1.16***<br>(0.28) | 1.18***<br>(0.29) | 1.21***<br>(0.29) | 1.09***<br>(0.26) | 1.16***<br>(0.28) | 0.91***<br>(0.22) | 0.87***<br>(0.31)  |
| Malaria Ecology                         |                         |                   | -0.20<br>(0.23)   |                   |                   |                   |                   |                   |                   | 0.01<br>(0.21)     |
| Ecological Diversity                    |                         |                   |                   | 0.20<br>(0.12)    |                   |                   |                   |                   |                   | 0.04<br>(0.13)     |
| Agricultural<br>Suitability (avg.)      |                         |                   |                   |                   | -0.20<br>(0.13)   |                   |                   |                   |                   | -0.35**<br>(0.16)  |
| Agricultural<br>Suitability (std.)      |                         |                   |                   |                   | 0.26*<br>(0.13)   |                   |                   |                   |                   | 0.17<br>(0.17)     |
| Caloric Suitability<br>(Pre-1500 ,avg.) |                         |                   |                   |                   |                   | -0.01<br>(0.15)   |                   |                   |                   | 0.23<br>(0.16)     |
| Caloric Suitability<br>(Pre-1500 ,std.) |                         |                   |                   |                   |                   | 0.19<br>(0.17)    |                   |                   |                   | 0.09<br>(0.18)     |
| Temperature (Spatial<br>Corr., Avg.)    |                         |                   |                   |                   |                   |                   | -0.07<br>(0.14)   |                   |                   | -0.03<br>(0.17)    |
| Temperature<br>(Volatility, Avg)        |                         |                   |                   |                   |                   |                   | -0.34<br>(0.30)   |                   |                   | -0.45<br>(0.30)    |
| Pct. Area within<br>100kms of Sea       |                         |                   |                   |                   |                   |                   |                   | 0.02<br>(0.14)    |                   | -0.14<br>(0.18)    |
| Coast Length                            |                         |                   |                   |                   |                   |                   |                   | 0.15<br>(0.33)    |                   | 0.29<br>(0.30)     |
| Ruggedness (Avg.)                       |                         |                   |                   |                   |                   |                   |                   |                   | -0.47**<br>(0.19) | -0.59***<br>(0.20) |
| Pre-Industrial<br>Mobility (avg.)       |                         |                   |                   |                   |                   |                   |                   |                   | 0.13<br>(0.35)    | 0.66<br>(0.53)     |
| Pre-Industrial<br>Mobility (std.)       |                         |                   |                   |                   |                   |                   |                   |                   | 0.46***<br>(0.16) | 0.26<br>(0.20)     |
| Main Controls                           | Yes                     | Yes               | Yes               | Yes               | Yes               | Yes               | Yes               | Yes               | Yes               | Yes                |
| First-stage F-statistic                 |                         | 29.97             | 24.50             | 29.02             | 29.16             | 28.00             | 30.19             | 28.64             | 33.65             | 17.21              |
| Adjusted- $R^2$                         | -0.01                   | -1.03             | -1.25             | -0.96             | -0.96             | -1.07             | -0.85             | -1.00             | -0.52             | -0.43              |
| Observations                            | 130                     | 130               | 130               | 130               | 130               | 130               | 130               | 130               | 130               | 130                |

Notes: Standardized coefficients. Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

## C Persistence

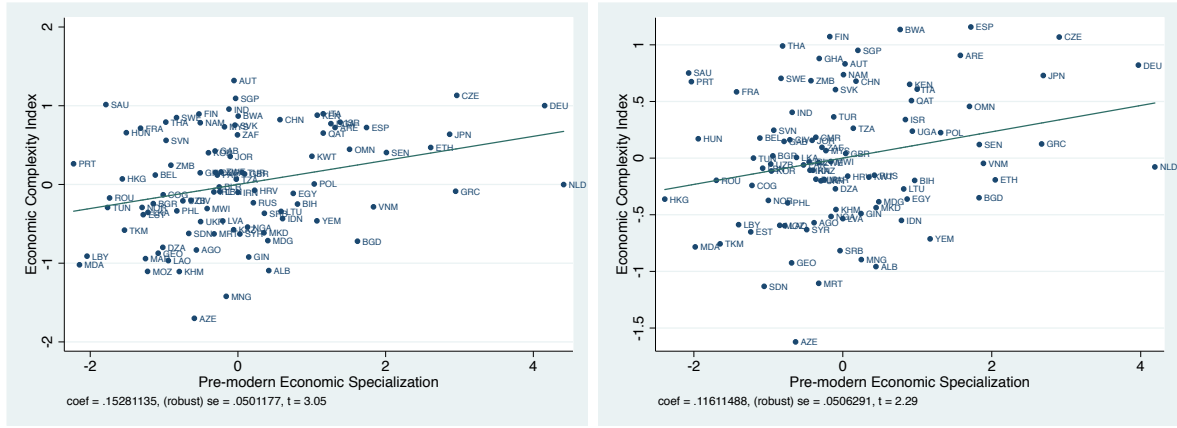


Figure C.1: Pre-modern Economic Specialization and Contemporary Economic Complexity

Table C.1: Pre-modern Economic Specialization and Contemporary Economic Complexity

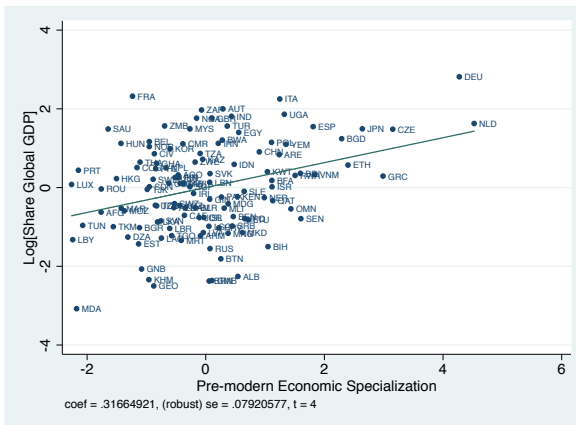
|                                    | Contemporary Development  |                   |                                 |                  |
|------------------------------------|---------------------------|-------------------|---------------------------------|------------------|
|                                    | Economic Complexity Index | # Goods Exported  | Ratio # Goods Exported/Imported | Share Global GDP |
|                                    | (1)                       | (2)               | (3)                             | (4)              |
| Pre-modern Economic Specialization | 0.19***<br>(0.07)         | 0.21***<br>(0.08) | 0.19**<br>(0.08)                | 0.16**<br>(0.06) |
| Continental FE                     | Yes                       | Yes               | Yes                             | Yes              |
| Geographical Controls              | Yes                       | Yes               | Yes                             | Yes              |
| Years Since Neolithic Transition   | Yes                       | Yes               | Yes                             | Yes              |
| Adjusted- $R^2$                    | 0.62                      | 0.46              | 0.46                            | 0.58             |
| Observations                       | 95                        | 80                | 80                              | 120              |

Notes: Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.

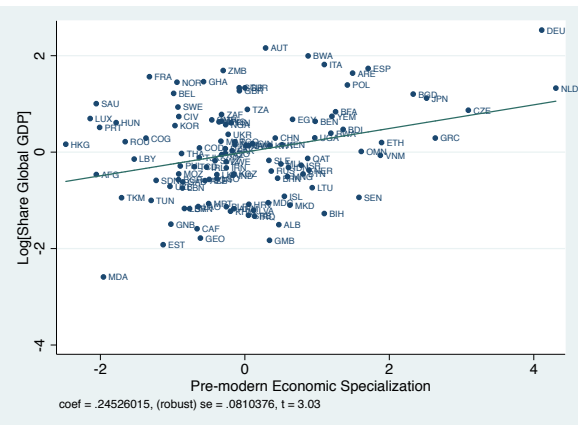
Table C.2: Pre-modern Economic Specialization and Share of Global GDP

|                                    | Log[Share of Global GDP] (2005) |                   |                   |                   |                   |                   |                   |                  |
|------------------------------------|---------------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|------------------|
|                                    | Main Measure                    |                   |                   |                   |                   |                   | Share             | Cardinal         |
|                                    | (1)                             | (2)               | (3)               | (4)               | (5)               | (6)               | (7)               | (8)              |
| Pre-modern Economic Specialization | 0.21***<br>(0.06)               | 0.19***<br>(0.06) | 0.24***<br>(0.06) | 0.23***<br>(0.06) | 0.19***<br>(0.06) | 0.18***<br>(0.06) | 0.20***<br>(0.07) | 0.13**<br>(0.06) |
| Continental FE                     | Yes                             | Yes               | Yes               | Yes               | Yes               | Yes               | Yes               | Yes              |
| Geographical Controls              | No                              | Yes               | Yes               | Yes               | Yes               | Yes               | Yes               | Yes              |
| Caloric Suitability Controls       | No                              | No                | Yes               | Yes               | Yes               | Yes               | Yes               | Yes              |
| Geographical Diversity Controls    | No                              | No                | No                | Yes               | Yes               | Yes               | Yes               | Yes              |
| Disease Controls                   | No                              | No                | No                | No                | Yes               | Yes               | Yes               | Yes              |
| Years Since Neolithic Transition   | No                              | No                | No                | No                | No                | Yes               | Yes               | Yes              |
| Adjusted- $R^2$                    | 0.35                            | 0.52              | 0.60              | 0.61              | 0.70              | 0.70              | 0.70              | 0.69             |
| Observations                       | 120                             | 120               | 120               | 120               | 120               | 120               | 120               | 120              |

Notes: Heteroskedasticity robust standard error estimates are reported in parentheses; \*\*\* denotes statistical significance at the 1% level, \*\* at the 5% level, and \* at the 10% level, all for two-sided hypothesis tests.



(a) Basic Controls (Column 3)



(b) All Controls (Column 6)

Figure C.2: Pre-modern Economic Specialization and Share of Global GDP