Population Diversity, Division of Labor and the Emergence of Trade and State*

Emilio Depetris-Chauvin† and Ömer Özak‡

October 28, 2015

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 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 the positive effect of population diversity on economic specialization and the emergence of trade-related institutions, which, in turn, facilitated the historical formation of states. Additionally, it provides suggestive evidence that regions historically inhabited by pre-modern societies with high levels of economic specialization have a larger occupational heterogeneity and are more developed today.

Keywords: Economic Specialization, Division of Labor, Trade, State Formation, Population Diversity, Population Heterogeneity, Genetic Diversity, Diversity, Emergence of State, Persistence, Out of Africa

JEL Classification: D74, F10, N47, O10, O17, Z10

*We wish to thank Klaus Desmet, Oded Galor, 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 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.

†Departamento de Economía, Universidad de los Andes. E-mail: e.depetris@uniandes.edu.co

‡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 and centralized institutions in the pre-colonial era may partially explain observed differences in contemporary economic outcomes across countries and regions (Berg, 1991; Bockstette et al., 2002; Findlay and O'Rourke, 2007; Greif, 1993; Putterman and Weil, 2010; Smith, 1776).[1] 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 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, facilitated the historical formation of states. Additionally, it provides suggestive evidence that regions historically inhabited by pre-modern societies with high levels of economic specialization have a larger occupational heterogeneity and are more developed today.

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

[1] 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). Similarly, pre-modern era experience with centralized institutions has been associated with higher incomes (Michalopoulos and Papaioannou, 2013), better public good provision (Gennaioli and Rainer, 2007), lower conflict prevalence (Depetris-Chauvin, 2014), and higher levels of autocracy (Hariri, 2012).

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. Additionally, the emergence of trade and this increased economic potential facilitated the emergence of states by increasing the means to maintain them and, on the other hand, increasing the gains of (i) solving coordination problems, (ii) providing public goods, as well as (iii) the expropriation of part of the surplus by a small sector of the population. Thus, the theory predicts that during the pre-modern era economic specialization, trade and the state should emerge and be more prevalent among diverse populations.

To empirically test this hypothesis, this research constructs a novel dataset of ethnic level measures of 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. The research constructs for over 1100 ethnicities novel measures of the number of economic activities in which specialization existed in the pre-modern era. Additionally, in order to overcome the lack of historical population diversity data, the analysis exploits data on genetic diversity for 267 ethnic groups. Underlying this measure of diversity is a fundamental sampling process generated by the serial founder effect behind the dispersion of anatomically modern humans out of East Africa more than $60,000$ years ago (Ramachandran et al., 2005). In particular, the successive divisions of an original population into various subpopulations generates a loss in the diversity in cultural traits, preferences, knowledge, abilities, and other intergenerationally transmitted characteristics, similar to the loss in genetic diversity, especially in an era when knowledge, culture, among others, were passed orally between generations.[2] Thus, this measure captures population diversity in a general sense and not only applies to genes.

The research establishes the robust positive causal effect of population diversity on the emergence and prevalence 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 state formation or propensity to trade on population compositions.

[2]E.g., phoneme and phenotypic diversity decrease due to this same process (Atkinson, 2011; Manica et al., 2007).

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 or the state 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 (genetic) 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, trade and the state 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. 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 mediating effect of specialization on the emergence and prevalence of the state. 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 the effect of population diversity on the emergence of states is (fully) mediated by its effect on economic specialization. Thus, suggesting that population diversity has no direct effect on the emergence of states, and only an indirect one through its effect on economic specialization. Moreover, this result provides suggestive evidence on the direction of causality between the emergence of trade and statehood.

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 and occupational heterogeneity. Moreover, the persistent positive effect of economic specialization on contemporary development is only partially mediated through pre-modern statehood. 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 statehood and 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) economic specialization on the emergence of states. 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.[3]

Third, this research contributes to the literature on the emergence of states and sheds new light on the old question regarding the relation between states and trade (Fukuyama, 2011;

[3]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).

Hobbes, 1651; Locke, 1690; Rousseau, 1755). While this is not the first research to propose a potential effect of trade on statehood,[4] it is the first to employ measures of pre-modern trade to causally establish this effect.

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, trade, and state centralization. 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, Trade, and Statehood

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 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

[4] See e.g., Fenske (2014) or Litina (2014).

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).

Additionally, this research suggests that the emergence of division of labor and trade facilitated the emergence of states. Indeed, full-time specialists are found in almost every study of early states (Claessen and Skalník, 1978). This pattern was particularly salient in centralized pre-modern societies such as the Aztecs. As documented by the *Matrícula of Huexotzinco*, a great deal of specialization existed among the Aztecs in mid-1500 CE (Prem and Carrasco Pizana, 1974): almost 1600 specialists are classified in different professions such as wood workers, stone cutters, basket makers, hunters, fishermen, and even doctors. Historical records and archaeological evidence provide evidence of well-developed regional exchange and the existence of market places even before the arrival of the Spaniards. When describing the Aztec's Tlatelolco Market, Díaz del Castillo (1796) wrote "All the things which are sold there... are so numerous and of such a different quality and the great market place... was so crowded with people that one would not have been able to see and enquire about it all in two days". Consistent with this narrative, a strand of literature in archeology proposes an adaptationist model of specialization, exchange, and state formation, which proposes the hypothesis 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).

The role of trade on early state formation has been well discussed in the historical, anthropological, political science, and economics literature (Bates, 1983; Fenske, 2014; Service, 1978). Although it is impossible to provide irrefutable empirical evidence on whether it was long-distance or local regional trade that mattered most for the early emergence of statehood, particular emphasis has been placed on the role of long-distance trade as a driver of centralization (Braudel, 1972; Gluckman, 1941; Polanyi, 1957). In particular, many historical narratives link the rise of powerful centralized polities to their engagement to long-distance trade such as in the case of the Songhai Empire in the Western Sahel. Nevertheless, scholars have also pointed at local regional trade as preceding long-distance trade and state emergence. For instance, in their work on Early State Claessen and Skalník (1978) argue:

> *"Local trade within a region is almost by definition closely associated with the founding of the states. Differentiation in society, and the supra-local coordination capabilities of central authorities makes for an increase flow of goods and people throughout the state, specially to and from the capital."*

In this same line of argument, Reid (2002) argues that specialization and local trade were well developed before the Buganda Kingdom started to trade with coastal Arabs in early 19th century; and that this long-distance trade was actually built on an older local trade system, which included a variety of currencies, trade networks, and markets for several commodities such as salt, iron, and bananas. Likewise, Gluckman (1941) argues that regional trade of specialized products such as millet, cassava, wood, and iron was very important within the territory of the Lozi Kingdom before their engagement in long-distance ivory trade. Bisson (1982) presents archeological evidence, which suggests that trade in indigenous products was taking place long before the introduction of foreign products into the trading systems of the Kingdom of Zimbabwe and the Mutapa Empire.

Although the examples provided above illustrate the strong link between division of labor, trade and statehood, 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. 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, state centralization, genetic 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.

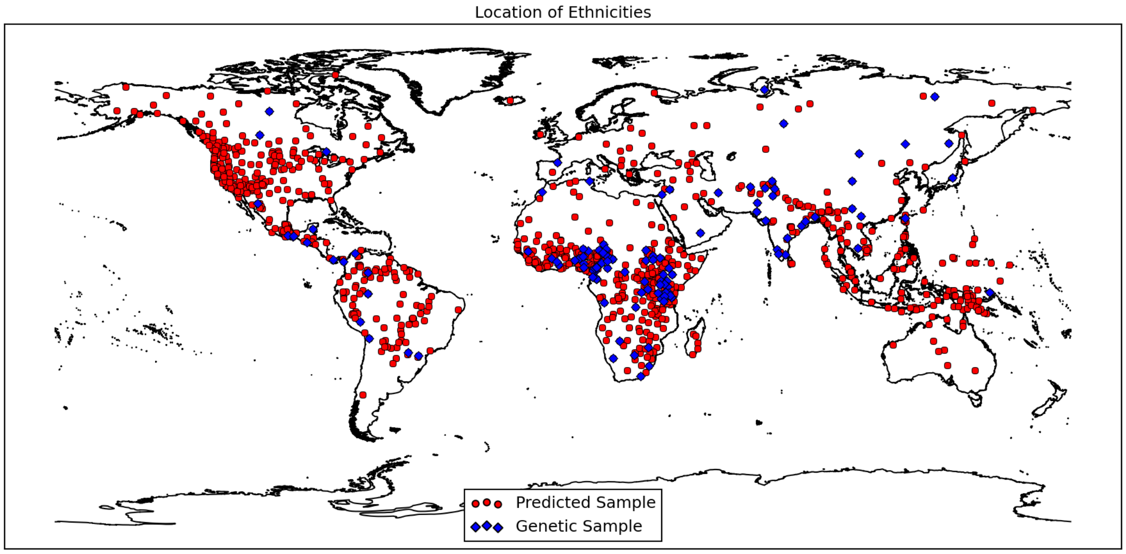


Figure 1: Location of Ethnicities employed in the Analysis (Full and Restricted Samples)

3.1 Dependent Variables: Economic Specialization, Trade and State Centralization

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.[5] Figure 1 depicts the location of the full sample of ethnicities used in the 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, the analysis employs

[5]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.

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), and complexity (v158.1). 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 the lack of the activity. In order to overcome this problem, the analysis uses information on the sexual division of labor for these same activities. In particular, variables v44-v54 assess for the same activities whether the activity was "absent or unimportant", "present", or if there existed gender division of labor in the activity.

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.[6] The main dependent variable in the analysis is the number

[6]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.

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).

In order to analyze the effect of population diversity on the emergence of the state, the analysis uses the measure of "Jurisdictional Hierarchy Beyond Local Community" (v33 in EA, v237 in SCCS), which has been previously used as a measure of the pre-industrial state presence (Fenske, 2014; Gennaioli and Rainer, 2007; Michalopoulos and Papaioannou, 2013). This variable measures levels of statehood discretely from "No Levels" to "Four Levels". For robustness, the analysis additionally shows that main results hold using Fenske's (2014) measure, which takes a value of 0 if there are "No Levels" of jurisdictional hierarchy beyond local community and 1 otherwise, or Gennaioli and Rainer's (2007) measure, which takes a value of 0 if there are "No Levels" or "One Level" and 1 otherwise.

3.2 Independent variables: Genetic Diversity and Predicted Genetic Diversity

This research 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 procedure, 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.

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.[7] In order to ensure comparability across populations, the analysis constrains the

[7]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*

construction of the genetic diversity to the set of 619 common SNP loci for which information exists for all ethnic groups.[8]

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.

3.3 Geographical Controls

An ethnicity's opportunities to trade, as well as its genetic diversity and its level of statehood 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 and B.1-B.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 expected heterozygosity, on economic specialization and trade, as well as the level of state centralization.

religion, language, ethnicity or blood type is better than others.

[8]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 ($p < 0.01$).

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.

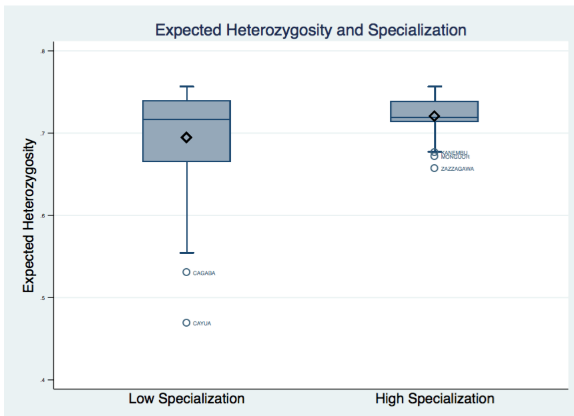


Figure 2: Population Diversity and Economic Specialization

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

$$Specialization_i = \alpha + \beta PD_i + G_i^{\prime}\Gamma + X_i^{\prime}\Delta + \epsilon_i \tag{1}$$

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 $G_i^{\prime}$ denotes a set of basic geographic controls whereas the vector $X_i^{\prime}$ includes a set of additional potential confounders that are discussed below in detail. Finally, ϵ_i is an error term that is allowed to be heteroskedastic. The proposed hypothesis in this paper implies $\beta > 0$.[9]

Table 1 analyzes the association between economic specialization and population diversity accounting for a basic set of geographic characteristics of ethnicities' homelands using OLS.

[9] 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.

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.

Table 1: Population Diversity and Economic Specialization

	Economic Specialization						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Population Diversity	0.27***	0.31***	0.27***	0.27***	0.25***	0.27***	0.36***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.09)
Absolute Latitude		0.15					0.80***
		(0.09)					(0.30)
Area			0.01				0.01
			(0.04)				(0.05)
Elevation (Avg.)				-0.03			0.31*
				(0.11)			(0.16)
Precipitation (Avg.)					-0.08		0.13
					(0.09)		(0.16)
Temperature (Avg.)						0.04	0.73***
						(0.08)	(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.

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 β 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 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.[10] 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. 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 precolonial 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).

[10] 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 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.[11]

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 linguistic diversity. Moreover, given that variation in soil quality can also be conducive to trade. 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).[12] Since population (density) potentially affects 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

[11] 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).

[12] The Caloric Suitability Index (CSI) measures for each cell of 10 kms × 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 available in the pre-1500CE period.

Table 2: Population Diversity and Economic Specialization

	Economic Specialization								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Population Diversity	0.36***	0.41***	0.36***	0.37***	0.34***	0.33***	0.31***	0.40***	0.31***
	(0.09)	(0.10)	(0.10)	(0.09)	(0.11)	(0.10)	(0.09)	(0.11)	(0.10)
Malaria Ecology		-0.36***							-0.41***
		(0.12)							(0.12)
Ecological Diversity			0.26***						0.20*
			(0.10)						(0.11)
Agricultural				0.00					0.13
Suitability (avg.)				(0.13)					(0.10)
Agricultural				0.22*					0.32**
Suitability (std.)				(0.13)					(0.14)
Caloric Suitability					-0.24*				-0.34**
(Pre-1500 ,avg.)					(0.14)				(0.14)
Caloric Suitability					0.30**				0.07
(Pre-1500 ,std.)					(0.11)				(0.14)
Temperature (Spatial						0.01			-0.05
Corr., Avg.)						(0.09)			(0.08)
Temperature						-0.58***			-0.11
(Volatility, Avg)						(0.19)			(0.20)
Pct. Area within							0.02		-0.16
100kms of Sea							(0.10)		(0.11)
Coast Length							0.49**		0.60***
							(0.22)		(0.20)
Ruggedness (Avg.)								-0.22	0.07
								(0.22)	(0.18)
Pre-Industrial								0.81*	1.06**
Mobility (avg.)								(0.41)	(0.46)
Pre-Industrial								-0.04	-0.36**
Mobility (std.)								(0.12)	(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
δ		0.83	1.26	0.89	1.03	1.01	1.02	0.84	1.18
β-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.

(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 δ (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 δ 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 β-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 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 genetically representative of the original population, the sampling process from subsequently smaller populations lead to a loss of genetic diversity, i.e., the serial founder effect. Therefore, the Out-of-Africa theory predicts that genetic diversity decreases along the different migratory routes that humans followed out of East Africa.[13]

Table 3: Population Diversity and Distance to Addis-Ababa

	Population Diversity (Expected Heterozygosity)								
	Full Sample								Specia-lization
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Pre-Industrial	-0.85***	-0.80***	-0.80***	-0.81***	-0.80***	-0.79***	-0.80***	-0.82***	-0.85***
Distance to Addis Ababa	(0.07)	(0.10)	(0.09)	(0.09)	(0.10)	(0.10)	(0.10)	(0.09)	(0.09)
Malaria Ecology			0.17***					0.20***	0.16**
			(0.05)					(0.07)	(0.08)
Agricultural				-0.00				-0.01	-0.01
Suitability (avg.)				(0.06)				(0.07)	(0.07)
Agricultural				0.08				0.13**	0.14*
Suitability (std.)				(0.06)				(0.06)	(0.07)
Caloric Suitability					0.02			0.07	0.09
Index (Pre-1500CE)					(0.07)			(0.07)	(0.09)
Caloric Suitability					-0.08			-0.13**	-0.13*
(Pre-1500 ,std.)					(0.05)			(0.06)	(0.07)
Pct. Area within						-0.00		0.13**	0.14**
100kms of Sea						(0.06)		(0.06)	(0.07)
Coast Length						0.03		0.04	0.01
						(0.06)		(0.05)	(0.07)
Ruggedness (Avg.)							-0.03	-0.19	-0.19
							(0.12)	(0.13)	(0.16)
Pre-Industrial							0.05	0.12	0.13
Mobility (avg.)							(0.20)	(0.23)	(0.25)
Pre-Industrial							-0.13	-0.07	-0.07
Mobility (std.)							(0.08)	(0.11)	(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

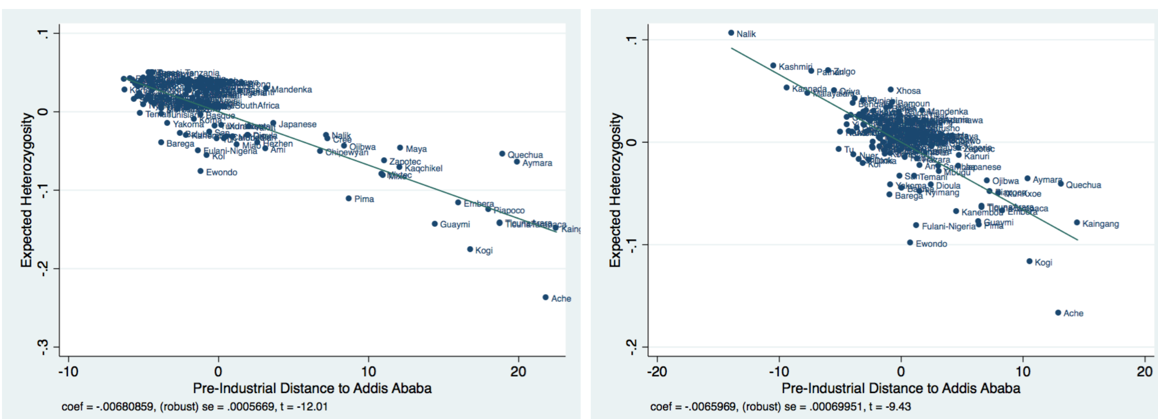
Notes: This table establishes the negative statistically and economically significant relation between expected heterozygosity and the distance to Addis Ababa after accounting for the set of basic geographical controls of Table 1 and an extended set of confounders and measures of isolation. Standardized 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.

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-

[13] This prediction has been supported empirically using data from various population samples (Ashraf and Galor, 2013b; Cann et al., 1987; Pemberton et al., 2013; Ramachandran et al., 2005).

coded genetic and ethnographic data is available.[14] 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.

Two facts stand out from the results in Table 3: (i) migratory distance to Addis Ababa alone explains 72 percent of the variation in 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 3(a) and 3(b) depict respectively the unconditional and conditional strong negative relationship between population diversity and the pre-industrial migratory distance to Addis Ababa.



(a) Unconditional (b) Conditional

Figure 3: 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.[15] In particular, the distance to Addis Ababa has the largest semi-partial R^2 in the analysis, e.g. in column 8 it is 0.3, which is 15 times larger then 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 that 2% of the total variation in genetic diversity.

[14] Similar results are obtained in the full sample of 267 ethnicities for which genetic data alone is available.
[15] Results not shown, but can be obtained from authors.

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

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 it (i) is the main predictor of population diversity, due to the serial founder effect and the Out-of-Africa theory, and (ii) only affects economic outcomes through its effect on diversity.

Table 4 presents the results of this instrumental variables (IV) analysis, in which population 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 the the measure of economic specialization used (see section 3.1). In particular, Table A.4 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.

Table 4: Population Diversity and Economic Specialization (IV)

	Economic Specialization									
	OLS	IV								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Population Diversity	0.36***	0.51***	0.53***	0.54***	0.54***	0.45***	0.49***	0.44***	0.56***	0.46***
	(0.09)	(0.12)	(0.12)	(0.12)	(0.13)	(0.14)	(0.13)	(0.11)	(0.14)	(0.14)
Malaria Ecology			-0.38***							-0.44***
			(0.11)							(0.11)
Ecological Diversity				0.26***						0.19*
				(0.10)						(0.10)
Agricultural					-0.01					0.08
Suitability (avg.)					(0.14)					(0.10)
Agricultural					0.22*					0.31**
Suitability (std.)					(0.12)					(0.12)
Caloric Suitability						-0.21				-0.28**
Index (Pre-1500CE)						(0.14)				(0.13)
Caloric Suitability						0.29***				0.09
(Pre-1500 ,std.)						(0.11)				(0.12)
Temperature (Spatial							0.01			-0.03
Corr., Avg.)							(0.10)			(0.07)
Temperature							-0.53***			-0.03
(Volatility, Avg)							(0.19)			(0.18)
Pct. Area within								0.01		-0.17
100kms of Sea								(0.10)		(0.11)
Coast Length								0.46**		0.57***
								(0.21)		(0.17)
Ruggedness (Avg.)									-0.28	0.02
									(0.22)	(0.18)
Pre-Industrial									0.91**	1.23***
Mobility (avg.)									(0.40)	(0.44)
Pre-Industrial									-0.03	-0.37**
Mobility (std.)									(0.12)	(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.

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 could be biased since it is based on a small non-random sample of ethnicities, for which both genetic and specialization data was available. 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

$$Specialization_i = \alpha + \beta \hat{PD}_i + G_i'\Gamma + X_i'\Delta + \epsilon_i \tag{2}$$

where the only difference with respect to equation (1) is the inclusion of $\hat{PD}_i$, 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).[16]

Based on this extended sample, the analysis replicates in columns 1 to 10 of Table 5 the main econometric specifications of Tables 1, 2, and 4. 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 4).

[16] 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 5: Predicted Population Diversity and Economic Specialization

	Economic Specialization										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Predicted Population	0.44***	0.42***	0.49***	0.42***	0.41***	0.46***	0.40***	0.42***	0.42***	0.53***	0.60***
Diversity	(0.02)	(0.03)	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)	(0.05)	(0.04)	(0.07)	(0.21)
Malaria Ecology			-0.18***							-0.18***	-0.14**
			(0.05)							(0.07)	(0.07)
Ecological Diversity				0.15***						0.10***	0.09***
				(0.04)						(0.04)	(0.05)
Agricultural					-0.05					-0.14***	-0.17***
Suitability (avg.)					(0.05)					(0.07)	(0.08)
Agricultural					0.09**					0.03	-0.02
Suitability (std.)					(0.05)					(0.06)	(0.07)
Caloric Suitability						0.08**				0.09***	0.09***
(Pre-1500 ,avg.)						(0.05)				(0.06)	(0.07)
Caloric Suitability						0.12***				0.12***	0.12***
(Pre-1500 ,std.)						(0.05)				(0.06)	(0.07)
Temperature (Spatial							-0.01			-0.04	-0.08
Corr., Avg.)							(0.05)			(0.06)	(0.08)
Temperature							-0.20***			0.01	0.04
(Volatility, Avg)							(0.10)			(0.15)	(0.14)
Pct. Area within								0.00		-0.14***	-0.11**
100kms of Sea								(0.05)		(0.07)	(0.07)
Coast Length								0.00		0.01	-0.02
								(0.13)		(0.12)	(0.12)
Ruggedness (Avg.)									0.14**	0.16**	0.18**
									(0.11)	(0.16)	(0.18)
Pre-Industrial									0.20**	0.46***	0.20*
Mobility (avg.)									(0.16)	(0.19)	(0.21)
Pre-Industrial									-0.05	-0.22***	-0.13*
Mobility (std.)									(0.09)	(0.12)	(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.

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 specialization by more than a half of a standard deviation.[17]

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 6: Heterogeneous Effects of Predicted Population Diversity on Economic Specialization

	Economic Specialization					
	(1)	(2)	(3)	(4)	(5)	(6)
Predicted Population Diversity	0.42***	0.33***	0.34***	0.13**	0.28***	0.28***
	(0.03)	(0.04)	(0.04)	(0.08)	(0.04)	(0.05)
Predicted Population Diversity		1.07***				
× Ecological Diversity		(0.43)				
Predicted Population Diversity			1.91***			
× Agricultural Suitability (std.)			(0.66)			
Predicted Population Diversity				2.24***		
× Temperature (Volatility, avg.)				(0.74)		
Predicted Population Diversity					2.77***	
× Ruggedness (Avg.)					(0.63)	
Predicted Population Diversity						1.54***
× Pre-Industrial Mobility (std.)						(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 6 analyzes the potential heterogenous effects of population diversity on economic specialization. In particular, it shows the main effect of population diversity and its interac-

[17] Table A.6 shows the point estimates of the reduced form economic specialization-distance to Addis Ababa for all the specifications in Table 5. The point estimates for pre-industrial distance to Addis Ababa are remarkably stable and strongly statistically significant.

tion 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.[18] As can be seen there, all main effects and interactions are positive 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 Pre-industrial Development

This section establishes that the positive causal effect of population diversity on economic specialization is robust to accounting for pre-industrial development. Accounting for other sources of pre-industrial development overcomes the potential concern that population diversity is capturing the effect of factors like the transition to agriculture or the history of settlement 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, which might be the actual source of economic specialization.

Table 7 explores the robustness of the results to accounting for pre-industrial economic development. Column 1 replicates the results of column 7 in Table 1 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 for the emergence

[18]The estimated coefficients are again reported as standardized betas, which simplifies the comparison 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 7: Predicted Population Diversity, Economic Specialization and Pre-Industrial Development

	Economic Specialization					
	Full Sample				Community Size Sample	
	(1)	(2)	(3)	(4)	(5)	(6)
Predicted Population Diversity	0.42***	0.40***	0.35***	0.43***	0.40***	0.26***
	(0.03)	(0.08)	(0.03)	(0.03)	(0.03)	(0.05)
Origtime		0.02				
		(0.10)				
Years Since Neolithic Revolution			0.30***			
			(0.04)			
Population Density (1500CE)				0.12***		
				(0.06)		
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes
Community Size FE	No	No	No	No	No	Yes
Adjusted-R^2	0.21	0.21	0.29	0.23	0.21	0.46
Observations	934	925	924	910	509	509

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, statehood 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.

of social stratification or a dominant elite.[19] 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.

Column 3 analyzes the potential confounding effect of the long-lasting influence of the Neolithic Revolution. As argued by Diamond (1997), an earlier transition from hunting and gathering practices to agriculture 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

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

(Ashraf and Galor, 2011). In line with these findings, column 3 shows that the time elapsed since the Neolithic Revolution positively affects economic specialization. Reassuringly, the estimated effect of predicted population diversity on economic specialization remains positive strongly statistically and economically significant thus suggesting that the omission of the Neolithic transition-timing was not spuriously driving the main results.[20]

Column 4 analyzes the potential confounding effect of economic development during the Malthusian era as measured by population density in 1500CE (Klein Goldewijk et al., 2011). This overcomes the potential concern that the causal effect of population diversity on economic specialization reflects its effect on development and that it is higher development which generates higher levels of specialization. Reassuringly, accounting for population density does not affect the estimated effect of population diversity on economic specialization. On the other hand, the results reveal that population density is positively correlated with economic specialization, which could be explained by the positive effect that market size has on economic specialization or the positive effect that economic specialization has on economic development. Interestingly, while economic specialization and population density are positively correlated, population diversity has no effect on population density in this sample.

Columns 5 and 6 further explore the potential confounding effect of population density on the results. In particular, the analysis accounts for the effect of the mean size of the local community (v31 from the Ethnographic Atlas), which is coded in the order of increasing settlement size with values ranging from 1 (fewer than 50 people) to 8 (cities of more than 50,000 people). Column 5 shows that the previously documented effect of predicted population diversity on economic specialization still holds in the sample of 509 ethnic groups for which data on mean size of the local communities is available. Column 6 establishes that the causal effect of predicted population diversity on economic specialization remains positive statistically and economically significant after accounting for the mean size of the local community via the inclusion of seven dummies.[21] While the decrease in the estimated effect could be capturing the direct positive effect of population diversity on the mean size of the local community, it could also be explained by the positive effect of economic specialization on development and the mean size of the local community.

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

[20]Alternatively, accounting for the degree of dependence on agriculture does not alter the results either.

[21]Results are similar if instead of dummies, the order variable for local community sizes is included in column 6.

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, the existence of writing and records, the degree of technological specialization, and a measure of sociocultural complexity. Reassuringly, these measures of trade correlate strongly and positively with the measure of economic specialization (see Table A.3). Table 8 presents the point estimates for the regression specifications given by:

$$Y_i = \alpha + \beta \hat{PD}_i + G_i'\Gamma + X_i'\Delta + \epsilon_i, \tag{3}$$

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 8 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.

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).

Finally, columns 7 and 8 establish the positive effect of population diversity on the degree of technological specialization and sociocultural complexity. The degree of technological so-

Table 8: Predicted Population Diversity, Economic Specialization and Pre-Industrial Development in the SCCS

	Pre-Industrial Measures of Trade							
	Economic Special-ization	Impor-tance of Trade	Intercom-munity Trade as Food Source	Money	Credit	Writing and Records	Techno-logical Special-ization	Com-plexity
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Predicted Population Diversity	0.37***	0.25***	0.42***	0.44***	0.19***	0.19***	0.31***	0.38***
	(0.06)	(0.07)	(0.08)	(0.08)	(0.07)	(0.06)	(0.06)	(0.07)
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additonal Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.32	0.04	0.19	0.23	0.19	0.26	0.29	0.31
Observations	168	168	165	165	153	168	168	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.

phistication is an ordinal measure that takes the values 1 (none), 2 (pottery only), 3 (loom weaving only), 4 (metalwork only), and 5 (smiths, weavers, potters) and can be seen as an additional measure of specialization. On the other hand, the measure of sociocultural complexity measures the economic and socio-political complexity of an ethnic group by combining information on its population density, fixity of residence, urbanization, dependence on agriculture, political integration, and social stratification among others. The estimated effects are agin economically and statistically significant and imply that a one standard deviation increase in population diversity increases technological specialization by 0.3 and sociocultural complexity by 0.4 standard deviations respectively.

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

4.8 Population Diversity and State Centralization

This section explores the potential effect of population diversity on the emergence of states in the pre-modern era. In particular, it analyzes the reduced form relation between population diversity and statehood levels in pre-modern societies.[22] Additionally, it establishes that economic specialization mediates the causal effect of population diversity on the emergence of pre-modern states.

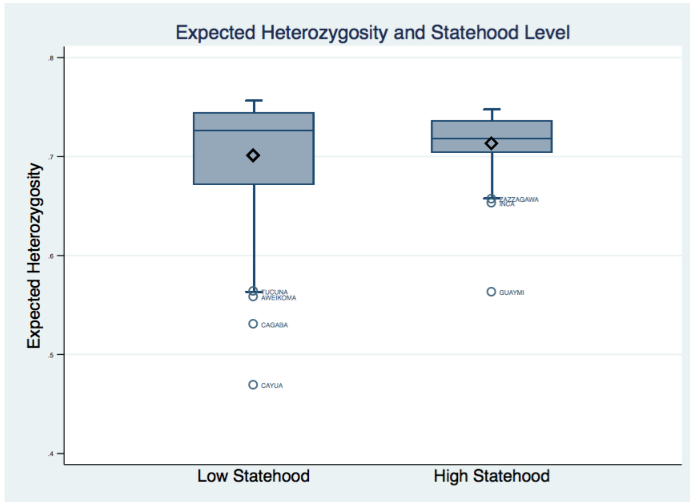


Figure 4: Population Diversity and Statehood

Figure 4 depicts the distribution of population diversity for groups above and below the mean level of Statehood. Clearly, the figure supports the hypothesis that groups with highly centralized states also have higher population diversity. Table 9 analyzes the effect of population diversity on statehood more formally. In order to better establish the effect of population diversity on statehood and the potential mediating effect of trade, as well as to account for other confounding factors in the emergence of the state, the specifications shown in Table 9 are similar to the ones previously employed to uncover the effect of population diversity on economic specialization (Tables 1 and 5).

In particular, column 1 of Table 9 reports the unconditional statistical relationship between predicted population diversity and statehood level. The point estimate is highly statistically significant and economically large: a one standard deviation increase in predicted population diversity increases statehood by two-fifths of a standard deviation. Column 2 accounts for the baseline set of geographic controls discussed in Table 1 with qualitatively similar results. Columns 3 to 9 account for additional geographical characteristics of ethnic homelands that might confound the effect of population diversity on statehood. Again, the point estimates remain remarkably stable across specifications. Moreover, accounting for the full set of controls (column 10) and for continental fixed effects (column 11) does not alter

[22]Statehood levels are measured on a scale from "No Levels" (0) to "Four Levels" (4) as measured by the "Jurisdictional Hierarchy Beyond Local Community" in the Ethnographic Atlas (Murdock, 1967).

Table 9: Predicted Population Diversity and State Centralization

	Statehood Level										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Predicted Population	0.39***	0.41***	0.49***	0.40***	0.41***	0.47***	0.40***	0.41***	0.41***	0.60***	0.44***
Diversity	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.04)	(0.04)	(0.05)	(0.04)	(0.09)	(0.18)
Malaria Ecology			-0.25***							-0.28***	-0.26***
			(0.05)							(0.07)	(0.06)
Ecological Diversity				0.14***						0.08**	0.06**
				(0.04)						(0.04)	(0.04)
Agricultural					0.07*					-0.02	-0.01
Suitability (avg.)					(0.05)					(0.08)	(0.06)
Agricultural					0.08*					0.05	0.05
Suitability (std.)					(0.05)					(0.06)	(0.06)
Caloric Suitability						0.19***				0.15***	0.15***
(Pre-1500 ,avg.)						(0.05)				(0.07)	(0.06)
Caloric Suitability						0.08**				0.09**	0.05
(Pre-1500 ,std.)						(0.04)				(0.05)	(0.06)
Temperature (Spatial							-0.05			-0.10**	-0.09
Corr., Avg.)							(0.06)			(0.08)	(0.07)
Temperature							-0.01			0.10	0.13
(Volatility, Avg)							(0.10)			(0.17)	(0.13)
Pct. Area within								-0.00		-0.13***	-0.02
100kms of Sea								(0.05)		(0.07)	(0.07)
Coast Length								-0.02		-0.00	-0.02
								(0.12)		(0.14)	(0.11)
Ruggedness (Avg.)									-0.01	0.07	0.04
									(0.11)	(0.19)	(0.14)
Pre-Industrial									0.09	0.35***	0.10
Mobility (avg.)									(0.16)	(0.21)	(0.19)
Pre-Industrial									-0.00	-0.22***	-0.10
Mobility (std.)									(0.09)	(0.14)	(0.11)
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	0.15	0.23	0.26	0.25	0.23	0.27	0.23	0.23	0.23	0.31	0.36
Observations	912	912	912	912	912	912	912	912	912	912	912

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 statehood centralization. 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.

the qualitative results.[23] Furthermore, Table 9 provides evidence on the effect of other de-

[23]Appendix B.2 establishes the robustness of these results to (i) the measure of statehood, (ii) the estimation method, and (iii) the sample employed. First, the results are robust to two additional definitions of state centralization: (a) the measure proposed by Fenske (2014), which takes a value of 0 if there are "No Levels" of jurisdictional hierarchy beyond local community and 1 otherwise; and (b) the measure proposed by Gennaioli and Rainer (2007), which takes a value of 0 if there are "No Levels" or "One Level" and 1

terminants of statehood highlighted previously in the literature. In particular, higher levels of ecological diversity (column 4), agricultural suitability (column 5), and caloric suitability (column 6) positively predict the existence of state centralization. On the other hand, a more pervasive disease environment substantially reduces the likelihood of ethnic groups having a centralized state (column 3).

While the results presented in Table 9 lend credence to the proposed hypothesis that higher levels of population diversity were conducive to economic specialization, which incentivized the emergence of trade-facilitating institutions and of states in pre-modern societies, they cannot reject an alternative hypothesis in which population diversity facilitates the emergence of states (through some non-trade related channel), which in turn would promote trade and specialization. In order to tackle the question of order of causality of the effect of population diversity on economic specialization and statehood, this research follows two strategies: (i) analyze the heterogeneous effects of diversity on statehood and (ii) uncover the direct and indirect effects of population diversity on statehood using an additional instrumental variable strategy.

Given the potential complementarities between a diverse population and a diverse environment, the theory predicts a heterogeneous effect of population diversity on trade and, thus, on statehood. In particular, diverse populations should be more able to take advantage of diverse environments, and thus the effect of population diversity on trade and statehood should be higher among ethnicities that are more diverse and live in more diverse environments. Table 10 establishes the positive effect of the interaction between population diversity and various measures of diversity of the natural environment on the level of statehood, accounting for an ethnic homeland's geographical characteristics. In particular, the positive effect of population diversity increases when an ethnic homeland has larger ecological diversity (column 2) or higher variation in agricultural suitability (column 3), temperature (column 4), ruggedness (column 5) or pre-industrial mobility (column 6).[24] These heterogeneous effects of population diversity on statehood are qualitatively similar to the heterogeneous effects of diversity on economic specialization (Table 6) and provide additional support to the theory. Moreover, the results imply that alternative theories of the effect of population diversity on statehood levels need to explain how non-trade related channels generate these heterogeneous effects.

The previous results are based on the reduced form relation between population diversity and both economic specialization and statehood. The rest of this section explores the struc-

otherwise (Table B.3). Second, the results remain qualitatively unchanged when an ordered probit model is estimated (Tables B.4 and B.5). Finally, the results are again qualitatively similar, both in the OLS and IV cases, when restricting the sample to 142 ethnicities for which observed genetic data exists (Table B.6).

[24]Similar effects for higher variation in caloric suitability and precipitation.

Table 10: Heterogeneous Effects of Predicted Population Diversity on State Centralization

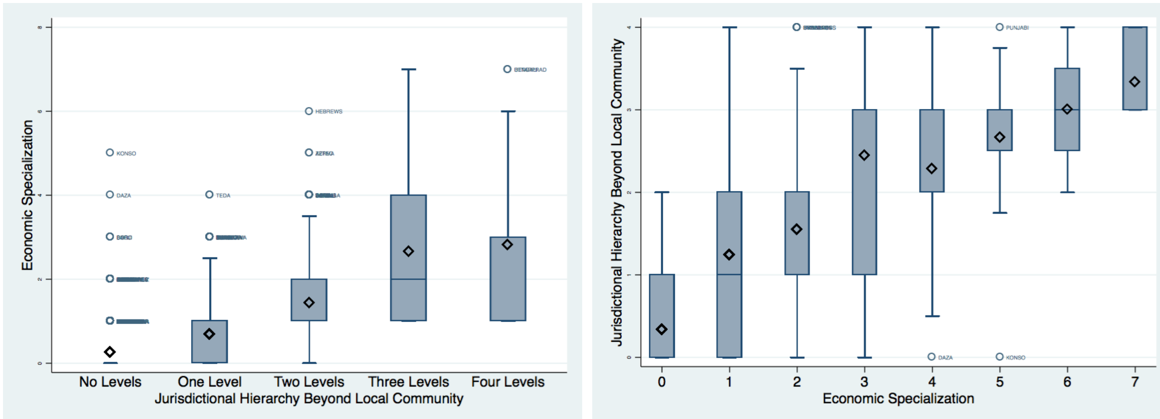
	Statehood Level					
	(1)	(2)	(3)	(4)	(5)	(6)
Predicted Population Diversity	0.41***	0.35***	0.33***	0.27***	0.28***	0.26***
	(0.03)	(0.04)	(0.04)	(0.10)	(0.04)	(0.05)
Predicted Population Diversity		0.65*				
× Ecological Diversity		(0.43)				
Predicted Population Diversity			1.98***			
× Agricultural Suitability (std.)			(0.81)			
Predicted Population Diversity				1.10**		
× Temperature (Volatility, Avg)				(0.84)		
Predicted Population Diversity					2.62***	
× Ruggedness (Avg.)					(0.62)	
Predicted Population Diversity						1.71***
× Pre-Industrial Mobility (std.)						(0.57)
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.23	0.25	0.25	0.23	0.25	0.24
Observations	912	912	912	912	912	912

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 statehood centralization. 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.

tural relation between these variables. As a first step, Figure 5(a) depicts the distribution of economic specialization for each level of statehood, while Figure 5(b) depicts the distribution of statehood for each level of economic specialization. The figures suggest a positive relation between the level of specialization and statehood. Nevertheless, increases in economic specialization seem to be more strongly correlated with higher levels of statehood.

Table 11 explores the causal relation between economic specialization and statehood further. It establishes the positive effect of population diversity (column 1) and economic specialization (column 2) on the level of statehood after accounting for all the geographical controls and continental fixed effects. In particular, the OLS estimates suggest that a one standard deviation increase in population diversity or economic specialization would increase statehood by about 0.4 standard deviations. Clearly, given the potential reverse causality of statehood on economic specialization, the estimate in column 2 might be biased.

If, as proposed in this paper, population diversity affects economic specialization, and has no direct effect on the level of statehood, then it would be a valid instrument for economic specialization. Column 3 shows the estimated effect of economic specialization on statehood



(a) Distribution of Economic Specialization for each State Centralization level

(b) Distribution of State Centralization for each Economic Specialization level

Figure 5: Economic Specialization and State Centralization

under this hypothesis. The estimated effect of economic specialization on statehood increases by 50%. Thus, under this (identification) hypothesis, a one standard deviation increase in economic specialization would increase the level of statehood by 0.6 standard deviations.

Column 4 provides supporting evidence to the hypothesis that population diversity affects statehood levels *only* through its effect on economic specialization. In particular, in an OLS horse race between economic specialization and population diversity, only economic specialization remains statistically and economically significant. Moreover, compared to column 1, the estimated coefficient on population diversity falls by almost 70% (without changes in the standard error), suggesting economic specialization *is* the channel through which population diversity affects statehood. Figure 6 shows the union of confidence intervals of the estimated effect of economic specialization on statehood when the exogeneity assumption on population diversity is violated (Conley et al., 2012). The red line is the estimated value of the direct effect of population diversity on statehood in column 4. The figure shows that even if the direct effect of population diversity was twice as large as in column 4, the causal effect of economic specialization on statehood would remain positive statistically and economically significant.

Although it is reassuring that economic specialization has a positive effect on statehood, the OLS results cannot be used to reject a possible direct causal effect of population diversity on statehood. Moreover, it is difficult to find an instrument based on theoretical arguments, which affects economic specialization *without* having a potential direct effect on statehood. 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

Table 11: Predicted Population Diversity, Economic Specialization and State Centralization

	Statehood Level				
	OLS	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)
Predicted Population Diversity	0.42***			0.15	0.15
	(0.14)			(0.12)	(0.12)
Economic Specialization		0.43***	0.66***	0.43***	0.42***
		(0.03)	(0.19)	(0.03)	(0.04)
All Controls	Yes	Yes	Yes	Yes	Yes
Continental FE	Yes	Yes	Yes	Yes	Yes
Breusch-Pagan F-stat					51.23
Breusch-Pagan p-value					0.00
First-stage F-statistic			15.25		54.85
Hansen's J-statistic					28.17
Hansen's J p-value					0.17
Adjusted-R^2	0.36	0.48	0.45	0.48	0.48
Observations	912	912	912	912	912

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 statehood is fully mediated by its effect on economic specialization. Moreover, it establishes that economic specialization causes statehood. These results account for the full set of geographical controls in Table 5 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.

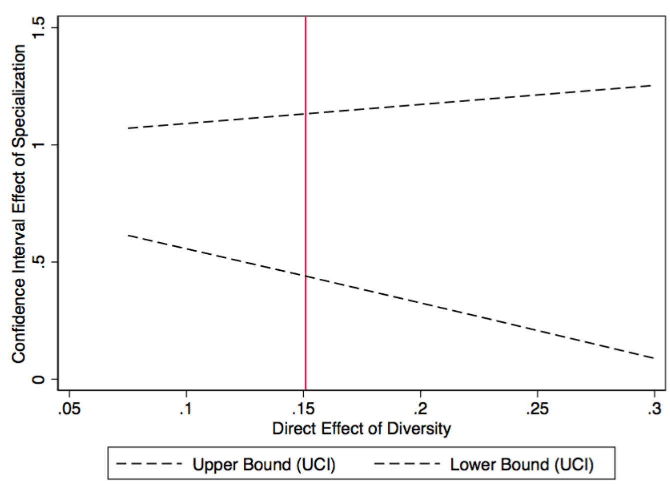


Figure 6: Estimated Effect of Economic Specialization on State
(Union of Confidence Intervals under Plausible Exogeneity of Population Diversity)

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), Column 5 in Table 11 establishes that population diversity has no direct effect on statehood, i.e., population diversity *only* affects statehood through its effect on economic specialization. Moreover, the Breusch-Pagan test suggests the presence of heterokedasticity, ensuring the conditions for identification are satisfied. Furthermore, Hansen's J test for overidentification restrictions suggests that the instruments are valid. Reassuringly, using different measures of statehood does not alter the result that population diversity has only an indirect causal effect on statehood via its direct effect on economic specialization (Table 12).

Table 12: Predicted Population Diversity, Economic Specialization and State Centralization

	Statehood Level		Any State		Centralized State	
	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Predicted Population Diversity	0.15	0.15	0.03	0.04	0.15	0.16
	(0.12)	(0.12)	(0.15)	(0.15)	(0.13)	(0.13)
Economic Specialization	0.43***	0.42***	0.23***	0.21***	0.39***	0.37***
	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)
Continental FE	Yes	Yes	Yes	Yes	Yes	Yes
All Controls	Yes	Yes	Yes	Yes	Yes	Yes
First-stage F-statistic		54.85		54.85		54.85
Hansen's J-statistic		28.17		44.66		20.25
J-stat p-value		0.17		0.00		0.57
Adjusted-R^2	0.48	0.48	0.33	0.33	0.38	0.38
Observations	912	912	912	912	912	912

Notes: This table establishes the robustness of the results in Table 11 to the measure of statehood. Thus, it 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 statehood is fully mediated by its effect on economic specialization. Moreover, it establishes that economic specialization causes statehood. These results account for the full set of geographical controls in Table 5 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.

These results suggest that population diversity and economic specialization played an essential role in the emergence of the state. Moreover, they support the view that the direction of causality runs from economic specialization and trade to state formation. Additionally, the results provide support to the view that economic development and inequality were essential to the development of the state, supporting both the functionalist and institutionalist theories of the emergence of the state (Boix, 2015).

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. Thus, if these institutions persist across time, it is conceivable that pre-modern economic specialization might have a persistent effect on economic development. Moreover, pre-modern economic specialization may affect contemporary development if it generates learning by doing processes or the emergence of certain cultural traits.

Table 13 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 in a horserace with pre-modern levels of state centralization, that both pre-modern economic specialization and statehood have a positive correlation with contemporary economic development. In line with the results of section 4.8, the inclusion of pre-modern statehood levels decreases the size and significance of the estimated effect of pre-modern economic specialization on contemporary economic development. This suggests that the effect of pre-modern economic specialization on contemporary economic development is at least partly mediated through it positive effect on the emergence and size of states.

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, even after accounting for the mediating effect of statehood. Thus, the Old World sample suggests that a one-standard deviation increase in pre-modern economic development generates 0.2 standard deviations increase in

Table 13: Pre-colonial Economic Specialization, State Centralization and Contemporary Economic Development

	Log(Average Light Intensity + 0.01)							
	Whole World				Old World		Africa	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Economic Specialization	0.08***	0.20***	0.16***	0.05*	0.21***	0.09**	0.11**	0.09*
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.05)	(0.05)
Statehood Level				0.25***		0.25***		0.05
				(0.03)		(0.04)		(0.05)
Main Controls	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Continental FE	No	Yes	Yes	Yes	Yes	Yes	No	No
Adjusted-R^2	0.26	0.35	0.39	0.42	0.32	0.36	0.23	0.23
Observations	912	912	912	912	577	577	329	329

Notes: This table establishes the persistent positive statistically and economically significant effect of pre-modern economic specialization on economic development. Additionally, it establishes that part of the effect of economic specialization works through its effect on pre-modern statehood levels. These results account for the main set of geographical controls in Table 5 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.

log-light intensity, of which about 50% works through pre-modern statehood levels. This result is in line with the view that traditional structures survived the colonization period and still matter today (Depetris-Chauvin, 2014; Gennaioli and Rainer, 2007; Hariri, 2012; Jha, 2013; Michalopoulos and Papaioannou, 2013). Columns 7 and 8 show that similar results are obtained when focusing exclusively on Africa. Nonetheless, the association between the light intensity and state centralization is not significant in the horse race with pre-modern economic specialization.

Clearly, the positive correlation between pre-modern economic specialization, statehood 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 and for regional fixed effects. Columns 1 and 5 establish that even after accounting for this expanded set of confounders, pre-modern economic specialization has a positive statistically and economically significant effect on contemporary development. Columns 2 and 6 provide additional evidence that the effect of economic specialization on development works through its effect on statehood levels. Additionally, columns 3 and 7 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). 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. Similar results are obtained when instrumenting both economic specialization and statehood. Although these results cannot fully determine its causal nature, they do suggest that pre-modern economic specialization has a positive effect on contemporary development. Moreover, economic specialization seems to have both a direct effect on development and an indirect one through its effect on pre-modern states.

Table 14: Pre-colonial Economic Specialization, State Centralization 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**	0.02	0.10***	0.02	0.09**	0.04*	0.11**	0.05
	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.04)	(0.04)
Statehood Level		0.11**		0.14***		0.12*		0.11**
		(0.05)		(0.04)		(0.07)		(0.05)
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
First-stage F-statistic			70.93	27.31			35.90	16.07
Hansen's J-statistic			51.62	94.44			45.42	84.01
J-stat p-value			0.04	0.04			0.07	0.07
Adjusted-R^2	0.24	0.25	0.22	0.23	0.22	0.23	0.20	0.21
Observations	910	910	910	910	575	575	575	575

Notes: This table establishes the persistent positive statistically and economically significant effect of pre-modern economic specialization on economic development. Additionally, it establishes that part of the effect of economic specialization works through its effect on pre-modern statehood levels. These results account for the full set of geographical controls in Table 5 and regional 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.

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

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

	Contemporary Occupational Heterogeneity						
	OLS					IV	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Economic Specialization	0.17**	0.17**	0.18**	0.16***	0.16***	0.15*	0.16*
	(0.07)	(0.08)	(0.08)	(0.03)	(0.03)	(0.09)	(0.09)
Statehood Level					-0.01		0.03
					(0.02)		(0.10)
Main Controls	No	Yes	Yes	Yes	Yes	Yes	Yes
All Additional Controls	No	No	Yes	Yes	Yes	Yes	Yes
Regional FE	No	No	No	Yes	Yes	Yes	Yes
First-stage F-statistic						126.58	139.08
Hansen's J-statistic						15.65	32.62
J-stat p-value						0.79	0.85
Adjusted-R^2	0.02	0.03	-0.01	0.02	0.01	-0.05	-0.06
Observations	187	187	187	187	187	187	187

Notes: This table establishes the persistent positive statistically and economically significant effect of pre-modern economic specialization on contemporary occupational heterogeneity. Additionally, it establishes pre-modern statehood levels do not affect contemporary occupational heterogeneity. These results account for the full set of geographical controls in Table 5 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.

Afro-barometer. Columns 1-4 in Table 15 establish that pre-modern economic specialization has a positive statistically and economically significant effect on contemporary occupational heterogeneity without and with controls, including all geographical controls and regional fixed effects. Additionally, column 5 establishes that in a horse race with pre-modern statehood levels, only pre-modern economic specialization is significantly associated with contemporary occupational heterogeneity. Finally, columns 6 and 7 follow Lewbel (2012) and instrument pre-modern economic specialization and statehood levels without affecting the qualitative results. Moreover, the F-statistic for the fist stage suggests that the instruments are strongly correlated with specialization and statehood, while Hansen's over-identification tests suggests that the instruments are valid. Although the statistical significance is lower in the IV analysis, the coefficients across all specifications are basically identical, suggesting that pre-modern economic specialization has a positive statistically and economically significant effect on contemporary occupational heterogeneity. In particular, a one-standard deviation increase in pre-colonial economic specialization increases contemporary occupational heterogeneity by about 0.2 standard deviations. Additionally, the results suggest that this effect is not mediated by statehood levels. Given the positive correlation between contemporary occupational heterogeneity and economic development, this result suggest a novel channel through which pre-modern economic specialization affects comparative development.

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 statehood and 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) economic specialization on the emergence of states. 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 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, it finds that this persistence is only partly due to the effect of specialization on pre-modern states. On the other hand, 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. This is 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.

Bisson, M. S. (1982). Trade and tribute. archaeological evidence for the origin of states in south central africa (commerce et tribut. documents archéologiques sur l'origine des états du sud de l'afrique centrale), *Cahiers d'études africaines* pp. 343–361.

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

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. G., Hansen, C. B. and Rossi, P. E. (2012). Plausibly exogenous, *Review of Economics and Statistics* **94**(1): 260–272.

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.

Depetris-Chauvin, E. (2014). State history and contemporary conflict: Evidence from sub-saharan africa, *Technical report*, Brown University.

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.

Díaz del Castillo, B. (1796). *Historia verdadera de la conquista de la Nueva España*, Imprenta de Don Benito Cano, Madrid.

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.

Fukuyama, F. (2011). *The origins of political order: from prehuman times to the French Revolution*, 1st ed edn, Farrar, Straus and Giroux, New York.

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.

Hariri, J. G. (2012). The autocratic legacy of early statehood, *American Political Science Review* **106**(3): 471–494.

Henderson, J. V., Storeygard, A. and Weil, D. N. (2012). Measuring economic growth from outer space, *The American Economic Review* **102**(2): 994–1028.

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.

Hobbes, T. (1651). *Leviathan, or, The matter, forme, & power of a common-wealth ecclesiasticall and civill*, Printed for Andrew Ckooke , at the Green Dragon in St. Pauls Church-yard, London.

Jha, S. (2013). Trade, institutions, and ethnic tolerance: Evidence from south asia, *American Political Science Review* **107**(04): 806–832.

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

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

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models, *Journal of Business & Economic Statistics* **30**(1): 67–80.

Litina, A. (2014). The geographical origins of early state formation, *Technical report*.

Locke, J. (1690). *Two treatises of government: in the former, the false principles, and foundation of Sir Robert Filmer, and his followers, are detected and overthrown. The latter is an essay concerning the true original, extent, and the end of civil government*, A. Churchill, London.

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 ed edn, McGraw Hill College, New York.
URL: *http://www.loc.gov/catdir/description/mh023/98018510.html*

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.

Prem, H. J. and Carrasco Pizana, P. (1974). *Matrícula de Huexotzinco: ms. mex. 387 der Bibliothèque Nationale Paris : Edition, Kommentar, Hieroglyphenglossar*, Akademische Druck- u. Verlagsanstalt, Graz, Austria.

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.

Reid, R. J. (2002). *Political power in pre-colonial Buganda: economy, society & welfare in the nineteenth century*, James Currey, Oxford.

Rousseau, J.-J. (1755). *Discours sur l'origine et les fondemens de l'inegalité parmi les hommes*, Marc Michel Rey, A Amsterdam.

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 Antrhopology of Political Evolution*, ISHI, Philadelphia, PA, pp. 21–34.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*, Whitestone, 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

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

	Economic Specialization and Trade Measures									
	Economic Specialization	Economic Specialization (Share)	Economic Specialization (Dev)	Importance of Trade	Intercommunity Trade as Food Source	Money	Credit Source	Writing and Records	Technological Specialization	Complexity
Economic Specialization	1.00									
Economic Specialization (Share)	0.94***	1.00								
Economic Specialization (Dev)	0.87***	0.74***	1.00							
Importance of Trade	0.20***	0.19**	0.13*	1.00						
Intercommunity Trade as Food 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 Records	0.55***	0.55***	0.47***	0.24***	0.31***	0.38***	0.38***	1.00		
Technological Specialization	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.2.1 Robustness to Measure of Economic Specialization

Table A.4: 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***	0.46***	0.33***	0.37***	0.13**	0.31**
	(0.05)	(0.14)	(0.05)	(0.12)	(0.06)	(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.2.2 Robustness to Continental Fixed Effects

Table A.5: 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***	0.57***	0.58***	0.51***	0.51***	0.53***	0.59***	0.55***	0.57***	0.58***
	(0.15)	(0.15)	(0.15)	(0.15)	(0.16)	(0.16)	(0.15)	(0.15)	(0.15)	(0.17)
Malaria Ecology			-0.13***							-0.09*
			(0.05)							(0.05)
Ecological Diversity				0.12***						0.09***
				(0.03)						(0.03)
Agricultural					-0.07*					-0.16***
Suitability (avg.)					(0.04)					(0.04)
Agricultural					0.04					0.01
Suitability (std.)					(0.05)					(0.05)
Caloric Suitability						0.06*				0.11***
Index (Pre-1500CE)						(0.03)				(0.03)
Caloric Suitability						0.08*				0.07
(Pre-1500 ,std.)						(0.04)				(0.05)
Temperature (Spatial							0.01			0.03
Corr., Avg.)							(0.03)			(0.04)
Temperature							-0.25***			-0.14*
(Volatility, Avg)							(0.06)			(0.08)
Pct. Area within								0.05		-0.02
100kms of Sea								(0.03)		(0.04)
Coast Length								0.01		0.01
								(0.03)		(0.03)
Ruggedness (Avg.)									0.10	0.08
									(0.06)	(0.07)
Pre-Industrial									0.15	0.28**
Mobility (avg.)									(0.10)	(0.11)
Pre-Industrial									-0.04	-0.14*
Mobility (std.)									(0.08)	(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.2.3 Reduced Form Analysis: Distance to Addis Ababa and Economic Specialization

Table A.6: Distance to Addis Ababa and Economic Specialization

	Economic Specialization										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Pre-Industrial Dist.	-0.45***	-0.42***	-0.45***	-0.42***	-0.41***	-0.45***	-0.40***	-0.42***	-0.42***	-0.47***	-0.36***
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)
Malaria Ecology			-0.10**							-0.07	-0.06
			(0.04)							(0.05)	(0.05)
Ecological Diversity				0.15***						0.10***	0.09***
				(0.03)						(0.03)	(0.03)
Agricultural					-0.05					-0.17***	-0.14***
Suitability (avg.)					(0.04)					(0.04)	(0.04)
Agricultural					0.13***					0.09**	0.07
Suitability (std.)					(0.04)					(0.04)	(0.04)
Caloric Suitability						0.09***				0.15***	0.11***
Index (Pre-1500CE)						(0.03)				(0.03)	(0.03)
Caloric Suitability						0.08*				0.06	0.07
(Pre-1500 ,std.)						(0.05)				(0.05)	(0.05)
Temperature (Spatial							0.02			0.02	0.03
Corr., Avg.)							(0.02)			(0.04)	(0.04)
Temperature							-0.32***			-0.23***	-0.14*
(Volatility, Avg)							(0.06)			(0.07)	(0.08)
Pct. Area within								0.02		-0.05	0.01
100kms of Sea								(0.03)		(0.04)	(0.04)
Coast Length								0.04		0.04	0.04
								(0.04)		(0.04)	(0.03)
Ruggedness (Avg.)									0.12*	0.05	0.05
									(0.07)	(0.07)	(0.07)
Pre-Industrial									0.27***	0.41***	0.28**
Mobility (avg.)									(0.10)	(0.11)	(0.11)
Pre-Industrial									-0.12	-0.21***	-0.14*
Mobility (std.)									(0.08)	(0.08)	(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.7: Distance to Addis Ababa and Economic Specialization

	Economic Specialization									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Pre-Industrial Distance	-0.48***	-0.37***	-0.33***	-0.35***	-0.36***	-0.36***	-0.40***	-0.39***	-0.36***	-0.36***
to Addis Ababa	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.11)	(0.10)	(0.10)	(0.11)	(0.10)
Malaria Ecology			-0.09*							-0.06
			(0.05)							(0.05)
Ecological Diversity				0.13***						0.09***
				(0.03)						(0.03)
Agricultural					-0.05					-0.14***
Suitability (avg.)					(0.04)					(0.04)
Agricultural					0.10**					0.07
Suitability (std.)					(0.04)					(0.04)
Caloric Suitability						0.05*				0.11***
Index (Pre-1500CE)						(0.03)				(0.03)
Caloric Suitability						0.09**				0.07
(Pre-1500 ,std.)						(0.04)				(0.05)
Temperature (Spatial							0.00			0.03
Corr., Avg.)							(0.03)			(0.04)
Temperature							-0.25***			-0.14*
(Volatility, Avg)							(0.06)			(0.08)
Pct. Area within								0.06**		0.01
100kms of Sea								(0.03)		(0.04)
Coast Length								0.04		0.04
								(0.03)		(0.03)
Ruggedness (Avg.)									0.10	0.05
									(0.06)	(0.07)
Pre-Industrial									0.16	0.28**
Mobility (avg.)									(0.11)	(0.11)
Pre-Industrial									-0.06	-0.14*
Mobility (std.)									(0.08)	(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.8: Heterogeneous Effects of Distance to Addis Ababa on Economic Specialization

	Economic Specialization					
	(1)	(2)	(3)	(4)	(5)	(6)
Pre-Industrial Distance	-0.07***	-0.05***	-0.06***	-0.02**	-0.05***	-0.05***
to Addis Ababa	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)
Ecological Diversity		1.55***				
		(0.32)				
Pre-Industrial Distance to Addis Ababa		-0.07***				
× Ecological Diversity		(0.02)				
Agricultural			3.84***			
Suitability (std.)			(0.91)			
Pre-Industrial Distance to Addis Ababa			-0.24***			
× Agricultural Suitability (std.)			(0.07)			
Temperature				-0.01		
(Volatility, Avg)				(0.20)		
Pre-Industrial Distance to Addis Ababa				-0.05***		
× Temperature (Volatility, Avg)				(0.01)		
Ruggedness (Avg.)					0.00***	
					(0.00)	
Pre-Industrial Distance to Addis Ababa					-0.00***	
× Ruggedness (Avg.)					(0.00)	
Pre-Industrial						6.93***
Mobility (std.)						(1.99)
Pre-Industrial Distance to Addis Ababa						-0.44***
× Pre-Industrial Mobility (std.)						(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.9: 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***	-0.44***	-0.35***	-0.43***	-0.41***	-0.28***
	(0.03)	(0.06)	(0.02)	(0.03)	(0.03)	(0.03)
Origtime		-0.03				
		(0.07)				
Years Since Neolithic Revolution			0.27***			
			(0.04)			
Population Density (1500CE)				0.11***		
				(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.10: 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***	-0.44***	-0.35***	-0.43***	-0.41***	-0.27***
	(0.03)	(0.06)	(0.02)	(0.03)	(0.03)	(0.03)
Origtime		-0.03				
		(0.07)				
Years Since Neolithic Revolution			0.27***			
			(0.04)			
Population Density (1500CE)				0.11***		
				(0.04)		
Mean Size of Local Communities						0.45***
						(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.11: Distance to Addis Ababa, Pre-Industrial Development, and Economic Specialization

	Pre-Industrial Measures of Trade							
	Economic Specialization	Importance of Trade	Intercommunity Trade as Food Source	Money	Credit Source	Writing and Records	Technological Specialization	Complexity
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pre-Industrial	-0.38***	-0.23***	-0.39***	-0.45***	-0.18**	-0.21***	-0.29***	-0.36***
Distance to Addis Ababa	(0.05)	(0.06)	(0.07)	(0.06)	(0.08)	(0.06)	(0.05)	(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 State: Additional Results and Supporting Material

B.1 Summary Statistics

Table B.1: Summary Statistics on Reduced Sample

	Mean	Std	Min	Max	N
Any Centralized State	0.63	(0.48)	0.00	1.00	142
Centralized State	0.35	(0.48)	0.00	1.00	142
Jurisdictional Hierarchy Beyond Local Community	2.26	(1.27)	1.00	5.00	142
Economic Specialization	1.36	(1.42)	0.00	7.00	115
Economic Specialization (Share)	0.21	(0.19)	0.00	0.80	115
Economic Specialization (Dev)	9.01	(3.98)	2.00	25.00	115
Population Diversity	0.71	(0.05)	0.47	0.76	142
Absolute Latitude	15.04	(14.88)	0.04	68.67	142
Area	0.15	(0.77)	0.00	8.97	142
Elevation (Avg.)	806.21	(667.02)	27.79	3581.35	142
Precipitation (Avg.)	92.36	(56.12)	11.77	334.73	142
Temperature (Avg.)	20.97	(7.97)	-13.44	28.31	142
Malaria Ecology	8.08	(8.97)	0.00	29.36	142
Ecological Diversity	0.26	(0.22)	0.00	0.67	142
Agricultural Suitability (avg.)	0.79	(0.31)	0.00	1.00	142
Agricultural Suitability (std.)	0.07	(0.10)	0.00	0.45	142
Caloric Suitability Index (Pre-1500CE)	2710.62	(1014.21)	0.00	5231.99	142
Caloric Suitability (Pre-1500 ,std.)	401.84	(354.84)	0.00	1520.41	142
Temperature (Spatial Corr., Avg.)	0.92	(0.18)	0.00	1.00	142
Temperature (Volatility, Avg)	0.82	(0.46)	0.27	2.87	142
Pct. Area within 100kms of Sea	0.20	(0.34)	0.00	1.00	142
Coast Length	0.40	(1.96)	0.00	19.65	142
Ruggedness (Avg.)	104.95	(140.76)	1.27	1076.01	142
Pre-Industrial Mobility (avg.)	0.28	(0.06)	0.07	0.39	142
Pre-Industrial Mobility (std.)	0.05	(0.04)	0.01	0.25	142

Table B.2: Summary Statistics on Full Sample

	Mean	Std	Min	Max	N
Any Centralized State	0.53	(0.50)	0.00	1.00	912
Centralized State	0.25	(0.43)	0.00	1.00	912
Jurisdictional Hierarchy Beyond Local Community	1.91	(1.09)	1.00	5.00	912
Economic Specialization	0.85	(1.21)	0.00	7.00	912
Economic Specialization (Share)	0.13	(0.17)	0.00	1.00	912
Economic Specialization (Dev)	7.75	(3.61)	1.00	25.00	912
Predicted Population Diversity	0.68	(0.05)	0.54	0.76	912
Absolute Latitude	20.74	(16.65)	0.02	71.22	912
Area	0.07	(0.37)	0.00	8.97	912
Elevation (Avg.)	750.74	(665.80)	1.06	4417.96	912
Precipitation (Avg.)	106.28	(71.27)	0.00	499.24	912
Temperature (Avg.)	19.08	(8.62)	-15.31	29.58	912
Malaria Ecology	5.60	(8.06)	0.00	33.95	912
Ecological Diversity	0.19	(0.21)	0.00	0.82	912
Agricultural Suitability (avg.)	0.76	(0.34)	0.00	1.00	912
Agricultural Suitability (std.)	0.07	(0.10)	0.00	0.47	912
Caloric Suitability Index (Pre-1500CE)	2664.64	(1280.01)	0.00	6955.56	912
Caloric Suitability (Pre-1500 ,std.)	361.12	(332.26)	0.00	2436.89	912
Temperature (Spatial Corr., Avg.)	0.86	(0.28)	0.00	1.00	912
Temperature (Volatility, Avg)	0.98	(0.57)	0.00	3.08	912
Pct. Area within 100kms of Sea	0.30	(0.42)	0.00	1.00	912
Coast Length	0.35	(3.00)	0.00	81.92	912
Ruggedness (Avg.)	137.71	(161.09)	0.05	1137.67	912
Pre-Industrial Mobility (avg.)	0.27	(0.07)	0.06	0.47	912
Pre-Industrial Mobility (std.)	0.06	(0.05)	0.00	0.27	912

B.2 State: Additional Results and Supporting Material

Table B.3: Population Diversity and Centralization (Alternative Measures)

	Any Centralized State			Centralized State		
	(1)	(2)	(3)	(4)	(5)	(6)
Population Diversity	0.22***	0.27***	0.27***	0.10	0.24***	0.26***
	(0.07)	(0.08)	(0.08)	(0.06)	(0.08)	(0.08)
Absolute Latitude		0.69***	0.62**		0.85***	0.87***
		(0.16)	(0.27)		(0.17)	(0.25)
Area		0.08*	0.05		0.13*	0.09
		(0.04)	(0.13)		(0.06)	(0.17)
Elevation (Avg.)		0.27**	0.13		0.12	0.36
		(0.11)	(0.27)		(0.12)	(0.25)
Precipitation (Avg.)		-0.03	-0.10		0.11	-0.03
		(0.10)	(0.12)		(0.10)	(0.12)
Temperature (Avg.)		0.60***	0.44		0.64***	-0.26
		(0.18)	(0.37)		(0.18)	(0.38)
Malaria Ecology			-0.17			0.00
			(0.13)			(0.12)
Agricultural			0.13			-0.03
Suitability (avg.)			(0.11)			(0.11)
Agricultural			0.04			0.14
Suitability (std.)			(0.10)			(0.10)
Caloric Suitability			-0.26**			-0.19
(Pre-1500 ,avg.)			(0.13)			(0.12)
Caloric Suitability			0.29**			0.40***
(Pre-1500 ,std.)			(0.14)			(0.10)
Pct. Area within			-0.00			0.23*
100kms of Sea			(0.11)			(0.12)
Coast Length			-0.01			-0.04
			(0.14)			(0.18)
Ruggedness (Avg.)			0.09			-0.36
			(0.22)			(0.24)
Pre-Industrial			0.27			1.23***
Mobility (avg.)			(0.43)			(0.44)
Pre-Industrial			-0.11			-0.24
Mobility (std.)			(0.16)			(0.15)
Altonji et al		-5.00	-5.06		-1.71	-1.63
δ		0.68	0.71		0.18	0.22
β-Oster		0.44	0.37		0.74	0.53
R^2	0.05	0.19	0.24	0.01	0.23	0.37
Adjusted-R^2	0.04	0.15	0.15	0.00	0.19	0.29
Observations	142	142	142	142	142	142

Table B.4: Predicted Expected Heterozygosity and State Centralization (Ordered Probit)

	Statehood Level (Ordered Probit)										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Predicted Population Diversity	9.80***	11.21***	13.23***	11.22***	11.29***	13.13***	11.96***	11.36***	11.17***	17.04***	8.77**
	(0.81)	(1.02)	(1.14)	(1.01)	(1.02)	(1.01)	(1.17)	(1.03)	(1.03)	(1.46)	(4.13)
Malaria Ecology			-0.03***							-0.04***	-0.04***
			(0.01)							(0.01)	(0.01)
Ecological Diversity				0.70***						0.37*	0.31
				(0.19)						(0.21)	(0.21)
Agricultural Suitability (avg.)					0.39***					0.07	0.07
					(0.14)					(0.16)	(0.19)
Agricultural Suitability (std.)					0.55					0.19	0.79
					(0.49)					(0.56)	(0.69)
Caloric Suitability (Pre-1500 ,avg.)						0.00***				0.00***	0.00***
						(0.00)				(0.00)	(0.00)
Caloric Suitability (Pre-1500 ,std.)						0.00*				0.00*	0.00
						(0.00)				(0.00)	(0.00)
Temperature (Spatial Corr., Avg.)							-0.40**			-0.68***	-0.43
							(0.17)			(0.25)	(0.34)
Temperature (Volatility, Avg)							0.34*			0.58***	0.62***
							(0.18)			(0.22)	(0.23)
Pct. Area within 100kms of Sea								0.02		-0.45***	-0.04
								(0.11)		(0.17)	(0.21)
Coast Length								-0.02		-0.01	-0.00
								(0.02)		(0.02)	(0.02)
Ruggedness (Avg.)									0.00	0.00**	0.00
									(0.00)	(0.00)	(0.00)
Pre-Industrial Mobility (avg.)									-0.22	4.94**	1.42
									(1.96)	(2.30)	(2.42)
Pre-Industrial Mobility (std.)									-0.19	-6.65***	-4.03*
									(1.92)	(2.24)	(2.33)
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
Pseudo-R^2	0.07	0.12	0.13	0.13	0.12	0.15	0.12	0.12	0.12	0.17	0.19
Observations	912	912	912	912	912	912	912	912	912	912	912

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 Statehood Centralization (Ordered Probit)

	Statehood Level (ordered probit)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Population Diversity	3.72**	9.40***	10.48***	9.29***	9.19***	10.78***	9.57***	9.10***	9.78***	8.46***
	(1.87)	(2.77)	(2.80)	(2.88)	(2.73)	(2.97)	(2.67)	(2.75)	(2.71)	(2.78)
Malaria Ecology			-0.03**							-0.01
			(0.01)							(0.02)
Ecological Diversity				0.82*						0.22
				(0.44)						(0.51)
Agricultural					0.71*					1.31**
Suitability (avg.)					(0.38)					(0.61)
Agricultural					2.11*					1.64
Suitability (std.)					(1.09)					(1.32)
Caloric Suitability						-0.00				-0.00**
(Pre-1500 ,avg.)						(0.00)				(0.00)
Caloric Suitability						0.00***				0.00***
(Pre-1500 ,std.)						(0.00)				(0.00)
Temperature (Spatial							-1.15			-1.14
Corr., Avg.)							(0.76)			(1.07)
Temperature							-1.54***			-1.45**
(Volatility, Avg)							(0.51)			(0.67)
Pct. Area within								0.78**		0.18
100kms of Sea								(0.35)		(0.50)
Coast Length								0.14		0.16*
								(0.11)		(0.09)
Ruggedness (Avg.)									0.00	0.00
									(0.00)	(0.00)
Pre-Industrial									-3.47	5.68
Mobility (avg.)									(8.79)	(9.94)
Pre-Industrial									3.83	-10.30**
Mobility (std.)									(3.83)	(4.86)
Main Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo-R^2	0.01	0.15	0.16	0.16	0.16	0.19	0.17	0.17	0.15	0.23
Observations	142	142	142	142	142	142	142	142	142	142

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 Statehood Centralization

	Statehood Level					
	OLS			IV		
	(1)	(2)	(3)	(4)	(5)	(6)
Expected	0.13**	0.27***	0.20**	0.15*	0.31**	0.20
Heterozygosity	(0.06)	(0.07)	(0.08)	(0.09)	(0.13)	(0.13)
Main Controls	No	Yes	Yes	No	Yes	Yes
All Additional Controls	No	No	Yes	No	No	Yes
Altonji et al		-1.88	-2.78			
δ		0.35	0.56			
β-Oster		0.42	0.22			
First-stage F-statistic				167.28	69.24	76.00
R^2	0.02	0.30	0.47	0.01	0.29	0.47
Adjusted-R^2	0.01	0.26	0.39	0.01	0.26	0.39
Observations	142	142	142	142	142	142

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.7: Population Diversity and Statehood Centralization

	Any Centralized State					
	OLS			IV		
	(1)	(2)	(3)	(4)	(5)	(6)
Expected	0.22***	0.27***	0.22**	0.30***	0.35***	0.27**
Heterozygocity	(0.07)	(0.08)	(0.09)	(0.08)	(0.09)	(0.11)
Main Controls	No	Yes	Yes	No	Yes	Yes
All Additional Controls	No	No	Yes	No	No	Yes
Altonji et al		-5.00	53.52			
δ		0.70	1.69			
β-Oster		0.40	0.21			
First-stage F-statistic				167.28	69.24	76.00
R^2	0.05	0.19	0.27	0.04	0.18	0.27
Adjusted-R^2	0.04	0.15	0.16	0.04	0.15	0.16
Observations	142	142	142	142	142	142

Notes: This table establishes the causal positive statistically and economically significant relation between statehood and population diversity as instrumented by 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 B.8: Population Diversity and Centralization

	Any Centralized State (Probit)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Expected	2.03***	2.79***	3.17***	2.76***	2.80***	2.83***	2.73***	2.79***	2.97***	2.76***
Heterozygocity	(0.69)	(0.83)	(0.82)	(0.85)	(0.82)	(0.88)	(0.81)	(0.84)	(0.85)	(0.89)
Malaria Ecology			-0.01**							-0.01*
			(0.00)							(0.01)
Ecological Diversity				0.35**						0.22
				(0.16)						(0.18)
Agricultural					0.11					0.27
Suitability (avg.)					(0.14)					(0.22)
Agricultural					0.55					-0.03
Suitability (std.)					(0.48)					(0.51)
Caloric Suitability						-0.00				-0.00**
Index (Pre-1500CE)						(0.00)				(0.00)
Caloric Suitability						0.00***				0.00**
(Pre-1500 ,std.)						(0.00)				(0.00)
Temperature (Spatial							-0.23			-0.23
Corr., Avg.)							(0.25)			(0.34)
Temperature							-0.44**			-0.31
(Volatility, Avg)							(0.22)			(0.25)
Pct. Area within								0.19		-0.01
100kms of Sea								(0.13)		(0.15)
Coast Length								-0.00		-0.01
								(0.03)		(0.03)
Ruggedness (Avg.)									0.00	0.00
									(0.00)	(0.00)
Pre-Industrial									-1.77	-0.54
Mobility (avg.)									(2.97)	(3.52)
Pre-Industrial									1.58	-1.98
Mobility (std.)									(1.79)	(2.08)
Main Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo-R^2	0.04	0.18	0.20	0.20	0.18	0.21	0.20	0.19	0.18	0.27
Observations	142	142	142	142	142	142	142	142	142	142

Marginal effects

Table B.9: Predicted Population Diversity and State Centralization

	Any Centralized State										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Predicted Expected	0.41***	0.40***	0.45***	0.40***	0.41***	0.47***	0.43***	0.42***	0.41***	0.60***	0.27*
Heterozygosity	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.09)	(0.16)
Malaria Ecology			-0.12***							-0.21***	-0.23***
			(0.05)							(0.07)	(0.05)
Ecological Diversity				0.07**						0.03	0.01
				(0.04)						(0.04)	(0.03)
Agricultural					0.08**					-0.01	0.01
Suitability (avg.)					(0.05)					(0.08)	(0.05)
Agricultural					0.01					-0.02	0.04
Suitability (std.)					(0.04)					(0.05)	(0.05)
Caloric Suitability						0.20***				0.16***	0.18***
(Pre-1500 ,avg.)						(0.06)				(0.07)	(0.05)
Caloric Suitability						0.02				0.06	-0.01
(Pre-1500 ,std.)						(0.04)				(0.06)	(0.05)
Temperature (Spatial							-0.01			-0.08	-0.05
Corr., Avg.)							(0.06)			(0.08)	(0.07)
Temperature							0.23***			0.27***	0.27***
(Volatility, Avg)							(0.12)			(0.18)	(0.12)
Pct. Area within								-0.09***		-0.17***	-0.03
100kms of Sea								(0.05)		(0.07)	(0.06)
Coast Length								-0.04**		-0.01	-0.00
								(0.10)		(0.12)	(0.06)
Ruggedness (Avg.)									0.10	0.22***	0.17**
									(0.12)	(0.19)	(0.12)
Pre-Industrial									-0.00	0.18	-0.00
Mobility (avg.)									(0.17)	(0.22)	(0.17)
Pre-Industrial									-0.11	-0.25***	-0.15**
Mobility (std.)									(0.09)	(0.14)	(0.09)
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	0.16	0.20	0.20	0.20	0.20	0.23	0.20	0.20	0.20	0.26	0.30
Observations	912	912	912	912	912	912	912	912	912	912	912

Notes: This table establishes the causal positive statistically and economically significant relation between statehood centralization and population diversity as predicted by 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. 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 B.10: Heterogeneous Effects of Predicted Population Diversity on State Centralization

	Any Centralized State					
	(1)	(2)	(3)	(4)	(5)	(6)
Predicted Expected Heterozygosity	0.40***	0.39***	0.35***	0.48***	0.30***	0.29***
	(0.04)	(0.05)	(0.04)	(0.09)	(0.05)	(0.06)
Predicted Population Diversity		0.13				
× Ecological Diversity		(0.42)				
Predicted Population Diversity			1.67***			
× Agricultural Suitability (std.)			(0.61)			
Predicted Population Diversity				-0.40		
× Temperature (Volatility, Avg)				(0.72)		
Predicted Population Diversity					2.13***	
× Ruggedness (Avg.)					(0.55)	
Predicted Expected Heterozygosity						1.39***
× Pre-Industrial Mobility (std.)						(0.53)
Main Controls	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.20	0.20	0.21	0.20	0.21	0.20
Observations	912	912	912	912	912	912

Notes: This table establishes the causal positive statistically and economically significant relation between statehood centralization and population diversity as predicted by the distance to Addis Ababa (see section 4.2). 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.