Preliminary

Ethnic Inequality*

Alberto Alesina
Harvard University and Igier

Stelios Michalopoulos
Brown University

Elias Papaioannou
Dartmouth College

March 2012

Abstract

This study explores the consequences and origins of contemporary differences in well-being across ethnic groups within countries. First, we construct country-level measures of ethnic inequality combining anthropological data on the spatial distribution of ethnic/linguistic groups with satellite images on light density at night. Second, we show that ethnic inequality is strongly negatively correlated with per capita income; this result pertains even when we condition on fractionalization, income inequality, and numerous other country characteristics. Third, when we explore the roots of ethnic inequality, we find that differences in geographic endowments across ethnic homelands explain a sizable portion of contemporary ethnic inequality. Fourth, we show that deeply rooted inequality in geographic endowments across ethnic regions in inversely related to contemporary development. Fifth, we show that the strong negative correlation between ethnic inequality and well-being obtains also when we solely explore within country variation using micro data from the Afrobarometer surveys.

Keywords: Ethnicities, Fractionalization, Development, Inequality, Geography

JEL classification Numbers: O10, O40, O43.

*For comments we thank Nathan Nunn and participants in seminars at Dartmouth and the Athens University of Economics and Business.

1 Introduction

Ethnic diversity has costs and benefits. On the one hand, diversity in skills, education, endowments can enhance productivity by promoting trade and innovation. On the other hand, ethnic diversity is often associated with inadequate public goods provision, poor policies, conflict, civil wars, and hatred. In fact a large literature shows a negative (though not always robust) effect of ethnolinguistic fragmentation on various aspects of economic development, with the possible exception of wealthy economies.[1]

This paper puts forward and tests an alternative conjecture. Our thesis is that what matters for development are economic differences between ethnic groups coexisting in the same country (region), rather than the degree of fractionalization. Inequality in income along ethnic lines is likely to increase animosity, impede institutional development, and lead to state capture. In addition differences in the level of development across ethnic groups is often associated with discriminatory policies of one (or more) groups against the others. As such ethnic inequality may lead to inadequate public goods provision. Moreover as Chua (2005) argues the presence of an economically dominant minority ethnic group may lower support for market institutions, as the majority of the population usually feels that the benefits of capitalism go to just a couple of ethnic groups.

The first (and perhaps main) contribution of this paper is to provide a measure of within-country differences in well-being across ethnic homelands, which we label "ethnic inequality". Internationally comparable data on income levels of ethnic groups for all countries are not available. Thus in order to construct country-level indicators of ethnic inequality, we combine geo-referenced data on the geographic location of ethnic groups within countries (using information from the *Ethnologue* and the digitized version of *Atlas Narodov Mira*) with satellite images on light density at night, which are available at a fine grid;[2] by doing so we are able to construct ethnic inequality proxy measures for all countries in the world in a fully consistent manner.[3] Interestingly the cross-ethnic group inequality index is weakly correlated with the commonly employed (and notoriously poorly measured) measure of income inequality at the national level (Gini coefficient). Ethnic inequality is only modestly positively correlated with

[1] We review this body of work below. See Alesina and La Ferrara (2005) for a survey.

[2] For some countries there are census or survey data on wages, earnings, and access to basic public goods at the ethnicity level. Yet, such data are available for only a sub-set of countries. Most importantly, since survey methods and censuses differ, even these data are not particularly useful for cross-country studies (see Banerjee and Duflo (2003) for a discussion of measurement issues of cross-country Gini coefficient indicators).

[3] Our approach proxying income per capita within countries across (ethnic and other) regions with luminosity builds on recent research showing that light density at night reflects well economic development both across countries over time and within countries across regions (see, among others, Henderson, Storeygard, and Weil (2011); Chen and Nordhaus (2011); and Michalopoulos and Papaioannou (2011)). We discuss the main features of our data below.

the standard measures of ethnolinguistic fragmentation that do not distinguish between differences in income across ethnic groups. To isolate the cross-ethnic component of inequality from the overall level of inequality across regions within a country, we utilize the fine level of the luminosity data and construct a measure of geographical (spatial) inequality.

Second, we relate our newly constructed indicators of ethnic inequality with income per capita across countries. We find a remarkably strong negative association between ethnic inequality and development. Underdevelopment goes in tandem with ethnic inequality. Moreover, we find that ethnic inequality is strongly inversely related to GDP per capita, even when we condition on the overall degree of spatial inequality. This result suggests that differences in development across ethnic groups rather than the general inequality in development within countries is correlated with slower development. We also show that the significantly negative correlation between ethnolinguistic fragmentation and development that has been shown in previous works (e.g. Easterly and Levine (1997); La Porta *et al.* (1999); Alesina *et al.* (2003)) weakens considerably and becomes statistically indistinguishable from zero when we also include in the specification the newly constructed index of ethnic inequality. These results thus show that it is the unequal concentration of wealth in some ethnically defined groups that is especially detrimental for development rather than ethnic diversity per se.

Third, given the strong negative correlation between development and ethnic inequality, we search for deeply rooted the factors that affect contemporary wealth differences across ethnic groups within countries. To examine the deep roots of ethnic inequality, we construct fine as well as a composite index reflecting differences in geographic endowments (elevation, land quality, sea distance, water sources) across ethnic homelands. We find that differences in geography across ethnic regions explain a sizable portion of contemporary ethnic inequality. This result appears remarkably strong and indicates that differences in geography across ethnic homelands have affected contemporary inequality in well-being.

Fourth, we show that contemporary development is considerably lower in countries with large differences in geographic endowments across ethnic homelands. This implies that ethnic inequality may have crucially shaped the development path. In the same vein, two stage-least-squares specifications show that the component of contemporary ethnic inequality (as captured by luminosity) explained by differences in ethnic-specific geographic differences is negatively correlated with per capita GDP. And while inequality in geographic characteristics across ethnic homelands may affect development at the country level via other mechanisms, the result retains significance when we account for the overall degree of spatial inequality in geographic endowments (which itself is not negatively associated with development) and other country characteristics.

Finally we provide within country results showing a similar negative association between ethnic inequality and direct measures of well being and public goods provision. In particular we present within country specifications associating regional development with ethnic inequality in the most ethnically heterogenous and unequal part of the world, Sub-Saharan Africa. Using micro level data from the 2005 Afrobarometer Surveys spanning 17 countries we show that ethnic inequality in inversely related to development and public goods provision at the region level.

Related Literature Our paper contributes mostly to the literature on the effects of ethnic diversity on various aspects of development. Easterly and Levine (1997) first showed a strong negative correlation between ethnolinguistic fragmentation and cross-country growth rates. Alesina *et al.* (2003) construct more refined data on ethnic, linguistic, and religious fractionalization and show that properly measuring fractionalization is key in understanding its effects. Subsequent studies tend to document an inverse relationship between ethnolinguistic fragmentation and various measures of well-being, such as income per capita, public goods provision, institutional quality, etc. though the correlations are not always robust.[4] Recently several authors have constructed more refined indicators of fractionalization, taking into account the degree of cultural and genetic similarities (Fearon (2003); Desmet, Ortuno-Ortin, and Wacziarg (2011)), the polarization of groups (e.g. Montalvo and Reynal-Querol (2005)), and the presence of a dominant ethnic group (e.g. Fearon, Kasara, and Laitin (2007)). Alesina and Zhuravskaya (2011) build a new index of segregation and show that ethnic segregation -the clustering of ethnic groups across regions within a country- is detrimental to long-run institutional development and reduces trust amongst ethnic groups. Alesina, Easterly, and Matuszeski (2011), Englebert, Tarango and Carter (2002), and Michalopoulos and Papaioannou (2011) show that ethnic partitioning is associated with a lower level of development and higher degree of civil conflict.

Very few authors have looked at the consequences and driving forces of economic inequality between ethnic groups inside a country. Baldwin and Huber (2010) use survey-level data on well-being (mostly from the Afrobarometer and the World Value Surveys) to construct an index of between-group heterogeneity for 46 democratically governed countries. They then show that between group economic inequality is associated with a lower degree of pubic goods provision. Chua (2003) argues that the presence of economically dominant ethnic minorities leads to ethnic hatred, conflict, and institutional capture. She builds her argument using case-studies where inequality mainly between a dominant ethnic group and the other ethnicities in a country

[4] See Alesina and La Ferrara (2005) for a survey.

has spurred conflict and undermined the consolidation of free market institutions. Examples of ethnic hatred against economically dominant minorities include the Ibo in Cameroon, the Tutsi in Rwanda, the Kikuyu in Kenya and Chinese minorities in many East Asian countries.

Our paper makes contact to several other strands of literature. First, our paper is related to the research on inequality and development, which dates back (at least) to the influential work of Simon Kuznets (see Perotti (2006) and Benabou (2006) for recent surveys). Yet due to data problems and theoretical ambiguities, the empirical studies produce conflicting and in general insignificant correlations between cross-country Gini coefficients and development or economic growth (see Banerjee and Duflo (2003)). Our work contributes to this body of work by emphasizing a neglected component of income disparities inside a country, the ethnic component. Since the negative association between ethnic inequality and contemporary development retains its statistical and economic significance when we condition on the overall degree of spatial inequality (in luminosity) implies that ethnic, rather than the overall inequality, is detrimental to development.

Second, our paper is related to empirical works studying the deep origins of contemporary development. The literature has mainly focused on the impact of colonization and early institutions (see for a review Acemoglu, Johnson, and Robinson (2005)), cultural features, such as family ties, trust, social capital and religion (see for reviews Guiso, Sapienza, and Zingales (2006) and Fernandez (2011)), and geography (Sachs (2010); Nunn and Puga (2011)).[5] A notable difference of our paper with this body of research, is that the effects of geography on development go through the effect of inequality in endowments rather than its level. In this regard our work relates to Ashraf and Galor (2011, 2012) who study the effects of cultural and genetic diversity in the process of economic development before and after the Industrial Revolution.

Third, our paper relates to the literature on optimal country size (see Alesina and Spolaore (2005) and Alesina, Spolaore, and Wacziarg (2005) for reviews). This body of research models the advantages and drawbacks of country size and fractionalization on development mostly via trade and animosity; in contrast our results showing a strong negative association between per capita GDP and inequality in geographic endowments across ethnic homelands, suggests that heterogeneity in geography may be detrimental to economic performance, especially when it interacts with ethnic differences in endowments.

Finally our results showing a strong within-country negative association between ethnic inequality and various measures of local public goods and regional development across Sub-Saharan African states adds to the literature on the deep roots of African development (see,

[5] Frankel and Romer (1999) and Alcala and Ciccone (2004), among others, show that geography affects current productivity levels via affecting trade.

Miguel (2010) and Collier (2009) for reviews). In particular our results that highlight the role of income disparities across ethnic groups is mostly related to works emphasizing the negative development effects of the artificial, non-organic, nature of African states (see for example Herbst (2001) and Michalopoulos and Papaioannou (2011)).

Structure The paper is organized as follows. In the next section we explain in detail the construction of the ethnic inequality variable that combines information on ethnicities' homelands with satellite images on light density at night. We also present summary statistics of the newly constructed variables and some basic correlations between the cross ethnic inequality indicators and measures of fractionalization, income inequality, and development. In Section 3 we report the results of our analysis associating income per capita with ethnic inequality. In Section 4 we examine the origins of contemporary ethnic inequality. In Section 5 we report least squares estimates associating contemporary development with inequality in geography across ethnic homelands. We also report two-stage least squares estimates that link income per capita with the component of ethnic inequality predicted by geographic differences across ethnic homelands. In Section 6 we examine the within country across regions association between ethnic inequality and wealth, as well as public goods provision in 17 Sub-Saharan countries using micro-level data from the Afrobarometer. In the last section we summarize and discuss some directions for future research.

2 Data, Descriptive Statistics, and Basic Correlations

2.1 Data

2.1.1 Location of ethnic groups

We identify the location of ethnic groups using two data sets that report the spatial distribution of ethnic (linguistic) groups around the world. First we identify ethnic homelands with the Geo-Referencing of Ethnic Groups (GREG), which is the digitized version of the classical *Soviet Atlas Narodov Mira* (Weidmann, Rød and Cederman (2010)). GREG portrays the homelands of $1,276$ ethnic groups around the world. The information pertains to the early 1960's so for many countries, in Africa in particular, it corresponds to the time of independence.[6] The GIS data set uses the political boundaries of 1964 to allocate groups to different countries. To be able to use contemporary country-level statistics we project the ethnic homelands within the political boundaries of the 2000 Digital Chart of the World (ignoring tiny homelands of less

[6] The original Atlas Narodov Mira consists of 57 ethnographic maps, covering all regions of the world at various scales. The original sources of the maps are the following three: (1) ethnographic and geographic maps assembled by the Institute of Ethnography at the USSR Academy of Sciences, (2) population census data, and (3) ethnographic publications of government agencies.

than 1 square kilometer). There are 2,127 ethnic homelands within all contemporary countries. Most areas (1,623) are coded as pertaining to a single group whereas in the remaining 495 there can be up to three groups sharing the same territory. In case of overlapping regions we assign the respective homeland to each of the groups located there. The size of the polygons varies considerably: The smallest polygon occupies an area of 1.10 km^2 (this is the Chinese (Han) in Laos) and the largest polygon extends over 7,335,476 km^2 (this is the case of American English in the US). The median (mean) size of a group is 4,185 (61,270) km^2. In the GREG dataset the median (mean) country has 6 (11.68) ethnic groups with the most diverse being Indonesia with 95 mapped ethnicities.

Our second source is the $15th$ edition of the *Ethnologue* (Gordon (2005)) which maps 7,570 linguistic (rather than ethnic) groups. Unlike the GREG in case of the *Ethnologue* we do not need to reassign the linguistic homelands since the initial GIS mapping is done using the political boundaries of 2000. Although the *Ethnologue* is a global data set with detailed linguistic mappings, its coverage for some continents (e.g. Latin America) is quite limited while for others (i.e. Africa) is very detailed. Another limitation of the *Ethnologue* is that it corresponds to the early 1990's, and thus the location of ethnic groups maybe partially affected by national policies and institutions, ethnic conflict, or other features. Each ethnolinguistic polygon in the *Ethnologue* delineates a traditional linguistic region; populations away from their homelands (e.g., in cities, refugee populations, etc.) are not mapped. Linguistic groups of unknown location, widespread languages, and extinct languages, are not mapped; the only exception for not mapping widespread languages is the case of the English language, which is mapped for the United States. *Ethnologue* also records areas where languages overlap; in this case we assign the polygon where two say languages are spoken to both linguistic groups. *Ethnologue* provides a more refined language aggregation compared to the GREG. As a result the median (mean) homeland extends to 728 (12,986) km^2. The smallest mapped language is that of the Domari group in Israel which covers 1.18 km^2 with the largest language group being again the English language in the US covering 9,327,331 km^2. The median (mean) country in the *Ethnologue* has 8.50 (43.71) language groups with Papua New Guinea being the most linguistically diverse with 808 groups.

2.1.2 Luminosity data

Comparable data on income per capita at the ethnicity level across all countries in the world do not exist. Therefore as a proxy for ethnic development we use satellite image data on light density at night to measure economic activity at the ethnic (linguistic) homeland level. The luminosity data come from the Defense Meteorological Satellite Program's Operational

Linescan System (DMSP-OLS) that reports images of the earth at night captured from 20 : 00 to 21 : 30 local time. The measure is a six-bit digital number calculated for every 30-second area pixel/cell (approximately 1 square kilometer), which is averaged across overlapping raw input pixels on all evenings in a year. The index ranges from 0 -indicating either the absence of light or a very low degree of luminosity that cannot be captured by the satellite sensors- to 63, which usually coincides with the value at the capital cities in the developed countries. The data is available since 1992. To construct light density at the desired level of aggregation we first average the digital number of luminosity across pixels that fall within the boundaries of an ethnic group (results are similar if we use the median value of light intensity). To obtain a per-capita measure of development we then divide the luminosity value of each ethnic region with the respective population, using data from the Gridded Population of the World' dataset (GPW, CIESIN, 2005) that reports geo-referenced pixel-level population estimates across the globe for 1990 and 2000.[7]

This approach builds on the recent contribution of Henderson *et al.* (2012) and subsequent works (e.g. Chen and Nordhaus (2011), Michalopoulos and Papaioannou (2011)) which show a strong positive correlation between luminosity and development at various levels of aggregation. Henderson *et al.* (2012) show that satellite light density captures well abrupt changes in economic activity both at an aggregate and at a local scale (as for example during the Rwandan genocide or when large deposits of rubies and sapphires were accidently discovered in Madagascar). In the same vein, Michalopoulos and Papaioannou (2011) show a strong correlation between light density at night and GDP per capita across African countries as well as a significant within country negative correlation between light density and infant mortality. Moreover, using micro-level data from the Demographic and Health Surveys and from the 2005 Afrobarometer Surveys they show a positive within-country across-ethnicity correlation between luminosity and various public goods provision measures (such as access to clean water, the presence of a sewage system, a composite wealth index, etc.). Other studies showing that luminosity captures well economic activity and public goods provision include **?**, **?**, and **?**, among others (see Henderson *et al.* (2012) for additional references).

Figures 1*a* and 1*b* illustrate the strong cross-country association between (log) GDP per capita and (log) luminosity per capita in 2000. The simple correlation is 0.77 and the R^2 of the unconditional model is 0.60. The elasticities are 0.51 and 0.34 for the unconditional and region fixed effects specification, respectively. Both estimates are highly significant, as the estimate is more than 9 times larger than the standard error.

[7] To construct luminosity per capita within an ethnic homeland in 1992 we use the population estimates of 1990, whereas for luminosity per capita in 2000 and 2009 we use the population estimates of 2000.

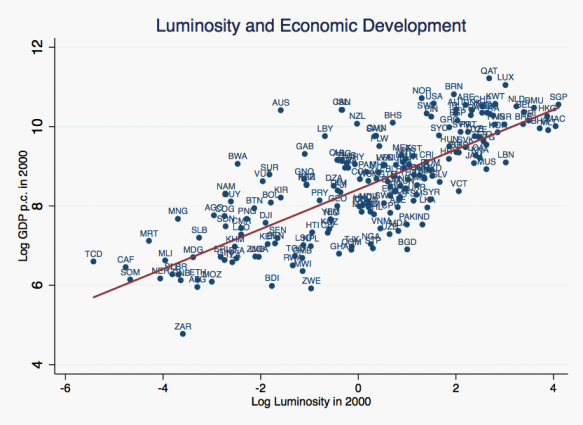


Figure 1*a*

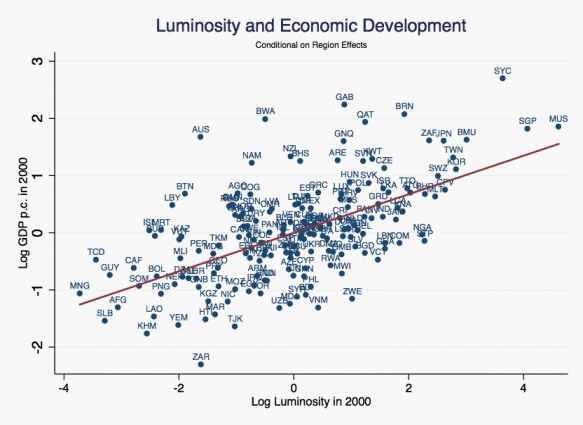


Figure 1*b*

An important caveat has to be kept in mind. Differences in luminosity within a country may reflect both income differences across citizens in these regions and/or differences in the provision of public goods -access to electricity in particular for private use, streets lamps etc. Thus to some extent we may be capturing at the same time both differences in per capita income and differences in how different groups are favored by public policies. Our hypothesis is that these two features should go handy, since a larger degree of income inequality across groups may foster animosity and the unequal provision of public goods.[8] In Section 6 we report within African countries results associating ethnic inequality at the region level with both an index of well-being and various proxy measures of public goods provision in an effort to isolate these two effects.

2.1.3 Ethnic inequality

We estimate the level development at the ethnic homeland level as average luminosity per capita, and then we aggregate the values at the country level to construct a Gini index that reflects inequality across ethnic groups (ethnic inequality) within every country. Thus, the ethnic Gini coefficient does not capture differences in individual income, but differences in mean income (as reflected in luminosity per capita) across ethnic homelands. For the two different databases (GREG and *Ethnologue*) we construct Gini coefficients (and coefficients of variation) for each country using cross-ethnic-homeland data in 1992, 2000, and 2009. Since in many countries there are some tiny ethnic homelands, we also construct the Gini coefficient of ethnic inequality excluding small ethnicities, defined as groups capturing less than 1% of the 2000 population in a country. For example, in Kenya the *Atlas Narodov Mira* (the *Ethnologue*)

[8] See Alesina, Baqir and Easterly (2001) and Alesina, Baqir and Hoxby (2005) for a discussion of the effects of ethnic fragmentation on disagreements over the provision and allocation of public goods within US localities and Miguel and Gugerty (2005) for evidence associating ethnic diversity with public goods provision in Kenya.

maps 19 (53) ethnic (linguistic) areas. Yet 7 ethnic (37 linguistic) areas are less than one percent of the Kenya's population as of 2000 (see Table 1). We thus construct the ethnic Gini index using all ethnic groups (19 and 53), but also just using the 12 large ethnic and 16 large linguistic areas in Kenya, respectively.[9] We also construct all measures of ethnic inequality excluding homelands where capital cities fall. This is useful as we account for extreme values of lights in the capitals and for the fact that in the capital cities there is likely to be higher ethnic mixing than the one observed in the data.

2.1.4 Spatial inequality

We also construct Gini coefficients (and coefficients of variation) for each country using pixels/cells of (approximately) the same size. Specifically we first compute luminosity per capita across grids/cells of 2.5 x 2.5 decimal degrees (approximately $250km$ x $250km$) for 1992, 2000, and 2009. The median (mean) virtual country extends $25,662$ $(29,676)$ km^2. The size of the median virtual country is similar to the size of an ethnic homeland in the GREG dataset when we exclude those groups with less than 1 percent of a country's population. Then we estimate the Gini coefficient across the "virtual" countries falling in each country. This index (overall spatial Gini coefficient) is intended to capture the overall, rather than the purely ethnic-specific, component of spatial inequality in development. Of course, the overall spatial inequality index partially reflects inequality in luminosity per capita across ethnic homelands. We thus (almost) always include both the ethnic inequality and the overall spatial inequality index in the empirical specifications.

2.1.5 Example

Figures 2, $3a - 3b$, and 4 give a graphical illustration of the construction of the cross-ethnic inequality index for Afghanistan. The GREG map (reproduced in Figure 2) portrays the spatial distribution of 31 ethnic groups. The largest group is the Afghan (which includes the Pashtuns and Pathans) that mostly reside in the southern and central-southern regions of the country. The group takes up 51% of the population in 2000. The second largest group is the Tajiks which compose 22% of the total population and are located in the northeastern regions as well as in scattered pockets in the western part of the country. The smallest group are the Yazghulems in the northeastern part of the country taking up a tiny 0.0001% of the population. There are

[9] There a few small countries in the sample with only one ethnic group. For these countries inequality is zero. According to the GREG maps there are 25 countries with GDP data from the Penn World Tables (Edition 7) with just one ethnic group (e.g. Comoros, Madagascar, Korea, Malta, Sao Tome and Principe). According to the Ethnologue there are 31 countries with just one linguistic group. Since *Ethnologue*'s coverage in South and Central America is limited, we have many countries in these regions with just one group (e.g. Haiti, Cuba, Uruguay). See Table 1 for a complete listing. As we show below our results are robust to including or excluding these countries from the analysis.

also 8 territories in which groups overlap. In four of those the Afghan groups (Pashtuns and Pathans) overlap with the Tajiks, while in two other regions they overlap with the Hazara-Berberi and in one region with the Persians; in one region the Brahui share the same homeland with the Baloch.

Using this mapping we first estimate for each ethnic homeland luminosity per capita. For a group that appears in multiple pockets we derive the weighted average of light density per capita assigning as weights the fraction of each pocket's surface area with respect to the total area of the ethnic group. Figure 3*a* maps the distribution of lights per capita across ethnic homelands in Afghanistan. Regional development, as reflected in luminosity per capita, is minimal in the center of the country, where the Hazara-Berberi tribes reside and in the eastern provinces where the Nuristani, the Pamir Tajiks, the Pashai, and the Kyrgyz tribes are located. Luminosity is clearly higher in the Pashtun/Pathans homelands and to some lesser extent in the Tajik regions. Second, using lights per capita across the 31 ethnic homelands we estimate the Gini coefficient and the coefficient of variation in 1992 (the first year that the luminosity data is available), in 2000 and in 2009. In 2000 the Gini coefficient estimated from GREG is 0.935 remarkably close to the estimate when we use the *Ethnologue* linguistic maps (0.901). We also estimated the ethnic inequality indicators excluding ethnicities constituting less than 1% of a country's population (see Figure 3*b*). Doing so the Gini coefficient is 0.458. When we use the *Ethnologue* maps the ethnic inequality Gini index is almost identical, 0.45; see Table 1.

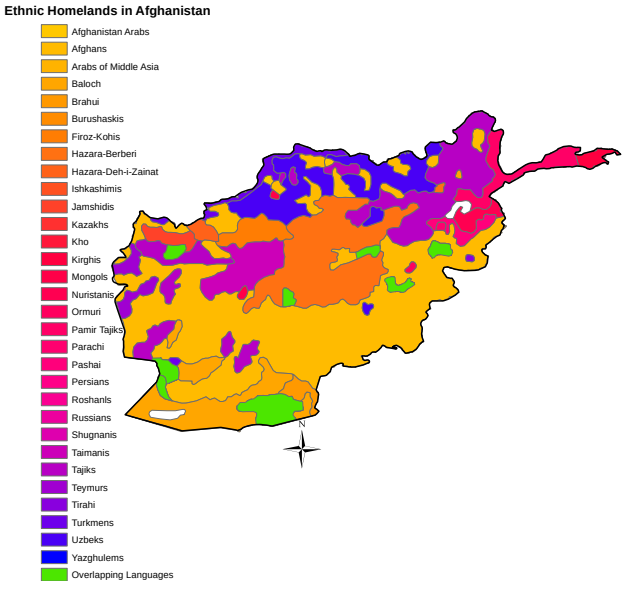


Figure 2: Ethnic Homelands in Afghanistan

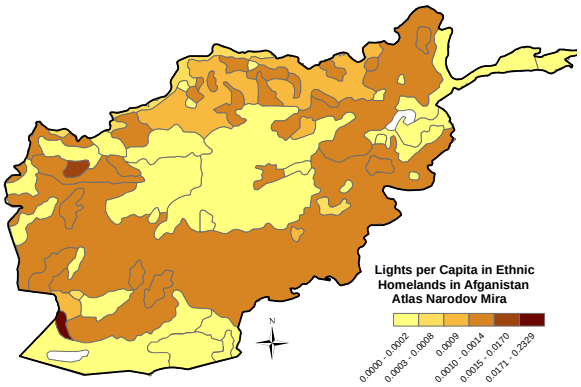


Figure 3*a*: Lights across Ethnic Homelands

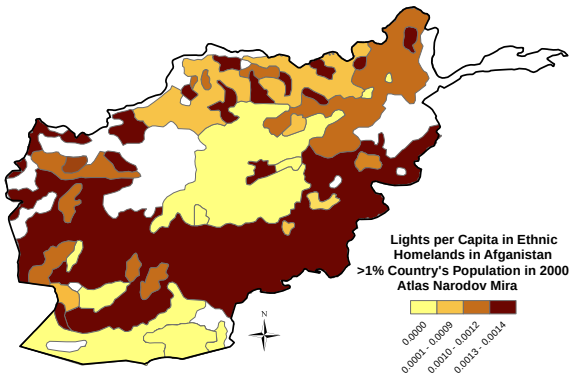


Figure 3*b*: Lights across Ethnic Homelands

Figure 4 portrays the construction of overall spatial inequality index for Afghanistan. First, we split the world into boxes of 2.5 x 2.5 decimal degrees. Then we intersect with countries' boundaries in 2000. As a result there are 24 pixels in Afghanistan. Around a third of them are rectangular in terms of decimal degrees while the rest are smaller, since their contours follow Afghanistan's borders. Second, we estimate for each pixel luminosity per capita, exactly as we did when we used ethnolinguistic homelands as the unit of analysis, by dividing average luminosity with per capita income. Third, we calculate the Gini coefficient (and the coefficient of variation) across these pixels/territories for each country. The resulting measure (overall spatial inequality index) reflects spatial inequality in lights per capita across (randomly carved) pixels.

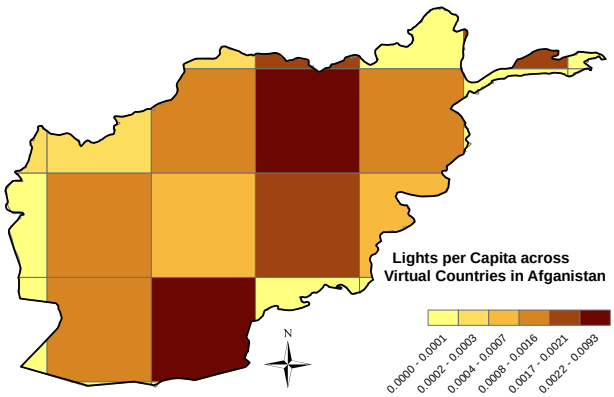


Figure 4: Lights across Pixels/Boxes

2.2 Descriptive Statistics

Table 1 reports the values of the cross-ethnic group inequality index for all countries in 2000 using both the GREG and the *Ethnologue* mapping. The table also gives the number of

ethnic groups in each country and reports the values for the overall spatial inequality index for countries with data on per capita GDP (from the latest vintage of the Penn World Tables). The Data Appendix gives detailed variable definitions and sources.

According to the *Ethnologue*'s more detailed mapping of ethnic homelands the countries with the highest cross-ethnic-group inequality (where Gini exceeds 0.90) are: Angola, Burkina Faso, Central African Republic, Ivory Coast, Cameroon, Congo, Ethiopia, Gabon, Ghana, Liberia, Nigeria, Somalia, Zaire; and outside Africa Afghanistan, Australia, Brazil, Colombia, Indonesia, Laos, Nepal, Philippines, Papua New Guinea, Venezuela, and Vietnam. The countries with the highest overall spatial inequality in light density (Gini higher than 0.90) are Australia, Somalia, Chad, Mali, Zaire, and Sudan. The countries with lowest overall spatial inequality in light density (Gini lower than 0.10) are: Trinidad and Tobago, Rwanda, Comoros, Belgium, and many other very small countries (such as Bahrain, Samoa, Jamaica).[10]

Table 2, Panels A and B report the correlation structure of the ethnic Gini coefficients between the two global maps and in three different points in time. A couple of interesting patterns emerge. First, the correlation of the Gini coefficients across the two alternative mapping of groups is strong, around $0.75 - 0.80$. Second, in the relatively short period where luminosity data are available ($1992 - 2009$), ethnic inequality appears very persistent, as the correlations of the Gini coefficients over time exceed 0.9.[11] Given the high inertia, in our empirical analysis we will exploit the cross-country variation. Third, the correlation between ethnic inequality and the overall spatial inequality (constructed using luminosity across cells of 2.5 x 2.5 decimal degrees) is high, but far from perfect (around 0.5 to 0.6). This is useful since in our empirical analysis we will be able to condition on the overall degree of spatial inequality in development, when we examine the correlation between ethnic inequality and development. Figures 5a and 5b illustarte this plotting ethnic inequality against the overall degree of spatial inequality (see also Appendix Figures $2a - 2b$ that present the cross-country distribution of ethnic inequality conditioning on the overall degree of spatial inequality). A few interesting patterns emerge. On the one hand, the Democratic Republic of Congo (Zaire), Sudan, and Chad have much higher ethnic inequality as compared to the overall spatial inequality (which is also very high). On the other hand, USA, Australia, Canada, Russia and Chile, score low in ethnic inequality

[10] Appendix Figures $1a - 1b$ provide a graphical illustration of the distribution of ethnic inequality across the world using the GREG and the *Ethnologue* mapping of ethnic homelands, respectively; in Figure $1c$ we present the Gini coefficient of inequality within countries across pixels/boxes. In all figures darker colors indicate a higher degree of inequality. The countries with the highest between-ethnic-group inequality are Sudan, Chad, Afghanistan, Laos, and Myanmar (Gini index higher than 0.90).

[11] There are however some interesting changes between 1992 and 2009. We observe large negative changes in the ethnic Gini coefficient (decrease in ethnic inequality by more than 0.3) in Somalia, Sudan, Laos, Gambia, and Botswana. Instead we observe large positive changes in ethnic inequality (the Gini coefficient increases by more than 0.3) in Myanmar, Sierra Leone, and Yemen.

as compared to the overall degree of spatial inequality that is quite high. Costa Rica, Albania, Slovenia, Panama, and Rwanda score very high in ethnic inequality, while in contrast the overall degree of spatial inequality is very low.

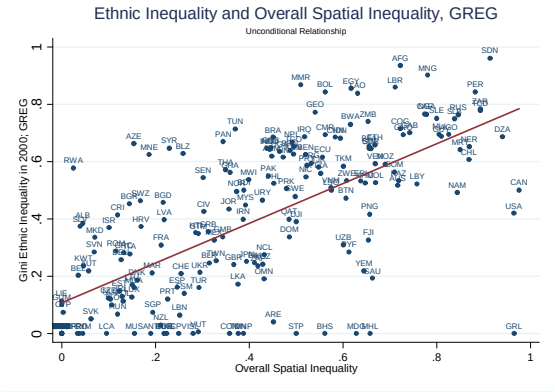


Figure 5*a*

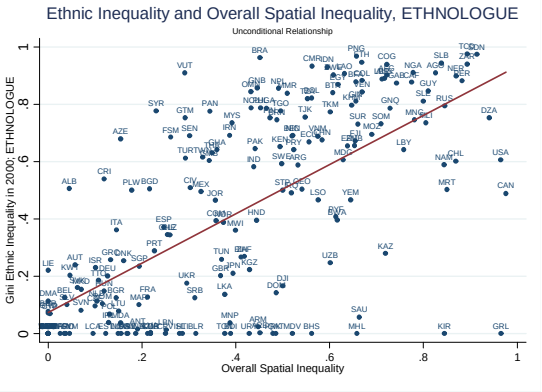


Figure 5*b*

2.3 Basic Correlations

2.3.1 Fractionalization

Table 2 - Panel *C* reports the cross-country correlation between the various ethnic inequality measures and spatial inequality with the widely used measures of ethnic, linguistic, and religious fractionalization that reflect the probability that two randomly chosen individuals will not be part of the same (ethnic, linguistic, or religious) group (data come from Alesina *et al.*. (2003)). The table also reports the correlation of ethnic inequality with the recently complied segregation measures by Alesina and Zhruravskaya (2011) that reflect the clustering of groups within countries.

There is a positive correlation between ethnic inequality and the linguistic and ethnic fractionalization measures $(0.45 - 0.58)$ though not with the religious fractionalization index.[12] Religious affiliation is (or was) in many countries not a free choice so it is quite likely to be endogenous. In fact the more religiously homogenous countries are the ones where freedom is less tolerated. Alesina *et al.*. (2003) note that this index of fractionalization, contrary to ethnic and linguistic fractionalization, shows no correlation with the level of per capita income. Thus from now on, we do not condition on religious fragmentation. Figures 6*a*− 6*d* give a graphical illustration of the positive association between ethnic inequality and ethnic-linguistic fragmentation. Since there is a regional component on fragmentation and ethnic inequality, we

[12] The linguistic index uses only languages to differentiate groups, the ethnic index uses language and other physical characteristics, like skin colors. See Alesina *at al.* (2003) for more details.

also report the correlation conditioning on continental fixed effects. The positive association is also present within continents.

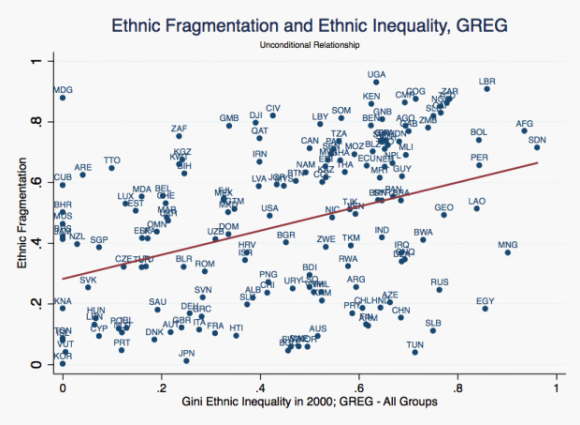


Figure 6*a*

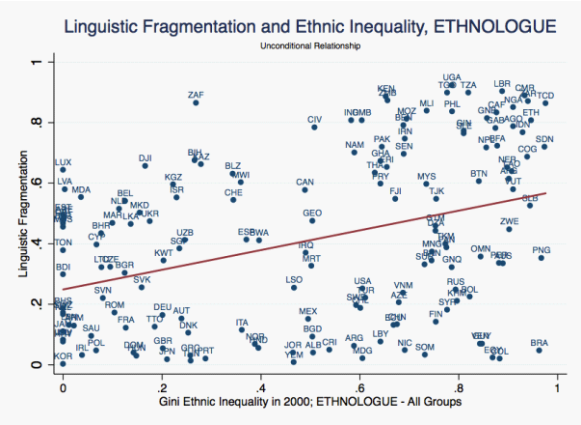


Figure 6*b*

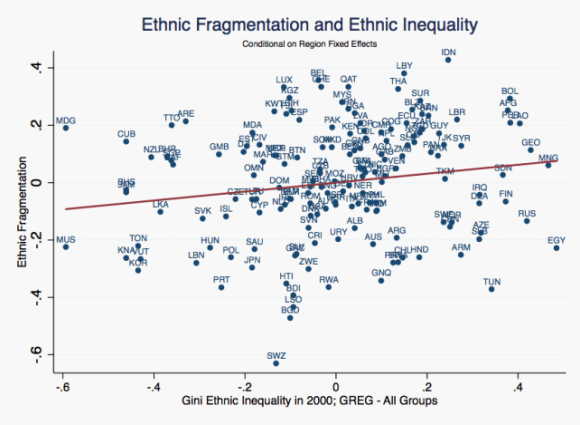


Figure 6*c*

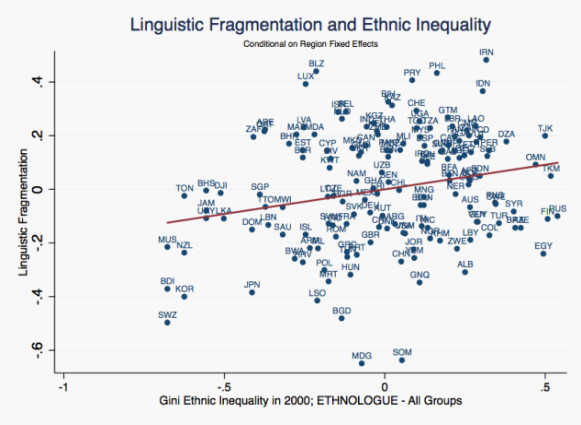


Figure 6*d*

Ethnic and linguistic fragmentation are also positively correlated with the overall spatial inequality index (0.45 and 0.35). The correlation between ethnic inequality and the ethnic and linguistic segregation measures is also positive but somewhat smaller ($0.20 - 0.35$). Ethnic inequality tends to go in tandem with segregation across ethnolinguistic groups. This is reasonable since more mixing of groups would naturally lead to a reduction of ethnic based inequality, which instead is more likely to survive when groups are geographically separated. Again there is no significant association with the religious segregation index.[13]

Income Inequality We then examine the association between ethnic inequality with income inequality, as reflected in the standard Gini coefficient. The income Gini coefficient is

[13] We also examined the association between the Montalvo and Reynal-Querol (2005) polarization and the ethnic inequality index. The correlation is positive, but small (around 0.10) and statistically insignificant. The same applies to the revised polarization measures of Desmet *et al.* (2011).

taken from Easterly (2004), who using survey and census data compiled from WIDER (UN's World Institute for Development Economics Research) constructs adjusted cross-country Gini coefficient indicators for more than a hundred countries over the period 1965 – 2000. There is a weak to moderate positive correlation between income inequality and ethnic inequality. Figures 7*a* and 7*b* illustrate this association using the GREG and the *Ethnologue* mapping of ethnic homelands, respectively. The correlation coefficient between ethnic inequality and economic inequality ranges from 0.25 and 0.40. Yet as Figures 7*c* and 7*d* show this correlation weakens further and becomes statistically insignificant once we simply condition on regional (continental) dummies.

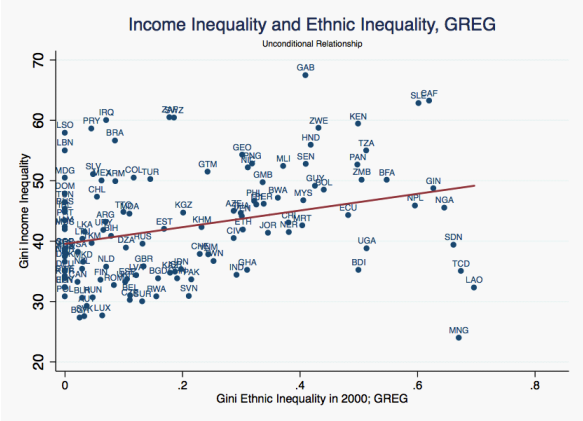


Figure 7*a*

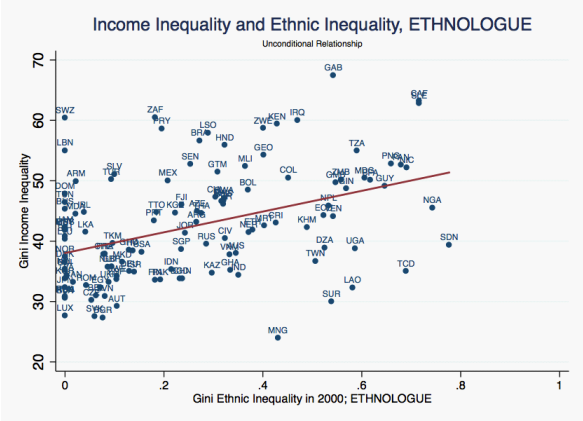


Figure 7*b*

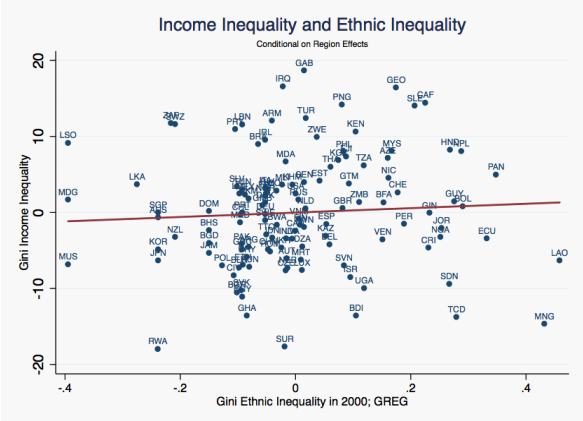


Figure 7*c*

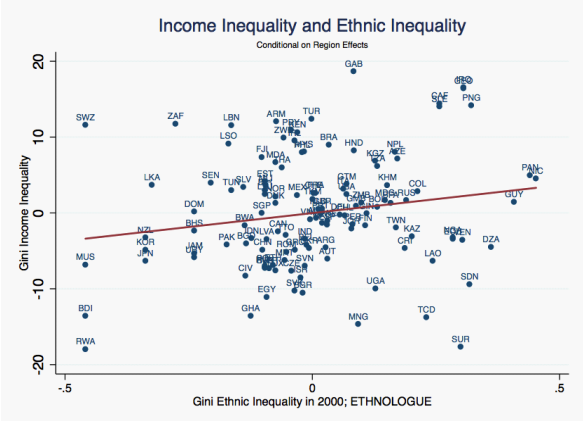


Figure 7*d*

Development Table 2 - Panel *D* reports correlations of our variables of interest (ethnic inequality and the overall degree of spatial inequality) with the log of per capita GDP in 2000 (using data from the latest vintage of the Penn World Tables), a rule of law and a control of corruption index (using data from World Bank's Governance Matters Database (Kaufman *et al.* (2008); see the Data Appendix for detailed variable definitions). Ethnic inequality is

strongly inversely related to GDP per capita. The correlation between our benchmark measures of ethnic inequality (that excludes tiny ethnicities/languages) and log per capita GDP in 2000 is -0.65 and -0.58 with the GREG and the *Ethnologue* mapping, respectively. Figures 8*a* - 8*d* illustrate this association.[14]

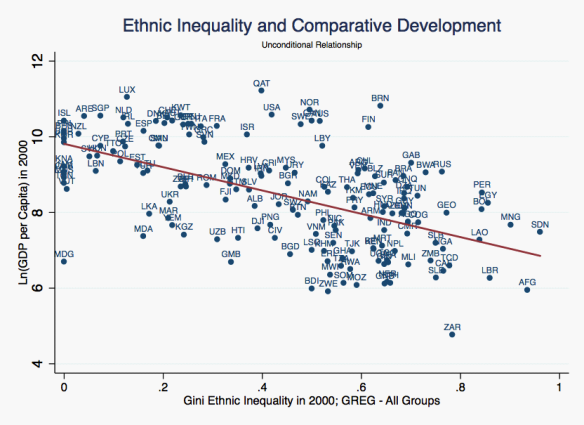


Figure 8*a*

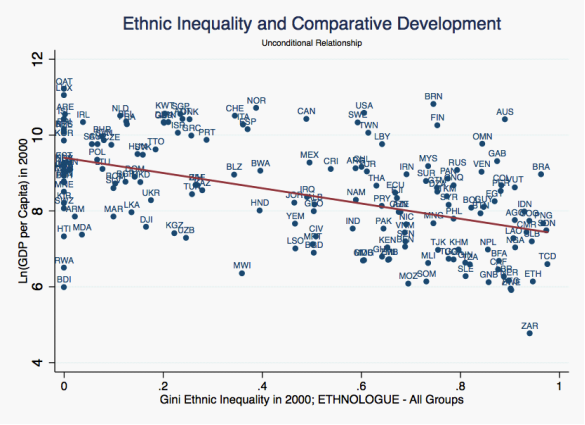


Figure 8*b*

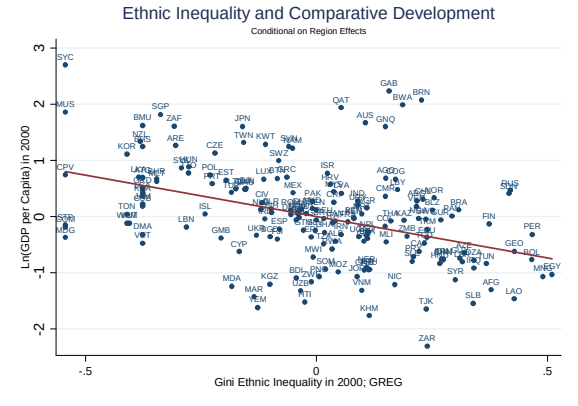


Figure 8*c*

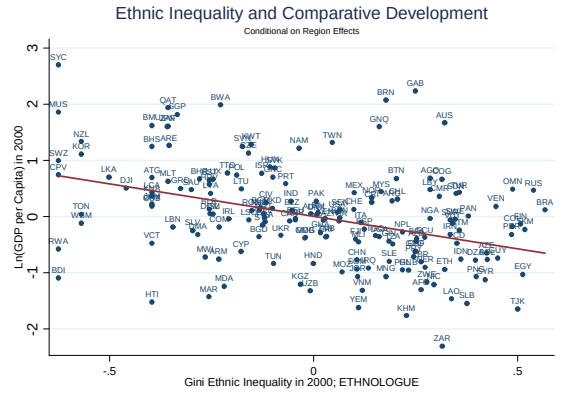


Figure 8*d*

Underdeveloped countries tend also to have wider income disparities across regions. Yet the correlation between the overall spatial inequality index and GDP per capita is smaller in magnitude (-0.44). This suggests that ethnic rather than overall spatial inequality correlates stronger with development. Given the strong correlation of economic and institutional development, it comes at no surprise that ethnic inequality is also strongly negatively correlated with the rule of law and the control of corruption indicators (correlations around $0.45 - 0.50$). Similarly there is significantly negative association between ethnic inequality and human capital measures, such as average years of schooling, enrollment rates, literacy, etc.[15]

[14] The correlation is somewhat weaker in 2009, 0.60 and 0.51 with the GREG and the Ethnologue maps respectively; the correlation is a bit stronger in 1992 (0.67 and 0.60 respectively).

[15] For brevity we do not report these correlations, but the graphs are available upon request.

Summary Overall these correlations clearly show that ethnic inequality is strongly negatively associated with economic development. Moreover, while ethnic inequality correlates positively with ethnolinguistic fragmentation and the overall degree of spatial inequality, the correlation is moderate allowing us to proceed into a regression analysis where we will be able to explore the role of ethnic inequality conditioning on these correlated features.

3 Ethnic Inequality and Development

3.1 Benchmark Estimates

In Table 3 we report LS regressions correlating ethnic inequality with economic development, as reflected in the log of per capita GDP in year 2000. In Panel *A* we use the ethnic inequality measure using the GREG (*Atlas Narodov Mira*) database, while Panel *B* reports otherwise identical specifications using the detailed mapping of languages of *Ethnologue*. In all specifications we include region fixed effects to (partly) account for broad continental differences in the spatial variation of ethnic (and spatial) inequality and development.[16]

The unconditional coefficient of the ethnic inequality index in column (1) is negative and highly significant. The estimates in column (2) show a highly negative association between development and the overall degree of spatial inequality. In column (3) we include both the ethnic inequality index and the measure reflecting the overall spatial inequality in per capita luminosity. The ethnic inequality index continues to enter with a highly significant estimate; moreover the coefficient on the ethnic inequality Gini index drops only slightly in absolute value. In contrast the estimate on the overall spatial inequality index drops by more than half in both permutations. This hints that the ethnic component of regional inequality is the relatively stronger correlate of underdevelopment.

In columns (4)-(6) we examine whether the strong negative association between ethnic inequality and GDP p.c. simply reflects ethnic heterogeneity; to do so we augment the specification with the log number of ethnic/linguistic groups of each country. In line with previous works (e.g. Alesina *et al.* (2003)) income per capita is significantly lower in countries with many ethnic (Panel *A*) and linguistic (Panel *B*) groups (column (4)); yet the estimates in (5) and (6) clearly show that it is ethnic inequality rather than ethnic-linguistic heterogeneity that correlates with underdevelopment. In columns (7)-(9) we further examine whether ethnic inequality or ethnolinguistic fractionalization correlate with underdevelopment, using the Alesina *et al.* (2003) index of ethnic (in Panel *A*) and linguistic (in Panel *B*) fragmentation. Development is

[16] We follow the World Bank regional classification and group countries in one of the following regions: East Asia and the Pacific, Latin America and the Caribbean, Middle East and North Africa, Europe and Central Asia, North America, South Asia, Sub-Saharan Africa and Western Europe.

lower in ethnically (or linguistically) heterogeneous countries. Yet once we include the ethnic inequality (Gini) index the coefficient on the fragmentation measures drops (in absolute value) considerably.

In columns (10)-(12) we examine whether the strong negative association between ethnic inequality and income per capita is driven by inequality in population density across ethnic homelands; to do so we construct Gini coefficients of population density combining the population estimates from the Gridded Population of the World' dataset (GPW, CIESIN, 2005) with the mapping of ethnic-linguistic groups. The Gini index of population density in 2000 that captures the unequal distribution of population across ethnic homelands enters with a negative and significant coefficient, implying that under-development goes in tandem with the unequal clustering of population across ethnic regions. Yet once we include in the specification the ethnic inequality index (in (11)) and the overall spatial inequality index (in (12)), the population density Gini coefficient index turns insignificant. In contrast the ethnic inequality measure retains its economic and statistical significance.

The most conservative estimate on the ethnic inequality index in Panel A of Table 3 (1.04) implies that a reduction in the ethnic Gini coefficient by 0.10 (approximately half a standard deviation) is associated by approximately 10% (0.10 log points) increase in per capita GDP. The standardized beta coefficient of the ethnic inequality index that measures the increase in standard deviation units of log GDP per capita to a one-standard-deviation increase in the ethnic Gini is around $0.22 - 0.35$. This is quite large and quite similar to the works examining the effect of institutions on long-run development (e.g. Acemoglu, Johnson, and Robinson (2001, 2002)).[17]

3.2 Sensitivity

In Table 4 we augment the specifications with additional control variables and we experiment with the ethnic Gini coefficient indicators that exclude tiny groups that constitute less than 1% of a country's population. In all specifications we control for the overall degree of spatial inequality in lights per capita and for country size, measured by the log of population and the log of land area. Conditioning on size is important, as ethnic heterogeneity, ethnic inequality, and the overall degree of spatial inequality are naturally larger in larger (in terms of population and land area) countries. We also control for the absolute value of latitude, because development is on average higher far from the equator (e.g. Hall and Jones (1999)) and because ethnolinguistic

[17] The standardized "beta" coefficient of the ethnic inequality index is twice as large as the analogous coefficient of the overall spatial inequality Gini index or the ethnic fragmentation measure. This in turn further illustrates that it is ethnic inequality rather than the overall degree of inequality or fractionalization the key correlate of under-development.

fragmentation is more pervasive in areas close to the equator (e.g. Michalopoulos (2012)). In many specifications we also control for ethnolinguistic fragmentation. The ethnic inequality index enters with a stable negative estimate; the coefficient is more than two standard errors lower than zero across all permutations. In (3), (6), (9), and (12) we condition on a rich set of geographic controls; to avoid concerns of self-selecting the conditioning set, we follow the baseline specification of Nunn and Puga (2011) and include an index of terrain ruggedness, distance to the coast, the percentage of arable land, an index of soil quality, and the percentage of tropical land (the Data Appendix gives detailed variable definitions; below we report results with an alternative set of geographic controls).

The negative correlation between ethnic inequality and income per capita remains quite strong. Thus while still an unobserved or omitted country-wide factor may jointly affect development and ethnic inequality, the estimates clearly point out that the strong correlation between under-development and ethnic inequality does not reflect (observable) differences in geographic endowments or continental disparities (captured by the region fixed effects). Overall, the correlation between GDP per capita and ethnic inequality is stronger and more robust than the correlation between GDP per capita and spatial inequality that becomes insignificant in many specifications. Moreover, in all specifications the usual ethnolinguistic fragmentation indicators enter with a statistically insignificant coefficient. This result hints that ethnic disparities in well being rather than ethnic fractionalization fosters (or is fostered by) underdevelopment.[18]

4 The Origins of Ethnic Inequality

4.1 Geography and History

Given the strong correlation between ethnic inequality and development, it is intriguing to examine the origins of contemporary differences in economic well-being across ethnic groups within countries. Since there are very few, if any, studies and theories on the determinants of ethnic inequality, we searched for potential correlates from recent works examining the deep causes of development that place an emphasis on geographical features and history. Yet we found very little evidence that contemporary differences in ethnic inequality are driven by geographic features, such as mean elevation, access to the sea, terrain ruggedness, soil quality, etc. (see Appendix Table 4 - Panel *A*).

[18] We performed several additional sensitivity checks. In the appendix we report some of these robustness checks. In Appendix Table 1 we drop from the estimation countries with just one ethnic group according to either the Atlas Narodov Mira (in (1)-(6)) or according to the *Ethnologue* database (in (7)-(12)) in otherwise identical specifications Table 4. In Appendix Table 2 we report specifications using the ethnic inequality indicators that exclude from the estimation capital cities to account for the extreme (in some cases) values of luminosity in large metropolitan areas. The results are robust,

Likewise we found only weak evidence that ethnic inequality is related to historical features, related to legal tradition that has been transplanted via colonization (see La Porta *et al.* (1997, 1998)) or the conditions that European settlers were facing at the time of colonization, which have shaped post independence development (Acemoglu, Johnson, and Robinson (2001)). There is some weak evidence that ethnic inequality is somewhat lower in British common law countries and in countries with high settler mortality, yet the estimates are not always statistically significant. In the same vein we found weak (and in general insignificant) correlations between ethnic inequality and institutions (executive constraints) in the initial years after independence (see Acemoglu *et al.* (2008)), state antiquity (see Bockstette, Chanda and Putterman (2002)), and the share of Europeans at the time of colonization (see Hall and Jones (1999) and Easterly and Levine (2010)). There is also an insignificant association between ethnic inequality and state artificiality, as reflected in the percentage of the population that is partitioned by the national border and an index of border straightness (see Appendix Table 4 - Panel *B*).

4.2 Unequal Distribution of Geography

Yet, geography plays a major role in explaining contemporary ethnic inequality. What matters, however, for ethnic inequality is not the level of geographic endowments at the country level. In contrast conceptually it should be the ethnic-specific inequality in the distribution of geographic features which should matter.

4.2.1 Data on Inequality in Geography

The construction of the inequality in geographic endowments measures across ethnic (linguistic) areas is analogous to the compilation of the ethnic inequality indicators. First, we obtain geo-referenced data on elevation, land's suitability for agriculture (land quality), and presence of water bodies (lakes, rivers, and other streams). We also estimate the distance of an ethnic (linguistic) region to the closest sea coast.[19] Second,we construct for each ethnic (linguistic) area the mean value for each of the four geographic measures. Third, we aggregate the data at the country-level, so as to construct Gini coefficients -for GREG and for *Ethnologue*- reflecting inequality in elevation, in land quality, in water access, and distance to the sea across ethnic (linguistic) homelands for each country. Exactly as we did for the ethnic inequality measures, we construct the geographic Gini coefficients including all ethnicities (languages) in each country and also excluding tiny ethnic (linguistic) areas that make up less than one percent of a country's population. Appendix Table 5 reports the values (and summary statistics) of the

[19] This is done by calculating the distance to the closest coast from each point within an ethnic homeland and then averaging across all these points.

ethnic inequality in geography Gini coefficients for all countries using the spatial distribution of ethnicities by GREG. In order to isolate the effects of inequality in geography across ethnic homelands from the overall degree of inequality in geographic endowments, we also split the world (and countries) into pixels (artificial regions) of 2.5 x 2.5 decimal degrees, and exactly as we did when we calculated the overall degree of spatial inequality in luminosity per capita, we estimate for each country a Gini coefficient that reflects spatial inequality in elevation, land quality, presence of lakes and rivers, and distance to the sea.

Appendix Table 6 gives the correlation of the geographic inequality Gini coefficients across ethnic homelands (using the GREG maps) and across pixels. The table also gives the correlation of the inequality measures with the level of elevation, land quality, presence of water, and distance to the sea. There are some interesting patterns. First, as expected there is a positive -though not perfect- association between the ethnic component of inequality in all geographic features and the overall degree of inequality across random pixels (around 0.7). This pattern is similar to the correlation of the benchmark ethnic inequality indicators and overall degree of spatial inequality index when we used luminosity per capita. Second, all ethnic inequality in geographic endowments indicators are positively correlated. The same applies for the Gini coefficients that were constructed based on random pixels rather than ethnic homelands. This suggests that there may be a common factor of inequality in geographic endowments, at least according to these four dimensions. Third, there is no systematic association between inequality in geographic endowments -either across ethnic homelands or across artificial boxes- and the level of elevation, land quality, access to the sea, and presence of water bodies. This is useful as it shows that the Gini coefficients along these four dimensions do not capture level effects of geography.

4.2.2 Preliminary Evidence

In Table 5 we explore the association between contemporary ethnic inequality, and the four measures of inequality in geographic endowments across ethnic (linguistic) homelands using both the GREG (in (1)-(6)) and the *Ethnologue* (in (7)-(12)) mapping of ethnic (linguistic) groups. Columns (1), (4), (7), and (10) give unconditional specifications. The ethnic Gini coefficients in geographic endowments enter with positive estimates suggesting that ethnic-specific differences in endowments translate into a higher degree in contemporary disparities in ethnic development. In (2), (5), (8), and (11) we condition on the overall degree of spatial inequality in geographic endowments with the four Gini coefficients in elevation, land quality, presence of water bodies, and distance to the sea coast based on random pixels. In columns (3), (6), (9), and (12) we also condition for the mean value of elevation, land quality, distance

to the sea coast, and water bodies. In all these specifications all ethnic Gini coefficients enter with positive estimates suggesting that ethnic inequality in development is partly explained by an unequal distribution in geography.

Of the four ethnic inequality in endowment measures, the Gini coefficient in elevation and the Gini coefficient in water access appear the stronger correlates of ethnic inequality. Yet in many specifications the Gini coefficient in distance to the sea coast and the Gini coefficient in land's quality for agriculture also enter with statistically significant positive estimates. Overall the message from Table 5 is that exogenous differences in geography across ethnic regions have long-lasting effects.

4.2.3 A Composite Index

We have not identified a clearly dominant geographic feature "leading" the correlation between ethnic-specific income inequality and inequality in endowments across ethnic homelands (though elevation and presence of water bodies seem to be somewhat more important). Moreover, there is multi-collinearity between geographic endowments. In order to circumvent these problems we aggregate the four indexes of ethnic inequality in geographic endowments via a principal component (factor) analysis. The use of factor analysis techniques looks appealing because we have many variables (Gini coefficients) that aim at capturing the same concept (with some degree of noise), in our application inequality in geographic endowments.

Table 6 gives the results of the principal component analysis. We report results both with the *Ethnologue* and the GREG mapping of linguistic/ethnic groups, using either all areas in each country and also excluding tiny ethnolinguistic cleavages. The first principal component explains more than half of the common variance of the four measures of ethnic inequality in geographic endowments. The second principal component explains around 20% of the total variance, while in total the third and fourth principal components explain a bit less than a fourth of the total variance. The eigenvalue of the first principal component is greater than two (one being the rule of thumb), while the eigenvalues of the other principal components are less than one. Thus we focus on the first principal component.

In Figures $9a - 9d$ we plot the ethnic inequality indicators against the first principal component of ethnic-specific inequality in geographic endowments. There is a remarkably strong positive association. As geographic inequality is (to a first-approximation) exogenous these graphs suggest that differences in geography explain a sizable portion of contemporary differences in development across ethnic (linguistic) homelands.[20]

[20] A possible source of endogeneity of geographic endowments may have to do with the fact that in ancient time stronger (i.e. more developed) ethnic groups battled and conquered better lands. This hypothesis is hard to test with available data but should be kept in mind. If this were the case, current ethnic inequality would

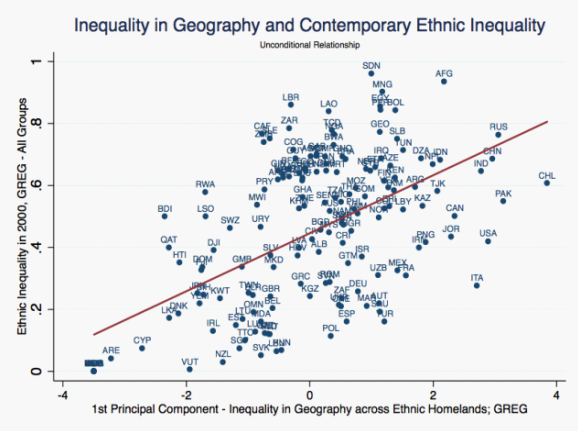


Figure 9*a*

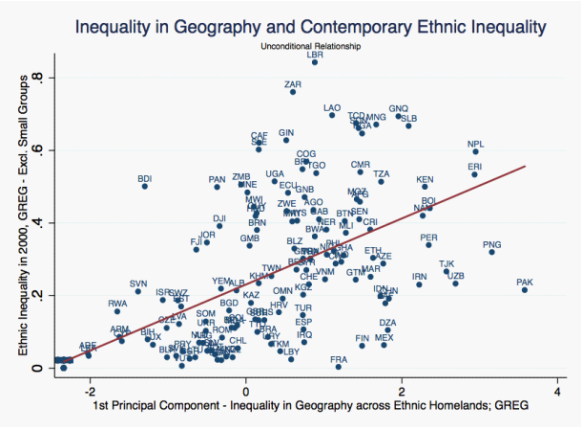


Figure 9*b*

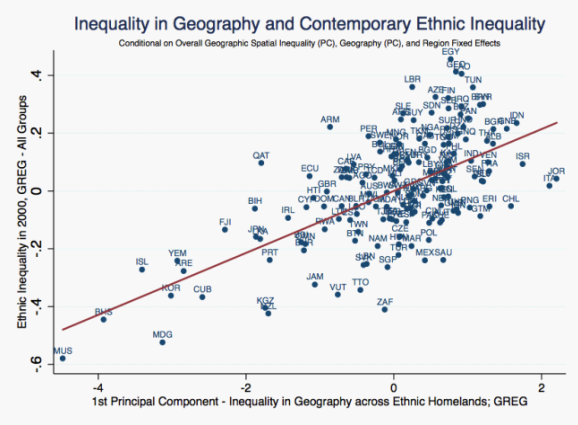


Figure 9*c*

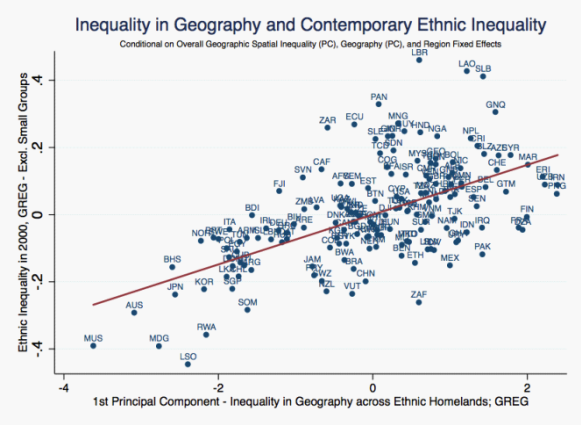


Figure 9*d*

In Table 7 we formally examine the effect of ethnic-specific geographic inequality, as captured in the composite index of inequality in geographic endowments across linguistic homelands, on contemporary ethnic inequality in income per capita. Columns (1), (4), (7), and (10) show that the strong correlation illustrated in Figures $9a - 9d$ is not driven by continental differences (absorbed by the region fixed effects). In all permutations the composite index of ethnic differences in endowments enters with a positive and highly significant coefficient. In (2), (5), (8), and (11) we control for the overall degree of spatial inequality in geographic endowments augmenting the specifications with the first-principal component of the Gini coefficients in geography when we use artificial pixels rather than ethnic homelands (Table 6 - Panel *E* gives the factor loadings). This has little effect on the coefficients on the ethnic inequality in geographic endowments index. In columns (3), (6), (9), and (12) we also control for the level effects of the four geographical features, augmenting the specification with mean elevation, the

be due not only to geographic endowments, but also to other types of endowments of ethnic groups, possibly genetic.

average degree of land quality, distance to the sea coast, and the percentage of each country's area by water.[21] Since inequality in geographic endowments across ethnic homelands (or even across random pixels) is uncorrelated with the mean values of geography, this has little effect on our results. The estimate on the ethnic inequality in geography index in column (3) implies that conditional on region fixed effects, the overall degree of spatial inequalities in geography, and level differences in geography across countries, a one-standard-deviation increase in ethnic inequality in geography (1.5 points) translates into an 11 percentage points increase in the ethnic inequality index (approximately half a percentage deviation; see Table 1).

5 Ethnic Inequality in Geographic Endowments and Contemporary Development

5.1 LS Estimates

Given the strong positive association between ethnic-specific income inequality and inequality in geographic endowments across ethnic homelands, it is interesting to examine whether ethnic-level differences in geography are systematically linked to contemporary development. This is useful because the strong negative correlation between ethnic inequality and development shown earlier may (partially at least) be driven by reverse causation. While endogeneity due to omitted variables cannot get eliminated, since geography is predetermined examining the effect of inequality in geographic endowments across ethnic homelands in development is useful in sorting out the direction of causation.

Table 8 reports estimates regressing log per capita GDP in 2000 on the composite index capturing inequality in geographic endowments across ethnic homelands, conditioning always on continental fixed effects. The coefficient on the ethnic inequality in geographic endowments index is negative across all permutations. The estimate is statistically significant at standard confidence levels (usually at the 99% level). This suggests that countries where ethnic groups differ considerably in the degree of their homeland's geographic endowments are less developed today. The estimate in column (3) implies that a one-standard-deviation increase in geographic inequality across ethnic homelands is associated with a lower degree of income per capita by approximately 0.18 standard deviations, approximately 1.4 log points). Given that the composite index reflecting inequality in geographic endowments is exogenous, the estimates in Table 8 are not driven by reverse causation. While still the correlation between development and inequality in geographical endowments across ethnic homelands may be driven by some unobserved characteristic, the fact that the correlation retains significance once we control for

[21] The results are similar if instead of using the four mean values of geography, we augment the specification with the first (and also the second) principal component of geography in levels (results not shown).

the overall degree in spatial inequality in geography and the level effects of geography suggests that causality runs from ethnic-specific inequality in geography to economic development at the country level.

5.2 Two Stage Least Squares Estimates

Given the strong positive effect that inequality in geographic endowments across ethnic homelands exerts on contemporary ethnic inequality (Table 7) and the negative association between geographic inequality and development (Table 8), it is intriguing to combine the two sets of results into an instrumental variables two-stage approach that under instrument validity will identify the one-way effect of ethnic inequality on contemporary development.

Formally, identification requires that (a) exogenous differences in geographic endowments across ethnic homelands are significantly associated with ethnic inequality (i.e. there is a strong first-stage fit); and (b) that conditional on other characteristics, inequality in geographic endowments across ethnic homelands affects development only via its effect on ethnic inequality (i.e. the exclusion restriction is satisfied). The results in Table 8 show that there is strong positive association between geographic heterogeneity across ethnic homelands and contemporary ethnic-specific economic inequality, as reflected in satellite light density per capita. Thus the first assumption for instrument validity seems to hold. How about the second assumption? While ethnic-specific inequality in geographic endowments may affect a country's development via channels beyond ethnic inequality (e.g. trade, financial development), in many specifications we condition on the overall degree of spatial inequality in geography (as well as the level effects of geography. By doing so, we purge from the ethnic-specific inequality measure the purely spatial component and therefore mitigate concerns that our ethnic-specific geographic inequality index captures non-ethnicity specific channels. Moreover, intuitively it is reasonable to assume that differences in geographic endowments across ethnic homelands affects development primarily by shaping differences in economic performance across ethnic groups.

Table 9 reports 2SLS regressions associating inequality in geography across ethnic homelands with ethnic inequality in a first-stage model and the component of ethnic inequality explained by geographic disparities across ethnic regions with log per capita GDP in 2000. The specifications follow Tables 7 and 8 that in some sense report the corresponding first-stage estimates and the corresponding reduced-form estimates of the 2SLS estimates. The 2SLS coefficient on the ethnic inequality index in the simple specifications in (1), (4), (7), and (10), is negative and highly significant. This implies that the component of contemporary ethnic-specific income inequality shaped by inequality in geographic endowments across ethnic homelands is significantly inversely related to income per capita across countries. Of course

inequality in geographic endowments may affect development via other than ethnic inequality channels (such as trade, for example). In this case the exclusion restriction would be violated. To (partly at least) account for this, in all other columns we condition on the overall degree of spatial inequality in geographic endowments as well as the level effect of geography. The 2SLS coefficient on ethnic inequality retains its economic and statistical significance, implying that the exogenous component of ethnic inequality driven by geographic differences across ethnic homelands is inversely related to contemporary development.[22]

6 Analysis with Micro Data within Sub-Saharan African Countries

In this section we take a more micro approach that explores within country (across region) variation on living conditions and access on public goods to examine the association between ethnic inequality and development. By doing so we also use "hard" measures of development as opposed the proxy of luminosity. Specifically we provide evidence that the uncovered negative relationship between ethnic-specific economic inequality and economic performance also obtains across districts within Sub-Saharan African countries. Our focus on Africa is natural. First, Africa is by far the most ethnically and linguistically heterogeneous part of the world. Second, ethnic patronage politics are a key factor of recent economic performance across many African countries. Third, it seems that a considerable portion of Africa's growth tragedy in the period $1960-1990$ can be attributed to ethnic fractionalization (Easterly and Levine (1997)).[23] Fourth, we have micro data on both the ethnic identity and economic conditions across many African regions by the Afrobarometer Surveys. These data on public goods, living conditions, and ethnic identity are useful for many reasons. To begin with unlike the cross-country analysis where as a measure of economic well-being we use satellite light density at night, in the African sample we employ survey-level data on self-reported economic living conditions and access to public goods. Then instead of assigning parts of a country to one or more groups via the use of linguistic/ethnic maps that clearly contain some degree of subjectivity and er-

[22] While we cannot formally test the exclusion restriction, to investigate whether geographic inequality affects development through other channels we run LS regressions of log per capita GDP on ethnic inequality and the composite index that captures inequality in geographic endowments across ethnic homelands. The results are presented in Appendix Table 7 reports these results. Across all permutations the coefficients on the ethnic inequality index is negative and highly significant, while in contrast the composite index capturing inequality in geography across ethnic regions enters with a small and statistically indistinguishable from zero estimate. Thus while not definite, these results suggest that -conditional on the overall degree of spatial inequality within a country and- differences in endowments across ethnic homelands affect development primarily via shaping ethnic inequality.

[23] Note, however, that our cross-country results showing a negative association between ethnic inequality and development are not driven by a particular region of the world (see the discussion above and the results reported in Appendix Table 3).

ror, we use self-reported data on ethnic identity minimizing measurement error. In addition, since we are able to focus within countries and exploit variation across districts this allows us to control for country-specific characteristics assuaging, albeit not entirely resolving, concerns that the pattern uncovered across countries may be driven by omitted country-level unobservables. Also, the use of a richer dataset allows to shed some light on the mechanism via which ethnic inequality affects development. Finally, as we explain below the data allows us to even account for ethnicity fixed effects (since groups are present in more than one region) and thus further account for the concern that the correlation between ethnic inequality and economic performance is driven by a specific group being systematically discriminated or favored vis a vis other groups in the country.

6.1 Data

We follow Nunn and Wantchenkon (2011) and use the individual-level survey data from the third round of the Afrobarometer survey, which was conducted in 2005. These surveys are based on interviews conducted in a random sample of either $1,200$ or $2,400$ individuals of voting age in 17 Sub-Saharan African countries.[24] Out of the $21,822$ respondents $20,751$ have a clearly identified ethnic identity. These individuals reside in 1265 districts across the 17 countries. In each district there are on average 2.77 ethnic groups ranging from districts which are ethnically homogeneous to those that are extremely diverse. An average ethnic group resides across 30 districts within a country. The Hausa in Nigeria, for example, may be found in 82 districts.

The ethnic inequality index at the regional level is constructed using individual's responses to the quality of their living conditions. The $1-5$ ordered index of living conditions reflects respondent's view of their present living conditions which can be: (i) very bad, (ii) fairly bad, (iii) neither good nor bad, (iv) fairly good, or (v) very good. Using this data we construct the average living conditions for each ethnic group in a given district. Then we calculate the Gini coefficient of living conditions across ethnicities in each district; moreover to isolate the ethnic component of inequality from the overall degree of income inequality we also calculate the overall Gini coefficient in living conditions across all respondents within each district.

As explanatory variables capturing development we employ the percentage of urban households in a district, and measures capturing local public goods provision, namely access to clean piped water, access to a sewage system, and access to electricity;[25] the Data Appendix

[24] These countries are: Benin, Botswana, Ghana, Kenya, Lesotho, Madagascar, Malawi, Mali, Mozambique, Namibia, Nigeria, Senegal, South Africa, Tanzania, Uganda, Zambia, and Zimbabwe. The dataset is available at www.afrobarometer.org

[25] We also estimated specifications using the average living conditions in a region as the dependent variable finding similar results. Since we use the living conditions index to construct the ethnic inequality and the overall

gives detailed variable definitions, while Appendix Table 8 reports summary statistics.

6.2 Results

In Table 11 we examine the within country correlation between ethnic inequality and regional development, reflected in urbanization (in columns (1)-(2)) or in public goods (in columns (5)-(8)). In all specifications we condition on the log number of ethnic groups and on the average level of living conditions. Columns (1) and (2) show that districts with high between ethnic group inequality are systematically less urbanized. In contrast to the ethnic inequality Gini coefficient that is negatively associated with urbanization, ethnic diversity enters with a positive estimate. This most likely reflects the fact that urban places are more likely to attract members of different ethnic groups. The correlation between ethnic inequality and urbanization is not driven by differences in average living conditions across districts (which as expected enters always with a positive sign). In column (2) we include in the specification the Gini in living conditions across all respondents in the region to isolate the effect of ethnic inequality from the overall degree of income inequality. The coefficient on the ethnic Gini is remarkably stable and retains its statistical significance; in contrast the overall Gini inequality index enters with a small and statistically indistinguishable from zero estimate.

In columns (3)-(8) we explore the relationship between ethnic inequality and various proxy measures of local public goods provision. In (3) and (4) the dependent variable is the share of respondents within a district that reports having access to piped water in their village; in (5) and (6) we look at access to a sewage system, while in (7) and (8) we focus on access to an electricity grid. The ethnic inequality Gini coefficient is always negative and highly significant. It suggests that ethnically unequal districts experience systematically lower provision of public goods. The results in even-numbered columns where we control on the overall Gini coefficient, further show that, unlike the overall inequality in well-being, it is the ethnic-specific component of inequality that correlates negatively with access to public goods.

The nice feature of the results in Table 10 is that we can account for common to all regions inside a country factors with the inclusion of country fixed effects. Yet we can explore further the richness of the Afrobarometer Surveys, since we have individual level data. In Table 11 we report similar to Table 10 empirical specifications associating ethnic inequality at the region level with development, running, however, individual level regressions. This offers two main advantages. First, we now condition on a comprehensive set of individual characteristics.[26] Second, since members of the same ethnic group are present in different

income inequality Gini coefficients, we do not report these models.

[26] Following Nunn and Watchekon (2011) we are including in the specifications the respondent's age and age squared, a gender indicator, 5 living-conditions fixed effects, 10 education fixed effects, 18 religion fixed effects,

districts within a country and thus are exposed to different degrees of district-level ethnic inequality, we can include in the specification ethnicity fixed effects. This is useful as it ensures that the strong negative association between ethnic inequality and public goods is not driven by a single (or a couple) of ethnic groups.[27] The ethnicity fixed effects specifications therefore examine whether individuals from the same ethnic group perform better or worse if they reside in regions characterized with a high degree of ethnic group inequality.

The individual-level regression analysis in Table 11 strengthens both our cross-country results (where we used luminosity and geography at the ethnic homeland level to construct country-level measures of ethnic inequality) and the within country results (in Table 10) showing a strong inverse relationship between ethnic inequality and development and public goods across Sub-Saharan African regions. Within countries and conditional on many individuals characteristics, members of the same ethnic group that are found in more ethnically unequal districts are less likely to live in an urban location, are less likely to have access to clean water, are also less likely to have a sewage system, and are less likely to have electricity. Importantly this correlation is neither driven by the overall district-level inequality nor by the average living conditions in the district. The economic magnitudes of the estimates imply a considerable economic effect. Increasing ethnic inequality by one standard deviation decreases access to piped water by 4.8%, access to a sewage system by 5.2% and the presence of an electricity grid by 3.2%.

7 Conclusions

Our thesis in this paper is that ethnic differences in economic performance rather than the degree of fractionalization impedes economic development. As Chua forcefully argues the presence of economically dominant ethnic minorities has led to ethnic hatred, conflict, institutional capture, poor public goods provision, and has undermined public support for free market institutions in many parts of the world. While a large literature has examined the effects of various aspects of fractionalization (such as fragmentation, polarization, segregation) on economic performance, there is little -if any- works studying the inter-linkages between ethnic inequality and economic development.

and 25 occupation fixed effects.

[27] There are 252 ethnic groups whose members are located in more than one district. For example, the Pular in Senegal are found in 23 out of the 31 country's districts. In the district of Matam where the Pular coexist with the Soninke, the Wolof and the Mandinka ethnic inequality in living conditions is only 0.017 compared to a 10 times larger Gini ethnic inequality index of 0.114 in the district of Sedhiou where the Pular reside with the Wolof, the Mandinka, the Manjack, the Diola and the Bambara. In Sedhiou district all Pular respondents report having no access to local public goods like electricity, piped water and sewage system, whereas the Pular in the ethnically more equal Matam although they report no access to a sewage system, 72% of them have access to an electricity grid and a similar number reports having access to piped water.

This paper is a first effort to fill this gap. Our analysis proceeds in five steps. First, combining anthropological maps on the spatial distribution of ethnic groups within countries in the 1960s and in the 1990s with satellite images on light density at night we construct ethnic inequality Gini coefficients that reflect inequality in well-being (and public good provision) across ethnic groups within the same country. Interestingly ethnic inequality is weakly correlated with the standard measures of income inequality and only modestly correlated with ethnolinguistic fractionalization. Second, we show that the newly constructed proxy of ethnic inequality is strongly negatively correlated with the level of per capita GDP across countries. The correlation retains its economic and statistical significance when we condition on the overall degree of spatial inequality in development (which is also negatively associated with economic development). Moreover, once we include in the empirical specification both the ethnic inequality index and the widely-used ethnolinguistic fragmentation indicators, the latter loses significance. This result suggests that it is inequality across ethnic groups that is correlated with poor economic performance rather than fractionalization. Third, we examine the roots of contemporary differences in well being across ethnicities. Using detailed data on geographic endowments across ethnic homelands in land's suitability for agriculture, elevation, proximity to the coast, and presence of water bodies, we construct fine and composite indicators of ethnic inequality in geography. These indexes are critical determinants of contemporary ethnic inequality. Fourth, we show that inequality in geographic endowments across ethnic homelands is also inversely related to contemporary development. This result, therefore, implies that the negative correlation between contemporary ethnic inequality and development is not driven by reverse causation. Fifth, we show a similar negative association between ethnic inequality and development exploring solely within country across regions variation in 17 Sub-Saharan countries using micro-level data on well-being and public goods provision.

Our results suggest that works assessing the consequences and origins of the societal structure should focus on the unequal distribution of income and endowments across ethnic groups. Moreover, future theoretical and empirical research should also establish channels through which ethnic inequality and development are linked.

8 References

TO BE ADDED

9 Data Appendix

9.1 Main Data

Income level: Log of per capita GDP at PPP (Chain Index) in 2000. *Source: Penn World Tables, Edition 7. Heston, Summers, and Aten (2011).*

Rule of Law: The rule of law index is "capturing perceptions of the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence." The standardized index which corresponds in 2000 ranges from -2.5 to $+2.5$ with higher values indicating better functioning institutions. *Source: World Bank Governance Matters Indicators Database (Kaufman, Kraay, and Mastruzzi (2005)). available at: http://info.worldbank.org/governance/wgi/index.asp*

Control of Corruption: The control of corruption index is "capturing perceptions of the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as capture of the state by elites and private interests." The standardized index which corresponds in 2000 ranges from -2.5 to $+2.5$ with lower values indicating a higher degree of corruption. *Source: World Bank Governance Matters Indicators Database (Kaufman, Kraay, and Mastruzzi (2005)). available at: http://info.worldbank.org/governance/wgi/index.asp*

Income Inequality. Adjusted Gini coefficient index averaged over the period $1965-1998$. *Source: Easterly (2007); based on WIDER.*

Ethnic/Linguistic/Religious Fragmentation: Index of ethnic/linguistic/religious heterogeneity, constructed as one minus the Herfindahl index of the share of the largest ethnic/linguistic/religious groups. It reflects the probability that two randomly selected individuals follow different ethnic/linguistic/religious groups. *Source: Alesina et al. (2003).*

Ethnic/Linguistic/Religious Seggregation: Index ranging from zero to one capturing ethnic/linguistic.religious seggregation (clustering) within countries. If each region is comprised of a separate group, then the index is equal to 1, and this is the case of full segregation. If every region has the same fraction of each group as the country as a whole, the index is equal to 0, this is the case of no segregation. The index is increasing in the square deviation of regional-level fractions of groups relative to the national average. The index gives higher weight to the deviation of group composition from the national average in bigger regions than in smaller regions." *Source: Alesina and Zhuravskaya (2012).*

Light Density at Night: Light density is calculated averaging light density observations across pixels that fall within each country in 1992, in 2000, and in 2009.
Source: Available at http://www.ngdc.noaa.gov/dmsp/global_composites_v2.html.

Water Area: Total area covered by rivers or lakes in square kilometers. *Source: Con-*

structed using the "Inland water area features" dataset from Global Mapping International, Colorado Springs, CO, USA. Global Ministry Mapping System.

Elevation: Average elevation in kilometers. *Source: National Oceanic and Atmospheric Administration (NOAA) and U.S. National Geophysical Data Center, TerrainBase, release 1.0 (CD-ROM), Boulder, Colorado. http://www.sage.wisc.edu/atlas/data.php?incdataset=Topography*

Land Suitability for Agriculture: Average land quality for cultivation within each country. The index is the product of two components capturing the climatic and soil suitability for farming. *Source: ?; Original Source: Atlas of the Biosphere.*
Available at http://www.sage.wisc.edu/iamdata/grid_ data_ sel.php.

Distance to the Sea Coast: The geodesic distance from the centroid of each country to the nearest coastline, measured in 1000*s* of km's. *Source: Global Mapping International, Colorado Springs, Colorado, USA. Series name: Global Ministry Mapping System. Series issue: Version 3.0*

Population: Log population in 2000. Source:

Land Area: Log surface area. *Source: Nunn and Puga (2011).*

9.2 Data used in the Sensitivity Analysis reported in the Appendix Tables

Soil quality: Percentage of each country with fertile soil. *Source: Nunn and Puga (2011).*

Terrain Ruggedness: The terrain ruggedness index quantifies topographic heterogeneity. The index is the average across all grid cells in the country not covered by water. The units for the terrain ruggedness index correspond to the units used to measure elevation differences. Ruggedness is measured in hundreds of metres of elevation difference for grid points 30 arc-seconds (926 metres on the equator or any meridian) apart. *Source: Nunn and Puga (2011).*

Tropical: Percentage of each country with tropical climate. *Source: Nunn and Puga (2011).*

Desert: The percentage of the land surface area of each country covered by sandy desert, dunes, rocky or lava flows. *Source: Nunn and Puga (2011).*

Latitude: Absolute latitude is expressed in decimal degrees, for the geographical centroid of the country. *Source: Nunn and Puga (2011).*

Common Law: Indicator variable that identifies countries that have a common law legal system. *Source: Nunn and Puga (2011).*

European Descent: *Source: Nunn and Puga (2011).*

Settler Mortality: *Source: Acemoglu, Johnson, and Robinson (2001).*

State Antiquity: Normalized State Antiquity Index in 1950, using a 1% discount rate.

Source: Bockstette, Chanda, Putterman (2002) and Putterman (2007).

Border Straightness Index: The 0-1 index reflects how straight national borders are. *Source: Alesina, Easterly, and Matuszeski (2011).*

Ethnic Partitioning: Percentage of the population of a country that belongs to partitioned ethnic groups. *Source: Alesina, Easterly, and Matuszeski (2011).*

Regional Fixed Effects: There region constants correspond to: South Asia, East Asia and Pacific, Latin America and the Caribbean, North America, Western Europe, Europe and Central Asia, Middle East and Northern Africa, Sub-Saharan Africa. The classification follows World Bank's World Development Indicators.

9.3 Afrobarometer Data

Living Conditions: Respondent's view of their present living conditions which can be: (i) very bad, (ii) fairly bad, (iii) neither good nor bad, (iv) fairly good, or (v) very good. For the district-level analysis responses are averaged across individuals within a district. *Source: 2005 Afrobarometer Surveys*

Urban Household: an indicator for whether the respondent lives in an urban location. For the district-level analysis responses are averaged across individuals within a district. *Source: 2005 Afrobarometer Surveys*

Access to an electricity grid: Individual response to the question on "whether in the enumeration area there is an electricity grid that most houses could access". For the district-level analysis responses are averaged across individuals within a district. *Source: 2005 Afrobarometer Surveys*

Access to piped water: Individual response to the question on "whether in the enumeration area there is a piped water system that most houses could access". For the district-level analysis responses are averaged across individuals within a district. *Source: 2005 Afrobarometer Surveys*

Access to sewage system: Individual response to the question on "whether in the enumeration area there is a sewage system that most houses could access". For the district-level analysis responses are averaged across individuals within a district. *Source: 2005 Afrobarometer Surveys.*

Table 2: Correlation Structure

Panel A: Ethnic Inequality Indicators (all ethnic areas)

	Ethnic Gini						Spatial Gini		
	GREG			Ethnologue					
Ethnic Gini 2009 (GREG)	1.0000								
Ethnic Gini 2000 (GREG)	0.9627*	1.0000							
Ethnic Gini 1992 (GREG)	0.9418*	0.9475*	1.0000						
Ethnic Gini 2009 (ETHN)	0.7499*	0.7379*	0.7429*	1.0000					
Ethnic Gini 2000 (ETHN)	0.7463*	0.7452*	0.7483*	0.9914*	1.0000				
Ethnic Gini 1992 (ETHN)	0.7477*	0.7441*	0.7816*	0.9602*	0.9640*	1.0000			
Spatial Gini 2009	0.6592*	0.6738*	0.6327*	0.6786*	0.6828*	0.6533*	1.0000		
Spatial Gini 2000	0.6620*	0.6962*	0.6400*	0.6745*	0.6896*	0.6533*	0.9673*	1.0000	
Spatial Gini 1992	0.6805*	0.7103*	0.6737*	0.6935*	0.7003*	0.6950*	0.9305*	0.9368*	1.0000

Panel B: Ethnic Inequality Indicators (excluding tiny ethnic areas)

	Ethnic Gini						Spatial Gini		
	GREG			Ethnologue					
Ethnic Gini 2009 (GREG)	1.0000								
Ethnic Gini 2000 (GREG)	0.9625*	1.0000							
Ethnic Gini 1992 (GREG)	0.9316*	0.9455*	1.0000						
Ethnic Gini 2009 (ETHN)	0.7551*	0.7560*	0.7573*	1.0000					
Ethnic Gini 2000 (ETHN)	0.7169*	0.7481*	0.7532*	0.9758*	1.0000				
Ethnic Gini 1992 (ETHN)	0.7166*	0.7469*	0.7756*	0.9381*	0.9552*	1.0000			
Spatial Gini 2009	0.4966*	0.5132*	0.5122*	0.5850*	0.5998*	0.5885*	1.0000		
Spatial Gini 2000	0.4989*	0.5371*	0.5230*	0.5910*	0.6180*	0.5999*	0.9673*	1.0000	
Spatial Gini 1992	0.5072*	0.5400*	0.5487*	0.5891*	0.6096*	0.6273*	0.9305*	0.9368*	1.0000

Panel C: Correlation with Measures of Ethnic-Linguistic-Religious Fragmentation & Segregation

Ethnic Gini 2000 - All (GREG)	1.0000								
Ethnic Gini 2000 - All (ETHN)	0.7452*	1.0000							
Overall Spatial Gini 2000	0.6962*	0.6896*	1.0000						
Ethnic Fragmentation	0.4397*	0.4685*	0.4631*	1.0000					
Linguistic Fragmentation	0.3932*	0.4080*	0.3423*	0.6885*	1.0000				
Religious Fragmentation	-0.0423	-0.0125	-0.0218	0.1629*	0.2881*	1.0000			
Ethnic Seggregation	0.2751*	0.4449*	0.1902	0.4813*	0.3705*	-0.0442	1.0000		
Linguistic Seggregation	0.2138*	0.3695*	0.1969	0.3945*	0.3056*	-0.0363	0.8422*	1.0000	
Religious Seggregation	0.2594*	0.2474*	0.1957	0.2502*	0.2957*	0.0896	0.2205	0.1276	1.0000

Panel D: Correlation with Development and Income Inequality

Ethnic Gini 2000 (GREG)	1.0000								
Ethnic Gini 2000 (ETHN)	0.7481*	1.0000							
Ethnic Gini 2000 - All (GREG)	0.7173*	0.6052*	1.0000						
Ethnic Gini 2000 - All (ETHN)	0.6521*	0.8235*	0.7452*	1.0000					
Overall Spatial Gini 2000	0.5371*	0.6180*	0.6962*	0.6896*	1.0000				
Income Inequality (Gini coeff.)	0.3010*	0.4177*	0.2576*	0.3238*	0.2877*	1.0000			
Log real GDP p.c. in 2000	-0.6432*	-0.5757*	-0.5272*	-0.4909*	-0.4604*	-0.3784*	1.0000		
Rule of Law in 2000	-0.5007*	-0.4944*	-0.4723*	-0.4443*	-0.3937*	-0.4021*	0.7970*	1.0000	
Control of Corruption in 2000	-0.4982*	-0.4679*	-0.4460*	-0.4203*	-0.3582*	-0.4057*	0.7381*	0.9409*	1.0000

Table 3a: Ethnic Inequality and Economic Development (in 2000), Atlas Naradov Mira

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Ethnic Gini	-1.4708***		-1.2514***		-1.2514***	-1.0175**		-1.4888***	-1.4102***		-1.9376***	-1.6757***
	(0.2490)		(0.3253)		(0.4154)	(0.4827)		(0.2650)	(0.3276)		(0.3320)	-0.3793
	-5.91		*-3.85*		*-3.01*	*-2.11*		*-5.62*	*-4.30*		*-5.84*	*-4.42*
Spatial Gini		-1.2624***	-0.3476			-0.3563			-0.1291			-0.4999
		(0.2999)	(0.3716)			(0.3771)			(0.3622)			(0.3758)
		-4.21	*-0.94*			*-0.94*			*-0.36*			*-1.33*
Log Number of Languages				-0.2985***	-0.0686	-0.0714						
				(0.0581)	(0.0931)	(0.0939)						
				-5.14	*-0.74*	*-0.76*						
Ethnic Fragmentation							*-0.5307*	*0.0009*	*0.019*			
							(0.3480)	(0.3428)	(0.3475)			
							-1.53	*0.00*	*0.05*			
Ethni Gini in Population Density										-0.6977**	0.8405**	0.9371**
										(0.2817)	(0.3900)	(0.3737)
										-2.48	*2.16*	*2.51*
adjusted R-squared	0.6510	0.6250	0.6510	0.6300	0.6500	0.6500	0.5960	0.6610	0.6590	0.5890	0.6590	0.6600
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	177	177	177	177	177	177	164	164	164	177	177	177

Table 3b: Ethnic Inequality and Economic Development (in 2000), Ethnologue

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Ethnic Gini			-0.9379***		-1.2834***	-1.0418**		-1.2665***	-1.0527***		-1.5685***	-1.3406***
	(0.2251)		(0.2741)		(0.3631)	(0.4463)		(0.2457)	(0.2865)		(0.3321)	-0.3524
	-5.14		*-3.42*		*-3.53*	*-2.33*		*-5.16*	*-3.67*		*-4.72*	*-3.80*
Spatial Gini		-1.2624***	-0.4219			-0.4098			-0.4361			-0.4915
		(0.2999)	(0.3531)			(0.3634)			(0.3505)			(0.3522)
		-4.21	*-1.19*			*-1.13*			*-1.24*			*-1.40*
Log Number of Languages				-0.1802***	0.0323	0.0249						
				(0.0462)	(0.0723)	(0.0742)						
				-3.9	*0.45*	*0.34*						
Linguistic Fragmentation							-0.5800*	-0.11	-0.1125			
							(0.3156)	(0.2749)	(0.2768)			
							-1.84	*-0.4*	*-0.41*			
Ethni Gini in Population Density										-0.6851**	0.6839*	0.7284*
										(0.2722)	(0.4071)	(0.4132)
										-2.52	1.68	1.76
adjusted R-squared	0.6456	0.6245	0.6463	0.6168	0.6439	0.6444	0.5952	0.6613	0.6621	0.5943	0.6513	0.653
Region Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	177	177	177	177	177	177	167	167	167	177	177	177

Table 4: Ethnic Inequality and Economic Development (in 2000)
Sensitivity Analysis. Controls and Alternative Measures of Ethnic Inequality

	Atlas Narodov Mira (GREG)						Ethnologue					
	All Ethnic Areas			Excl. Small Ethnic Areas			All Ethnic Areas			Excl. Small Ethnic Areas		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Ethnic Gini	-1.4885***	-1.8572***	-1.4785***	-1.1315***	-1.3194***	-1.1765***	-0.9968**	-1.3239***	-1.1608***	-0.7560**	-1.0002***	-0.7310**
	(0.4398)	(0.5039)	(0.5039)	(0.3700)	(0.3680)	(0.3425)	(0.4083)	(0.4108)	(0.4301)	(0.3138)	(0.3246)	-0.3102
	-3.38	*-3.69*	*-2.93*	*-3.06*	*-3.59*	*-3.43*	*-2.44*	*-3.22*	*-2.70*	*-2.41*	*-3.08*	*-2.36*
Spatial Gini	-0.8949**	-0.8610*	-0.8866**	-0.7205	-0.6471	-0.6928*	-0.8420*	-0.9951**	-0.9231**	-0.8392*	-0.9945**	-1.0443**
	(0.4353)	(0.4402)	(0.4039)	(0.4532)	(0.4655)	(0.4069)	(0.4511)	(0.4529)	(0.4090)	(0.4682)	(0.4649)	(0.4361)
	-2.06	*-1.96*	*-2.19*	-1.59	*-1.39*	*-1.7*	*-1.87*	*-2.20*	*-2.26*	*-1.79*	*-2.14*	*-2.39*
Frafmentation		0.4998	0.3088		0.1054	-0.0279		-0.0937	0.0381		-0.107	0.0118
		(0.3997)	(0.3781)		(0.3650)	(0.3381)		(0.2680)	(0.2940)		(0.2833)	(0.3078)
		1.25	*0.82*		*0.29*	*-0.08*		*-0.35*	*0.13*		*-0.38*	*0.04*
adjusted R-squared	0.670	0.680	0.705	0.664	0.669	0.703	0.656	0.671	0.701	0.655	0.670	0.697
Region Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Simple	Simple	Rich	Simple	Simple	Rich	Simple	Simple	Rich	Simple	Simple	Rich
Observations	177	173	173	177	173	173	177	173	173	177	173	173

Table 5. On the Origins of Ethnic Inequality
From Inequality in Geographic Endowments to Inequality in Development across Ethnic Homelands

	Atlas Narodov Mira (GREG)						Ethnologue					
	All Ethnic Areas			Excl. Small Ethnic Areas			All Ethnic Areas			Excl. Small Ethnic Areas		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Gini Water	0.5545***	0.4771***	0.4413***	0.3044***	0.2469***	0.2292***	0.6274***	0.6526***	0.6550***	0.2906***	0.3125***	0.2690***
	(0.0600)	(0.0644)	(0.0692)	(0.0558)	(0.0585)	(0.0580)	(0.0653)	(0.0739)	(0.0772)	(0.0639)	(0.0719)	(0.0634)
	9.25	*7.41*	*6.38*	*5.46*	*4.22*	*3.95*	*9.6*	*8.83*	*8.49*	*4.55*	*4.35*	*4.25*
Gini Sea Distance	0.0432	0.2143**	0.2079*	0.0900	0.1610	0.1424	-0.0685	-0.0367	-0.2152	0.2160**	0.2156*	0.2728**
	(0.0977)	(0.1039)	(0.1207)	(0.0959)	(0.1059)	(0.1099)	(0.1006)	(0.1319)	(0.1390)	(0.1084)	(0.1247)	(0.1270)
	0.44	*2.06*	*1.72*	*0.94*	*1.52*	*1.30*	*-0.68*	*-0.28*	*-1.55*	*1.99*	*1.73*	*2.15*
Gini Elevation	0.3614	0.2040	0.8454*	1.0384***	1.0924***	1.5219***	0.8322	0.9385	2.5429***	0.9612*	1.0607*	2.0005***
	(0.3197)	(0.3828)	(0.5020)	(0.2840)	(0.3720)	(0.4132)	(0.5578)	(0.6747)	(0.8231)	(0.5012)	(0.6179)	(0.5896)
	1.13	*0.53*	*1.68*	*3.66*	*2.94*	*3.68*	*1.49*	*1.39*	*3.09*	*1.92*	*1.72*	*3.39*
Gini Land Suitability for Agriculture	0.2671***	0.0807	0.0306	0.1024	0.0222	0.1734	0.4137***	0.2427**	0.3965**	0.1228	0.0162	-0.0002
	(0.0873)	(0.1178)	(0.2181)	(0.1186)	(0.1275)	(0.1544)	(0.0892)	(0.1219)	(0.1577)	(0.1154)	(0.1484)	(0.1619)
	3.06	*0.68*	*0.14*	*0.86*	*0.17*	*1.12*	*4.64*	*1.99*	*2.51*	*1.06*	*0.11*	*0.00*
adjusted R-squared	0.626	0.632	0.628	0.619	0.628	0.641	0.711	0.72	0.734	0.634	0.642	0.659
Region Fixed Effe	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Spatial	Level/ Spatial	No	Spatial	Level/ Spatial	No	Spatial	Level/ Spatial	No	Spatial	Level /Spatial
Observations	167	163	163	167	163	163	167	163	163	167	163	163

Table 6 - Principal Component Analysis

	Eigenvalue	Variance Explained		Variable	Factor Loadings			
		Marginal	Cummulative		1st PC	2nd PC	3rd PC	4th PC
Panel A: Gini Coefficient GREG - All Groups								
1st Principal Component	2.2601	0.5650	0.5650	Gini Water	0.5494	-0.2610	-0.0778	-0.7899
2nd Principal Component	0.7956	0.1989	0.7639	Gini Sea Distance	0.4667	-0.6889	0.1288	0.5395
3rd Principal Component	0.5269	0.1317	0.8957	Gini Elevation	0.4999	0.4241	-0.7026	0.2767
4th Principal Component	0.4173	0.1043	1.0000	Gini Land Quality	0.4801	0.5268	0.6955	0.0913
Panel B: Gini Coefficient GREG - Excluding Small Ethnic Groups								
1st Principal Component	2.2506	0.5627	0.5627	Gini Water	0.5554	-0.3418	0.0131	-0.758
2nd Principal Component	0.8019	0.2005	0.7631	Gini Sea Distance	0.5073	-0.5816	0.0403	0.6346
3rd Principal Component	0.5784	0.1446	0.9077	Gini Elevation	0.4604	0.5562	0.6848	0.0983
4th Principal Component	0.3691	0.0923	1.0000	Gini Land Quality	0.4715	0.4853	-0.7275	0.1141
Panel C: Gini Coefficient ETHNOLOGUE - All Groups								
1st Principal Component	2.2411	0.5603	0.5603	Gini Water	0.5788	-0.2466	-0.0099	-0.7772
2nd Principal Component	0.8179	0.2045	0.7647	Gini Sea Distance	0.4933	-0.6288	0.2072	0.5642
3rd Principal Component	0.6110	0.1527	0.9175	Gini Elevation	0.4832	0.3592	-0.7564	0.2556
4th Principal Component	0.3300	0.0825	1.0000	Gini Land Quality	0.4337	0.644	0.6204	0.1107
Panel D: Gini Coefficient ETHNOLOGUE - Excluding Small Ethnic Groups								
1st Principal Component	2.27744	0.5694	0.5694	Gini Water	0.5824	-0.2501	-0.1075	-0.7659
2nd Principal Component	0.754243	0.1886	0.7579	Gini Sea Distance	0.5157	-0.6211	0.0489	0.5881
3rd Principal Component	0.67432	0.1686	0.9265	Gini Elevation	0.4499	0.5597	-0.6493	0.2505
4th Principal Component	0.294	0.0735	1	Gini Land Quality	0.4386	0.4883	0.7513	0.0686

Panel E: Gini Coefficient - Overall Spatial Inequality

		0.4974	0.4974	Gini Water	0.5578	-0.2744	0.2536	-0.7411
2nd Principal Component		0.2167	0.7141	Gini Sea Distance	0.5081	-0.5525	0.1588	0.6413
3rd Principal Component	0.6441	0.1610	0.8752	Gini Elevation	0.5052	0.2362	-0.83	0.0088
4th Principal Component	0.4993	0.1248	1.0000	Gini Land Quality	0.4188	0.7508	0.4707	0.1983

Table 7: On the Origins of Contemporary Ethnic Inequality
From Inequality in Geographic Endowments to Inequality in Development across Ethnic Homelands

	Atlas Narodov Mira (GREG)						Ethnologue					
	All Ethnic Areas			Excl. Small Ethnic Areas			All Ethnic Areas			Excl. Small Ethnic Areas		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1st Principal Component	0.1032***	0.1176***	0.1067***	0.0692***	0.0773***	0.0678***	0.1494***	0.1618***	0.1471***	0.0877***	0.0973***	0.1091***
Gini in Ethnic Geographic	(0.0092)	(0.0097)	(0.0097)	(0.0075)	(0.0075)	(0.0074)	(0.0111)	(0.0114)	(0.0115)	(0.0083)	(0.0086)	(0.0146)
Endowments	*11.25*	*12.16*	*10.95*	*9.17*	*10.38*	*9.17*	*13.44*	*14.14*	*12.84*	*10.63*	*11.37*	*7.46*
1st Principal Component		-0.0267**	-0.0356***		-0.0205***	-0.0235***		-0.0210*	-0.0356***		-0.0212**	0.0041
Gini in Geographic		(0.0103)	(0.0111)		(0.0075)	(0.0080)		(0.0124)	(0.0134)		(0.0093)	(0.0156)
Endowments		*-2.59*	*-3.2*		*-2.75*	*-2.94*		*-1.7*	*-2.65*		*-2.27*	*0.26*
adjusted R-squared	0.5660	0.5780	0.6010	0.6140	0.6260	0.6180	0.6530	0.6560	0.6510	0.6310	0.6400	0.5760
Observations	163	163	163	163	163	163	163	163	163	163	163	163
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	No	No	Geography	No	No	Geography	No	No	Geography	No	No	Geography

Table 8. Ethnic Inequality and Development
Inequality in Geographic Endowments across Ethnic Homelands and Contemporary Development

	Atlas Narodov Mira (GREG)						Ethnologue					
	All Ethnic Areas			Excl. Small Ethnic Areas			All Ethnic Areas			Excl. Small Ethnic Areas		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1st Principal Component	-0.1652***	-0.2048***	-0.1558**	-0.2060***	-0.2224***	-0.1629***	-0.1358***	-0.1673***	-0.1423***	-0.1598***	-0.1759***	-0.1406***
Gini in Ethnic Geographic	(0.0445)	(0.0588)	(0.0601)	(0.0467)	(0.0509)	(0.0478)	(0.0505)	(0.0634)	(0.0503)	(0.0481)	(0.0552)	(0.0459)
Endowments	*-3.71*	*-3.48*	*-2.59*	*-4.41*	*-4.37*	*-3.41*	*-2.69*	*-2.64*	*-2.83*	*-3.32*	*-3.19*	*-3.06*
1st Principal Component		0.0735	0.0822		0.0415	0.0464		0.0537	0.0755		0.0357	0.0542
Gini in Geographic		(0.0667)	(0.0653)		(0.0514)	(0.0486)		(0.0646)	(0.0553)		(0.0579)	(0.0511)
Endowments		*1.1*	*1.26*		*0.81*	*0.95*		*0.83*	*1.36*		*0.62*	*1.06*
adjusted R-squared	0.628	0.629	0.6510	0.641	0.641	0.66	0.617	0.616	0.65	0.625	0.623	0.654
Observations	163	163	163	163	163	163	163	163	163	163	163	163
Additional Controls	No	No	Geography	No	No	Geography	No	No	Geography	No	No	Geography
Region Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 9: 2SLS Estimates
Inequality in Geographic Endowments across Ethnic Homelands, Ethnic Inequality, and Contemporary Development

	Atlas Narodov Mira (GREG)						Ethnologue					
	All Ethnic Areas			Excl. Small Ethnic Areas			All Ethnic Areas			Excl. Small Ethnic Areas		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1st Principal Component	-1.6013***	-1.7416***	-1.4602***	-2.9751***	-2.8751***	-2.4043***	-0.9087***	-1.0340***	-0.9671***	-1.8210***	-1.8076***	-1.5793***
Gini in Ethnic	(0.3987)	(0.4686)	(0.5223)	(0.6756)	(0.6560)	(0.6689)	(0.3168)	(0.3650)	(0.3093)	(0.5249)	(0.5294)	(0.4563)
Geographic Endowments	*-4.02*	*-3.72*	*-2.80*	*-4.4*	*-4.38*	*-3.59*	*-2.87*	*-2.83*	*-3.13*	*-3.47*	*-3.41*	*-3.46*
1st Principal Component		0.0269	0.0301		-0.0175	-0.0101		0.032	0.041		-0.0026	0.0109
Gini in Geographic Endowments		(0.0524)	(0.0492)		(0.0446)	(0.0424)		(0.0555)	(0.0459)		(0.0508)	(0.0447)
		0.51	*0.61*		*-0.39*	*-0.24*		*0.58*	*0.89*		*-0.05*	*0.24*
Observations	163	163	163	163	163	163	163	163	163	163	163	163
Region Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	No	No	Geography	No	No	Geography	No	No	Geography	No	No	Geography
First-Stage F-score												

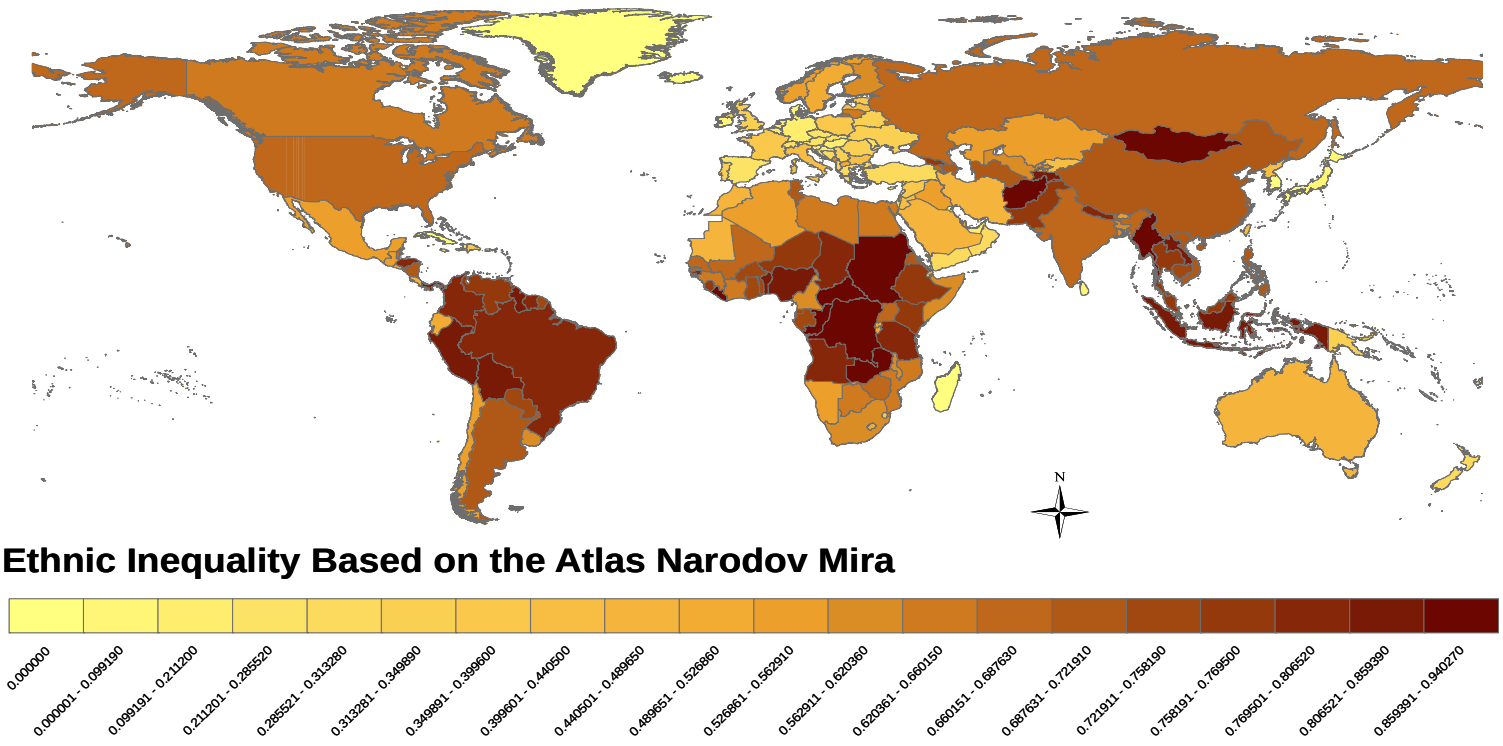
Table 10: Ethnic Inequality and Development: District-level Evidence from the Afrobarometer
District-Level Ethnic Inequality and Local Public Good Provision and Development within Countries

	Percentage of Urban Households within District		Average Access to Piped Water within District		Average Presence of Sewage System within District		Average Access to an Electricity Grid within District	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gini in Living Conditions across Ethnicities within District	-0.5969***	-0.5926***	-0.6521**	-0.6585**	-0.6529***	-0.6403***	-0.6523***	-0.6547**
	(0.2011)	(0.1921)	(0.2700)	(0.2612)	(0.1880)	(0.1754)	(0.2524)	(0.2606)
	-2.97	*-3.09*	*-2.41*	*-2.52*	*-3.47*	*-3.65*	*-2.58*	*-2.51*
Ln (Number of Ethnicities Within District)	0.2577***	0.2575***	0.1902***	0.1906***	0.1606***	0.1600***	0.2384***	0.2386***
	(0.0333)	(0.0326)	(0.0336)	(0.0334)	(0.0468)	(0.0452)	(0.0497)	(0.0479)
	7.75	*7.89*	*5.66*	*5.71*	*3.43*	*3.54*	*4.80*	*4.98*
Gini in Living Conditions across Respondents within District		-0.0109		0.0163		-0.0323		0.0064
		(0.2281)		(0.1703)		(0.1829)		(0.2419)
		-0.05		*0.1*		*-0.18*		*0.03*
Average Living Conditions Within District	0.0755***	0.0751***	0.0569	0.0574	0.0388	0.0377	0.0710**	0.0712**
	(0.0160)	(0.0179)	(0.0370)	(0.0420)	(0.0284)	(0.0249)	(0.0304)	(0.0301)
	4.72	*4.21*	*1.54*	*1.37*	*1.36*	*1.52*	*2.34*	*2.36*
Adjusted R-squared	0.171	0.171	0.204	0.204	0.181	0.181	0.312	0.312
Observations	1265	1265	1252	1252	1246	1246	1257	1257
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

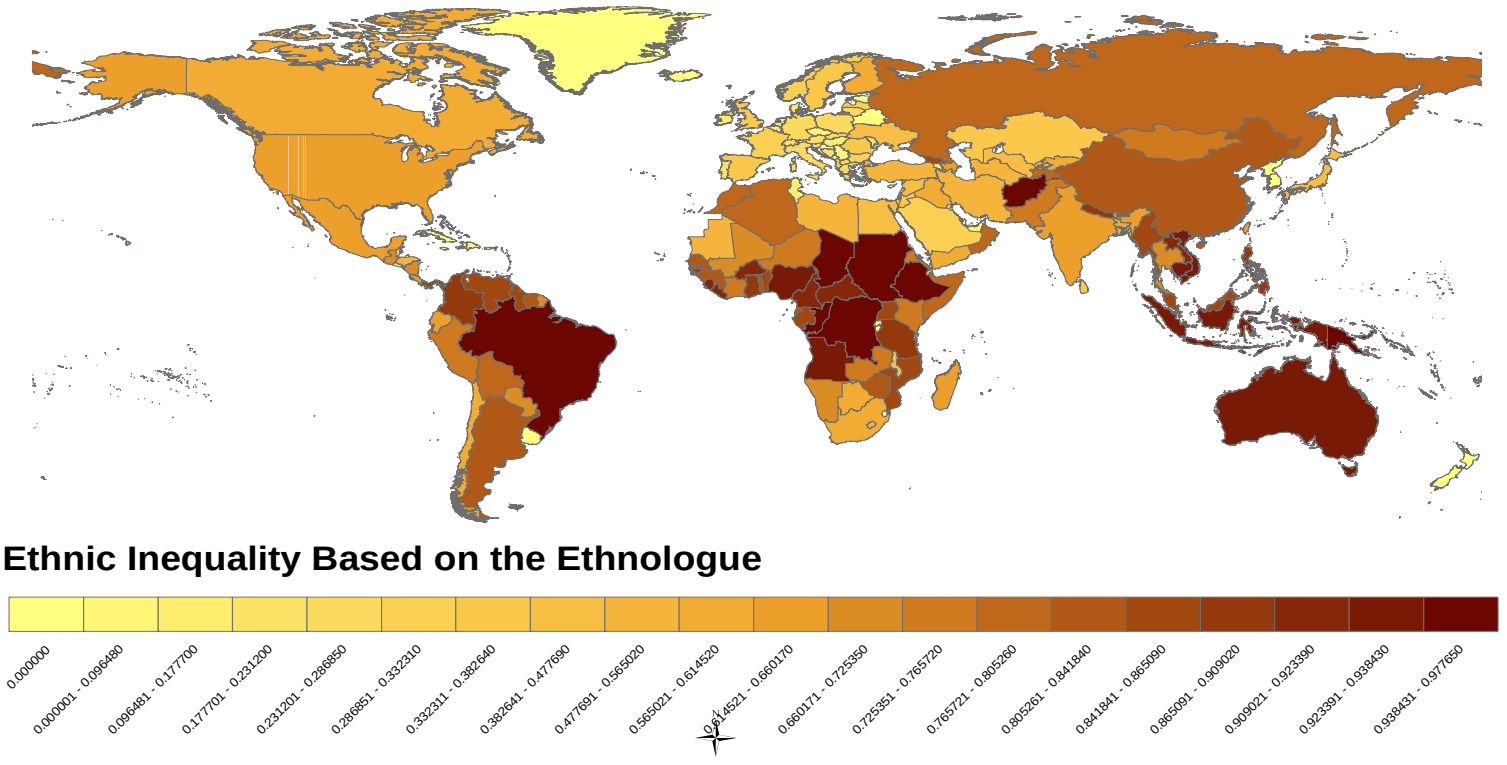
Table 11: Ethnic Inequality and Development: Household-level Evidence from the Afrobarometer
District-Level Ethnic Inequality and Local Public Good Provision and Development within Countries

	Urban Household		Access to Piped Water in the Enumeration area		Sewage Sytem is Present in the Enumeration area		Access to Electricity Grid in the Enumeration area	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gini in Living Conditions across Ethnicities within District	-0.8379***	-0.9115***	-0.6340***	-0.6055***	-0.6038***	-0.6581***	-0.4436**	-0.3994**
	(0.1918)	(0.1990)	(0.2135)	(0.2203)	(0.1834)	(0.1893)	(0.1765)	(0.1862)
	-4.39	*-4.6*	*-2.97*	*-2.75*	*-3.29*	*-3.48*	*-2.51*	*-2.15*
Ln (Number of Ethnicities Within District)	0.2690***	0.2697***	0.1637***	0.1634***	0.1311***	0.1316***	0.1849***	0.1846***
	(0.0270)	(0.0272)	(0.0301)	(0.0300)	(0.0351)	(0.0350)	(0.0300)	(0.0298)
	9.99	*9.93*	*5.44*	*5.46*	*3.73*	*3.76*	*6.16*	*6.19*
Living Conditions within District	0.0587**	0.0697***	0.0478*	0.0436	0.0211	0.029	0.0371	0.0304
	(0.0247)	(0.0258)	(0.0273)	(0.0288)	(0.0281)	(0.0297)	(0.0244)	(0.0258)
	2.38	*2.7*	*1.75*	*1.51*	*0.75*	*0.98*	*1.52*	*1.18*
Gini in Living Conditions across Respondents within District		*0.2886*		*-0.1109*		*0.2122*		*-0.1745*
		(0.213)		*(0.189)*		*(0.224)*		*(0.165)*
		1.36		*0.59*		*0.95*		*1.05*
Adjusted R-squared	0.345	0.345	0.349	0.349	0.274	0.274	0.384	0.384
Observations	20078	20078	19778	19778	19464	19464	19832	19832
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Districts	1265	1265	1252	1252	1246	1246	1257	1257
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Ethnicities	252	252	252	252	252	252	252	252
Ethnicity Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

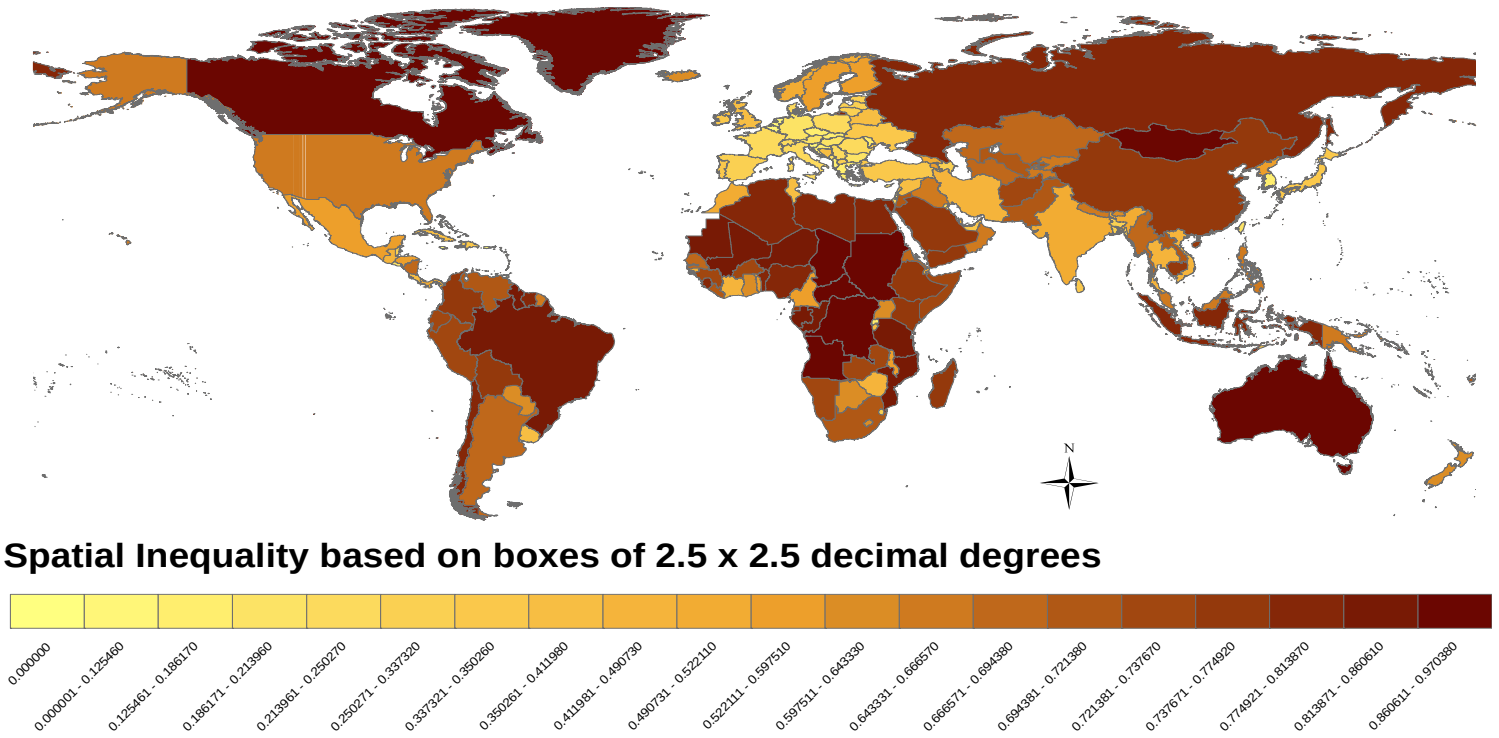
Appenidx Figure 1a



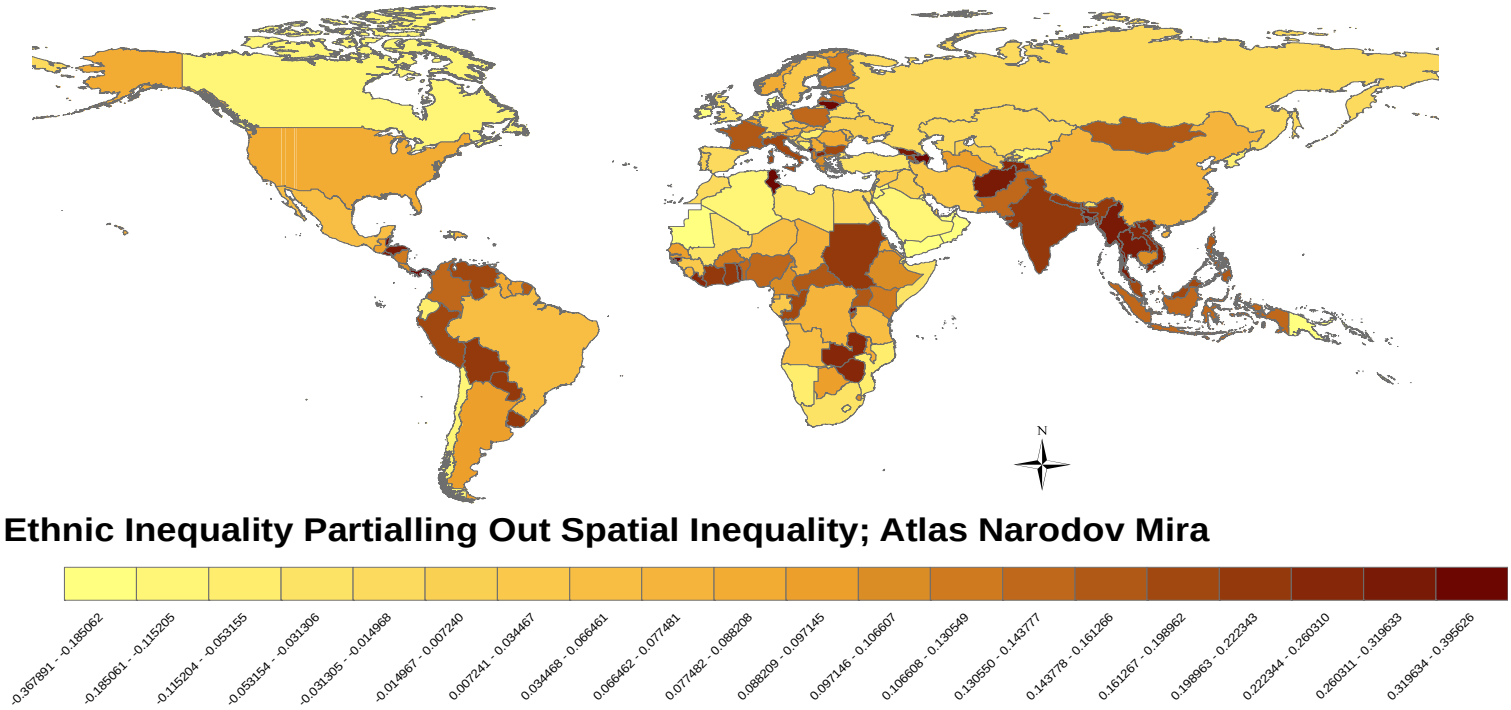
Appenidx Figure 1b



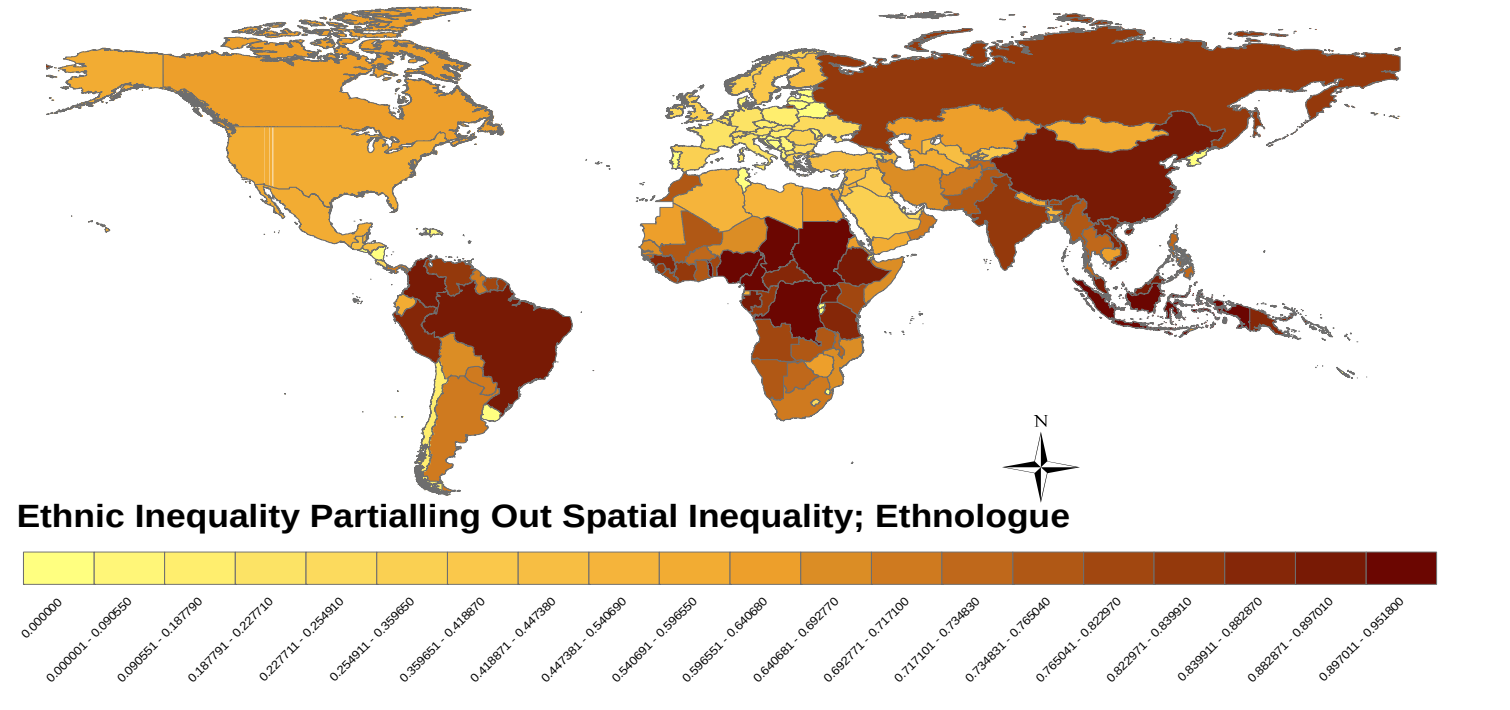
Appenidx Figure 1c



Appenidx Figure 2a



Appenidx Figure 2b



Appendix Table 1: Ethnic Inequality and Economic Development (in 2000)
Sensitivity Analysis. Excluding Countries with Just one Ethnic or Linguistic Group

	Atlas Narodov Mira (GREG)						Ethnologue					
	All Ethnic Areas			Excl. Small Ethnic Areas			All Ethnic Areas			Excl. Small Ethnic Areas		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Ethnic Gini	-1.4325***	-1.4923***	-1.0715**	-1.6119***	-1.7242***	-1.5913***	-0.8893**	-1.1557***	-1.0046**	-0.9875***	-1.0586***	-0.7838**
	(0.5090)	(0.5387)	(0.5405)	(0.4188)	(0.4038)	(0.3864)	(0.4466)	(0.4211)	(0.4792)	(0.3201)	(0.3246)	-0.3162
	-2.81	*-2.77*	*-1.98*	*-3.85*	*-4.27*	*-4.12*	*-1.99*	*-2.74*	*-2.10*	*-3.09*	*-3.26*	*-2.48*
Spatial Gini	-1.1123**	-1.1906**	-1.1820***	-0.741	-0.6684	-0.7683*	-1.2542**	-1.3290***	-1.1367**	-1.1547**	-1.2918***	-1.2063***
	(0.4692)	(0.4648)	(0.4456)	(0.5071)	(0.5066)	(0.4310)	(0.4864)	(0.4910)	(0.4396)	(0.4658)	(0.4683)	(0.4282)
	-2.37	*-2.56*	*-2.65*	*-1.46*	*-1.32*	*-1.78*	*-2.58*	*-2.71*	*-2.59*	*-2.48*	*-2.76*	*-2.82*
Fragmentation		0.2670	0.2130		-0.0005	-0.079		-0.1570	-0.017		-0.2119	-0.0196
		(0.4446)	(0.4404)		(0.3902)	(0.3513)		(0.2769)	(0.3254)		(0.2870)	(0.3124)
		0.6	*0.48*		*0*	*-0.22*		*-0.57*	*-0.05*		*-0.74*	*-0.06*
adjusted R-squared	0.666	0.680	0.696	0.669	0.677	0.710	0.664	0.677	0.695	0.676	0.686	0.704
Region Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Simple	Simple	Rich	Simple	Simple	Rich	Simple	Simple	Rich	Simple	Simple	Rich
Observations	156	154	154	134	132	132	152	148	148	140	136	136

Appendix Table 2: Ethnic Inequality and Economic Development (in 2000)
Sensitivity Analysis. Excluding from the Estimation of Ethnic Inequality Regions of Capital Cities

	Atlas Narodov Mira (GREG)						Ethnologue					
	All Ethnic Areas			Excl. Small Ethnic Areas			All Ethnic Areas			Excl. Small Ethnic Areas		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Ethnic Gini	-1.0923***	-1.2417***	-1.0532***	-0.8439**	-0.8856*	-0.6078	-1.0048***	-1.0733***	-0.8525***	-0.6803*	-0.7732*	-0.7234*
	(0.3627)	(0.3440)	(0.3174)	(0.4258)	(0.4674)	(0.5043)	(0.2942)	(0.2997)	(0.2968)	(0.3974)	(0.4048)	-0.424
	-3.01	*-3.61*	*-3.32*	*-1.98*	*-1.89*	*-1.21*	*-3.42*	*-3.58*	*-2.87*	*-1.71*	*-1.91*	*-1.71*
Spatial Gini	-1.1223**	-1.0583**	-0.9449**	-1.2976***	-1.4184***	-1.2771***	-1.1852***	-1.1060**	-1.0248**	-1.3415***	-1.2659**	-1.1303**
	(0.5034)	(0.4974)	(0.4077)	(0.4897)	(0.4900)	(0.4379)	(0.4404)	(0.4557)	(0.4130)	(0.4661)	(0.4987)	(0.4372)
	-2.23	*-2.13*	*-2.32*	*-2.65*	*-2.89*	*-2.92*	*-2.69*	*-2.43*	*-2.48*	*-2.88*	*-2.54*	*-2.59*
Fragmentation		0.1128	0.0894		0.1295	0.1264		-0.0906	-0.098		-0.1709	-0.2377
		(0.3539)	(0.3521)		(0.4374)	(0.4692)		(0.3635)	(0.3750)		(0.3909)	(0.4253)
		0.32	*0.25*		*0.3*	*0.27*		*-0.25*	*-0.26*		*-0.44*	*-0.56*
adjusted R-squared	0.6675	0.683	0.7183	0.6573	0.6759	0.6927	0.6751	0.6848	0.7017	0.6558	0.6648	0.6794
Region Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Simple	Simple	Rich	Simple	Simple	Rich	Simple	Simple	Rich	Simple	Simple	Rich
Observations	157	154	154	131	128	128	153	149	149	138	134	134

Appendix Table 3: Ethnic Inequality and Economic Development (in 2000)
Sensitivity Analysis. Excluding from the Estimation Iteratively Regions

	Atlas Narodov Mira (GREG)						Ethnologue					
	no SSA	no MENA	no EAP	no ECA	no LCA	No WE-A	no SSA	no MENA	no EAP	no ECA	no LCA	No WE-A
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Ethnic Gini	-1.6152***	-1.2695***	-1.3032***	-1.4744***	-1.3094***	-1.3957***	-0.9822***	-0.8610***	-0.8178***	-1.1483***	-1.0578***	-1.0298***
	-0.3871	-0.423	-0.3454	-0.4372	-0.4274	-0.389	-0.3051	-0.2979	-0.3084	-0.3193	-0.3289	-0.304
	-4.17	-3	-3.77	-3.37	-3.06	-3.59	-3.22	-2.89	-2.65	-3.6	-3.22	-3.39
Spatial Gini	-0.3785	-0.9203*	-0.731	-0.4365	-0.9243	-0.8441*	-0.7761	-1.0104*	-0.8974*	-0.4587	-0.9570*	-0.9742*
	-0.5094	-0.5017	-0.4842	-0.5186	-0.5754	-0.5089	-0.5161	-0.5126	-0.497	-0.5291	-0.5689	-0.5112
	-0.74	-1.83	-1.51	-0.84	-1.61	-1.66	-1.5	-1.97	-1.81	-0.87	-1.68	-1.91
Fragmentation	0.3289	-0.218	-0.2361	0.1769	-0.1128	-0.0631	0.3496	-0.1212	-0.0887	0.3672	-0.0003	0.0844
	-0.3569	-0.3779	-0.3359	-0.3624	-0.388	-0.3981	-0.3712	-0.3664	-0.3367	-0.3494	-0.3757	-0.392
	0.92	-0.58	-0.7	0.49	-0.29	-0.16	0.94	-0.33	-0.26	1.05	0	0.22
adjusted R-squared	0.5517	0.6738	0.729	0.6913	0.6847	0.5609	0.5251	0.6667	0.7181	0.6886	0.6842	0.5539
Region Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Simple	Simple	Simple	Simple	Simple	Simple	Simple	Simple	Simple	Simple	Simple	Simple
Observations	126	153	152	146	141	147	126	153	152	146	141	147

Appendix Table 4
On the Origins of Contemporary Ethnic Inequality

Panel A: Geography

	Atlas Narodov Mira (GREG)				Ethnologue			
	All	Excl. 1%	All	Excl. 1%	All	Excl. 1%	All	Excl. 1%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Latitude	-0.0034*	-0.0055***	-0.0043*	-0.0061***	-0.0113***	-0.0076***	-0.0091***	-0.0071***
	(0.0017)	(0.0016)	(0.0023)	(0.0020)	(0.0020)	(0.0015)	(0.0024)	(0.0017)
Log Land Area	0.011	-0.0132	0.0326***	0.0043	0.0514***	-0.0049	0.0447***	0.002
	(0.0111)	(0.0093)	(0.0081)	(0.0074)	(0.0121)	(0.0105)	(0.0107)	(0.0076)
Log Population	-0.0086	-0.0208	-0.0073	-0.0181	-0.0221	-0.0381**	-0.0246	-0.0345**
	(0.0153)	(0.0144)	(0.0130)	(0.0123)	(0.0175)	(0.0158)	(0.0168)	(0.0148)
Area under Water	0.0000	0.0000			0.0000	0.0000		
	(0.0000)	(0.0000)			(0.0000)	(0.0000)		
Distance to the Coast	0.0001**	0.0001			-0.0001	-0.0001		
	(0.0000)	(0.0001)			(0.0001)	(0.0001)		
Elevation	0.0000	0.0000			0.0001	0.0000		
	(0.0000)	(0.0000)			(0.0000)	(0.0000)		
Land Suitability for Agriculture	0.0839	0.054			-0.0800	0.0435		
	(0.0644)	(0.0546)			(0.0768)	(0.0619)		
Ruggdeness			-0.0064	-0.0019			0.0106	0.0027
			(0.0100)	(0.0109)			(0.0134)	(0.0101)
Soil Quality			-0.0002	0.0004			0.0008	0.0005
			(0.0006)	(0.0006)			(0.0007)	(0.0006)
Distance to the Sea			0.0887*	0.0736*			0.0245	-0.0152
			(0.0516)	(0.0436)			(0.0573)	(0.0335)
Desert Area			-0.0045***	-0.0036***			-0.0043**	-0.0051***
			(0.0013)	(0.0012)			(0.0018)	(0.0012)
Tropical Area			-0.0003	-0.0002			0.0005	0.0002
			-0.0006	-0.0006			-0.0008	-0.0006
Overall Spatial Gini	0.4743***	0.2300**	0.4524***	0.2474***	0.4322***	0.3684***	0.5939***	0.4787***
	-0.0972	-0.0943	-0.0985	-0.0885	-0.1201	-0.103	-0.1074	-0.0931
adjusted R-squared	0.521	0.543	0.596	0.524	0.617	0.596	0.663	0.646
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	163	163	177	177	163	163	177	177

Appendix Table 4
On the Origins of Contemporary Ethnic Inequality

Panel B: Historical Features

	Atlas Narodov Mira (GREG)				Ethnologue			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Latitude	-0.0066***	-0.0074***	-0.0034	-0.0044	-0.0118***	-0.0085***	-0.0092***	-0.0066***
	(0.0024)	(0.0023)	(0.0023)	(0.0030)	(0.0026)	(0.0021)	(0.0022)	(0.0019)
Log Land Area	0.0323**	-0.0013	-0.0041	-0.0257	0.0665***	0.0065	0.0417*	-0.0205
	(0.0139)	(0.0130)	(0.0172)	(0.0173)	(0.0159)	(0.0148)	(0.0230)	(0.0143)
Log Population	-0.0204	-0.0315	-0.0452*	-0.0376	-0.0171	-0.0333*	-0.0184	-0.0128
	(0.0188)	(0.0200)	(0.0261)	(0.0312)	(0.0204)	(0.0183)	(0.0311)	(0.0287)
Common Law Dummy	-0.0867*	-0.0004			-0.0207	0.0516		
	(0.0490)	(0.0424)			(0.0475)	(0.0440)		
Log Settler Moratlity	0.028	0.0331			0.0329	0.0431**		
	(0.0187)	(0.0201)			(0.0222)	(0.0199)		
State Antiguity			-0.1246	-0.0593			-0.0629	0.1281
			(0.1132)	(0.1606)			(0.1255)	(0.0804)
Executive Constraints			-0.0049	0.0315			0.0451	0.0158
at independence			(0.0488)	(0.0511)			(0.0595)	(0.0421)
European Descent			-0.0006	-0.0007			-0.0004	0.0008
			(0.0012)	(0.0009)			(0.0011)	(0.0007)
Ethnic Partitioning			0.0000	0.0013			0.0000	-0.0015*
			(0.0007)	(0.0008)			(0.0011)	(0.0008)
Border Straightness			1.2963	1.1281			0.9350	1.0043
			(1.1515)	(1.1675)			(1.5277)	(1.0220)
Overall Spatial Gini	0.3301***	0.1800	0.4150***	0.2027	0.5062***	0.3887***	0.6648***	0.5403***
	-0.1239	-0.1136	-0.1108	-0.1435	-0.1288	-0.1183	-0.1506	-0.1291
adjusted R-squared	0.503	0.478	0.65	0.562	0.645	0.535	0.658	0.657
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	77	77	82	82	77	77	82	82

Appendix Table 6 - Pairwise Correlation Structure Geographic Variables

Gini Ethnic Elevation	1											
Gini Geograpphic Elevation	0.7643*	1										
Mean Elevation	0.5822*	0.4387*	1									
Gini Ethnic Land Quality	0.4732*	0.3186*	0.2600*	1								
Gini Geographic Land Quality	0.3847*	0.3240*	0.2328*	0.9027*	1							
Mean Land Quality	-0.0219	-0.0956	0.003	-0.5222*	-0.6009*	1						
Gini Ethnic Area under Water	0.4800*	0.1358	0.2546*	0.4281*	0.2761*	0.087	1					
Gini Geographic Area under Water	0.2468*	0.3786*	0.0277	0.2665*	0.2897*	-0.1585*	0.2308*	1				
Mean Area under Water	0.1965*	0.123	0.0233	0.2952*	0.3607*	-0.1873*	0.1592*	0.1389	1			
Gini Ethnic Sea Distance	0.3232*	0.1372	-0.0922	0.2854*	0.2186*	0.012	0.5597*	0.3047*	0.2859*	1		
Gini Geographic Sea Distance	0.1908*	0.3239*	-0.2236*	0.1212	0.1755*	-0.064	0.1313	0.5183*	0.2072*	0.5958*	1	
Mean Distance to Sea Coast	0.2178*	0.1322	0.4729*	0.3193*	0.3365*	-0.1896*	0.3002	0.2256*	0.143	-0.1176	-0.3747*	1

Appendix Table 7: Indirect Test of Exclusion Restriction
Inequality in Geographic Endowments across Ethnic Homelands, Ethnic Inequality, and Contemporary Development

	Atlas Narodov Mira (GREG)						Ethnologue					
	All Ethnic Areas			Excl. Small Ethnic Areas			All Ethnic Areas			Excl. Small Ethnic Areas		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Ethnic Inequality	-1.6225***	-1.5535***	-1.3502***	-1.6661***	-1.6320***	-1.3481**	-1.4245***	-1.3806***	-1.2434***	-1.6058***	-1.5661***	-1.2765**
Ethnic Gini in 2000	(0.3974)	(0.4080)	(0.4237)	(0.5198)	(0.5159)	(0.5255)	(0.3513)	(0.3518)	(0.3575)	-0.5074	-0.5156	-0.5155
	-4.08	*-3.81*	*-3.19*	*-3.21*	*-3.16*	*-2.57*	*-4.05*	*-3.92*	*-3.48*	*-3.16*	*-3.04*	*-2.48*
1st Principal Component	0.081	0.0574	0.0642	-0.0055	-0.0195	-0.0264	0.1149	0.0952	0.1152*	0.0537	0.0384	0.0338
Gini in Ethnic Geographic	(0.0715)	(0.0783)	(0.0729)	(0.0535)	(0.0540)	(0.0515)	(0.0702)	(0.0719)	(0.0685)	-0.0607	-0.0622	-0.0601
Endowments	*1.13*	*0.73*	*0.88*	*-0.1*	*-0.36*	*-0.51*	*1.64*	*1.32*	*1.68*	*0.88*	*0.62*	*0.56*
Spatial Gini			-0.7473			-1.0333**			-0.9577**			-0.8865**
Overall Spatial Inequality in 2000			-0.4657			-0.4447			-0.4622			-0.4461
			-1.6			*-2.32*			*-2.07*			*-1.99*
adjusted R-squared	0.688	0.687	0.691	0.686	0.688	0.695	0.687	0.686	0.694	0.682	0.682	0.687
Observations	163	163	163	163	163	163	163	163	163	163	163	163
Region Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Appendix Table 8: Summary Statistics in the Afrobarometer Sample

Panel A: District - Level

	Obs	Mean	Std. Dev.	Min	Max
Access to Sewage System	1246	0.23	0.38	0.00	1.00
Access to an Electricity Grid	1257	0.53	0.45	0.00	1.00
Access to Piped Water	1252	0.46	0.45	0.00	1.00
Percentage of Urban Respondents	1265	0.31	0.42	0.00	1.00
Living Conditions	1265	2.69	0.67	1.00	5.00
Ln (number of Ethnic Groups)	1265	0.76	0.71	0.00	3.14
Ethnic Gini Coefficient in Living Conditions	1265	0.07	0.08	0.00	0.33
Gini Coefficient in Living Conditions	1293	0.19	0.08	0.00	0.38

Panel B: Individual - level

	Obs	Mean	Std. Dev.	Min	Max
Access to Sewage System	19464	0.22	0.42	0.00	1.00
Access to an Electricity Grid	19832	0.54	0.50	0.00	1.00
Access to Piped Water	19778	0.49	0.50	0.00	1.00
Living in an Urban	20078	0.37	0.48	0.00	1.00
Average Living Conditions within District	20078	2.61	1.19	1.00	5.00
Ln (number of Ethnic Groups within District)	20078	1.28	0.86	0.00	3.14
Ethnic Gini Coefficient in Living Conditions Across Ethnic Groups within District	20078	0.10	0.08	0.00	0.33
Gini Coefficient in Living Conditions Across Districts	20078	0.21	0.06	0.00	0.38