THE UNIVERSITY OF WARWICK

Sep 2013 No.155

Ethnic Inequality

Alberto Alesina
Harvard University, NBER and IGIER
Stelios Michalopoulos
Brown University and NBER

WORKING PAPER SERIES

Centre for Competitive Advantage in the Global Economy

Department of Economics

CAGE
Competitive Advantage in the Global Economy

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Ethnic Inequality*

Alberto Alesina
Harvard University, NBER and IGIER

Stelios Michalopoulos
Brown University and NBER

Elias Papaioannou
London Business School and CEPR and NBER

First Draft: October 2012

Revised July 2013

Abstract

This study explores the consequences and origins of between-ethnicity economic inequality both across and within countries. First, combining satellite images of nighttime luminosity with the historical homelands of ethnolinguistic groups we construct measures of ethnic inequality for a large sample of countries and show that the latter is strongly inversely related to comparative development. Second, differences in geographic endowments across ethnic homelands explain a sizable portion of ethnic inequality contributing to its persistence over time. Third, exploiting across-district within-African countries variation using individual-level data on ethnic identification and well-being from the Afrobarometer Surveys we find that between-ethnic-group inequality is systematically linked to regional under-development. In this sample we also explore the channels linking ethnic inequality to (under) development, finding that ethnic inequality maps to political inequality, heightened perceptions of discrimination and undersupply of public goods.

Keywords: Ethnicity, Diversity, Inequality, Development, Geography

JEL classification Numbers: O10, O40, O43.

*We thank Christian Dippel, Nathan Nunn, Debraj Ray, Andrei Shleifer, Enrico Spolaore, Pierre Yared, Romain Wacziarg, David Weil, Michele Lenza, Oeindrila Dube and Ivo Welch for valuable comments and suggestions. We also would like to thank for useful feedback seminar participants at Dartmouth, the Athens University of Economics and Business, UBC, Brown, CREi, Oxford, Bocconi, NYU, Paris School of Economics, Warwick, LSE, Nottingham, the NBER Summer Institute Meetings in Political Economy, the CEPR Development Economics Workshop, the conference "How Long is the Shadow of History? The Long-Term Persistence of Economic Outcomes" at UCLA, and the Nemmers Conference in the Political Economy of Growth and Development at Northwestern University. All errors are our own responsibility.

1 Introduction

Ethnic diversity has costs and benefits. On the one hand, diversity in skills, education, and endowments can enhance productivity by promoting trade and innovation, especially in advanced economies. On the other hand, ethnic diversity is often associated with poor policies, low public goods provision, conflict, and hatred. In fact a large literature shows a negative effect of ethnolinguistic fragmentation on various aspects of economic performance, with the possible exception of wealthy economies (see Alesina and Ferrara (2005) for a review). Income inequality may also have both positive and negative effects on development. On the negative side, a higher degree of income inequality may lead to conflict, crime, prevent the poor from acquiring education and/or lead to expropriation and lofty taxation discouraging investment. On the positive side, however, income inequality may spur innovation and entrepreneurship by motivating individuals. Further complicating the relationship between the two, a positive correlation between inequality and development may simply reflect Simon Kuznetz's conjecture that industrialization translates into higher levels of inequality at the early stage of development; while at later stages the association becomes negative. Given the theoretical ambiguities (and data issues) perhaps it comes at no surprise that it has been very hard to detect empirically a robust association between inequality and development (see Benabou (2005) and Galor (2011) for surveys).

This paper puts forward and tests an alternative conjecture that focuses on the inter-section of ethnic diversity and inequality. Our thesis is that what mostly matters for development are economic differences between ethnic groups coexisting in the same country, rather than the degree of fractionalization per se or income inequality conventionally measured (i.e., independent of ethnicity). Inequality in income along ethnic lines is likely to lead to political inequality, increase animosity, and lead to discriminatory policies of one (or more) groups against the others. Furthermore, differences in preferences across both ethnic and income lines may lead to inadequate public goods provision, as groups' ideal allocation of public goods will be quite distant. Moreover, the presence of an economically dominant ethnic minority may lower support for democracy and free-market institutions, as the majority of the population usually feels that the benefits of capitalism go to just a handful of ethnic groups. Ethnic inequality often implies that well-being depends on group affiliation; hence it is more likely to generate envy and perceptions that the system is "unfair", more so than the conventionally measured economic inequality, since the latter can be more easily be thought of as the result of ability or effort.

The first contribution of this paper is to provide measures of within-country differences in well-being across ethnic groups, coined as "ethnic inequality". Information on income levels of ethnic groups for all countries is not available. Hence, to construct country-level indicators of ethnic inequality for the largest possible sample, we combine maps on the location of ethnolinguistic

groups with satellite images of light density at night, which are strong proxies of development and are available at a fine grid (see Henderson, Storeygard, and Weil (2012)). The cross-ethnic group inequality index is weakly correlated with the commonly employed -and notoriously poorly measured- income inequality measures at the country level. To isolate the cross-ethnic component of inequality from the overall inequality across regions, we also construct proxies of spatial inequality.

Second, we document a strong negative association between ethnic inequality and real GDP per capita at the country level. This correlation holds even when we condition on the overall degree of spatial inequality, which is also inversely related to well-being (an novel finding by itself). We also uncover that the negative correlation between ethnolinguistic fragmentation and development weakens considerably when we account for ethnic inequality; this suggests that it is the unequal concentration of wealth across ethnic lines that is detrimental for development rather than diversity per se.

Third, in an effort to shed light on the roots of ethnic inequality we construct measures reflecting differences in geographic endowments across ethnic homelands and show that the latter is a strong predictor of ethnic inequality. In contrast there is no link between contemporary ethnic inequality and historical variables capturing the type of colonization, state formation, legal origin, etc. Fourth, we show that contemporary development at the country level is also inversely related to inequality in geographic endowments across ethnic homelands.

Fifth, we examine the link between ethnic inequality and development using individual-level data from the Afrobarometer surveys, exploiting within-country across-district variation. Besides the immediate econometric benefits of looking within -rather than across countries- (effectively accounting for country-level fixed factors related to national institutions and policies, historical legacies, etc), this is quite useful for validating the cross-country results (based on luminosity and ethnographic maps) using detailed micro-level data. Specifically, we construct measures capturing between-group and within-group inequality utilizing information from roughly $20,000$ respondents in 16 Sub-Saharan countries on ethnic identification and well-being. Our analysis shows that -conditional on numerous individual characteristics- respondents residing in ethnically unequal districts are less educated and have lower standards of living. This pattern also holds when we exploit within-ethnicity variation (this is feasible as we observe respondents from the same ethnicity residing in multiple districts), effectively accounting for both unobserved ethnic features and migration.

Sixth, we exploit the richness of the Afrobarometer data to make some progress towards understanding the mechanisms linking ethnic inequality to (lack of) development in the most unequal and fragmented part of the world, Sub-Saharan Africa. We begin our analysis on the

channels showing that ethnic inequality goes in tandem with political inequality and discrimination across ethnic lines. We then examine the link between ethnic inequality and public goods provision building on the large literature in political economy showing that in ethnically diverse countries (as well as regions and cities within the United States and some other countries) there is under-provision of public goods. Our analysis shows that even when one looks within the same country access to basic public goods, such as piped water, a sewage system, and electrification is systematically lower in districts characterized by high levels of ethnic inequality. While this association does not necessarily reflect a causal relationship, it pertains even when controlling for numerous individual characteristics (including proxies of wealth and well-being) and ethnicity fixed effects. Finally, we examine the link between ethnic inequality and individual beliefs on the merits of democracy. This investigation is motivated by the influential thesis of Chua (2003), who argues that the spread of democracy (and free market institutions) since the 1990s does not necessarily lead to improved economic efficiency, when wealth is concentrated across ethnic lines. The uncovered evidence is supportive of this conjecture. Respondents found in ethnically unequal districts feel unfairly treated by the government and are less satisfied with the functioning of democratic institutions.

Related Works While both the literature linking ethnolinguistic diversity and economic performance and the literature studying the interplay between inequality and development are voluminous, there have been very few works examining the role of ethnic inequality on well-being. Exploiting data from 29 developing countries (from the Demographic and Health Surveys), Kyriacou (2013) finds that socioeconomic ethnic group inequalities reduce government quality. Cederman, Weidman, and Gleditch (2011) combine proxies of local economic activity from the *G-Econ* database with ethnolinguistic maps to construct an index of ethnic inequality for a subset of "politically relevant ethnic groups" (as defined by Ethnic Power Relations Dataset) and then show that in highly unequal countries, both rich and poor groups fight more often than those groups whose wealth is closer to the country average.

Structure The paper is organized as follows. In section 2 we describe the construction of the ethnic inequality index and present summary statistics. In Section 3 we report the results of our analysis associating income per capita with ethnic inequality across 173 countries. In Section 4 we examine the geographic origins of contemporary differences in ethnic inequality across countries; we also report estimates associating contemporary development with inequality in geographic endowments across ethnic homelands. In Section 5 we first examine the within-country across-district association between ethnic inequality and well-being in 16 Sub-Saharan countries using individual-level data from the Afrobarometer Surveys. We then use these data to

explore the channels linking ethnic inequality to development. In the last section we summarize and discuss potential avenues for future research.

2 Data and Descriptive Statistics

Since data on income-wages at the ethnicity level are not readily available for a large number of countries, to construct proxies of ethnic inequality for a large number of countries we combine information from ethnographic maps on the location of ethnic-linguistic groups with satellite images on light density at night that are available at a fine grid and can thus be aggregated at the historical ethnic homeland level. In this Section we discuss the construction of the cross-country measures capturing inequality in development (as reflected in luminosity) across ethnic homelands and across regions within 173 countries.

2.1 Location of ethnic groups

We identify the location of ethnic groups employing two data sets/maps. First we use the Geo-Referencing of Ethnic Groups (GREG), which is the digitized version of the *Soviet Atlas Narodov Mira* (Weidmann, Rod, 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.[1] The GIS data set uses the political boundaries of 1964 to allocate groups to different countries. We thus project the ethnic homelands to the political boundaries of the 2000 Digital Chart of the World (ignoring polygons of less than 1 km^2); this results in $2,125$ ethnic homelands within contemporary countries. Most areas $(1,630)$ are coded as pertaining to a single group whereas in the remaining 495 there can be up to three groups. For example, in Northeast India along an area of $4,380$ km^2 the Assamese, the Oriyas and the Santals overlap. In these cases we assign the respective homeland to all groups. The size of ethnic homelands varies considerably. The smallest polygon occupies an area of 1.15 km^2 (French in Monaco) and the largest extends over $7,335,476$ km^2 (American English in the US). The median (mean) group size is $4,198$ $(61,506)$ km^2. The median (mean) country has 8 (11.52) ethnicities with the most diverse being Indonesia with 94 groups.

Our second source is the $15th$ edition of *Ethnologue* (Gordon (2005)) that maps $7,570$ linguistic groups (using the political boundaries of 2000 for the geo-referencing). In spite of the detailed linguistic mapping, *Ethnologue's* coverage for some continents (e.g., Latin America) is limited while for others (i.e. Africa and Asia) is very detailed. *Ethnologue's* mapping corresponds to the early 1990's; thus the location of ethnic groups may be affected by national policies, conflict,

[1] The original *Atlas Narodov Mira* consists of 57 ethnographic maps. The original sources are: (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.

or other features. Each polygon in the *Ethnologue* delineates a traditional linguistic region; populations away from their homelands (in cities, refugee camps) are not mapped. Groups of unknown location, widespread and extinct languages are not mapped, the only exception is the English in the United States. *Ethnologue* also records areas where languages overlap; in this case we assign the polygon to all languages. *Ethnologue* provides a more refined linguistic aggregation compared to the GREG. As a result the median (mean) homeland extends to 728 (12,986) km^2. The smallest language is the Domari in Israel which covers 1.18 km^2 with the largest group is the English in the US covering 9,327,331 km^2. The median (mean) country has 9 (41.9) groups with Papua New Guinea being the most diverse country with 791 linguistic groups.

GREG attempts to map major immigrant groups whereas Ethnologue generally does not. This is important for countries in the New World. For example, in Argentina GREG reports 16 groups, among them Germans, Italians, and Chileans, whereas Ethnologue reports 20 purely indigenous groups, such as the Toba and the Quechua. For Canada *Ethnologue* lists 77 mostly indigenous groups, like the Blackfoot and the Chipewyan with only English and French being historically non-indigenous; in contrast GREG that lists 23 groups is featuring many non-indigenous groups, such as Swedes, Russians, Norwegians and Germans. Hence, the two ethnolinguistic mappings capture different cleavages, at least in some continents.[2]

2.2 Luminosity

Since comparable data on income per capita at the ethnicity level across all countries in the world do not exist, following Henderson, Storeygard, and Weil (2012) and subsequent studies (e.g., (Chen and Nordhaus (2011), Pinkovskiy (2011), Michalopoulos and Papaioannou (2012, 2013)) we use satellite image data on light density at night as a proxy.[3] The luminosity data come from the Defense Meteorological Satellite Program's Operational Linescan System that reports images of the earth at night. The six-bit number that ranges from 0 to 63 is available approximately at every square kilometer since 1992. To construct luminosity at the desired level of aggregation we average all observations falling within the boundaries of an ethnic group and then divide with the population of each area using data from the Gridded Population of the World that reports geo-referenced pixel-level population estimates for 1990 and 2000.[4]

[2] We are including all groups in our analysis without attempting to make a distinction as to which cleavage is more salient.

[3] These -and other works- show that luminosity is a strong correlate of development at various levels of aggregation (countries, regions, ethnic homelands).

[4] The data is constructed using subnational census and other surveys of population at various levels (city, neighborhood, region). See for details: http://sedac.ciesin.columbia.edu/data/collection/gpw-v

2.3 Ethnic Inequality

We proxy the level development in ethnic homeland i with average luminosity per capita, y_i, and construct the Gini index that reflects inequality across ethnic groups (ethnic inequality) for each country. Specifically, the Gini coefficient for a country's population consisting of n groups with values of luminosity per capita for the historical homeland of ethnicity i, y_i, where $i = 1$ *to* n are indexed in non-decreasing order ($y_i \leq y_{i+1}$), is calculated as follows:

$$G = \frac{1}{n}(n + 1 - 2\frac{\sum_{i=1}^{n}(n + 1 - i)y_i}{\sum_{i=1}^{n} y_i})$$

Note that the ethnic Gini index captures differences in mean income -as captured in luminosity per capita at the ethnic homeland- across groups. For each of the two different linguistic maps we construct Gini coefficients for the maximum sample of countries using cross-ethnic-homeland data in 1992, 2000, and 2009. As a robustness we also construct the Gini coefficient dropping the capital cities and excluding small ethnicities, defined as those capturing less than 1% of the 2000 population in a country. For example, in Kenya the *Atlas Narodov Mira* (the *Ethnologue*) 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. 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.

2.4 Spatial inequality

Since we use ethnic homelands (rather than individual-level) data to measure between-group inequality, the ethnic inequality measures also reflect regional disparities in income (that may not be related to ethnic diversity). To isolate the between-ethnicity component from regional inequality, we thus also construct Gini coefficients reflecting the overall (rather than the ethnic) degree of spatial inequality in each country. Since we couldn't find a widely-accepted way to measure spatial inequality, we construct for robustness two measures of the overall degree of spatial inequality in each country.

Spatial Gini Coefficient 1. This index is based on aggregating luminosity per capita across roughly equally-sized boxes. We first generate a global grid of pixels of 2.5 by 2.5 decimal degrees (that extends from -180 to 180 degrees longitude and from 85 degrees latitude to -65 degrees latitude). Second, we intersect the resulting global grid with the 2000 *Digital Chart of the World* that portrays contemporary national borders; this results in $4,512$ pixels across the globe falling within country boundaries. The median (mean) pixel extends to $25,967$ ($29,780$) km^2, being comparable to the size of ethnic homelands in the GREG dataset, when we exclude those groups with less than 1 percent of a country's population ($20,338$ km^2). Third, for each

pixel (of $2.5x2.5$ decimal degrees) we compute luminosity per capita in 1992, 2000, and 2009. Fourth, we aggregate the data at the country level estimating a Gini coefficient that captures the overall degree of economic spatial inequality.

Spatial Gini Coefficient 2 Virtual countries created by the 2.5 by 2.5 degree boxes are on average somewhat larger than ethnic homelands; moreover, because of the fixed grid dimensionality, smaller countries end up having fewer boxes. Hence, to capture spatial inequality at a level of aggregation similar to the one in the data we also constructed an index of spatial inequality based on Thiessen polygons. The latter have the unique property that each polygon contains only one input point, and any location within a polygon is closer to its associated point than to the point of any other polygon. Importantly, we use as input points the centroids of the linguistic homelands according to the *Ethnologue* dataset. Thus, Thiessen polygons have the *exact* same centroid as the actual linguistic homelands in the *Ethnologue* database; the key difference being that ethnic homelands rather than being symmetric polygons have idiosyncratic shapes. We then intersect the $7,570$ Thiessen polygons with the country boundaries in 2000 obtaining a total of $9,116$ grids. We then construct a spatial Gini coefficient that reflects inequality in lights per capita across Thiessen polygons.[5] The mean size of the Thiessen polygons is 14.809 km, very similar to the mean size of homelands in the *Ethnologue* ($12,964$ km^2).

It is important to realize that both proxies of the overall degree of spatial inequality also reflect inequality across ethnic homelands, since (i) there is clearly some degree of measurement error on the exact boundaries of ethnic regions and (ii) because population mixing is in practice higher than the one we observe in the data. Moreover, in countries with large groups the spatial Gini coefficients may also (partially) capture within-ethnic-group inequality. We thus (almost) always include both the ethnic inequality and the overall spatial inequality index in the empirical specifications.[6]

2.5 Example

Figures $1a - 1b$ provide an illustration of the construction of the ethnic inequality measures for Afghanistan. The *Atlas Narodov Mira* maps 31 ethnicities (Figure $1a$). The Afghan is the largest group that consists of the Pashtuns and the Pathans residing in the southern and central-southern regions. This group takes up 51% of the population in 2000. The second largest group is the Tajik, who compose 22% of the population and are located in the north-eastern regions and in scattered

[5] To focus on non-trivial grids in terms of size for both the Thiessen polygons and the 2.5 by 2.5 decimal degree boxes we drop those polygons capturing an area of less than 100 square kilometers.

[6] In principle one could generate within-group inequality measures using the finer structure of the luminosity data. However, within-group mobility and risk sharing issues makes a luminosity-based, within-group inequality index less satisfactory. We perform a proper decomposition of between and within-group inequality in the last section using micro-level data from Africa. The micro-level data also allows us to account for migration (as quite often we observe individuals from the same ethnic groups in more than two regions in a country).

pockets in the western part of the country. There are 8 territories in which groups overlap. We first estimate for each of the 31 homelands luminosity per capita. For groups appearing in multiple pockets we derive the weighted average of lights per capita using as weights the fraction of each pocket's surface area to the total area of the group in the country. Figure 1*b* portrays the distribution of lights per capita. Regional development is low in the center of the country, where the Hazara-Berberi reside and in the eastern provinces, where the Nuristani, the Pamir Tajiks, the Pashai, and the Kyrgyz tribes are located. Luminosity is higher in the Pashtun/Pathans homelands and to some lesser extent in the Tajik regions. Second, using lights per capita across all homelands we estimate the Gini coefficient in 1992, in 2000, and in 2009. In 2000 the Gini coefficient estimated from GREG is 0.93 very close to the estimate when we use *Ethnologue* that maps 39 groups (0.90). We also estimated the ethnic inequality measures excluding the ethnic homeland where the capital, Kabul, falls; and we also estimated Gini coefficients of ethnic inequality excluding groups constituting less than 1% of the country's population.

Figures $2a - 2b$ illustrate the construction of the overall spatial inequality indicators (Gini coefficients) using the two different methods. When we divide the globe into pixels of 2.5 x 2.5 decimal-degree boxes we get 24 areas in Afghanistan (Figure 2*a*). When we use Thiessen polygons we get 56 pixels in Afghanistan (Figure 2*b*). After estimating for each pixel, average luminosity per capita, we aggregate at the country level calculating the Gini coefficient across these pixels. The resulting measures, overall spatial inequality Gini index 1 and 2 for Afghanistan equal 0.722 and 0.827, respectively.

2.6 Descriptive Analysis

2.6.1 Ethnic Inequality around the World

Table 1 - Panel *A* reports summary statistics, while Appendix Table 1, Panels *A* and *B* report the correlation structure of the ethnic Gini coefficients between the two global maps in different points in time. The correlation of the Gini coefficients across the two alternative mappings is strong, around $0.75 - 0.80$. 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. Given the high inertia, in our empirical analysis we will exploit the cross-country variation. The correlation between ethnic inequality and the overall spatial inequality indicators is high, but far from perfect; ranging between $0.6 - 0.8$.

Figures $3a-3d$ illustrate the global distribution of ethnic and spatial inequality. Africa (and South Asia) are the most ethnically unequal place(s) in the world. In contrast Western Europe is the region with the lowest level of ethnic inequality. According to the *Atlas Narodov Mira*, the countries with the highest ethnic group inequality are Sudan, Afghanistan, and Mongolia

(Gini index higher than 0.90). 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: Chad, Sudan, Papua New Guinea, Brazil, Ethiopia, Angola, Nigeria, Zimbabwe, Zaire, Cameroon, Laos and Indonesia. The countries with the highest overall spatial inequality in light density according the measure based on Thiessen polygons (spatial Gini 2 is higher than 0.90) are Chad, Papua New Guinea, Zaire, Gabon, Congo, the Central African Republic, and Sudan.

Since we are primarily interested in uncovering the explanatory power of ethnic inequality beyond the overall spatial inequality in most specifications we control for the latter. Figures $3e-3f$ portray the global distribution of ethnic inequality partialling out the effect of the overall degree of spatial inequality. In Figures $4a$ - $4b$ we plot ethnic inequality against the overall degree of spatial inequality. A few interesting patterns emerge. On the one hand, Sudan, Afghanistan, and Mongolia have much higher ethnic inequality as compared to the overall spatial inequality (which is also very high). In contrast, USA and Canada score low in ethnic inequality as compared to the overall degree of spatial inequality (which is high). On the other hand, Azerbaijan, Syria, Albania, Tunisia, Haiti, and Rwanda score quite high in ethnic inequality, while in contrast the overall degree of spatial inequality is quite low.

2.6.2 Basic Correlations

Ethnic Diversity Appendix Table 1 - Panel C reports the correlation structure between the various ethnic inequality and spatial inequality measures with the widely-used ethnolinguistic fragmentation measures. There is a positive correlation between ethnic inequality and linguistic-ethnic fractionalization $(0.38-0.45)$ (data come from Alesina, Devleeschauwer, Easterly, Kurlat, and Wacziarg (2003))). In contrast, there is no systematic association between ethnic inequality and religious fractionalization. Figures $5a-$ $5b$ provide a graphical illustration (including continental fixed effects does not change the pattern). The correlation between ethnic inequality and the segregation measures compiled by Alesina and Zhuravskaya (2011) is also positive $(0.20-0.45)$. Ethnic inequality tends to go in tandem with segregation. This is reasonable since more mixing of groups may lead to a reduction of ethnic-based inequality, which instead is more likely to persist when groups are geographically separated. We also examine the association between ethnic inequality and spatial inequality with the ethnic polarization indicators of Montalvo and Reynal-Querol (2005) and Esteban, Mayoral, and Ray (2012), failing to detect a systematic association. These results show that the ethnic inequality measure captures a dimension distinct from already-proposed aspects of a country's ethnic composition.

Income Inequality We then examined the association between ethnic inequality and income inequality, as reflected in the standard Gini coefficient (Appendix Table 1 - Panel D).

The income Gini coefficient is taken from Easterly (2007) who using survey and census data compiled from the WIDER (UN's World Institute for Development Economics Research) constructs adjusted cross-country Gini coefficients for more than a hundred countries over the period $1965-2000$. Figures $6a$ and $6b$ illustrate this association using the GREG and the *Ethnologue* mapping of ethnic homelands, respectively. The correlation between ethnic inequality and economic inequality is moderate, around $0.25-0.30$. Yet this correlation weakens considerably and becomes statistically insignificant once we simply condition on continental constants.

3 Ethnic Inequality and Development

In Table 3 we report cross-country LS estimates associating the log of per capita GDP in 2000, with ethnic inequality (Appendix Table 1 - Panel D reports the unconditional correlation of ethnic inequality with various proxy measures of economic and institutional development). In Panel A we use the ethnic inequality measure using the *Atlas Narodov Mira* database, while in Panel B we use the measures derived from *Ethnologue's* mapping. In all specifications we include region fixed effects to account for continental differences in the coverage of ethnic groups and the huge variation in economic development.

The coefficient of the ethnic inequality index in column (1) is negative and highly significant. Figures $7a-7b$ illustrate the unconditional association.[7] The estimates in columns (2) and (4) also reveal a negative association between development and the overall degree of spatial inequality, as reflected on the Gini coefficient based on pixels of 2.5 *by* 2.5 degrees and the Gini coefficient based on Thiessen polygons that have the same centroid as ethnic homelands in the *Ethnologue*. This suggests that underdevelopment goes in tandem with regional inequalities.[8] In columns (3) and (5) we include both the ethnic inequality index and the spatial Gini coefficients. The ethnic inequality index continues to enter with a highly significant estimate that falls only slightly in absolute value. In contrast the coefficient on the overall spatial inequality drops considerably in all permutations; moreover the estimate becomes statistically indistinguishable from zero. This suggests that the ethnic component of regional inequality is the relatively stronger correlate of underdevelopment.[9]

In columns (6)-(9) we add the log number of ethnic/linguistic groups in the empirical model.

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

[8] This -to the best of our knowledge novel- result is interesting by itself and deserves future work in understanding the inter-play between development and regional disparities in income (see Bolton and Roland (1997) for a theoretical exposition).

[9] There is clearly some degree of measurement error on the exact boundaries of ethnic homelands, while by construction there is no error on the spatial inequality measures. Additionally, to the extent that populations mix, the overall spatial inequality index also captures part of ethnic inequality. Both observations suggest that the coefficient of ethnic inequality on development is likely to be an underestimate of the true magnitudes.

In line with previous works, income per capita is significantly lower in countries with many ethnic (Panel A) and linguistic (Panel B) groups (column (6)); yet the estimates in columns (7)-(9) clearly show that it is ethnic inequality rather than ethnolinguistic heterogeneity that correlates with underdevelopment. In columns (10)-(11) we examine whether the significantly negative association between ethnic inequality and income per capita is driven by an unequal clustering of population across ethnic homelands; to do so we construct Gini coefficients of population combining the population estimates in 2000 from the Gridded Population of the World dataset with the mapping of ethnolinguistic groups. The population Gini index enters with a significantly negative estimate, implying that under-development is associated with an unequal clustering of population across ethnic regions. Yet once we include in the specification the ethnic inequality index and the overall spatial inequality indicators (in (11)), the population Gini coefficient index turns insignificant. The same applies with the spatial Gini coefficient. In contrast the ethnic inequality measure retains its economic and statistical significance.

The most conservative estimate on the ethnic inequality index in Panel A (1.08) implies that a reduction in the ethnic Gini coefficient by 0.25 (approximately one standard deviation, from the level of Nigeria where the ethnic Gini is 0.76 to the level of Namibia where the ethnic Gini is 0.50) is associated with a 31% (0.27 log points) increase in per capita GDP. The standardized beta coefficient of the ethnic inequality index is around $0.20 - 0.30$, quite similar to the works on the role of institutions on development (e.g., Acemoglu, Johnson, and Robinson (2001)).

3.1 Sensitivity Analysis

Other Aspects of the Ethnic Composition In Table 3 we investigate whether other dimensions of the distribution of the population across groups, related to fractionalization and polarization, rather than inequality across ethnic lines affect comparative development. In columns (1) and (5) we augment the specification with a fractionalization index (from Alesina, Devleeschauwer, Easterly, Kurlat, and Wacziarg (2003)) whereas in columns (2) and (6) we experiment with Fearon's (2003) cultural fragmentation index that adjusts the fractionalization index for linguistic distances among ethnic groups. Doing so has no effect on the coefficient on ethnic inequality that retains its economic and statistical significance. Moreover, the fractionalization indicators enter with unstable and statistically insignificant estimates.

Motivated by recent works highlighting the importance of polarization in columns (3), (4), (7), and (8) we condition on two alternative measures of ethnic polarization (from Montalvo and Reynal-Querol (2005) and Esteban, Mayoral, and Ray (2012); the latter adjusts for linguistic differences across groups). Ethnic inequality correlates strongly with development, while the polarization measures enter with insignificant estimates. We also estimated specifications including

both the polarization and the fractionalization indicators; in all perturbations the coefficient on ethnic inequality retains its statistical and economic significance.

Alternative Measures and Geographic Controls In Table 4 we augment the specification with additional controls and experiment with alternative ethnic inequality proxies. In columns (3), (4), (9), and (10) we use ethnic Gini coefficients that exclude ethnic regions where capitals fall. In columns (5), (6), (11), and (12) we use ethnic Gini indicators that exclude groups that constitute less than 1% of a country's population.[10] In all specifications we control for the overall degree of spatial inequality in lights per capita using the spatial Gini index that is based on Thiessen polygons and ethnic fractionalization.

In odd-numbered columns we control for a country's size with the log of population in 2000 and log land area, as ethnic heterogeneity, ethnic inequality, and the overall degree of spatial inequality are likely to be increasing in size. 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 diversity is higher in areas close to the equator (e.g., Michalopoulos (2012)). The ethnic inequality index enters with a negative and significant estimate across all permutations. In even-numbered columns 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 (2012) and include (on top of the size controls and latitude) an index of terrain ruggedness, distance to the coast, an index of gem quality, the percentage of each country with fertile soil and the percentage of tropical land (the Data Appendix gives detailed variable definitions). The negative correlation between ethnic inequality and income per capita remains strong. The coefficient on the ethnic inequality measures is quite similar to the more parsimonious specifications with the size controls only. Thus while still an unobserved or omitted country-wide factor may jointly affect development and ethnic inequality, the estimates clearly point out that the correlation does not reflect (observable) mean differences in geographical characteristics or continental disparities.

Other We performed numerous additional sensitivity checks. For example, we dropped from the estimation (typically small) countries with just one ethnic (or linguistic) group. The results remain stable (see Appendix Table 2). We also repeated estimation excluding iteratively observations from each continent or from each income group (following World Bank's classification); the results remain intact. We also estimate quantile (and median) regressions to explore potential heterogeneity in the correlation between ethnic inequality and development.[11] The

[10] Note that a priori there is no reason in excluding small groups, since ethnic hatred may be directed to tiny groups that control a significant portion of the economy (Chua (2003)).

[11] For example in the similar to column (3)-Table 2a specification, quantile regression estimates are in the range of -1.0 to -1.95.

coefficient on the ethnic inequality index is quite stable across quantiles.

4 Inequality in Geographic Endowments and Ethnic Inequality

4.1 On the Origins of Ethnic Inequality

Given the strong correlation between ethnic inequality and under-development it is intriguing to examine the roots of ethnic inequality.

We started our exploration of the origins of inequality across ethnic lines by examining the association between the ethnic inequality proxies and commonly-used historical variables that have been found to correlate with contemporary development. There is little evidence linking contemporary differences in ethnic inequality to the legal tradition (La Porta, Lopez-de-Silanes, Shleifer, and Vishny (1998)), the conditions that European settlers faced at the time of colonization (Acemoglu, Johnson, and Robinson (2001)), the share of Europeans in the population (Hall and Jones (1999) and Putterman and Weil (2010)), the inclusiveness of early institutions (Acemoglu, Johnson, Robinson, and Yared (2008)), state history (Bockstette, Chanda, and Putterman (2002)), and borders' design (Alesina, Easterly, and Matuszeski (2011)); for brevity we report these results in Appendix Table 3. These insignificant associations suggest that the strong negative correlation between ethnic inequality and development does not reflect the aforementioned aspects of history.

Then, motivated by the findings of Michalopoulos (2012) that ethnolinguistic diversity increases with geographic heterogeneity, we conjecture that geographic and ecological endowments play a role in explaining contemporary differences in income across ethnic lines. To the extent that land endowments affect the diffusion and adoption of technology, then ethnic-specific inequality in the distribution of geographic features would manifest itself in contemporary differences in well-being across groups. To construct proxies of geographic inequality, we first obtain geo-referenced data on elevation, land's suitability for agriculture, distance to the coast and presence of water bodies (lakes, rivers, and other streams) and construct for each ethnic area the mean value of each of these measures. We then derive Gini coefficients at a country level that reflect group-specific inequality in each of these dimensions. Following the same procedure to the one regarding the construction of spatial inequality in luminosity (see Section 2), we estimated measures of the overall degree of inequality in geographic endowments, constructing for each of the four geographic features two spatial Gini coefficients: one based on the *2.5 x 2.5* decimal degrees pixels and one based on Thiessen polygons.

In Table 5 we explore the association between ethnic inequality and these measures of inequality in geographic endowments across ethnic homelands. Across all permutations, all four ethnic Gini coefficients in geographic endowments enter with positive estimates suggesting that

ethnic-specific differences in endowments translate into larger contemporary disparities in ethnic development. Depending on the specification details -*GREG* or *Ethnologue* mapping, whether we use all homelands or drop ethnic regions where capitals fall or small groups, whether we condition on the level of geography and the overall degree of spatial inequality in each of the four geographic features- different Gini coefficients of geographic inequality enter with significant estimates. Thus while we cannot precisely identify which geographic feature(s) matter most, the message from Table 6 is that differences in geography across ethnic regions translate into differences in contemporary ethnic inequality.

We thus aggregate the four indexes of ethnic inequality in geographic endowments via principal components. The use of factor analysis techniques is 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. In line with this, there is strong positive correlation between the four Gini coefficients (see Appendix Table 4). Table 6 reports the results of the principal component analysis. The first principal component explains more than half of the common variance of the four measures of inequality in geographic endowments. The second principal component explains around 20% of the total variance, while jointly the third and fourth principal components explain a bit less than a fourth of the total variance. Interestingly, all four inequality measures load positively on the first principal component. Moreover, the eigenvalue of the first principal component is close or greater than two (one being the rule of thumb), while the eigenvalues of the other principal components are less than one. We thus focus on the first principal component, which given the significant positive loadings of all Gini coefficients, we label it "*inequality in geographic endowments across ethnic homelands*".[12]

In Figures $8a - 8b$ we plot the ethnic inequality in luminosity against the first principal component of inequality in ethnic-specific geographic endowments. There is a strong positive association, suggesting that differences in geography explain a sizable portion of contemporary differences in development across ethnic homelands.

In Table 7 we formally assess the role of ethnic-specific geographic inequality, as captured by the composite index of inequality in geographic endowments across ethnic-linguistic homelands on contemporary ethnic inequality. Columns (1) and (4) show that the strong correlation illustrated in the figures is not driven by continental differences. In columns (2) and (5) 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 (using Thiessen polygons with the same centroid as ethnic homelands). This has little effect on the coefficient of the

[12] We also estimated and incorporated in the factor analysis inequality across ethnic homelands on distance to the capital city and on the presence of natural resources (namely oil, gold, and diamonds). The results, however, are similar, in the sense that the first principal component is quite similar and adding more geographic Gini indicators does not affect much the estimates.

ethnic inequality in geographic endowments that retains its economic and statistical significance. In contrast the Gini coefficient based on Thiessen polygons that captures the spatial degree in geographic inequality enters with a small and statistically insignificant estimate. In columns (3) and (6) we control for the level effects of geography, augmenting the specification with mean elevation, land area under water, distance to the coast, and land suitability for agriculture. In all permutations the composite index reflecting differences in geographic endowments across ethnic homelands enters with a positive and highly significant coefficient.[13] The estimate in column (3) implies that a one-standard-deviation increase in the inequality in geography across ethnic homelands index (1.56 points, say from Mozambique to Malawi) translates into an 15 percentage points increase in the ethnic inequality index (somewhat more than half a standard deviation; see Table 1*A*).

4.2 Geographic Inequality and Development

Given the strong positive association between ethnic inequality -as reflected in lights per capita across ethnic homelands- and inequality in geographic endowments, it is interesting to examine whether contemporary development is systematically linked to the unequal distribution of geographic endowments across ethnic homelands. We thus estimated LS specifications associating the log of real GDP p.c. in 2000 with the composite index of ethnic-specific inequality in geography. While omitted-variables concerns cannot be eliminated, examining the role of inequality in geographic endowments across ethnic homelands on comparative development is useful in assuaging concerns that the estimates in Tables $2-4$ are driven by reverse causation. Moreover, geographic inequality can be thought of as an alternative "primitive" measure of economic differences across linguistic homelands (compared to the ethnic inequality index based on luminosity).

Results Table 8 reports the results. The coefficient on the proxy of ethnic inequality in geographic endowments in (1) and (4) is negative and highly significant suggesting that countries with sizable inequalities in geographic endowments across ethnic homelands are less developed. In columns (2) and (5) we condition on the overall degree of inequality in geography with the spatial Gini index based on Thiessen polygons, while in (3) and (6) we also control for land quality, elevation, land area under water, and distance to the coast. The coefficient on the inequality in geographic endowments across ethnic homelands index is negative in all permutations. The coefficient is statistically different than zero in all but one specifications. In contrast the estimate on the principal component that reflects the overall spatial inequality in geographic endowments

[13] Appendix Table 5 shows that the results are similar when we exclude from the estimation ethnic regions where capital cities fall and small ethnic groups consisting less than 1% of a country's population.

is quantitatively small, changes sign and is statistically insignificant.[14] The estimates in columns (1) and (3) imply that a one-standard-deviation increase in geographic inequality across ethnic homelands (1.5 points) decreases income per capita by approximately 30% (0.27 log points). These results further show that inequality across ethnic regions is a feature of under-development.

Further Evidence We also estimated two-stage-least-squares estimates associating geographic inequality across ethnic homelands to ethnic inequality in lights per capita in the first-stage and the component of ethnic inequality explained by geographic disparities across ethnic regions with log per capita GDP in 2000 in the second stage. While the 2SLS estimates do not necessarily identify the causal effect of ethnic inequality on development, they may be useful in accounting for measurement error in the proxy measure of development (lights per capita). The results (reported in Appendix Table 7) show the 2SLS estimate on the ethnic Gini coefficient is highly significant and quite similar in magnitude to the LS estimate.

We also estimated specifications linking development to both the ethnic inequality measure (based on lights per capita) and the composite index capturing inequality in geographic endowments across ethnic homelands. The results, shown in Appendix Table 8, show that once we condition on contemporary ethnic inequality differences in endowments across ethnic homelands lose their explanatory power. While some peculiar type of measurement error may explain this finding, this result indicates that inequality in geographic endowments across ethnic homelands affects contemporary development via its role on ethnic inequality.

Discussion The results in this section should not be interpreted as proving that unequal geography across ethnic lines necessarily "causes" ethnic inequality (and under-development). It is possible that certain groups for a plethora of reasons (e.g., faster early development, superior military technology, or genetic differences) conquered higher quality territories. In this regard the correlation between inequality in geographic endowments across ethnic lines and ethnic inequality in development (captured by lights per capita) indicates the sizable persistence of inequality. Hence, one might view an unequal ethnic geography as a manifestation of deeper ethnic differences. Nevertheless, even in this case, it is the presence of an inherently unequal geography in a country that partially allows these primordial ethnic differences to become salient (otherwise there would be no "better land" for stronger groups to conquer and every group would have the same land endowment).

[14] Appendix Table 6 reports otherwise identical specifications using the inequality measures that exclude from the estimation capitals and small ethnic groups. The results are similar.

5 Individual-Level Evidence from Sub-Saharan Africa and a Primer on the Channels

In this section we take a micro approach that explores within-country across-district variation in ethnic inequality and various development outcomes in a sample of 16 Sub-Saharan African countries. The analysis serves two purposes. First, it is quite useful exploring the association between inequality across (and within) ethnic groups and well-being using micro data and exploiting within (rather than across) country variation. Hence, instead of assigning parts of a country to a single (or more) groups with linguistic maps, we use self-reported data on ethnic identity and living conditions minimizing measurement error on the construction of ethnic inequality measures and also accounting for migration (as we observe people from the same ethnic group in many regions). Second, the plethora of questions of the Afrobarometer Surveys allows us to shed light on the channels that link ethnic inequality to well-being.

Our focus on Sub-Saharan Africa is natural, as Africa is by far the most ethnically diverse part of the world, while ethnic inequality is also quite high. Moreover, previous works suggest that a considerable portion of Africa's growth tragedy may be attributed to its ethnic diversity and patronage politics across ethnic lines (e.g., Easterly and Levine (1997), Franck and Rainer (2012)). In the same vein, the literature on the origins of African political and economic development - mostly in political science- places a key role to ethnic disparities in income (e.g., Robinson (2001)).

5.1 Ethnic Inequality and Well Being

5.1.1 Data

We use individual-level survey data from the 3rd round of the Afrobarometer surveys, that cover 16 Sub-Saharan African in 2005.[15] The surveys are based on interviews of a random sample of either $1,200$ or $2,400$ individuals in each country. We consider all individuals that have a clearly identified ethnic identity and answer the questions on individual well-being.[16] This is the case for $20,617$ out of $25,200$ respondents; individuals reside in $1,298$ districts. Overall there are 213 ethnic groups across the 16 countries. In each district there are on average 3 ethnic groups (range from 1 to 23 ethnicities). A (nice) feature of the data -that we exploit below- is that individuals from the same ethnic group are present in more than one districts.

We construct inequality measures at the district level using individual responses on an ordered $(1-5)$ living conditions index where a score of 1 indicates very bad conditions; 2 fairly bad;

[15] These countries are: Benin, Botswana, Ghana, Kenya, Lesotho, Madagascar, Malawi, Mozambique, Mali, Namibia, Nigeria, Senegal, South Africa, Tanzania, Uganda, and Zambia. The 3rd round of the Surveys was also conducted in Cape Verde and Zimbabwe but in these countries the ethnic identification question was not asked.

[16] We also drop individuals from ethnic groups with less than 10 respondents in each country (as in this case both the ethnic-specific mean and the variance of living conditions are likely to be rather imprecisely estimated).

3 neither good nor bad; 4 fairly good, and 5 very good. Based on these data we calculate Theil indicators reflecting the between-group and the within-group components of overall inequality (the results are similar with the Gini index; we prefer the Theil index as it can be decomposed to a between ethnic group and a within ethnic group component). Table 1-Panel B reports summary statistics of the overall, the between, and the within ethnic group inequality measures as well as the other measures we consider below.

The formulae below describe the between-group and the within-group components of inequality in district d. Namely, for n groups within a district, $y_{i,j}$ denotes the living conditions of individual j belonging to group i; y_i reflects the mean living conditions of group i, and Y denotes the total living conditions in a given district; n_i is the number of individuals belonging to group i and N is the total number of individuals in the district.

$$T_{between} = \sum_i \frac{y_i}{Y} \ln(\frac{\frac{y_i}{Y}}{\frac{n_i}{N}}) \qquad T_{within} = \sum_i \frac{y_i}{Y} \sum_j \frac{y_{i,j}}{y_i} \ln(\frac{\frac{y_{i,j}}{y_i}}{\frac{1}{n_i}})$$

5.1.2 District-Level Estimates

Table 9 reports the results of the within-country analysis that associates ethnic inequality at the district-level with well-being, as reflected in the $1-5$ living conditions index (in Panel A) and a $0-9$ education index (in Panel B). We augment the specification with the within-ethnic-group Theil index, so as to jointly examine the role of between and within-group inequality. In all specifications we control for the log number of ethnicities and the log of the number of respondents in each district (the Data Appendix gives variable definitions and sources for all variables, Table 1-Panels B and C report summary statistics while Appendix Table 9 provides the pairwise correlations). Odd-numbered columns report estimates across all districts, while in even-numbered columns we drop districts with only one ethnicity (as in these cases the between-ethnic-group inequality is zero by construction).

Specifications (1) and (2) in Panel A show that district-level development -as reflected in the average living condition across respondents- is significantly lower in districts with high levels of between-group inequality. The coefficient on the between-group inequality index implies that a 5% decrease in the Theil index (approximately two standard deviations; see Table 1-Panel B) is associated with a 0.31 point increase in the average living conditions in a region (close to half a standard deviation). The coefficient on the within-ethnicity Theil index is also negative and highly significant, suggesting that inequality within groups is also a feature of regional underdevelopment. The estimate on the within-group Theil index suggests that a 5% decrease in inequality (a bit more than one standard deviation) is associated with a 0.27 points increase in the average level of living conditions in a district. The results are similar when we use the average level of education at the district level as the dependent variable (in Panel B). Regions

inhabited by more educated respondents are characterized by a lower degree of both between and within-group disparities.

5.1.3 Individual-level Estimates

In columns (4) to (6) we report specifications associating between-group and within-group inequality at the district level with living conditions (Panel A) and education at (Panel B) at the individual level. Moving to the individual level allows us to condition on numerous individual characteristics (Table 1C reports summary statistics). Following Nunn and Wantchekon (2011) we control for the respondent's age and age squared, a gender indicator, a dummy variable indicating urban households, 22 religion fixed effects, and 25 occupational constants. We also control for the share of the district's population that is of the same ethnicity as the respondent. The estimates on the between and the within ethnic inequality measures in columns (3) and (4) suggest that the negative correlations with the development outcomes (living conditions and education) are not driven by (observable) individual characteristics.

Since members of the same ethnic group are present in more than one district, we also explore whether conditional on ethnic-specific (observable and unobservable) characteristics, inequality across ethnic lines is an important correlate of individual well-being and education. Overall almost all (210 out of the 213) ethnic groups in our sample may be found in more than one district. The median ethnicity can be found in 20 districts. Conditioning on ethnicity fixed effects seems a priori important, because recent works show that ethnic-specific historical traits, related, for example, to the slave trades (e.g. Nunn (2008)), pre-colonial political centralization (e.g., Gennaioli and Rainer (2007) and Michalopoulos and Papaioannou (2013)), and ethnic partitioning (e.g., Michalopoulos and Papaioannou (2011)) have long-lasting effects on development. The inclusion of country-ethnicity fixed effects also ensures that the negative association between ethnic inequality and development is not driven by certain ethnic groups -that may either dominate politics in one country or suffer from discrimination.

The estimates on columns (5) and (6), imply that conditional on an array of individual characteristics, respondents from the same ethnic group report worse living conditions when they reside in districts characterized by larger ethnic inequality; also individuals in ethnically unequal regions are also less educated. The estimate in column (6) in Panel B on the between-group Theil index (-3.03) suggests that a one standard deviation increase in ethnic inequality is associated with 0.25 standard deviation drop in the level of education. Interestingly within-group inequality is also negatively related to both living conditions and education; the coefficient on the within ethnic inequality Theil in column (6) of Panel B suggests that a one standard deviation increase in the within-group inequality measure is associated with a 0.42 drop in education.

Sensitivity Analysis We perturbed the empirical model in various ways to explore the robustness of these results. First, rather than conditioning on the log number of ethnicities we constructed a standard fractionalization measure. This has no effect on our main results. Second, we repeated estimation using the mean log deviation index or the Gini coefficient. The results are virtually unchanged. Third, we repeated the analysis restricting estimation to urban districts. This is useful as inequality is higher in urban places and, unlike the cross-country setting, we can properly account for increased population mixing usually associated with urbanization. Across all permutations, the coefficient on the between-ethnic-group Theil index is negative and significant at the 99% confidence level. Fourth, since the living conditions and the education measures take discrete values we estimated ordered probit models with maximum likelihood again finding similar results. To avoid cluttering the exposition the results are available upon request.

Summary To the best of our knowledge the results in Table 9 are the first piece of evidence showing in a systematic manner with micro level data that inequality, both across ethnic lines and within ethnicities, is inversely related to economic well-being in Sub-Saharan Africa.

5.2 Channels

We now exploit the richness of the questionnaires of the Afrobarometer Surveys to explore several channels linking ethnic inequality and well-being.

5.2.1 Economic Conditions and Political Power

Given the large amount of evidence on ethnic-based patronage politics and discrimination in Africa, we begin our analysis on the mechanisms linking ethnic inequality to under-development examining the link of a group's well-being and its political power. Economic differences across groups may translate into differences in the influence that ethnicities exert on the political landscape (and vice versa). We test this conjecture tabulating respondents' answers to a question aiming to measure the political influence of their group to that of other groups in the same country; the question gauging the political influence of the group reads: "*Think about the conditions of [respondent's identity group]. Do they have less, the same, or more influence in politics than other groups in this country?*" Higher values in this index -that ranges from 1 to 5- indicate more influence (see the Data Appendix for details).

Figure 9*a* portrays the relationship between the average of a group's mean living conditions and its perceived political leverage (partialling out country-specific constants and the size of each group). The evident positive association indicates that the economic standing of a group is closely linked to its political power. Interestingly, the standardized "beta" coefficient of a group's

mean living conditions (0.36) is quite similar to the (similarly positive) "beta" coefficient of the log number of respondents of the group (0.43), implying that the role of economic well-being on political power is similar to that of group size; the latter is consistent with the recent findings of Francois, Rainer, and Trebbi (2013) who document that a group's size determines its representation in the government across most Sub-Saharan African countries.

A similar picture emerges when we look at how a group's living conditions relate to whether the group's respondents feel discriminated by the government; we derive a proxy of group discrimination using the average value to the following question: "*How often are [respondent's identity group] treated unfairly by the government?*" The index ranges from 0 to 3 with higher values indicating a higher frequency of unfair treatment. Figure 9*b* shows a strong association between group-specific living conditions and group's unfair treatment by the central government.

We also experimented with alternative measures of a group's economic conditions. Specifically, instead of measuring group's economic status by taking the average of living conditions across individuals, we use information from another question that reflects perceptions of group members regarding the economic power of their own group relative to other groups in the country. The question reads: "*Think about the condition of [respondent's identity group]. Are their economic conditions worse, the same as, or better than other groups in this country?*" Higher values (range 1 to 5) indicate higher economic prosperity. There is a strong link between group-specific economic conditions (versus other groups) and political influence or (lack thereof) discrimination by the government. Please see Figures 10*a*, 10*b*, 10*c*, respectively.

5.2.2 Public Goods Provision

The Argument A large literature provides compelling evidence that public goods provision, redistributive policies, and effective governance are less prevalent in ethnically/racially diverse communities (e.g., Alesina, Glaeser, and Sacerdotte (2001)) and countries (Desmet, Ortuño-Ortín, and Wacziarg (2012)). The more dispersed preferences are across groups then the desired public goods will be more distant to the chosen one (see Luttmer (2001) for evidence). Hence, if the level of income shapes preferences for public goods then group differences in economic conditions will make group preferences diverge leading to lower public goods provision and increased political tensions.[17] In the same vein, one of the main empirical patterns of urban economics is that the rich quite often want to "isolate"; since this desire may be especially strong when wealth

[17] Baldwin and Huber (2010) provide empirical evidence in line with this idea for 46 democracies. Similarly, Deshpande (2000) and Anderson (2011) focus on income inequality across castes in India and associate between-caste inequality to public goods. See also Loury (2002) for an overview of works studying the implications of the evolution of racial inequality in the US.

is correlated with group identity, ethnic inequality may lead to segregation.[18] The link between political inequality and economic inequality (shown above) may limit the supply of public goods to geographically isolated minorities, with an insufficient supply to poorer and less powerful groups. In contrast to the least well-off, the "rich" may get the desired services (of public goods) by directly purchasing the respective private goods; obtaining, for example, a private electricity generator instead of contributing to the extension of the public grid. This phenomenon would exacerbate the under-provision of public goods.

Results Table 10*a* presents results relating inequality to access to various type of public goods, namely piped water, sewage systems and electrification at the enumeration area. The unit of analysis in columns (1), (4) and (7) is the district, while in the other columns the unit of analysis is the individual. To further account for individual features, following Nunn and Wantchekon (2011) in these specifications we also include as additional controls five dummies for household's living conditions and nine indicators for education. To economize on space we focus on districts with more than one ethnic group (the pattern is the same for the full sample as already shown in Table 9). In all specifications the between-group inequality measure enters with a significantly negative sign. On the contrary, the explanatory power of within-group inequality index is small and the coefficient is often statistically indistinguishable from zero. For example, in column (5) the beta coefficients for the between and within-group inequality are 0.049 and 0.006, respectively. The results in Table 10*a* thus show that in districts marked by sizable inequalities across ethnic lines, there is under-provision of public goods. Since we do not have random assignment on ethnic inequality, these estimates do not necessarily identify causal effects; yet it is important to note that we control for numerous individual level characteristics, including education, living conditions, occupation, and in many specifications even for ethnicity fixed factors.

In Table 10*b* we report otherwise identical to Table 10*a* specifications but we now include as an additional regressor the mean value of the living conditions index in each district. (Note that in the individual-level regressions we continue controlling for individual level living conditions and education). Access to public goods is negatively related to between-group inequality; in contrast the correlation between public goods provision and within-group inequality is weak. This suggests that the effect of access to public goods goes above and beyond the effect of ethnic inequality on average well being, since we are controlling for it in these regressions.

Examples A few examples are useful to illustrate the pattern. We focus on the most restrictive specifications that include ethnicity fixed-effects. The Pular in Senegal are found in 28

[18] Alesina and Zhuravskaya (2011) show that ethnic and linguistic segregation correlate negatively with proxies of effective governance.

of the 31 country's districts. In the district of Matam, where the Pular coexist with the Soninke, the Wolof and the Mandinka, between-group inequality is minimal (0.0005) whereas in the district of Sedhiou, where the Pular coexist with the Wolof, the Mandinka, the Manjack, the English, the Diola, and the Bambara, between-group inequality is 0.0145. In the Sedhiou district all Pular report having no access to electricity, piped water and sewage system, whereas in Matam 72% of the Pular have access to an electricity grid and access to clean water. Another example is the Herero that are found in 32 of the 87 districts in Namibia. In the district of Otjiwarongo, where we observe respondents from 5 groups, between-group inequality is minimal, 0.0038. In this region all Herero reply having access to electricity, sewage system, and clean water. On the contrary, in the district of Otjinene where between-group inequality is more than 10 times larger, 0.0408, only 43% of Herero reply having access to either an electricity grid or a sewage system (in both regions within-group inequality is quite similar).

5.2.3 Fairness, Markets and Democracy

The argument In an influential work, Chua (2003) notes cases in which the spread of democracy and free market institutions in the 1990s has led to animosity and institutional capture by amplifying pre-existing ethnic tensions. Her thesis is that in the presence of ethnic inequality (and especially when a small fraction of the population controls most wealth in the economy), then democratization may result in a backlash, as most individuals perceive representative and capitalist institutions as unfair, captured, and corrupt. [19]

When the less-privileged, but more populous groups, come to power they may want to turn the cards around, pursuing ethnic politics aiming to compensate their group for the perceived injustice. The resulting belief that markets are unfair may also lead to under-supply of effort (as in model by Alesina and Angeletos (2005) and Benabou and Tirole (2006)). The strong correlation documented above between well being and political influence goes in the same direction. It undermines the perception of "equity" in the society.

Results Tables 11a and 11b report the results of our analysis examining the link between ethnic inequality (and within-ethnic-group) inequality and perceptions on the functioning of democratic institutions. In columns (1)-(3) the dependent variable captures whether the respondent feels that in some instances a non-democratic government is preferable to a democratic one. In columns (4)-(6) the dependent variable is a $0-4$ range index reflecting the degree of satisfaction with the way democracy works in the respondent's country, whereas in columns (7)-(9)

[19] Chua (2003) presents case-study evidence supporting her thesis. She discusses, among others, the influence of Chinese minorities in Philippines, Indonesia, and other Eastern Asian countries; the dominant role of (small) Lebanese communities in Western Africa and the similarly strong influence of Indian societies in Eastern Africa. Other examples, include the I(g)bo in Nigeria and the Kikuyu in Kenya.

the dependent variable is an ordered variable (range from 1 to 5) where reflecting respondents' belief on whether the constitution expresses the values and hopes of people. (So in all specifications higher values in the dependent variable reflect greater support/satisfaction with the democracy/constitution). To get an idea of the overall variation in the outcome of interest 23% of the respondents either (strongly) believe that the constitution does not reflect the values of the people of the country with a similar 19% answering the opposite.

The evidence in Table 11*a* reveals that support and especially satisfaction with democracy is overall negatively related to across-group inequality and to some extent to within-group inequality. It is worthwhile to note that amongst the individual controls we always include the living conditions and education level of the respondent, hence the negative association between ethnic inequality and support for democracy operates beyond the influence of the former on the well-being of the respondent. In Table 11*b* we also control for the overall level of well being of the district as a whole. Doing so weakens considerably the role of cross-group inequality and eliminates any influence of the within-group component. This finding suggests that the link between ethnic inequality and disaffection with democratic institutions works primarily (though not entirely) through the relationship of ethnic inequality and district-level wellbeing documented above.

5.2.4 Other Channels

We also examined the link between ethnic inequality and proxies of social-civic capital investigating whether ethnic inequality is related to trust. The results are inconclusive. It seems that in Africa at least ethnic inequality and trust are not systematically linked.

Finally, we correlated inequality across and within ethnic groups with proxies of conflict at the regional level, using geo-referenced data on conflict from the ACLED database. While many have linked (ethnic and overall) inequality with conflict (see Horowitz (1985) for an overview), as shown theoretically by Esteban and Ray (2011) and Esteban, Mayoral, and Ray (2012) the relationship is not straightforward; for instance, a very rich and powerful group may be so strong that any insurrection from (much) weaker groups is unthinkable. Likewise, a change in ethnic inequality starting from a relatively equal equilibrium may generate violence. Given the theoretical ambiguities, perhaps it comes at no surprise that we could not detect a systematic relationship between ethnic (or within group) inequality and conflict.[20]

[20] To economize on space, we do not report these results, which are available upon request.

6 Conclusion

This study shows that ethnic differences in economic performance rather than the degree of diversity or the overall level of inequality are negatively correlated with economic development. While a large literature has examined (a) the interplay between inequality and development and (b) the effects of various aspects of the ethnic composition (such as fragmentation, polarization, segregation) on economic performance, there is little -if any- work studying the linkages between ethnicity, inequality, and comparative development. This paper is a first effort to fill this gap.

In the first part of the paper we examine the role of ethnic inequality on development using cross-country data. First, combining linguistic maps on the spatial distribution of ethnic groups within countries with satellite images of light density at night we construct Gini coefficients reflecting inequality in well-being across ethnic lines for a large number of countries. Ethnic inequality is weakly correlated with the standard measures of income inequality and only modestly correlated with ethnolinguistic fractionalization, polarization, and segregation. Second, we show that the newly constructed proxy of ethnic inequality is strongly negatively correlated with per capita GDP across countries. The correlation retains its significance when we condition on the overall degree of spatial inequality in development, which is also negatively associated with economic development (a new finding by itself). Including in the empirical specification both the ethnic inequality index and the widely-used ethnolinguistic fragmentation indicators, the latter loses significance, suggesting that it is inequality across ethnic lines that is correlated with poor economic performance rather than fractionalization. Third, we conduct an initial step exploring the roots of contemporary differences in well-being across ethnic groups within countries. In this regard, we construct indicators of ethnic inequality in geographic endowments and show that contemporary differences in development across ethnic homelands have a significant geographic component. The latter is also inversely related to contemporary development.

In the second part of the paper we take a micro approach exploiting within-country differences in ethnic inequality for 16 Sub-Saharan countries covered by the Afrobarometer Surveys. First, using individual-level data on ethnic self-identification and well-being we find a negative association between ethnic inequality and regional development. The negative correlation between ethnic inequality and well-being retains significance even when we control for numerous individual characteristics and account for unobserved ethnic-specific traits. Second, we explore the mechanisms linking ethnic inequality to regional development. We show that economic inequality goes in tandem with political inequality. Our analysis further establishes that ethnic inequality is closely linked to the under-provision of basic public goods and to some extent to the lack of confidence in democratic institutions.

Overall, we view our work as a first step towards mapping and understanding the conse-

quences and origins of contemporary differences in income across ethnic groups. Future research should further explore the channels via which ethnic inequality interacts with comparative development. Moreover, theoretical work is needed to shed light on the interaction between ethnic identity, inequality, and various aspects of economic efficiency. We plan on tackling some of these questions in future work.

References

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

Acemoglu, D., S. Johnson, J. A. Robinson, and P. Yared (2008): "Income and Democracy," *American Economic Review*, 98(3), 808–842.

Alesina, A., and G.-M. Angeletos (2005): "Fairness and Redistribution: US vs Europe," *Ameican Economic Review*, 95(4), 960–980.

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg (2003): "Fractionalization," *Journal of Economic Growth*, 8(2), 155–194.

Alesina, A., W. Easterly, and J. Matuszeski (2011): "Artificial States," *Journal of the European Economic Association*, 9(2), 246–277.

Alesina, A., and E. L. Ferrara (2005): "Ethnic Diversity and Economic Performance," *Journal of Economic Literature*, 43, 762–800.

Alesina, A., E. Glaeser, and B. Sacerdotte (2001): "Why Doesn't the United States Have a European Welfare State?," *Brookings Papers on Economic Activity*, 2, 187–278.

Alesina, A., and E. Zhuravskaya (2011): "Segregation and the Quality of Government in a Cross-Section of Countries," *American Economic Review*, 101(5), 1872–1911.

Anderson, S. (2011): "Caste as an Impediment to Trade," *American Economic Journal: Applied Economics*, 3(1), 239–263.

Baldwin, K., and J. D. Huber (2010): "Economic versus Cultural Differences: Forms of Ethnic Diversity and Public Goods Provision," *American Political Science Review*, 104(4), 644–662.

Benabou, R. (2005): "Inequality, Technology, and the Social Contract," in *Handbook of Economic Growth*, ed. by P. Aghion, and S. Durlauf, vol. 1, chap. 25, pp. 1595–1638. Elsevier, North-Holland, Netherlands.

Benabou, R., and J. Tirole (2006): "Belief in a Just World and Redistribute Politics," *Quarterly Journal of Economics*, 121(2), 699–746.

Bockstette, V., A. Chanda, and L. Putterman (2002): "States and Markets: The Advantage of an Early Start," *Journal of Economic Growth*, 7(4), 347–369.

Bolton, P., and G. Roland (1997): "The Breakup of Nations: A Political Economy Analysis," *Quarterly Journal of Economics*, 112(4), 1057–1090.

Cederman, L.-E., N. B. Weidman, and K. S. Gleditch (2011): "Horizontal Inequalities and Ethnonationalist Civil War: A Global Comparison," *American Political Science Review*, 105(1), 478–495.

Chen, X., and W. D. Nordhaus (2011): "Using Luminosity Data as a Proxy for Economic Statistics," *Proceeedings of the National Academy of Sciences*, 108(21), 8589–8594.

Chua, A. (2003): *World on Fire: How Exporting Free-Market Democracy Breds Ethnic Hatred and Global Instability*. William Heinemann, London, UK.

Deshpande, A. (2000): "Does Caste Still Define Disparity? A Look at Inequality in Kerala, India," *American Economic Review, Papers and Proceedings*, 90(2), 322–325.

Desmet, K., I. Ortuño-Ortín, and R. Wacziarg (2012): "The Political Economy of Ethnolinguistic Cleavages," *Journal of Development Economics*, 97(2), 322–338.

Easterly, W. (2007): "Inequality Does Cause Underdevelopment," *Journal of Development Economics*, 84(2), 755–776.

Easterly, W., and R. Levine (1997): "Africa's Growth Tragedy: Policies and Ethnic Divisions," *Quarterly Journal of Economics*, 112(4), 1203–1250.

Esteban, J., L. Mayoral, and D. Ray (2012): "Ethnicity and Conflict: An Empirical Study," *American Economic Review*, 102(4), 1310–1342.

Esteban, J., and D. Ray (2011): "Linking Conflict to Inequality and Polarization," *American Economic Review*, 101(4), 1345–1374.

Franck, R., and I. Rainer (2012): "Does the Leader's Ethnicity Matter? Ethnic Favoritism, Education and Health in Sub-Saharan Africa," *American Political Science Review*, forthcoming.

Francois, P., I. Rainer, and F. Trebbi (2013): "How Is Power Shared In Africa?," Working Paper, NBER, No. 18425.

Galor, O. (2011): "Inequality, Human Capital and the Process of Development," in *Handbook of the Economics of Education*, ed. by S. M. Eric A. Hanushek, and L. Woessmann. Elsevier, Amesterdam, Holland.

Gennaioli, N., and I. Rainer (2007): "The Modern Impact of Precolonial Centralization in Africa," *Journal of Economic Growth*, 12(3), 185–234.

Gordon, R. G. (2005): "Ethnologue: Languages of the World," .

Hall, R. E., and C. I. Jones (1999): "Why Do Some Countries Produce So Much More Output Per Worker Than Others?," *Quarterly Journal of Economics*, 114(1), 83–116.

Henderson, V. J., A. Storeygard, and D. N. Weil (2012): "Measuring Economic Growth from Outer Space," *American Economic Review*, 102(2), 994–1028.

Heston, A., R. Summers, and B. Aten (2011): "Penn World Table Version 7.0," Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.

Horowitz, D. L. (1985): *Ethnic Groups in Conflict*. University of California Press, Berkeley and Los Angeles, California.

Kaufmann, D., A. Kraay, and M. Mastruzzi (2005): "Governance Matters IV : Governance Indicators for 1996-2004," *Policy Research Working Paper Series, The World Bank*.

Kyriacou, A. P. (2013): "Ethnic Group Inequalities and Governance: Evidence from Developing Countries," *Kyklos*, 66(1), 78–101.

La Porta, R., F. Lopez-de-Silanes, A. Shleifer, and R. Vishny (1998): "Law and Finance," *Journal of Political Economy*, 106(6), 1113–1155.

——— (1999): "The Quality of Government," *Journal of Law Economics and Organization*, 15(1), 222–279.

Loury, G. C. (2002): *The Anatomy of Racial Inequality*. Harvard University Press, Cambridge, MA.

Luttmer, E. F. (2001): "Group Loyalty and the Taste for Redistribution," *Journal of Political Economy*, 109(3), 500–528.

Michalopoulos, S. (2012): "The Origins of Ethnolinguistic Diversity," *American Economic Review*, 102(4), 1508–1539.

Michalopoulos, S., and E. Papaioannou (2011): "Divide and Rule or the Rule of the Divided? Evidence from Africa," NBER Working Paper 17184.

——— (2012): "National Institutions and Sub-national Development in Africa," NBER Working Paper 18275.

——— (2013): "Pre-colonial Ethnic Institutions and Contemporary African Development," *Econometrica*, 81(1), 113–152.

Montalvo, J.-G., and M. Reynal-Querol (2005): "Ethnic Polarization, Potential Conflict, and Civil Wars," *American Economic Review*, 95(3), 796–816.

Nunn, N. (2008): "The Long Term Effects of Africa's Slave Trades," *Quarterly Journal of Economics*, 123(1), 139–176.

Nunn, N., and D. Puga (2012): "Ruggedness: The Blessing of Bad Geography in Africa," *Review of Economics and Statistics*, 94(1), 20–36.

Nunn, N., and L. Wantchekon (2011): "The Slave Trade and the Origins of Mistrust in Africa," *American Economic Review*, 101(7), 3221–3252.

Pinkovskiy, M. (2011): "Economic Discontinuities at Borders: Evidence from Satellite Data on Lights at Night," mimeo MIT.

Putterman, L., and D. N. Weil (2010): "Post-1500 Population Flows and the Long-Run Determinants of Economic Growth and Inequality," *Quarterly Journal of Economics*, 125(4), 1627–1682.

Robinson, J. A. (2001): "Social Identity, Inequality, and Conflict," *Economics of Governance*, 2(2), 85–99.

Weidmann, N. B., J. K. Rod, and L.-E. Cederman (2010): "Representing Ethnic Groups in Space: A New Dataset," *Journal of Peace Research*, 47(4), 491–499.

7 Data Appendix

7.1 Cross-Country 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).*

Population: Log population in 2000. *Source: Penn World Tables, Edition 7. Heston, Summers, and Aten (2011).*

Land Area: Log surface area. *Source: Nunn and Puga (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 (Kaufmann, 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 Fractionalization: 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 ethnolinguistic/religious groups. *Source: Alesina et al. (2003).*

Ethnic/Linguistic/Religious Segregation: Index ranging from zero to one capturing ethnic/linguistic/religious segregation (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).

Ethnolinguistic Polarization 1: Index of ethnolinguistic polarization that achieves a maximum score when a country is occupied by two groups of the same population. *Source: Montalvo and Reynal-Querol (2005a,b).*

Ethnolinguistic Polarization 2: The polarization index accounts for the degree of similarity between linguistic groups using the *Ethnologue* linguistic tree. *Source: Esteban, Mayoral, and Ray (2012).*

Cultural Fragmentation: Index of ethnolinguistic fractionalization that accounts for the degree of similarity between linguistic groups using the *Ethnologue* linguistic tree. *Source: Fearon (2003).*

Soil quality: Percentage of each country with fertile soil. *Source: Nunn and Puga (2012).*

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 (2012).*

Tropical: The percentage of the land surface of each country with tropical climate. *Source: Nunn and Puga (2012).*

Desert: The percentage of the land surface area of each country covered by sandy desert, dunes, rocky or lava flows. *Source: Nunn and Puga (2012).*

Latitude: Absolute latitude is expressed in decimal degrees, for the geographical centroid of the country. *Source: Nunn and Puga (2012).*

Gem-Quality Diamond Extraction: Carats of gem-quality diamond extraction between 1958 and 2000, normalized by land area. *Source: Nunn and Puga (2012).*

Common Law: Indicator variable that identifies countries that have a common law legal system. *Source: La Porta, Lopez-de-Silanes, Shleifer, and Vishny (1999) and Nunn and Puga (2012).*

European Descent: The variable, calculated from version 1.1 of the migration matrix of *Putterman and Weil (2010)*, estimates the percentage of the year 2000 population in every country that is descended from people who resided in Europe in 1500. *Source: Nunn and Puga (2012).*

Settler Mortality: Log of mortality rates faced by European colonizers in late 19th century. *Source: Acemoglu, Johnson, and Robinson (2001).*

State Antiquity: Normalized state antiquity Index in 1950, using a 1% discount rate. *Source: Bockstette, Chanda, and Putterman (2002).*

Border Straightness Index: The $0-1$ index reflects how straight -and thus most likely to be non-organic- 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: The region constants correspond to: South Asia, East Asia and Pacific, Latin America and the Caribbean, North America, Western Europe, Eastern Europe and Central Asia, Middle East and Northern Africa, and Sub-Saharan Africa. The classification follows World Bank's World Development Indicators.

Light Density at Night: Light density is calculated averaging light density observations across pixels that fall within each territory (ethnic/linguistic homeland, Thiessen polygon, and pixel) and then dividing by population.
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: Constructed 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: Michalopoulos (2012); 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*

7.2 Micro-Level Data from Afrobarometer Surveys (3^{rd} Round)

Living Conditions: Respondent's view of their present living conditions. The question (Q4B) reads "*In general, how would you describe your own present living conditions?*". The answers 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 all individuals in each district. *Source: 2005 Afrobarometer Surveys.*

Education: Respondent's education/schooling. The question (Q90) reads "*What is the highest level of education you have completed?*". The answers are: 0=No formal schooling, 1=In-

formal schooling (including Koranic schooling), 2=Some primary schooling, 3=Primary school completed, 4=Some secondary school/ High school, 5=Secondary school completed/High school, 6=Post-secondary qualifications, other than university e.g. a diploma or degree from a technical/polytechnic/college, 7=Some university, 8=University completed, 9=Post-graduate. For the district-level analysis responses are averaged across all individuals in each district. *Source: 2005 Afrobarometer Surveys.*

Access to piped water: Response to the question (Q116e) 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 within a district. Question was filled in conjunction with field supervisor. For the district-level analysis responses are averaged across all individuals in each district. *Source: 2005 Afrobarometer Surveys*

Access to sewage system: Response to the question (Q116f) on "*whether in the enumeration area there is a sewage system that most houses could access*". For the district-level analysis responses are averaged within a district. Question was filled in conjunction with field supervisor. For the district-level analysis responses are averaged across all individuals in each district. *Source: 2005 Afrobarometer Surveys.*

Access to an electricity grid: Response to the question (Q116d) on "*whether in the enumeration area there is an electricity grid that most houses could access*". For the district-level analysis responses are averaged within a district. Question was filled in conjunction with field supervisor. For the district-level analysis responses are averaged across all individuals in each district. *Source: 2005 Afrobarometer Surveys*

Support for Democracy: Ordered (0, 1, 2) variable capturing support for democracy, based on the following question (Q37). "*Which of these three statements is closest to your own opinion? A: Democracy is preferable to any other kind of government. B: In some circumstances, a non-democratic government can be preferable. C: For someone like me, it doesn't matter what kind of government we have.*" We assign a value of 2 for individuals who reply A, a value of 1 for individuals who choose B, and a 0 for individuals who choose C. For the district-level analysis responses are averaged across all individuals in each district. *Source: 2005 Afrobarometer Surveys.*

Satisfaction with Democracy: Ordered (0, 1, 2, 3, 4) variable reflecting satisfaction with democracy. The question (Q47) reads. "Overall, how satisfied are you with the way democracy works in [Ghana/Kenya/etc.] are you?" The replies are 0=My country is not a democracy, 1=Not at All Satisfied, 2=Not Very Satisfied, 3=Fairly Satisfied, 4=Very Satisfied. For the district-level analysis responses are averaged across all individuals in each district. *Source: 2005 Afrobarometer Surveys.*

Constitution: Ordered variable (1, 2, 3, 4, 5) reflecting people's beliefs (satisfaction) about whether the constitution expresses the values and hopes of people. The question (Q52A) reads: Constitution expresses values and hopes. Do you disagree or agree with the following statement: Our constitution expresses the values and hopes of the [Ghanaian/Kenyan/etc.] people. : 1=Strongly Disagree, 2=Disagree, 3=Neither Agree Nor Disagree, 4=Agree, 5=Strongly Agree. For the district-level analysis responses are averaged across all individuals in each district. *Source: 2005 Afrobarometer Surveys.*

Ethnic-level Economic Conditions (power): Ethnic-level average of individual responses to the following question (Q80A) reflecting economic conditions of each group in the country. "Think about the condition of _ _ _ _ _ _ _ _ _ _ _ _ _ _ [respondent's identity group]. Are their economic conditions worse, the same as, or better than other groups in this country? " 1=Much Better, 2=Better, 3=Same, 4=Worse, 5=Much Worse. *Source: 2005 Afrobarometer Surveys.*

Ethnic-level Political Conditions (power): Ethnic-level average of individual responses to the following question (Q80B) reflecting political conditions of each group in the country. "Think about the condition of _ _ _ _ _ _ _ _ _ _ _ _ _ _ [respondent's identity group]. Are their political conditions worse, the same as, or better than other groups in this country? " 1=Much Better, 2=Better, 3=Same, 4=Worse, 5=Much Worse. *Source: 2005 Afrobarometer Surveys.*

Ethnic-level Unfairness (Discrimination): Ethnic-level average of individual responses to the following question (Q81) reflecting unfairness (discrimination) of each group by the government of each country."How often are _ _ _ _ _ _ _ _ _ _ _ _s [respondent's identity group] treated unfairly by the government?" 0=Never, 1=Sometimes, 2=Often, 3=Always. *Source: 2005 Afrobarometer Surveys.*

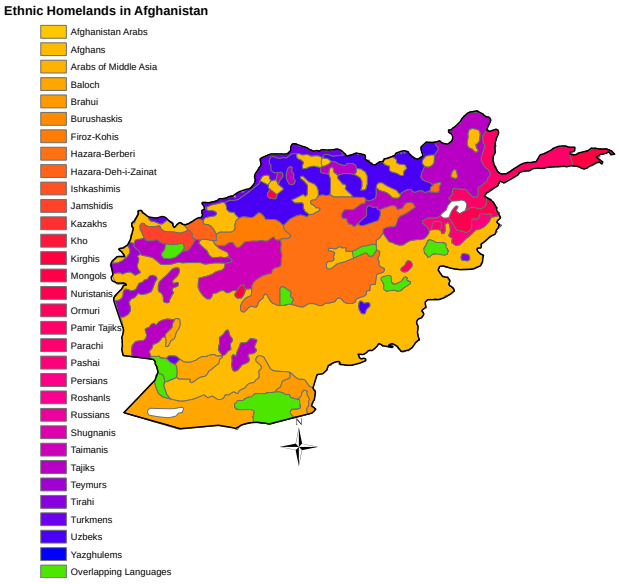

Figure 1a
Ethnic Homelands in Afghanistan

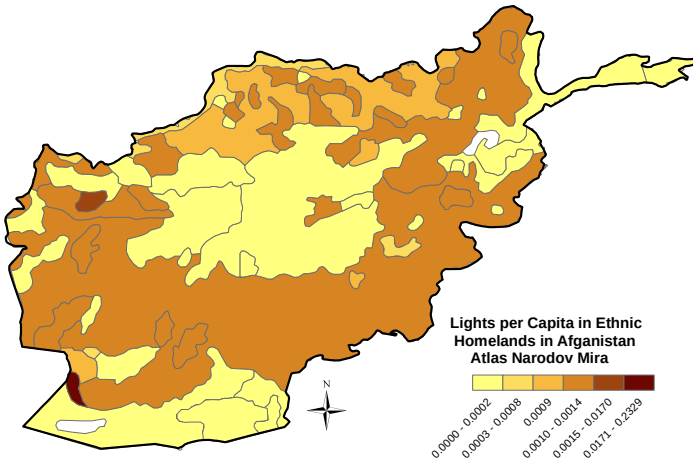

Figure 1b
Lights per Capita across Ethnic Homelands

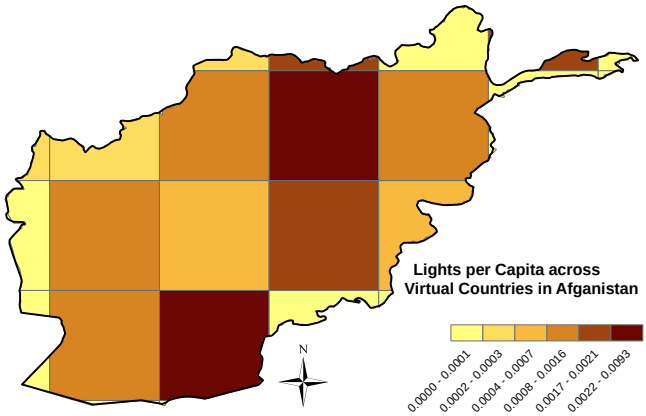

Figure 2a: Lights across 2.5 by 2.5 dd Boxes

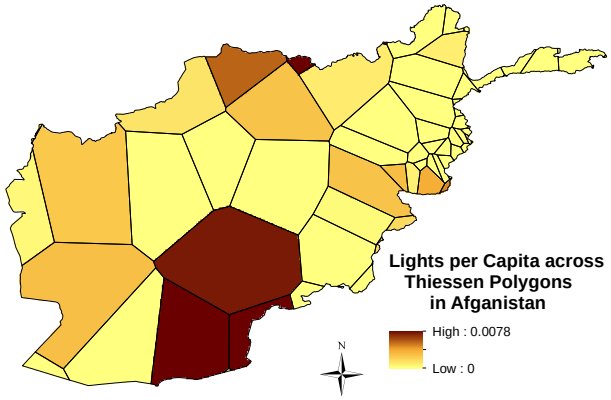

Figure 2b: Lights across Thiessen Polygons

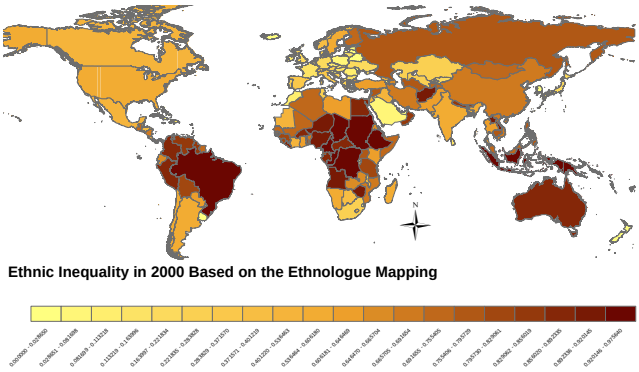

Figure 3a

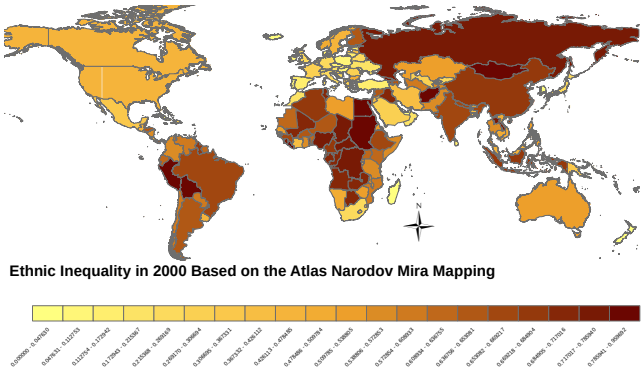

Figure 3b

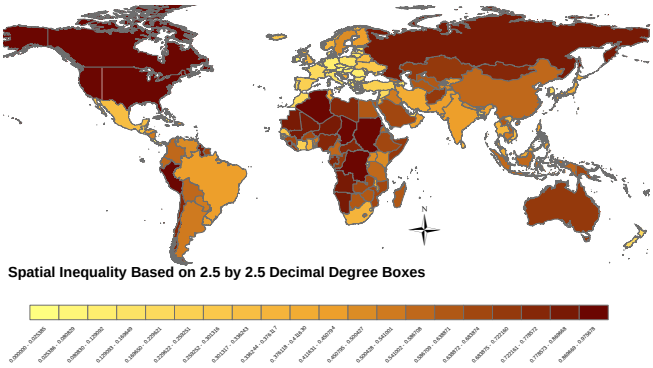

Figure 3c

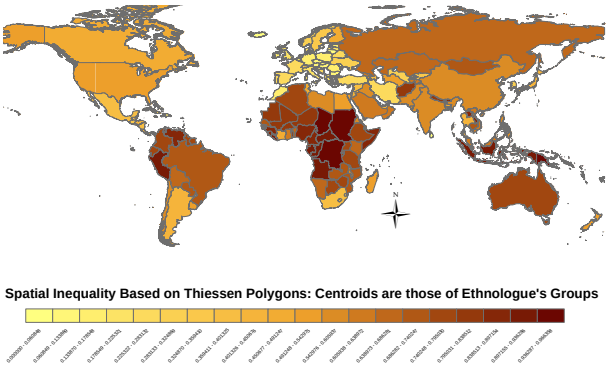

Figure 3d

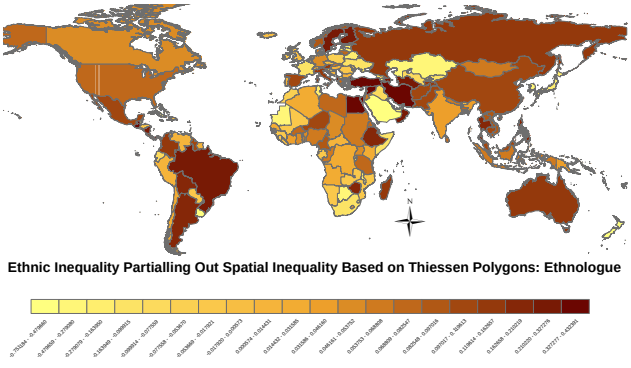

Figure 3e

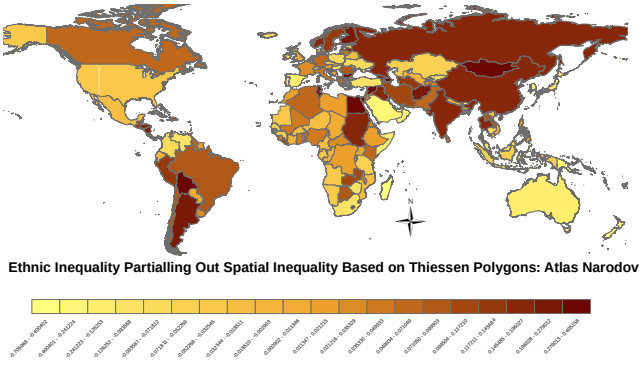

Figure 3f

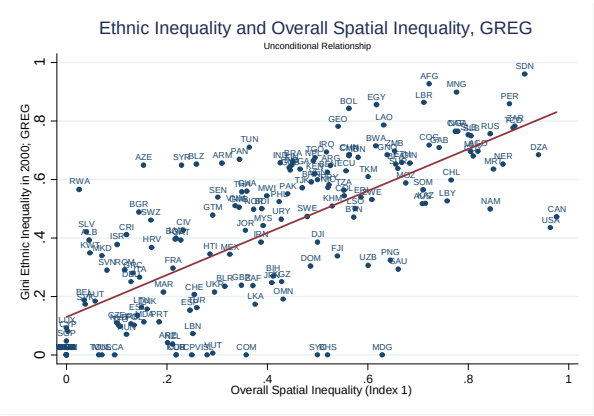

Figure 4*a*

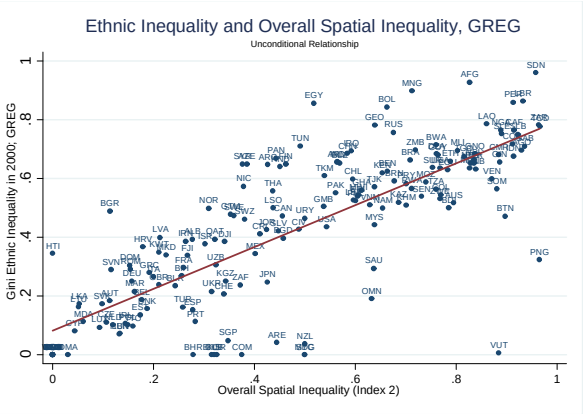

Figure 4*b*

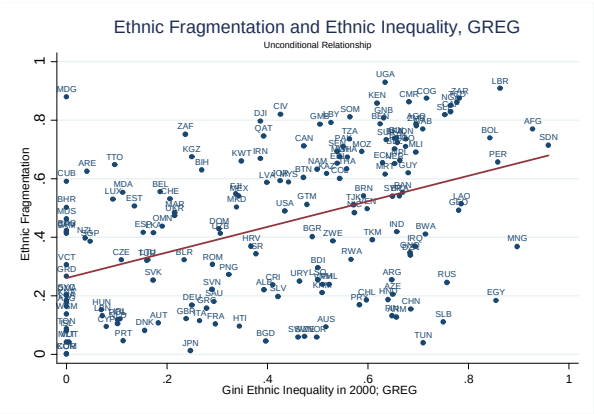

Figure 5*a*

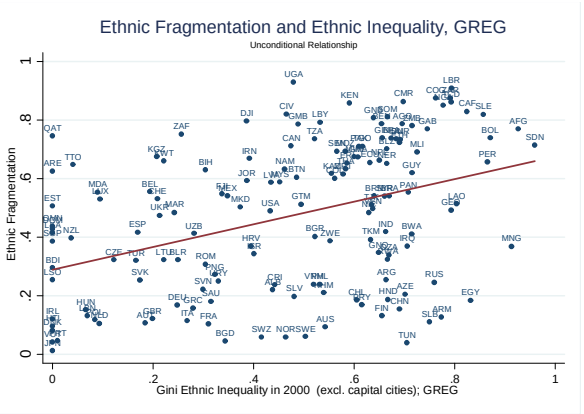

Figure 5*b*

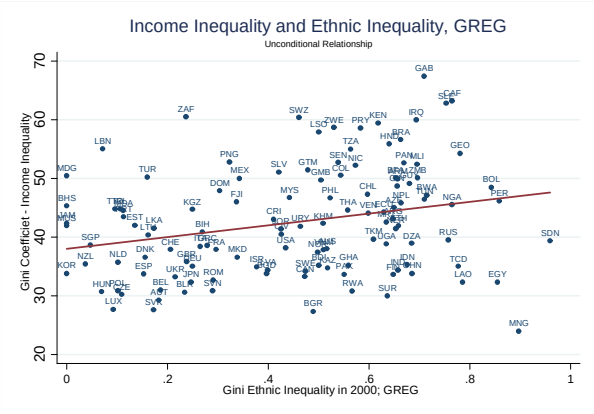

Figure 6*a*

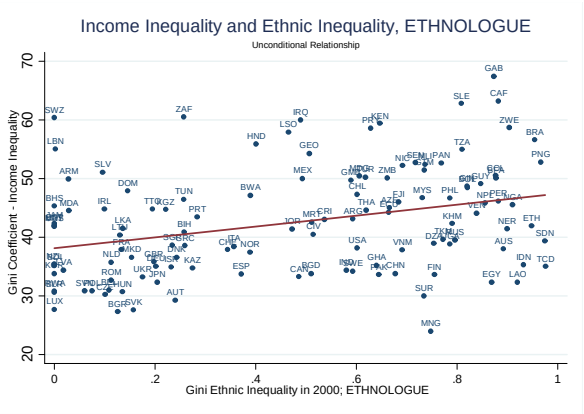

Figure 6*b*

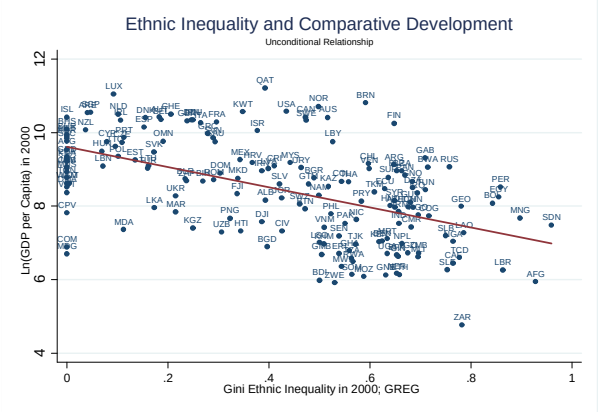

Figure 7*a*

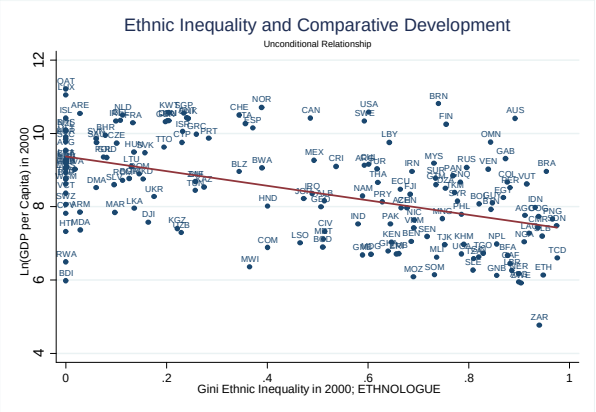

Figure 7*b*

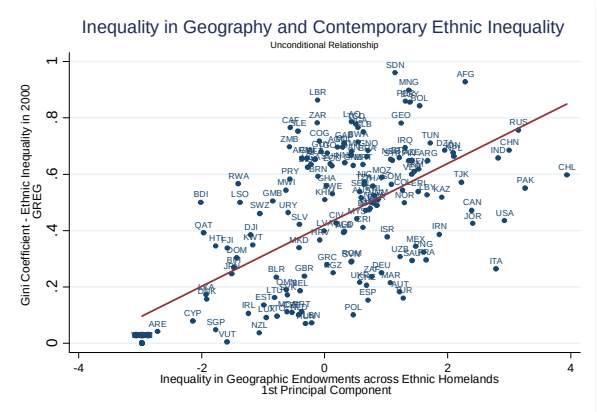

Figure 8*a*

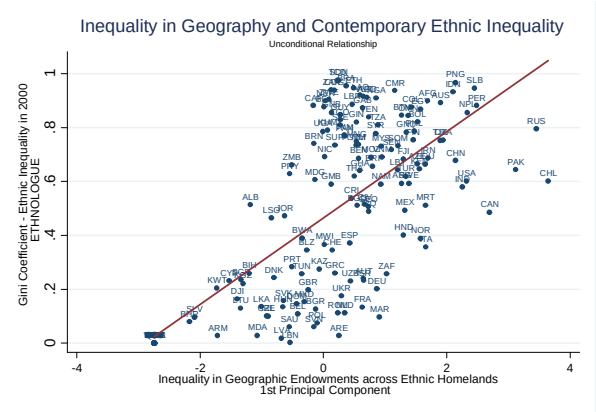

Figure 8*b*

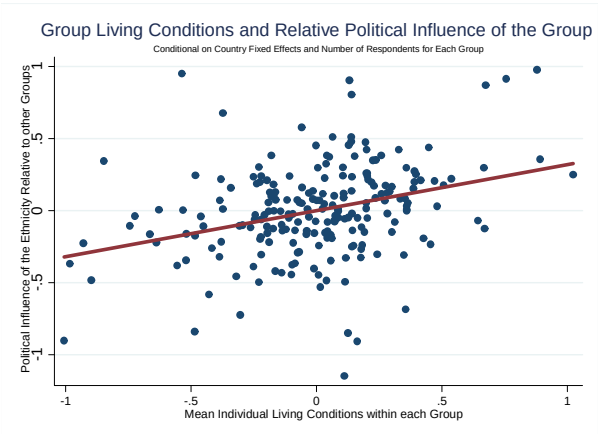

Figure 9a

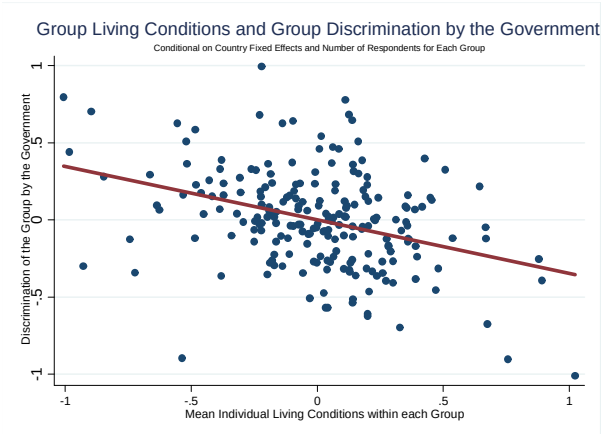

Figure 9b

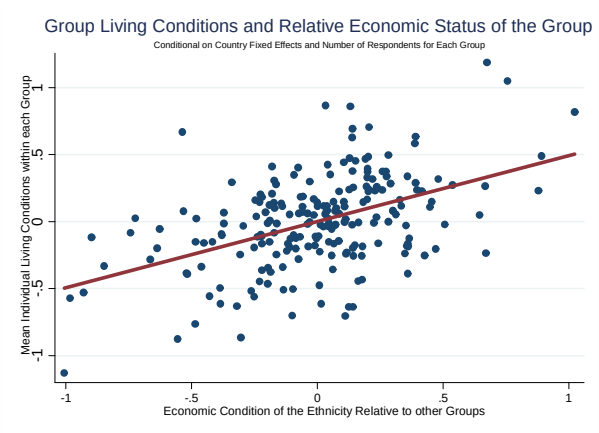

Figure 10a

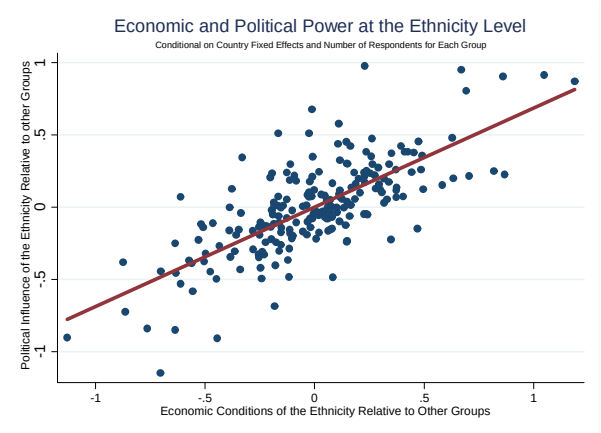

Figure 10b

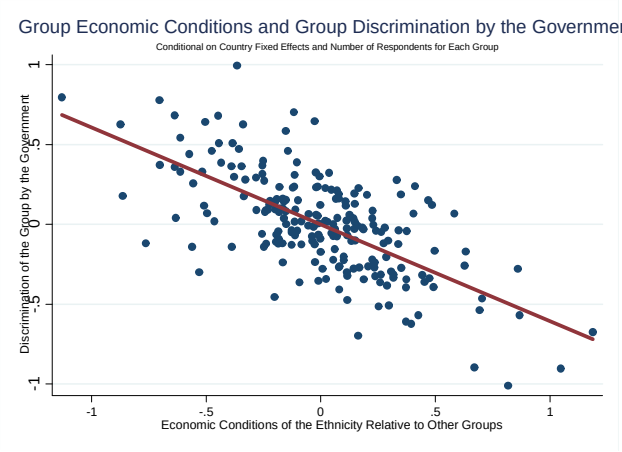

Figure10c

Table 1A: Summary Statistics - Cross Country Inequality Measures

	Obs.	mean	st. dev.	p25	median	p75	min	max
Number of Ethnicities (GREG)	173	11.52	14.17	1.00	3.00	8.00	13.00	94.00
Ethnic Gini in 2009 (GREG), All Groups	173	0.42	0.26	0.00	0.19	0.47	0.63	0.96
Ethnic Gini in 2000 (GREG), All Groups	173	0.42	0.26	0.00	0.19	0.47	0.65	0.96
Ethnic Gini in 1992 (GREG), All Groups	173	0.48	0.29	0.00	0.21	0.56	0.72	0.97
Number of Languages (ETHNOLOGUE)	173	41.91	99.78	1.00	3.00	9.00	36.00	791.00
Ethnic Gini in 2009 (ETHNOLOGUE), All Groups	173	0.45	0.33	0.00	0.13	0.47	0.77	0.97
Ethnic Gini in 2000 (ETHNOLOGUE), All Groups	173	0.46	0.34	0.00	0.11	0.51	0.77	0.98
Ethnic Gini in 1992 (ETHNOLOGUE), All Groups	173	0.50	0.35	0.00	0.15	0.55	0.83	0.99
Number of Pixels	173	24.28	63.84	1.00	4.00	8.00	22.00	637.00
Spatial Gini in 2009, Pixels	173	0.40	0.26	0.00	0.17	0.40	0.60	0.98
Spatial Gini in 2000, Pixels	173	0.40	0.26	0.00	0.17	0.40	0.60	0.98
Spatial Gini in 1992, Pixels	173	0.45	0.27	0.00	0.21	0.47	0.68	0.95
Number of Thiessen Polygons	173	50.79	98.05	1.00	7.00	17.00	54.00	698.00
Spatial Gini in 2009, Thiessen Polygons	173	0.47	0.29	0.00	0.21	0.48	0.71	0.97
Spatial Gini in 2000, Thiessen Polygons	173	0.48	0.29	0.00	0.23	0.46	0.73	0.97
Spatial Gini in 1992, Thiessen Polygons	173	0.52	0.31	0.00	0.23	0.52	0.81	0.99

Table 1B: Summary Statistics - Afrobarometer Sample - District Level

	Obs.	mean	st. dev.	p25	median	p75	min	max
Theil Index - Overall Inequality	1298	0.05	0.05	0.00	0.04	0.08	0.00	0.31
Theil Index - Between-Group Inequality	1298	0.04	0.04	0.00	0.03	0.07	0.00	0.21
Theil Index - Within-Group Inequality	1298	0.01	0.02	0.00	0.00	0.02	0.00	0.31
Living Conditions Index	1298	2.71	0.67	2.25	2.75	3.17	1.00	4.88
Education	1298	4.01	1.42	3.00	4.00	5.00	1.00	8.75
Access to Sewage System	1277	0.24	0.39	0.00	0.00	0.50	0.00	1.00
Access to Clean Piped Water	1283	0.47	0.45	0.00	0.44	1.00	0.00	1.00
Access to Electricity Grid	1289	0.54	0.45	0.00	0.60	1.00	0.00	1.00
Ethnic Fractionalization	1298	0.71	0.27	0.50	0.76	1.00	0.11	1.00
Number of Ethnic Groups per District	1298	2.74	2.24	1.00	2.00	3.00	1.00	18.00
Number of Respondents per District	1298	16.29	23.16	7.00	8.00	16.00	1.00	357.00
Support for Democracy	1284	1.57	0.39	1.38	1.67	1.86	0.00	2.00
Satisfaction with Democracy	1286	2.60	0.67	2.13	2.67	3.04	0.00	4.00
Constitution Expresses People's Hopes-Values	1290	3.52	0.61	3.15	3.59	3.98	1.00	5.00

Table 1C: Summary Statistics - Afrobarometer Sample - Individual Level

	Obs.	mean	st. dev.	p25	median	p75	min	max
Theil Index - Overall Inequality	21138	0.07	0.05	0.03	0.07	0.11	0.00	0.31
Theil Index - Between-Group Inequality	21138	0.06	0.05	0.02	0.06	0.09	0.00	0.21
Theil Index - Within-Group Inequality	21138	0.01	0.02	0.00	0.00	0.02	0.00	0.31
Living Conditions Index	21138	2.63	1.20	2.00	3.00	4.00	1.00	5.00
Education	21138	4.09	2.02	3.00	4.00	5.00	1.00	10.00
Access to Sewage System	20487	0.24	0.43	0.00	0.00	0.00	0.00	1.00
Access to Clean Piped Water	20818	0.51	0.50	0.00	1.00	1.00	0.00	1.00
Access to Electricity Grid	20878	0.55	0.50	0.00	1.00	1.00	0.00	1.00
Ethnic Fractionalization	21138	0.60	0.35	0.25	0.69	0.94	0.00	1.00
Number of Ethnic Groups per District	21138	4.85	3.93	2.00	4.00	7.00	1.00	18.00
Number of Respondents per District	21138	49.19	63.35	12.00	24.00	63.00	1.00	357.00
Support for Democracy	18199	1.58	0.73	1.00	2.00	2.00	0.00	2.00
Satisfaction with Democracy	18024	2.56	1.06	2.00	3.00	3.00	0.00	4.00
Constitution Expresses People's Hopes-Values	17733	3.51	1.18	3.00	4.00	4.00	1.00	5.00

Panel A reports summary statistics for the main ethnic inequality and overall spatial inequality measures employed in the cross-country analysis. Section 3 gives details on the construction of these measures.
Panel B reports summary statistics for all measures, employed in the cross-district analysis within African countries (Afrobarometer sample). Panel C reports summary statistics for all measures, employed in the cross-individual analysis within African countries (Afrobarometer sample). The Data Appendix gives detailed variable definitions and sources.

Table 2a - Baseline Estimates: Ethnic Inequality and Economic Development (in 2000), Atlas Naradov Mira

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Ethnic Inequality	-1.4707***		-1.4003***		-1.4985***		-1.1868***	-1.1099**	-1.2446**	-1.3447***	-1.4580***
[Gini Coeff., GREG]	(0.2504)		(0.3633)		(0.4018)		(0.4069)	(0.5045)	(0.4835)	(0.5148)	(0.4905)
	-5.87		*-3.85*		*-3.73*		*-2.92*	*-2.20*	*-2.57*	*-2.61*	*-2.97*
Spatial Inequality 1		-1.1508***	-0.1112					-0.1186		-0.0592	
[Gini Coeff., Pixels]		(0.2786)	(0.3780)					(0.3809)		(0.3820)	
		-4.13	*-0.29*					*-0.31*		*-0.15*	
Spatial Inequality 2				-1.1612***	0.0373				0.0880		0.1060
[Gini Coeff., Thiessen Polyg]				(0.2559)	(0.3961)				(0.4084)		(0.4155)
				-4.54	*0.09*				*0.22*		*0.26*
Log Number of Languages						-0.3037***	-0.0887	-0.0893	-0.0912	-0.1675	-0.1723
[GREG]						(0.0605)	*(0.0930)*	*(0.0936)*	*(0.0949)*	(0.1286)	(0.1304)
						-5.02	*-0.95*	*-0.95*	*-0.96*	*-1.30*	*-1.32*
Ethnic Inequality in Population										0.5164	0.5305
[Gini Coeff., GREG]										*(0.4472)*	*(0.4465)*
										1.15	*1.19*
adjusted R-squared	0.661	0.626	0.659	0.631	0.659	0.643	0.661	0.659	0.659	0.659	0.659
observations	173	173	173	173	173	173	173	173	173	173	173
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 2b - Baseline Estimates: Ethnic Inequality and Economic Development (in 2000), Ethnologue

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Ethnic Inequality	-1.1281***		-1.0245***		-1.0839***		-1.2806***	-1.1734**	-1.2309***	-1.0657***	-1.2554***
[Gini Coeff., ETHNO]	(0.2267)		(0.2975)		(0.3817)		(0.3694)	(0.4625)	(0.4528)	(0.3747)	(0.4300)
	-4.98		*-3.44*		*-2.84*		*-3.47*	*-2.54*	*-2.72*	*-2.84*	*-2.92*
Spatial Inequality 1		-1.1508***	-0.2035					-0.1857		-0.1531	
[Gini Coeff., Pixels]		(0.2786)	(0.3524)					(0.3617)		(0.3508)	
		-4.13	*-0.58*					*-0.51*		*-0.44*	
Spatial Inequality 2				-1.1612***	-0.0732				-0.0884		0.1967
[Gini Coeff., Thiessen Polyg]				(0.2559)	(0.4343)				(0.4354)		(0.4539)
				-4.54	*-0.17*				*-0.20*		*0.43*
Log Number of Languages						-0.1730***	0.0389	0.0357	0.0399		
[ETHNO]						(0.0467)	*(0.0741)*	*(0.0759)*	*(0.0751)*		
						-3.70	*0.53*	*0.47*	*0.53*		
Ethnic Inequality in Population										0.6678*	0.7013*
[Gini Coeff., ETHNO]										*(0.3727)*	*(0.3742)*
										1.79	*1.87*
adjusted R-squared	0.654	0.626	0.653	0.631	0.652	0.625	0.653	0.651	0.651	0.66	0.661
observations	173	173	173	173	173	173	173	173	173	173	173
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

The table reports cross-country OLS estimates. The dependent variable is the log of real GDP per capita in 2000. The ethnic Gini coefficients reflect inequality in lights per capita across ethnic homelands. In Table 3A we use the digitized version of the Atlas Narodov Mira (GREG) to aggregate lights per capita across ethnic homelands. In Table 3B we use the digitized version of the Ethnologue database to aggregate lights per capita across linguistic homelands.

The spatial Gini coefficient 1 captures the degree of spatial inequality across 2.5 by 2.5 decimal degree boxes/pixels in each country (boxes/pixels intersected by national boundaries are of smaller size). The spatial Gini coefficient 2 captures the degree of spatial inequality across Thiessen polygons in each country. Thiessen polygons have the unique property that each polygon contains only one input point, and any location within a polygon is closer to its associated point than to the point of any other polygon. The input points are the centroids of the linguistic homelands according to the Ethnologue dataset. To construc the spatial inequality 2 index we intersect the 7,570 Thiessen polygons with the country boundaries of 2000 and compute the spatial Gini across the resulting polygons within each country.

All specifications include continental fixed effects (constants not reported). The Data Appendix gives detailed variable definitions and data sources. Heteroskedasticity-adjusted standard errors are reported in parentheses below the estimates. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Table 3 - Sensitivity Checks A: Ethnic Inequality and Economic Development (in 2000)
Conditioning on Ethno-linguistic Fragmentation and Polarization

	Atlas Narodov Mira (GREG)				Ethnologue			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ethnic Inequality [Gini Coeff.]	-1.4977***	-1.6771***	-1.5567***	-1.5782***	-1.1099***	-0.8248**	-1.1186***	-1.1438***
	(0.4011)	(0.3955)	(0.4075)	(0.3924)	(0.3716)	(0.3417)	(0.3744)	(0.3756)
	-3.73	*-4.24*	*-3.82*	*-4.02*	*-2.99*	*-2.41*	*-2.99*	*-3.04*
Ethnic Fragmentation	-0.0097				0.1256			
	(0.3590)				(0.3431)			
	-0.03				*0.37*			
Cultural Fragmentation		-0.3417				0.0339		
		(0.3520)				*(0.3469)*		
		-0.97				*0.10*		
Ethno-linguistic Polarization 1			0.464				*0.5817*	
			(0.9745)				*(0.9820)*	
			0.48				*0.59*	
Ethno-linguistic Polarization 2				1.9997				2.1173*
				(1.2688)				(1.1802)
				1.58				*1.79*
Spatial Inequality 2 [Gini Coeff.]	0.0397	0.0739	0.0641	0.0563	-0.0862	-0.468	-0.0577	-0.0568
	(0.4163)	(0.4827)	(0.3928)	(0.3845)	(0.4408)	(0.4760)	(0.4230)	(0.4224)
	0.10	*0.15*	*0.16*	*0.15*	*-0.20*	*-0.98*	*-0.14*	*-0.13*
adjusted R-squared	0.657	0.694	0.653	0.658	0.65	0.675	0.646	0.651
observations	173	150	172	172	173	150	172	172
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

The table reports cross-country OLS estimates. The dependent variable is the log of real GDP per capita in 2000. In columns (1) and (5) we control for ethnic fragmentation using an index that reflects the likelihood that two randomly chosen individuals in one country will be members of the same group (from Alesina et al., 2003). In columns (2) and (6) we control for cultural (linguistic) fragmentation using an index (from Fearon, 2003) that accounts for linguistic distances among ethnic groups. In columns (3) and (7) we control for ethnic polarization, using the Montalvo and Reynal-Querol (2005) index. In columns (4) and (8) we control for ethnic polarization using a polarization index that accounts for linguistic distances among ethnic groups (from Duclos, Esteban, and Rey (2004) and Esteban and Rey (2011, 2012)).
In all specification we control for the overall degree of spatial inequality in a country using the Gini coefficient of lights per capita based on Thiessen polygons that use as input points the centroids of the linguistic homelands according to the Ethnologue dataset. All specifications include continental fixed effects (constants not reported). The Data Appendix gives detailed variable definitions and data sources. Heteroskedasticity-adjusted standard errors are reported in parentheses below the estimates. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Table 4 - Sensitivity Checks B: Ethnic Inequality and Economic Development (in 2000)
Additional Controls and Alternative Measures of Ethnic Inequality

	Atlas Narodov Mira (GREG)						Ethnologue					
	All Ethnic Areas		Excl. Capitals		Excl. Small Groups		All Ethnic Areas		Excl. Capitals		Excl. Small Groups	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Ethnic Inequality	-1.5123***	-1.3073***	-1.1601***	-1.0370***	-2.0069***	-1.5247***	-1.0650***	-0.8779***	-1.1723***	-0.8062**	-1.3370***	-1.2490***
[Gini Coeff.]	(0.4357)	(0.3815)	(0.4023)	(0.3309)	(0.5701)	(0.5501)	(0.3480)	(0.2997)	(0.3747)	(0.3552)	(0.4800)	(0.4535)
	-3.47	*-3.43*	*-2.88*	*-3.13*	*-3.52*	*-2.77*	*-3.06*	*-2.93*	*-3.13*	*-2.27*	*-2.79*	*-2.75*
Spatial Inequality 2	0.0743	0.2679	-0.7642	-0.275	-0.1669	-0.0063	-0.2035	0.0377	-0.4793	-0.2985	-0.2473	0.1042
[Gini Coeff.]	(0.4710)	(0.4043)	(0.5395)	(0.4501)	(0.4365)	(0.3968)	(0.4908)	(0.4370)	(0.4822)	(0.4757)	(0.5105)	(0.4425)
	0.16	*0.66*	*-1.42*	-0.61	*-0.38*	*-0.02*	*-0.41*	*0.09*	*-0.99*	*-0.63*	*-0.48*	*0.24*
Ethnic	0.0943	0.1826	0.0709	0.2411	0.5099	0.5054	0.1308	0.2437	-0.1721	0.1237	0.1909	0.3108
Fragmentation	(0.3716)	(0.3162)	(0.3904)	(0.3437)	(0.4040)	(0.3667)	(0.3614)	(0.3134)	(0.3958)	(0.3544)	(0.3578)	(0.3151)
	0.25	*0.58*	*0.18*	0.7	*1.26*	*1.38*	*0.36*	*0.78*	*-0.43*	*0.35*	*0.53*	*0.99*
adjusted R-squared	0.663	0.723	0.676	0.741	0.673	0.723	0.654	0.715	0.672	0.723	0.654	0.7173
Observations	173	173	152	152	173	173	173	173	147	147	173	173
Region Fixed Effec	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Simple	Rich	Simple	Rich	Simple	Rich	Simple	Rich	Simple	Rich	Simple	Rich

The table reports cross-country OLS estimates. The dependent variable is the log of real GDP per capita in 2000. In columns (1)-(6) we use the digitized version of the Atlas Narodov Mira (GREG) to aggregate lights per capita across ethnic homelands and construct the ethnic inequality measures. In columns (7)-(12) we use the digitized version of the Ethnologue database to aggregate lights per capita across linguistic homelands and construct the ethnic inequality measures. For the construction of the ethnic inequality measures (Gini coefficients) in columns (3), (4), (9), and (10) we exclude ethnic areas where capital cities fall. For the construction of the ethnic inequality measures (Gini coefficients) in columns (5), (6), (11), and (12) we exclude small ethnic groups consisting of less than one percept of country's population.
Odd-numbered columns include as controls absolute latitude, log land area, and log population in 2000 (simple set of controls). Even-numbered columns include as controls absolute latitude, log land area, log population in 2000, an index of terrain ruggedness, the percentage of each country with fertile soil, the percentage of each country with tropical climate, average distance to nearest ice-free coast, and an index of gem-quality diamond extraction (rich set of controls). In all specifications we control for ethnic fragmentation using an index that reflects the likelihood that two randomly chosen individuals in one country will be members of the same group (from Alesina et al., 2003).
In all specification we control for the overall degree of spatial inequality in a country using the Gini coefficient of lights per capita based on Thiessen polygons that use as input points the centroids of the linguistic homelands according to the Ethnologue dataset. All specifications include continental fixed effects (constants not reported). The Data Appendix gives detailed variable definitions and data sources. Heteroskedasticity-adjusted standard errors are reported in parentheses below the estimates. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Table 5. The Origins of Contemporary Ethnic Inequality
Inequality in Geographic Endowments across Ethnic Homelands and Contemporary Ethnic Inequality in Development across Ethnic

	Atlas Narodov Mira (GREG)						Ethnologue					
	All Ethnic Areas			Excl. Capitals			All Ethnic Areas			Excl. Capitals		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Gini Coef. - Sea Distance	0.0489	0.1379	0.1857	0.0552	0.1090	0.1961	0.0355	-0.1886	-0.1756	0.0897	0.0276	0.0231
	(0.0956)	(0.1231)	(0.1156)	(0.1274)	(0.1611)	(0.1494)	(0.0969)	(0.1567)	(0.1622)	(0.1143)	(0.1639)	(0.1481)
Gini Coef. - Elevation	0.4343	0.8727	0.6973	0.0953	0.1046	0.0754	0.6292	3.0353**	1.7750	0.0027**	0.0038***	0.0019
	(0.3304)	(0.6673)	(0.6334)	(0.1206)	(0.1262)	(0.1178)	(0.4782)	(1.2705)	-1.2926	(0.0010)	(0.0011)	(0.0012)
Gini Coeff. -Land Quality	0.2528***	-0.1028	-0.121	0.3067***	0.1227	0.1227	0.3680***	0.6496***	0.6000***	0.3338***	0.4759**	0.4116**
	(0.0910)	(0.1775)	(0.1813)	(0.0935)	(0.1816)	(0.1872)	(0.0978)	(0.2038)	(0.1654)	(0.0922)	(0.1873)	(0.1658)
Gini Coeff. - Water Area	0.5632***	0.4985***	0.4744***	0.4476***	0.4011***	0.3914***	0.6698***	0.5550***	0.5862***	0.5238***	0.4206***	0.4330***
	(0.0550)	(0.0773)	(0.0733)	(0.0970)	(0.1275)	(0.1280)	(0.0621)	(0.0857)	(0.0875)	(0.0786)	(0.0997)	(0.0934)
adjusted R-squared	0.674	0.6674	0.6408	0.5967	0.5988	0.6078	0.7511	0.7537	0.7595	0.7169	0.714	0.7398
Observations	173	169	162	151	150	150	168	166	160	144	143	142
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Spatial	Spatial & Level	No	Spatial	Spatial & Level	No	Spatial	Spatial & Level	No	Spatial	Spatial & Level

The table reports cross-country OLS estimates, associating contemporary ethnic inequality with inequality in geographic endowments across ethnic homelands. The dependent variable is the ethnic Gini coefficient that reflects inequality in lights per capita across ethnic-linguistic homelands, using the digitized version of Atlas Narodov Mira (GREG) in (1)-(6) and Ethnologue in (7)-(12). To construct the inequality measures in geographic endowments we first estimate the distance of the centroid of each ethnic homeland to the closest sea coast, average elevation, average soil quality, and the area of each homeland covered by water (lakes, rivers, and other streams) and then construct Gini coefficients capturing inequality in each of these geographic features for each country.
In columns (1)-(3) and (7)-(9) we use all ethnic-linguistic homelands; in columns (4)-(6) and (10)-(12) we exclude ethnic-linguistic regions where capital cities fall. In columns (2), (5), (8), and (11) we control for the overall degree of spatial inequality in geographic endowments using the Gini coefficient of each of these features (distance to the closest sea coast, elevation, soil quality, water area) based on Thiessen polygons that use as input points the centroids of the linguistic homelands according to the Ethnologue dataset. In columns (3), (6), (9), and (12) we also control for the mean value of distance to closest sea coast, elevation, soil quality, and area under water. All specifications include continental fixed effects (constants not reported). The Data Appendix gives detailed variable definitions and data sources. Heteroskedasticity-adjusted standard errors are reported in parentheses below the estimates. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Table 6 - Principal Component Analysis

	Eigenvalue	Variance Explained	Variable	Factor Loadings 1st PC	2nd PC	3rd PC	4th PC
Panel A: Gini Coefficient GREG - All Groups (173 countries)							
1st Principal Component	2.474	0.618	Gini Sea Distance	0.482	-0.666	0.080	0.563
2nd Principal Component	0.708	0.177	Gini Elevation	0.491	0.444	-0.720	0.208
3rd Principal Component	0.481	0.120	Gini Land Quality	0.479	0.531	0.689	0.120
4th Principal Component	0.337	0.084	Gini Water Area	0.545	-0.277	-0.027	-0.791
Panel B: Gini Coefficient GREG - Excluding Capitals (151 countries)							
1st Principal Component	2.420	0.567	Gini Sea Distance	0.537	-0.413	0.153	0.720
2nd Principal Component	0.753	0.200	Gini Elevation	0.540	-0.259	0.468	-0.650
3rd Principal Component	0.515	0.128	Gini Land Quality	0.402	0.873	0.230	0.152
4th Principal Component	0.312	0.105	Gini Water Area	0.508	0.020	-0.840	-0.190
Panel C: Gini Coefficient ETHNOLOGUE - All Groups (168 countries)							
1st Principal Component	2.389	0.597	Gini Sea Distance	0.498	-0.633	0.186	0.563
2nd Principal Component	0.739	0.185	Gini Elevation	0.478	0.355	-0.770	0.229
3rd Principal Component	0.583	0.146	Gini Land Quality	0.445	0.643	0.610	0.128
4th Principal Component	0.289	0.072	Gini Water Area	0.571	-0.246	0.008	-0.783
Panel D: Gini Coefficient ETHNOLOGUE - Excluding Capitals (144 countries)							
1st Principal Component	1.804	0.451	Gini Sea Distance	0.594	0.188	-0.468	0.627
2nd Principal Component	1.052	0.263	Gini Elevation	0.186	0.865	0.456	-0.095
3rd Principal Component	0.703	0.176	Gini Land Quality	0.466	-0.455	0.722	0.234
4th Principal Component	0.441	0.110	Gini Water Area	0.629	-0.096	-0.228	-0.737
Panel F: Gini Coefficient - Overall Spatial Inequality Index 1 - Pixel 2.5 x 2.5 degrees (164 countries)							
1st Principal Component	2.132	0.533	Gini Sea Distance	0.502	-0.559	0.131	0.647
2nd Principal Component	0.814	0.203	Gini Elevation	0.497	0.404	-0.759	0.117
3rd Principal Component	0.580	0.145	Gini Land Quality	0.462	0.616	0.637	0.045
4th Principal Component	0.474	0.118	Gini Water Area	0.536	-0.381	0.032	-0.752
Panel E: Gini Coefficient - Overall Spatial Inequality Index 2 - Thiessen Polygons (169 countries)							
1st Principal Component	2.082	0.521	Gini Sea Distance	0.480	-0.608	0.377	0.507
2nd Principal Component	0.849	0.212	Gini Elevation	0.515	0.347	-0.662	0.420
3rd Principal Component	0.598	0.150	Gini Land Quality	0.451	0.639	0.612	-0.116
4th Principal Component	0.471	0.118	Gini Water Area	0.549	-0.319	-0.212	-0.744

The table reports the results of the principal component analysis that is based on four measures (Gini coefficients) reflecting inequality in geographic endowments in distance to the coast, elevation, land suitability for agriculture, and area under water across ethnic homelands (Panels A, B, C, and D), pixels of 2.5 x 2.5 decimal degrees (in Panel E) and Thiessen polygons that use as input points the centroids of the linguistic homelands according to the Ethnologue dataset (Panel F). Column (1) reports the eigenvalue of each principal component and column (2) gives the percentage of the total variance explained by each principal component. The other columns give the factor loadings in the four principal components of the Gini coefficient reflecting inequality in distance to the coast, elevation, land suitability for agriculture, and area under water.

Table 7: The Origins of Contemporary Ethnic Inequality
Inequality in Geographic Endowments across Ethnic Homelands and Contemporary Ethnic Inequality

	Atlas Narodov Mira (GREG)			Ethnologue		
	(1)	(2)	(3)	(4)	(5)	(6)
Inequality in Geographic Endowments across Ethnic Homelands (PC)	0.1227***	0.1089***	0.0973***	0.1613***	0.1523***	0.1588***
	(0.0088)	(0.0178)	(0.0171)	(0.0095)	(0.0196)	(0.0199)
	14.00	*6.13*	*5.70*	*16.99*	*7.76*	*7.98*
Spatial Inequality in Geographic Endowments (PC)		0.0109	0.0068		0.0075	-0.006
		(0.0182)	(0.0194)		(0.0202)	(0.0228)
		0.6	*0.35*		*0.37*	*-0.26*
Adjusted R-squared	0.623	0.616	0.605	0.695	0.689	0.688
Observations	173	169	162	168	166	160
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	No	No	Geography	No	No	Geography

The table reports cross-country OLS estimates, associating contemporary ethnic inequality with inequality in geographic endowments across ethnic homelands. The dependent variable is the ethnic Gini coefficient that reflects inequality in lights per capita across ethnic-linguistic homelands in 2000, using the digitized version of Atlas Narodov Mira (GREG) (in columns (1)-(3)) and Ethnologue (in columns (4)-(6)).
The main independent variable is a composite index capturing inequality in geographic endowments across ethnic homelands. The index is the first principal component of inequality across ethnic-linguistic homelands in distance to the coast, elevation, land suitability for agriculture, and area under water. In columns (2), (3), (5), and (6) we control for the overall degree of spatial inequality in geographic endowments using a composite index that aggregates (via principal components) Gini coefficients on distance to the coast, elevation, land suitability for agriculture, water area across Thiessen polygons that use as input points the centroids of the linguistic homelands according to the Ethnologue dataset. In columns (3) and (6) we also control for the mean value of distance to the coast, elevation, land suitability for agriculture, and area under water.
All specifications include continental fixed effects (constants not reported). The Data Appendix gives detailed variable definitions and data sources. Heteroskedasticity-adjusted standard errors are reported in parentheses below the estimates. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

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

	Atlas Narodov Mira (GREG)			Ethnologue		
	(1)	(2)	(3)	(4)	(5)	(6)
Inequality in Geographic Endowments across Ethnic Homelands (PC)	-0.1789***	-0.2311***	-0.1770**	-0.1611***	-0.058	-0.1526*
	(0.0405)	(0.0851)	(0.0891)	(0.0459)	(0.0956)	(0.0860)
	-4.42	*-2.71*	*-1.99*	*-3.51*	*-0.61*	*-1.78*
Spatial Inequality in Geographic Endowments (PC)		0.0755	0.048		-0.0898	0.0268
		(0.0970)	(0.1113)		(0.0986)	(0.1093)
		0.78	*0.43*		*-0.91*	*0.24*
adjusted R-squared	0.629	0.646	0.668	0.623	0.639	0.673
Observations	173	169	162	168	166	160
Region Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	No	No	Geography	No	No	Geography

The table reports cross-country OLS estimates, associating contemporary economic development with inequality in geographic endowments across ethnic homelands. The dependent variable is the log of real GDP per capita in 2000.
The main independent variable is a composite index capturing inequality in geographic endowments across ethnic homelands, using the digitized version of Atlas Narodov Mira (GREG) (in columns (1)-(3)) and Ethnologue (in columns (4)-(6)). The index is the first principal component of inequality across ethnic-linguistic homelands in distance to the coast, elevation, land suitability for agriculture, and area under water. In columns (2), (3), (5), and (6) we control for the overall degree of spatial inequality in geographic endowments using a composite index that aggregates (via principal components) Gini coefficients on distance to the coast, elevation, land suitability for agriculture, water area across Thiessen polygons that use as input points the centroids of the linguistic homelands according to the Ethnologue dataset. In columns (3) and (6) we also control for the mean value of distance to the coast, elevation, land suitability for agriculture, and area under water.
All specifications include continental fixed effects (constants not reported). The Data Appendix gives detailed variable definitions and data sources. Heteroskedasticity-adjusted standard errors are reported in parentheses below the estimates. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Table 9 - Ethnic Inequality and Regional Development within African Countries
District-level and Individual-Level Analysis Using Data from the Afrobarometer Surveys

	District-Level		Individual-Level			
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Living Conditions			
Between Group Ineq.	-5.4994***	-6.2017***	-5.9472***	-5.8051***	-4.7228***	-5.1732***
[Theil Index]	(0.8809)	(1.0308)	(1.0874)	(0.8644)	(0.8488)	(0.8604)
Within Group Ineq.	-2.6220**	-5.7490***	-2.3695***	-5.4641***	-1.4248***	-4.5081***
[Theil Index]	(0.9085)	(1.0771)	(0.4606)	(0.5605)	(0.3351)	(0.5255)
adjusted R-squared	0.356	0.500	0.154	0.166	0.188	0.188
Observations	1298	842	20617	16580	20617	16580
			Panel B: Education			
Between Group Ineq.	-4.4948***	-4.7501***	-3.2780***	-2.7169**	-3.4919***	-3.0329***
[Theil Index]	(0.9452)	(1.2602)	(1.0931)	(1.1152)	(1.0357)	(1.0617)
Within Group Ineq.	-3.5337***	-3.9965***	-1.5350***	-1.8955***	-1.3886***	-2.0192***
[Theil Index]	(0.5205)	(0.8417)	(0.5110)	(0.6833)	(0.5003)	(0.6909)
adjusted R-squared	0.506	0.547	0.484	0.483	0.498	0.495
Observations	1298	842	20617	16580	20617	16580
Country Fixed Effects	Yes	Yes	Yes	Yes	No	No
Country-Ethnicity Fixed Effects	No	No	No	No	Yes	Yes
District-level Controls	Yes	Yes	Yes	Yes	Yes	Yes
Individual-level Controls	No	No	Yes	Yes	Yes	Yes
Sample; Districts	All	>1 group	All	>1 group	All	>1 group

The table reports OLS estimates, associating two proxies of well-being (living conditions in Panel A and education in Panel B) with inequality between and within ethnic groups at the district level, as reflected in the Theil index. The dependent variable in Panel A is a 1-5 living conditions index; the dependent variable in Panel B is a 1-10 education index. Columns (1)-(2) report district-level estimates, while columns (3)-(8) report individual level estimates. The between-ethnic-group and the within-ethnic-group Theil indicators are based on individuals' responses on living conditions. The district-level conditioning set includes the log number of ethnic groups in each district, the log number of respondents in each district, and district-level ethnic fractionalization. The individual-level conditioning set includes age, age squared, a gender indicator variable, 22 religion fixed effects and 25 occupation fixed effects. Odd-numbered columns report estimates in the full sample. In even-numbered columns we exclude from the estimation districts with respondents from just one ethnic group. All specifications include country fixed effects (constants not reported). Specifications (5) and (6) include ethnicity-country fixed effects (constants not reported).

The Data Appendix gives detailed variable definitions and data sources. All variables are constructed using data from the 3rd round of the Afrobarometer Surveys. Clustered at the district level standard errors are reported in parentheses below the estimates. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Table 10a - Ethnic Inequality and Public Goods Provision within African Countries

	Piped Water			Sewage System			Electrification		
	District	Individual Level		District	Individual Level		District	Individual Level	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Between Group Ineq.	-2.5917**	-1.9828***	-1.8395***	-1.9295***	-1.3094**	-1.3981***	-1.7328**	-1.1828*	-1.2205**
[Theil Index]	(0.9664)	(0.6336)	(0.5793)	(0.4999)	(0.5217)	(0.5135)	(0.7864)	(0.6140)	(0.5995)
Within Group Ineq.	-0.9725**	-0.4500	-0.4652	-0.4209	0.058	0.1477	-1.2173***	-0.2566	-0.4046
[Theil Index]	(0.3637)	(0.3774)	(0.3563)	(0.5185)	(0.3194)	(0.3325)	(0.2345)	(0.3421)	(0.3107)
adjusted R-squared	0.238	0.392	0.441	0.219	0.399	0.421	0.307	0.44	0.478
Observations	837	16378	16378	833	16067	16067	839	16429	16429
Country Fixed Effects	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No
Country-Ethnicity Fixed Effects	No	No	Yes	No	No	Yes	No	No	Yes
District Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample, Districts	>1 groups	>1 groups	>1 groups	>1 groups	>1 groups	>1 groups	>1 groups	>1 groups	>1 groups

Table 10b - Ethnic Inequality and Public Goods Provision within African Countries

	Piped Water			Sewage System			Electrification		
	District	Individual Level		District	Individual Level		District	Individual Level	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Between Group Ineq.	-2.1828*	-1.8972***	-1.8137***	-1.7909***	-1.3979***	-1.4289***	-1.3118	-1.3686**	-1.2675**
[Theil Index]	(1.0963)	(0.6526)	(0.5900)	(0.5177)	(0.5370)	(0.5290)	(0.8746)	(0.6350)	(0.6173)
Within Group Ineq.	-0.5935*	-0.3748	-0.4443	-0.2897	-0.0195	0.1229	-0.8267***	-0.4204	-0.4431
[Theil Index]	(0.3305)	(0.4178)	(0.3875)	(0.5116)	(0.3330)	(0.3385)	(0.1815)	(0.3633)	(0.3399)
Mean Living Conditions at the District	0.0659	0.0137	0.0043	0.0228	-0.0143	-0.0052	0.0678*	-0.0297	-0.0079
	(0.0451)	(0.0307)	(0.0288)	(0.0243)	(0.0229)	(0.0240)	(0.0332)	(0.0302)	(0.0269)
adjusted R-squared	0.238	0.392	0.441	0.219	0.399	0.421	0.307	0.44	0.478
Observations	837	16378	16378	833	16067	16067	839	16429	16429
Country Fixed Effects	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No
Country-Ethnicity Fixed Effects	No	No	Yes	No	No	Yes	No	No	Yes
District Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample, Districts	>1 groups	>1 groups	>1 groups	>1 groups	>1 groups	>1 groups	>1 groups	>1 groups	>1 groups

The table reports OLS estimates, associating measures of public goods (access to piped water, access to a sewage system, and access to an electricity grid) with inequality between and within ethnic groups at the district level, as reflected in the Theil index. Columns (1), (4), and (7) report district-level estimates, while columns (2), (3), (5), (6), (8), and (9) report individual level estimates. The dependent variable in columns (1)-(3) is a dummy variable that takes on the value of one for households that have access to clean piped water; in columns (4)-(6) is a dummy variable that takes on the value of one for households that have access to a sewage system; in columns (7)-(9) is a dummy variable that takes on the value of one for households that have access to an electricity grid. The between-ethnic-group and the within-ethnic-group Theil indicators are based on individuals' responses on living conditions.

The district-level conditioning set includes the log number of ethnic groups in each district, the log number of respondents in each district, and district-level ethnic fractionalization. The individual-level conditioning set includes age, age squared, a gender indicator variable, 22 religion fixed effects, 25 occupation fixed effects, 5 living conditions fixed effects, and 9 education fixed effects. In all specifications we exclude from the estimation districts with respondents from just one ethnic group. Panel B reports otherwise identical to Panel A specifications including also in the set of explanatory variables the mean value of the living conditions index in the district. All specifications include country fixed effects (constants not reported). Specifications (3), (6), and (9) include ethnicity-country fixed effects (constants not reported). The Data Appendix gives detailed variable definitions and data sources. All variables are constructed using data from the 3rd round of the Afrobarometer Surveys. Clustered at the district level standard errors are reported in parentheses below the estimates. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Table 11a - Ethnic Inequality and Beliefs on Democracy within African Countries

	Support for Democracy			Satisfaction with Democracy			Constituion Reflects People's Values-Hopes		
	District	Individual Level		District	Individual Level		District	Individual Level	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Between Group Ineq.	-1.4171*	-0.6833	-0.9821	-2.5130**	-1.9787***	-1.8511***	-1.8538*	-1.6343*	-1.4638*
[Theil Index]	(0.7702)	(0.6387)	(0.6341)	(0.8794)	(0.7619)	(0.6824)	(0.9479)	(0.8803)	(0.8169)
Within Group Ineq.	-0.5123	-0.5371*	-0.4089	-1.3277**	-1.2481**	-0.9595**	-0.0716	-0.6032	0.1742
[Theil Index]	(0.3996)	(0.3228)	(0.3101)	(0.4655)	(0.5182)	(0.4654)	(0.8013)	(0.5818)	(0.4933)
adjusted R-squared	0.244	0.085	0.098	0.414	0.156	0.184	0.187	0.052	0.071
Observations	838	14381	14381	837	14231	14231	839	13923	13923
Country Fixed Effects	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No
Country-Ethnicity Fixed Effects	No	No	Yes	No	No	Yes	No	No	Yes
District Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Sample, Districts	>1 group	>1 group	>1 group	>1 group	>1 group	>1 group	>1 group	>1 group	>1 group

Table 11b - Ethnic Inequality and Beliefs on Democracy within African Countries

	Support for Democracy			Satisfaction with Democracy			Constitution Reflects People's Values-Hope		
	District	Individual Level		District	Individual Level		District	Individual Level	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Between Group Ineq.	-1.2924*	-0.6556	-1.0286	-1.7639*	-1.181	-1.5854**	-0.933	-0.4923	-0.8835
[Theil Index]	(0.6844)	(0.6489)	(0.6442)	(0.9030)	(0.7817)	(0.6962)	(0.9094)	(0.9260)	(0.8553)
Within Group Ineq.	-0.3964	-0.5134	-0.4458	-0.6305	-0.5556	-0.7441	0.7744	0.3774	0.6381
[Theil Index]	(0.3520)	(0.3497)	(0.3268)	(0.5149)	(0.5320)	(0.4956)	(0.7115)	(0.5962)	(0.5442)
Mean Living Conditions at the District	0.0202	0.0044	-0.0077	0.1215**	0.1271***	0.0445	0.1492***	0.1756***	0.0932**
	(0.0301)	(0.0257)	(0.0260)	(0.0429)	(0.0386)	(0.0343)	(0.0427)	(0.0482)	(0.0410)
adjusted R-squared	0.243	0.085	0.098	0.421	0.157	0.184	0.2	0.054	0.072
Observations	838	14381	14381	837	14231	14231	839	13923	13923
Country Fixed Effects	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No
Country-Ethnicity Fixed Effects	No	No	Yes	No	No	Yes	No	No	Yes
District Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample, Districts	>1 group	>1 group	>1 group	>1 group	>1 group	>1 group	>1 group	>1 group	>1 group

The table reports OLS estimates, associating measures reflecting individual's beliefs on support and satisfaction with democracy with inequality between and within ethnic groups at the district level, as reflected in the Theil index. Columns (1), (4), and (7) report district-level estimates, while columns (2), (3), (5), (6), (8), and (9) report individual level estimates. The dependent variable in columns (1)-(3) is a trichotomous (0-3 range) variable that reflects' individual's beliefs on support for democratic rule; in columns (4)-(6) the dependent variable is an ordered (0-4 range) index that reflects individual's satisfaction with the functioning of democratic institutions; in columns (7)-(9) the dependent variable is an ordered (1-5 range) index that captures individual's beliefs on whether the constitution reflects people's hope and values. The between-ethnic-group and the within-ethnic-group Theil indicators are based on individuals' responses on living conditions. The district-level conditioning set includes the log number of ethnic groups in each district, the log number of respondents in each district, and district-level ethnic fractionalization. The individual-level conditioning set includes age, age squared, a gender indicator variable, 22 religion fixed effects, 25 occupation fixed effects, 5 living conditions fixed effects, and 9 education fixed effects.

In all specifications we exclude from the estimation districts with respondents from just one ethnic group. Panel B reports otherwise identical to Panel A specifications including also in the set of explanatory variables the mean value of the living conditions index in the district.All specifications include country fixed effects (constants not reported). Specifications (3), (6), and (9) include ethnicity-country fixed effects (constants not reported). The Data Appendix gives detailed variable definitions and data sources. All variables are constructed using data from the 3rd round of the Afrobarometer Surveys. Clustered at the district level standard errors are reported in parentheses below the estimates. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Appendix Table 1: Correlation Structure - Cross-Country Measures

Panel A: Ethnic Inequality Indicators (all ethnic areas)

	Ethnic Inequality Indicators - Gini Coefficients						Overall Spatial Inequality Indicators - Gini Coefficients					
	GREG			Ethnologue			Spatial Gini 1			Spatial Gini 2		
Ethnic Gini 2009 (GREG)	1											
Ethnic Gini 2000 (GREG)	0.9545*	1										
Ethnic Gini 1992 (GREG)	0.9382*	0.9519*	1									
Ethnic Gini 2009 (ETHN)	0.7679*	0.7564*	0.7686*	1								
Ethnic Gini 2000 (ETHN)	0.7625*	0.7619*	0.7696*	0.9914*	1							
Ethnic Gini 1992 (ETHN)	0.7719*	0.7686*	0.7958*	0.9759*	0.9775*	1						
Spatial Gini 2009 (Thiessen)	0.7784*	0.7760*	0.7755*	0.8180*	0.8205*	0.8082*	1					
Spatial Gini 2000 (Thiessen)	0.7724*	0.7903*	0.7777*	0.8053*	0.8218*	0.8012*	0.9835*	1				
Spatial Gini 1992 (Thiessen)	0.7870*	0.7986*	0.8136*	0.8242*	0.8289*	0.8397*	0.9639*	0.9604*	1			
Spatial Gini 2009 (Pixels)	0.6826*	0.7002*	0.6633*	0.7161*	0.7214*	0.7096*	0.7699*	0.7845*	0.7638*	1		
Spatial Gini 2000 (Pixels)	0.6851*	0.7190*	0.6642*	0.7129*	0.7273*	0.7066*	0.7710*	0.8002*	0.7676*	0.9736*	1	
Spatial Gini 1992 (Pixels)	0.6860*	0.7187*	0.6886*	0.7347*	0.7446*	0.7535*	0.7807*	0.7970*	0.8129*	0.9384*	0.9463*	1

Appendix Table 1: Correlation Structure - Cross-Country Measures

Panel B: Ethnic Inequality Indicators (excl. capitals)

	Ethnic Inequality Indicators - Gini Coefficients						Overall Spatial Inequality Indicators - Gini Coefficients					
	GREG			Ethnologue			Spatial Gini 1			Spatial Gini 2		
Ethnic Gini 2009 (GREG)	1											
Ethnic Gini 2000 (GREG)	0.9440*	1										
Ethnic Gini 1992 (GREG)	0.8965*	0.9209*	1									
Ethnic Gini 2009 (ETHN)	0.6673*	0.6748*	0.6438*	1								
Ethnic Gini 2000 (ETHN)	0.6735*	0.6942*	0.6507*	0.9905*	1							
Ethnic Gini 1992 (ETHN)	0.6399*	0.6627*	0.6918*	0.9533*	0.9521*	1						
Spatial Gini 2009 (Thiessen)	0.6750*	0.6898*	0.6205*	0.8343*	0.8405*	0.7942*	1					
Spatial Gini 2000 (Thiessen)	0.6788*	0.7183*	0.6356*	0.8160*	0.8368*	0.7843*	0.9835*	1				
Spatial Gini 1992 (Thiessen)	0.6842*	0.7189*	0.6672*	0.8385*	0.8503*	0.8360*	0.9639*	0.9604*	1			
Spatial Gini 2009 (Pixels)	0.6118*	0.6319*	0.5494*	0.6884*	0.7013*	0.6513*	0.7699*	0.7845*	0.7638*	1		
Spatial Gini 2000 (Pixels)	0.6023*	0.6410*	0.5475*	0.6746*	0.6960*	0.6434*	0.7710*	0.8002*	0.7676*	0.9736*	1	
Spatial Gini 1992 (Pixels)	0.6117*	0.6494*	0.5657*	0.7152*	0.7333*	0.7038*	0.7807*	0.7970*	0.8129*	0.9384*	0.9463*	1

Appendix Table 1: Correlation Structure - Cross-Country Measures

Panel C: Correlation of Ethnic Inequality Indicators with Measures of Ethnic-Linguistic-Religious Fragmentation, Polarization, and Segregation

Ethnic Gini 2000 - All (GRE	1											
Ethnic Gini 2000 - All (ETH	0.7619*	1										
Spatial Gini 2000 (Thiessen)	0.7903*	0.8218*	1									
Spatial Gini 2000 (Pixels)	0.7190*	0.7273*	0.8002*	1								
Ethnic Fragmentation	0.4464*	0.4666*	0.5099*	0.4640*	1							
Linguistic Fragmentation	0.3878*	0.4123*	0.4653*	0.3506*	0.6885*	1						
Religious Fragmentation	-0.057	-0.0035	0.044	0.0041	0.1629*	0.2748*	1					
Ethnic Segregation	0.2944*	0.4468*	0.3348*	0.2064*	0.4813*	0.3705*	-0.0442	1				
Linguistic Segregation	0.2437*	0.3711*	0.2266*	0.2131*	0.3945*	0.3056*	-0.0363	0.8422*	1			
Religious Segregation	0.2552*	0.2449*	0.2249*	0.2097	0.2502*	0.2911*	0.0811	0.2205	0.1276	1		
Ethnic Polarization 1	0.1042	0.0955	0.0144	0.0942	0.3065*	0.2617*	-0.1019	0.1196	0.1781	0.0251	1	
Ethnic Polarization 2	0.0497	0.0335	-0.0345	0.0254	0.1697*	0.1032	-0.0389	0.0654	0.1151	-0.0012	0.5161*	1

Appendix Table 1: Correlation Structure - Cross-Country Measures

Panel D: Correlation of Ethnic Inequality Indicators with Measures of Development and Income Inequality

Ethnic Gini - All (GREG)	1.0000											
Ethnic Gini - Excl. Capitals (	0.9404*	1.0000										
Ethnic Gini - Excl. Small (G	0.6992*	0.6137*	1.0000									
Ethnic Gini - All (ETHN)	0.7619*	0.7096*	0.6666*	1								
Ethnic Gini - Excl. Capitals (	0.7229*	0.6942*	0.6560*	0.9831*	1							
Ethnic Gini - Excl. Small (E'	0.6337*	0.6069*	0.7785*	0.8183*	0.7687*	1						
Spatial Gini 2000 (Thiessen)	0.7903*	0.7183*	0.7104*	0.8218*	0.8368*	0.7914*	1					
Spatial Gini 2000 (Pixels)	0.7190*	0.6410*	0.5593*	0.7273*	0.6960*	0.6448*	0.8002*	1				
Income Inequality (Gini coet	0.2643*	0.2608*	0.3063*	0.3260*	0.3151*	0.4228*	0.3452*	0.2887*	1			
Log real GDP p.c. in 2000	-0.5294*	-0.5193*	-0.6552*	-0.4950*	-0.5250*	-0.5795*	-0.5611*	-0.4556*	-0.3751*	1		
Rule of Law in 2000	-0.4933*	-0.5324*	-0.5081*	-0.4464*	-0.4869*	-0.4999*	-0.4982*	-0.4108*	-0.3998*	0.7952*	1	
Control of Corruption in 200	-0.4570*	-0.4944*	-0.4984*	-0.4207*	-0.4447*	-0.4753*	-0.4658*	-0.3677*	-0.4041*	0.7400*	0.9423*	1

The table reports the correlation structure between the main variables employed in the cross-country empirical analysis. Panel A gives the correlation between the main ethnic inequality measures and the overall spatial inequality measures in 1992, 2000, and 2009.
Panel B gives the correlation between ethnic inequality and the overall spatial inequality measures in 1992, 2000, and 2009. For the estimation of the ethnic inequality measures (Gini coefficients) we exclude ethnic regions where capital cities fall.
Panel C gives the correlation between the main ethnic inequality measures and the overall spatial inequality measures in 2000 with measures reflecting ethnic, linguistic, and religious fragmentation, segregation, and polarization.
Panel D gives the correlation between the main ethnic inequality measures and the overall spatial inequality measures in 2000 with income inequality and measures capturing economic and institutional development in 2000.
The Data Appendix gives detailed variable definitions and data sources. * indicates statistical significance at the 5% level.

Appendix Table 2 - Additional Sensitivity Checks: Ethnic Inequality and Economic Development (in 2000)
Excluding Countries with One Ethnic-Linguistic Group

Panel A: Conditioning on Overall Spatial Inequality Index based on Thiessen Polygons

	Atlas Narodov Mira (GREG)						Ethnologue					
	All Ethnic Areas		Excl. Capitals		Excl. Small Groups		All Ethnic Areas		Excl. Capitals		Excl. Small Groups	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Ethnic Inequality	-1.6096***	-1.3812***	-1.1601***	-1.0370***	-2.0382***	-1.4606***	-1.1754***	-0.7805**	-1.1723***	-0.8062**	-1.3301***	-0.9689*
[Gini Coeff.]	(0.4575)	(0.3835)	(0.4023)	(0.3309)	(0.5897)	(0.5397)	(0.3749)	(0.3514)	(0.3747)	(0.3552)	(0.5041)	(0.4964)
Spatial Inequality 2	-0.4639	-0.0641	-0.7642	-0.275	-0.8093*	-0.4447	-0.4912	-0.3364	-0.4793	-0.2985	-0.6475	-0.3788
[Gini Coeff.]	(0.5269)	(0.4456)	(0.5395)	(0.4501)	(0.4791)	(0.4388)	(0.4748)	(0.4669)	(0.4822)	(0.4757)	(0.5049)	(0.4473)
Ethnic	0.0863	0.2430	0.0709	0.2411	0.4818	0.534	-0.1566	0.1251	-0.1721	0.1237	-0.1395	0.1456
Fragmentation	(0.3834)	(0.3359)	(0.3904)	(0.3437)	(0.4254)	(0.3974)	(0.3721)	(0.3373)	(0.3958)	(0.3544)	(0.3780)	(0.3440)
adjusted R-squared	0.680	0.741	0.676	0.741	0.691	0.739	0.677	0.729	0.672	0.723	0.674	0.7288
Observations	153	153	152	152	153	153	148	148	147	147	148	148
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Simple	Rich	Simple	Rich	Simple	Rich	Simple	Rich	Simple	Rich	Simple	Rich

Appendix Table 2 - Additional Sensitivity Checks: Ethnic Inequality and Economic Development (in 2000), cont.
Excluding Countries with One Ethnic-Linguistic Group

Panel B: Conditioning on Overall Spatial Inequality Index based on Pixels of same Size (2.5 x 2.5 degrees)

	Atlas Narodov Mira (GREG)						Ethnologue					
	All Ethnic Areas		Excl. Capitals		Excl. Small Groups		All Ethnic Areas		Excl. Capitals		Excl. Small Groups	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Ethnic Inequality	-1.3891***	-1.0916***	-1.1161***	-0.9068***	-1.9173***	-1.2843**	-1.0929***	-0.7285**	-1.1124***	-0.7595***	-1.2700***	-0.8987**
[Gini Coeff.]	(0.4090)	(0.3465)	(0.3490)	(0.2934)	(0.5602)	(0.5102)	(0.3121)	(0.2925)	(0.3027)	(0.2829)	(0.3931)	(0.4043)
Spatial Inequality 1	-1.1909**	-0.9672**	-1.5238***	-1.2222***	-1.3685***	-1.1437***	-1.3388***	-1.0038**	-1.3646***	-1.0080**	-1.3792***	-0.9862**
[Gini Coeff.]	(0.5053)	(0.4173)	(0.4953)	(0.4028)	(0.4729)	(0.4032)	(0.4747)	(0.4440)	(0.4659)	(0.4347)	(0.4688)	(0.4195)
Ethnic	0.1211	0.2454	0.1247	0.2408	0.5101	0.5178	-0.1528	0.0998	-0.1663	0.0986	-0.1355	0.1206
Fragmentation	(0.3782)	(0.3285)	(0.3804)	(0.3273)	(0.4144)	(0.3778)	(0.3627)	(0.3376)	(0.3835)	(0.3516)	(0.3659)	(0.3418)
adjusted R-squared	0.689	0.748	0.692	0.753	0.701	0.749	0.690	0.736	0.687	0.731	0.687	0.7354
Observations	153	153	152	152	153	153	148	148	147	147	148	148
Region Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Simple	Rich	Simple	Rich	Simple	Rich	Simple	Rich	Simple	Rich	Simple	Rich

Both panels report cross-country OLS estimates. In all specifications we drop countries with just one ethnic group (in (1)-(6)) or just one linguistic group (in (7)-(12)). The dependent variable is the log of real GDP per capita in 2000. In all specifications in Panel A we control for the overall degree of spatial inequality in a country using the Gini coefficient of lights per capita based on Thiessen polygons that use as input points the centroids of the linguistic homelands according to the Ethnologue dataset. In all specifications in Panel B we control for the overall degree of spatial inequality in a country using the Gini coefficient of lights per capita based on polygons that have the same size (2.5 x 2.5 degrees).
In columns (1)-(6) we use the digitized version of the Atlas Narodov Mira (GREG) to aggregate lights per capita across ethnic homelands and construct the ethnic inequality measures. In columns (7)-(12) we use the digitized version of the Ethnologue database to aggregate lights per capita across linguistic homelands and construct the ethnic inequality measures. For the construction of the ethnic inequality measures (Gini coefficients) in columns (3), (4), (9), and (10) we exclude ethnic areas where capital cities fall. For the construction of the ethnic inequality measures (Gini coefficients) in columns (5), (6), (11), and (12) we exclude small ethnic groups consisting of less than one percept of country's population. Odd-numbered columns include as controls absolute latitude, log land area, and log population in 2000 (simple set of controls). Even-numbered columns include also control for an index of terrain ruggedness, the percentage of each country with fertile soil, the percentage of each country with tropical climate, average distance to nearest ice-free coast, and an index of gem-quality diamond extraction (rich set of controls). In all specifications we control for ethnic fragmentation using an index that reflects the likelihood that two randomly chosen individuals in one country will be members of the same group.
All specifications include continental fixed effects (constants not reported). The Data Appendix gives detailed variable definitions and data sources. Heteroskedasticity-adjusted standard errors are reported in parentheses below the estimates. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Appendix Table 3A: Geography, History and Contemporary Ethnic Inequality, Atlas Naodov Mira

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Land Area	0.0247**	0.0235**	0.0204**	0.0071	0.0147	0.0171	0.0034	0.0024
	(0.0098)	(0.0096)	(0.0094)	(0.0152)	(0.0101)	(0.0111)	(0.0124)	(0.0113)
Log Population	0.003	0.0029	0.0004	-0.0046	-0.0030	-0.0052	-0.0166	-0.0164
	(0.0117)	(0.0127)	(0.0125)	*(0.0173)*	*(0.0148)*	*(0.0149)*	(0.0134)	(0.0212)
Latitude	0.0008	0.0009	0.0004	-0.0021	0.0045**	0.0030	0.0026	0.0053**
	(0.0018)	(0.0022)	(0.0023)	(0.0032)	(0.0022)	(0.0022)	(0.0023)	(0.0025)
Ruggedness		-0.0031	-0.0051	0.0200	(0.0022)	0.0013	0.0023	0.0128
		(0.0104)	(0.0107)	(0.0249)	(0.0107)	(0.0116)	(0.0120)	(0.0098)
Soil Quality		0.0002	0.0003	0.0004	0.0002	-0.0003	0.0005	0.0002
		(0.0006)	(0.0006)	(0.0010)	(0.0006)	(0.0007)	(0.0007)	(0.0008)
Tropical Climate		0.0001	0.00001	-0.0003	0.0007	0.0009*	0.0006	0.0011**
		(0.0005)	(0.0005)	(0.0006)	(0.0005)	(0.0005)	(0.0006)	(0.0006)
Gem Stones		0.0001	0.0002	0.0001	0.0002	0.0003	0.0001	0.0002
		(0.0001)	(0.0003)	(0.0002)	(0.0004)	(0.0002)	(0.0002)	(0.0005)
Distance to the Coast		0.0486	0.0476	0.0677	0.0226	0.1407***	0.0393	0.0107
		(0.0422)	(0.0412)	*(0.0606)*	*(0.0437)*	*(0.0483)*	(0.0432)	(0.0381)
Common Law			-0.0410					
			(0.0320)					
Log Settler Mortality				0.0133				
				(0.0205)				
European Descent					-0.0011*			
					(0.0006)			
Executive Constraints at Independence						-0.0012		
						(0.0411)		
State Antiquity Index							0.1211	
							(0.0829)	
Ethnic Partitioning								-0.0004
								(0.0005)
Border Straightness								-0.0213
								(0.8016)
Spatial Inequality 2 [Gini Coeff.]	0.6669***	0.6515***	0.6429***	0.6650***	0.7010***	0.5805***	0.7564***	0.7448***
	(0.0816)	(0.0875)	(0.0891)	(0.1092)	(0.0783)	(0.0884)	(0.0846)	(0.1047)
adjusted R-squared	0.684	0.676	0.687	0.665	0.668	0.631	0.723	0.653
Observations	173	173	173	77	157	133	142	113
Region Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Appendix Table 3B: Geography, History and Contemporary Ethnic Inequality, Ethnologue

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Land Area	0.0247**	0.0235**	0.0204**	0.0071	0.0147	0.0171	0.0034	0.0024
	(0.0098)	(0.0096)	(0.0094)	(0.0152)	(0.0101)	(0.0111)	(0.0124)	(0.0113)
Log Population	0.003	0.0029	0.0004	-0.0046	-0.0030	-0.0052	-0.0166	-0.0164
	(0.0117)	(0.0127)	(0.0125)	*(0.0173)*	*(0.0148)*	*(0.0149)*	(0.0134)	(0.0212)
Latitude	0.0008	0.0009	0.0004	-0.0021	0.0045**	0.0030	0.0026	0.0053**
	(0.0018)	(0.0022)	(0.0023)	(0.0032)	(0.0022)	(0.0022)	(0.0023)	(0.0025)
Ruggedness		-0.0031	-0.0051	0.0200	(0.0022)	0.0013	0.0023	0.0128
		(0.0104)	(0.0107)	(0.0249)	(0.0107)	(0.0116)	(0.0120)	(0.0098)
Soil Quality		0.0002	0.0003	0.0004	0.0002	-0.0003	0.0005	0.0002
		(0.0006)	(0.0006)	(0.0010)	(0.0006)	(0.0007)	(0.0007)	(0.0008)
Tropical Climate		0.0001	0	-0.0003	0.0007	0.0009*	0.0006	0.0011**
		(0.0005)	(0.0005)	(0.0006)	(0.0005)	(0.0005)	(0.0006)	(0.0006)
Gem Stones		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Distance to the Coast		0.0486	0.0476	0.0677	0.0226	0.1407***	0.0393	0.0107
		(0.0422)	(0.0412)	*(0.0606)*	*(0.0437)*	*(0.0483)*	(0.0432)	(0.0381)
Common Law			-0.0410					
			(0.0320)					
Log Settler Mortality				0.0133				
				(0.0205)				
European Descent					-0.0011*			
					(0.0006)			
Executive Constraints at Independence						-0.0012		
						(0.0411)		
State Antiquity Index							0.1211	
							(0.0829)	
Ethnic Partitioning								-0.0004
								(0.0005)
Border Straightness								-0.0213
								(0.8016)
Spatial Inequality 2 [Gini Coeff.]	0.6669***	0.6515***	0.6429***	0.6650***	0.7010***	0.5805***	0.7564***	0.7448***
	(0.0816)	(0.0875)	(0.0891)	(0.1092)	(0.0783)	(0.0884)	(0.0846)	(0.1047)
adjusted R-squared	0.6652	0.659	0.6606	0.6239	0.6493	0.622	0.6826	0.6502
Observations	173	173	173	77	157	133	142	113
Region Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

The table reports cross-country OLS estimates, associating contemporary ethnic inequality with various geographic and historical variables. The dependent variable is the ethnic Gini coefficient that reflects inequality in lights per capita across ethnic-linguistic homelands, using the digitized version of Atlas Narodov Mira (GREG) in Panel A and Ethnologue in Panel B. In all specifications we control for the overall degree of spatial inequality in a country using the Gini coefficient of lights per capita based on Thiessen polygons that use as input points the centroids of the linguistic homelands according to the Ethnologue dataset. All specifications include continental fixed effects (constants not reported). The Data Appendix gives detailed variable definitions and data sources. Heteroskedasticity-adjusted standard errors are reported in parentheses below the estimates. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Appendix Table 4: Correlation Structure Inequality Measures in Geographic Endowments

	Distance to the Sea			Elevation			Land Quality			Water Area		
Ethnic Gini - Sea Distance	1											
Spatial Gini 1 - Sea Distance	0.7971*	1										
Mean Sea Distance	-0.0306	-0.1910*	1									
Ethnic Gini - Elevation	0.3879*	0.2692*	0.2612*	1								
Spatial Gini 1 - Elevation	0.3313*	0.2698*	0.2131*	0.8776*	1							
Mean Elevation	0.0254	-0.0979	0.5010*	0.6012*	0.5662*	1						
Ethnic Gini - Land Quality	0.3702*	0.2460*	0.3698*	0.5181*	0.4110*	0.3367*	1					
Spatial Gini 1 - Land Quality	0.3007*	0.2311*	0.3577*	0.4475*	0.4061*	0.3360*	0.9253*	1				
Mean Land Quality	0.0299	0.0005	-0.1677*	-0.0075	0.0180	0.0238	-0.4825*	-0.5423*	1			
Ethnic Gini - Water Area	0.6298*	0.4833*	0.3505*	0.5288*	0.3904*	0.3606*	0.5002*	0.4133*	0.1217	1		
Spatial Gini 1 - Water Area	0.5081*	0.4924*	0.3819*	0.4074*	0.3746*	0.3735*	0.3315*	0.3049*	0.0735	0.7775*	1	
Mean Water Area	0.2944*	0.2538*	0.1472	0.2078*	0.1995*	0.0365	0.2996*	0.3041*	-0.1798*	0.1648*	0.1236	1

The table reports the correlation structure between the main geographic variables employed in the cross-country analysis within African countries. Specifically the table gives the correlation between inequality in geographic endowments across ethnic homelands, inequality in geographic endowments across pixels of 2.5x2.5 degrees, and the level of geography, as reflected in distance to the sea, elevation, an index of land (soil) suitability (quality) for agriculture, and water area. * indicate statistical significance at the 5% level.

Appendix Table 5: The Origins of Contemporary Ethnic Inequality. Sensitivity Analysis
Inequality in Geographic Endowments across Ethnic Homelands and Contemporary Ethnic Inequality

	Atlas Narodov Mira (GREG)			Ethnologue		
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Excluding Capitals			
Inequality in Geographic Endowments across Ethnic Homelands (PC)	0.0984***	0.0846***	0.0841***	0.1422***	0.1220***	0.1334***
	(0.0102)	(0.0158)	(0.0147)	(0.0123)	(0.0180)	(0.0175)
	9.67	*5.34*	*5.73*	*11.53*	*6.78*	*7.64*
Spatial Inequality in Geographic Endowments (PC)		0.0256	0.0122		0.0282	-0.0048
		(0.0209)	(0.0211)		(0.0224)	(0.0235)
		1.22	*0.58*		*1.25*	*-0.20*
adjusted R-squared	0.538	0.542	0.582	0.676	0.674	0.700
Observations	151	150	150	144	143	142
			Panel B: Excluding Small Groups			
Inequality in Geographic Endowments across Ethnic Homelands (PC)	0.0760***	0.0710***	0.0680***	0.0934***	0.0957***	0.0960***
	(0.0072)	(0.0090)	(0.0090)	(0.0079)	(0.0105)	(0.0101)
	10.62	*7.92*	*7.57*	*11.76*	*9.08*	*9.47*
Spatial Inequality in Geographic Endowments (PC)		0.0024	0.0043		-0.0062	-0.0079
		(0.0079)	(0.0096)		(0.0101)	(0.0127)
		0.31	*0.45*		*-0.62*	*-0.62*
adjusted R-squared	0.6165	0.6227	0.6106	0.6365	0.6348	0.6202
Observations	173	169	162	169	167	161
Region Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	No	No	Geography	No	No	Geography

The table reports cross-country OLS estimates, associating contemporary ethnic inequality with inequality in geographic endowments across ethnic homelands. The dependent variable is the ethnic Gini coefficient that reflects inequality in lights per capita across ethnic-linguistic homelands in 2000, using the digitized version of Atlas Narodov Mira (GREG) (in columns (1)-(3)) and Ethnologue (in columns (4)-(6)). To construct the ethnic inequality index (Gini coefficient) we exclude ethnic regions where capital cities fall (in Panel A) and ethnic regions where small ethnicities consisting less than one percent of a country's population reside (in Panel B).
The main independent variable is a composite index capturing inequality in geographic endowments across ethnic homelands. The index is the first principal component of inequality across ethnic-linguistic homelands in distance to the coast, elevation, land suitability for agriculture, and area under water. In columns (2), (3), (5), and (6) we control for the overall degree of spatial inequality in geographic endowments using a composite index that aggregates (via principal components) Gini coefficients on distance to the coast, elevation, land suitability for agriculture, water area across Thiessen polygons that use as input points the centroids of the linguistic homelands according to the Ethnologue dataset. In columns (3) and (6) we also control for the mean value of distance to the coast, elevation, land suitability for agriculture, and area under water.
All specifications include continental fixed effects (constants not reported). The Data Appendix gives detailed variable definitions and data sources. Heteroskedasticity-adjusted standard errors are reported in parentheses below the estimates. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Appendix Table 6: Inequality in Geographic Endowments across Ethnic Homelands and Contemporary Development. Sensitivity Analysis

	Atlas Narodov Mira (GREG)			Ethnologue		
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Excluding Capitals			
Inequality in Geographic Endowments across Ethnic Homelands (PC)	-0.1176**	-0.1152*	-0.1268*	-0.0861	-0.008	-0.0593
	(0.0476)	(0.0676)	(0.0704)	(0.0621)	(0.0956)	(0.0980)
	-2.47	*-1.7*	*-1.80*	*-1.39*	*-0.08*	*-0.6*
Spatial Inequality in Geographic Endowments (PC)		0.0052	0.0153		-0.1313	-0.0876
		(0.1094)	(0.1244)		(0.1011)	(0.1185)
		0.05	*0.12*		*-1.30*	*-0.74*
adjusted R-squared	0.633	0.633	0.664	0.638	0.640	0.662
Observations	151	150	150	144	143	142
			Panel B: Excluding Small Groups			
Inequality in Geographic Endowments across Ethnic Homelands (PC)	-0.2075***	-0.1951***	-0.1709***	-0.1781***	-0.1346*	-0.1732**
	(0.0431)	(0.0576)	(0.0591)	(0.0446)	(0.0695)	(0.0671)
	-4.81	*-3.39*	*-2.89*	*-4.00*	*-1.94*	*-2.58*
Spatial Inequality in Geographic Endowments (PC)		-0.0061	0.0005		-0.0267	0.0352
		(0.0557)	(0.0718)		(0.0735)	(0.0874)
		-0.11	*0.01*		*-0.36*	*0.40*
adjusted R-squared	0.6397	0.6550	0.6780	0.6259	0.6393	0.6729
Observations	173	169	162	169	167	161
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	No	No	Geography	No	No	Geography

The table reports cross-country OLS estimates, associating contemporary economic development with inequality in geographic endowments across ethnic homelands. The dependent variable is the log of real GDP per capita in 2000. To construct the ethnic inequality index and the inequality in geographic endowments across ethnic homelands (Gini coefficients) we exclude ethnic regions where capital cities fall (in Panel A) and ethnic regions where small ethnicities consisting less than one percent of a country's population reside (in Panel B).

The main independent variable is a composite index capturing inequality in geographic endowments across ethnic homelands, using the digitized version of Atlas Narodov Mira (GREG) (in columns (1)-(3)) and Ethnologue (in columns (4)-(6)). The index is the first principal component of inequality across ethnic-linguistic homelands in distance to the coast, elevation, land suitability for agriculture, and area under water. In columns (2), (3), (5), and (6) we control for the overall degree of spatial inequality in geographic endowments using a composite index that aggregates (via principal components) Gini coefficients on distance to the coast, elevation, land suitability for agriculture, water area across Thiessen polygons that use as input points the centroids of the linguistic homelands according to the Ethnologue dataset. In columns (3) and (6) we also control for the mean value of distance to the coast, elevation, land suitability for agriculture, and area under water.

All specifications include continental fixed effects (constants not reported). The Data Appendix gives detailed variable definitions and data sources. Heteroskedasticity-adjusted standard errors are reported in parentheses below the estimates. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Appendix Table 7: Inequality in Geographic Endowments across Ethnic Homelands, Ethnic Inequality, and Contemporary Development. 2SLS Estimates

	Atlas Narodov Mira (GREG)			Ethnologue		
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: All Ethnic Homelands			
Inequality in Geographic Endowments across Ethnic Homelands (PC)	-1.4574***	-2.1223***	-1.8196**	-0.9983***	-0.3808	-0.9613*
	(0.3034)	(0.7983)	(0.8996)	(0.2655)	(0.6024)	(0.5002)
	-4.8	*-2.66*	*-2.02*	*-3.76*	*-0.63*	*-1.92*
Spatial Inequality in Geographic Endowments (PC)		0.0986	0.0603		-0.0869	0.021
		(0.1015)	(0.1086)		(0.0981)	(0.0982)
		0.97	*0.56*		*-0.89*	*0.21*
First-Stage F-score	195.96	37.54	32.43	138.19	24.99	26.43
Observations	173	169	162	168	166	160
			Panel B: Excluding Capitals			
Inequality in Geographic Endowments across Ethnic Homelands (PC)	-1.1956**	-1.3624*	-1.5067*	-0.6051	-0.0657	-0.4444
	(0.4778)	(0.7382)	(0.7919)	(0.4015)	(0.7518)	(0.6787)
	-2.5	*-1.85*	*-1.90*	*-1.51*	*-0.09*	*-0.65*
Spatial Inequality in Geographic Endowments (PC)		0.04	0.0337		-0.1295	-0.0898
		(0.1146)	(0.1228)		(0.1139)	(0.1062)
		0.35	*0.27*		*-1.14*	*-0.85*
First-Stage F-score	93.418	28.529	32.801	133.040	45.965	58.354
Observations	151	150	150	144	142	142
			Panel C: Excluding Small Groups			
Inequality in Geographic Endowments across Ethnic Homelands (PC)	-2.7288***	-2.7499***	-2.5126***	-1.9062***	-1.4074**	-1.8042***
	(0.5617)	(0.8320)	(0.8745)	(0.4444)	(0.7026)	(0.6425)
	-4.86	*-3.31*	*-2.87*	*-4.29*	*-2.00*	*-2.81*
Spatial Inequality in Geographic Endowments (PC)		0.0006	0.0113		-0.0355	0.0208
		(0.0547)	(0.0715)		(0.0668)	(0.0791)
		0.01	*0.16*		*-0.53*	*0.26*
First-Stage F-score	112.78	62.693	57.242	138.19	82.427	89.658
Observations	173	169	162	169	167	161
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	No	No	Geography	No	No	Geography

Table Notes

The table reports cross-country two-stage-least-squares (2SLS) estimates, associating contemporary inequality in lights per capita across ethnic homelands with inequality in geographic endowments across ethnic homelands in the first stage and the component of ethnic inequality explained by inequality in geographic endowments across ethnic homelands with economic development in the second stage. The dependent variable in the second stage is the log of real GDP per capita in 2000. The dependent variable in the first stage is the ethnic Gini coefficient that reflects inequality in lights per capita across ethnic-linguistic homelands in 2000, using the digitized version of Atlas Narodov Mira (GREG) (in columns (1)-(3)) and Ethnologue (in columns (4)-(6)). In Panel A we use all ethnic-linguistic homelands. In Panel B we exclude ethnic-linguistic regions where capital cities fall. In Panel C we exclude ethnic-linguisic regions where small ethnicities consisting less than one percent of a country's population reside.
The main independent variable in the first stage is a composite index capturing inequality in geographic endowments across ethnic homelands, using the digitized version of Atlas Narodov Mira (GREG) (in columns (1)-(3)) and Ethnologue (in columns (4)-(6)). The index is the first principal component of inequality across ethnic-linguistic homelands in distance to the coast, elevation, land suitability for agriculture, and area under water. In columns (2), (3), (5), and (6) we control for the overall degree of spatial inequality in geographic endowments using a composite index that aggregates (via principal components) Gini coefficients on distance to the coast, elevation, land suitability for agriculture, water area across Thiessen polygons that use as input points the centroids of the linguistic homelands according to the Ethnologue dataset. In columns (3) and (6) we also control for the mean value of distance to the coast, elevation, land suitability for agriculture, and area under water.
All specifications include continental fixed effects (constants not reported). The Data Appendix gives detailed variable definitions and data sources. Heteroskedasticity-adjusted standard errors are reported in parentheses below the estimates. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Appendix Table 8: Inequality in Geographic Endowments across Ethnic Homelands, Ethnic Inequality, and Contemporary Development. OLS Estimates

	Atlas Narodov Mira (GREG)			Ethnologue		
	All Homelands	Excl. Capitals	Excl. Small Groups	All Homelands	Excl. Capitals	Excl. Small Groups
	(1)	(2)	(3)	(4)	(5)	(6)
Contemporary Ethnic Inequality	-1.4885***	-1.3292***	-1.6339***	-1.2844***	-1.5078***	-1.2004**
	(0.4017)	(0.3833)	(0.5041)	(0.3620)	(0.4006)	(0.4696)
	-3.71	*-3.47*	*-3.24*	*-3.55*	*-3.76*	*-2.56*
Inequality in Geographic Endowments across Ethnic Homelands (PC)	0.0038	0.0131	-0.0832	0.0462	0.1284	-0.066
	(0.0617)	(0.0561)	(0.0525)	(0.0680)	(0.0841)	(0.0582)
	0.06	*0.23*	*-1.59*	*0.68*	*1.53*	*-1.13*
adjusted R-squared	0.659	0.659	0.663	0.652	0.678	0.641
Observations	173	151	173	168	144	169
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

The table reports cross-country OLS estimates, associating contemporary economic development with contemporary ethnic inequality and inequality in geographic endowments across ethnic homelands. The dependent variable is the log of real GDP per capita in 2000. In columns (1) and (4) we construct the ethnic inequality measures and the inequality in geographic endowments across ethnic homelands (Gini coefficients) using all ethnic-linguistic homelands. In columns (2) and (5) we exclude ethnic-linguistic regions where capital cities fall. In columns (3) and (6) we exclude ethnic-linguistic regions where small ethnicities consisting less than one percent of a country's population reside. The main independent variables are and index capturing contemporary differences in development (as reflected in lights per capita in 2000) across ethnic homelands and a composite index capturing inequality in geographic endowments across ethnic homelands. The index is the first principal component of inequality across ethnic-linguistic homelands in distance to the coast, elevation, land suitability for agriculture, and area under water. In columns (1)-(3) we use the digitized version of Atlas Narodov Mira (GREG) and in columns (4)-(6) we are using the Ethnologue maps. All specifications include continental fixed effects (constants not reported). The Data Appendix gives detailed variable definitions and data sources. Heteroskedasticity-adjusted standard errors are reported in parentheses below the estimates. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Appendix Table 9: Correlation Structure Afrobarometer Data at the District Level

	Ethnic Inequality Indicators						Development Proxy Measures				
Theil Index - Overall	1										
Theil Index - Between	0.9077*	1									
Theil Index - Within	0.5687*	0.1710*	1								
Mean Log Deviation - Overall	0.9893*	0.8951*	0.5683*	1							
Mean Log Deviation - Between	0.8985*	0.9863*	0.1762*	0.9072*	1						
Mean Log Deviation - Within	0.5568*	0.1637*	0.9864*	0.5649*	0.1653*	1					
Living Conditions Index	-0.2169*	-0.2095*	-0.0986*	-0.1603*	-0.1518*	-0.0781*	1.0000				
Education	-0.0114	-0.0308	0.0336	0.0196	0.0019	0.0422	0.3983*				
Sewage System	-0.0230	-0.0302	0.0053	-0.0076	-0.0158	0.0132	0.1696*	0.3944*	1.0000		
Clean Piped Water	-0.0098	-0.0133	0.0031	0.0039	-0.0020	0.0130	0.1319*	0.2747*	0.5517*	1	
Access to Electricity Grid	-0.0231	-0.0429	0.0299	0.0010	-0.0153	0.0324	0.2331*	0.4803*	0.5185*	0.5280*	1

The table reports the correlation structure between the main variables employed in the cross-region analysis within African countries (Afrobarometer Sample). The Data Appendix gives detailed variable definitions and data sources. * indicate statistical significance at the 5% level.