Ethnic Inequality*

Alberto Alesina
Harvard University and NBER and IGIER

Stelios Michalopoulos
Brown University and NBER

Elias Papaioannou
London Business School and CEPR and NBER

October 2012

Abstract

This study explores the consequences and origins of contemporary differences in well-being across ethnic groups within countries. We construct measures of ethnic inequality combining ethnolinguistic maps on the spatial distribution of groups with satellite images of light density at night. Ethnic inequality is strongly inversely related to per capita income; this pattern holds when we condition on the overall degree of spatial inequality -that is also associated with underdevelopment. We further show that differences in geographic endowments across ethnic homelands explain a sizable portion of contemporary ethnic inequality. This deeply-rooted inequality in geographic attributes across ethnic regions is also negatively related to comparative development. We also show that ethnic inequality goes in tandem with lower levels development also within countries. Using micro-level data from the Afrobarometer surveys we show that individuals from the same ethnic group are worse off when they reside in districts with a high degree of ethnic inequality.

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, 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, Bocconi, NYU, 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 public policies, low public goods provision, conflict, civil wars, and hatred. In fact a large literature shows a negative effect of ethnolinguistic fragmentation on various aspects of economic development, with the possible exception of wealthy economies.[1]

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

The first contribution of this paper is to provide measures of within-country differences in well-being across ethnic groups, "ethnic inequality". Information on income levels of ethnic groups for all countries are not available. Hence to construct country-level indicators of ethnic inequality for the largest possible sample, we combine ethnolinguistic maps on the location of ethnic groups with satellite images of light density at night, which are good 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 utilize the fine level of the luminosity data and also construct proxies of spatial inequality. Second, we find a remarkably strong negative association between ethnic inequality and real GDP per capita. This correlation holds even when we condition on the overall degree of spatial inequality, which is also inversely related to economic well-being. We also document 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

[1] See Alesina and La Ferrara (2005) for a review. We discuss the most closely related works in the next section.

is a strong predictor of ethnic inequality. Fourth, contemporary development is also inversely related to inequality in geographic endowments across ethnic homelands. Fifth, we examine the link between ethnic inequality and development using micro data from the Afrobarometer surveys, exploiting within-country across-district variation. Instead of assigning groups to ethnic homelands, we construct measures capturing between and within-group inequality utilizing individual-level data from more than $20,000$ respondents on ethnic identification and well-being. Individuals residing in ethnically unequal districts have lower standards of living and are less educated. Access to basic public goods, such as piped water and a sewage system, is also 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 we explore the richness of the data and control for numerous individual characteristics and ethnicity fixed effects (this is feasible as we observe respondents from the same ethnicity residing in multiple districts).

The paper is organized as follows. In the next section we discuss works linking ethnic inequality to under-development and place our findings within the existing literature. In section 3 we describe the construction of the ethnic inequality variable and present summary statistics. In Section 4 we report the results of our analysis associating income per capita with ethnic inequality across 173 countries. In Section 5 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 6 we examine the within-country across-district association between ethnic inequality and well-being and public goods in 17 Sub-Saharan countries using individual-level data from the Afrobarometer Surveys. In the last section we summarize.

2 Theoretical Channels and Related Works

Civil Conflict The idea that inequality spurs conflict leading to under-development dates back at least to Thomas Hobbes and Karl Marx. Yet, a large empirical literature does not uncover a robust association between income inequality and conflict (see Blattman and Miguel (2010)).[2] Horowitz (1985) offers an explanation for this (non) finding arguing that "*In much of Asia and Africa, it is only modest hyperbole to assert that the Marxian prophecy has had an ethnic fulfillment*".[3] Horowitz (1985) studies several cases in which civil wars and insurgent movements are ethnic rather than class-based. The literature on conflict has thus switched focus from income inequality to ethnic fractionalization (e.g., Collier and Hoeffler (1998)). Nevertheless, there is no systematic association between conflict and ethnic (or religious) fractionalization, as reflected in

[2] See Benabou (2005) and Galor (2011) for surveys of the vast research on inequality and development.

[3] This citation is borrowed from Esteban, Mayoral, and Ray (2012).

a Herfindahl-Hirschman index that reflects the likelihood that two randomly chosen individuals in a country will be members of different groups (e.g., Fearon and Laitin (2003)).

The apparent puzzle is resolved from a theoretical point of view by Esteban and Ray (1994, 1999) and empirically by Montalvo and Reynal-Querol (2005) and Esteban, Mayoral, and Ray (2012). They show that, unlike fractionalization, ethnic polarization is a key predictor of conflict. In recent work Esteban and Ray (2011a) provide a unified theoretical framework investigating how various aspects of a society's ethnic composition interact with civil conflict. The additional insight is that, besides the degree of polarization, inter-group distances, i.e., how "deep" ethnic divisions are, also shape conflict intensity. "Distance" measures are intended to capture the losses from having the public goods and policies that the other groups favor. The empirical literature has constructed indices of polarization and fragmentation taking into account cultural distances across groups using *Ethnologue's* language tree (Fearon (2003), Esteban, Mayoral, and Ray (2012), and Desmet, Ortuño-Ortín, and Wacziarg (2012)). Yet, an alternative and rather compelling measure of distance between ethnic groups is the level of income. Thus our ethnic inequality measures that capture income heterogeneity (or access to public goods) across ethnic lines naturally fit in this theoretical paradigm. In some sense the newly constructed indicators allow us testing a joint Horowitzian/Marxian view that arises from the interaction between class struggles and ethnicity. Robinson (2001) summarizes this idea in the context of Africa arguing that "*the effect of inequality on conflict depends on the socioeconomic composition of the ethnic groups. If these are unbalanced in the sense that one group contains mostly capitalists while the other contains mostly workers, then ethnic conflict increases with inequality*".

To the extent that poverty and inadequate public goods provision have an ethnic component, ethnic inequality may spur opposition towards the status-quo by enhancing group cohesion and increasing the salience of group identity. Caselli and Coleman (2012) develop a theory of ethnic conflict, where group identification is endogenous and ethnic markers, such as color, religion, and occupation, help enforce group membership. Income differences across groups may naturally act as an additional ethnic marker.

Taxation, Effective Governance and Public Goods Provision While differences in preferences do not always lead to conflict, they may still generate political strife, government stalemate, and under-provision of public goods. As long as preferences across groups differ, see Luttmer (2001), then average and median distance from the chosen public good is higher the more dispersed preferences are. A large literature that uses US city data 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)). Desmet, Ortuño-Ortín, and Wacziarg (2012) detect a robust cross-country negative cor-

relation between linguistic fractionalization and redistribution when focusing on deep linguistic cleavages. If income differences are correlated with racial and linguistic divisions then preferences across groups will be even more distant thus yielding lower public good provision and creating political tensions.[4] Baldwin and Huber (2010) provide empirical evidence in line with this idea for 46 democracies. Using survey-level data they show a negative correlation between an index of between-group inequality in welfare and the provision of public goods across countries. Finally, one of the key empirical results of urban economics is that the rich quite often want to "isolate"; since this desire may be especially strong when wealth is correlated with group identity, ethnic inequality may lead to segregation. Alesina and Zhuravskaya (2011) show that ethnic and linguistic segregation correlate negatively with proxies of effective governance and trust.

Social Capital One key component of good governance is social capital that is based on trust, political participation, and civic engagement (see Putnam (1993) for an early contribution and Guiso, Sapienza, and Zingales (2011) for a survey). Alesina and Ferrara (2000) develop a theory where civic participation is decreasing when inequality correlates with cultural heterogeneity (that manifests itself in diverse preferences). Yet, due to data limitations in their empirical analysis they relate group membership and participation in social activities across U. S. localities to income inequality and racial/ethnic fragmentation -rather than their intersection. Subsequent works both in the US and across countries similarly document a negative association between proxies of social capital and ethnic fragmentation and income inequality. Tesei (2011) examines the role of racial inequality on trust across US metropolitan areas and finds that racial inequality -rather than the overall degree of fragmentation- is the key correlate of trust.

Fairness and Markets Chua (2003) discusses cases in which the spread of democracy and free market institutions has led to animosity and institutional capture by amplifying pre-existing ethnic tensions. She notes that in many countries a small market-dominant ethnic minority controls a significant portion of the economy. Examples include Chinese minorities in Philippines, Indonesia, and other Eastern Asian countries; Lebanese communities in Western Africa; the I(g)bo in Nigeria or the Kikuyu in Kenya; whites in many countries in Latin America and in South Africa. Chua argues that since the benefits from trade and financial liberalization policies, privatization, and other free-market policies are mostly seized by these market-oriented minorities, ethnic inequality and in turn tensions rise. Most importantly, liberalization policies and democratization may result in a backlash, as most individuals perceive capitalist institutions as unfair, captured, and corrupt. Under democratic rule when the less-privileged, but more

[4] See Loury (2002) for an overview of the evidence on the evolution of racial inequality in the US. An analogous case is India, where inequality is concentrated across castes. See Deshpande (2000) and Anderson (2011) for empirical analyses associating between-caste inequality and public goods.

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.[5]

3 Data and Descriptive Statistics

3.1 Location of ethnic groups

We identify the location of ethnic groups employing two data sets. 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.[6] 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, 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.

[5] Alesina and Angeletos (2005) and Benabou and Tirole (2006) develop models linking the perception of fairness and the supply of effort.

[6] 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.

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.[7]

3.2 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., Michalopoulos and Papaioannou (2012)) we use satellite image data on light density at night as a proxy. 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.

We proxy the level development at the ethnic homeland level with average luminosity per capita; we then aggregate the values at the country level constructing a Gini index that reflects inequality across ethnic groups (ethnic inequality) within each country. Note that the ethnic Gini coefficient does not capture differences in individual income, but differences in mean income across groups. For each of the two different linguistic maps we construct Gini coefficients for all countries using cross-ethnic-homeland data in 1992, 2000, and 2009. For robustness we also construct the Gini coefficient dropping the capital 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

[7] Presently, we are including all groups in our analysis without attempting to make a distinction as to which cleavage is more salient.

and 16 large linguistic areas in Kenya, respectively.

3.3 Spatial inequality

Since we use ethnic homelands rather than individual-level data to measure between-group inequality, our ethnic inequality measures also capture regional disparities in income. Yet spatial inequality may not necessarily reflect ethnic-level differences. To isolate the between-ethnicity component of regional inequality, we thus 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.

Spatial Gini Coefficient 1 The first index is based on 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 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, rather than the purely ethnic-specific, component of spatial inequality in development.

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.[8] The mean size of the Thiessen polygons is 14.809 km, very

[8] 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.

similar to the mean size of homelands in the *Ethnologue* ($12,964\ km^2$).

The two proxies of the overall degree of spatial inequality also reflect inequality across ethnic homelands, since there is clearly some degree of measurement error on the exact boundaries of ethnic regions and because population mixing is in practice higher than the one we observe in the data.[9] 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.[10]

3.4 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 are the Tajiks, who compose 22% of the population and are located in the north-eastern regions and in scattered pockets in the western part of the country. There are 8 territories in which groups overlap. In four of those the Afghan groups overlap with the 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 $1b$ portrays the distribution of lights per capita. Regional development is low in the center, 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 a 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 $2a$). When we use Thiessen polygons

[9] In the last section we construct regional measures of ethnic inequality using individual-level data on ethnic identification and well-being, accounting therefore for these issues .

[10] 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.

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.

3.5 Descriptive Analysis

3.5.1 Ethnic Inequality around the World

Table 1 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 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 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 4*a* - 4*b* 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). On the other hand, USA and Canada score low in ethnic inequality as compared to the overall degree of spatial inequality. Azerbaijan, Syria, Albania, Tunisia, Haiti, and Rwanda score quite high in ethnic inequality, while in contrast the overall degree of spatial inequality is very low.

3.5.2 Basic Correlations

Ethnic Diversity Appendix Table 1 - Panel C reports the correlation 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, though somewhat smaller $(0.20 - 0.45)$. 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 examined 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. Yet this correlation weakens considerably and becomes statistically insignificant once we simply condition on continental constants.

4 Ethnic Inequality and Development

In Table 3 we report cross-country LS estimates associating the log of per capita GDP in year 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. The coefficient of the ethnic inequality index in column (1) is negative and

highly significant. Figures $7a - 7b$ illustrate the unconditional association.[11]

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.[12] 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.[13]

In columns (6)-(9) we add the log number of ethnic/linguistic groups. 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 the ethnic Gini is 0.76 to the level of Namibia where the ethnic Gini is 0.50) is associated by 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)).

[11] 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).

[12] This -to the best of our knowledge novel- result is in line with the model of Bolton and Roland (1997).

[13] Note that 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.

4.1 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.[14] 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

[14] 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)).

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 (captured by the region fixed effects).

In Appendix Table 2 we further explore the robustness of our estimates dropping (typically small) countries with just one ethnic group. Across all permutations ethnic inequality enters with a negative and highly significant estimate.

5 Inequality in Geographic Endowments and Ethnic Inequality

5.1 On the Origins of Ethnic Inequality

In this section we begin an exploration of the origins of ethnic inequality. We started with commonly used historical variables that have been found to correlate with development. Appendix Table 2 shows that 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)). These insignificant associations suggest that the strong negative correlation between ethnic inequality and development does not reflect the aforementioned aspects of history.

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 ethnic inequality. To the extent 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 for 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, 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 exactly geographic feature matters the most, the message from Table 6 is that exogenous 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 a principal component analysis. 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 to 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*".

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 remarkably strong positive association. As geographic inequality is to a first-approximation exogenous these graphs suggest that differences in geography explain a sizable portion of contemporary differences in development (or public goods provision) across ethnic homelands.[15]

In Table 7 we formally assess the role of ethnic-specific geographic inequality, as captured

[15] A possible source of endogeneity may be that in ancient times stronger (and perhaps more advanced) groups conquered territories of better quality. If this was indeed the case, current ethnic inequality would be due not only to geographic endowments but also to other deeply-rooted traits.

by the composite index of inequality in geographic endowments across 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 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. Appendix Table 4 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. 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*).

5.2 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 endogeneity due to omitted variables 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.

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 is quantitatively small, changes sign and is statistically insignificant. Appendix Table 5 reports otherwise identical specifications using the inequality measures that exclude from the estimation capitals and small ethnic groups. The results are similar. 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.[16]

6 Micro Evidence from Sub-Saharan Africa

In this section we take a micro approach that explores across-district variation in ethnic inequality and development within Sub-Saharan African countries. Our focus on Africa is natural. First, Africa is by far the most ethnically diverse part of the world, while ethnic inequality is also quite high. Second, existing studies suggest that a considerable portion of Africa's growth tragedy may be attributed to its ethnic diversity and ethnic patronage politics (Easterly and Levine (1997), Franck and Rainer (2012)). Third, 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)). Fourth, we have high-quality micro-level data on both the ethnic identity and economic conditions that allow us performing a detailed exploration of the role of ethnic inequality on individual well-being. Hence, instead of assigning parts of a country to a single (or more) groups via the use of linguistic maps, we use self-reported data on ethnic identity and living conditions minimizing measurement error.

[16] 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) and geography. 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 power explaining output per worker.

6.1 Data

We use individual-level survey data from the 3rd round of the Afrobarometer surveys, conducted across 17 Sub-Saharan African in 2005.[17] 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. This is the case for $20,984$ out of $25,200$ respondents, who reside in 1301 districts. In each district there are on average 3 ethnic groups (range from 1 to 23 ethnicities). A feature of the data that we will exploit is that individuals from the same ethnic group reside in different districts. We construct inequality measures using individual responses on an ordered $(1-5)$ living conditions index. (1) indicates very bad conditions; (2) fairly bad; (3) neither good nor bad; (4) fairly good, and (5) very good. We calculate Theil indicators reflecting the overall degree of inequality in a district and most importantly the between-group and the within-group component. As dependent variables we use measures of well-being, namely the $1-5$ living condition index and a $0-9$ education index based on years of schooling, and public goods indicators reflecting access to piped water, sewage system, and an electricity grid.

6.2 Results

6.2.1 District-level Analysis

Table 9 reports the results of the within-country analysis that associates district-level living conditions, education, urbanization, and public goods provision to ethnic inequality. Table 1-Panel *B* reports summary statistics, while Appendix Table 9 reports the correlation structure of the main variables. We augment the specification with the within-group Theil index, so as to jointly examine the role of between and within-group inequality. We also control for fractionalization using the log number of ethnicities in a district. In odd-numbered columns we report estimates in the full sample, while in even-numbered columns we exclude from the estimation regions with only one ethnicity.[18]

The specifications in (1) and (2) show that regional development 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) is associated with a 0.31 point increase in the average living conditions in a region (a bit more than half a standard deviation). The coefficient on the within-ethnicity Theil index is

[17] These countries are: Benin, Botswana, Ghana, Kenya, Lesotho, Madagascar, Malawi, Mali, Mozambique, Namibia, Nigeria, Senegal, South Africa, Tanzania, Uganda, Zambia, and Zimbabwe. Nunn and Wantchekon (2011) use a subset of this dataset to assess the role of the slave trades on trust.

[18] The fractionalization index enters with a significantly positive coefficient, capturing most likely a city-effect where both development and mixing are higher. As we show below the negative correlation between ethnic inequality and development and public goods access also pertains across urban households/regions.

also negative and highly significant, suggesting that inequality within groups is also a feature of under-development. This finding is in line with the theory of Esteban and Ray (2011b) that link inequalities within ethnicities to conflict; the main idea being that when within-group inequality is high then the group has both the capital (from the wealthy) and the necessary labor (from the poor) to orchestrate a coup/war. The estimate on the within-group Theil index suggests that a 5% decrease (one standard deviation) is associated with a 0.28 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 (columns (3)-(4)). Regions inhabited by more educated respondents are characterized by a lower degree of both between and within-group disparities. The standardized "beta" coefficients are -0.05 and -0.035, respectively. In specifications (5)-(10) we examine the relationship between inequality and public goods provision. In columns (5) and (6) the dependent variable is the fraction of respondents within a district with access to piped water; in (7) and (8) we look at access to a sewage system, while in (9) and (10) on access to an electricity grid. The ethnic inequality index enters with a negative and significant coefficient in all permutations. The same pattern is found for the within-group inequality that is also negatively associated with the provision of basic public goods.

6.2.2 Individual-level Analysis

In Table 10 - Panel *A* we report specifications associating ethnic inequality at the district level with living conditions, education, and access to public goods at the individual level. This allows us conditioning on numerous individual characteristics (Appendix Table 10 reports the summary statistics). Following Nunn and Wantchekon (2011) we control for the respondent's age and age squared, a gender indicator, 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. Specifications (1)-(4) show that respondents residing in regions marked by high ethnic inequality have a lower quality of living conditions and are less educated. Moreover, the linear probability estimates in (5)-(10) show that respondents in ethnically unequal districts are less likely to have access to piped water, a sewage system, and an electricity gird.

Since members of the same ethnic group are present in more than one district, we also explore whether conditional on ethnic-specific characteristics, inequality across ethnic lines is an important correlate of individual well-being. Overall 265 out of the 328 ethnic groups in our sample may be found in more than one district. Table 10-Panel *B* reports otherwise identical to Panel *A* estimates replacing the country fixed effects with ethnicity-country constants. 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 politi-

cal centralization(Gennaioli and Rainer (2007) and Michalopoulos and Papaioannou (2012)), and ethnic partitioning (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. Specifications (1) and (2) show 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; individuals in ethnically unequal regions are also less educated ((3)-(4)). Specifications (5)-(10) further show that members of the same group found in more ethnically unequal districts are less likely to live in an urban household, are less likely to have access to clean water, a sewage system, and an electricity grid. The economic magnitudes of the estimates on the between-group Theil index (that is always significant at the 99% level) are large. Lowering ethnic inequality by one standard deviation increases the likelihood of household's access to piped water and a sewage system by 4.6 % and the presence of an electricity grid by 3.0 %.

Examples A couple of examples are useful to illustrate our results. The Pular in Senegal are found in 28 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).

6.2.3 Sensitivity Analysis

We perturbed the empirical model in various ways to explore the robustness of these results. In the Supplementary Appendix we report the main sensitivity checks. First, rather than conditioning on the log number of ethnicities we constructed a standard fractionalization measure. As Appendix Table 11 demonstrates this has no effect on our main results. Second, we repeated estimation using the mean log deviation index. Appendix Table 12 reports the results. Given the high correlation of the mean log deviation and the Theil indicators, it comes at no surprise

that the results are virtually unchanged. Third, we repeated the analysis restricting estimation to urban households. This is useful as inequality is higher in urban places and, unlike the cross-country setting, here we can properly account for increased population mixing in urban places. Appendix Table 13 reports the results. 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 ML models finding similar results (Appendix Table 14).

7 Conclusion

This study shows that ethnic differences in economic performance rather than the degree of diversity 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 inter-linkages between ethnicity, inequality, and comparative development. This paper is a first effort to fill this gap.

First, combining linguistic maps on the spatial distribution of groups within countries with satellite images of light density at night we construct Gini coefficients reflecting inequality in well-being (and/or public goods provision) 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. 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 groups 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. In contrast there is no systematic association between ethnic inequality and historical traits that have been linked to contemporary development. Fourth, we show that inequality in geographic endowments across ethnic homelands is inversely related to contemporary development. Thus, while the significant negative correlation between ethnic inequality and development may be driven by omitted variables and/or other forms of endogeneity, it does not seem to be solely

an outcome of reverse causation. Finally, we show a similar negative association between ethnic inequality and development exploring solely within-country across-district variation in 17 Sub-Saharan countries using micro-level data on well-being, public goods provision, and ethnic identification. Exploiting information from more than $20,000$ respondents, our analysis shows that, conditional on an array of individual characteristics and ethnicity fixed effects, respondents from the same ethnic group report worse living conditions, lower levels of formal education, and inadequate access to basic public goods when they reside in districts characterized with a higher degree of ethnic group inequality.

We view our work as a first step towards mapping and understanding the consequences and origins of contemporary differences in income across ethnic groups. Future research should explore the channels via which ethnic inequality and development are linked and investigate the historical, cultural, and politico-economic origins of ethnic inequality. 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 (2000): "Participation in Heterogeneous Communities," *Quarterly Journal of Economics*, 115(2), 847–904.

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

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.

Blattman, C., and E. Miguel (2010): "Civil War," *Journal of Economic Literature*, 48(1), 3–57.

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.

Caselli, F., and W. J. Coleman (2012): "On the Theory of Ethnic Conflict," mimeo London School of Economics.

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

Collier, P., and A. Hoeffler (1998): "On Economic Causes of Civil War," *Oxford Economic Papers*, 50(4), 563–573.

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 (1994): "On the Measurement of Polarization," *Econometrica*, 62(4), 819–851.

——— (1999): "Conflict and Distribution," *Journal of Economic Theory*, 87(1), 379–415.

——— (2011a): "Linking Conflict to Inequality and Polarization," *American Economic Review*, 101(4), 1345–1374.

——— (2011b): "A Model of Ethnic Conflict," *Journal of the European Economic Association*, 9(3), 496–521.

Fearon, J. (2003): "Ethnic Structure and Cultural Diversity by Country," *Journal of Economic Growth*, 8, 195–222.

Fearon, J., and D. Laitin (2003): "Ethnicity, Insurgency and Civil War," *American Political Science Review*, 97, 75–90.

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.

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," .

Guiso, L., P. Sapienza, and L. Zingales (2011): "Civic Capital as the Missing Link," in *Handbook of Social Economics*, ed. by J. Benhabib, A. Bisin, and M. Jackson, vol. 1, pp. 417–480. Elsevier.

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, CA.

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

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): "The Long-Run Effects of the Scramble for Africa," NBER Working Paper No. 17620.

——— (2012): "Pre-colonial Ethnic Institutions and Contemporary African Development," *Econometrica*, forthcoming.

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.

Putnam, R. D. (1993): *Making Democracy Work. Civic Traditions in Modern Italy*. Pinceton University Press, Princeton, NJ.

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.

Tesei, A. (2011): "Racial Fragmentation, Income Inequality and Social Capital Formation: New Evidence from the US," mimeo, UPF.

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.

8 Data Appendix

8.1 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*

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

Living Conditions: Respondent's view of their present living conditions. The question (Q4B) is "*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) is "*What is the highest level of education you have completed?*". The answers are: 0=No formal schooling, 1=Informal 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. *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. *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. *Source: 2005 Afrobarometer Surveys*

Urban Household: Indicator for whether the respondent comes from an urban location (question answered by interviewer). For the district-level analysis responses are averaged across individuals within a district. *Source: 2005 Afrobarometer Surveys.*

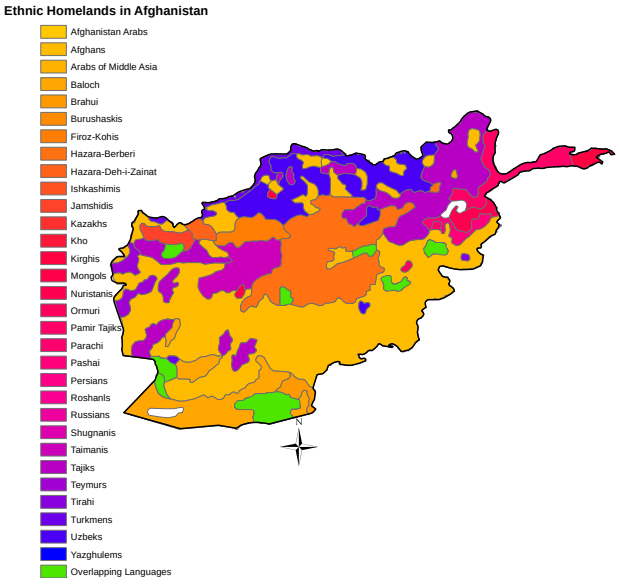

Figure 1*a*: Ethnic Homelands in Afghanistan

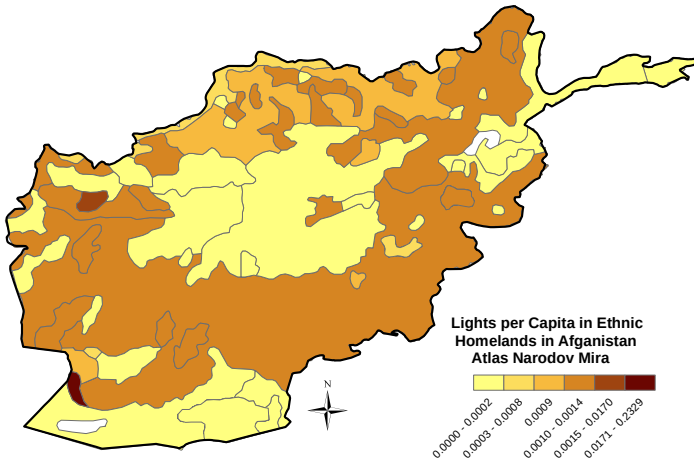

Figure 1*b*: Lights per Capita across Ethnic Homelands

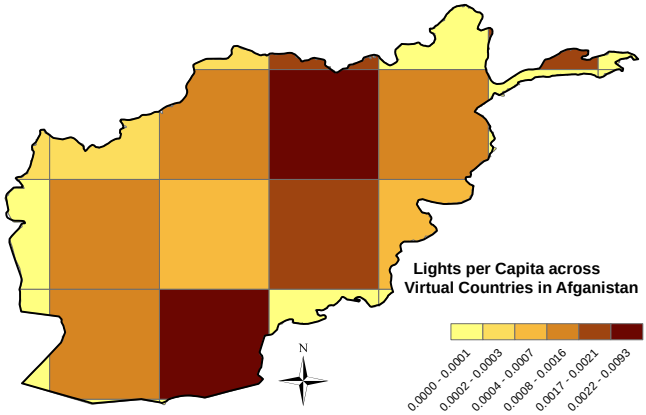

Figure 2*a*: Lights across 2.5 by 2.5 dd Boxes

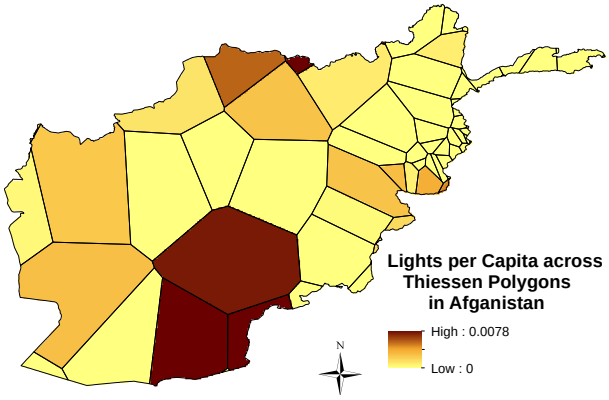

Figure 2*b*: Lights across Thiessen Polygons

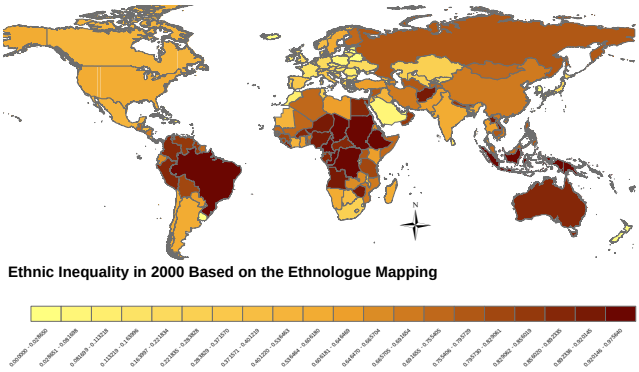

Figure 3*a*

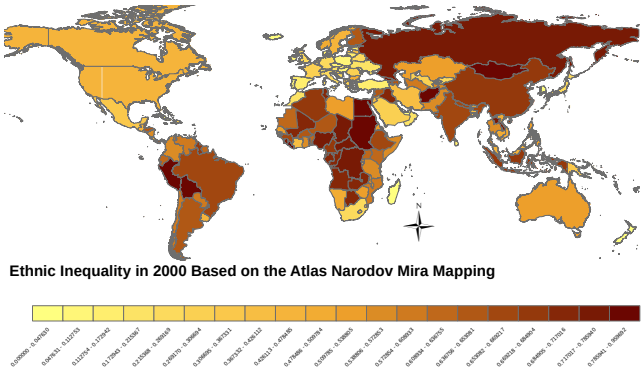

Figure 3*b*

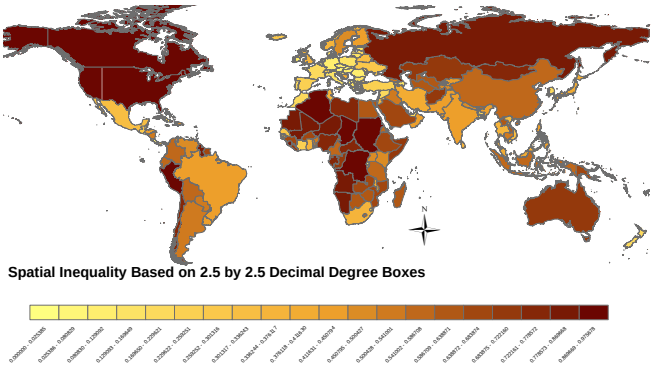

Figure 3*c*

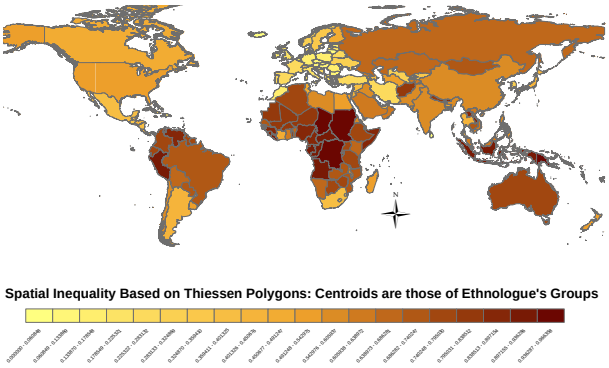

Figure 3*d*

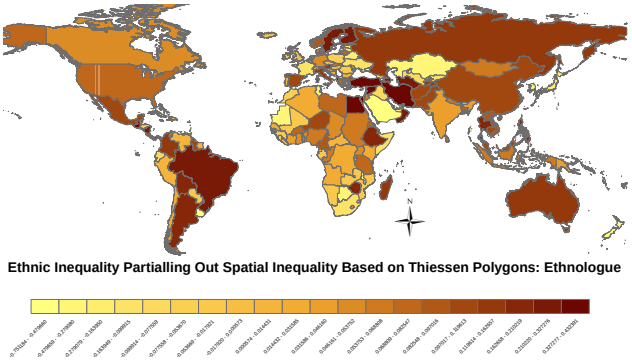

Figure 3*e*

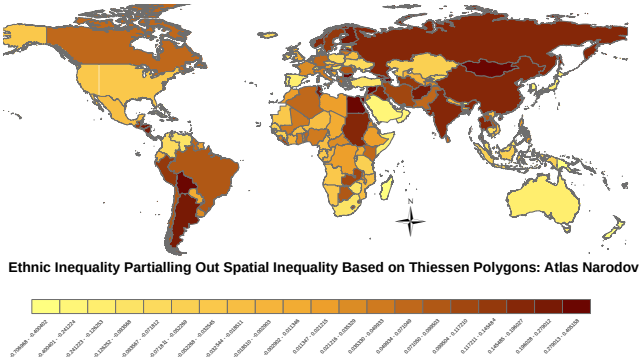

Figure 3*f*

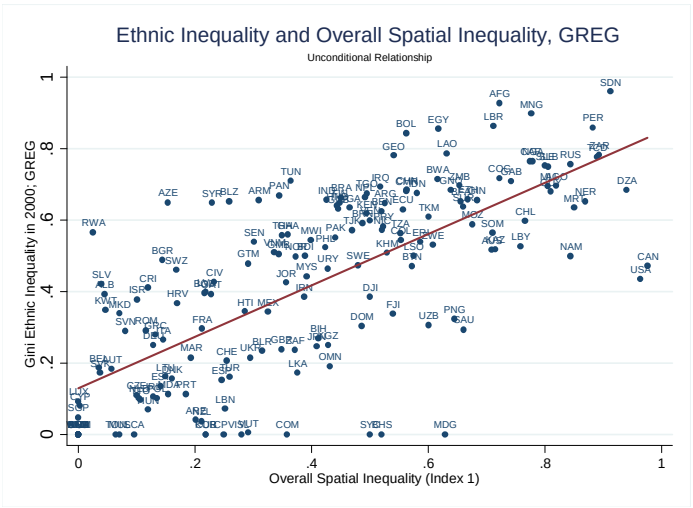

Figure 4a

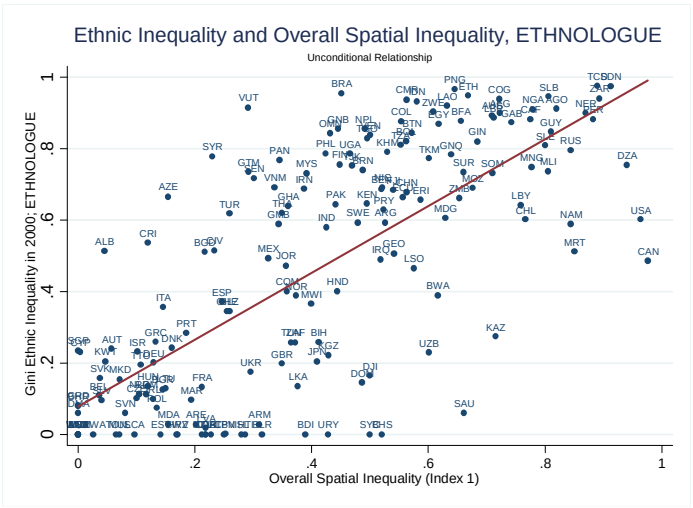

Figure 4b

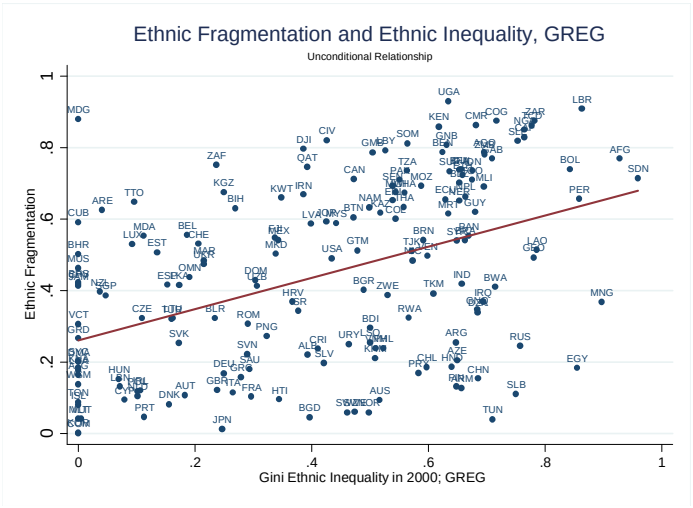

Figure 5a

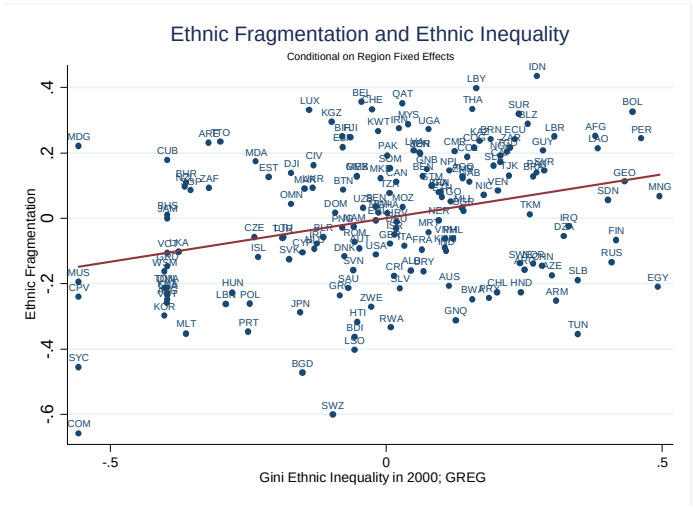

Figure 5b

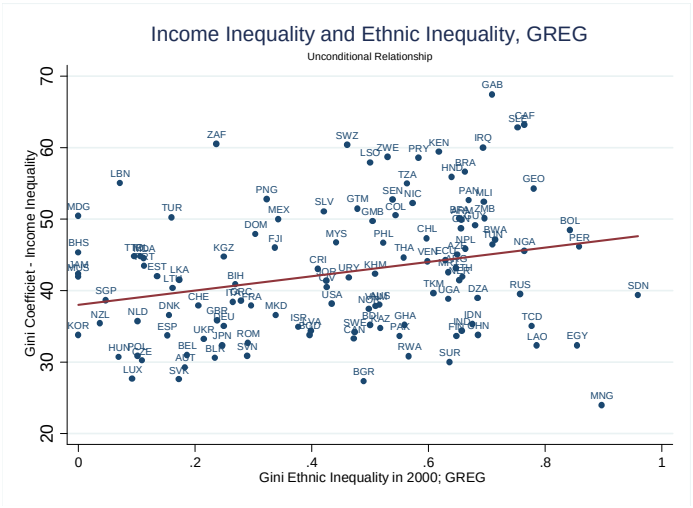

Figure 6a

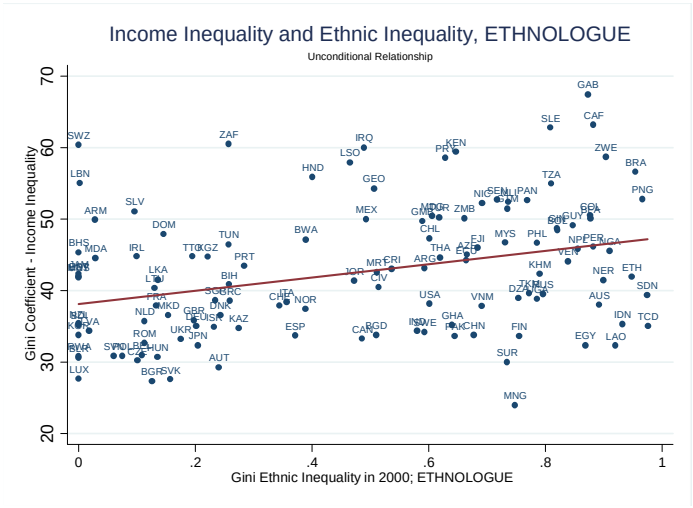

Figure 6b

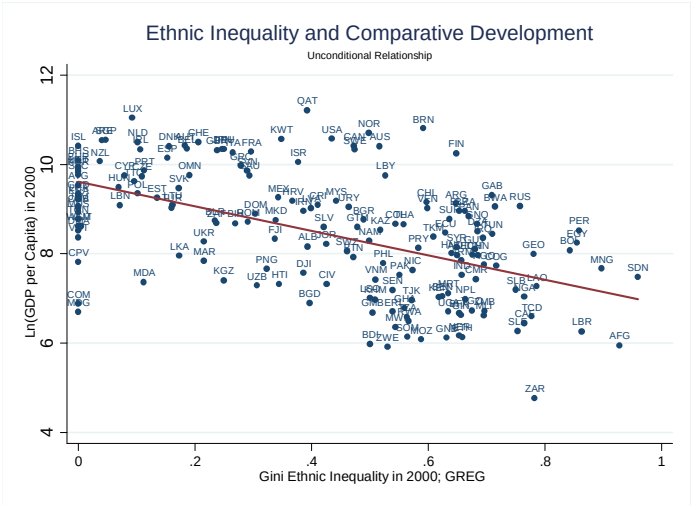

Figure 7a

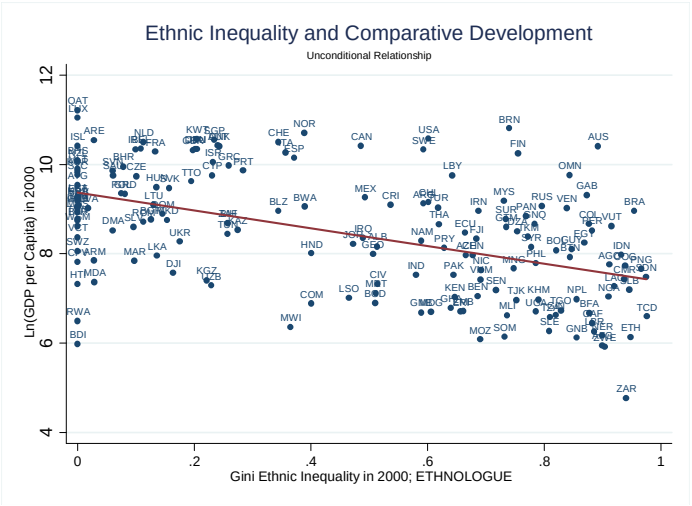

Figure 7b

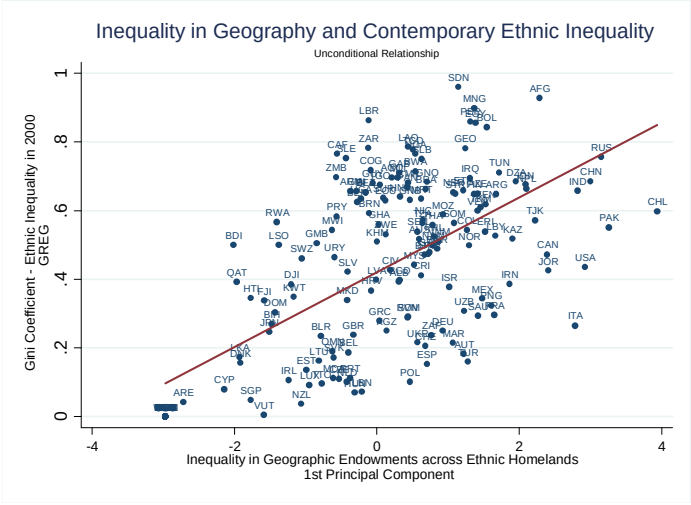

Figure 8a

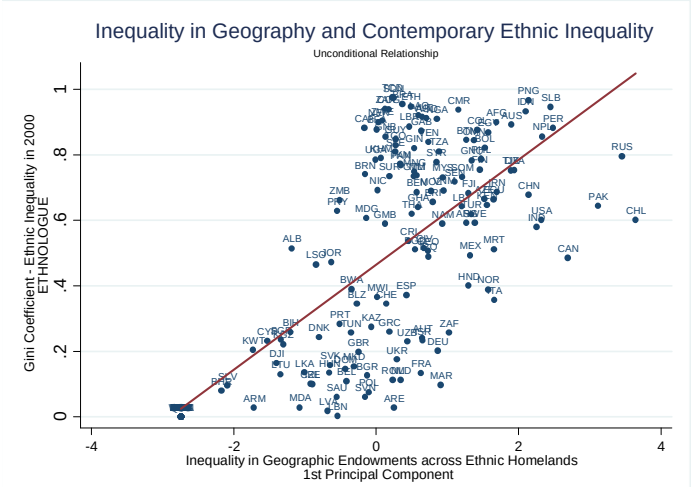

Figure 8b

Table 1A: Summary Statistics - Cross Country Inequality Measures

	Obs.	mean	st. dev.	p25	median	p75	min	max
Number of Ethnicities (GREG)	173	11.520	14.167	1.00	3.00	8.00	13.00	94.00
Ethnic Gini in 2009 (GREG), All Groups	173	0.423	0.259	0.00	0.19	0.47	0.63	0.96
Ethnic Gini in 2000 (GREG), All Groups	173	0.420	0.263	0.00	0.19	0.47	0.65	0.96
Ethnic Gini in 1992 (GREG), All Groups	173	0.476	0.288	0.00	0.21	0.56	0.72	0.97
Number of Languages (ETHNOLOGUE)	173	41.908	99.780	1.00	3.00	9.00	36.00	791.00
Ethnic Gini in 2009 (ETHNOLOGUE), All Groups	173	0.454	0.333	0.00	0.13	0.47	0.77	0.97
Ethnic Gini in 2000 (ETHNOLOGUE), All Groups	173	0.456	0.338	0.00	0.11	0.51	0.77	0.98
Ethnic Gini in 1992 (ETHNOLOGUE), All Groups	173	0.497	0.355	0.00	0.15	0.55	0.83	0.99
Number of Pixels	173	24.283	63.837	1.00	4.00	8.00	22.00	637.00
Spatial Gini in 2009, Pixels	173	0.404	0.262	0.00	0.17	0.40	0.60	0.98
Spatial Gini in 2000, Pixels	173	0.405	0.263	0.00	0.17	0.40	0.60	0.98
Spatial Gini in 1992, Pixels	173	0.453	0.273	0.00	0.21	0.47	0.68	0.95
Number of Thiessen Polygons	173	50.792	98.053	1.00	7.00	17.00	54.00	698.00
Spatial Gini in 2009, Thiessen Polygons	173	0.472	0.290	0.00	0.21	0.48	0.71	0.97
Spatial Gini in 2000, Thiessen Polygons	173	0.475	0.291	0.00	0.23	0.46	0.73	0.97
Spatial Gini in 1992, Thiessen Polygons	173	0.517	0.310	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	1301	0.052	0.053	0.00	0.00	0.04	0.09	0.31
Theil Index - Between-Group Inequality	1301	0.012	0.023	0.00	0.00	0.00	0.02	0.31
Theil Index - Within-Group Inequality	1301	0.040	0.045	0.00	0.00	0.03	0.07	0.21
Mean Log Deviation - Overall Inequality	1301	0.057	0.057	0.00	0.00	0.05	0.10	0.29
Mean Log Deviation - Between-Group Inequality	1301	0.013	0.025	0.00	0.00	0.00	0.02	0.29
Mean Log Deviation - Within-Group Inequality	1301	0.044	0.048	0.00	0.00	0.03	0.08	0.23
Living Conditions Index	1301	2.709	0.665	1.00	2.25	2.75	3.17	4.88
Education	1301	4.017	1.420	1.00	3.00	4.00	5.00	8.75
Access to Sewage System	1280	0.244	0.390	0.00	0.00	0.00	0.50	1.00
Access to Clean Piped Water	1286	0.470	0.450	0.00	0.00	0.44	1.00	1.00
Access to Electricity Grid	1292	0.543	0.454	0.00	0.00	0.60	1.00	1.00
Urbanization	1301	0.320	0.425	0.00	0.00	0.00	0.88	1.00
Number of Ethnic Groups	1301	2.915	2.538	1.00	1.00	2.00	4.00	23.00

Pane 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-region analysis within African countries (Afrobarometer sample). Section 6.1 gives details on the construction of these measures.

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 Analysis Using Data from the Afrobarometer Surveys

	Living Conditions		Education		Piped Water		Sewage System		Electricity Grid		Urbanization	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Between-Group Ineq.	-4.9156***	-6.2329***	-4.8300***	-5.0611***	-2.3742***	-2.4819***	-1.8750***	-1.8256***	-1.6597**	-1.5788**	-1.8920***	-1.9359***
[Theil Index]	(0.7928)	(1.0646)	(1.0006)	(1.1533)	(0.5946)	(0.6418)	(0.4573)	(0.4840)	(0.7693)	(0.7102)	(0.5349)	(0.5335)
	-6.20	*-5.85*	*-4.83*	*-4.39*	*-3.99*	*-3.87*	*-4.10*	*-3.77*	*-2.16*	*-2.22*	*-3.54*	*-3.63*
Within-Group Ineq.	-2.7822***	-5.8355***	-3.3152***	-3.7319***	-0.9359***	-1.0316***	-0.6892	-0.4841	-1.2840***	-1.1552***	-0.6372*	-0.4502
[Theil Index]	(0.9196)	(1.0642)	(0.5246)	(0.9003)	(0.2653)	(0.3009)	(0.4133)	(0.5473)	(0.2040)	(0.2944)	(0.3226)	(0.4504)
	-3.03	*-5.48*	*-6.32*	-4.15	*-3.53*	*-3.43*	*-1.67*	*-0.88*	*-6.29*	*-3.92*	*-1.98*	*-1.00*
Log Number of Groups	0.2881***	0.0740**	0.6018***	0.5680***	0.2008***	0.1715***	0.1697***	0.1551***	0.2412***	0.2291***	0.2756***	0.2935***
	(0.0802)	(0.0321)	(0.0741)	(0.0982)	(0.0248)	(0.0282)	(0.0462)	(0.0448)	(0.0460)	(0.0568)	(0.0252)	(0.0320)
	3.59	*2.31*	*8.13*	5.78	*8.11*	*6.09*	*3.67*	*3.47*	*5.25*	*4.03*	*10.93*	*9.17*
adjusted R-squared	0.359	0.517	0.509	0.552	0.215	0.243	0.216	0.220	0.314	0.304	0.175	0.186
Observations	1301	861	1301	861	1286	856	1280	852	1292	858	1301	861
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	All	>1 groups	All	>1 groups	All	>1 groups	All	>1 groups	All	>1 groups	All	>1 groups

The table reports cross-regional OLS estimates, associating various proxy measures of economic development and public goods provision with living conditions inequality between and within ethnic groups, as reflected in the Theil index. The dependent variable is columns (1) and (2) is the average value of a composite living conditions index; in columns (3) and (4) is the average value of an education index; in columns (5) and (6) is the percentage of households in a region with access to clean piped water; in columns (7) and (8) is the percentage of households in a region with access to a sewage system; in columns (9) and (10) is the percentage of households in a region with access to an electricity grid; and in columns (11) and (12) is the percentage of urban households in a region.

The between-ethnic-group and the within-ethnic-group Theil indicators are based on individuals' responses on living conditions. In all specifications we control for the log number of ethnic groups in each region. All specifications include country fixed effects (constants not reported). Odd-numbered columns report estimates in the full sample. In even-numbered columns we exclude from the estimation regions with respondents from just one ethnic group. The Data Appendix gives detailed variable definitions and data sources. All variables are constructed using data from the 3rd round of the Afrobarometer Surveys. Standard errors reported in parentheses below the estimates are clustered at the country level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Table 10a - Ethnic Inequality and Regional Development within African Countries
Individual-Level Analysis Using Data from the Afrobarometer Surveys

	Living Conditions		Education		Piped Water		Sewage System		Electricity Grid		Urban Household	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Between Group Ineq.	-4.4921***	-5.6231***	-5.5814***	-5.3468***	-2.6112***	-2.6819***	-2.6151***	-2.4384***	-1.8300***	-1.7784***	-3.2946***	-3.1542***
[Theil Index]	(0.7708)	(0.8581)	(1.2431)	(1.2523)	(0.7195)	(0.6987)	(0.5697)	(0.5482)	(0.6578)	(0.6360)	(0.7652)	(0.7541)
	-5.83	*-6.55*	*-4.49*	*-4.27*	*-3.63*	*-3.84*	*-4.59*	*-4.45*	*-2.78*	*-2.80*	*-4.31*	*-4.18*
Within Group Ineq.	-3.0798***	-5.7650***	-1.6567***	-1.5315**	-0.7725**	-0.6488*	-0.3155	-0.0153	-0.7908***	-0.6465**	-0.2343	0.0835
[Theil Index]	(0.7793)	(0.8610)	(0.5612)	(0.6343)	(0.3009)	(0.3599)	(0.3198)	(0.3837)	(0.2730)	(0.3087)	(0.3459)	(0.3966)
	-3.95	*-6.7*	*-2.95*	*-2.41*	*-2.57*	*-1.8*	*-0.99*	*-0.04*	*-2.90*	*-2.09*	*-0.68*	*0.21*
Log Number of	0.1259***	0.0339	0.3235***	0.3281***	0.1698***	0.1648***	0.1385***	0.1360***	0.1895***	0.1800***	0.2573***	0.2708***
Groups	(0.0339)	(0.0369)	(0.0452)	(0.0532)	(0.0267)	(0.0312)	(0.0274)	(0.0272)	(0.0262)	(0.0275)	(0.0257)	(0.0288)
	3.72	*0.92*	*7.16*	6.17	*6.36*	*5.28*	*5.06*	*5.00*	*7.24*	*6.55*	*10.02*	*9.41*
adjusted R-squared	0.157	0.171	0.472	0.467	0.283	0.301	0.26	0.279	0.321	0.316	0.312	0.323
Observations	20984	17254	20984	17254	20656	17041	20332	16713	20724	17100	20984	17254
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	All	>1 groups	All	>1 groups	All	>1 groups	All	>1 groups	All	>1 groups	All	>1 groups

The table reports OLS estimates, associating various proxy measures of individual well-being (living conditions, education, urban) and household's access on basic public goods (piped water, seage system, electricity grid) with living conditions inequality between and within ethnic groups at the region level, as reflected in the Theil index. The dependent variable is columns (1) and (2) is a 1-5 living conditions index; in columns (3) and (4) is a 1-10 education index; in columns (5) and (6) is an indicator (dummy variable) reflecting household's access to clean piped water; in columns (7) and (8) is an indicator (dummy variable) reflecting household's access to a sewage system; in columns (9) and (10) is an indicator (dummy variable) reflecting household's access to an electricity grid; and in columns (11) and (12) is an indicator (dummy variable) for urban households.
The between-ethnic-group and the within-ethnic-group Theil indicators are based on individuals' responses on living conditions. In all specifications we control for the log number of ethnic groups in each region and the share of the district's population that is the same ethnicity as the respondent. The individual-level controls are for 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 regions with respondents from just one ethnic group. All specifications include 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. Double-clustered standard errors at the ethnicity and district level are reported in parentheses below the estimates. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Table 10b - Ethnic Inequality and Regional Development within African Countries and within African Ethnicities
Individual-Level Analysis Using Data from the Afrobarometer Surveys

	Living Conditions		Education		Piped Water		Sewage System		Electricity Grid		Urban Household	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Between-Group Ineq.	-4.0126***	-5.2839***	-4.9041***	-4.6823***	-2.5521***	-2.6067***	-2.4418***	-2.2573***	-1.8425***	-1.8005***	-2.9829***	-3.1542***
[Theil Index]	(0.8018)	(0.9068)	(1.1194)	(1.1137)	(0.6908)	(0.6803)	(0.5786)	(0.5542)	(0.6460)	(0.6170)	(0.7162)	(0.7541)
	-5.00	*-5.83*	*-4.38*	*-4.20*	*-3.69*	*-3.83*	*-4.22*	*-4.07*	*-2.85*	*-2.92*	*-4.16*	*-4.18*
Within-Group Ineq.	-1.7938***	-4.5017***	-1.6604***	-2.0550***	-0.7330**	-0.7601**	-0.4206	-0.1260	-0.9181***	-0.9159***	-0.3754	0.0835
[Theil Index]	(0.5400)	(0.7180)	(0.6353)	(0.7030)	(0.2900)	(0.3645)	(0.3374)	(0.4105)	(0.2644)	(0.3002)	(0.3256)	(0.3966)
	-3.32	*-6.27*	*-2.61*	-2.92	*-2.53*	*-2.09*	*-1.25*	*-0.31*	*-3.47*	*-3.05*	*-1.15*	*0.21*
Log Number of Groups	0.0856***	0.0033	0.3164***	0.3099***	0.1625***	0.1486***	0.1423***	0.1352***	0.1859***	0.1637***	0.2646***	0.2708***
	(0.0301)	(0.0332)	(0.0487)	(0.0553)	(0.0242)	(0.0276)	(0.0292)	(0.0290)	(0.0265)	(0.0270)	(0.0242)	(0.0288)
	2.84	*0.1*	*6.49*	5.61	*6.7*	*5.38*	*4.88*	*4.66*	*7.02*	*6.07*	*10.94*	*9.41*
adjusted R-squared	0.202	0.211	0.497	0.493	0.353	0.374	0.31	0.327	0.383	0.378	0.366	0.323
Observations	20984	17254	20984	17254	20656	17041	20332	16713	20724	17100	20984	17254
Country-Ethnicity FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	All	>1 groups	All	>1 groups	All	>1 groups	All	>1 groups	All	>1 groups	All	>1 groups

The table reports OLS estimates, associating various proxy measures of individual well-being (living conditions, education, urban) and household's access on basic public goods (piped water, seage system, electricity grid) with living conditions inequality between and within ethnic groups at the region level, as reflected in the Theil index. The dependent variable is columns (1) and (2) is a 1-5 living conditions index; in columns (3) and (4) is a 1-10 education index; in columns (5) and (6) is an indicator (dummy variable) reflecting household's access to clean piped water; in columns (7) and (8) is an indicator (dummy variable) reflecting household's access to a sewage system; in columns (9) and (10) is an indicator (dummy variable) reflecting household's access to an electricity grid; and in columns (11) and (12) is an indicator (dummy variable) for urban households.
The between-ethnic-group and the within-ethnic-group Theil indicators are based on individuals' responses on living conditions. In all specifications we control for the log number of ethnic groups in each region and the share of the district's population that is the same ethnicity as the respondent. The individual-level controls are for 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 regions with respondents from just one ethnic group. All specifications 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. Double-clustered standard errors at the ethnicity and district level 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.

Appndix 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	
Urbanization	0.0571*	0.0430	0.0496	0.0850*	0.0722*	0.0576*	0.1935*	0.3807*	0.5098*	0.4407*	0.5154*	1
Log Number of Ethnic Groups	0.5880*	0.5097*	0.3814*	0.6011*	0.5259*	0.3777*	0.0530	0.1444*	0.1079*	0.1565*	0.1147*	0.2344*

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.

Appendix Table 10: Summary Statistics - Afrobarometer Sample - Individual Level

	Obs.	mean	st. dev.	p25	median	p75	min	max
Theil Index - Overall Inequality at the District Level	21526	0.074	0.053	0.00	0.04	0.07	0.11	0.31
Theil Index - Between-Group Inequality	21526	0.012	0.016	0.00	0.00	0.01	0.02	0.31
Theil Index - Within-Group Inequality	21526	0.062	0.048	0.00	0.03	0.06	0.09	0.21
Mean Log Deviation - Overall Inequality	21526	0.080	0.055	0.00	0.04	0.08	0.12	0.29
Mean Log Deviation - Between-Group Inequality	21526	0.012	0.018	0.00	0.00	0.01	0.02	0.29
Mean Log Deviation - Within-Group Inequality	21526	0.068	0.049	0.00	0.03	0.07	0.10	0.23
Living Conditions Index (1-5)	21526	2.635	1.200	1.00	2.00	3.00	4.00	5.00
Education (1-10)	21526	4.098	2.021	1.00	3.00	4.00	5.00	10.00
Access to Sewage System (0/1)	20859	0.244	0.430	0.00	0.00	0.00	0.00	1.00
Access to Clean Piped Water (0/1)	21195	0.507	0.500	0.00	0.00	1.00	1.00	1.00
Access to Electricity Grid (0/1)	21263	0.546	0.498	0.00	0.00	1.00	1.00	1.00
Urban Household (0/1)	21526	0.384	0.486	0.00	0.00	0.00	1.00	1.00
Number of Ethnic Groups in a District	21526	5.375	4.663	1.00	2.00	4.00	7.00	23.00
Share of District's Population of the Same Ethnicity	21508	0.586	0.350	0.00	0.23	0.67	0.93	1.00

The table reports summary statistics for the main measures, employed in the individual-level specifications examining the association between ethnic inequality and various measures of development within African countries (Afrobarometer sample). Section 6.1 gives details on the construction of these measures.

Appendix Table 11 - Ethnic Inequality and Regional Development within African Countries and within African Ethnicities
Individual-Level Analysis Using Data from the Afrobarometer Surveys
Sensitivity Analysis: Conditioning on Ethnic Fractionalization at the Regional Level

	Living Conditions		Education		Piped Water		Sewage System		Electricity Grid		Urban Household	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Between-Group Ineq.	-4.4628***	-5.3674***	-4.9939***	-5.0035***	-2.7037***	-2.8161***	-2.6550***	-2.5419***	-1.8820***	-1.9331***	-3.2485***	-3.2184***
[Mean Log Deviation]	(0.8585)	(0.9078)	(1.2149)	(1.2037)	(0.7237)	(0.7136)	(0.6030)	(0.5796)	(0.6518)	(0.6309)	(0.7378)	(0.7227)
	-5.20	*-5.91*	*-4.11*	*-4.16*	*-3.74*	*-3.95*	*-4.40*	*-4.39*	*-2.89*	*-3.06*	*-4.40*	*-4.45*
Within-Group Ineq.	-1.6345***	-4.4571***	-0.6175	-1.7498**	-0.2258	-0.5902	-0.0036	0.0763	-0.3037	-0.7678**	0.4418	0.0824
[Mean Log Deviation]	(0.5342)	(0.7442)	(0.7032)	(0.7596)	(0.2842)	(0.3676)	(0.3640)	(0.4146)	(0.3033)	(0.3156)	(0.3588)	(0.4063)
	-3.06	*-5.99*	*-0.88*	-2.3	*-0.79*	*-1.61*	*-0.01*	*0.18*	*-1.00*	*-2.43*	*1.23*	*0.20*
Ethnic Fractionalization	0.3213***	0.0571	0.5526***	0.4652***	0.3224***	0.2524***	0.3170***	0.2949***	0.3207***	0.2242***	0.5390***	0.4907***
	(0.0888)	(0.0996)	(0.1414)	(0.1586)	(0.0591)	(0.0627)	(0.0849)	(0.0849)	(0.0692)	(0.0659)	(0.0727)	(0.0813)
	3.62	*0.57*	*3.91*	2.93	*5.46*	*4.02*	*3.74*	*3.47*	*4.63*	*3.4*	*7.42*	*6.04*
adjusted R-squared	0.203	0.211	0.494	0.49	0.343	0.364	0.302	0.32	0.368	0.364	0.339	0.351
Observations	20984	17254	20984	17254	20656	17041	20332	16713	20724	17100	20984	17254
Country-Ethnic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	All	>1 groups	All	>1 groups	All	>1 groups	All	>1 groups	All	>1 groups	All	>1 groups

The table reports OLS estimates, associating various proxy measures of individual well-being (living conditions, education, urban) and household's access to basic public goods (piped water, sewage system, electricity grid) with living conditions inequality between and within ethnic groups at the region level, as reflected in the Theil index. In all specifications we control for an (Herfindal-Hirschman) ethnic fractionalization index at the regional level that reflects the likelihood that two randomly chosen individuals in a region will not be members of the same ethnic group. The dependent variable is columns (1) and (2) is a 1-5 living conditions index; in columns (3) and (4) is a 1-10 education index; in columns (5) and (6) is an indicator (dummy variable) reflecting household's access to clean piped water; in columns (7) and (8) is an indicator (dummy variable) reflecting household's access to a sewage system; in columns (9) and (10) is an indicator (dummy variable) reflecting household's access to an electricity grid; and in columns (11) and (12) is an indicator (dummy variable) for urban households.

The between-ethnic-group and the within-ethnic-group Theil indicators are based on individuals' responses on living conditions. In all specifications we control for the log number of ethnic groups in each region and the share of the district's population that is the same ethnicity as the respondent. The individual-level controls are for 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 regions with respondents from just one ethnic group. All specifications 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. Double-clustered standard errors at the ethnicity and district level are reported in parentheses below the estimates. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level,

Appendix Table 12 - Ethnic Inequality and Regional Development within African Countries and within African Ethnicities
Individual-Level Analysis Using Data from the Afrobarometer Surveys
Sensitivity Analysis: Measuring Between and Within Ethnic Group Inequality with the Mean Log Deviation

	Living Conditions		Education		Piped Water		Sewage System		Electricity Grid		Urban Household	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Between-Group Ineq.	-3.3611***	-4.3029***	-4.1434***	-3.8250***	-2.2331***	-2.2566***	-2.2366***	-2.0582***	-1.7579***	-1.6877***	-2.6927***	-2.4762***
[Mean Log Deviation]	(0.7014)	(0.7996)	(1.0475)	(1.0461)	(0.6434)	(0.6235)	(0.5322)	(0.4986)	(0.5789)	(0.5436)	(0.6538)	(0.6411)
	-4.79	*-5.38*	*-3.96*	*-3.66*	*-3.47*	*-3.62*	*-4.20*	*-4.13*	*-3.04*	*-3.10*	*-4.12*	*-3.86*
Within-Group Ineq.	-0.9775**	-2.8864***	-1.2753**	-1.3643**	-0.6284**	-0.5886*	-0.3759	-0.0675	-0.7782***	-0.7144**	-0.1969	-0.002
[Mean Log Deviation]	(0.4898)	(0.6314)	(0.5848)	(0.6558)	(0.2617)	(0.3158)	(0.3012)	(0.3462)	(0.2399)	(0.2779)	(0.3235)	(0.3804)
	-2.00	*-4.57*	*-2.18*	-2.08	*-2.4*	*-1.86*	*-1.25*	*-0.20*	*-3.24*	*-2.57*	*-0.61*	*-0.01*
Log Number of Group Ethnicities	0.0687**	0.0100	0.3106***	0.3124***	0.1620***	0.1497***	0.1428***	0.1354***	0.1861***	0.1654***	0.2609***	0.2707***
	(0.0313)	(0.0342)	(0.0501)	(0.0561)	(0.0247)	(0.0279)	(0.0293)	(0.0289)	(0.0267)	(0.0272)	(0.0250)	(0.0271)
	2.20	*0.29*	*6.21*	5.57	*6.57*	*5.36*	*4.87*	*4.68*	*6.96*	*6.09*	*10.44*	*10.00*
adjusted R-squared	0.201	0.206	0.496	0.492	0.352	0.372	0.31	0.327	0.383	0.377	0.366	0.379
Observations	20984	17254	20984	17254	20656	17041	20332	16713	20724	17100	20984	17254
Country-Ethnic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	All	>1 groups	All	>1 groups	All	>1 groups	All	>1 groups	All	>1 groups	All	>1 groups

The table reports OLS estimates, associating various proxy measures of individual well-being (living conditions, education, urban) and household's access on basic public goods (piped water, sewage system, electricity grid) with living conditions inequality between and within ethnic groups at the region level, as reflected in the mean log deviation index. The dependent variable is columns (1) and (2) is a 1-5 living conditions index; in columns (3) and (4) is a 1-10 education index; in columns (5) and (6) is an indicator (dummy variable) reflecting household's access to clean piped water; in columns (7) and (8) is an indicator (dummy variable) reflecting household's access to a sewage system; in columns (9) and (10) is an indicator (dummy variable) reflecting household's access to an electricity grid; and in columns (11) and (12) is an indicator (dummy variable) for urban households.
The between-ethnic-group and the within-ethnic-group mean log deviation indicators are based on individuals' responses on living conditions. In all specifications we control for the log number of ethnic groups in each region and the share of the district's population that is the same ethnicity as the respondent. The individual-level controls are for 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 regions with respondents from just one ethnic group. All specifications 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. Double-clustered standard errors at the ethnicity and district level are reported in parentheses below the estimates. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Appendix Table 13 - Ethnic Inequality and Regional Development within African Countries and within African Ethnicities
Sensitivity Analysis: Restricting Estimation to Urban Households

	Living Conditions		Education		Piped Water		Sewage System		Electricity Grid	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Between-Group Inequality	-3.3611***	-4.3029***	-4.1434***	-3.8250***	-2.2331***	-2.2566***	-2.2366***	-2.0582***	-1.7579***	-1.6877***
[Theil Index]	(0.7014)	(0.7996)	(1.0475)	(1.0461)	(0.6434)	(0.6235)	(0.5322)	(0.4986)	(0.5789)	(0.5436)
	-4.79	*-5.38*	*-3.96*	*-3.66*	*-3.47*	*-3.62*	*-4.20*	*-4.13*	*-3.04*	*-3.10*
Within-Group Inequality	-0.9775**	-2.8864***	-1.2753**	-1.3643**	-0.6284**	-0.5886*	-0.3759	-0.0675	-0.7782***	-0.7144**
[Theil Index]	(0.4898)	(0.6314)	(0.5848)	(0.6558)	(0.2617)	(0.3158)	(0.3012)	(0.3462)	(0.2399)	(0.2779)
	-2.00	*-4.57*	*-2.18*	-2.08	*-2.4*	*-1.86*	*-1.25*	*-0.20*	*-3.24*	*-2.57*
Log Number of Ethnicities	0.0687**	0.0100	0.3106***	0.3124***	0.1620***	0.1497***	0.1428***	0.1354***	0.1861***	0.1654***
	(0.0313)	(0.0342)	(0.0501)	(0.0561)	(0.0247)	(0.0279)	(0.0293)	(0.0289)	(0.0267)	(0.0272)
	2.20	*0.29*	*6.21*	5.57	*6.57*	*5.36*	*4.87*	*4.68*	*6.96*	*6.09*
adjusted R-squared	0.201	0.206	0.496	0.492	0.352	0.372	0.31	0.327	0.383	0.377
Observations	20984	17254	20984	17254	20656	17041	20332	16713	20724	17100
Country-Ethnicity FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	All	>1 groups	All	>1 groups	All	>1 groups	All	>1 groups	All	>1 groups

The table reports OLS estimates, associating various proxy measures of individual well-being (living conditions, education, urban) and household's access on basic public goods (piped water, sewage system, electricity grid) with living conditions inequality between and within ethnic groups at the region level, as reflected in the Theil index. Estimation is restricted to urban households. The dependent variable is columns (1) and (2) is a 1-5 living conditions index; in columns (3) and (4) is a 1-10 education index; in columns (5) and (6) is an indicator (dummy variable) reflecting household's access to clean piped water; in columns (7) and (8) is an indicator (dummy variable) reflecting household's access to a sewage system; in columns (9) and (10) is an indicator (dummy variable) reflecting household's access to an electricity grid; and in columns (11) and (12) is an indicator (dummy variable) for urban households.

The between-ethnic-group and the within-ethnic-group Theil indicators are based on individuals' responses on living conditions. In all specifications we control for the log number of ethnic groups in each region and the share of the district's population that is the same ethnicity as the respondent. The individual-level controls are for 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 regions with respondents from just one ethnic group. All specifications 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. Double-clustered standard errors at the ethnicity and district level are reported in parentheses below the estimates. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Appendix Table 14: Ethnic Inequality and Regional Development within African Countries and within African Ethnicities
Individual-Level Analysis Using Data from the Afrobarometer Surveys
Sensitivity Analysis: Ordered Probit Estimates

	Living Conditions Index				Education/Schooling			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Between Ethnic Group Inequality	-4.2681***	-5.4589***	-4.0019***	-5.3222***	-3.9535***	-3.7997***	-3.5671***	-3.4350***
[Theil Index]	(0.7398)	(0.7810)	(0.7395)	(0.7917)	(0.8914)	(0.9074)	(0.7919)	(0.7999)
	-5.77	*-6.99*	*-5.41*	*-6.72*	*-4.43*	*-4.19*	*-4.50*	*-4.29*
Within Ethnic Group Inequality	-3.0422***	-5.7637***	-1.8210***	-4.6217***	-1.1349***	-1.0831**	-1.2022***	-1.5351***
[Theil Index]	(0.5006)	(0.5949)	(0.3376)	(0.5016)	(0.3970)	(0.4773)	(0.4096)	(0.4848)
	-6.08	*-9.69*	*-5.39*	*-9.21*	*-2.86*	*-2.27*	*-2.94*	*-3.17*
Log Number of Ethnic Groups	0.1251***	0.0358	0.0884***	0.0067	0.2302***	0.2345***	0.2309***	0.2279***
	(0.0290)	(0.0323)	(0.0241)	*(0.0281)*	*(0.0287)*	*(0.0343)*	(0.0278)	(0.0330)
	4.31	*1.11*	*3.67*	*0.24*	*8.03*	*6.84*	*8.29*	*6.91*
Log Likelihood	-29,600	-24,100	-29,000	-23,700	-35,000	-28,900	-34,500	-28,400
observations	20984	17254	20984	17254	20984	17254	20984	17254
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed-Effects	Yes	Yes	No	No	Yes	Yes	No	No
Country/Ethnicity Fixed-Effects	No	No	Yes	Yes	No	No	Yes	Yes

The table reports ordered probit estimates (coefficients), associating individual's living conditions and education with inequality between and within ethnic groups at the region level, as reflected in the Theil index. The dependent variable is columns (1)-(4) is a 1-5 living conditions index and in columns (5)-(8) a 1-10 education index. The between-ethnic-group and the within-ethnic-group Theil indicators are based on individuals' responses on living conditions. In all specifications we control for the log number of ethnic groups in each region and the share of the district's population that is the same ethnicity as the respondent. The individual-level controls are for 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 regions with respondents from just one ethnic group. The specifications in columns (1), (2), (5), and (6) include country fixed effects (constants not reported). The specifications in columns (3), (4), (7), and (8) 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. Double-clustered standard errors at the ethnicity and district level are reported in parentheses below the estimates. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.