Ethnic Inequality

Alberto Alesina

Harvard University, Innocenzo Gasparini Institute for Economic Research, Centre for Economic Policy Research, and National Bureau of Economic Research

Stelios Michalopoulos

Harvard University, Brown University, Centre for Economic Policy Research, and National Bureau of Economic Research

Elias Papaioannou

London Business School, Centre for Economic Policy Research, and National Bureau of Economic Research

This study explores the consequences and origins of between-ethnicity inequality for a large sample of countries. First, combining satellite images of nighttime luminosity with the homelands of ethnolinguistic groups, we construct measures of ethnic inequality. Second, we uncover a strong inverse association between ethnic inequality and contemporary development above and beyond its relationship with cross-region and cross–administrative unit inequality. Third, we establish that differences in geographic endowments across ethnic homelands explain a sizable fraction of the variation in economic disparities across groups. Fourth, we show that inequality in geographic endowments across ethnic homelands is a negative correlate of development.

I. Introduction

Ethnic diversity has costs and benefits. On the one hand, diversity in skills, education, and endowments can enhance productivity by promoting innovation. On the other hand, diversity is often associated with poor and ethnically targeted policies, inefficient provision of public goods, and ethnic-based hatred and conflict. In fact, a large literature finds a negative impact of ethnolinguistic fragmentation on various aspects of economic performance, with the possible exception of wealthy economies (see Alesina and La Ferrara [2005] for a review). Income inequality may also have both positive and negative effects on development. On the negative side, a higher degree of income inequality may lead to conflict and crime, prevent the poor from acquiring education, or lead to expropriation and lofty taxation discouraging investment. On the positive side, income inequality may spur innovation and entrepreneurship by motivating individuals and by providing the necessary pools of capital for capital-intensive modes of production. Further complicating the relationship between the two, a positive correlation between inequality and development may reflect Simon Kuznetz's (1955) conjecture that industrialization translates into higher levels of inequality at the early stages of development, while at later stages, the association becomes negative. Given the theoretical ambiguities (and data issues), perhaps it comes as no surprise that it has been very hard to detect empirically a robust association between inequality and development (see Benabou [2005] and Galor [2011] for surveys).

    This paper puts forward and tests an alternative conjecture that focuses on the intersection of ethnic diversity and inequality. Our thesis is that what matter most for comparative development are economic differences between ethnic groups coexisting in the same country rather than the degree of fractionalization per se or income inequality conventionally measured (i.e., independent of ethnicity).[1]

    The first contribution of this study is to provide measures of within-country differences in well-being across ethnic groups, defined as "ethnic inequality." To overcome the sparsity of income data along ethnic lines and

    [1] Stewart (2002) and Chua (2003) are early precedents. Providing case study evidence, they argue that horizontal inequalities across ethnic/religious/racial groups are important features of underdeveloped and conflict-prone societies. Yet to the best of our knowledge, there have been very few systematic empirical works, if any, that directly examine this conjecture. We discuss parallel studies that touch on this issue below.

in order to construct country-level indicators of ethnic inequality for the largest possible set of states, we combine ethnographic and linguistic maps on the location of groups with satellite images of light density at night, which are available at a fine grid. Recent studies have shown that luminosity is a strong proxy of development (e.g., 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 and is modestly correlated with ethnic fractionalization. To isolate the cross-ethnic component of inequality from the overall regional inequality, we also construct proxies of spatial inequality and measures capturing regional differences in well-being across first- and second-level administrative units.

Second, we document a strong negative association between ethnic inequality and real GDP per capita across countries. This correlation holds even when we control for the overall degree of spatial inequality and inequality across administrative regions. The latter is also inversely related to a country's economic performance, a novel finding in itself. We also uncover that the negative correlation between ethnolinguistic fragmentation and development weakens considerably (and becomes statistically indistinguishable from zero) when we account for ethnic inequality; this suggests that it is the unequal concentration of wealth across ethnic lines that correlates with development rather than diversity per se.

Third, in an effort to shed light on the roots of ethnic inequality, we explore its geographic underpinnings. In particular, motivated by recent work showing that linguistic groups tend to reside in distinct land endowments (see Michalopoulos 2012), we construct Gini coefficients reflecting differences in various geographic attributes across ethnic homelands and show that the latter is a strong predictor of ethnic inequality. On the contrary, there is no link between contemporary ethnic inequality and often-used historical variables capturing the type of colonization and legal origin, among others.

Fourth, we show that contemporary development at the country level is also inversely related to inequality in geographic endowments across ethnic homelands. Yet, once we condition on between-group income inequality, differences in geographic endowments are no longer a significant correlate of underdevelopment. These results suggest that geographic differences across ethnic homelands influence comparative development mostly via shaping economic inequality across groups.

Mechanisms and related works.—Income disparities along ethnic lines are likely to lead to political inequality based on ethnic affiliation, increase between-group animosity, and lead to discriminatory policies of one (or more) group against the others. In line with this idea, in recent work Huber and Suryanarayan (2013) document that party ethnification in India is more pronounced in states with a high degree of inequality

across subcastes.[2] Furthermore, differences in preferences along both ethnic and income lines may lead to inadequate public goods provision, as groups' ideal allocations of resources will be quite distant. Baldwin and Huber (2010) provide empirical evidence linking between-group inequality to the underprovision of public goods for 46 democracies. In Alesina, Michalopoulos, and Papaioannou (2014), we show that there is a strong inverse link between ethnic inequality and public goods within 18 sub-Saharan African countries (and that this effect partly stems from political inequality and ethnic-based discrimination).[3] Ethnic inequality may also impede institutional development and the consolidation of democracy (Robinson 2001). In line with this conjecture, Kyriacou (2013) exploits survey data from 29 developing countries and shows that socioeconomic ethnic group inequalities reduce government quality.

Chua (2003) presents case study evidence arguing that the presence of an economically dominant ethnic minority may lower support for democracy and free-market institutions, as the majority of the population usually feels that the benefits of capitalism go to just a handful of ethnic groups. She discusses, among others, the influence of Chinese minorities in the Philippines, Indonesia, Malaysia, and other eastern Asian countries; the dominant role of (small) Lebanese communities in western Africa; and the similarly strong influence of Indian societies in eastern Africa. Other examples include the I(g)bo in Nigeria and the Kikuyu in Kenya. Finally, to the extent that ethnic inequality implies that well-being depends on one's ethnic identity, then it is more likely to generate envy and perceptions that the system is "unfair" and reduce interpersonal trust, more so than the conventionally measured economic inequality, since the latter can be more easily thought of as the result of ability or effort. Consistent with the view that ethnic inequality is detrimental to the formation of social ties across groups, Tesei (2014) finds that greater racial inequality across US metropolitan areas is associated with low levels of social capital.

---

[2] Ethnic inequality may impede development by spurring civil conflict (Horowitz 1985). However, Esteban and Ray (2011) show that the effect of ethnic inequality on conflict is ambiguous, as it also depends on within-group inequality. Recent works in political science provide opposing results. Cederman, Weidman, and Gleditch (2011) combine proxies of local economic activity from the G-Econ database with ethnolinguistic maps to construct an index of ethnic inequality for a subset of "politically relevant ethnic groups" (as defined by the Ethnic Power Relations Dataset) and then show that in highly unequal countries, both rich and poor groups fight more often than those groups whose wealth is closer to the country average. However, in parallel work, Huber and Mayoral (2013) find no link between inequality across ethnic lines and conflict.

[3] Similarly, Deshpande (2000) and Anderson (2011) focus on income inequality across castes in India and associate between-caste inequality to public goods provision. See also Loury (2002) for an overview of works studying the evolution of racial inequality in the United States and its implications.

Organization.—The paper is organized as follows. In Section II, we describe the construction of the ethnic (and regional) inequality measures and present summary statistics and the basic correlations. In Section III, we report the results of our analysis associating income per capita with ethnic inequality across 173 countries. Besides reporting various sensitivity checks, we also examine the link between development and inequality across administrative regions. In Section IV, we explore the geographic origins of contemporary differences in economic performance across groups. In Section V, we report estimates associating contemporary development with inequality in geographic endowments across ethnic homelands. In Section VI, we summarize our findings and discuss avenues for future research.

II.   Data

To construct proxies of ethnic inequality for the largest set of countries, we combine information from ethnographic-linguistic maps on the location of groups with satellite images of light density at night that are available at a fine grid. In this section, we discuss the construction of the cross-country measures reflecting inequality in development (as captured by luminosity per capita) across ethnic homelands within 173 countries. We also describe in detail the construction of the other measures of spatial inequality and discuss the main patterns.

A.   *Ethnic Inequality Measures*

1.   Location of Ethnic Groups

We identify the location of ethnic groups employing two data sets/maps.[4] 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 928 ethnic groups around the world. The information pertains to the early 1960s, so for many countries, in Africa in particular and to a lesser extent in Southeast Asia, it corresponds to the time of independence.[5] The 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; this results in 2,129 ethnic homelands within

---

[4] Note that across all units of analysis in the construction of the respective indexes, we exclude polygons of less than 1 square kilometer (km²) to minimize measurement error in the drawing of the underlying maps.

[5] The original *Atlas Narodov Mira* consists of 57 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 at the time.

contemporary countries. Most areas (1,637) are coded as pertaining to a single group, whereas in the remaining 492 homelands, there can be up to three overlapping groups. For example, in northeastern India over an area of 4,380 km², the Assamese, the Oriyas, and the Santals overlap. The luminosity of a region where multiple groups reside contributes to the average luminosity of each group. The size of ethnic homelands varies considerably. The smallest polygon occupies an area of 1.09 km² (French in Monaco), and the largest extends over 7,335,476 km² (American English in the United States). The median (mean) group size is approximately 4,200 (61,000) km². The median (mean) country in our sample has eight (11.5) ethnicities, with the most diverse being Indonesia with 95 groups.

Our second source is the fifteenth edition of the *Ethnologue* (Gordon 2005), which maps 7,581 language-country groups worldwide in the mid/late 1990s, using the political boundaries of 2000. In spite of the comprehensive linguistic mapping, *Ethnologue*'s coverage of the Americas and Australia is rather limited while for others (i.e., Africa and Asia), it is very detailed. Each polygon delineates a traditional linguistic region; populations away from their homelands (in cities, refugee camps) are not mapped. Groups of unknown location, as well as widespread and extinct languages, are not mapped either; the only exception is the English in the United States. *Ethnologue* also records areas where languages overlap. *Ethnologue* provides a more refined linguistic aggregation compared to the GREG. As a result, the median (mean) homeland extends to 726 (12,676) km². The smallest language is the Domari in Israel, which covers 1.18 km², and the largest group is the English in the United States, covering 7,330,520 km². The median (mean) country has nine (42.3) groups, with Papua New Guinea being the most diverse with 809 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 (e.g., the Toba and the Quechua). For Canada, *Ethnologue* lists 77 mostly indigenous groups, such as the Blackfoot and the Chipewyan, with only English and French being nonindigenous; in contrast, GREG lists 23 groups featuring many nonindigenous ones, such as Swedes, Russians, Norwegians, and Germans. Hence, the two ethnolinguistic mappings capture different cleavages, at least in some continents. Though we have performed various sensitivity checks, for our benchmark results we are including all groups without attempting to make a distinction as to which cleavage is more salient.[6]

---

[6] A thorough exploration of ethnic inequality across different linguistic cleavages is relegated to the online supplementary appendix.

It is important to note that the underlying maps do include regions where groups overlap, and we take that into account in our measure, as we show below. However, both maps do not capture relatively recent within-country migrations toward the urban centers, for example. The reason is that the original sources attempt to trace the historical homeland of each group. Hence, actual ethnic mixing is likely higher than what the ethnographic maps reflect. This will induce measurement error to our proxies of ethnic inequality. Nevertheless, under the assumption that in a given urban center the respective indigenous group is relatively more populous than recent migrant ones, assigning the observed luminosity per capita to this group is not entirely ad hoc. Moreover, there is a large literature documenting that migrant workers channel systematically a fraction of their earnings back to their homelands. This would imply that although we do not observe migrant workers in our data set to the extent that they send remittances to their families and influence their livelihoods, this will be reflected in the luminosity per capita of the ancestral homelands, which we directly measure. Moreover, to at least partially account for this issue, we have constructed all inequality measures also excluding the regions where capitals fall.

2.   Data on Luminosity and Population

Comparable data on income per capita at the ethnicity level are scarce. Hence, following Henderson et al. (2012) and subsequent studies (e.g., Chen and Nordhaus 2011; Michalopoulos and Papaioannou 2013, 2014; Pinkovskiy 2013; Hodler and Raschky 2014; Pinkovskiy and Sala-i-Martin 2014), we use satellite image data on light density at night as a proxy. These—and other works—show that luminosity is a strong correlate of development at various levels of aggregation (countries, regions, ethnic homelands). The luminosity data come from the Defense Meteorological Satellite Program's Operational Linescan System that reports images of the earth at night (from 20:30 till 22:00). 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 by the population of each area using data from the Gridded Population of the World, which reports georeferenced pixel-level population estimates for 1990 and 2000.[7]

---

[7] The data are constructed using subnational censuses and other population surveys at various levels (city, neighborhood, region). For details, see http://sedac.ciesin.columbia .edu/data/collection/gpw-v. In the online supplementary appendix, we present various sensitivity checks that examine the role of measurement error in both the population estimates and the luminosity data.

3. New Ethnic Inequality Measures

We proxy the level of economic development in ethnic homeland i with mean luminosity per capita, y_i; and we then construct an ethnic Gini coefficient for each country that reflects inequality across ethnolinguistic regions. Specifically, the Gini coefficient for a country's population consisting of n groups with values of luminosity per capita for the historical homeland of group i, y_i, where $i = 1$ to n are indexed in nondecreasing order $(y_i \leq y_{i+1})$, is calculated as follows:

$$G = \frac{1}{n}\left[n + 1 - 2\frac{\sum_{i=1}^{n}(n + 1 - i)y_i}{\sum_{i=1}^{n}y_i}\right].$$

The ethnic Gini index captures differences in mean income—as captured by luminosity per capita at the ethnic homeland—across groups. For each of the two different ethnic-linguistic maps (*Atlas Narodov Mira* and *Ethnologue*), we construct Gini coefficients for the maximum sample of countries using cross–ethnic homeland data in 1992, 2000, and 2012. For each mapping we construct three ethnic Ginis. First, for our baseline estimates we use information from all groups. Second, we construct the Gini coefficient dropping the capital cities. This allows us to account both for extreme values in luminosity and also for population mixing, which is naturally higher in capitals. Third, we compile measures excluding small ethnicities, defined as those representing less than 1 percent of the 2000 population in a country.[8]

B. Measures of Spatial Inequality

Since we use ethnic homelands (rather than individual-level) data to measure between-group inequality, the ethnic inequality measures also reflect regional disparities in income or public goods provision that may not be related to ethnicity per se. To isolate the between-ethnicity component of inequality from the regional one, we also construct Gini coefficients reflecting (i) the overall degree of spatial inequality and (ii) inequality across (first- and second-level) administrative units for each country. Moreover, in an attempt to assess the accuracy of the underlying groups' mappings, we perturbed the original homelands and compiled Gini coefficients based on these altered ethnic homelands.

[8] For example, in Kenya the *Atlas Narodov Mira* (the *Ethnologue*) maps 18 (53) ethnic (linguistic) groups. Yet eight ethnic (37 linguistic) areas host less than 1 percent of Kenya's population as of 2000. So we construct Gini coefficients (i) using all ethnic groups (18 and 53), (ii) dropping ethnic regions where the capital (Nairobi) falls (17 and 52), and (iii) using the 11 ethnic and 16 linguistic groups, respectively, whose populations exceed 1 percent of Kenya's population.

1.   Overall Spatial Inequality Index

Our baseline index reflecting the overall degree of spatial inequality is based on aggregating (via the Gini coefficient formula) luminosity per capita across roughly equally sized pixels in each country. We first generate a global grid of 2.5×2.5 decimal degrees (that extends from -180 to 180 degrees longitude and from 75 degrees latitude to -65 degrees latitude). Second, we intersect the resulting grid with the 2000 Digital Chart of the World, which portrays contemporary national borders. This results in 4,865 pixels across the globe falling within country boundaries. The median (mean) box extends to roughly 22,500 (27,500) km². Note that boxes intersected by the coastline and national boundaries are smaller. Third, for each box we compute luminosity per capita in 1992, 2000, and 2012. Fourth, we aggregate the data at the country level estimating a Gini coefficient that captures the overall degree of spatial inequality. The cross-country mean (median) number of pixels used for the estimation of the spatial Ginis is 24.9 (8), so these Ginis are quite comparable to the ethnic inequality measures.

2.   Inequality across Administrative Regions

We also compiled inequality measures across administrative units, using data from the GADM Global Administrative Areas database on the boundaries of administrative regions. Following a procedure similar to the derivation of the ethnic inequality and the overall spatial inequality indexes, we construct measures reflecting inequality (in lights per capita) across first-level and second-level administrative units. In our sample of 173 countries, the median (mean) number of first-level administrative units is 13 (17). A median (mean) first-level administrative unit spans roughly 7,197 (44,050) km², which is somewhat larger than the median size (4,578) for groups in the *Atlas Narodov Mira*. The cross-country median (mean) size of second-level administrative units is 110 (301) km². So, these units are much smaller than the *Ethnologue* or the *Atlas Narodov Mira* homelands.

3.   Inequality across Perturbed Ethnic Regions

We have also created ethnic Gini coefficients using perturbed ethnic regions. Using as inputs the centroids of ethnic-linguistic homelands, we generate Thiessen polygons that have the unique property that each polygon contains only one input point and that any location within a polygon is closer to its associated point than to a point within any other polygon. Thiessen polygons have the exact same centroids as the actual linguistic and ethnic homelands in the *Ethnologue* and GREG databases, respec-

tively, the key difference being that the actual homelands have idiosyncratic shapes.[9] We then construct a spatial Gini coefficient that reflects inequality in lights per capita across these sets of Thiessen polygons. The mean size of the Thiessen polygons based on the *Ethnologue* (GREG) database is 11,862 (58,784) km², very similar to the mean size of homelands in the *Ethnologue* (GREG)—12,676 (61,213) km².

Comment.—All three proxies of the spatial inequality also reflect inequality across ethnic homelands, since (i) there is clearly some degree of measurement error on the exact boundaries of ethnic regions, and (ii) population mixing is likely higher than the one we observe in the data. Moreover, in several countries, administrative boundaries follow ethnic lines, while in the case of large groups, the spatial Gini coefficients may also (partially) capture within–ethnic group inequality.[10]

C. Example

Figures 1 and 2 provide an illustration of the construction of the ethnic inequality measures for Afghanistan. The *Atlas Narodov Mira* (GREG) maps 31 ethnicities (fig. 1*A*), whereas the *Ethnologue* reports 39 languages (fig. 2*A*). According to GREG, the Afghans (Pashtuns) are the largest group residing in the southern and central-southern regions. This group makes up 51 percent of the population in 2000. The second-largest group are the Tajik people, who compose 22 percent of the population and are located in the northeastern regions and in scattered pockets in the western part of the country. There are eight territories in which groups overlap.

Figures 1*B* and 2*B* portray the distribution of lights per capita for each group, with lighter colors indicating more brightly lit homelands. The center of the country, where the Hazara-Berberi reside, is poor; the same applies to the eastern provinces, where the Nuristani, the Pamir Tajiks, the Pashai, and the Kyrgyz groups 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 esti-

---

[9] Note that there are very few instances in which the number of Thiessen polygons is not identical to the number of the underlying groups (e.g., there is a difference of one group for six out of the 173 countries in the *Ethnologue*). This difference is due to the fact that a handful of border/coastal groups have such a peculiar shape that their centroid falls out of the country's boundaries they belong to. Hence, since Thiessen polygons are based on the centroids of the ethnic-linguistic groups that fall within the country, those groups whose centroids fall outside are not taken into account. Note that this has virtually no effect on the results since when we focus on the countries where the number of Thiessen polygons is identical to the number of groups, the pattern is the same.

[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 make a luminosity-based, within-group inequality index less satisfactory.

A

Ethnic Homelands in Afghanistan
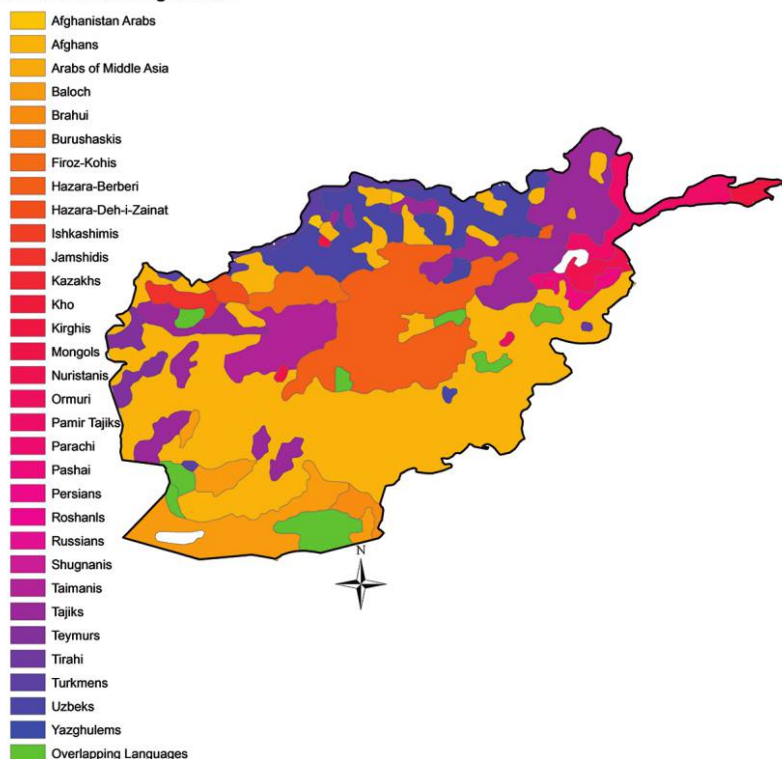
FIG. 1.—Illustration of the construction of the ethnic inequality measures for Afghanistan. The *Atlas Narodov Mira* (GREG) maps 31 ethnicities (panel *A*); the *Ethnologue* reports 39 languages (panel *B*).

mate the Gini coefficient in 1992, in 2000, and in 2012. In 2000 the Gini coefficient estimated from GREG (*Ethnologue*) is 0.95 (0.90).[11] We also estimated the ethnic inequality measures excluding the ethnic homeland where the capital, Kabul, falls. In this case the ethnic Ginis are similar (0.95 with GREG and 0.91 with *Ethnologue*). For robustness, we also estimated Gini coefficients of ethnic inequality excluding groups constituting less than 1 percent of the country's population. In this case the Gini coefficient with the GREG mapping is based on just four groups, while the *Ethnologue*-based Gini is based on seven ethnic homelands.

[11] Since the Gridded Population of the World reports zero population for some ethnic areas, the Gini index with the GREG mapping is based on 27 ethnic observations, while the Gini coefficient with the *Ethnologue* mapping is based on 39 linguistic groups (no gaps in this case).

B
Ethnologue Languages in Afghanistan
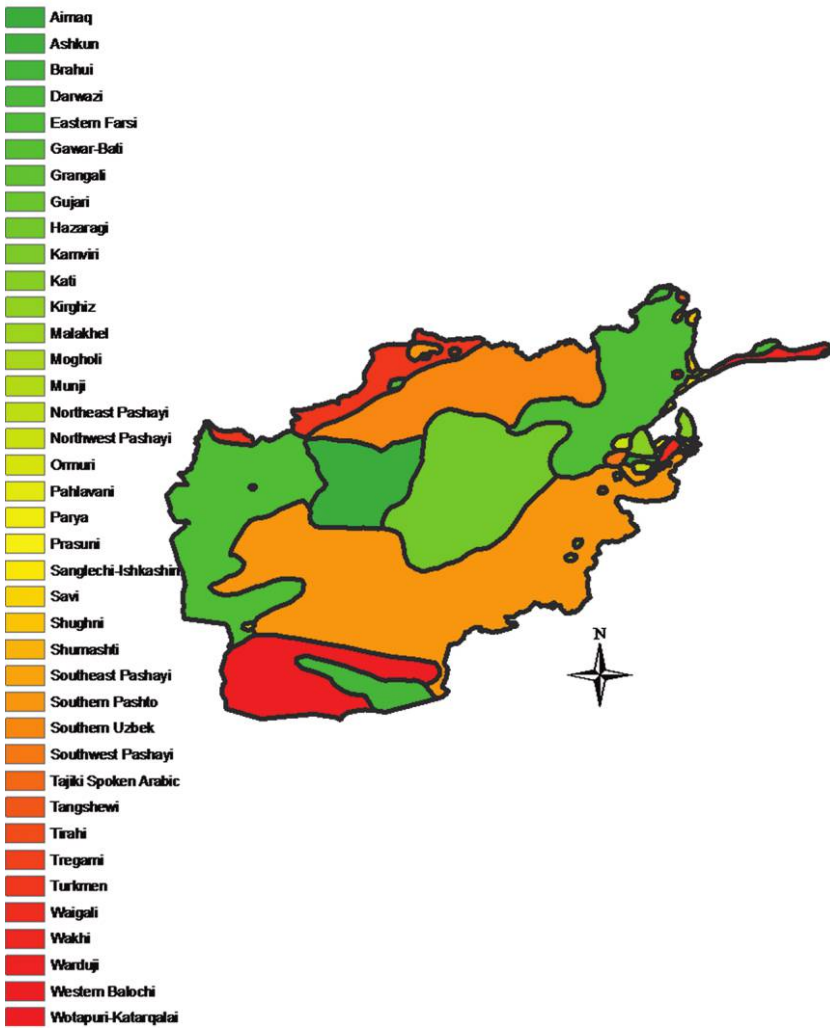
FIG. 1 (*Continued*)

Figure 3 illustrates the construction of the overall spatial inequality. When we divide the globe into boxes of 2.5×2.5 decimal degree boxes, we get 24 areas in Afghanistan. The estimated Gini index capturing the overall degree of spatial inequality in Afghanistan is 0.73. For consistency we also estimated the overall spatial inequality (Gini) index excluding the pixel where the capital city falls or those boxes where less than 1 percent
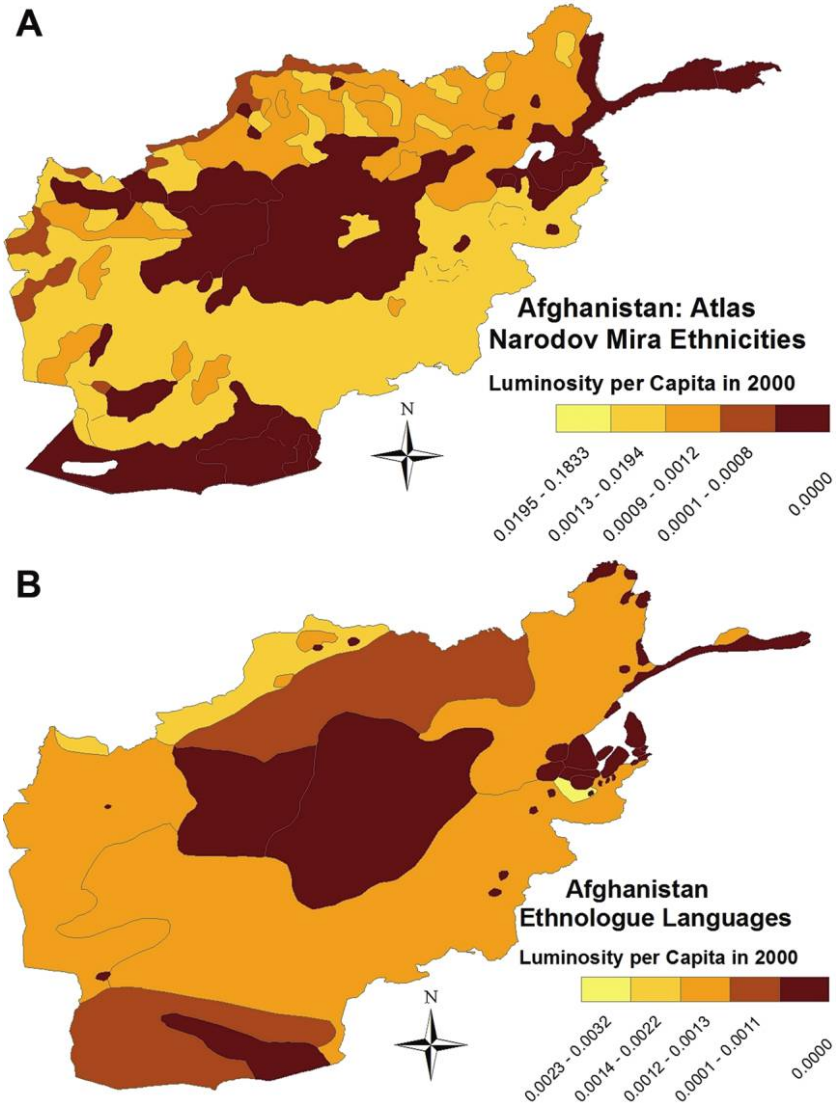
Fig. 2.—Distribution of lights per capita for each group based on GREG and *Ethnologue* mapping, with lighter colors indicating more brightly lit areas.

of the country's population lived in 2000. Figures 4*A* and 4*B* illustrate the construction of inequality measures across administrative regions using both the first-level and second-level units. There are 32 provinces (*velayat*) that constitute the first-level administrative units, and there are 328 second-level administrative units (*wuleswali*). After estimating aver-
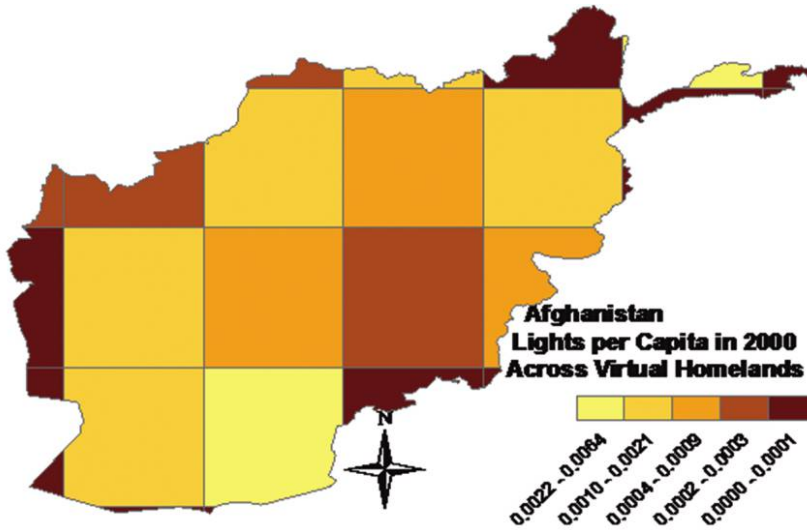
Fig. 3.—Construction of the overall spatial inequality. When we divide the globe into boxes of 2.5 × 2.5 decimal degree boxes, we get 24 areas in Afghanistan.

age luminosity per capita for each unit, we construct Gini indexes capturing inequality in development across administrative regions. Again, we construct these inequality measures using all regions, dropping the capital, and also excluding those units with less than 1 percent of total population. In our example, the first-level administrative unit Gini index is 0.76 and the second-level administrative unit Gini coefficient is 0.93. Figures 5A and 5B illustrate the derivation of the perturbed ethnic homelands Gini index for Afghanistan based on the *Atlas Narodov Mira* and the *Ethnologue,* respectively. There are 31 and 39 Thiessen polygons, as many as the number of ethnic and linguistic groups. The centroids of the Thiessen polygons are identical to the ones of the actual homelands, the only difference being that the actual homelands have rather peculiar shapes.

D. Descriptive Evidence

1. Ethnic Inequality around the World

Table 1 reports summary statistics for the baseline ethnic inequality measures and the proxies of the overall degree of spatial inequality and regional inequality across administrative units. The average and median values of the ethnic Gini coefficients are quite similar with both mappings in each year (around 0.42–0.49 in 2000). The average (median) value of the overall spatial Gini coefficient in 2000 is similar, 0.42 (0.43). The
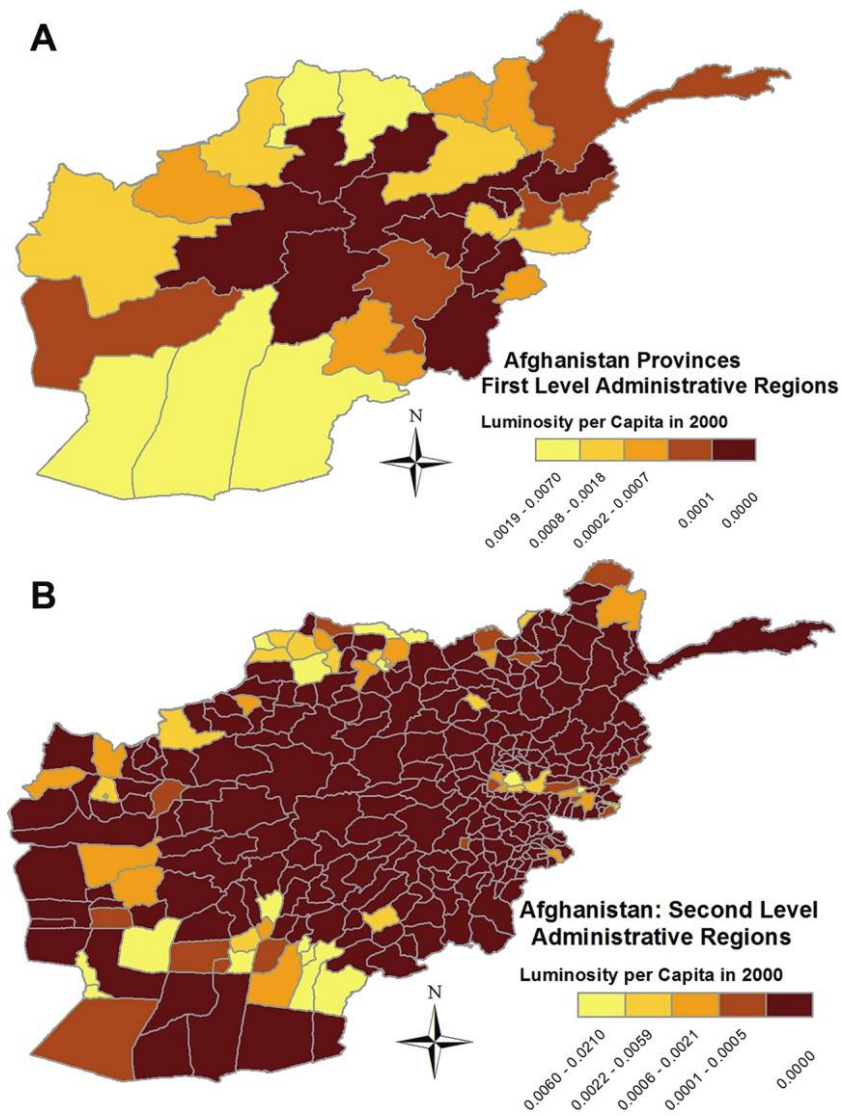
FIG. 4.—Construction of inequality measures across administrative regions using both the first-level and second-level units.

Gini coefficients based on administrative regions are, on average, smaller when estimated across first-level units (mean 0.37) and larger when estimated at the finer second level (mean 0.57). Moreover, regional inequality seems to be slightly trending downward, as all Gini coefficients are smaller in 2012 (and in 2000). This may be driven by the expansion of
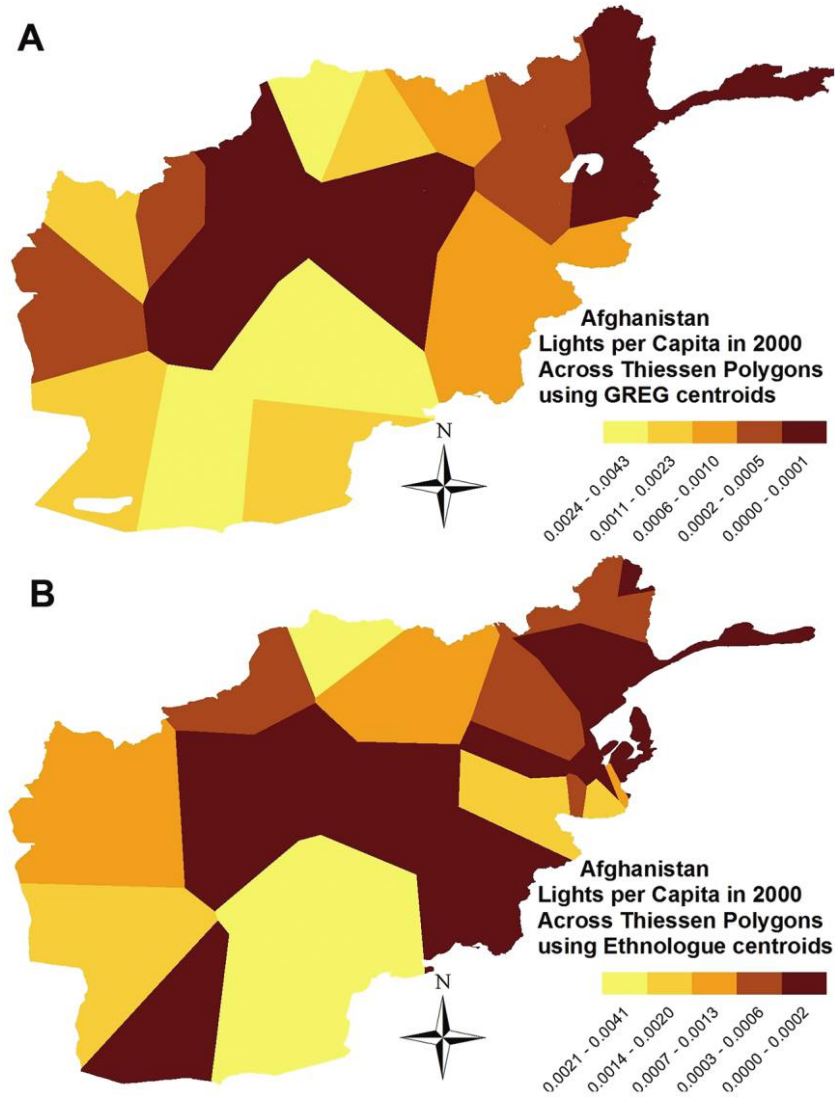
Fig. 5.—Perturbed ethnic homelands for Afghanistan based on the *Atlas Narodov Mira* (panel *A*) and the *Ethnologue* (panel *B*).

electrification (and regional convergence) in many underdeveloped and developing countries (mostly in Africa and South Asia).

Figures 6*A* and 6*B* illustrate the global distribution of ethnic inequality with the GREG and *Ethnologue* mappings, respectively. Sub-Saharan Africa and East and South Asia host the most ethnically unequal countries. For example, with the *Ethnologue* mapping, the mean (median) of the

TABLE 1
SUMMARY STATISTICS: CROSS-COUNTRY INEQUALITY MEASURES

	Observations	Mean	Standard Deviation	Minimum	25th Percentile	Median	75th Percentile	Maximum
A. *Atlas Narodov Mira* (GREG)								
Number of ethnic homelands	173	10.994	13.579	1	3	8	12	94
Ethnic Gini in 2012	173	.406	.248	.00	.21	.44	.57	.96
Ethnic Gini in 2000	173	.424	.260	.00	.20	.48	.63	.97
Ethnic Gini in 1992	173	.473	.280	.00	.27	.54	.70	.96
B. *Ethnologue*								
Number of linguistic homelands	173	38.619	94.029	1	3	8	35	753
Ethnolinguistic Gini in 2012	173	.439	.325	.00	.13	.44	.74	.98
Ethnolinguistic Gini in 2000	173	.446	.333	.00	.12	.49	.77	.98
Ethnolinguistic Gini in 1992	173	.487	.346	.00	.14	.54	.80	.99
C. Pixels								
Number of pixels	173	19.572	36.838	1.00	4.00	8.00	20	292
Spatial Gini in 2012	173	.411	.270	.00	.18	.42	.60	.98
Spatial Gini in 2000	173	.421	.269	.00	.18	.43	.64	.97
Spatial Gini in 1992	173	.463	.271	.00	.21	.49	.70	.96

444

D. First-Level Administrative Units

N first-level administrative units	173	16.873	14.962	1.00	8.00	12.00	20	88
Administrative unit (first-level)								
Gini in 2012		.353	.202	.00	.20	.30	.45	.93
Administrative unit (first-level)								
Gini in 2000		.368	.215	.00	.21	.31	.48	.94
Administrative unit (first-level)								
Gini in 1992		.422	.241	.00	.21	.38	.59	.94

E. Second-Level Administrative Units

N second-level administrative units	135	290.104	626.285	11.00	48.00	99.00	248	5,478
Administrative unit (second-level)								
Gini in 2012		.562	.237	.14	.35	.56	.78	.95
Administrative unit (second-level)								
Gini in 2000		.572	.250	.15	.36	.55	.83	.96
Administrative unit (second-level)								
Gini in 1992		.639	.249	.13	.43	.68	.87	.98

NOTE.—The table reports summary statistics for the main ethnic inequality, overall spatial inequality, and administrative unit inequality measures employed in the cross-country analysis. Section II and the Data Appendix give details on the construction of the ethnic inequality measures (Gini coefficients). All Gini coefficients are estimated using luminosity per capita across ethnic homelands (in panel A), across linguistic homelands (in panel B), pixels/boxes of roughly the same size (in panel C), first-level administrative units (in panel D), and second-level administrative units (in panel E).

445
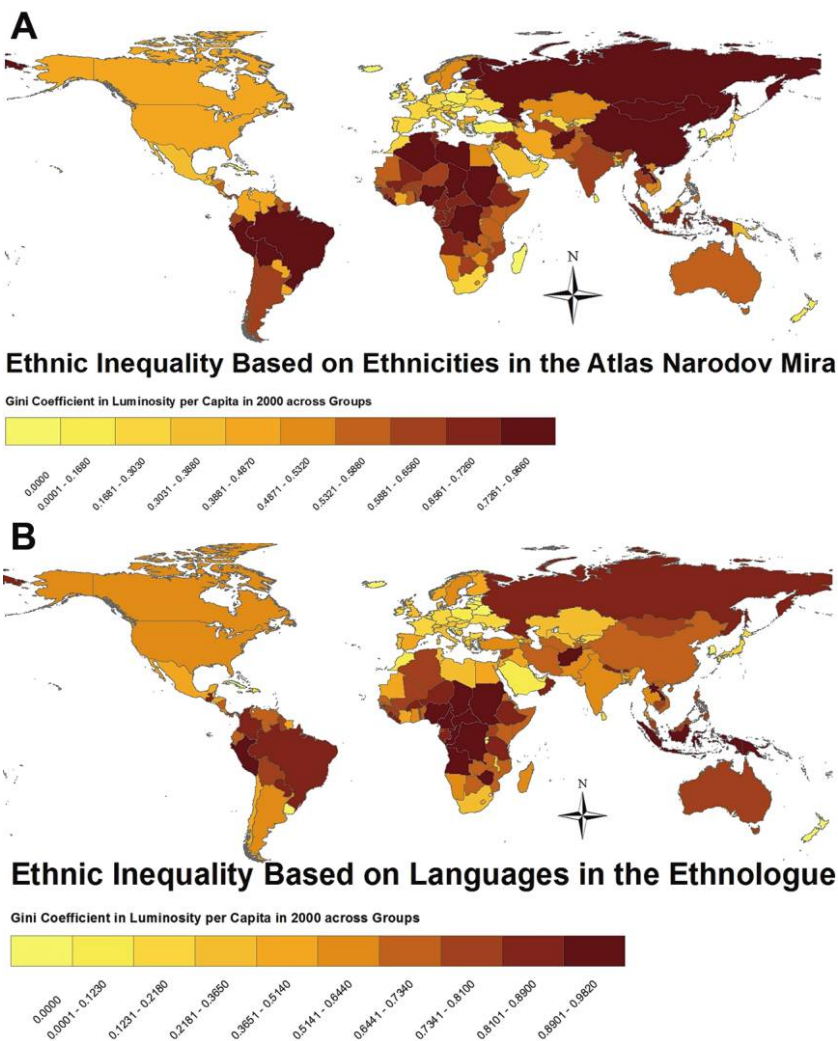
FIG. 6.—Global distribution of ethnic inequality with the GREG and the *Ethnologue* mapping (panels *A* and *B*), world distribution of the overall degree of spatial inequality and regional inequality across first-level administrative units (panels *C* and *D*), and global distribution of ethnic inequality partialing out the effect of the overall spatial inequality (panels *E* and *F*).

baseline ethnic inequality index for sub-Saharan African countries is 0.63 (0.728), while for South and East Asian countries the corresponding mean (and median) value of the ethnic Gini index is 0.59 (0.69).[12] In

[12] Specifically, ethnic inequality is particularly high across South Asia (in total seven countries, namely, Afghanistan, Bangladesh, Bhutan, India, Nepal, Pakistan, and Sri

C

Spatial Inequality Based on 2.5 by 2.5 Decimal Degree Boxes

Gini Coefficient in Luminosity per Capita in 2000 across Virtual Countries

0.0000 - 0.0410 0.0411 - 0.1420 0.1421 - 0.2270 0.2271 - 0.3350 0.3351 - 0.4380 0.4381 - 0.5100 0.5101 - 0.5980 0.5981 - 0.6850 0.6851 - 0.7940 0.7941 - 0.9670

D

Administrative-Level Inequality Based on First-Level Regions

Gini Coefficient in Luminosity per Capita in 2000 across First-Level Administrative Regions

0.0000 - 0.1240 0.1241 - 0.1820 0.1821 - 0.2280 0.2281 - 0.2630 0.2631 - 0.3130 0.3131 - 0.3840 0.3841 - 0.4400 0.4401 - 0.5690 0.5691 - 0.7110 0.7111 - 0.9370

FIG. 6 (*Continued*)

contrast, western Europe is the region with the lowest level of ethnic inequality (mean and median values of ethnic Gini around 0.24). According to the *Atlas Narodov Mira*, the five most ethnically unequal countries are Sudan, Afghanistan, Mongolia, Zambia, and Central African Repub-

_____

Lanka). The mean and median Gini index is 0.635 on the basis of the *Ethnologue* and 0.55 when we use the GREG. Ethnic inequality is also high in the 21 countries of the East Asia and Pacific region, but only when we use the *Ethnologue*, where the mean is 0.58.

E

Ethnic Inequality based on Atlas Narodov Mira Groups: Partialling Out Spatial Inequality across 2.5 by 2.5 Decimal Degree Boxes

Residual of Ethnic Inequality on Virtual Countries Inequality

-0.5739 - 0.2371   -0.2370 - 0.1384   -0.1383 - 0.0993   -0.0992 - 0.0575   -0.0574 - 0.0059   -0.0060 - 0.0526   0.0527 - 0.0884   0.0885 - 0.1759   0.1760 - 0.2259   0.2260 - 0.3004

F

Ethnic Inequality based on Ethnologue Groups: Partialling Out Spatial Inequality across 2.5 by 2.5 Decimal Degree Boxes

Residual of Ethnic Inequality on Virtual Countries Inequality

-0.6175 - 0.3099   -0.3098 - 0.1962   -0.1961 - 0.0907   -0.0906 - 0.0392   -0.0391 - 0.0144   0.0145 - 0.0651   0.0652 - 0.1299   0.1300 - 0.1825   0.1826 - 0.3050   0.3051 - 0.5776

FIG. 6 (*Continued*)

lic, with an average Gini coefficient in luminosity across ethnic home-
lands of 0.91. According to the *Ethnologue*'s more detailed mapping of
language groups, the countries with the highest cross–ethnic group in-
equality (where Gini exceeds 0.95) are Democratic Republic of Congo,
Papua New Guinea, Sudan, Ethiopia, and Chad.

Figures 6*C* and 6*D* plot the world distribution of the overall degree of spatial inequality and regional inequality across first-level administrative units, respectively. As is evident, spatial and regional inequality is much higher in Asia and Africa as compared to western Europe and Latin America. The countries with the highest overall spatial inequality according to the measure based on the 2.5×2.5 decimal degree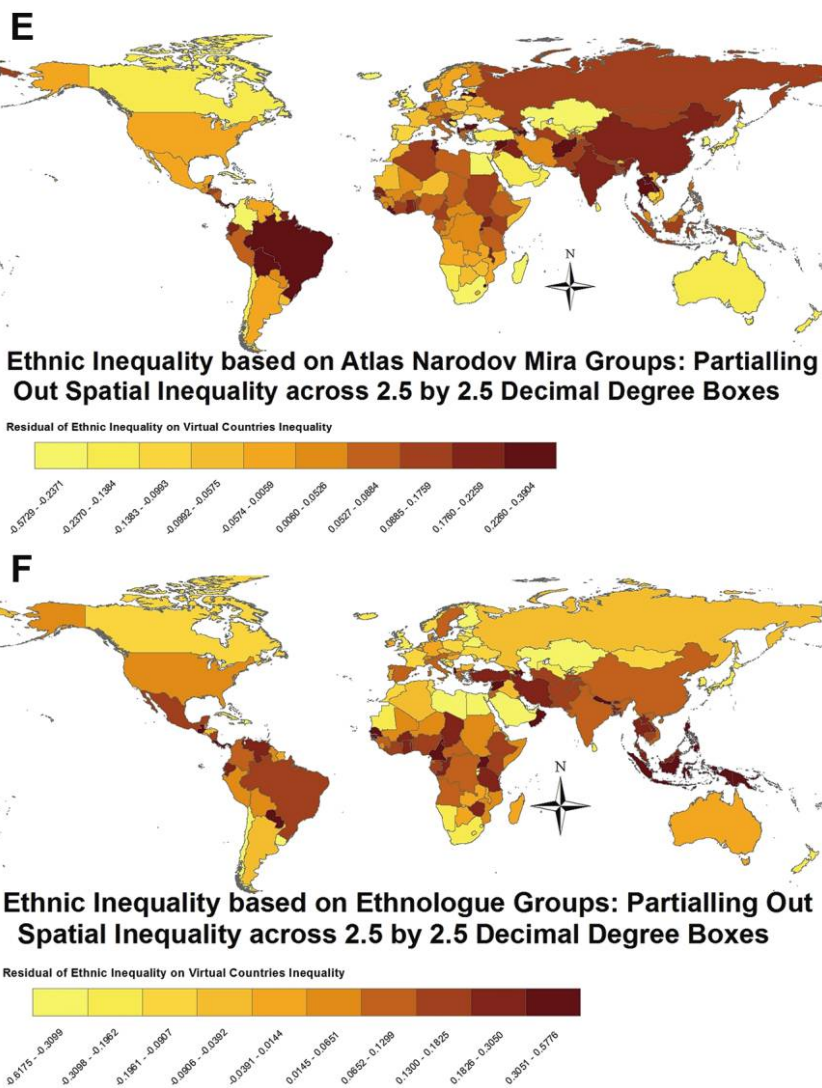 boxes are Russia, Mongolia, Sudan, Peru, and Egypt; in all these countries the spatial Gini coefficient exceeds 0.90. The countries with the highest regional inequality across first-level administrative units are Libya, Chad, and Guinea (Gini around 0.90). We should stress that in some countries first-level administrative units cover large territories (in terms of both population and land area). Hence, inequality measured across these units may not adequately capture existing regional inequalities. To partly account for this, we have also constructed Gini coefficients using second-level administrative regions that in many countries are numerous. However, an important caveat to keep in mind throughout the analysis is that in several countries regional inequalities and, more importantly, ethnic disparities in income may occur at much finer levels of aggregation (e.g., neighborhoods) than what our ancestral-ethnic homeland approach allows for.[13]

Appendix table 1 reports the correlation structure of the ethnic Gini coefficients between the two global maps at different points in time. A couple of interesting patterns emerge. First, the correlation of the Gini coefficients across the two alternative mappings is strong but not overwhelming. The correlation with the baseline measures that use all ethnic areas is around .75, but when we drop small groups or capitals, the correlation falls to .65. In line with our discussion above, these correlations suggest that the two maps capture somewhat different aspects of ethnic-linguistic cleavages. Second, in the 20-year period for which luminosity data are available (1992–2012), ethnic inequality appears very persistent, as the correlations of the Gini coefficients over time exceed .90. Given the high inertia, in our empirical analysis below we will exploit cross-country variation. Third, not surprisingly, the correlation between ethnic inequality and the Gini coefficient capturing the overall degree of

---

[13] A case exemplifying this situation is that of South Africa, a country with sizable income differences between ethnic groups. Since segregation, after the fall of apartheid, occurs at a much finer level than in the ancestral homelands, our data cannot capture this phenomenon. South Africa looks also quite equal when inequality is measured across first-level administrative units (0.22). This is very similar to the ethnic Ginis, which are 0.20 with GREG and 0.28 with *Ethnologue*. However, regional inequality in South Africa is significantly higher when estimated across second-level administrative units (Gini index is 0.40, very similar to the global mean and median values).

spatial inequality and regional inequality across (first-level) administrative units is positive, but again far from perfect. In particular, the correlation of the ethnic Gini with the overall spatial Gini (based on artificial boxes) ranges between .55 and .70, while the correlation of the ethnic Gini coefficients with the administrative unit Ginis is lower, around .50.

Since we are primarily interested in documenting the explanatory power of ethnic inequality beyond the overall spatial inequality in most specifications, we control for the latter. Figures 6E and 6F portray the global distribution of ethnic inequality partialing out the effect of the overall spatial inequality.

2. Basic Correlations

Ethnic diversity.—Appendix table 2, panel A, reports the correlation structure between the various ethnic inequality and spatial inequality measures and the widely used ethnolinguistic fragmentation indexes. We observe a positive correlation between ethnic inequality and linguistic-ethnic fractionalization (.35–.45). Figures 7A and 7B provide a graphical illustration of the association between the two proxies of ethnic inequality and the ethnic and linguistic fragmentation measures of Alesina et al. (2003) and Desmet, Ortuño-Ortín, and Wacziarg (2012), respectively. The correlation between ethnic inequality and the segregation measures compiled by Alesina and Zhuravskaya (2011) is also positive (.20–.45). Ethnic inequality tends to go in tandem with segregation. This is reasonable since economic differences between groups are more likely to persist when groups are also geographically separated. We also examine the association between ethnic inequality and spatial inequality with the ethnic polarization indicators of Montalvo and Reynal-Querol (2005), failing to detect a systematic association. These patterns suggest 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 (app. table 2, panel B). The income Gini coefficient is taken from Easterly (2007), who using survey and census data compiled by the 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 8A and 8B illustrate this association using the GREG and the *Ethnologue* mapping, respectively. The correlation between ethnic inequality and economic inequality is moderate, around .25–.30. Yet this correlation weakens considerably and becomes statistically insignificant once we simply condition on continental constants.
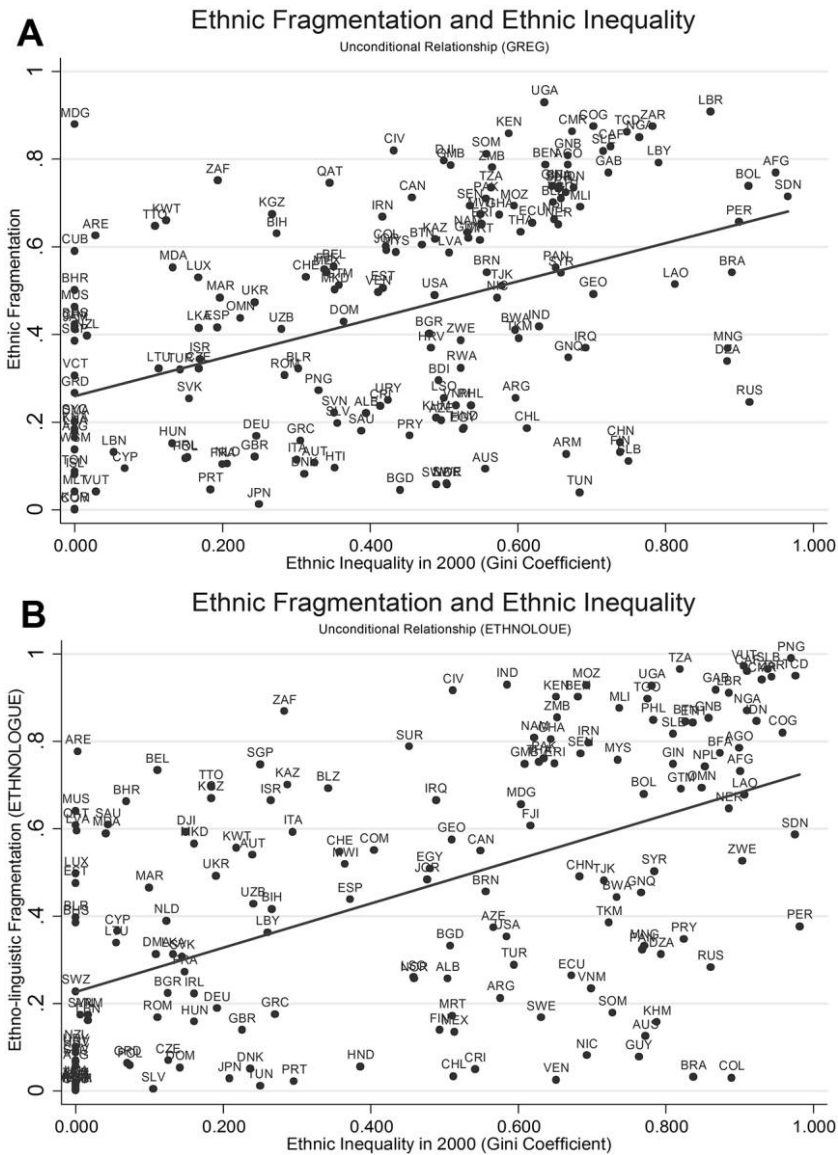
FIG. 7.—Graphical illustration of the association between the two proxies of ethnic inequality and the ethnic and linguistic fragmentation measures of Alesina et al. (2003) and Desmet et al. (2012), respectively.
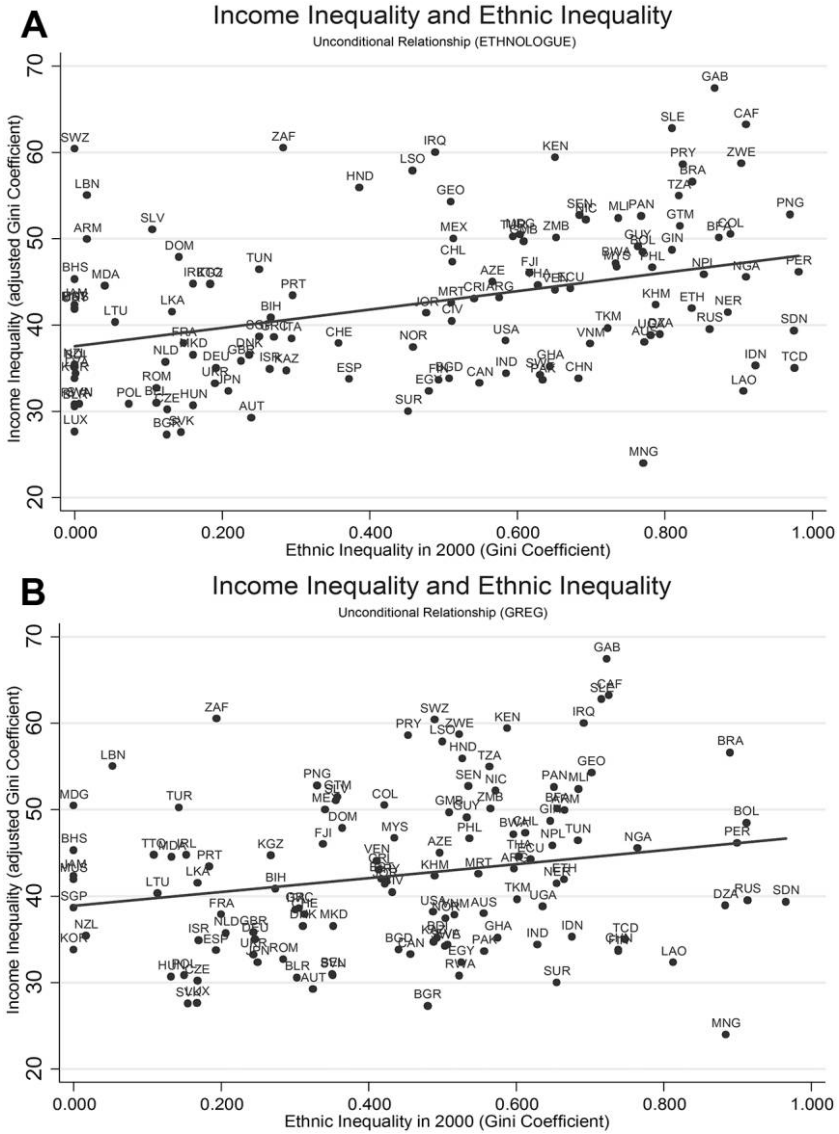
Fig. 8.—Association between income inequality and ethnic inequality using the *Ethnologue* and GREG mapping of groups' homelands.

III. Ethnic Inequality and Development

A. Baseline Estimates

In table 2 we report cross-country ordinary least squares (OLS) estimates, relating the log of per capita GDP in 2000 with ethnic inequality. In panel

A we use the ethnic inequality measure based on the *Atlas Narodov Mira* mapping, while in panel B we use the measures derived from *Ethnologue*'s mapping. In all specifications we include region-specific constants (following the World Bank's classification) to account for continental differences in ethnic inequality and comparative economic development.

The coefficient of the ethnic inequality index in column 1 is negative and significant at the 1 percent level. Figures 9*A*–9*D* illustrate the unconditional and the conditional on regional fixed effects association. Specification 2 also reveals a negative association between economic development and the overall degree of spatial inequality, as reflected in the Gini coefficient based on pixels of 2.5×2.5 degrees. This suggests that underdevelopment goes in tandem with regional inequalities. In column 3 we include both the ethnic inequality Gini index and the spatial Gini coefficient. The estimate on the ethnic inequality Gini is stable with both the GREG and the *Ethnologue* mappings. In contrast, the coefficient on the overall spatial inequality measure drops considerably and becomes statistically indistinguishable from zero in both models. This suggests that the ethnic component of spatial inequality is the relatively stronger negative correlate of development.

In column 4 we associate the log of per capita GDP with 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; yet the estimates in column 5, where we jointly include in the empirical model the proxies of ethnic inequality and fractionalization, show that it is income differences along ethnic lines rather than ethnolinguistic heterogeneity per se that correlate with underdevelopment. The results are similar when we jointly include in the specification the ethnic Gini index, the overall spatial inequality measure, and the fractionalization measure in column 6. Although, as a result of the small number of observations and multicollinearity (see app. table 1), these results should be interpreted with caution, only the ethnic inequality measure enters with a statistically significant estimate.

In columns 7 and 8 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 or by the skewness in the size of ethnic homelands; to do so we construct Gini coefficients of population and land area that capture inequality in the size of ethnic homelands. The ethnic inequality Gini index retains its economic and statistical significance, while both the population and the homeland size ethnic Ginis enter with estimates statistically indistinguishable from zero. This suggests that the association between ethnic inequality and underdevelopment is not driven by inequality in the size of ethnic homelands captured by either the population of each group or the area of each homeland. In column 9 we also control for a country's size including in the empirical model the log of population in 2000 and

TABLE 2
BASELINE ESTIMATES: ETHNIC INEQUALITY AND ECONOMIC DEVELOPMENT IN 2000

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				A. Atlas Narodov Mira (GREG)					
Ethnic inequality (Gini)	−1.3911***		−1.3900***		−.9518**	−.9276*	−1.3449***	−1.1032**	−1.1172**
	(.2588)		(.3416)		(.3953)	(.4845)	(.4943)	(.5188)	(.5492)
Spatial inequality (Gini)		−.9973***	−.0015			−.0315	−.0046	.0104	−.5592
		(.2774)	(.3510)			(.3568)	(.3539)	(.3591)	(.4749)
Log number of ethnicities				−.3136***	−.1429	−.1433		−.2174*	−.1863
				(.0612)	(.0908)	(.0917)		(.1277)	(.1440)
Ethnic inequality in population (Gini)							.6517	1.1554	1.0858
							(1.1500)	(1.1554)	(1.2546)
Ethnic inequality in size (area; Gini)							−.7933	−.7949	−.8276
							(1.1732)	(1.1060)	(1.2297)
Log land area									.1442
									(.0879)
Log population (in 2000)									−.1368
									(.0829)
Adjusted R^2	.654	.623	.652	.646	.657	.655	.65	.656	.662
Observations	173	173	173	173	173	173	173	173	173
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
				B. Ethnologue					
Ethnic inequality (Gini)	−1.1603***		−1.0745***		−1.1688***	−1.0636**	−1.3452***	−1.2368***	−1.7186***
	(.2328)		(.2652)		(.3587)	(.4241)	(.3469)	(.4377)	(.4549)
Spatial inequality (Gini; pixels)		−.9973***	−.1549			−.1564	−.1760	−.1992	−.5351
		(.2774)	(.3021)			(.3136)	(.3010)	(.3165)	(.4235)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log number of languages				−.1921*** (.0466)	.0021 (.0688)	−.0025 (.0711)		−.0378 (.0826)	.0726 (.0967)
Ethnic inequality in population (Gini)							.8012 (.9324)	.8460 (.9387)	.8453 (.9144)
Ethnic inequality in size (area; Gini)							−.4898 (.8948)	−.4574 (.9045)	−.2958 (.8853)
Log land area									.1366* (.0808)
Log population									−.2105** (.0840)
Adjusted R^2	.654	.623	.652	.632	.652	.65	.651	.649	.665
Observations	173	173	173	173	173	173	173	173	173
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note.—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-linguistic homelands. In panel A, we use the digitized version of the *Atlas Narodov Mira* (GREG) to aggregate lights per capita across ethnic homelands. In panel B, we use the digitized version of the *Ethnologue* database to aggregate lights per capita across linguistic homelands. The overall spatial inequality index (Gini coefficient) captures the degree of spatial inequality across 2.5×2.5 decimal degree boxes/pixels in each country (boxes intersected by national boundaries are of smaller size). Section II gives details on the construction of the ethnic inequality and spatial inequality (Gini) indexes. The log number of ethnicities in cols. 4–6, 8, and 9 denotes the logarithm of the number of ethnic and linguistic groups in each country according to the GREG (in panel A) and the *Ethnologue* (in panel B). Columns 7, 8, and 9 include as controls a Gini index capturing inequality in population across ethnic (linguistic) homelands and a Gini index capturing inequality in land area across ethnic (linguistic) homelands. Column 9 includes the log of a country's land area and the log of population in 2000. All specifications include regional fixed effects (constants not reported). The Data Appendix gives detailed variable definitions and data sources. Robust (heteroskedasticity-adjusted) standard errors are reported in parentheses below the estimates.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

455
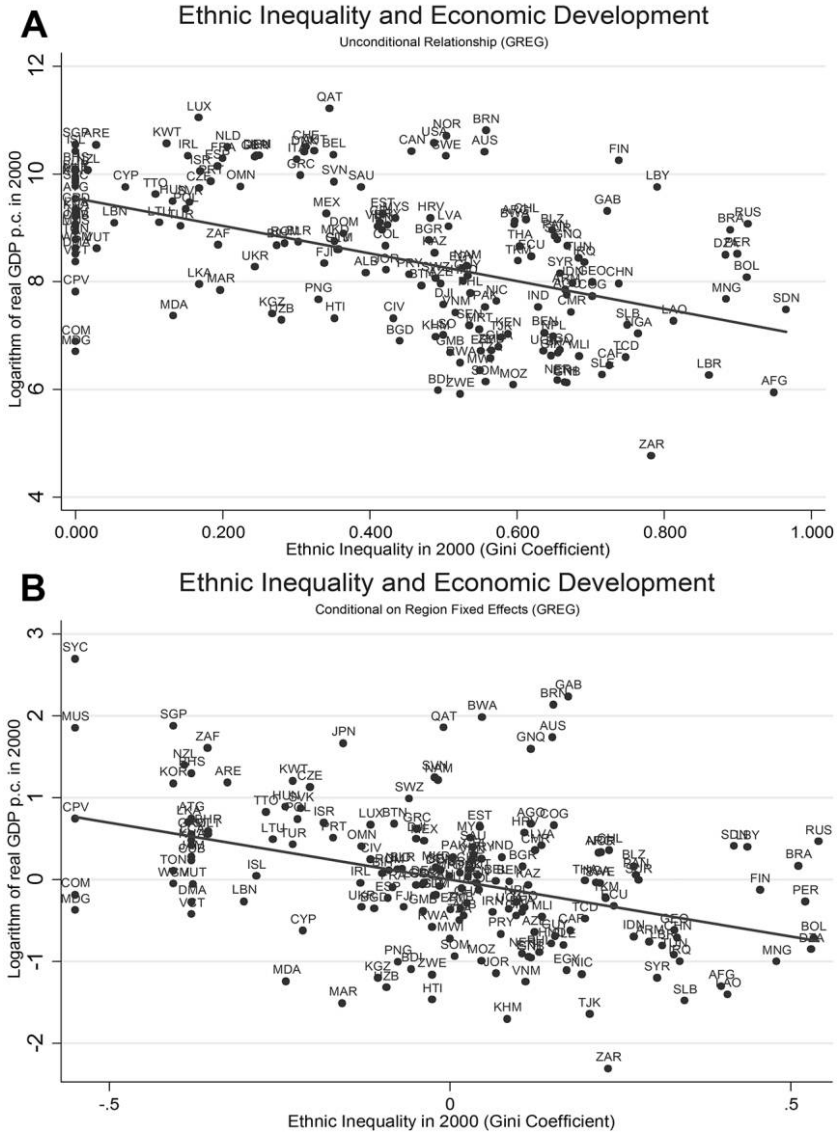
Fig. 9.—Unconditional and conditional on regional fixed effects association between ethnic inequality and GDP per capita across countries.

log of land area, as ethnic heterogeneity, ethnic inequality, and the over-all degree of spatial inequality are likely to be increasing in size. Doing so has little effect on our results. Ethnic inequality remains a systematic correlate of underdevelopment.
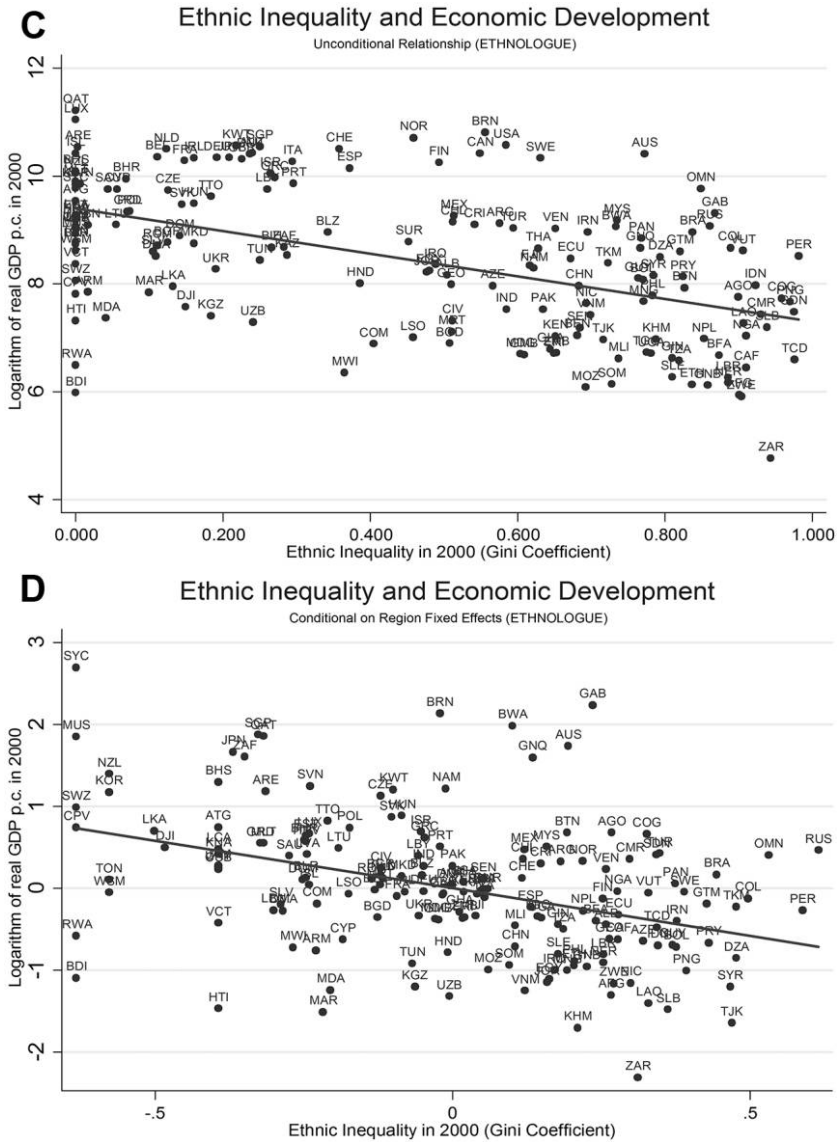
FIG. 9 (*Continued*)

The estimate on the ethnic inequality index with the *Atlas Narodov Mira* mapping in panel A (col. 9) implies that a reduction in the ethnic Gini coefficient by 0.25 (one standard deviation, from the level of Nigeria, where the ethnic Gini is 0.76, to the level of Namibia, where the ethnic Gini is 0.53) is associated with a 28 percent (0.25 log points) increase in per capita GDP (these countries have very similar overall spatial Ginis

of around 0.8). The standardized beta coefficient of the ethnic inequality index is around 0.20–0.30, quite similar to that of the works on the role of institutions on development (e.g., Acemoglu, Johnson, and Robinson 2001).

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, polarization, and genetic diversity, rather than income inequality across ethnic lines, influence comparative development. In columns 1 and 6 we augment the specification with the Alesina et al. (2003) and Desmet et al. (2012) ethnic and linguistic fractionalization measures, respectively. Doing so has no effect on the coefficient on ethnic inequality, which retains its economic and statistical significance. Moreover, the fractionalization indicators enter with unstable and statistically insignificant estimates, suggesting that it is differences in well-being across ethnic lines that explain underdevelopment rather than fragmentation per se.[14] In columns 2 and 7 we experiment with Fearon's (2003) cultural fragmentation index that adjusts the fractionalization index for linguistic distances among ethnic groups. Cultural fractionalization enters with a statistically insignificant estimate, while the ethnic inequality Gini index retains its economic and statistical significance.

Motivated by recent works highlighting the importance of polarization (Montalvo and Reynal-Querol 2005; Esteban, Mayoral, and Ray 2012), in columns 3 and 8 we condition on an index of ethnic polarization. Ethnic inequality correlates strongly with development, while the polarization measures enter with insignificant estimates.[15]

Building on the recent work of Alesina and Zhuravskaya (2011) showing that countries with a high degree of ethnolinguistic segregation tend to have low-quality national institutions and inefficient bureaucracies, in columns 4 and 9 we include in the specifications their measures of ethnic and linguistic segregation, respectively. The sample falls considerably, as these measures are available for approximately 90 countries. While there is some evidence that ethnic segregation is a feature of underdevelopment, the coefficient on the ethnic inequality proxy continues to be quite stable and significant at standard confidence levels.

In columns 5 and 10 we condition on a proxy of within-country genetic diversity, based on migratory distance of each country's capital from

[14] When we do not include the ethnic inequality Ginis, the ethnic and linguistic fragmentation measures enter with negative and significant (at the 10 percent to 5 percent level) estimates (approximately −0.55).

[15] The same applies if we use alternative measures of ethnic-linguistic polarization. Overall polarization is significantly related to civil conflict but not to income per capita. 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.

Ethiopia. Since Ashraf and Galor (2013) argue that the effect of genetic diversity on development is nonlinear, we enter the latter in a quadratic fashion (though this has no effect on our results). In all permutations the ethnic inequality proxy enters with a stable (around -1) and highly significant estimate.

Overall the results in table 3 show that the strong negative association between ethnic inequality and income across countries is not mediated by differences in the societies' ethnic or genetic composition.[16]

Alternative measures of ethnic inequality and geographic controls.—In table 4 we augment the main specification with an array of geographic traits and experiment with alternative measures of ethnic inequality. In columns 1 and 4 we use the baseline ethnic inequality measures based on all homelands. In columns 2 and 5 we use ethnic Ginis that exclude from the estimation regions where capitals fall. Note that the sample drops as in these models we do not consider monoethnic and monolinguistic countries. In columns 3 and 6 we introduce ethnic Ginis that exclude groups with less than 1 percent of a country's population. Note that, a priori, there is no reason to exclude small groups, since ethnic hatred may be directed to minorities that, nevertheless, control a significant portion of the economy (Chua 2003). Moreover, when these groups are dropped, the sample of ethnic homelands used to estimate the ethnic Ginis drops considerably.[17] To avoid concerns of self-selecting the conditioning set, we follow the baseline specification of Nunn and Puga (2012) and include (on top of log population and log land area) an index of terrain ruggedness, distance to the coast, an index of gem quality, the percentage of each country with fertile soil, and the percentage of tropical land (the Data Appendix gives variable definitions). To isolate the role of ethnic inequality on development from regional inequalities and ethnic fragmentation, in all specifications we control for the overall degree of spatial inequality in lights per capita and ethnic-linguistic fractionalization.

The negative correlation between ethnic inequality and income per capita remains strong. This applies to all proxies of ethnic inequality. While compared to the unconditional specifications the estimate on ethnic Gini declines somewhat, it retains significance at standard confidence levels. Thus, while still an unobserved or omitted countrywide factor may

---

[16] We also experimented with the newly constructed index of birthplace diversity of Alesina, Harnoss, and Rapoport (2016), again finding that the link between ethnic inequality and underdevelopment is robust.

[17] On average, the number of ethnic (linguistic) groups per country falls from 11 (39) to 4.2 (7). While the median number of groups per country across the 173 countries is eight (with both GREG and *Ethnologue*), when we drop groups consisting of less than one of a country's population, the medians fall to 3 (GREG) and 4 (*Ethnologue*). In contrast to the ethnic inequality measures, the spatial Gini and the administrative unit Gini coefficients do not get affected much when we drop small (in terms of population) pixels and administrative regions.

TABLE 3
ETHNIC INEQUALITY AND ECONOMIC DEVELOPMENT ETHNIC: INEQUALITY AND OTHER FEATURES OF ETHNOLINGUISTIC FRAGMENTATION, POLARIZATION, AND DIVERSITY

	ATLAS NARODOV MIRA (GREG)					ETHNOLOGUE				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ethnic inequality (Gini)	−1.3911*** (.3511)	−1.6116*** (.3876)	−1.3987*** (.3478)	−1.4369*** (.5274)	−1.3684*** (.3551)	−1.0716*** (.2932)	−.9620*** (.2942)	−1.0893*** (.2619)	−1.0477* (.5808)	−1.0108*** (.2901)
Ethnic-linguistic fragmentation	.0055 (.3549)					−.0061 (.2943)				
Cultural fragmenta-tion		−.3721 (.3486)					.0108 (.3637)			
Ethnolinguistic polarization			.4442 (1.0061)					.6216 (1.0004)		
Ethnic-linguistic segregation				−1.3294* (.7294)					−.1890 (.8974)	
Genetic diversity					183.0360** (84.5187)					195.3181** (81.1672)
Genetic diversity²					−130.1618** (60.2510)					−140.5393** (58.3610)

460

Spatial inequality (Gini)	−.0021	.2096	−.0335	.2086	.0615	−.1557	−.1692	−.1878	.1651	−.1015
	(.3541)	(.3717)	(.3540)	(.4144)	(.3407)	(.3046)	(.3086)	(.3072)	(.4348)	(.3038)
Adjusted R^2	.650	.684	.646	.731	.678	.650	.669	.646	.674	.675
Observations	173	150	172	96	157	173	150	172	92	157
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

NOTE.—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, based on the digitized version of the *Atlas Narodov Mira* (GREG) in cols. 1–5 and based on the *Ethnologue* in cols. 6–10. The overall spatial inequality index (Gini coefficient) captures the degree of spatial inequality across 2.5 × 2.5 decimal degree boxes/pixels in each country (boxes intersected by national boundaries and the coastline are of smaller size). Section II gives details on the construction of the ethnic inequality and spatial inequality (Gini) indexes. In columns 1 and 6, we control for ethnic and linguistic fragmentation using indicators that reflect the likelihood that two randomly chosen individuals in one country will not be members of the same group (the ethnic fragmentation index in col. 1 comes from Alesina et al. [2003] and the linguistic fragmentation index in col. 6 comes from Desmet et al. [2012]). In cols. 2 and 7, we control for cultural (linguistic) fragmentation using an index (from Fearon 2003) that accounts for linguistic distances among groups. In cols. 3 and 8, we control for ethnic polarization, using the Montalvo and Reynal-Querol (2005) index. In cols. 4 and 9, we control for ethnic and linguistic segregation, respectively, using the measures of Alesina and Zhuravskaya (2011). In cols. 5 and 10, we control for the genetic diversity index of Ashraf and Galor (2013). All specifications include regional fixed effects (constants not reported). The Data Appendix gives detailed variable definitions and data sources. Robust standard errors are reported in parentheses below the estimates.

 * Statistically significant at the 10 percent level.
 ** Statistically significant at the 5 percent level.
 *** Statistically significant at the 1 percent level.

461

TABLE 4
ETHNIC INEQUALITY AND ECONOMIC DEVELOPMENT IN 2000: ADDITIONAL CONTROLS AND ALTERNATIVE MEASURES OF ETHNIC INEQUALITY

	ATLAS NARODOV MIRA (GREG)			ETHNOLOGUE		
	All Ethnic Areas (1)	Excluding Capitals (2)	Excluding Small Groups (3)	All Ethnic Areas (4)	Excluding Capitals (5)	Excluding Small Groups (6)
Ethnic inequality (Gini)	−.9172***	−.6367*	−.9835**	−.7692***	−.6262**	−.9356**
	(.3287)	(.3266)	(.4503)	(.2902)	(.2909)	(.3955)
Spatial inequality (Gini)	−.5698	−1.2035***	−1.1254*	−.6952*	−1.3366***	−1.0495*
	(.3686)	(.3700)	(.6237)	(.3825)	(.3535)	(.6105)
Ethnic-linguistic fragmentation	.1409	.3185	.3793	.1075	−.0597	.1411
	(.3145)	(.3009)	(.3444)	(.2671)	(.2696)	(.2525)
Adjusted R^2	.724	.760	.736	.723	.759	.734
Observations	173	155	173	173	147	173
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Simple controls	Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes

NOTE.—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, based on the digitized version of the *Atlas Narodov Mira* (GREG) in cols. 1–3 and on the *Ethnologue* in cols. 4–6. The overall spatial inequality index (Gini coefficient) captures the degree of spatial inequality across 2.5 × 2.5 decimal degree boxes/pixels in each country (boxes intersected by national boundaries and the coastline are of smaller size). For the construction of the ethnic and the spatial inequality measures (Gini coefficients) in cols. 1 and 4, we use all ethnic (linguistic) homelands (and pixels); in cols. 2 and 5, we exclude ethnic areas (and pixels) where capital cities fall; in cols. 3 and 6, we exclude polygons (linguistic, ethnic, boxes) with less than 1 percent of a country's population. Section II gives details on the construction of the ethnic inequality and spatial inequality (Gini) indexes. In all specifications we control for ethnic-linguistic fragmentation using indicators reflecting the likelihood that two randomly chosen individuals in one country will not be members of the same group (the ethnic fragmentation index in cols. 1–3 comes from Alesina et al. [2003] and the linguistic fragmentation index in cols. 4–6 comes from Desmet et al. [2012]). In all specifications we include as controls log land area and log population in 2000 (simple set of controls), a measure 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 (geographic set of controls). All specifications include regional fixed effects (constants not reported). The Data Appendix gives detailed variable definitions and data sources. Robust standard errors are reported in parentheses below the estimates.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

jointly affect development and ethnic inequality, the estimates clearly point out that the correlation does not reflect (observable) mean differences in commonly employed geographical characteristics.

B. Inequality across Administrative Units, Ethnic Inequality, and Development

We now examine the relationship between ethnic inequality and comparative development, accounting for regional disparities across administrative units. In this regard, as described in Section II, we have constructed Gini coefficients reflecting inequality in lights per capita across first- and second-level administrative units. This variable is quite useful in many ways. First, as administrative units are well defined, the regional Ginis are easily interpretable. Second, examining the link between spatial inequality across administrative regions and development is interesting by itself. A vast literature that goes back at least to the work of Williamson (1965) has studied theoretically and empirically the interlinkages between development and spatial (regional) inequality. (See the reviews of Kanbur and Venables [2008] and Kim [2009] for recent works.) Third, since in some countries ethnic boundaries have formed the basis for the delineation of administrative units, we can directly test whether the strong cross-country correlation between inequality across ethnic homelands and GDP per capita reflects an inverse relationship between inequality across politically defined regions and comparative development.

Table 5 reports the results. Let us start with panel A, where we use Gini coefficients of regional inequality estimated across first-level administrative units. On average, there are 18 first-level administrative units in each country. Examples of first-level unit regions include the German *lander* (16); the US (50), Brazilian (27), and Indian (35) states; the Swiss cantons (26); and the Chinese provinces and autonomous regions (32). The coefficient on the administrative unit Gini index in the unconditional specification (in col. 1) is negative and highly significant (−1.60). This suggests that underdevelopment is characterized by large regional differences in well-being (or public goods provision). This is in accord with our earlier results (e.g., table 2, col. 2) showing a similar pattern when using the overall spatial inequality Gini. In columns 2 and 6 we include both the administrative unit and the ethnic inequality Ginis (using the *Atlas Narodov Mira* and *Ethnologue* mappings, respectively). Both inequality measures enter with negative and significant estimates (magnitude around −1). In columns 3 and 7 we control for ethnic and linguistic fractionalization (using the Alesina et al. [2003] and Desmet et al. [2012] measures, respectively). In line with our previous estimates, once we account for inequalities across ethnic (and now also across administrative)

TABLE 5
ETHNIC INEQUALITY, ADMINISTRATIVE UNIT INEQUALITY, AND ECONOMIC DEVELOPMENT

	Unconditional	Atlas Narodov Mira (GREG)					Ethnologue		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A. Inequality across Administrative Units (First-Level)									
Administrative unit inequality (Gini)	−1.6717***	−1.0139**	−1.0185**	−1.5074***	−1.2651***	−1.0080**	−1.0150**	−1.5177***	−1.2752***
	(.4004)	(.4188)	(.4230)	(.4096)	(.4307)	(.4446)	(.4486)	(.4333)	(.4494)
Ethnic inequality (Gini)		−1.0252***	−1.0401***	−1.0478***	−.8087**	−.8527***	−.8280***	−.7952**	−.6096**
		(.2611)	(.2765)	(.3557)	(.3197)	(.2487)	(.2856)	(.3158)	(.2941)
Ethnic-linguistic fragmentation			.0584	−.0206	.1337		−.0548	−.1103	.0239
			(.3479)	(.3387)	(.3062)		(.2783)	(.2736)	(.2572)
Adjusted R^2	.642	.667	.665	.686	.741	.667	.665	.686	.738
Observations	173	173	173	173	173	173	173	173	173
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Simple controls	No	No	No	Yes	Yes	No	No	Yes	Yes
Geographic controls	No	No	No	No	Yes	No	No	No	Yes

B. Inequality across Administrative Units (Second-Level)

Administrative unit inequality (Gini)	−.6095*	−.1579	−.1713	−1.0886***	−.6976*	−.2834	−.3099	−1.1967***	−.7995**
	(.3391)	(.3421)	(.3409)	(.3572)	(.3675)	(.3518)	(.3535)	(.3627)	(.3482)
Ethnic inequality (Gini)		−1.1068***	−1.0495***	−1.0863***	−.7882**	−.7086**	−.6176*	−.8881***	−.6873**
		(.3315)	(.3346)	(.3812)	(.3445)	(.2974)	(.3323)	(.3389)	(.3415)
Ethnic-linguistic fragmentation			−.2237	−.5142	.025		−.1827	−.2027	.0599
			(.4018)	(.3772)	(.3514)		(.3207)	(.2937)	(.2557)
Adjusted R^2	.677	.698	.696	.726	.773	.692	.690	.723	.773
Observations	135	135	135	135	135	135	135	135	135
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Simple controls	No	No	No	Yes	Yes	No	No	Yes	Yes
Geographic controls	No	No	No	No	Yes	No	No	No	Yes

NOTE.—The table (both panels A and B) 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, based on the digitized version of the *Atlas Narodov Mira* (GREG) in cols. 2–5 and on the *Ethnologue* in cols. 6–9. The administrative unit Gini index reflects inequality in lights per capita across administrative regions. In panel A we use first-level administrative units. In panel B we use second-level administrative units. Section II gives details on the construction of the ethnic inequality and spatial inequality (Gini) indexes. In specifications 3–5 and 7–9, we control for ethnic-linguistic fragmentation using indicators reflecting the likelihood that two randomly chosen individuals in one country will not be members of the same group (the ethnic fragmentation index in cols. 3–5 comes from Alesina et al. [2003] and the linguistic fragmentation index in cols. 7–9 comes from Desmet et al. [2012]). Specifications 4, 5, 8, and 9 include as controls log land area and log population in 2000 (simple set of controls). Specifications 5 and 10 include as controls a measure 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 (geographic set of controls). All specifications include regional fixed effects (constants not reported). The Data Appendix gives detailed variable definitions and data sources. Robust standard errors are reported in parentheses below the estimates.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

465

regions, there is no systematic link between ethnolinguistic fragmentation and development. In columns 4, 5, 8, and 9 we control for country size (log population and log land area) and the rich set of geographic features. The results remain intact. Across all permutations, both the ethnic inequality measure and the Gini index capturing inequality across first-level administrative units enter with negative and highly significant coefficients. The "standardized" beta coefficients that summarize in terms of standard deviations the change in the outcome variable (log of per capita GDP) induced by a one standard deviation change in the independent variables are comparable for the two inequality measures, around 0.20.

Table 5, panel B, reports similar specifications in which administrative-level inequality is estimated across second-level units. The Global Administrative Areas database does not report second-level administrative units for all countries; hence the sample drops to 135 (we mostly lose small countries, such as Singapore, Jamaica, and Swaziland). The results are similar if we assign to these countries the first-level administrative unit Gini coefficients. As the median (mean) number of such units is 110 (301), the respective Ginis are estimated using a very fine aggregation. Examples include the German (*regierungsbezirk*) government regions (40), the French *département* (96), and the Brazilian municipalities (5,503). The coefficient on the administrative region Gini index in column 1 is negative and significant at the 90 percent level; yet its magnitude is considerably smaller than the analogous one with the first-level administrative Gini index (-0.61). (The implied beta coefficient is -0.10.) The coefficient on the administrative region Gini drops considerably and loses its statistical significance once we include the ethnic inequality proxy (cols. 2 and 5) and condition on ethnolinguistic fragmentation (cols. 3 and 6). In contrast, the ethnic inequality measure retains its statistical and economic significance. The coefficient on the ethnic Gini is unaffected when we condition on size and geography (in cols. 4, 5, 8, and 9).

The evidence in table 5 reveals two important findings. First, in a large cross section of countries there is a clear negative association between economic performance and regional inequalities across first-level administrative units. This new (to the best of our knowledge) finding adds to the literature in urban economics and economic geography that studies the relationship between regional economic disparities and the process of development.[18] Second, and more important given our focus, the strong cross-country link between ethnic inequality and underdevelop-

---

[18] Note that because of the lack of comparable regional income data across countries, empirical works on spatial inequalities have mostly been country specific, and the few existing comparative studies have relied on small samples (e.g., Ezcurra 2013; Lessmann 2014).

ment does not capture the similarly negative association between GDP per capita and economic differences across politically defined spatial units.

C. Perturbing Ethnic Homelands

We now explore whether the pattern uncovered so far survives a horse race between ethnic inequality constructed using the original mappings and ethnic inequality based on slightly modified ethnic homelands.[19] Showing that our original ethnic inequality measures dominate the Gini index based on perturbed ethnic homelands would suggest that not only are the centroids of the groups correctly identified in the original maps but that also the specific boundaries delineated are more precise than the Thiessen-based ones. Effectively, this sensitivity check investigates how precisely drawn the groups' boundaries are in the underlying data sets.

Table 6 reports the results of the "horse race" regressions, examining the link between the log of per capita GDP and ethnic inequality, conditional on the perturbed ethnic homelands Gini index. Across all specifications the ethnic Gini index enters with a negative and significant estimate that is quite similar (around -0.9) to the more parsimonious specifications in tables 2–4. In contrast, the Gini index based on the perturbed ethnic areas (Thiessen polygons) enters with an estimate unstable and statistically indistinguishable from zero. It is perhaps instructive to point out that the perturbed linguistic homelands of *Ethnologue* seem to have little predictive power on GDP per capita beyond the role of ethnic inequality based on the *Ethnologue* homelands themselves, whereas for the case of GREG the perturbed ethnic inequality index enters with a (consistent) negative sign and has a moderate magnitude. This pattern is in line with the idea that the *Ethnologue* compared to GREG's mapping may have less measurement error since the former draws from a wealth of resources that are up to date and more precisely documented, unlike GREG, which derives from maps of the 1960s.

D. Further Robustness Checks

We have performed numerous sensitivity checks to investigate the robustness of the strong cross-country association between ethnic inequality and underdevelopment. We report and discuss in detail these robust-

---

[19] As explained in Sec. II, we are creating the modified groups' homelands generating two alternative sets of Thiessen polygons, one using as input points the centroids of the linguistic homelands according to the *Ethnologue* data set and the other using the respective centroids of the *Atlas Narodov Mira*. Thiessen polygons have the exact same centroids as the actual linguistic and ethnic homelands in the *Ethnologue* and GREG databases, respectively.

TABLE 6
ETHNIC INEQUALITY AND DEVELOPMENT CONDITIONING ON PERTURBED ETHNIC HOMELANDS

	ATLAS NARODOV MIRA (GREG)				ETHNOLOGUE			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ethnic inequality (Gini)	−.9510*	−.9566*	−.8652*	−.9276**	−1.4389***	−1.4365***	−1.2413***	−.8063*
	(.5004)	(.5074)	(.5174)	(.4383)	(.4075)	(.4115)	(.4159)	(.4166)
Perturbed ethnic inequality (Gini)	−.5371	−.5457	−.8223	−.2645	.3302	.3376	.0259	−.1845
	(.5258)	(.5262)	(.5469)	(.4829)	(.4640)	(.4718)	(.4950)	(.4410)
Ethnic-linguistic fragmentation		.0446	−.0107	.1614		−.0213	−.0147	.1510
		(.3521)	(.3498)	(.3159)		(.2962)	(.2962)	(.2656)
Adjusted R^2	.654	.652	.662	.721	.653	.650	.657	.718
Observations	173	173	173	173	173	173	173	173
Region fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Simple controls	No	No	Yes	Yes	No	No	Yes	Yes
Geographic controls	No	No	No	Yes	No	No	No	Yes

NOTE.—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, based on the digitized version of the *Atlas Narodov Mira* (GREG) in cols. 1–4 and on the *Ethnologue* in cols. 5–8. The perturbed ethnic inequality measures (Gini coefficients) capture the degree of spatial inequality across Thiessen polygons in each country that use as input points the centroids of the ethnic-linguistic homelands according to the *Atlas Narodov Mira* (in cols. 1–4) and to the *Ethnologue* (in cols. 5–8). 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 a point of any other polygon. In specifications 2–4 and 6–8, we control for ethnic-linguistic fragmentation using indicators reflecting the likelihood that two randomly chosen individuals in one country will not be members of the same group (the ethnic fragmentation index in cols. 2–4 comes from Alesina et al. [2003] and the linguistic fragmentation index in cols. 6–8 comes from Desmet et al. [2012]). Specifications 3, 4, 7, and 8 include as controls log land area and log population in 2000 (simple set of controls). Specifications 4 and 8 include as controls 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 (geographic set of controls). All specifications include regional fixed effects (constants not reported). The Data Appendix gives detailed variable definitions and data sources. Robust standard errors are reported in parentheses below the estimates.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

ness checks in the online supplementary appendix. Specifically, we show that the results are similar when (i) we do not include region fixed effects; (ii) we estimate ethnic Ginis without taking into account observations either from capitals or from small groups; (iii) we drop from the estimation (typically small) countries with just one ethnic or linguistic group; (iv) we use radiance-calibrated luminosity data to construct all inequality measures (so as to account for top coding in the lights data that occurs at the major urban centers); (v) we account for the resolution of population estimates at the grid level that are used to compile the inequality measures; (vi) we use inequality measures (based on lights) nonstandardized by population and control for inequality in the distribution of population across ethnic areas; (vii) we perform the analysis at various nodes of *Ethnologue*'s linguistic tree (this approach follows Desmet et al. [2012], who show that the impact of ethnic fractionalization on growth, public goods, and conflict depends on the level of linguistic aggregation); (viii) we try accounting for measurement error of the underlying mapping of groups estimating two-stage least-squares models that extract the common component of ethnic inequality from both *Ethnologue* and the *Atlas Narodov Mira*; (ix) on top of the rich set of geographic variables, we also condition on various historical controls; and (x) we drop iteratively from the estimation a different continent/region and focus within each region separately. The regional analysis reveals that the development-ethnic inequality nexus is nonexistent for countries in western Europe and North America and weak in Latin America. On the contrary, the association is especially strong within East and South Asia as well as for countries in the Middle East and North Africa.

IV. On the Origins of Ethnic Inequality

Given the strong correlation between ethnic inequality and underdevelopment, we have investigated the roots of inequality across ethnic lines.

A. *Historical* (*Colonial*) *Origins*

We started by examining the association between ethnic inequality and commonly used historical correlates of contemporary development. There is little evidence linking contemporary differences in well-being across ethnic groups to the legal tradition (La Porta et al. 1998); the conditions that European settlers faced at the time of colonization, as captured by settler mortality (Acemoglu et al. 2001) or precolonial population densities (Acemoglu, Johnson, and Robinson 2002); the share of Europeans in the population (Hall and Jones 1999; Putterman and Weil 2010); and border design and state artificiality (Alesina, Easterly, and Matuszeski 2011). For brevity, we report these results in the

online supplementary appendix.[20] These insignificant associations suggest that the strong negative correlation between ethnic inequality and development does not reflect the aforementioned aspects of history.

B.   Geographic Origins

Motivated by the insight of Michalopoulos (2012) that differences in land endowments gave rise to location-specific human capital, leading to the formation of ethnolinguistic groups, we investigated whether differences in geographic and ecological attributes play a role in explaining contemporary income disparities across ethnic lines. To the extent that land endowments shape ethnic human capital and affect the diffusion and adoption of technology and innovation (e.g., Diamond 1997), ethnic-specific inequality in the distribution of geographic features would manifest itself in contemporary differences in well-being across groups.[21]

To construct proxies of geographic inequality, we obtained georeferenced data on elevation, land suitability for agriculture, distance to the coast, precipitation, and temperature and calculated for each ethnic area the mean value.[22] We then derived Gini coefficients at the country level that reflect group-specific inequality in each of these (five) dimensions. We also estimated measures of the overall degree of inequality in geographic endowments, constructing for each of the five geographic traits spatial Gini coefficients across boxes of 2.5 × 2.5 decimal degrees and across administrative units.

Preliminary evidence.—In table 7 we explore the association between ethnic inequality (in lights per capita) and these measures of inequality in geographic endowments across ethnic homelands. Specifications 1 and 5 simply condition on region fixed effects. To isolate the ethnic-specific component, in columns 2 and 6 we include in the empirical model Gini coefficients capturing the overall degree of spatial inequality across each of these five traits, while in columns 3 and 7 we include Gini coefficients of inequality in the same five geographic features across first-level administrative units. In specifications 4 and 8 we include as controls the country averages of each of the five variables. In almost all permutations, all five ethnic Ginis enter with positive estimates; this suggests that ethnic-specific differences in geo-ecological endowments

---

[20] There is also no association between ethnic inequality and proxies of the inclusiveness of early institutions (Acemoglu et al. 2008) and state history (Bockstette, Chanda, and Putterman 2002).

[21] Language differences between groups are likely to exacerbate the limited mobility across ethnic homelands induced by the underlying differences in ethnic-specific human capital.

[22] In the previous draft of the paper, we also used information on the share of each ethnic area covered by water bodies (lakes, rivers, and other streams). The results are similar; we omit this variable because luminosity gets magnified over water areas because of bleeding-blooming.

TABLE 7
ON 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)	(7)	(8)
Land quality (Gini)	.3035**	−.0077	−.0918	.184	.4042***	.1913	.4271**	.3930***
	(.1348)	(.2191)	(.1864)	(.1414)	(.1132)	(.1784)	(.1924)	(.1230)
Temperature (Gini)	1.651	−9.1956	7.3784	3.0874	19.5600***	47.6529***	39.7024***	36.8859***
	(7.2462)	(11.1686)	(10.7068)	(8.0507)	(6.8608)	(12.3794)	(10.6988)	(10.1117)
Precipitation (Gini)	.3421*	.7845*	.7969**	.2916	.1884	.5606	.3762	.2878
	(.2034)	(.4227)	(.3445)	(.2259)	(.2636)	(.4862)	(.4119)	(.2620)
Distance to the coast (Gini)	.2852**	.3954**	.1589	.4640***	.1433	.1819	.0012	.3462**
	(.1123)	(.1801)	(.1631)	(.1458)	(.1388)	(.1659)	(.1994)	(.1332)
Elevation (Gini)	.5002*	.6844**	.9012***	.3019	.4413***	.3449	.6396***	−.0356
	(.2674)	(.3255)	(.3064)	(.2772)	(.1584)	(.2232)	(.2013)	(.1601)
Adjusted R^2	.450	.467	.493	.491	.583	.611	.617	.667
Observations	164	164	164	164	164	164	164	164
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Spatial	Admin. unit	Levels	No	Spatial	Admin. unit	Levels

NOTE.—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 cols. 1–4 and *Ethnologue* in cols. 5–8. To construct the inequality measures in geographic endowments across ethnic homelands, we first estimate the distance from the centroid of each ethnic homeland to the closest seacoast, average elevation, precipitation, temperature, and land quality for agriculture and then construct Gini coefficients capturing inequality across ethnic homelands in each of these geographic features for each country. In cols. 2 and 6, 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 seacoast, elevation, precipitation, temperature, and land quality for agriculture) estimated across 2.5 × 2.5 decimal degree boxes/pixels in each country (boxes intersected by national boundaries and the coastline are of smaller size). In cols. 3 and 7, we control for the regional inequality across administrative units in geographic endowments using the Gini coefficient of each of these features (distance to the coast, elevation, precipitation, temperature, and land quality for agriculture) estimated across first-level administrative units in each country. Columns 4 and 8 include as controls the mean values (for each country) of distance to seacoast, elevation, precipitation, temperature, and land quality for agriculture. All specifications include regional fixed effects (constants not reported). The Data Appendix gives detailed variable definitions and data sources. Robust standard errors are reported in parentheses below the estimates.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

translate into larger disparities in ethnic contemporary development. Depending on the specification details—GREG or *Ethnologue* mapping, whether we condition on the level of each geographical trait and regional inequality in each of the five geographic features—different Gini coefficients of geographic inequality enter with significant estimates. For example, in the specifications using the GREG mapping, the Ginis capturing inequality in elevation and proximity to the coast enter with significant estimates, while in the *Ethnologue*-based models the Gini indicators reflecting inequality in land quality for agriculture and temperature are the key correlates of ethnic inequality. Moreover, the controls capturing inequality across random pixels or administrative regions all enter with statistically insignificant estimates (coefficients not shown). Thus, while we cannot precisely identify which geographic feature(s) matters most, the message from table 7 is that differences in geography across ethnic regions translate into differences in contemporary ethnic inequality.

A composite index of inequality in geographic endowments.—We thus aggregate the five Gini indexes of ethnic inequality in geographic endowments via principal components. The use of factor analysis techniques is appropriate in our context because we have many variables (Gini coefficients) that aim at capturing a similar concept (with some degree of noise), namely, inequality in ethnic-specific geographic attributes. Moreover, we are not sure about which aspects of geographic inequality should matter the most. Table 8 reports the results of the principal component analysis. The first principal component explains approximately 60 percent of the common variance of the five measures of inequality in geographic endowments across ethnic homelands and close to 50 percent when we estimate Gini coefficients across pixels of (roughly) the same size and across first-level administrative units. The second principal component explains around 20 percent of the total variance, while jointly the other principal components explain a bit less than a fourth of the total variance. All five inequality measures load positively on the first principal component. This applies to all inequality measures (across ethnic and linguistic homelands, administrative regions, and boxes). The eigenvalue of the first principal component is greater than two in all permutations (one being the rule of thumb), while the eigenvalues of the other principal components are close to and less than one. We thus focus on the first principal component, which, given the significant positive loadings of all Gini coefficients, we label as "inequality in geographic endowments across ethnic homelands."

Inequality in geography across ethnic homelands and ethnic inequality.—In figures 10A and 10B we plot the baseline index of ethnic inequality (based on lights per capita) against the first principal component of inequality in ethnic-specific geographic endowments. There is a strong positive correlation for both mappings (around .55), suggesting that differences in

TABLE 8
PRINCIPAL COMPONENT ANALYSIS: INEQUALITY IN GEOGRAPHIC ENDOWMENTS

PRINCIPAL COMPONENT	EIGENVALUE	VARIANCE EXPLAINED	VARIABLE	FACTOR LOADINGS				
				1st Principal Component	2nd Principal Component	3rd Principal Component	4th Principal Component	5th Principal Component
A. Gini Coefficient GREG: All Groups								
1st	2.982	.596	Gini land quality	.460	−.133	−.572	.647	−.159
2nd	.844	.169	Gini temperature	.482	−.358	.388	.097	.693
3rd	.746	.149	Gini precipitation	.485	.025	−.463	−.741	.038
4th	.251	.050	Gini sea distance	.297	.918	.143	.145	.167
5th	.177	.035	Gini elevation	.482	−.106	.536	−.058	−.682
B. Gini Coefficient *Ethnologue*: All Groups								
1st	2.942	.588	Gini land quality	.432	−.423	.585	.496	−.215
2nd	1.062	.213	Gini temperature	.448	−.305	−.688	.319	.362
3rd	.551	.110	Gini precipitation	.505	−.250	.183	−.781	.201
4th	.260	.052	Gini sea distance	.353	.710	.274	.196	.509
5th	.184	.037	Gini elevation	.484	.403	−.275	−.067	−.724
C. Gini Coefficient: Spatial Inequality Index								
1st	2.711	.542	Gini land quality	.482	−.444	.259	.118	.700
2nd	1.149	.230	Gini temperature	.476	−.060	−.678	.519	−.203
3rd	.629	.126	Gini precipitation	.497	−.350	.360	−.281	−.650
4th	.313	.063	Gini sea distance	.306	.677	.504	−.439	−.042
5th	.199	.040	Gini elevation	.448	.468	−.301	−.667	.212
D. Gini Coefficient: Administrative Unit Inequality								
1st	2.314	.463	Gini land quality	.484	−.414	.428	−.084	.637
2nd	1.270	.254	Gini temperature	.482	−.063	−.640	.586	.100
3rd	.771	.154	Gini precipitation	.574	−.200	.253	−.089	−.748
4th	.403	.081	Gini sea distance	.181	.717	.492	.455	.059
5th	.242	.048	Gini elevation	.415	.520	−.318	−.659	.150

NOTE.—The table reports the results of the principal component analysis that is based on five measures (Gini coefficients) reflecting inequality in geographic endowments in distance to the coast, elevation, precipitation, temperature, and land quality for agriculture across ethnic-linguistic homelands (panels A and B), pixels of 2.5 × 2.5 decimal degrees (in panel C), and first-level administrative regions (in panel D). Column 1 reports the eigenvalue of each principal component, and col. 2 gives the percentage of the total variance explained by each principal component. The other columns give the factor loadings for each of the five principal components.

geography explain a sizable portion of contemporary differences in development across ethnic homelands.

In table 9 we formally assess the role of ethnic-specific geographic inequality, as captured by the composite index of inequality in geographic endowments across ethnic-linguistic homelands, on contemporary ethnic inequality.[23] Columns 1 and 5 show that the strong correlation illustrated in the figures is not driven by continentwide differences. In columns 2 and 6 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 pixels of 2.5×2.5 decimal degrees. Likewise, in specifications 3 and 7 we add the first principal component of the geographic inequality measures across first-level administrative regions. This has little effect on the coefficient of the ethnic inequality in geographic endowments, which retains its economic and statistical significance (at the 99 percent level). Moreover, the two proxies of the overall degree of spatial inequality in geography enter with small coefficients that also have the "wrong sign" and are not always statistically significant. In columns 4 and 8 we control for the level effects of geography, augmenting the specification with the country average values of elevation, precipitation, temperature, distance to the coast, and land suitability for agriculture. The composite index reflecting differences in geographic endowments across ethnic homelands continues entering with a positive and significant coefficient. The estimate with the *Ethnologue* mapping (0.12) implies that a one standard deviation increase in the inequality in geography across ethnic homelands index (1.74 points, from Zambia to Ethiopia) translates into a 20 percentage point increase in the ethnic inequality index (exactly as the difference in ethnic inequality between Zambia and Ethiopia, somewhat more than half a standard deviation; see table 1).

In the online supplementary appendix, we show that the link between ethnic inequality and inequality in geographic endowments across ethnic homelands prevails (i) when we compile cross-country composite indicators of inequality in geographic endowments across ethnic lines using a richer set of geographic variables, (ii) when we condition on contemporary

---

[23] In this as well as in the subsequent tables, we also report bootstrap standard errors that account for the fact that the key independent variable—inequality in geographic endowments across ethnic homelands—is a "generated" regressor (as it is a principal component capturing a geography factor; see Wooldridge 2002). Our bootstrap method works as follows. A random sample with replacement is generated from the full sample of countries. In this random sample, we extract the first principal component of the five Gini indicators that capture inequality in geography across ethnic lines on elevation, precipitation, temperature, distance to coast, and land quality. We then use this principal component (from the random sample) in the regression (where the dependent variable is ethnic inequality). This process is repeated 10,000 times. Table 9 gives the standard deviation of the coefficient estimates across all (10,000) replications (for a similar approach, see the recent study by Ashraf and Galor [2013]). As can be seen, bootstrap standard errors are very similar to standard heteroskedasticity-adjusted (White) standard errors.
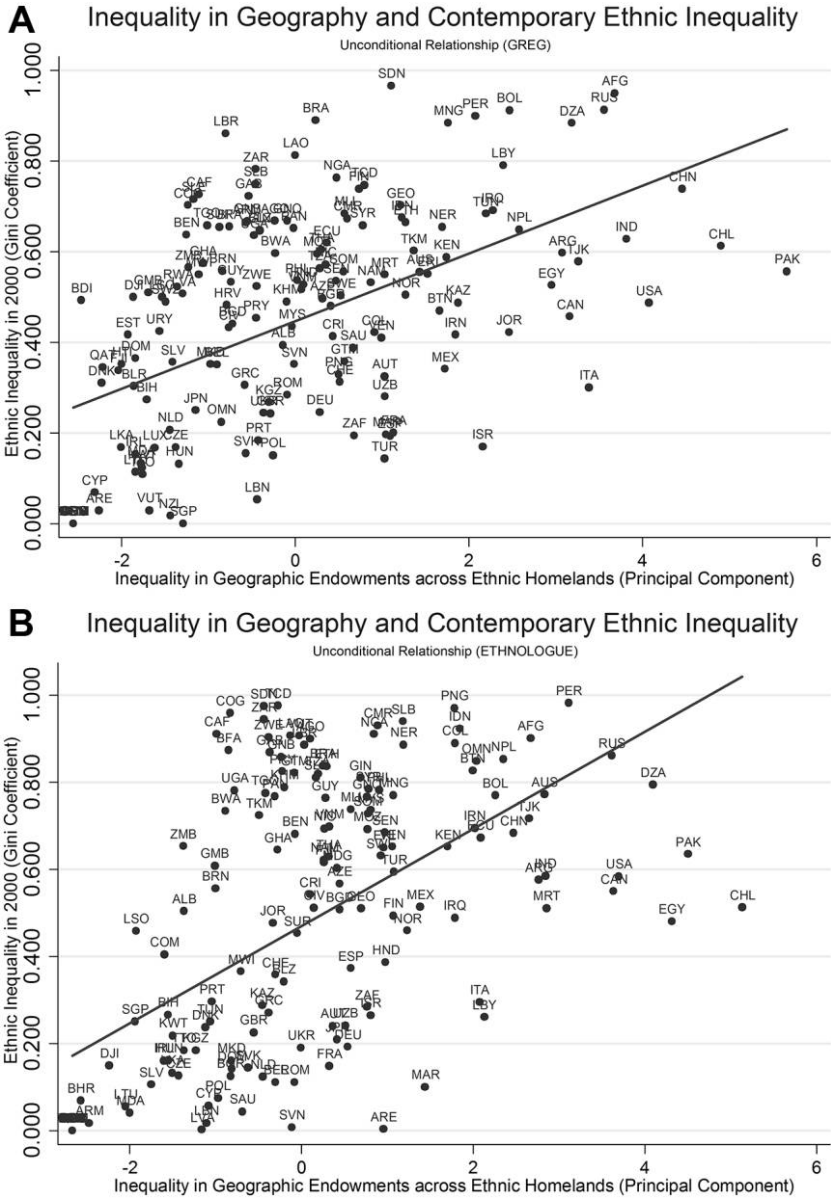
Fig. 10.—Panels *A* and *B* plot the baseline index of ethnic inequality (based on lights per capita) against the first principal component of inequality in ethnic-specific geographic endowments. Panels *C* and *D* plot the conditional on regional fixed effects association.
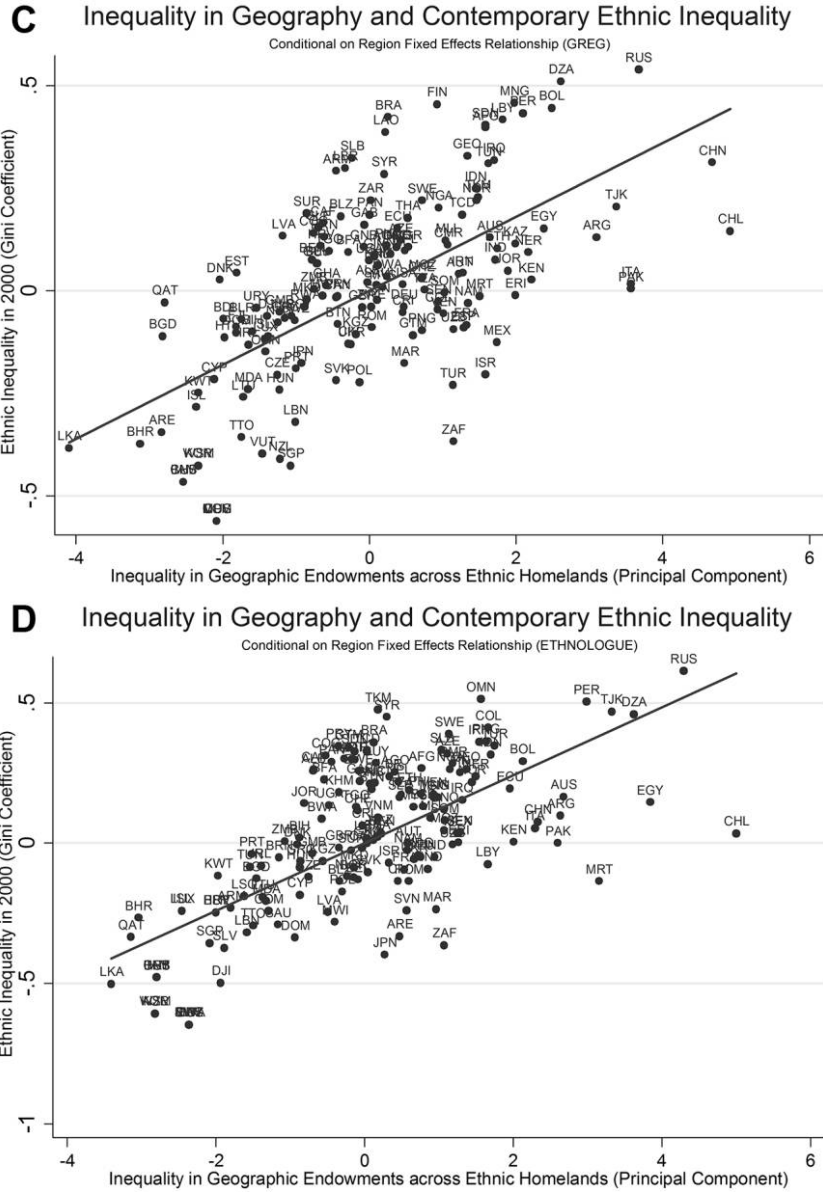
C Inequality in Geography and Contemporary Ethnic Inequality

Conditional on Region Fixed Effects Relationship (GREG)

D Inequality in Geography and Contemporary Ethnic Inequality

Conditional on Region Fixed Effects Relationship (ETHNOLOGUE)

Fɪɢ. 10 (*Continued*)

476

TABLE 9

ON 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)	(7)	(8)
Inequality in geography across ethnic homelands (principal components)	.0902***	.1053***	.1077***	.0898***	.1210***	.1458***	.1595***	.1181***
	(.0101)	(.0195)	(.0190)	(.0107)	(.0124)	(.0198)	(.0186)	(.0123)
	[.0102]	[.0192]	[.0196]	[.0108]	[.0118]	[.0200]	[.0192]	[.0124]
Spatial inequality in geography (principal components)		−.0195				−.0332*		
		(.0201)				(.0195)		
		[.0200]				[.01984]		
Inequality in geography across administrative units (principal components)			−.0251				−.0569***	
			(.0210)				(.0185)	
			[.0214]				[.0195]	
Adjusted R^2	.452	.453	.456	.483	.587	.594	.608	.656
Observations	164	164	164	164	164	164	164	164
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional controls	No	No	No	Levels	No	No	No	Levels

NOTE.—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 cols. 1–4 and *Ethnologue* in cols. 5–8. The composite index of inequality in geographic endowments is the first principal component of five inequality measures (Gini coefficients) measuring inequality across ethnic-linguistic homelands in distance to the coast, elevation, precipitation, temperature, and land quality for agriculture. The mapping of ethnic homelands follows the digitized version of GREG in cols. 1–4 and *Ethnologue* in cols. 5–8. Columns 2 and 6 include a composite index reflecting the overall degree of spatial inequality in geographic endowments. The composite index aggregates (via principal components) Gini coefficients across 2.5 × 2.5 decimal degree boxes/pixels in each country (boxes/pixels intersected by national boundaries and the coastline are of smaller size) of distance to the coast, elevation, precipitation, temperature, and land quality for agriculture. Columns 3 and 7 include a composite index reflecting regional inequality in geographic endowments across administrative units. The composite index aggregates (via principal components) Gini coefficients across first-level administrative units in distance to the coast, elevation, precipitation, temperature, and land quality for agriculture. Columns 4 and 8 include as controls the mean values (for each country) of distance to seacoast, elevation, precipitation, temperature, and land quality for agriculture. All specifications include regional fixed effects (constants not reported). The Data Appendix gives detailed variable definitions and data sources. Robust standard errors are reported in parentheses in brackets bootstrap standard errors that account for the use of a generated (principal component) regressor.

 * Statistically significant at the 10 percent level.
 ** Statistically significant at the 5 percent level.
 *** Statistically significant at the 1 percent level.

differences in development across space or administrative unions, and
(iii) when we iteratively drop different regions from the estimation.

V.   Geographic Inequality and Development

A.   Inequality in Geographic Endowments across Ethnic Homelands and Economic Development

Given the strong positive association between ethnic inequality and in-
equality in geographic endowments, it is interesting to examine whether
contemporary development is systematically linked to the unequal dis-
tribution of geographic endowments across ethnic homelands. We thus
estimated least-squares specifications associating the log of real GDP
per capita in 2000 with the composite index of ethnic-specific inequality
in geography (across the five geographic dimensions). While omitted-
variables concerns cannot be eliminated, examining the role of inequal-
ity in geographic endowments across ethnic homelands on comparative
development assuages 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 lin-
guistic homelands (compared to the ethnic inequality index based on
luminosity).

   Table 10 reports the results. The coefficient on the proxy of ethnic in-
equality in geographic endowments in columns 1 and 5 is negative
(around 0.13) and significant at the 99 percent confidence level. This
suggests that countries with sizable inequalities in geographic endow-
ments across ethnic homelands are, on average, less developed. In col-
umns 2 and 6 we condition on the overall degree of inequality in geog-
raphy with the spatial Gini index based on boxes, while in columns 3
and 7 we control for inequality in the geography across first-level admin-
istrative units. This allows examining whether the negative association
between development and geographic disparities across ethnic home-
lands—revealed in columns 1 and 5—captures the role of overall spatial
geographic inequalities, unrelated to ethnicity. The composite measures
capturing geographic inequalities across space and across administrative
regions enter with estimates statistically indistinguishable from zero (which
also have the opposite sign). In contrast, the composite index capturing in-
equality in geographic endowments across ethnic homelands retains its sta-
tistical and economic significance. These results further show that it is in-
equality across ethnic lines (in geography in this case) rather than across
space or administrative regions that correlates with underdevelopment.
The same applies when we control for the mean values of the five geo-
graphic variables (in cols. 4 and 8). The most conservative estimate implies
that a one standard deviation increase in geographic inequality across

TABLE 10
INEQUALITY IN GEOGRAPHIC ENDOWMENTS ACROSS ETHNIC HOMELANDS AND CONTEMPORARY DEVELOPMENT

	ATLAS NARODOV MIRA (GREG)				ETHNOLOGUE			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Inequality in geography across ethnic homelands (principal component)	−.1347***	−.2087***	−.1610**	−.0993***	−.1293***	−.1687**	−.1324**	−.1119****
	(.0383)	(.0754)	(.0695)	(.0367)	(.0432)	(.0708)	(.0659)	(.0364)
	[.0387]	[.0758]	[.0709]	[.0389]	[.0432]	[.0717]	[.0675]	[.0379]
Spatial inequality in geography (principal component)		.0954				.0527		
		(.0854)				(.0737)		
		[.0855]				[.0749]		
Inequality in geography across administrative units (principal component)			.0375				.0045	
			(.0903)				(.0764)	
			[.0907]				[.0776]	
Adjusted R^2	.635	.636	.633	.691	.633	.632	.630	.696
Observations	164	164	164	164	164	164	164	164
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional controls	No	No	No	Levels	No	No	No	Levels

NOTE.—The table reports cross-country OLS estimates, associating contemporary development with inequality in geographic endowments across ethnic homelands. The dependent variable is the log of real GDP per capita in 2000. The composite index of inequality in geographic endowments is the first principal component of five inequality measures (Gini coefficients) measuring inequality across ethnic-linguistic homelands in distance to the coast, elevation, precipitation, temperature, and land quality for agriculture. The mapping of ethnic homelands follows the digitized version of *Atlas Narodov Mira* (GREG) in cols. 1–4 and of *Ethnologue* in cols. 5–8. Columns 2 and 6 include a composite index reflecting the overall degree of spatial inequality in geographic endowments. The composite index aggregates (via principal components) Gini coefficients across 2.5 × 2.5 decimal degree boxes/pixels in each country (boxes/pixels intersected by national boundaries and the coastline are of smaller size) of distance to the coast, elevation, precipitation, temperature, and land quality for agriculture. Columns 3 and 7 include a composite index reflecting regional inequality in geographic endowments across administrative units. The composite index aggregates (via principal components) Gini coefficients across first-level administrative units in distance to the coast, elevation, precipitation, temperature, and land quality for agriculture. Columns 4 and 8 include as controls the mean values (for each country) of distance to the coast, elevation, precipitation, temperature, and land quality for agriculture. All specifications include regional fixed effects (constants not reported). The Data Appendix gives detailed variable definitions and data sources. Robust standard errors are reported in parentheses in the first row below the estimates. In the second row below the estimates we also report in brackets bootstrap standard errors that account for the use of a generated (principal component) regressor.
  * Statistically significant at the 10 percent level.
  ** Statistically significant at the 5 percent level.
  *** Statistically significant at the 1 percent level.

ethnic homelands (1.7 points) decreases income per capita by approximately 25 percent (0.22 log points).

In the online supplementary appendix, we show that the negative association between the log income per capita and inequality in geographic endowments across ethnic lines is present also when (i) we drop iteratively a different region and (ii) we control for contemporary differences in spatial development or inequality in lights per capita across administrative regions.

B. Geographic Inequality across Ethnic Homelands, Ethnic Inequality, and Economic Development

Given the strong negative correlation between development and ethnic inequality when the latter is proxied by differences in geographic endowments (table 10) or in disparities in luminosity per capita (tables 2–6), in table 11 we report specifications linking development to both measures. The results reveal that once we condition on contemporary ethnic income inequality, differences in geography across groups lose their power in explaining cross-country variation in development. While some peculiar type of measurement error may explain this finding, it indicates that ethnic-specific inequality in geographic endowments relates to contemporary development primarily via its influence on ethnic inequality.

Since geographic inequality across ethnic lines does not seem to exert an independent influence on GDP once we account for ethnic differences in well-being, we also estimated two-stage least-squares (2SLS) 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 with the log per capita GDP in 2000 in the second stage. While the 2SLS estimates do not identify causal effects, they account for measurement error in the proxy measure of development (lights per capita) and also isolate the geography-driven component of ethnic inequality. The 2SLS results (reported in the online supplementary appendix, table 23) reveal that the part of ethnic inequality that reflects geographic differences across ethnic homelands is a significant correlate of development.

Remark.—These results should not be interpreted as showing that unequal geography across ethnic lines necessarily "causes" ethnic inequality (and underdevelopment). It is possible that certain groups for a plethora of reasons (e.g., higher early development, superior military technology) conquered better-quality territories. In this regard the correlation between inequality in geographic endowments across ethnic lines and ethnic inequality (captured by lights per capita) indicates the sizable persistence of inequality. Hence, one might view an unequal ethnic geography as a manifestation of deeper ethnic differences. Nevertheless, even in this

INEQUALITY IN GEOGRAPHIC ENDOWMENTS ACROSS ETHNIC HOMELANDS, ETHNIC INEQUALITY, AND CONTEMPORARY DEVELOPMENT

	ATLAS NARODOV MIRA (GREG)				ETHNOLOGUE			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ethnic inequality (development)	−1.2984***	−1.2673***	−1.2959***	−1.0602***	−1.1603***	−1.1505***	−1.2258***	−.7805**
	(.3685)	(.3601)	(.3650)	(.3440)	(.3326)	(.3320)	(.3391)	(.3294)
	[.3732]	[.3646]	[.3730]	[.3491]	[.3342]	[.3363]	[.3444]	[.3387]
Inequality in geography across ethnic homelands (principal component)	−.0177	−.0753	−.0214	−.0041	.011	−.001	.0632	−.0197
	(.0495)	(.0719)	(.0661)	(.0455)	(.0560)	(.0752)	(.0737)	(.0552)
	[.0502]	[.0735]	[.0694]	[.0470]	[.0562]	[.0769]	[.0770]	[.0563]
Spatial inequality in geography (principal component)		.0707				.0145		
		(.0803)				(.0665)		
		[.0812]				[.0688]		
Inequality in geography across adminnstrative units (principal component)			.005				−.0653	
			(.0804)				(.0694)	
			[.0822]				[.0719]	
Adjusted R^2	.663	.662	.660	.707	.662	.660	.661	.705
Observations	164	164	164	164	164	164	164	164
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional controls	No	No	No	Levels	No	No	No	Levels

NOTE.—The table reports cross-country OLS estimates, associating contemporary development with ethnic inequality and inequality in geographic endowments across ethnic homelands. 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, based on the digitized version of the *Atlas Narodov Mira* (GREG) in cols. 1–4 and based on the *Ethnologue* in cols. 5–8. The composite index of inequality in geographic endowments is the first principal component of five inequality measures (Gini coefficients) measuring inequality across ethnic-linguistic homelands in distance to the coast, elevation, precipitation, temperature, and land quality for agriculture. The mapping of ethnic homelands follows the digitized version of GREG in cols. 1–4 and of *Ethnologue* in cols. 5–8. Columns 2 and 6 include a composite index reflecting the overall degree of spatial inequality in geographic endowments. The composite index aggregates (via principal components) Gini coefficients across 2.5 × 2.5 decimal degree boxes/pixels in each country (boxes/pixels intersected by national boundaries and the coastline are of smaller size) in distance to the coast, elevation, precipitation, temperature, and land quality for agriculture. Columns 3 and 7 include a composite index reflecting regional inequality in geographic endowments across administrative units. The composite index aggregates (via principal components) Gini coefficients across first-level administrative units in distance to the coast, elevation, precipitation, temperature, and land quality for agriculture. Columns 4 and 8 include as controls the mean values (for each country) of distance to the coast, elevation, precipitation, temperature, and land quality for agriculture. All specifications include regional fixed effects (constants not reported). The Data Appendix gives detailed variable definitions and data sources. Robust standard errors are reported in parentheses. In the first row below the estimates we also report in brackets bootstrap standard errors that account for the use of a generated (principal component) regressor.

* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.
*** Statistically significant at the 1 percent level.

case, it is the presence of an inherently unequal geography that partially allows these primordial ethnic differences to become salient (otherwise there would be no "better land" for stronger groups to conquer and every group would have the same land endowment).

VI.   Conclusion

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

First, combining linguistic maps on the spatial distribution of ethnic groups within countries with satellite images of light density at night, we construct Gini coefficients reflecting inequality in well-being across ethnic lines for a large number of countries. Ethnic inequality is weakly correlated with the standard measures of income inequality and modestly correlated with ethnolinguistic fractionalization, polarization, and segregation. Second, we show that the newly constructed proxy of ethnic inequality is negatively related to per capita GDP. The association retains its economic and statistical significance when we condition on inequality across administrative units, which is also inversely related to 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 lines that is correlated with poor economic performance rather than fractionalization per se. Third, we conduct a preliminary exploration of the roots of contemporary differences in well-being across ethnic groups. In this regard, we construct indicators of ethnic inequality in various geographic endowments and show that contemporary differences in development across ethnic homelands have a significant geographic component. Fourth, we show that geographic inequality across ethnic lines is also inversely related to contemporary development and that this correlation seems to operate via ethnic inequality.

Our study calls for future work on both the empirical and the theoretical fronts. One could employ our cross-country data and approach to examine the role of specific policies, such as trade openness and democratization, in shaping inequality across ethnic lines (and even administrative regions) over time. Furthermore, building on the literature on institutions (e.g., Acemoglu, Johnson, and Robinson 2005) and state formation

(e.g., Besley and Persson 2011), one could explore the role of initial—at independence—differences in standards of living across ethnic groups on the subsequent path of economic and political modernization. Future works should also employ within-country approaches that are more suitable for identifying the mechanisms at play. For instance, it is of great interest to understand the channels via which ethnic differences in income shape development. Does the link operate via the provision of public goods, via spurring conflict and animosity, or by shaping trust and beliefs? For example, in ongoing work (Alesina et al. 2014), we use a plethora of micro-level data from Africa to assess the role of between-group and within–ethnic group inequality on public goods provision, trust, and civic and political participation within (rather than across) countries. Moreover, given the large literature on inequality, fragmentation, and conflict, future work should explore in detail the role of ethnic-level income differences on conflict (as Mitra and Ray [2014] do in the case of Hindu-Muslim conflict in India). Another avenue of future research is to compile between-group inequality measures over time using detailed data from censuses, surveys, or tax records that are available for some developed countries, in the spirit of Atkinson, Piketty, and Saez (2011).

Appendix

Data

Country-Level Data

Income level: Log of real per capita GDP at purchasing power parity (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 et al. 2011).

Land area: Log surface area. Source: Nunn and Puga (2012).

Income inequality: Adjusted Gini coefficient index averaged over the period 1965–98. Source: Easterly (2007), based on the UN World Institute for Development Economics Research data.

Ethnic-linguistic fractionalization: Index of ethnic-linguistic heterogeneity. It reflects the probability that two randomly selected individuals belong to different ethnolinguistic/religious groups. For completeness we use two measures, one from Alesina et al. (2003), which in turn is based on the *CIA Factbook* and *Encyclopaedia Britannica*, and one from Desmet et al. (2012), which is based on *Ethnologue* (level 15).

Ethnic-linguistic segregation: Index ranging from zero to one capturing ethnic-linguistic segregation (clustering) within countries. If each region is composed of a separate group, then the index is equal to one; this is the case of complete segregation. If every region has the same fraction of each group as the country as a whole, the index is equal to zero; 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 (2011).

Ethnolinguistic polarization: Index of ethnolinguistic polarization that achieves its maximum score when a country consists of two groups of equal size. Source: Montalvo and Reynal-Querol (2005).

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

Genetic diversity: The expected heterozygosity (genetic diversity) of a country's contemporary population. The index is based on distances from East Africa to the year 1500 locations of the ancestral populations of the country's component ethnic groups in 2000 and on the pairwise migratory distances among these ancestral populations. The source countries of the ancestral populations are identified from the World Migration Matrix (Putterman and Weil 2010), and the modern capital cities of these countries are used to compute the aforementioned migratory distances. The measure of genetic diversity is then computed by applying (i) the coefficients obtained from regressing expected heterozygosity on migratory distance from East Africa at the ethnic group level, using a worldwide sample of 53 ethnic groups constituting the Human Genome Diversity Cell Line Panel, compiled by the Human Genome Diversity Project (HGDP) and the Centre d'Étude du Polymorphisme Humain (CEPH); (ii) the coefficients obtained from regressing pairwise genetic distance on pairwise migratory distance in a sample of 1,378 HGDP-CEPH ethnic group pairs; and (iii) the ancestry weights representing the fractions of the year 2000 national population that can trace their ancestral origins to different source countries in the year 1500. Source: Ashraf and Galor (2013).

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 meters of elevation difference for grid points 30 arc-seconds (926 meters 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).

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 et al. (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 the late nineteenth century. Source: Acemoglu et al. (2001).

Population density before colonization: Log of population density around the year 1500. Source: Acemoglu et al. (2002) and Nunn and Puga (2012).

Border straightness index: The 0-1 index reflects how straight—and thus likely to be nonorganic—national borders are. Source: Alesina et al. (2011).

Neolithic transition: The logarithm of the number of thousand years elapsed (as of the year 2000) since the majority of the population residing within a country's modern national borders began practicing sedentary agriculture as the primary mode of subsistence. This measure, reported by Putterman (2008), is compiled using a wide variety of both region- and country-specific archaeological studies as well as more general encyclopedic works on the transition from hunting and gathering to agriculture during the neolithic revolution. Source: Putterman and Weil (2010) and Ashraf and Galor (2013).

Ethnic partitioning: Percentage of the population of a country that belongs to partitioned ethnic groups. Source: Alesina et al. (2011).

Regional fixed effects: The region constants correspond to Southeast Asia and the Pacific, Latin America and the Caribbean, North America, western Europe, eastern Europe and central Asia, the Middle East and northern Africa, and sub-Saharan Africa. The classification follows the World Bank's World Development Indicators database.

Group-Level Data

Light density at night per capita: Light density is calculated by averaging luminosity observations across pixels that fall within each territory (ethnic-linguistic homelands, boxes of 2.5 × 2.5 decimal degrees, administrative units, and Thiessen polygons) and then dividing by population density. Source: National Oceanic and Atmospheric Administration, National Geophysical Data Center (http://ngdc .noaa.gov/eog/dmsp/downloadV4composites.html).

Population density: Average number of people per square kilometer for 1990 and 2000. Source: Center for International Earth Science Information Network, Columbia University, and Centro Internacional de Agricultura Tropical, 2005. Gridded Population of the World Version 3: Population Density Grids. Palisades, NY: Socioeconomic Data and Applications Center, Columbia University (http:// sedac.ciesin.columbia.edu/gpw).

Area: Total area (in square kilometers) of each territory (ethnic-linguistic homelands, boxes of 2.5 × 2.5 decimal degrees, administrative units, and Thiessen polygons).

Elevation: Average elevation above country minimum value in meters. Source: WorldClim—Global Climate Data. Data were originally collected by NASA-JPL SRTM. http://www.worldclim.org/current.

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 (http://www.sage.wisc.edu/iamdata/grid_data_sel.php).

Distance to the seacoast: The geodesic distance from the centroid of each country to the nearest coastline, measured in 1,000s of kilometers. Source: Global Mapping International, Colorado Springs, Colorado. Series name: Global Ministry Mapping System. Series issue: Version 3.0.

Average annual precipitation: Average annual precipitation (millimeters) for the approximate 1950–2000 time frame within the respective territory (ethnic-linguistic

homelands, boxes of 2.5 × 2.5 decimal degrees, administrative units, and Thiessen polygons). Source: WorldClim—Global Climate Data (http://www.worldclim.org/bioclim).

Average annual temperature: Average annual temperature for the approximate 1950–2000 time frame within the respective territory (ethnic-linguistic homelands, boxes of 2.5 × 2.5 decimal degrees, administrative units, and Thiessen polygons). Source: WorldClim—Global Climate Data (http://www.worldclim.org/bioclim).

Precipitation seasonality: Coefficient of variation of annual precipitation for the approximate 1950–2000 time frame within the respective territory (ethnic-linguistic homelands, boxes of 2.5 × 2.5 decimal degrees, administrative units, and Thiessen polygons). Source: WorldClim—Global Climate Data (http://www.worldclim.org/bioclim).

Temperature range: Range (estimated as the difference of the maximum value of the warmest month minus the minimum value of the coldest month) of annual temperature for approximately the period 1950–2000 within the respective territory (ethnic-linguistic homelands, boxes of 2.5 × 2.5 decimal degrees, administrative units, and Thiessen polygons). Source: WorldClim—Global Climate Data (http://www.worldclim.org/bioclim).

References

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *A.E.R.* 91 (5): 1369–1401.

———. 2002. "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution." *Q. J.E.* 107 (4): 1231–94.

———. 2005. "Institutions as a Fundamental Cause of Long-Run Growth." In *Handbook of Economic Growth*, edited by Philippe Aghion and Steven N. Durlauf, 109–39. Amsterdam: North-Holland.

Acemoglu, Daron, Simon Johnson, James A. Robinson, and Pierre Yared. 2008. "Income and Democracy." *A.E.R.* 98 (3): 808–42.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. 2003. "Fractionalization." *J. Econ. Growth* 8 (2): 155–94.

Alesina, Alberto, William Easterly, and Janina Matuszeski. 2011. "Artificial States." *J. European Econ. Assoc.* 9 (2): 246–77.

Alesina, Alberto, Johann Harnoss, and Hillel Rapoport. 2016. "Birthplace Diversity and Economic Prosperity." *J. Econ. Growth*, forthcoming.

Alesina, Alberto, and Eliana La Ferrara. 2005. "Ethnic Diversity and Economic Performance." *J. Econ. Literature* 43:762–800.

Alesina, Alberto, Stelios Michalopoulos, and Elias Papaioannou. 2014. "Inequality in Africa." Manuscript, Brown Univ., Harvard Univ., and London Bus. School.

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

Anderson, Siwan. 2011. "Caste as an Impediment to Trade." *American Econ. J.: Appl. Econ.* 3 (1): 239–63.

Ashraf, Quamrul, and Oded Galor. 2013. "The Out of Africa Hypothesis, Human Genetic Diversity, and Comparative Economic Development." *A.E.R.* 103 (1): 1–46.

Atkinson, Anthony B., Thomas Piketty, and Emmanuel Saez. 2011. "Top Incomes in the Long Run of History." *J. Econ. Perspectives* 49 (1): 3–71.

Baldwin, Kate, and John D. Huber. 2010. "Economic versus Cultural Differences: Forms of Ethnic Diversity and Public Goods Provision." *American Polit. Sci. Rev.* 104 (4): 644–62.

Benabou, Roland. 2005. "Inequality, Technology, and the Social Contract." In *Handbook of Economic Growth*, vol. 1, edited by Philippe Aghion and Steven Durlauf, 1595–1638. Amsterdam: North-Holland.

Besley, Timothy, and Torsten Persson. 2011. *Pillars of Prosperity: The Political Economics of Development Clusters*. Princeton, NJ: Princeton Univ. Press.

Bockstette, Valerie, Areendam Chanda, and Louis Putterman. 2002. "States and Markets: The Advantage of an Early Start." *J. Econ. Growth* 7 (4): 347–69.

Cederman, Lars-Eric, Nils B. Weidman, and Kristian Skrede Gleditch. 2011. "Horizontal Inequalities and Ethnonationalist Civil War: A Global Comparison." *American Polit. Sci. Rev.* 105 (1): 478–95.

Chen, Xi, and William D. Nordhaus. 2011. "Using Luminosity Data as a Proxy for Economic Statistics." *Proc. Nat. Acad. Sci.* 108 (21): 8589–94.

Chua, Amy. 2003. *World on Fire: How Exporting Free-Market Democracy Breeds Ethnic Hatred and Global Instability*. London: Heinemann.

Deshpande, Aswini. 2000. "Does Caste Still Define Disparity? A Look at Inequality in Kerala, India." *A.E.R. Papers and Proc.* 90 (2): 322–25.

Desmet, Klaus, Ignacio Ortuño-Ortín, and Romain Wacziarg. 2012. "The Political Economy of Ethnolinguistic Cleavages." *J. Development Econ.* 97 (2): 322–38.

Diamond, Jared. 1997. *Guns, Germs, and Steel: The Fates of Human Societies*. New York: Norton.

Easterly, William. 2007. "Inequality Does Cause Underdevelopment." *J. Development Econ.* 84 (2): 755–76.

Esteban, Joan, Laura Mayoral, and Debraj Ray. 2012. "Ethnicity and Conflict: An Empirical Study." *A.E.R.* 102 (4): 1310–42.

Esteban, Joan, and Debraj Ray. 2011. "Linking Conflict to Inequality and Polarization." *A.E.R.* 101 (4): 1345–74.

Ezcurra, Roberto. 2013. "Does Economic Globalization Affect Regional Inequality? A Cross-Country Analysis." *World Development* 52:92–103.

Fearon, James D. 2003. "Ethnic Structure and Cultural Diversity by Country." *J. Econ. Growth* 8:195–222.

Galor, Oded. 2011. "Inequality, Human Capital and the Process of Development." In *Handbook of the Economics of Education*, edited by Stephen Machin, Eric A. Hanushek, and Ludger Woessmann. Amsterdam: Elsevier.

Gordon, Raymond G., Jr., ed. 2005. *Ethnologue: Languages of the World*. 15th ed. Dallas: SIL Internat.

Hall, Robert E., and Charles I. Jones. 1999. "Why Do Some Countries Produce So Much More Output per Worker than Others?" *Q.J.E.* 114 (1): 83–116.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil. 2012. "Measuring Economic Growth from Outer Space." *A.E.R.* 102 (2): 994–1028.

Heston, Alan, Robert Summers, and Bettina Aten. 2011. "Penn World Table Version 7.0." Center Internat. Comparisons of Production, Income, and Prices, Univ. Pennsylvania.

Hodler, Roland, and Paul A. Raschky. 2014. "Regional Favoritism." *Q.J.E.* 129 (2): 995–1033.

Horowitz, Donald L. 1985. *Ethnic Groups in Conflict*. Berkeley: Univ. California Press.

Huber, John D., and Laura Mayoral. 2013. "Inequality, Ethnicity and Civil Conflict." Working paper, Dept. Polit. Sci., Columbia Univ.

Huber, John D., and Pavithra Suryanarayan. 2013. "Ethnic Inequality and the Ethnification of Political Parties: Evidence from India." Working paper, Dept. Polit. Sci., Columbia Univ.

Kanbur, Ravi, and Anthony J. Venables. 2008. "Spatial Inequality and Development—Overview of UNU-WIDER Project." Manuscript, Cornell Univ.

Kim, Sukkoo. 2009. "Spatial Inequality and Economic Development: Theories, Facts, and Policies." In *Urbanization and Growth*, edited by P. Annez, R. Buckley, and M. Spence, 1–39. Washington, DC: Internat. Bank Reconstruction and Development and World Bank.

Kuznets, Simon. 1955. "Economic Growth and Income Inequality." *A.E.R.* 45 (1): 1–28.

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

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert Vishny. 1998. "Law and Finance." *J.P.E.* 106 (6): 1113–55.

———. 1999. "The Quality of Government." *J. Law, Econ., and Org.* 15 (1): 222–79.

Lessmann, Christian. 2014. "Spatial Inequality and Development—Is There an Inverted-U Relationship?" *J. Development Econ.* 106 (3): 35–51.

Loury, Glenn C. 2002. *The Anatomy of Racial Inequality.* Cambridge, MA: Harvard Univ. Press.

Michalopoulos, Stelios. 2012. "The Origins of Ethnolinguistic Diversity." *A.E.R.* 102 (4): 1508–39.

Michalopoulos, Stelios, and Elias Papaioannou. 2013. "Pre-colonial Ethnic Institutions and Contemporary African Development." *Econometrica* 81 (1): 113–52.

———. 2014. "National Institutions and Subnational Development in Africa." *Q.J.E.* 129 (1): 151–213.

Mitra, Anirban, and Debraj Ray. 2014. "Implications of an Economic Theory of Conflict: Hindu-Muslim Violence in India." *J.P.E.* 122 (4): 719–65.

Montalvo, Jose-Garcia, and Marta Reynal-Querol. 2005. "Ethnic Polarization, Potential Conflict, and Civil Wars." *A.E.R.* 95 (3): 796–816.

Nunn, Nathan, and Diego Puga. 2012. "Ruggedness: The Blessing of Bad Geography in Africa." *Rev. Econ. and Statis.* 94 (1): 20–36.

Pinkovskiy, Maksim. 2013. "Economic Discontinuities at Borders: Evidence from Satellite Data on Lights at Night." Manuscript, Massachusetts Inst. Tech.

Pinkovskiy, Maksim, and Xavier Sala-i-Martin. 2014. "Lights, Camera, . . . Income! Estimating Poverty Using National Accounts, Survey Means, and Lights." Working Paper no. 19831, NBER, Cambridge, MA.

Putterman, Louis. 2008. "Agriculture, Diffusion and Development: Ripple Effects of the Neolithic Revolution." *Economica* 75 (300): 729–48.

Putterman, Louis, and David N. Weil. 2010. "Post-1500 Population Flows and the Long-Run Determinants of Economic Growth and Inequality." *Q.J.E.* 125 (4): 1627–82.

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

Stewart, Frances. 2002. "Horizontal Inequalities: A Neglected Dimension of Development." Annual Lecture 5, World Inst. Development Econ. Res, United National Univ., Helsinki.

Tesei, Andrea. 2014. "Trust, Racial Fragmentation, and Income Inequality: New Evidence from the US." Manuscript, Queen Mary Coll., Univ. London.

Weidmann, Nils B., Jan Ketil Rod, and Lars-Erik Cederman. 2010. "Representing Ethnic Groups in Space: A New Dataset." *J. Peace Res.* 47 (4): 491–99.

Williamson, Jeffrey G. 1965. "Regional Inequality and the Process of National Development: A Description of the Patterns." *Econ. Development and Cultural Change* 13 (4): 1–84.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press.