Strength in Numbers: Ethnic Group Size and Economic Development

Victoria Gerenutti Klarosk* Rogerio Santarrosa†

February 28, 2023

Abstract

We use the partition of ethnic groups across adjacent countries, caused by the Scramble for Africa, to study the effect of group size on economic development. The arbitrariness of the border design caused exogenous variation in the population size of split groups within countries and thus works as a natural experiment. We combine information on the historical homeland of ethnic groups, gridded geo-referenced estimates of population on the eve of colonization, and a measure of ethnic contemporary economic performance, as proxied by satellite images of light density at night. Using within group variation plus country fixed effects, we find that larger groups are currently more developed. We exploit several channels that can potentially mediate these effects. We show that these results are driven by the number of people and not simply by the amount of natural resources. Larger groups are more urban, more likely to have big cities in their homelands, and have more access to transportation infrastructure. We use contemporary African Demographic Surveys and find that individuals living in homelands of larger groups are more educated, have moved from agriculture to service sector, and are more likely to hold permanent working positions and skilled jobs. Finally, we also find that large groups are more likely to be politically organized and represented in the highest executive political positions in the country.

Keywords: Population size, economic development, ethnicity, agglomeration effects, political power, ethnic diversity, Africa
JEL classification: N00, N37, O10, O43, O55

*Insper Institute of Education and Research. Email: victoriagk@al.insper.edu.br
†Insper Institute of Education and Research. Email: rogeriobs2@insper.edu.br

1 Introduction

There is a strong association between ethnic diversity and Africa's underdevelopment (e.g., Easterly and Levine (1997), Alesina and La Ferrara (2005)). However, there are still significant challenges in interpreting this relationship as causal. Most of these empirical patterns have been established in cross-country analyses and must assume that ethnic fractionalization is not correlated to any other (non-observable) determinant of national economic performance. This concern is reinforced by recent evidence that ethnic classifications have been shaped by natural endowments, as land quality (Michalopoulos, 2012), and historical events, such as the slave trade (Nunn, 2008), which in turn affect development through other channels.

In this paper, we investigate a different dimension, and in some degree an implied consequence of this phenomenon. We test whether ethnic group size matters for its economic development. If ethnic diversity has in fact a negative effect on aggregate outcomes, it is expected that small groups would perform relatively worse than their larger counterparts.

There are several reasons (which we cover in detail in the next section) why larger groups might have an economic advantage over small ones. Many of them are hypothesized as fundamental causes of development for which there is no conclusive evidence in the literature. For instance, differences in cultural traits, language barriers, or taste preferences can impose obstacles to economic transactions between individuals from different groups. These barriers can go from mundane language misunderstandings to taste discrimination. In such cases, individuals from large groups may benefit from a higher likelihood of facing economic transactions with a similar fellow. Commercial transactions, transmission of information, coordination problems, division of labor, trust and other interactions can be facilitated by these similarities.

Under the presence of ethnic barriers, large groups will have advantage in benefiting from agglomeration effects. Several economic growth models highlight the relationship between local productivity and population size (Peters, 2022). These agglomeration spillovers can be

driven by several sources: for instance, labor market density, fixed local transportation costs, knowledge spillovers, local variety creation and increasing returns to scale in manufacturing (Greenstone et al., 2010; Peters, 2022). Political economy factors can also play a role in an ethnic diverse environment. If political competition follows ethnic lines (which has been extensively documented in Africa), large groups can become politically more powerful. In turn, they will be more able to allocate resources and public goods to favor their own group.

To establish the causal effect of group size on economic development, we use a historical natural experiment: the Scramble for Africa. This event started with the Berlin Conference of 1884/85 and was completed by the turn of the 20th century, when Europeans drew (artificially) colonial African borders before hardly settling in the continent. Considerable evidence indicates that the borders were designed with little information or regard for the geographic distribution of native institutions already in place in the continent. In fact, 80% of African borders has followed latitudinal and longitudinal lines (Alesina et al. 2011). This caused more than 200 ethnic homelands to be partitioned across two or more countries. In practice, this made a larger share of one split group (higher population) fall in one country while a lower share of the same tribe became part of another state. Intuitively, our identification strategy compares the economic outcomes of these two parts to establish the causal effect of group size.

Michalopoulos and Papaioannou (2016) relies on a similar identification strategy to test the effect of ethnic partition (comparing split versus non-split groups). They provide compelling evidence that country boundaries were set exogenously to the local political geography of the region by testing an extensive set of pre-colonial characteristics across split and non-split ethnic groups. The results support the largely arbitrary nature of the borders' drawing, at least with respect to ethnic organization. Here instead we use within group variation also caused by the same natural experiment.

We use several data sources to implement our empirical design. First, we use the Ethnolinguistic Map constructed by Murdock (1967) that depicts the spatial distribution of the

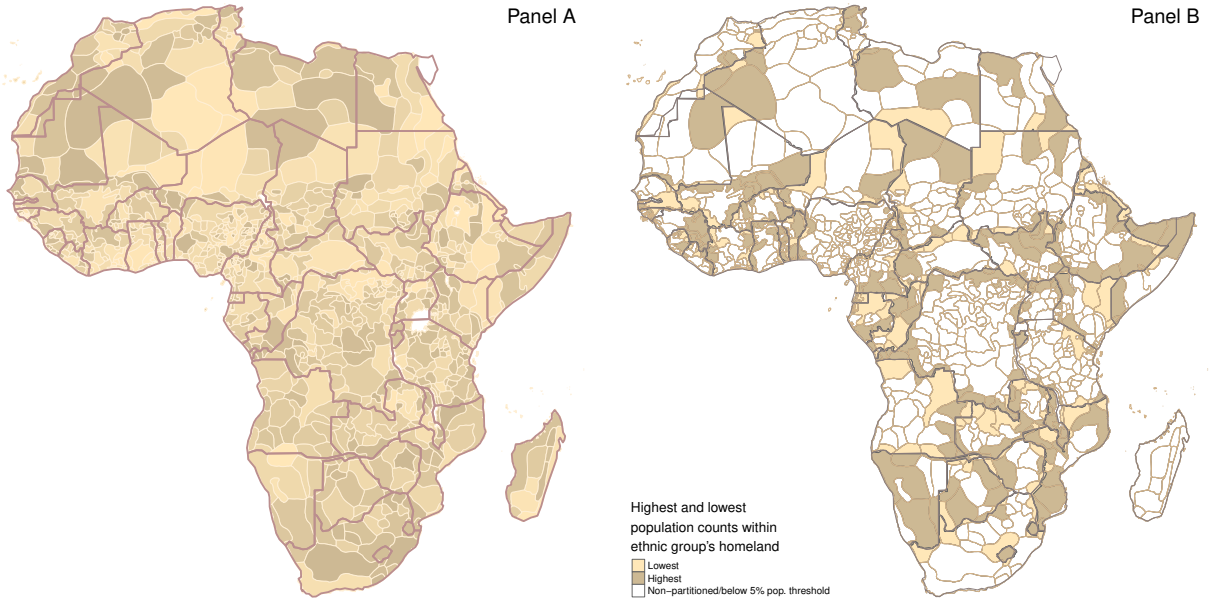
continent's ethnic groups at the time of the colonization, building on sources covering the period 1860-1940. We overlay Murdock's map with current country borders to obtain a map at the country-tribe level. Second, we use gridded historical population data (HYDE) to construct estimates of population size at each country-tribe in 1880. Importantly, these are estimates of the population just before the national borders were defined. Figure 1 shows the overlap of national borders and the pre-existing tribes (Panel A). Panel B specifically highlights the division of groups across countries in segments with higher versus lower population count. Third, since no granular, comparable measures of economic development are available for the whole African continent, we proxy economic activity by average light density at nighttime (Michalopoulos and Papaioannou 2013, Michalopoulos and Papaioannou 2014, Alesina et al. 2016).

The arbitrariness of the border design creates a natural experiment which allows us to assess how the exogenous size of split ethnic groups affects their development. For this, we leverage within-ethnicity cross-country variation in population size, as of 1880 (pre-scramble), while holding country and ethnic characteristics constant – our main specification includes country fixed effects[1] and ethnic fixed effects. This gives the causal intention-to-treat effect of ethnic group size on development. In our preferred specification, we find that a 10% growth in the group size increases light density by 2%. These estimates are economic meaningful. Had an ethnic group that were split equally in half kept untouched, their ethnic-country population would be 100% larger. As a result, their development level, measured by light index, would be 20% larger. This result is robust to a battery of geography and agriculture controls.[2]

[1]We are able to include country fixed effects because several split ethnic groups might be in a single country

[2]Geography controls are (log) distance to coast, indicator for being landlocked, terrain ruggedness index, average slope, average precipitation, and malaria suitability index. Agriculture controls are share of irritated area in 1880, share of grazing area in 1880, share of pasture area in 1880, share of cropland area in 1880, mean altitude, maximum altitude, crop suitability index, land suitability level, share of water sites, dominant soil, agro-zones, and hydro-basin class.

Figure 1: Traditional ethnic homelands, country borders, and set of partitioned ethnic groups



Note: Panel A displays the overlay of the ethnic groups' homelands in Murdock (1959, 1967), as colored areas, and current country borders, as lines. Panel B emphasizes the tribes identified as split according to the details discussed in section 4.1, and uses two shades to distinguish parts with highest and lowest population sizes within split ethnic groups. Areas of tribes partitioned into more than one country have lighter shade if their population is lower than the median population count, and darker if it is greater.

Next we assemble data from a variety of sources to examine potential channels through which group size can affect economic development. First, we examine the role of natural endowments. A larger group might be more likely to have access to natural resources which can facilitate agricultural production (water) or even mines/oil fields that can boost some economic activity. We do find that larger groups are more likely to have lakes and gold mines in their territory. In contrast, we do not find a higher likelihood of having oil fields, diamond fields or other gem mines. We investigate whether our effects are driven by the higher number of people or the increased amount of natural resources. Our effects are remain unchanged when accounting for natural resources as controls in our regressions. For additional robustness, throughout the reminder of the paper, we always control for the area of the homeland. If anything, the estimated coefficient on population is larger, and reveals that the effects are driven by the number of people and not by the larger areas in which big groups inhabit.

We also investigate the effect of group size on the dispersion of luminosity within ethnic

groups' homeland. We find that our effects are driven by the top percentiles of the light distribution. We also find significant effects on several measures of light dispersion, consistent with the development impact being driven by a few clusters in the homeland area.

To fully investigate the potential role of these top luminosity clusters, we use a multitude of geo-referenced data related to the formation of urban agglomerations. We show that larger groups have a greater share of urban population. Their homelands are also significantly more likely to hold urban clusters (measured by the continuity of buildings in areas occupied by more than 10,000 inhabitants), a metropolitan area, and also the country's capital. We also find that the average travel time to the closest city or health facility is lower across their territory. This is relevant since economic mobility in Africa is highly associated with migration to urban centers (mostly comprised by the capital). Moreover, distance to cities, and to the country's capital in particular, affects access to markets which can be used either to buy important inputs (e.g fertilizers) or sell agricultural surplus.

Further, we validate these patterns using the Demographic and Health Surveys (DHS) for African countries. We collect the most recent survey of each country for which we have the GPS of the respondent available. We overlay the GPS information into our country-tribe shapefile to identify the ethnic homeland location of each individual. Approximately, half of our sample of split ethnic groups are represented by some DHS survey. First, we validate that the significant effects of population size on luminosity hold for this subsample. Second, we test the effect of the ethnic population size on several household characteristics. We find significant impact on the likelihood of the household being urban and increase in the household wealth index, driven by both the ownership of private assets (e.g, television, bicycle, mobile phone) and public goods (e.g, water access and sanitation facilities). Third, we use individual-level data to find some evidence on structural economic transformation generated by these economies of scale and urbanization effects. We find that both men and women attain more years of education. Women, in particular, despite not changing the probability of working, are

more likely to hold permanent jobs. Both men and women are less likely to work in agriculture and more likely to work on sales and services. Men, in particular, are also more likely to have a skilled manual occupation. We corroborate and complement these findings using the round 7 of Afrobarometer. We find that large groups have better access to public goods such as electricity, transportation, piped water and sewage systems. On the other hand, we find no effects on school and health clinic units. In terms of social attitudes, individuals from larger groups have a better perception of the state capabilities, trust less in informal institutions, are better informed, and hold more equal gender norms.

In addition, we provide some evidence on infrastructure investment. We overlay our country-tribe map with several shapefiles on types of roads and railways. We find that larger groups are in fact more connected with roads and railways. These effects are consistent the larger number and greater relevance of cities in their territories. This result is also consistent with greater political capacity of such groups in obtaining public investments.

Finally, we combine Murdock's tribes to the Ethnic Power Relations Dataset (EPR). This last dataset consists of a selected list of politically relevant groups. They are groups for which there is at least one organization claiming to representing them at national politics. EPR also has information on the degree of political power that these groups possess. We find that larger groups are more likely to become politically relevant, and as a result, occupy dominant positions in the national government.

This paper is structured as follows. In the next section, we review part of the literature on the links between ethnic group size and economic development, focusing on papers that have investigated the African context. In section 3, we discuss the historical background of the Scramble for Africa. In section 4, we describe the data sources and the way we have used them to carry out our analysis. Section 5 shows the empirical strategy and the main results are presented in Section 6. The potential channels are investigated in Section 7. We end with concluding remarks in Section 8.

2 Related literature

Setting aside country and institutional specifics, why would a group benefit from being a majority? Or, equivalently, what kind of obstacles does a minority group face that could harm its economic success?

An initial explanation is borrowed from the Social Psychology literature and relates to the social identity theory. Previous work has argued that patterns of intergroup behavior are consistent with individuals attributing positive utility to the well-being of members of their own group and negative utility to that of other groups (Tajfel et al. 1971). This could imply that minorities engage less in community social activities due to their status and become segregated (Alesina and La Ferrara 2000). But beyond the explicit *taste* for interacting primarily with members of one's group, group affiliation may be important for trade in the presence of market imperfections. Greif (1993) points that, in Medieval times, alliances along ethnic lines were common across traders in order to exchange information about opportunistic behavior of others. Group membership according to ethnicity, therefore, may have helped sustain a reputation mechanism (Greif 1993). Put differently, ethnic affiliation may be instrumental for constructing *trust*.

High level of trust may be mapped to various desirable economic outcomes. When individuals trust each other, transaction costs are reduced, organizations run better, the need for formal regulation reduces, governments provide services more efficiently, policy promises become more credible, and financial systems develop better (Finseraas et al. 2019). Moreover, lack of trust in out-group individuals may penalize proportionally more minorities if it is detrimental to inter-group interaction. At the workplace, co-ethnicity is linked to efficiency in contact across employees (Hjort 2014), and thus minority employees may be less productive due to the need to interact more frequently with out-group individuals. By the same reason, minorities can be more penalized if lack of trust deters technology transmission across groups. This is because

a larger group of individuals potentially innovates more simply due to a mechanic effect of its size and complexity.

Although *taste* and *trust* may be important mechanisms liking group size and economic activity, cultural traits, such as norms, religion, and language, could have similar effects on contact, by obstructing out-group and facilitating in-group interaction across individuals. Additionally, Laitin and Ramachandran (2022) shows that individuals have more years of education and higher reading proficiency when their indigenous language is the language of instruction at school in African countries. Evidently, if minorities have lower bargaining power to choose the language of instruction at schools, they will be disproportionately affected by this channel. Although taste, trust and cultural traits may have similar effects on interaction, and may actually be very entangled, trust and group identity related to taste may be more easily malleable (Lowe 2021, Depetris-Chauvin et al. 2020), Finseraas et al. 2019) than cultural practices. By identifying this distinction, one can argue by which of these channels group size has greatest effect on economic development.

Until this point we have focused on explanations that favor an *economic mechanism* for linking the two objects. But the relationship between them may be complicated by concurring and complementary explanations related to *political capital*. Comparative politics works have pointed that ethnic favoritism in politics exist throughout the world and despite of the system of government (De Luca et al. 2018), but is particularly important in African countries (Dickens 2018). Politicians often engage in redistribution towards their own personal homelands or co-ethnics through public services investments, such as in educational inputs (Kramon and Posner 2016) and roads (Burgess et al. 2015).

This political channel may impact group development if its size influences its access to formal politics. Francois et al. (2015) show that political coalitions in Africa share ministerial positions highly proportionally to ethnic groups' population sizes, as elites bargain over political power. Therefore, large ethnic groups that have elite members as part of the government

coalition may be favored by redistribution or targeted public goods.

Finally, if minority groups face more obstacles to be formally represented and to have their preferences met by institutional politics, they may also rely more on customary institutions. This dynamic, in turn, may affect property rights and the rule of law. While these may have both positive and negative consequences for development - for example, by relying in informal institutions, individuals may have more trust in each other, but also lack sufficient rule of law and juridical complexity to foster innovation -, this is still an open research question.

3 Historical background

In the 1860s the French and British started systematically exploring West Africa. This action was followed by 40 years of Europeans signing treaties and agreements to partition the largely unexplored African continent into free-trade zones, protectorates, and colonies. The Berlin Conference defines a key moment during this historical process. It took place from November 1884 until February 1885, and although its concrete goal was to set the boundaries of Central Africa, it is regarded among historians and political scientists as the landmark of ethnic partitioning in the continent, as it set down the guidelines that would be used to delimit the zones.

Previous works have discussed this event in great detail. For instance, Michalopoulos and Papaioannou (2016) argue that the key principle followed by the European powers was to preserve the *status quo* and prevent conflict over Africa, which resulted in borders being designed with little regard for local conditions. In fact, many drawing decisions were made before information from explorers, geographers, and missionaries could arrive to Europe (Michalopoulos and Papaioannou 2016),[3] while in the Berlin conference "there was no African representation,

[3]This is evident in a speech delivered by British prime minister Robert Cecil in 1890:

> "*We have been engaged in drawing lines upon maps where no white man's feet have ever tord; we have been giving away mountains and rivers and lakes to each other, only hindered by the small impediment that we never knew exactly where the mountains and rivers and lakes were.*"(Times, August 7, 1890)

and African concerns were, if they mattered at all, completely marginal to the basic economic, strategic, and political interests of the negotiating European powers"(Asiwaju, 1985, p. 25).

Several explanations surfaced in African historiography to rationalize the largely accidental border design. First, it is emphasized that the continent had been little explored by Europeans at the time of the partitioning, and it was not uncommon for these powers to swipe pieces of land back and forth with limited idea of what they were worth. Similarly, neither locals or Europeans were foreseeing that these zones would became *states*: they were simply delimiting regions of influence. Moreover, demarcation on the ground was poor, and where it existed, locals could freely move across colonial borders in a way that resistance from the bottom does not seem to have been relevant or to have decidedly and systematically shaped the borders. Motivated by these accounts, Michalopoulos and Papaioannou (2016) run a battery of regressions of an ethnic partitioning indicator (as consequence of border design) over a set of geographical, ecological, and some documented pre-colonial characteristics. It shows that there is no systematic association between these variables and the probability of an ethnic group being partitioned into two or more countries, which gives a formal foundation to the argument of the artificiality of border design, at least with respect to the tribes homelands.

Following the independence of African states, national borders have mostly conformed with what was established during the colonization. In fact, with the independence of 40 African states happening within a short period of time over the 1950s and first half of the 1960s, nearly all of them signed the Charter of the Organization of the African Union (OUA) in 1964 accepting colonial borders (with exception of Somalia and Morocco). Europeans were interested in maintaining their influence in the region, whereas African leaders wanted to prevent any threat to their position by the realignment of the borders. Therefore, the issue of border redesign was a timely one to be avoided and the boundaries were, in many ways, "the most consequential part of the colonial state" (Herbst 2000).

4 Data description

4.1 Ethnic homelands and partitioned groups

The sources of ethnic-level data are the Murdock map of ethnoliguistic regions (Murdock 1959) and the Ethnolinguistic Atlas (Murdock 1967), which together compose a anthropological database that is largely used across social sciences. Murdock (1959, 1967) compiled rich and systematic portraits of diverse societies with information on various aspects of life, such as subsistence, political organization, or social norms, based on numerous ethnographies from the 19th century. For each of the 843 ethnic groups identified in the African continent, Murdock (1959, 1967) provides the geographical coordinates and maps which have been digitized and made available by Nunn and Wantchekon (2011). The compiled map portrays ethnic homelands as partitions of the whole continent, including uninhabited regions, just before the Scramble for Africa for the near entirety of the tribes. As we cannot perfectly track how the spatial distribution of these ethnicities changed since then, in all exercises that will follow we use the ethnic population figures as of 1880 (Pre-Scramble). Reassuringly, other papers have shown that the distribution of ethnic groups from Murdock (1959, 1967) has not changed drastically in the last 150 years. Using data from the Afrobarometer, Nunn and Wantchekon (2011) report a correlation of 0.55 between the location of the respondents in 2005 and the historical homeland of their ethnicity as identified in Murdock's map. Recently, Bahrami-Rad et al. (2021) made an effort to link ethnicities from Murdock's map to corresponding ones reported in Standard Demographic and Health Surveys (DHS), using the complete version of historical societies in Murdock's map (around 1200 groups). Using self-reported data from 790,000 individuals across 43 countries, it then validates the use of Murdock's map by showing that there are positive associations between the historical measures collected by ethnographers and the current practices reported in DHS by the descendants of these historical societies.

To identify partitioned ethnicities and assign each area to the respective country, we inter-

sect Murdock's map with the Global Administrative Database (GADM), that portrays contemporary national boundaries. We identify 1314 polygons of country-tribe units from the intersection, and from the 843 initial tribe homelands, 368 are assigned as potentially split after the overlay of maps. We take a conservative approach and restrict the sample of split ethnic groups to the ones which had at least 5% of their population laying in one side of the border, according to the historical (gridded) population estimates provided by the History database of the Global Environment (HYDE), the main data source for the historical population estimates. This choice accounts for potential imprecision of border lines in Murdock (1959, 1967), as well as the possibility that individuals that ended up in countries where their ethnic group was a small share may have more easily migrated to bordering countries where their ethnic group was larger. This choice reduces the number of split groups to 243 ethnic homelands, which corresponds to 540 country-tribe observations. Figure 1 panel A displays the simple overlay of both maps, and panel B emphasizes the partitioned tribes whose population shares are above the 5% threshold.

4.2 Satellite light density at night

With the aim of comparing subnational levels, we require detailed spatial data on economic development. To the best of our knowledge, geocoded high-resolution measures of economic development spanning all of Africa are not available. To overcome this limitation, we use satellite light density at night to proxy for local economic activity. The luminosity data come from the Visible Infrared Imaging Radiometer Suite (VIIRS) Day Night Band (DNB) on board of the Joint Polar-orbiting Satellite System (JPSS) satellites, and are made available by the Earth Observation Group (Elvidge et al. 2021).[4] We use the product Annual VNL V2 for 2019, produced from monthly cloud-free average radiance for every 15 arc second grid cells (approximately 500 meters at the Equator line), in the form of raster files. The satellite

[4]Other papers that proxy economic activity with nighttime light density have used a lower resolution and lower quality time series produced by the Operational Linescan Sensor (OLS) onboard Defense Meteorological Satellite Program (DMSP) satellites, which has been discontinued since 2013.

detects lights from human settlements, fires, gas flares, lightning, and the aurora, and latter processing uses the twelve-month median radiance to discard high and low radiance outliers, filtering out most fires and isolating the background. We construct average light density per square kilometer for 2019, averaging across pixels at the desired level of aggregation.

The measure available is expressed in terms of nano Watts per steradian per square centimeter, a measure of radiance per area. Therefore, at the pixel level (the finest grid), this measure reflects not only light availability (e.g. a pixel is lit or not), but *light density* (the intensity of light use), aggregated for the whole year in question.

Several other papers have used luminosity data to proxy for development. Earlier contributions and diffusion of this method can be credited to the work of Henderson et al. (2011), along with previous work that has shown that light density is a robust proxy of economic activity (see Elvidge et al. (1997), Doll et al. (2006), and Pinkovskiy (2011)). These studies establish a strong within-country correlation between light density at night and GDP levels and growth rates. Chen and Nordhaus (2011), although emphasizing some of the problematic issues with using satellite image data, argued that luminosity can be quite useful for regional analysis in countries with poor-quality income data.

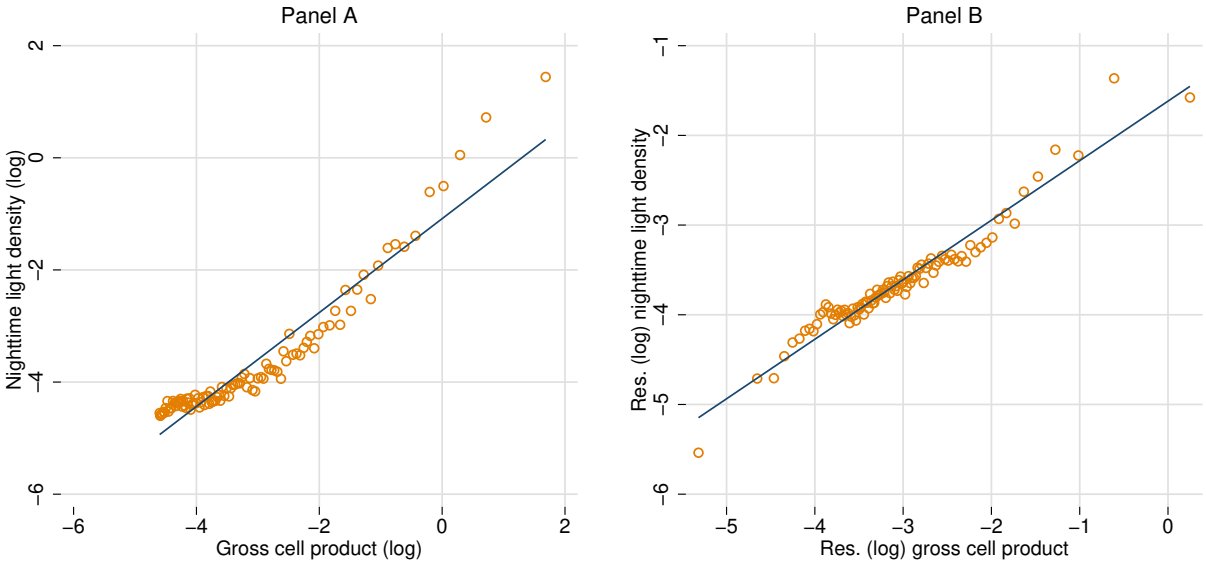
4.3 Cross-validation: satellite light density and regional development

Nordhaus et al. 2006 developed a geophysically scaled economic data set (hence, G-Econ), that consists of 1:1 degree grid cells for all terrestrial regions and is available for four years, 1990, 1995, 2000, and 2005. Its main purpose is to provide consistent economic activity data for subnational levels, comparable throughout the world. Importantly, its main innovation, the gross cell product (GCP), a version compatible with a regional gross product for the grid cell unit, takes into account the geographical intensity of economic activity without using measures of nighttime light intensity, as it uses primarily harmonized measures of the System of National Accounts. As a measure of gross value added, we can validate the use of light intensity by

aggregating it at the 1:1 grid cell level and comparing it to the available GCP.

We do this exercise in Figure 2, which uncovers the highly positive relationship between light density and the GCP in African countries, both unconditionally (panel A) and conditionally on country and year fixed effects and other country characteristics available at the grid level (panel B). These results reassure our choice of proxy for economic development, since we cannot use the GCP directly due to its level of aggregation (around 100 km by 100 km) being higher than some of the partitioned ethnic homelands.

Figure 2: Plots of light density versus gross cell product at the grid-cell level for Africa



Note: Panel A is a binned plot of grid-cell averages of nighttime light density and gross cell product (Nordhaus et al. 2006) for years 2000 and 2005 (last available) for the whole African continent calculated by the Peace Research Institute Oslo (PRIO). Panel B displays the binned residualized variables after controlling for year, country, existence of diamond mine and petroleum exploration, distance to the country's capital, and population count. Because of the skewed distribution of both measures and because of 0 values, we follow the literature and transform them using the operator $\ln(\cdot + 0.01)$.

4.4 Other data sources

Other data included in the exercises come mainly from georreferenced files, such as shapefiles and raster images, and are aggregated at the geographic level of the analysis. These are the Historical Database of the Global Environment - HYDE (historical population estimates, irrigated soil, pasture land, cropland, and others); the Global Agro-Ecological Zoning

- GAEZ/FAO (soil crop suitability, water availability, altitude, etc); The Malaria Project (malaria suitability, estimates of distance to nearest healthcare facility, and accessibility to cities); the Peace Research Institute Oslo - PRIO and USGS (occurrence of onshore petroleum, precious gemstones, and other minerals); and NASA (such as the global roads database, gROADS, and distance from the coast). In order to improve the precision of population figures today, we also collect gridded population estimates provided by the Socioeconomic Data and Applications Center (SEDAC/NASA). We complement these data with information made available by Nunn and Wantchekon (2011), such as colonial railways, Nunn and Puga (2012), such as terrain ruggedness and slope, and soil aridity and potential evapotranspiration (Trabucco and Zomer 2019). In order to aggregate the variables at the desired level, we use the weights given by the grid cell area definition in each of the data sets, when available.

5 Empirical strategy

The source of variation explored is the population size across partitioned groups. Therefore, we estimate the following regression model at the level of the tribe-country unit in the sample of partitioned ethnic groups

$$Y_{e,c,t} = \delta_e + \delta_c + \beta \log\left(\text{Population}_{e,c,1880}\right) + X'_{e,c}\phi + \varepsilon_{e,c} \tag{1}$$

where indices e, c denote, respectively, ethnicity and country. The outcome of interest is the average nighttime light intensity at the country-tribe level. The variable of interest, $\log\left(\text{Population}_{e,c,1880}\right)$, is calculated from gridded population estimates from the 1880s in order to capture effects disregarding endogenous migration that has happened since then. Parameters δ_e and δ_c are ethnic and country fixed-effects, respectively, and $X'_{e,c}$ is a vector of controls at the country-ethnicity level. As discussed in section 4.1, this approach is endorsed by previous works that have documented a significant persistence in location decisions since the Scramble for Africa throughout the continent. Therefore, our estimates can be interpreted as intention-to-treat. The inclusion of ethnicity- and country-specific effects allows us to overcome omitted

variable bias stemming from group-specific characteristics (e.g. precolonial institutions, precolonial development, etc), and country attributes, such as overall economic development, and country-specific policies and institutions, allowing us to emulate a quasi-natural experiment.

Table 1 presents descriptive statistics for the main variables at the ethnicity-country level. Although the light density at pixel level varies from 0 to 50664.59 $nW/cm^2/sr$ in Africa, the aggregated values by the mean at the country-tribe level are quite low (around 0.11 $nW/cm^2/sr$), while a direct measure of quantity, such as the share of lit pixels in the continent, is around 3%. Although consistent with low development levels, these values uncover large heterogeneity across regions. Both the distribution of light density and population size are highly skewed to smaller values. To smooth the light density distribution and given that it has some null values, we follow the literature and apply a logarithmic transformation after the addition of a small number, set to 0.01 as in Michalopoulos and Papaioannou (2014) and Alesina et al. (2016).

Table 1: Summary statistics at the ethnicity-country group level

	Mean	Median	SD	Min	Max	N
Main outcome and partitioned definitions						
Average luminosity in 2019, $nW/cm^2/sr$	0.11	0.00	0.64	0.00	13.25	1314
Ethnic group is split | population share ≥ 0, population share ≤ 1	0.64	1.00	0.48	0.00	1.00	1314
Ethnic group is split | population share $\geq 1\%$ in one country	0.53	1.00	0.50	0.00	1.00	1314
Ethnic group is split | population share $\geq 5\%$ in one country	0.41	0.00	0.49	0.00	1.00	1314
Ethnic group is split | population share $\geq 10\%$ in one country	0.33	0.00	0.47	0.00	1.00	1314
Population measures						
Total population in 1880, base 1000	82.76	22.60	262.07	0.00	5932.42	1314
Average population count in 1880 by km^2	601.73	270.47	1389.34	0.00	35662.68	1314
Population share in ethnic homeland area in 1880	0.04	0.01	0.09	0.00	0.97	1314
Geography controls						
Distance to coast (km)	619.77	561.50	445.20	0.68	1779.40	1314
Landlocked indicator	0.65	1.00	0.48	0.00	1.00	1314
Average altitude above sea level (m)	615.62	475.99	447.76	2.80	2902.05	1314
Ruggedness terrain index (100m)	0.68	0.38	0.83	0.00	6.76	1314
Average slope (%)	1.84	1.03	2.27	0.00	19.13	1314
Average precipitation (2019) in mm	84.84	82.90	52.54	0.02	315.06	1314
Malaria suitability index	0.19	0.15	0.17	0.00	0.75	1314
Share of area prone to seasonal water-logging	3.25	0.00	14.02	0.00	100.00	1314
Average anual evapotranspiration mm/year in 1970-2000	2064.46	1965.09	560.22	1064.26	3944.69	1314
Aridity index for 1970-2000	5946.39	5021.54	4610.09	0.58	25602.29	1314
Soil controls						
Share of irrigated area (%) in 1880	0.03	0.00	0.18	0.00	3.63	1314
Share of rain-fed area (%) in 1880	3.57	1.59	4.60	0.00	28.47	1314
Share of grazing area (%) in 1880	14.91	12.81	11.77	0.00	58.05	1314
Share of pasture area (%) in 1880	4.31	0.47	7.17	0.00	42.78	1314
Share of rangeland area (%) in 1880	9.19	4.09	11.60	0.00	58.05	1314
Share of cropland area (%) in 1880	3.60	1.63	4.61	0.00	28.47	1314
Soil suitability for rain-fed, low inputs crops (GAEZ v4)	2.51	2.25	1.07	0.31	8.40	1314
Suitability index for currently grown crop (GAEZ v4)	4711.07	5220.14	3648.82	0.00	10000.00	1314

6 Main Results

First, given the exogenous nature of the event of border design with respect to the tribes' homelands, one should not expect to find systematic association across ethnic population size and geography and land endowment variables that are invariant to land mass. Accordingly, appendix tables A1 and A2 show that, conditional on ethnicity and country, the tested outcomes do not exhibit systematic association with the size of the population, with the (marginal) exception of measures that reflect the median altitude of the area. This potentially uncovers the fact that larger land areas (positively associated with population size) have their centroids more distant from the coast, which could explain these differences in elevation. Moreover,

as distance to coast is only marginally significant, we do not draw much conclusion from it, although we note that the positive coefficient implies that larger groups are more distant from the shore, and this could only act in he opposite direction of the expected effect of group size on development (only making it smaller). Also, there are no consistent differences across land use and suitability. Nonetheless, we test the stability of the main estimates to the inclusion of this set of variables.

Table 2 displays the effect of the log of population size on the proxy for economic activity, and each column tests a slightly different specification that combines ethnicity and country fixed effects with two sets of control variables for geography and soil characteristics. As the sample is restricted to partitioned groups, the specifications compare economic outcomes across the *same* groups residing on both sides of country borders. In other words, in matching groups that share the same unobservable characteristics, and purging specific effects of ethnicity and country, we compare the level of development for the same group on both sides of the border, and therefore the estimator of the parameter of interest will represent a sort of quasi-experiment as it flushes away ethnic and country characteristics that might affect estimates.

The complete specification in column (7) documents that a 10% increase in population size increases nighttime light density by about 1.98%. This point estimate is overall stable over the inclusion of fixed effects, as columns (1) to (4) show. The baseline specification in column (4), including only fixed effects, is about 9.6% smaller then the point estimate after the inclusion of the full set of geography and soil controls. Taken together, these results are a robust evidence that ethnic groups' territorial portions with exogenously larger populations have today more intense economic activity, as proxied by nighttime light density, than their smaller counterparts.

Table 2: Effect of ethnic group size on economic development

	Light density (log)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Population in 1880 (log)	0.193***	0.130***	0.244***	0.179***	0.198***	0.181***	0.198***
	(0.038)	(0.043)	(0.032)	(0.032)	(0.037)	(0.032)	(0.039)
Ethnic FE		✓		✓	✓	✓	✓
Country FE			✓	✓	✓	✓	✓
Geography					✓		✓
Soil controls						✓	✓
# Observations	540	540	537	535	535	535	535
R^2	0.070	0.789	0.480	0.877	0.884	0.883	0.890
Mean Dep. Var.	-3.767	-3.767	-3.780	-3.789	-3.789	-3.789	-3.789
S.D. Dep. Var.	1.106	1.106	1.093	1.085	1.085	1.085	1.085

* p <0.10 ** p <0.05 *** p <0.01. The table presents the estimates for equation 1 in the sample of partitioned ethnic groups whose population shares across borders are above 5% (see section 4.1). Regressions are controlled by the variables depicted in tables A1 and A2. Standard errors double-clustered for ethnicity and country.

6.1 Robustness

6.1.1 Mechanic effects

Within split ethnic homelands, territories that ended up with a larger population in the eve of the colonization have sustained greater levels of luminosity today in comparison to their smaller sized counterparts. However, an immediate concern with the previous finding is that the impact of population size on light density may be solely mechanical, meaning it may not truly indicate an increase in the intensity of economic activity, but rather just a higher number of light bulbs. To address this, columns (1) to (4) of Table 3 show that, conditioning on *current population density*, the point estimates of population size in 1880 decreases by at most 34% and at least 19%, but remains highly significant in all specifications. This implies that population size in the past causes luminosity to increase proportionally more today, which we take as evidence of more intense economic activity.

6.1.2 Path dependence

In the same spirit, to test whether our findings may be driven by economic persistence throughout time, the next columns of Table 3 condition the effect of population size on a contemporary measure of economic complexity, which is the population density in 1880. Even

though these contemporary measures are very correlated, the effects of group size remain very significant and robust to this additional control. Meanwhile the proxy for economic development in the past loses significance in the complete specification, as shown in column (8) of Table 3. We interpret this finding as additional evidence that our main result is not a mere consequence of path dependence in development, but rather a consequence of the agglomeration of potentially co-ethnic population caused by the border design.

Table 3: No evidence of mechanic effect on light bulbs and of path dependence

	Light density (log)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Population in 1880 (log)	0.137***	0.153***	0.142***	0.160***	0.156***	0.166***	0.163***	0.175***
	(0.029)	(0.033)	(0.030)	(0.037)	(0.034)	(0.039)	(0.034)	(0.040)
Log of population density today	0.345***	0.357***	0.315***	0.326***				
	(0.063)	(0.060)	(0.068)	(0.065)				
Log of population density in 1880					0.241**	0.256**	0.199	0.217
					(0.116)	(0.118)	(0.128)	(0.129)
Ethnic FE	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Geography		✓		✓		✓		✓
Soil controls			✓	✓			✓	✓
# Observations	535	535	535	535	535	535	535	535
R^2	0.893	0.899	0.895	0.902	0.884	0.890	0.887	0.894
Mean Dep. Var.	-3.789	-3.789	-3.789	-3.789	-3.789	-3.789	-3.789	-3.789
S.D. Dep. Var.	1.085	1.085	1.085	1.085	1.085	1.085	1.085	1.085

* p <0.10 ** p <0.05 *** p <0.01. The table presents the estimates for equation 1 in the sample of partitioned ethnic groups whose population shares across borders are above 5% (see section 4.1). Regressions are controlled by the variables depicted in tables A1 and A2 and by respondent's age and squared age. Standard errors double-clustered for ethnicity and country.

6.1.3 The role of natural endowments

As one would expect, population count and land mass were positively and significantly correlated after the partition of the continent. This raises questions to whether natural endowments themselves may have had significant developmental role in the long run. This is because, generally speaking, the larger the area, the higher the probability of encountering economically meaningful natural resources. Although our findings do not support the notion that current economic activity is primarily driven by path dependence – and, hence, we do not expect that the availability of certain natural resources attracted individuals in the first place, and that this, in turn, explains the patterns we find in long run economic activity –, it is possible that larger groups, upon encountering a greater volume of natural resources and

having a larger population to exploit them, were able to make better use of these endowments for economic gain. For instance, access to large bodies of water likely facilitates agricultural production, while the existence of reserves of tradable natural resources directly affects economic activity, and the size of the workforce available to explore them is likely to be important for economically successful exploration, since these are activities intensive in labor to this day.

We assemble natural endowments data from many sources to extend the appendix Tables A1 and A2, and test the association between population size and availability of natural amenities, repeating the exercise from equation 1 on these new outcome variables. Table 4 shows that larger groups are more likely to have lakes and gold mines in their territory,[5] while there is no increased likelihood of having oil fields, diamond and other gems mines, or greater supply of rivers, as measured by their extension.

To address whether an ethnicity-country group's population size acts on current development through these channels, Table 4 reproduces the baseline regression while conditioning, at each column, on these new variables. Columns (1) through (6) show that the estimate of the parameter of interest is highly stable and remains significant. In column (7), we include the log of the land area of the ethnicity-country unit as an additional control, as this variable has the potential to address our primary objective of examining the impact of endowment supply on development, by the correlation discussed above. This new specification strengthens the link of population and economic activity. Finally, the last column of Table 4 shows that land area seems to manage the correlation of all such resources and current development, when included as the only additional control to the baseline model.

Although some of the tested endowments seem significant, none of them, nor their combination, seems to mediate significantly the baseline result. For the purpose of simplicity and since we cannot reject the importance of these resources for generating economic activity, all exercises to come will display the correlates of population size conditional on the land area of

[5]Data on onshore oil fields, diamond mines, and other gems mines are provided by PRIO, and refer to known reserves, while data on gold mines are made available by the USGS, and considers high volume reserves.

the ethnicity-country group's territory. With this, we intend to purge in the next sections the effect of natural endowments that, by design, are available to larger groups and could possibly have a direct impact on development.

Table 4: Correlation of population size and economically meaningful natural endowments

	Water availability			**Mineral reserves**			
	(1) River extension per km^2 of land	(2) Lakes	(3) Major inland water bodies	(4) Onshore oil fields	(5) Diamond mines	(6) Other gems mines	(7) Gold mines
Population in 1880 (log)	-0.000	0.099***	0.073*	-0.038	0.034	0.020	0.064**
	(0.002)	(0.026)	(0.037)	(0.026)	(0.032)	(0.026)	(0.026)
Geography and soil controls	✓	✓	✓	✓	✓	✓	✓
# Observations	535	535	535	535	535	535	535
R^2	0.758	0.764	0.837	0.796	0.720	0.754	0.682
Mean Dep. Var.	0.078	0.062	0.256	0.065	0.107	0.062	0.047
S.D. Dep. Var.	0.022	0.241	0.437	0.247	0.309	0.241	0.211

* p <0.10 ** p <0.05 *** p <0.01. The table presents the estimates for equation 1 in the sample of partitioned ethnic groups whose population shares across borders are above 5% (see section 4.1). Regressions are controlled by the variables depicted in tables A1 and A2 and by respondent's age and squared age. Standard errors double-clustered for ethnicity and country.

Table 5: Main table, controlling for natural endowments

	Dep. var.: Light density in log							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Population in 1880 (log)	0.204***	0.174***	0.200***	0.211***	0.215***	0.209***	0.421***	0.392***
	(0.039)	(0.038)	(0.039)	(0.041)	(0.040)	(0.041)	(0.116)	(0.124)
River length	3.792*					3.808	3.422	
	(2.196)					(2.294)	(2.088)	
Lake		-0.141				-0.175	-0.281	
		(0.183)				(0.183)	(0.182)	
Major water bodies		0.288**				0.298**	0.338**	
		(0.139)				(0.143)	(0.139)	
Onshore oild fields			-0.047			-0.035	0.094	
			(0.230)			(0.235)	(0.233)	
Diamond mines				-0.194		-0.150	-0.131	
				(0.122)		(0.122)	(0.116)	
Other gems mines					-0.384**	-0.444***	-0.419**	
					(0.159)	(0.164)	(0.156)	
Gold mines					-0.065	-0.077	-0.132	
					(0.185)	(0.179)	(0.169)	
Log(area excluding water bodies)							-0.244**	-0.217
							(0.120)	(0.129)
Ethnic FE	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Geography and soil controls	✓	✓	✓	✓	✓	✓	✓	✓
Observations	535	535	535	535	535	535	535	535
R-squared	0.892	0.892	0.890	0.891	0.892	0.897	0.901	0.894
Dep. Var. Mean	-3.789	-3.789	-3.789	-3.789	-3.789	-3.789	-3.789	-3.789
Dep. Var. SD	1.085	1.085	1.085	1.085	1.085	1.085	1.085	1.085

* p <0.10 ** p <0.05 *** p <0.01. The table presents the estimates for equation 1 in the sample of partitioned ethnic groups whose population shares across borders are above 5%. Regressions are controlled by the variables depicted in tables A1 and A2 and by respondent's age and squared age. Standard errors double-clustered for ethnicity and country. Dependent variables of columns (2) through (7) are indicators.

6.2 Heterogeneity and inequality

The summary statistics in Table 1 demonstrated that averaging the light density measure by the mean potentially masks high heterogeneity in the distribution of economic activity within regions. To address this point specifically, we next replicate the baseline model 1 to other measures calculated from the distribution of luminosity at the pixel level in each ethnicity-country territory.

Consistent with the descriptive evidence, Table 6 shows that population size, conditional on land area, makes a difference only at the highest 5 percentiles of the distribution of light density across groups. This implies that, for the same land endowment and size, the effect of population is potentially very concentrated in a low number, relative to the entirety, of extremely lit pixels, while, taking into account the previous evidence, these concentrated points are high enough to significantly drive the mean economic activity in each territory. This pattern might be consistent with economic activity being concentrated in urban centers, and with groups that were larger in the past leading structural changes that accompany the urbanization process.

Tables B3 and B4, in the appendix, document that areas with larger population have a more unequal distribution of light density, for any of the measures of dispersion we analyze, and that the baseline effect is remains regardless of the relative sizes of the splits of an ethnic group, but that point estimates are larger for bigger groups.

Table 6: Effects of ethnic group size on the distribution of light density

	***N-th* quantile of light density (log)**										
	(1) $N = 10$	(2) $N = 20$	(3) $N = 30$	(4) $N = 40$	(5) $N = 50$	(6) $N = 60$	(7) $N = 70$	(8) $N = 80$	(9) $N = 90$	(10) $N = 95$	(11) $N = 99$
Population in 1880 (log)	0.044	0.047	0.048	0.048	0.048	0.048	0.055	0.085	0.164	0.271*	0.483**
	(0.058)	(0.077)	(0.090)	(0.099)	(0.107)	(0.113)	(0.119)	(0.125)	(0.138)	(0.148)	(0.192)
Log(area)	0.069	0.126	0.168*	0.200**	0.227**	0.249**	0.259**	0.247*	0.202	0.127	0.017
	(0.056)	(0.076)	(0.090)	(0.099)	(0.107)	(0.113)	(0.120)	(0.127)	(0.142)	(0.150)	(0.200)
Ethnic FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Geography and soil	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
# Observations	535	535	535	535	535	535	535	535	535	535	535
R^2	0.691	0.704	0.710	0.714	0.717	0.719	0.722	0.727	0.750	0.783	0.839
Mean Dep. Var.	-3.841	-3.464	-3.209	-3.017	-2.861	-2.732	-2.614	-2.503	-2.366	-2.228	-1.701
S.D. Dep. Var.	0.455	0.639	0.762	0.855	0.929	0.992	1.054	1.118	1.227	1.369	1.837

* p <0.10 ** p <0.05 *** p <0.01. The table presents the estimates for equation 1 in the sample of partitioned ethnic groups whose population shares across borders are above 5% (see section 4.1). Instead of aggregating light density by the mean, the outcomes are the values that light density assumes in each quantile within country-homeland unit in log. Regressions are controlled by the variables depicted in tables A1 and A2 and by respondent's age and squared age. Standard errors double-clustered for ethnicity and country.

7 Channels

This section continues to characterize the forces that potentially mediate the link between partitioned ethnic group size and economic development. For each potential channel, we present and discuss suggestive evidence.

7.1 Economic and scale effects

The size of an ethnic group within a country can impact its economic performance in various ways. First, economic resources that sustain well-being, such as sewage system, access to electricity, public services, among others, are concentrated in *cities*. Therefore, agglomeration of individuals if not accompanied by structural changes in economic organization of the space, likely hinders economic development. Moreover, if agglomerations produce such changes, the space is not the only consequential part of belonging to larger groups, as individuals also transit to economic activities compatible with the urban area (McMillan and Harttgen 2014).

7.1.1 Urbanization

Following up on the evidence suggested by the distribution of economic activity in Section 6.2, Table 7 displays the results obtained from regressing nine measures of urbanization on the log of population density in 1880, conditional on land area and other baseline controls, as in equation 1. Consistent with the evidence in Table 6.2, these findings suggest that larger groups have become, in the long run, significantly more urbanized, as indicated by the higher shares of urban population (1), the number of administrative divisions (2), as well as the higher intensive (8) and extensive (9) margins of urban agglomerations. In the same direction, columns (3) and (4) show that larger groups have an increased likelihood of being home to the country's political capital cities and to other cities classified as metropolises. Table C5 further shows that larger groups are also more likely to have urban centers of any size in their territory, potentially indicating that the formation of urban clusters is an on-going process,

and individuals do not simply agglomerate in established focal big cities.

Table 7: Ethnic group size and Urbanization

	Measures of urbanisation								
	(1) Share of urban population	(2) Number of counties	(3) Country's capital within territory	(4) Metropolis within territory	(5) Travel time to cities (minutes)	(6) Travel time health facility by foot	(7) Travel time health facility by motor vehicle	(8) Urban cluster	(9) Quantity of urban clusters
Population in 1880 (log)	0.062**	3.180***	0.064***	0.073**	-65.902*	-155.561***	-66.518**	2.098***	2.942***
	(0.024)	(1.029)	(0.018)	(0.031)	(34.431)	(47.683)	(28.764)	(0.557)	(0.664)
Log(area)	✓	✓	✓	✓	✓	✓	✓	✓	✓
Ethnic FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Geography and soil resources	✓	✓	✓	✓	✓	✓	✓	✓	✓
# Observations	535	535	535	535	535	535	535	535	535
R^2	0.787	0.845	0.700	0.709	0.928	0.937	0.898	0.795	0.837
Mean Dep. Var.	0.331	7.030	0.028	0.064	307.481	346.341	136.732	3.662	6.772
S.D. Dep. Var.	0.253	11.051	0.165	0.244	474.234	595.500	286.315	5.876	7.643

* p <0.10 ** p <0.05 *** p <0.01. Regressions are controlled by the variables depicted in tables A1 and A2 and by respondent's age and squared age. Standard errors double-clustered for ethnicity and country. *Definitions*
Counties (GADM): Level 2 administrative units according to the Database of Global Administrative Areas (GADM), comparable to US counties.
Urban cluster (Africapolis): continuously built up areas (less than 200 metres between buildings and constructions) that have more than 10,000 inhabitants, according to country census. Clusters may or may not coincide with administrative divisions.
Travel time to cities (Malaria project): estimated time to commute by motor vehicle to the nearest high-density urban center, using road network from Open Street Maps and Google, and estimated surface frictions.

7.2 Individual and household level surveys

To examine the living conditions, attitudes and perceptions of individuals residing in partitioned homelands with different historical population sizes, we employ micro-level data from the most recent Demographic and Health Surveys (DHS) and round 7 of Afrobarometer, allowing us to control the regressions by a set of individual-level characteristics. Among the 36 African countries covered at any point in time by the DHS, 12 of them have absolutely no information for individuals' ethnic affiliation, and, of the remaining 24 countries, a sizeable fraction of individuals have not stated their ethnicity. In the Afrobarometer, individuals' ethnic group is asked in all the 33 countries surveyed in round 7. To consistently match the surveys' data sets to our baseline data at the ethnicity-country level, we perform a geographical match by assigning individuals' and households' clusters in each survey to a unit of ethnicity-country when the two overlap in the map. Additionally, for Afrobarometer, we are able to match 79.3% of individuals to their Murdock ethnic group correspondence using a combination of methods,[6] and conditioning on country.

[6]We employ primarily the Linking Ethnic Data from Africa (LEDA) algorithm of Müller-Crepon et al. (2022), explained in detail in section 7.4.2, complemented by the data sets made available by Nunn and Wantchekon (2011). Most individuals that could not be matched to any ethnicity from Murdock reported not identifying with any ethnic group, or stated their ethnicity in free text.

The estimates of interest are obtained by a regression model analogous to the baseline equation 1 at the individual or the household level i

$$Y_{i,e,c} = \delta_e + \delta_c + \beta \log\left(\text{Population1880}_{e,c}\right) + X'_{e,c}\phi + W'_i\lambda + \nu_{i,e,c} \tag{2}$$

All regressions are weighted by the individual or household weights provided, and controlled by respondent's age, and age squared. In the Afrobarometer sample, we additionally include controls for respondent's gender, and a dummy variable when the respondent and the interviewer are of the same ethnic group in order to reduce noise.

7.2.1 The DHS samples: Wealth and human capital

We focus on three main data sets: the women's and men's data sets, which contain information on individuals' occupation, educational attainment, health habits and other living conditions; and the household data sets, which provide information about family structure and assets holding. Pooling all most recent country data sets together, we have information for around 450k women and almost 192k men from 36 different countries.

Table 8 shows that about half of the split groups are represented in the latest surveys available, and that, for these, column (1) indicates that the effect of population size as of 1880 significantly impacts light density today. The next columns show that this effect and the result from the previous section are confirmed by the higher likelihood of the household being urban, and by the positive and significant association of population size with the proxies for wealth, such as the PCA index provided by the DHS original data, and our measures of wealth based on private and public assets reported in the surveys, aggregated according to the method proposed by Anderson (2008).[7] Panel B provides another evidence of economic activity and

[7]For the construction of the private assets index, we take into account the following information: availability, at the household, of radio, television, refrigerator, bicycle, motorcycle, car or truck, telephone line, type of fuel used for cooking, usage of bed net for sleeping, ownership of bank account, of mobile, computer, and scales of materials used in the house's construction. For the public assets index, we consider: availability of electricity, whether the toilet is shared with other households, type of drinking water supply, minutes to get to water, toilet type, and number of other households that use the toilet. The overall wealth index combines these private and public assets in one Anderson index.

living conditions being correlated with the urban space.

Table 8: Ethnic group size and living conditions

	Validation of sample	**Household living conditions**				
	Panel A: Baseline model					
	(1)	(2)	(3)	(4)	(5)	(6)
	Log(light density)	Urban household	Wealth index (DHS)	Anderson index for wealth	Anderson index for private assets	Anderson index for public assets
Population in 1880 (log)	0.682***	0.160***	0.048***	0.174***	0.145**	0.240***
	(0.159)	(0.053)	(0.015)	(0.061)	(0.062)	(0.086)
# Observations	250	190013	190013	190013	190013	190013
R^2	0.915	0.314	0.212	0.158	0.137	0.282
	Panel B: Inclusion of urban control					
	(1)	(2)	(3)	(4)	(5)	(6)
Population in 1880 (log)			0.024*	0.088*	0.067	0.116*
			(0.012)	(0.051)	(0.057)	(0.067)
Urban household=1			0.152***	0.539***	0.491***	0.772***
			(0.036)	(0.052)	(0.059)	(0.045)
Log(area)	✓	✓	✓	✓	✓	✓
Ethnic FE	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓
Geography and soil controls	✓	✓	✓	✓	✓	✓
# Observations			190,013	190,013	190,013	190,013
R^2			0.334	0.221	0.186	0.404
Mean Dep. Var.	-3.619	0.336	-0.008	-0.071	-0.033	-0.146
Sd Dep. Var.	1.032	0.472	0.169	0.840	0.870	0.862

* p <0.10 ** p <0.05 *** p <0.01. Weighted regressions. The table presents the estimates for equation 2 in the sample of partitioned ethnic groups whose population shares across borders are above 5% (see section 4.1). Regressions are controlled by the variables depicted in tables A1 and A2. Standard errors double-clustered for ethnicity and country.

Table 9 evidences that both men and women from larger groups have higher levels of education, but the effects are higher in the sample of women. Additionally, although the group's size does not seem to be related to the likelihood of being employed today, conditional on working, individuals from larger groups are less likely to be paid inkind, and more likely to hold permanent working positions when working, in comparison to seasonal positions. These last results are suggestive that individuals are relatively better off, once their working commitments are more stable in different seasons of the year, and that exchange relationships have a higher degree of monetization.

Table 9: Ethnic group size, human capital, and working conditions in the DHS samples aged 25 or above

	Panel A: Women's sample						
	Educational attainment				Working conditions		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Years of education	Able to read	Completed secondary school	Higher education	Currently working	Receives inkind payment	Permanent working position
Population in 1880 (log)	1.057***	0.093**	0.035**	0.027***	-0.006	-0.092**	0.176***
	(0.357)	(0.037)	(0.016)	(0.008)	(0.036)	(0.034)	(0.055)
Log(area)	✓	✓	✓	✓	✓	✓	✓
Ethnic FE	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓
Geography and soil controls	✓	✓	✓	✓	✓	✓	✓
# Observations	120,035	120,054	120,054	120,054	116,158	78,805	78,805
R^2	0.375	0.324	0.153	0.065	0.177	0.142	0.265
Mean Dep. Var.	4.707	0.521	0.107	0.050	0.710	0.162	0.524
Sd Dep. Var.	4.713	0.500	0.309	0.218	0.454	0.369	0.499
	Panel B: Men's sample						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Population in 1880 (log)	0.677*	0.027	0.026	0.036**	0.004	-0.140***	0.070
	(0.377)	(0.028)	(0.025)	(0.015)	(0.019)	(0.040)	(0.059)
Log(area)	✓	✓	✓	✓	✓	✓	✓
Ethnic FE	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓
Geography and soil controls	✓	✓	✓	✓	✓	✓	✓
# Observations	52,971	52,981	52,981	52,981	52,972	47,464	47,464
R^2	0.277	0.220	0.133	0.064	0.121	0.145	0.255
Mean Dep. Var.	6.361	0.696	0.183	0.096	0.904	0.176	0.630
Sd Dep. Var.	5.111	0.460	0.387	0.294	0.295	0.381	0.483

* p <0.10 ** p <0.05 *** p <0.01. Weighted regressions. The table presents the estimates for equation 2 in the sample of partitioned ethnic groups whose population shares across borders are above 5% (see section 4.1). Regressions are controlled by the variables depicted in tables A1, A2 and by respondent's age and squared age. Standard errors double-clustered for ethnicity and country.

Finally, we explore whether groups of different sizes are also associated with different occupational choices, that could reflect deeper changes in the structure of economic organization, as hypothesized in the beginning of this section. Table 10 shows that, for both male and female individuals, larger groups are associated with lower likelihood of working in rural occupations, such as agriculture, and higher likelihood of working with trade activities. For men, we also find a significant association with working in a skilled manual job. Although the estimates of population size on the index of concentration of sectors and occupations is negative in all cases, it is only marginally significant for men.

Table 10: Ethnic group size and occupation choice in the DHS samples aged 25 or above

	Panel A: Women's sample							
	Sector of occupation						Measures of diversification	
	(1) Agriculture	(2) Sales and services	(3) Domestic work	(4) Skilled manual	(5) Unskilled manual	(6) Professional	(7) HHI of occupations	(8) HHI of sectors
Population in 1880 (log)	-0.143*	0.112***	0.004	0.007	-0.001	0.006	-0.009	-0.038
	(0.072)	(0.040)	(0.007)	(0.007)	(0.012)	(0.009)	(0.028)	(0.051)
Log(area)	✓	✓	✓	✓	✓	✓	✓	✓
Ethnic FE	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Geography and soil controls	✓	✓	✓	✓	✓	✓	✓	✓
# Observations	113,641	113,641	113,641	113,641	113,641	113,641	248	248
R^2	0.290	0.182	0.071	0.076	0.101	0.043	0.838	0.778
Mean Dep. Var.	0.346	0.265	0.017	0.039	0.044	0.053	0.118	0.473
Sd Dep. Var.	0.476	0.441	0.131	0.194	0.205	0.225	0.158	0.192
	Panel B: Men's sample							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Population in 1880 (log)	-0.144**	0.076***	-0.001	0.045***	0.005	0.020	-0.063	-0.097*
	(0.054)	(0.023)	(0.008)	(0.015)	(0.021)	(0.018)	(0.049)	(0.048)
Log(area)	✓	✓	✓	✓	✓	✓	✓	✓
Ethnic FE	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Geography and soil controls	✓	✓	✓	✓	✓	✓	✓	✓
# Observations	51,794	51,794	51,794	51,794	51,794	51,794	234	234
R^2	0.247	0.123	0.095	0.088	0.089	0.040	0.681	0.743
Mean Dep. Var.	0.406	0.201	0.015	0.145	0.075	0.102	0.094	0.419
Sd Dep. Var.	0.491	0.401	0.120	0.352	0.263	0.302	0.158	0.195

* p <0.10 ** p <0.05 *** p <0.01. Weighted regressions. The table presents the estimates for equation 2 in the sample of partitioned ethnic groups whose population shares across borders are above 5% (see section 4.1). Regressions are controlled by the variables depicted in tables A1, A2 and by respondent's age and squared age. Standard errors double-clustered for ethnicity and country.

7.2.2 Afrobarometer: Social attitudes, perceptions, and access to public goods

As most outcomes that reflect social attitudes and perceptions in Afrobarometer are asked in scales, for each set of related variables we construct a combined normalized continuous score using the method of Anderson (2008). Table 11 shows the results obtained for outcomes that reflect social attitudes and perceptions. Although questions related to trust are only marginally significant for the case of informal institutions, namely the combination of religious and traditional ones, the other columns indicate that residing in ethnic homelands with higher population is positively associated with perceptions of higher state capacity, such as obedience to the law, to court rulings, and paying taxes, and with support for equality of gender roles. Moreover, these individuals report not only a higher periodicity of news consumption in any type of media outlet, but also greater access to public sector information, such as instructions of what taxes to pay and how to open a business, and to higher propensity of reporting corrupt

public servants in hypothetical situations.

Table 11: Ethnic group size and attitudes reported in Afrobarometer

	Social and cultural values					Access to information and mobilization		
	(1) Score: Trust in formal institutions	(2) Score: Trust in informal institutions	(3) Score: State capacity perception	(4) Score: Gender equality	(5) Score: Liberal leaning views	(6) Binary: Consumes news everyday	(7) Score: Access to public sector information	(8) Score: Reporting corrupt behavior
Population in 1880 (log)	0.062	-0.161*	0.207***	0.275***	-0.111	0.182***	0.240***	0.202***
	(0.112)	(0.081)	(0.064)	(0.097)	(0.098)	(0.059)	(0.078)	(0.061)
Log(area)	✓	✓	✓	✓	✓	✓	✓	✓
Ethnic FE	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Geography and soil resources	✓	✓	✓	✓	✓	✓	✓	✓
Survey controls	✓	✓	✓	✓	✓	✓	✓	✓
# Observations	16526	16526	16526	16526	16526	16526	16526	16526
R^2	0.190	0.181	0.140	0.121	0.130	0.168	0.127	0.142
Mean Dep. Var.	-0.016	-0.005	-0.024	-0.001	-0.015	0.581	-0.015	-0.010
S.D. Dep. Var.	1.004	1.003	1.008	0.998	1.005	0.493	0.998	1.000

* p <0.10 ** p <0.05 *** p <0.01. Respondent-level regressions weighted by cross-country sample weights provided by Afrobarometer. The table presents the estimates for equation 2 in the sample of partitioned ethnic groups whose population shares across borders are above 5% (see section 4.1). Regressions are controlled by the variables depicted in tables A1, A2, and survey controls are respondent's age, age squared, gender, and whether the interviewer is from the same ethnic groups as the respondent. Standard errors double-clustered for ethnicity and country.

Table 12 reports the effects of population size on objective measures of public goods provision, collected by the enumerators about the primary sampling units (PSU) of survey. Columns (1) to (3) corroborate the findings on the DHS sample, according to which individuals living in ethnic homelands with greater population in the past have today better access to electricity, water supply, and toilet facilities, as measured by the Anderson index of public goods. Accordingly, here we find high likelihoods of having electricity grid, piped water system, and sewage system in the PSU. to We find no evidence that population is correlated with cell phone service coverage, school and health clinics supply, and existence of market stalls, but we also find marginal correlation with existence of a police station and bank branch. At last, paid transport is highly positively correlated with population size in the past.

In Appendix Table D6, we test some of the same outcomes reported in the DHS subsection, and we obtain very similar results. Moreover, when including a covariate for the population size of individuals' ancestral homeland, Appendix Tables E7, E8, and E9 additionally indicate that the historical population of current living location is more predictive of all the results discussed above, while the historical ancestral population is not consistently predictive of any

of them.

Table 12: Ethnic group size and public infrastructure, according to Afrobarometer

	Available at the primary sampling area									
	(1) Electricity grid	(2) Piped water system	(3) Sewage system	(4) Cell phone service	(5) School	(6) Police station	(7) Health clinic	(8) Market stalls	(9) Bank	(10) Paid transport
Population in 1880 (log)	0.199**	0.277***	0.156**	-0.004	0.112	0.166*	-0.037	0.183	0.147*	0.198***
	(0.072)	(0.068)	(0.062)	(0.047)	(0.068)	(0.084)	(0.076)	(0.112)	(0.073)	(0.056)
Log(area)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Ethnic FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Geography and soil resources	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Survey controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
# Observations	17028	17028	17028	17028	17028	17028	17028	17028	17028	17028
R^2	0.428	0.356	0.356	0.248	0.199	0.202	0.226	0.257	0.220	0.287
Mean Dep. Var.	0.595	0.482	0.207	0.914	0.836	0.295	0.532	0.658	0.189	0.734
S.D. Dep. Var.	0.491	0.500	0.405	0.280	0.370	0.456	0.499	0.475	0.391	0.442

* p <0.10 ** p <0.05 *** p <0.01. Respondent-level regressions weighted by cross-country sample weights provided by Afrobarometer. The table presents the estimates for equation 2 in the sample of partitioned ethnic groups whose population shares across borders are above 5% (see section 4.1). Regressions are controlled by the variables depicted in tables A1, A2, and survey controls are respondent's age, age squared, gender, and whether the interviewer is from the same ethnic groups as the respondent. Standard errors double-clustered for ethnicity and country.

7.3 Road infrastructure

If population size matters for economic activity, regions occupied by larger groups may have the need to be better connected to different areas of a territory, to ensure the flow of goods and people. The first two columns of Table 13 show evidence that larger groups have better access to roads and railway networks, as measured by the distance of the groups' centroid to the nearest road and railway features. These effects seem to be driven, specifically, by truck roads in the first case, and no effect is found in railways constructed during colonial times, which may suggest that significant railway investments were made post-independence (likely in the form of expansion of the colonial investments (Jedwab and Storeygard 2019)).

Table 13: Ethnic group size and road infrastructure investment

	Log of distance		Log of the extension						
	(1) Nearest road	(2) Nearest railway	(3) Any road	(4) Motorway	(5) Trunk road	(6) Primary road	(7) Secondary road	(8) Railway	(9) Colonial railway
Population in 1880 (log)	-0.401**	-0.238***	0.894	-0.004	1.099**	-0.004	0.580	2.043**	0.395
	(0.174)	(0.079)	(0.634)	(0.002)	(0.522)	(0.002)	(0.350)	(0.782)	(0.271)
Log(area)	✓	✓	✓	✓	✓	✓	✓	✓	✓
Ethnic FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Geography and soil resources	✓	✓	✓	✓	✓	✓	✓	✓	✓
# Observations	535	535	535	535	535	535	535	535	535
R^2	0.734	0.897	0.804	0.796	0.831	0.796	0.843	0.756	0.857
Mean Dep. Var.	8.159	11.618	10.300	-4.599	2.432	-4.599	4.270	-0.881	-2.993
S.D. Dep. Var.	1.263	1.161	5.835	0.024	8.090	0.024	3.624	6.813	4.785

* p <0.10 ** p <0.05 *** p <0.01. The table presents the estimates for equation 1 in the sample of partitioned ethnic groups whose population shares across borders are above 5% (see section 4.1). Regressions are controlled by the variables depicted in tables A1 and A2 and by respondent's age and squared age. Standard errors double-clustered for ethnicity and country. Distances are calculated in meters from the country-ethnic group's polygon centroid to the nearest feature. Road data comes from Open Street Maps, and categories are definied as
Motorway road: High capacity highways designed to safely carry fast motor traffic. Usually with control of access.
Trunk road: Important roads that are not motorways.
Primary road: Highway linking large towns.
Secondary road: Highway which is not part of a major route, but nevertheless forming a link in the national route network.
Current railway data is provided by FAO Network, and colonial railways are provided by Nunn and Wantchekon (2011).

7.4 Political capital

In order to assess whether larger groups benefit from greater access to political power, we match each of the tribe-country groups to the ethnic-country groups defined in the Ethnic Power Relations family data sets (Vogt et al. 2015). We describe this collection of data and the matching procedure below in the next two subsections.

7.4.1 The Ethnic Power Relations Family

The EPR Core dataset provides annual data for *politically relevant* ethnic groups and their access to state power from 1946 to 2019. This core data set is complemented by geo-spatial information on ethnic groups' settlement patterns, data on ethnic groups' links to rebel organizations, on the trans-border relations of ethnic groups, and on intraethnic cleavages, composing the EPR Family of data sets.

The EPR data set defines ethnicity "as a subjectively experienced sense of commonality based on a belief in common ancestry and shared culture" (Vogt et al. 2015). Therefore, different markers may be used to indicate such shared ancestry and culture in each country, such as a common language, similar phenotypical features, or adherence to the same faith.

Moreover, an ethnic group is considered *politically relevant* if at least one political organization has claimed to represent its interests at the national level or if its members are subjected to state-led political discrimination. The data set uses the following nomenclatures to classify politically relevant actors: Monopoly or Dominance, when the group rules alone; Senior or Junior Partner, when the group shares power; and Powerless, Discrimination, or Self-Exclusion, when the group is excluded from power.[8]

We use the whole ethnicity-country yearly time series of EPR data until 2017. Among the many measures of an ethnic group's political status, we use the ones most directly associated with access to power and representation in formal politics, which are (i) whether the ethnic group occupies highest political positions in the country in 2017, having either monopoly or dominance over political power, or being senior partner; (ii) whether the group is powerless or discriminated (in which we include self-exclusion); (iii) the share of years the ethnic group occupies highest political positions since the country's independence, which is calculated as the ratio between the number of years a group is coded as in (i) and the difference between 2017 and the year of the country's independence (i.e. the year the country is coded into EPR).

7.4.2 The Linking Ethnic Data from Africa algorithm

As each data set employs a different definition for ethnic group, matching different data sets at the ethnicity level is challenging. The LEDA algorithm (Müller-Crepon et al. (2022)) allows one to link ethnic groups from different data sets based on the groups' match to the linguistic trees available in Ethnologue (Gordon Jr 2005). To match ethnicities from Murdock's map and EPR, we employ the algorithm's option to link ethnic groups based on the highest overlap in the language tree departing from the dialect level (e.g. how many nodes two groups share on the tree) and respecting country boundaries. For each group from Murdock, we assign the corresponding EPR group or groups, concluding that groups with no match (e.g. those whose overlap to any other group is zero) are not included in the EPR dataset, and are therefore not

[8]For more details about these categories, we refer to EPR's codebook (2021).

politically relevant in the national level. Doing this, the average language tree overlap that we consider in the matching procedure is 0.9 (in a scale from 0 to 1).

7.4.3 Results

Table 14 shows that larger groups are more likely to be politically relevant in the present (column (1)), and to access political power and be formally represented at the national level (column(3)). As a result, groups with largest populations tend to occupy dominant positions in the national government both in the present (column (2)) and throughout the years since independence (column (4)).

Table 14: Ethnic group size and political capital

	Political representation			
	(1) Politically relevant in country	(2) Occupies highest political positions	(3) Powerless or discriminated	(4) Share of years in highest positions since independence
Population in 1880 (log)	0.127**	0.150***	-0.184***	0.140***
	(0.049)	(0.047)	(0.050)	(0.038)
Log(area)	✓	✓	✓	✓
Ethnic FE	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
Geography and soil controls	✓	✓	✓	✓
# Observations	530	530	530	530
R^2	0.747	0.718	0.706	0.747
Mean Dep. Var.	0.506	0.191	0.611	0.168
S.D. Dep. Var.	0.500	0.393	0.488	0.331

* p <0.10 ** p <0.05 *** p <0.01. The table presents the estimates for equation 1 in the sample of partitioned ethnic groups whose population shares across borders are above 5% (see section 4.1). Regressions are controlled by the variables depicted in tables A1 and A2 and by respondent's age and squared age. Standard errors double-clustered for ethnicity and country.

8 Conclusion

We have studied the effect of group size on the ethnic-level economic performance. We take advantage of the Scramble for Africa that in practice divided the African continent between the European powers with negligible considerations about the location of pre-existing ethnic tribes. This caused a natural experiment where groups were split across countries and generated

exogenous variation on their size in each country.

We document that large groups have an economic advantage over their smaller counterparts. We find suggestive evidence in favor of the economies of scale and a political power mechanism. Although the availability of land endowments, another consequence of the scramble by design, exhibits unbalance across smaller and larger groups, it seem to be a minor mediating force to the link between population size and development in this context. Instead, we document that larger groups have potentially taken advantage of economies of agglomeration to form urban centers. As a result, living conditions, and road and railway connectivity are enhanced, and other fundamental basis for development, such as educational attainment, is improved, while choices of occupation also seem to reflect deeper changes in the structural organization of these societies.

Last but not least, larger groups are more likely to be politically relevant, indicating the existence of a political organization claiming to represent them. As a result, they are also more likely to have co-ethnics in the executive power. In an environment plagued with ethnic favoritism, this can imply advantage in patronage and allocation of public goods.

All together, our results highlight the relevance of ethnic barriers for the development process. The evidence also sheds light to another detrimental effect of the Scramble for Africa. By not taking into account the pre-existing settlement of ethnic groups, their partitioning have caused the presence of smaller groups within countries, who were eventually discriminated or less favored by economies of scale or political representation.

References

A. Alesina and E. La Ferrara. Participation in heterogeneous communities. *The Quarterly Journal of Economics*, 115(3):847–904, 2000. 7

A. Alesina and E. La Ferrara. Ethnic diversity and economic performance. *Journal of economic literature*, 43(3):762–800, 2005. 1

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

A. Alesina, S. Michalopoulos, and E. Papaioannou. Ethnic inequality. *Journal of Political Economy*, 124(2):428–488, 2016. 3, 16

M. L. Anderson. Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association*, 103(484):1481–1495, 2008. 26, 29

A. I. Asiwaju. *Partitioned Africans: Ethnic relations across Africa's international boundaries, 1884-1984*. C. Hurst & Co. Publishers, 1985. 10

D. Bahrami-Rad, A. Becker, and J. Henrich. Tabulated nonsense? testing the validity of the ethnographic atlas. *Economics Letters*, 204:109880, 2021. 11

R. Burgess, R. Jedwab, E. Miguel, A. Morjaria, and G. Padró i Miquel. The value of democracy: evidence from road building in kenya. *American Economic Review*, 105(6):1817–51, 2015. 8

X. Chen and W. D. Nordhaus. Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences*, 108(21):8589–8594, 2011. 13

G. De Luca, R. Hodler, P. A. Raschky, and M. Valsecchi. Ethnic favoritism: An axiom of politics? *Journal of Development Economics*, 132:115–129, 2018. 8

E. Depetris-Chauvin, R. Durante, and F. Campante. Building nations through shared experiences: Evidence from african football. *American Economic Review*, 110(5):1572–1602, 2020. 8

A. Dickens. Ethnolinguistic favoritism in african politics. *American Economic Journal: Applied Economics*, 10(3):370–402, 2018. 8

C. N. Doll, J.-P. Muller, and J. G. Morley. Mapping regional economic activity from night-time light satellite imagery. *Ecological Economics*, 57(1):75–92, 2006. 13

W. Easterly and R. Levine. Africa's growth tragedy: policies and ethnic divisions. *The quarterly journal of economics*, pages 1203–1250, 1997. 1

C. D. Elvidge, K. E. Baugh, E. A. Kihn, H. W. Kroehl, E. R. Davis, and C. W. Davis. Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. *International Journal of Remote Sensing*, 18(6):1373–1379, 1997. 13

C. D. Elvidge, M. Zhizhin, T. Ghosh, F.-C. Hsu, and J. Taneja. Annual time series of global viirs nighttime lights derived from monthly averages: 2012 to 2019. *Remote Sensing*, 13(5):922, 2021. 12

H. Finseraas, T. Hanson, Å. A. Johnsen, A. Kotsadam, and G. Torsvik. Trust, ethnic diversity, and personal contact: A field experiment. *Journal of Public Economics*, 173:72–84, 2019. 7, 8

P. Francois, I. Rainer, and F. Trebbi. How is power shared in africa? *Econometrica*, 83(2):465–503, 2015. 8

R. G. Gordon Jr. Ethnologue, languages of the world. *http://www. ethnologue. com/*, 2005. 33

M. Greenstone, R. Hornbeck, and E. Moretti. Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of political economy*, 118(3):536–598, 2010. 2

A. Greif. Contract enforceability and economic institutions in early trade: The maghribi traders' coalition. *American Economic Review*, pages 525–548, 1993. 7

V. Henderson, A. Storeygard, and D. N. Weil. A bright idea for measuring economic growth. *American Economic Review*, 101(3):194–99, 2011. 13

J. Herbst. Economic incentives, natural resources and conflict in africa. *Journal of African Economies*, 9(3):270–294, 2000. 10

J. Hjort. Ethnic divisions and production in firms. *The Quarterly Journal of Economics*, 129(4): 1899–1946, 2014. 7

R. Jedwab and A. Storeygard. Economic and political factors in infrastructure investment: evidence from railroads and roads in africa 1960–2015. *Economic History of Developing Regions*, 34(2): 156–208, 2019. 31

E. Kramon and D. N. Posner. Ethnic favoritism in education in kenya. *Quarterly Journal of Political Science*, 11(1):1–58, 2016. 8

D. D. Laitin and R. Ramachandran. Linguistic diversity, official language choice and human capital. *Journal of Development Economics*, 156:102811, 2022. 8

M. Lowe. Types of contact: A field experiment on collaborative and adversarial caste integration. *American Economic Review*, 111(6):1807–44, 2021. 8

M. S. McMillan and K. Harttgen. What is driving the'african growth miracle'? Technical report, National Bureau of Economic Research, 2014. 24

S. Michalopoulos. The origins of ethnolinguistic diversity. *American Economic Review*, 102(4):1508–39, 2012. 1

S. Michalopoulos and E. Papaioannou. Pre-colonial ethnic institutions and contemporary african development. *Econometrica*, 81(1):113–152, 2013. 3

S. Michalopoulos and E. Papaioannou. National institutions and subnational development in africa. *The Quarterly journal of economics*, 129(1):151–213, 2014. 3, 16

S. Michalopoulos and E. Papaioannou. The long-run effects of the scramble for africa. *American Economic Review*, 106(7):1802–48, 2016. 2, 9, 10

C. Müller-Crepon, Y. Pengl, and N.-C. Bormann. Linking ethnic data from africa (leda). *Journal of Peace Research*, 2022. 25, 33

G. P. Murdock. Africa, its peoples and their culture history. 1959. 4, 11, 12

G. P. Murdock. Ethnographic atlas: a summary. *Ethnology*, 6(2):109–236, 1967. 2, 4, 11, 12

W. Nordhaus, Q. Azam, D. Corderi, K. Hood, N. M. Victor, M. Mohammed, A. Miltner, and J. Weiss. The g-econ database on gridded output: methods and data. *Yale University, New Haven*, 6, 2006. 13, 14

N. Nunn. The long-term effects of africa's slave trades. *The Quarterly Journal of Economics*, 123(1): 139–176, 2008. 1

N. Nunn and D. Puga. Ruggedness: The blessing of bad geography in africa. *Review of Economics and Statistics*, 94(1):20–36, 2012. 15

N. Nunn and L. Wantchekon. The slave trade and the origins of mistrust in africa. *American Economic Review*, 101(7):3221–52, 2011. 11, 15, 25, 32

M. Peters. Market size and spatial growth—evidence from germany's post-war population expulsions. *Econometrica*, 90(5):2357–2396, 2022. 1, 2

M. Pinkovskiy. Economic discontinuities across borders: Evidence from satellite data on lights at night'. report, 2011. 13

H. Tajfel, M. G. Billig, R. P. Bundy, and C. Flament. Social categorization and intergroup behaviour. *European Journal of Social Psychology*, 1(2):149–178, 1971. 7

A. Trabucco and R. Zomer. Global aridity index and potential evapo-transpiration (et0) climate database v2. cgiar consortium for spatial information (cgiar-csi). *Published online, available from the CGIAR-CSI GeoPortal*, 2019. 15

M. Vogt, N.-C. Bormann, S. Rüegger, L.-E. Cederman, P. Hunziker, and L. Girardin. Integrating data on ethnicity, geography, and conflict: The ethnic power relations data set family. *Journal of Conflict Resolution*, 59(7):1327–1342, 2015. doi: 10.1177/0022002715591215. URL `http://dx.doi.org/10.1177/0022002715591215`. 32

A Appendix: Balance in geographical and soil characteristics

Table A1: Balance of natural resources supply – Geography

	(1) Distance to coast in log	(2) Landlocked indicator	(3) Mean altitude in log	(4) Ruggedness index	(5) Average slope in %	(6) Malaria suitability	(7) Share of sites prone to water logging	(8) Precipitation (current yearly average)	(9) Evapotranspiration in mm/year for 1970-2000	(10) Aridity index for 1970-2000
Population in 1880 (log)	0.019*	-0.021**	0.037***	0.028	0.077	-0.001	-1.077	-0.700	-5.818	-24.400
	(0.010)	(0.009)	(0.012)	(0.024)	(0.064)	(0.002)	(0.690)	(0.473)	(3.736)	(36.891)
Ethnic FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
# Observations	535	535	535	535	535	535	535	535	535	535
R^2	0.988	0.950	0.976	0.904	0.904	0.973	0.664	0.992	0.992	0.992
Mean Dep. Var.	5.946	0.622	6.064	0.640	1.716	0.200	3.485	84.970	2066.297	5896.364
S.D. Dep. Var.	1.206	0.485	0.975	0.724	1.945	0.158	14.451	52.249	540.913	4531.451

* p <0.10 ** p <0.05 *** p <0.01. The unit of analysis is the tribe-country group, and the sample is restricted to partitioned ethnic groups with at least 5% of its population as of 1880 falling at one side of the border. Each column reports a different regression of a geographic variable against the logarithmic of the population count in 1880. All regressions use fixed effects for ethnic group and country. Standard errors double-clustered for ethnicity and country.

Table A2: Balance of natural resources supply – Soil characteristics

	Historical land use						**Current measures**	
	(1) Share of irrigated land	(2) Share of rainfed land	(3) Share of graizing land	(4) Share of pasture land	(5) Share of rangeland	(6) Share of cropland	(7) Soil suitability for rain-fed, low inputs	(8) Suitability index for current crop
Population in 1880 (log)	0.006	-0.096	0.302	0.369**	-0.029	-0.090	0.075*	60.836
	(0.006)	(0.152)	(0.290)	(0.172)	(0.308)	(0.152)	(0.038)	(84.437)
Ethnic FE	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
# Observations	535	535	535	535	535	535	535	535
R^2	0.700	0.906	0.902	0.852	0.895	0.907	0.860	0.947
Mean Dep. Var.	0.028	3.157	14.283	3.267	9.647	3.185	2.445	4770.193
S.D. Dep. Var.	0.185	4.437	11.878	5.674	11.715	4.451	0.951	3598.435

* p <0.10 ** p <0.05 *** p <0.01. The unit of analysis is the tribe-country group, and the sample is restricted to partitioned ethnic groups with at least 5% of its population as of 1880 falling at one side of the border. Each column reports a different regression of a land endowment variable against the logarithmic of the population count in 1880. All regressions use fixed effects for ethnic group and country. Standard errors double-clustered for ethnicity and country.

B Appendix: Heterogeneity

Table B3: Effects of ethnic group size on dispersion measures

	Measures from the distribution of lights within country-homeland			
	(1)	(2)	(3)	(4)
	Log(S.D. of light density + 0.01)	S.D. of log(light density + 0.01)	Log(Mean abs. dev. of light density + 0.01)	S.D. of lit/unlit pixel
Population in 1880 (log)	0.387*	0.170***	0.427***	0.029***
	(0.193)	(0.048)	(0.138)	(0.009)
Log(area)	0.141	-0.100*	-0.203	-0.015
	(0.212)	(0.050)	(0.145)	(0.009)
Ethnic FE	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓
# Observations	535	535	535	535
R^2	0.868	0.913	0.897	0.923
Mean Dep. Var.	-2.337	0.429	-3.519	0.098
S.D. Dep. Var.	1.779	0.489	1.270	0.105

* p <0.10 ** p <0.05 *** p <0.01. The table presents the estimates for equation 1 in the sample of partitioned ethnic groups whose population shares across borders are above 5% (see section 4.1). Instead of aggregating light density by the mean, the outcomes are the values that light density assumes in each quantile within country-homeland unit in log. Geography controls are distance to coast (log), indicator for being landlocked, terrain ruggedness index, average slope, average precipitation, and malaria suitability index. Agriculture controls are share of irritated area in 1880, share of grazing area in 1880, share of pasture area in 1880, share of cropland area in 1880, mean altitude, maximum altitude, crop suitability index, land suitability level, share of water sites, dominant soil, agro-zones, and hydro-basin class. Standard errors double-clustered for ethnicity and country. Dependent variables of columns (2) through (7) are binary indicators.

Table B4: Heterogeneity according to smaller split's population size

	Dep. var.: Light density (log)		
		Control group population ≤ median	Control group population > median
	(1)	(2)	(3)
Population in 1880 (log)	0.333***	0.213*	0.763***
	(0.107)	(0.119)	(0.193)
Population in 1880 (log) × *Control*'s size	0.005*		
	(0.003)		
Log(area)	✓	✓	✓
Ethnic FE	✓	✓	✓
Country FE	✓	✓	✓
Controls	✓	✓	✓
# Observations	535	278	248
R^2	0.897	0.918	0.933
Mean Dep. Var.	-3.789	-4.010	-3.557
S.D. Dep. Var.	1.085	0.963	1.160

	Effect of a 10% increase in population	
Population from "control" group	Point estimate	95% CI
10th percentile	3.40%	[1.22%, 5.57%]
50th percentile	3.75%	[1.45%, 6.05%]
90th percentile	6.14%	[2.04%, 10.2%]

* p <0.10 ** p <0.05 *** p <0.01. Standard errors double-clustered for ethnicity and country.

C Appendix: Urbanization

Table C5: Ethnic group size and city size

	Quantity of urban clusters according to population size					
	(1) 10k to 20k	(2) 20k to 50k	(3) 50k to 100k	(4) 100k to 500k	(5) 500k to 2.5mi	(6) More than 2.5mi
Population in 1880 (log)	0.101**	0.117**	0.118**	0.096**	0.067**	0.030**
	(0.045)	(0.044)	(0.046)	(0.039)	(0.032)	(0.013)
Log(area)	✓	✓	✓	✓	✓	✓
Ethnic FE	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓
Geography and agricultural resources	✓	✓	✓	✓	✓	✓
# Observations	535	535	535	535	535	535
R^2	0.759	0.753	0.733	0.752	0.756	0.740
Mean Dep. Var.	0.512	0.480	0.258	0.198	0.073	0.017
S.D. Dep. Var.	0.500	0.500	0.438	0.399	0.260	0.129

* p <0.10 ** p <0.05 *** p <0.01. Standard errors double-clustered for ethnicity and country. *Definitions*

Counties (GADM): Level 2 administrative units according to the Database of Global Administrative Areas (GADM), comparable to US counties.

Urban cluster (Africapolis): continuously built up areas (less than 200 metres between buildings and constructions) that have more than 10,000 inhabitants, according to country census. Clusters may or may not coincide with administrative divisions.

Travel time to cities (Malaria project): estimated time to commute by motor vehicle to the nearest high-density urban center, using road network from Open Street Maps and Google, and estimated surface frictions.

D Appendix: Validation of DHS results in the Afrobarometer sample

Table D6: Ethnic group size and living conditions reported in Afrobarometer

	Cross-validation: living conditions									
	(1) Index of asset holdings	(2) Urban household	(3) Employed	(4) Occupation: services	(5) Occupation: agriculture	(6) Occupation: unskilled manual	(7) Occupation: skilled manual	(8) Occupation: upper or mid level	(9) Education: no education	(10) Education: university or more
Population in 1880 (log)	0.203**	0.136**	0.020	0.024	-0.167***	0.016	0.035***	0.037*	-0.101***	0.012
	(0.082)	(0.064)	(0.026)	(0.020)	(0.035)	(0.018)	(0.010)	(0.018)	(0.028)	(0.011)
Log(area)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Ethnic FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Geography and soil resources	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Survey controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
# Observations	18042	18042	18042	18042	18042	18042	18042	18042	18042	18042
R^2	0.262	0.335	0.123	0.056	0.234	0.060	0.062	0.045	0.311	0.063
Mean Dep. Var.	0.015	0.411	0.274	0.135	0.246	0.081	0.070	0.078	0.207	0.055
S.D. Dep. Var.	0.995	0.492	0.446	0.341	0.431	0.273	0.255	0.268	0.405	0.229

* $p<0.10$ ** $p<0.05$ *** $p<0.01$. Respondent-level regressions weighted by cross-country sample weights provided by Afrobarometer. The table presents the estimates for equation 1 in the sample of partitioned ethnic groups whose population shares across borders are above 5% (see section 4.1). Regressions are controlled by the variables depicted in tables A1 and A2, and survey controls are respondent's age, age squared, gender, and whether the interviewer is from the same ethnic groups as the respondent. Standard errors double-clustered for ethnicity and country.

E Appendix: Current location versus ancestral homeland

Table E7: Ethnic group size and attitudes reported in Afrobarometer

	Social and cultural values					Access to information and mobilization		
	(1) Score: Trust in formal institutions	(2) Score: Trust in informal institutions	(3) Score: State capacity perception	(4) Score: Gender equality	(5) Score: Liberal leaning views	(6) Binary: Consumes news everyday	(7) Score: Access to public sector information	(8) Score: Reporting corrupt behavior
Population in 1880 (log)	-0.002	-0.227**	0.279***	0.233**	0.113	0.216***	0.303**	0.202*
	(0.159)	(0.089)	(0.075)	(0.092)	(0.095)	(0.066)	(0.127)	(0.100)
Ancestral population in 1880 (log)	-0.003	-0.009	0.015	0.007	-0.011	0.001	-0.010	-0.007
	(0.010)	(0.011)	(0.012)	(0.014)	(0.011)	(0.005)	(0.013)	(0.012)
Log(area)	✓	✓	✓	✓	✓	✓	✓	✓
Ethnic FE	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Geography and soil resources	✓	✓	✓	✓	✓	✓	✓	✓
Survey controls	✓	✓	✓	✓	✓	✓	✓	✓
# Observations	12133	12133	12133	12133	12133	12133	12133	12133
R^2	0.219	0.190	0.159	0.130	0.149	0.165	0.135	0.141
Mean Dep. Var.	-0.028	0.025	-0.068	0.024	0.026	0.549	-0.049	-0.074
S.D. Dep. Var.	1.019	0.998	1.023	1.008	0.998	0.498	1.021	1.005

* $p<0.10$ ** $p<0.05$ *** $p<0.01$. Respondent-level regressions weighted by cross-country sample weights provided by Afrobarometer. The table presents the estimates for equation 1 in the sample of partitioned ethnic groups whose population shares across borders are above 5% (see section 4.1). Regressions are controlled by the variables depicted in tables A1 and A2, and survey controls are respondent's age, age squared, gender, and whether the interviewer is from the same ethnic groups as the respondent. Standard errors double-clustered for ethnicity and country.

Table E8: Ethnic group size and public infrastructure, according to Afrobarometer

	Available at the primary sampling area									
	(1) Electricity grid	(2) Piped water system	(3) Sewage system	(4) Cell phone service	(5) School	(6) Police station	(7) Health clinic	(8) Market stalls	(9) Bank	(10) Paid transport
Population in 1880 (log)	0.238***	0.264***	0.155**	-0.051	0.085	0.096	0.034	0.192*	0.153*	0.237***
	(0.085)	(0.066)	(0.073)	(0.084)	(0.079)	(0.074)	(0.060)	(0.105)	(0.081)	(0.071)
Ancestral population in 1880 (log)	0.012	0.002	0.003	-0.003	0.007	0.022**	0.014	0.008	0.011	0.009
	(0.007)	(0.007)	(0.006)	(0.005)	(0.006)	(0.009)	(0.008)	(0.008)	(0.008)	(0.010)
Log(area)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Ethnic FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Geography and soil resources	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Survey controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
# Observations	12526	12526	12526	12526	12526	12526	12526	12526	12526	12526
R^2	0.378	0.353	0.316	0.246	0.227	0.225	0.238	0.277	0.227	0.313
Mean Dep. Var.	0.512	0.418	0.174	0.896	0.847	0.286	0.544	0.615	0.188	0.706
S.D. Dep. Var.	0.500	0.493	0.379	0.305	0.360	0.452	0.498	0.487	0.390	0.456

* $p<0.10$ ** $p<0.05$ *** $p<0.01$. Respondent-level regressions weighted by cross-country sample weights provided by Afrobarometer. The table presents the estimates for equation 1 in the sample of partitioned ethnic groups whose population shares across borders are above 5% (see section 4.1). Regressions are controlled by the variables depicted in tables A1 and A2, and survey controls are respondent's age, age squared, gender, and whether the interviewer is from the same ethnic groups as the respondent. Standard errors double-clustered for ethnicity and country.

Table E9: Ethnic group size and living conditions reported in Afrobarometer

	Cross-validation: living conditions									
	(1) Index of asset holdings	(2) Urban household	(3) Employed	(4) Occupation: services	(5) Occupation: agriculture	(6) Occupation: unskilled manual	(7) Occupation: skilled manual	(8) Occupation: upper or mid level	(9) Education: no education	(10) Education: university or more
Population in 1880 (log)	0.248**	0.113	-0.008	0.042	-0.205***	0.017	0.016	0.047	-0.109***	0.005
	(0.108)	(0.074)	(0.032)	(0.031)	(0.033)	(0.019)	(0.011)	(0.028)	(0.026)	(0.017)
Ancestral population in 1880 (log)	0.022	0.009	-0.001	0.005	-0.001	-0.006	0.007***	-0.000	-0.004	0.002
	(0.014)	(0.008)	(0.005)	(0.004)	(0.006)	(0.004)	(0.002)	(0.003)	(0.004)	(0.002)
Log(area)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Ethnic FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Geography and soil resources	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Survey controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
# Observations	13211	13211	13211	13211	13211	13211	13211	13211	13211	13211
R^2	0.273	0.308	0.124	0.068	0.232	0.066	0.069	0.045	0.337	0.070
Mean Dep. Var.	-0.056	0.380	0.259	0.134	0.282	0.074	0.065	0.069	0.226	0.048
S.D. Dep. Var.	0.994	0.485	0.438	0.340	0.450	0.262	0.247	0.253	0.418	0.214

* $p<0.10$ ** $p<0.05$ *** $p<0.01$. Respondent-level regressions weighted by cross-country sample weights provided by Afrobarometer. The table presents the estimates for equation 1 in the sample of partitioned ethnic groups whose population shares across borders are above 5% (see section 4.1). Regressions are controlled by the variables depicted in tables A1 and A2, and survey controls are respondent's age, age squared, gender, and whether the interviewer is from the same ethnic groups as the respondent. Standard errors double-clustered for ethnicity and country.

F Appendix: Economic versus Political channels

Table F10: Ethnic group size and political capital

	Panel A: Log of population count Political representation			
	(1) Politically relevant in country	(2) Occupies highest political positions	(3) Powerless or discriminated	(4) Share of years in highest positions since independence
Population in 1880 (log)	0.127** (0.049)	0.150*** (0.047)	-0.184*** (0.050)	0.140*** (0.038)
# Observations	530	530	530	530
R^2	0.747	0.718	0.706	0.747
	Panel B: Population share			
	(1)	(2)	(3)	(4)
Population share in 1880	0.542 (0.396)	0.698* (0.407)	-0.644* (0.373)	0.833** (0.329)
# Observations	530	530	530	530
R^2	0.743	0.710	0.696	0.741
	Panel C: Comparison			
	(1)	(2)	(3)	(4)
Population in 1880 (log)	0.116** (0.052)	0.135** (0.050)	-0.174*** (0.050)	0.120*** (0.040)
Population share in 1880	0.317 (0.408)	0.436 (0.418)	-0.308 (0.367)	0.601* (0.330)
Log(area)	✓	✓	✓	✓
Geography and soil controls	✓	✓	✓	✓
# Observations	530	530	530	530
R^2	0.748	0.720	0.707	0.752
Mean Dep. Var.	0.506	0.191	0.611	0.168
S.D. Dep. Var.	0.500	0.393	0.488	0.331

* p <0.10 ** p <0.05 *** p <0.01. The table presents the estimates for equation 1 in the sample of partitioned ethnic groups whose population shares across borders are above 5% (see section 4.1). Geography controls are distance to coast (log), indicator for being landlocked, terrain ruggedness index, average slope, average precipitation, and malaria suitability index. Agriculture controls are share of irritated area in 1880, share of grazing area in 1880, share of pasture area in 1880, share of cropland area in 1880, mean altitude, maximum altitude, crop suitability index, land suitability level, share of water sites, dominant soil, agro-zones, and hydro-basin class. Standard errors double-clustered for ethnicity and country.

Table F11: Revisiting the baseline model

	Dependent variable: Light density (log)						
	Panel A: Baseline model						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Population count in 1880 (log)	0.266***	0.234***	0.434***	0.397***	0.422***	0.362***	0.392***
	(0.059)	(0.078)	(0.069)	(0.108)	(0.112)	(0.119)	(0.124)
# Observations	540	540	537	535	535	535	535
R^2	0.096	0.792	0.546	0.884	0.890	0.887	0.894
	Panel B: Alternative population measure						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Population share in 1880	3.018***	0.886	3.560***	1.030	1.321	0.876	1.272
	(0.657)	(0.577)	(0.953)	(0.912)	(0.953)	(0.923)	(0.974)
# Observations	540	540	537	535	535	535	535
R^2	0.074	0.785	0.427	0.868	0.875	0.877	0.884
	Panel C: Comparison						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Population count in 1880 (log)	0.214***	0.220***	0.415***	0.389***	0.409***	0.354***	0.375***
	(0.060)	(0.077)	(0.071)	(0.103)	(0.107)	(0.115)	(0.119)
Population share in 1880	2.145***	0.660	0.695	0.306	0.437	0.264	0.511
	(0.624)	(0.587)	(0.603)	(0.788)	(0.820)	(0.816)	(0.866)
Log(area)		✓		✓	✓	✓	✓
Ethnic FE		✓		✓	✓	✓	✓
Country FE			✓	✓	✓	✓	✓
Geography					✓		✓
Soil controls						✓	✓
# Observations	540	540	537	535	535	535	535
R^2	0.130	0.794	0.547	0.884	0.891	0.887	0.894
Mean Dep. Var.	-3.767	-3.767	-3.780	-3.789	-3.789	-3.789	-3.789
S.D. Dep. Var.	1.106	1.106	1.093	1.085	1.085	1.085	1.085

* p <0.10 ** p <0.05 *** p <0.01. The table presents the estimates for equation 1 in the sample of partitioned ethnic groups whose population shares across borders are above 5% (see section 4.1). Regressions are controlled by the variables depicted in tables A1 and A2 and by respondent's age and squared age. Standard errors double-clustered for ethnicity and country.