

Strength in Numbers: Ethnic Group Size and Economic Development

Victoria G. Klarosk* Rogerio Santarrosa*

*Insper

EEA-ESEM 2023

Roadmap

Introduction

Background & empirical strategy

Dimensions of economic development

- Economic consequences of scale

- Political capital

Conclusion

Roadmap

Introduction

Background & empirical strategy

Dimensions of economic development

- Economic consequences of scale

- Political capital

Conclusion

Motivation

- **National ethnic fragmentation** (Easterly and Levine, 1997. Alesina and La Ferrara, 2005.)

Motivation

- **National ethnic fragmentation** (Easterly and Levine, 1997. Alesina and La Ferrara, 2005.)
 - Barriers: communication,

Motivation

- **National ethnic fragmentation** (Easterly and Levine, 1997. Alesina and La Ferrara, 2005.)
 - Barriers: communication, in-group favoritism,

Motivation

- **National ethnic fragmentation** (Easterly and Levine, 1997. Alesina and La Ferrara, 2005.)
 - Barriers: communication, in-group favoritism, conflict of preferences

Motivation

- **National ethnic fragmentation** (Easterly and Levine, 1997. Alesina and La Ferrara, 2005.)
 - Barriers: communication, in-group favoritism, conflict of preferences
 - *Caveat*: cross-country analyses
 - threatens causal interpretation

Motivation

- **National ethnic fragmentation** (Easterly and Levine, 1997. Alesina and La Ferrara, 2005.)
 - Barriers: communication, in-group favoritism, conflict of preferences
 - *Caveat*: cross-country analyses
 - threatens causal interpretation

- Here: study of a dimension of this problem,
 - larger ethnic group \implies higher levels of economic development,taking advantage of a quasi-natural experiment

Potential gains of living around more co-ethnics

Economic gains of scale

- Social capital from shared ethnic affiliation
 - ability of working with co-ethnics, transaction costs (Hjort, 2014. Finseraas et al., 2019.)
- Achieve scale in production
 - preconceived by modern theories of growth
 - empirically hard to address (Peters, 2022. Sequeira, Nunn, and Qian, 2020.)

Construction of political capital

- Ethnic politics: advantage on political competition
 - redistribution of rents from office to co-ethnics (Amodio, Chiovelli, and Hohmann, 2022. Laitin and Ramachandran, 2022. De Luca et al., 2018. Dickens, 2018.)

This paper's approach

- Mostly “erratic” colonial borders with respect to traditional homelands
 - Same ethnic group partitioned across two or more different countries
 - Many ethnic groups within the same country
 - Large persistence of border design
- ⇒ Able to compare within ethnic group, across countries
- Fine-level estimates of historical population size
 - Discard endogenous migration *because of the border*
 - Fine-level measure of economic activity: nighttime light density

Roadmap

Introduction

Background & empirical strategy

Dimensions of economic development

- Economic consequences of scale

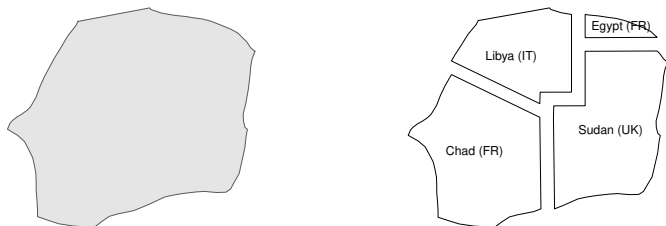
- Political capital

Conclusion

The Scramble for Africa

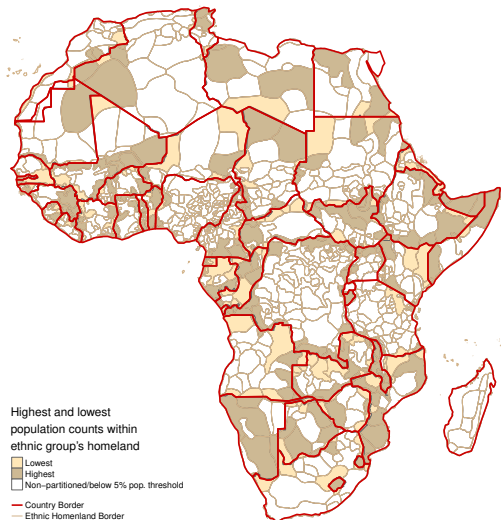
- 1860: Exploration of West Africa
- 1880-1905: Colonial border design
 - The Berlin Conference (1884-1885)
- Well documented “border artificiality” that endured independence
 - Laitin and Ramachandran (2022), Anderson (2018), Michalopoulos and Papaioannou (2016)

Figure 1: Example - The Bideyat homeland as in Murdock (1959), before and after the Scramble



Map of ethnic homelands partitioned by modern borders

Figure 2: "As if" many natural experiments



Data

Main

- Historical homelands: Murdock (1959)
- Modern boundaries: GADM
- Population size of ethnic groups in 1880: Estimates from HYDE (Historical Database of the Global Environment)
 - Combination of methods
- Nighttime light density: VIIRS Day-Night Band (Elvidge et al., 2021)
 - [check](#)

Balancedness and controls

- Geography and ecological variables, and historical land use [detail](#)

Multiple data sources to investigate mechanisms

- Malaria project, GADM, Africapolis, PRIO, USGS, OpenStreet Maps, DHS, Afrobarometer, EPR

Empirical approach

- Specification at the **ethnic homeland-country level** e, c

$$\text{AvgLum}_{e,c} = \delta_e + \delta_c + \beta_2 \log(\text{Population}_{e,c}^{1880}) + \beta_3 X_{e,c} + \varepsilon_{e,c} \quad (1)$$

- $\text{AvgLum}_{e,c}$: $\log(\text{Average light density} + 0.01)$
 - Highly correlated with Gross Cell Product (Nordhaus et al., 2006), which uses primarily national accounts [validation](#)
- δ_e, δ_c : ethnicity and country fixed effects
- $X_{e,c}$: controls for geography and soil's quality and historical use

Summary statistics

	Mean	SD	Min	Max	N
Full sample					
Average luminosity (current), $nW/cm^2/sr$	0.11	0.64	0.00	13.25	1314
Share of lit pixels (current)	0.03	0.08	0.00	0.98	1314
Average luminosity per capita (current), $nW/cm^2/sr$	0.06	1.06	0.00	34.39	1313
Total population in 1880, base 1000	82.76	262.07	0.00	5932.42	1314
Population share in ethnic homeland area in 1880	0.04	0.09	0.00	0.97	1314
Individuals per km^2 in 1880 (base 1000)	0.60	1.39	0.00	35.66	1314
Ethnic group is split PopSh \geq 5% in one country	0.41	0.49	0.00	1.00	1314
Paper sample: split ethnic groups (PopSh\geq5%)					
Average luminosity (current), $nW/cm^2/sr$	0.07	0.59	0.00	13.25	535
Share of lit pixels (current)	0.02	0.05	0.00	0.39	535
Average luminosity per capita (current), $nW/cm^2/sr$	0.11	1.62	0.00	34.39	535
Total population in 1880, base 1000	49.54	95.43	0.19	1054.11	535
Population share in ethnic homeland area in 1880	0.04	0.09	0.00	0.94	535
Individuals per km^2 in 1880 (base 1000)	0.48	0.78	0.00	7.62	535

Main result

Table 2: Effect of population size on nighttime light density

	Light density (log)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Population in 1880 (log)	0.188*** (0.038)	0.137*** (0.042)	0.242*** (0.032)	0.179*** (0.032)	0.181*** (0.031)	0.174*** (0.031)	0.171*** (0.031)
Ethnic FE		✓		✓	✓	✓	✓
Country FE			✓	✓	✓	✓	✓
Geography					✓		✓
Land use						✓	✓
# Observations	535	535	535	535	535	535	535
Adj. R^2	0.067	0.614	0.424	0.723	0.728	0.730	0.734
Mean Dep. Var.	-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

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$. The table presents the estimates for equation 1 in the sample of partitioned ethnic groups whose population shares across borders are above 5%. Standard errors double-clustered for ethnicity and country.

Robustness

- Overall balanced geography and historical land use [check](#)

Robustness

- Overall balanced geography and historical land use [check](#)
- Alternative light measures [check](#)

Robustness

- Overall balanced geography and historical land use [check](#)
- Alternative light measures [check](#)
- Alternative proxies for historical population [check](#)

Robustness

- Overall balanced geography and historical land use [check](#)
- Alternative light measures [check](#)
- Alternative proxies for historical population [check](#)
- Rule out mechanic effect of people on light bulbs [check](#)

Robustness

- Overall balanced geography and historical land use [check](#)
- Alternative light measures [check](#)
- Alternative proxies for historical population [check](#)
- Rule out mechanic effect of people on light bulbs [check](#)
- Effect *beyond path dependence* [check](#)

Robustness

- Overall balanced geography and historical land use [check](#)
- Alternative light measures [check](#)
- Alternative proxies for historical population [check](#)
- Rule out mechanic effect of people on light bulbs [check](#)
- Effect *beyond path dependence* [check](#)
- Validation of main effect with survey data [check](#)

Robustness

- Overall balanced geography and historical land use [check](#)
- Alternative light measures [check](#)
- Alternative proxies for historical population [check](#)
- Rule out mechanic effect of people on light bulbs [check](#)
- Effect *beyond path dependence* [check](#)
- Validation of main effect with survey data [check](#)
- Robust to considering only borders with different colonizers [check](#)

Robustness

- Overall balanced geography and historical land use [check](#)
- Alternative light measures [check](#)
- Alternative proxies for historical population [check](#)
- Rule out mechanic effect of people on light bulbs [check](#)
- Effect *beyond path dependence* [check](#)
- Validation of main effect with survey data [check](#)
- Robust to considering only borders with different colonizers [check](#)
- Spurious correlation of larger area and natural endowments
 - Wealth-creating natural assets
 - Population still dominates endowments [see](#)

Accounting for area

Table 3: Baseline table + conditioning on land area

	Light density (log)				
	(1)	(2)	(3)	(4)	(5)
Population in 1880 (log)	0.179*** (0.032)	0.181*** (0.031)	0.174*** (0.031)	0.171*** (0.031)	0.429*** (0.127)
Log(area excluding water bodies)					-0.282** (0.137)
Ethnic FE	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓
Geography		✓		✓	✓
Soil controls			✓	✓	✓
# Observations	535	535	535	535	535
R ²	0.877	0.885	0.883	0.890	0.896

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$. The table presents the estimates for equation 1 in the sample of partitioned ethnic groups whose population shares across borders are above 5%. Standard errors double-clustered for ethnicity and country.

Roadmap

Introduction

Background & empirical strategy

Dimensions of economic development

- Economic consequences of scale

- Political capital

Conclusion

Dimensions of economic development

(1) Economic effects

- Positive returns to population size mediated by *improvements in space* [urbanization]?
 - Supply of public services and goods
 - Transition to trade/industry?

(2) Interplay of political representation

- Better access to political power?

Agglomeration gains via urbanization

Table 4: Effects of population on urbanization measures

	Measures of urbanisation					
	(1)	(2)	(3)	(4)	(5)	(6)
	Share of urban population	Number of counties	Country's capital within territory	Metropolis within territory	Urban cluster	Quantity of urban clusters
Population in 1880 (log)	0.049* (0.026)	2.323*** (0.815)	0.056*** (0.018)	0.058* (0.031)	2.089*** (0.509)	2.943*** (0.724)
Log(area)	✓	✓	✓	✓	✓	✓
Ethnic FE	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓
Geography and soil controls	✓	✓	✓	✓	✓	✓
# Observations	535	535	535	535	535	535
Adj. R^2	0.474	0.652	0.244	0.272	0.538	0.593
Mean Dep. Var.	0.331	7.030	0.028	0.064	3.662	6.772
S.D. Dep. Var.	0.253	11.051	0.165	0.244	5.876	7.643

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$. 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.

- Higher probability of having any size of urban cluster [detail](#)

Corroborated by higher access to city infrastructure

Table 5: Ethnic group size and objective measures of infrastructure

	All outcomes in log					
	Distance to				Commute time to health facility	
	(1)	(2)	(3)	(4)	(5)	(6)
	Nearest road	Nearest railway	Nearest colonial railway	School (average)	By foot	By motor vehicle
Population in 1880 (log)	-0.436** (0.191)	-0.241*** (0.077)	-0.048 (0.041)	-0.146*** (0.042)	-0.148*** (0.045)	-0.172*** (0.057)
Ethnic FE	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓
Log(area), geography and agricultural controls	✓	✓	✓	✓	✓	✓
# Observations	535	535	535	535	535	535
Adj. R^2	0.358	0.748	0.886	0.897	0.930	0.897
Mean Dep. Var.	8.159	11.618	12.635	10.284	5.272	4.221
S.D. Dep. Var.	1.263	1.161	1.014	1.109	0.953	1.073

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$. Standard errors double-clustered for ethnicity and country. Distances to nearest features are calculated in meters from the ethnic-country pair's polygon centroid. Distance to school is calculated as the average distance of raster points to schools in the homeland-country polygon.

- Survey evidence: Access to electricity, Sewage system, Piped water supply, and Paid transport infrastructure (see), and health outcomes (see)

Organization of education and occupation

- Spatial match of DHS clusters with homeland-country areas
- Last survey available per country
- Sample of women and men aged 25 and above: economic achievements at adulthood
- Same equation as before at individual level

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

where individual characteristics include controls for age and age squared, and regressions are weight by individual weights provided.

Organization of education and occupation

Table 6: Survey evidence on educational attainment and employment

	Panel A: Women's sample					
	Education			Working conditions		
	(1)	(2)	(3)	(4)	(5)	(6)
	Reading proficiency	Years of schooling	Post-sec education	Currently working	In-kind payment for work	Permanent working position
Population in 1880 (log)	0.055* (0.027)	0.618** (0.284)	0.022** (0.008)	0.029 (0.041)	-0.064 (0.046)	0.130* (0.068)
# Observations	120,023	122,471	122,495	118,555	79,508	79,508
Adj. R^2	0.319	0.365	0.051	0.176	0.139	0.260
Mean Dep. Var.	0.427	4.567	0.046	0.702	0.161	0.525
S.D. Dep. Var.	0.495	4.665	0.211	0.457	0.368	0.499
	Panel B: Men's sample					
	(1)	(2)	(3)	(4)	(5)	(6)
Population in 1880 (log)	0.014 (0.022)	0.497* (0.284)	0.035*** (0.011)	0.012 (0.016)	-0.141*** (0.041)	0.099** (0.045)
# Observations	53,836	54,202	54,215	54,206	48,578	48,578
Adj. R^2	0.204	0.271	0.060	0.115	0.140	0.251
Mean Dep. Var.	0.572	6.340	0.096	0.904	0.175	0.627
S.D. Dep. Var.	0.495	5.116	0.294	0.295	0.380	0.484

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$. Standard errors double-clustered for ethnicity and country. Weighted regressions. Adults aged 25 and above. Individual controls are gender, age, and age squared. All regressions controlled by ethnic FE, country FE, and baseline controls.

Organization of education and occupation (continuation)

Table 7: Survey evidence on occupational distribution

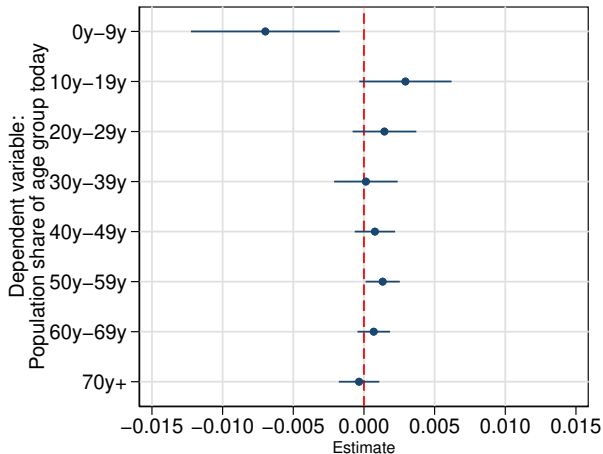
Panel A: Women's sample						
Occupation						
	(1)	(2)	(3)	(4)	(5)	(6)
	Agriculture	Sales and services	Unskilled manual	Skilled manual	Professional	Domestic work
Population in 1880 (log)	-0.091 (0.066)	0.099*** (0.035)	0.004 (0.011)	0.003 (0.005)	0.002 (0.008)	0.008 (0.007)
# Observations	116,298	116,298	116,298	116,298	116,298	116,298
Adj. R^2	0.289	0.175	0.100	0.075	0.039	0.067
Mean Dep. Var.	0.341	0.265	0.043	0.038	0.051	0.018
S.D. Dep. Var.	0.474	0.441	0.203	0.190	0.219	0.132
Panel B: Men's sample						
	(1)	(2)	(3)	(4)	(5)	(6)
Population in 1880 (log)	-0.105** (0.045)	0.041* (0.023)	0.009 (0.019)	0.041*** (0.013)	0.016 (0.015)	0.001 (0.007)
# Observations	52,996	52,996	52,996	52,996	52,996	52,996
Adj. R^2	0.245	0.118	0.086	0.084	0.036	0.087
Mean Dep. Var.	0.409	0.201	0.073	0.145	0.101	0.015
S.D. Dep. Var.	0.492	0.400	0.260	0.352	0.301	0.123

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$. Standard errors double-clustered for ethnicity and country. Weighted regressions. Adults aged 25 and above. Individual controls are gender, age, and age squared. All regressions controlled by ethnic FE, country FE, and baseline controls.

- Consistent evidence across different surveys and samples [show](#)

Demographic composition

Figure 3: Ethnic group size and age distribution today



Note: each row represents a different equation.

Source: SEDAC

Dimensions of economic development

(1) Economic effects

- Positive returns to population size mediated by *improvements in space* [urbanization]?
 - Supply of public services and goods
 - Transition to trade/industry?

(2) Interplay of political representation

- Better access to political power?

Political capital

- Ethnic Power Relations data set [Executive power]
 - Class. of actors: highest positions (monopoly, senior, dominant), junior partner, powerless (excluded) or discriminated
- Match of groups on the basis of linguistic similarity

Political capital

- Ethnic Power Relations data set [Executive power]
 - Class. of actors: highest positions (monopoly, senior, dominant), junior partner, powerless (excluded) or discriminated
- Match of groups on the basis of linguistic similarity

	Political representation			
	(1)	(2)	(3)	(4)
	Politically relevant in country	Occupies highest political positions	Powerless or discriminated	Share of years in highest positions since independence
Population in 1880 (log)	0.128** (0.053)	0.160*** (0.049)	-0.197*** (0.060)	0.148*** (0.039)
Ethnic FE	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
Log(area) + baseline controls	✓	✓	✓	✓
# Observations	530	530	530	530
Adj. R^2	0.380	0.308	0.269	0.385
Mean Dep. Var.	0.506	0.191	0.611	0.168
S.D. Dep. Var.	0.500	0.393	0.488	0.331

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$. Standard errors double-clustered for ethnicity and country. Years since independence: years since first coding of country into EPR.

Discussion: economic scale *versus* bargain power

Table 8: Revisiting the baseline model

	Panel A: Baseline			Panel B: Alternative			Panel C: Comparison		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Population count in 1880 (log)	0.451*** (0.111)	0.359*** (0.121)	0.429*** (0.127)				0.442*** (0.109)	0.353*** (0.118)	0.420*** (0.128)
Population share in 1880				0.827 (0.880)	0.979 (0.868)	0.979 (0.868)	0.378 (0.793)	0.224 (0.785)	0.294 (0.799)
Log(area)	✓	✓	✓	✓	✓	✓	✓	✓	✓
Ethnic FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Geography	✓		✓		✓	✓	✓		✓
Land use		✓	✓	✓	✓	✓		✓	✓
# Observations	535	535	535	535	535	535	535	535	535
Adj. R^2	0.748	0.737	0.747	0.713	0.717	0.717	0.747	0.736	0.746
Mean Dep. Var.	-3.789	-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	1.085

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

- Additional discussion [show](#)

Roadmap

Introduction

Background & empirical strategy

Dimensions of economic development

- Economic consequences of scale

- Political capital

Conclusion

Conclusion

- **A 10% increase in historical population leads to 1.8% increase in light density today**
 - stronger effect when conditioning on land area: 4.3%
 - robust to many threats to identification
 - causal interpretation given design
- **Structural changes in economic organization**
 - Urbanization; Access to public goods; More years of schooling; Composition of occupation
 - Demographic transition
- **Interplay of a political capital channel**
 - More likely to be represented in the national executive power
- **Novel methodology to investigate how population size causes structural transformations in the economy**, contributing to a recent empirical literature that links population size to productivity (Peters, 2022, Sequeira, Nunn, and Qian, 2020)

Thank you!

Victoria Gerenutti Klarosk
victoriagk@al.edu.insper.br

References

- Alesina, Alberto and Eliana La Ferrara (2005). "Ethnic diversity and economic performance". *Journal of economic literature* 43.3, pp. 762–800.
- Amodio, Francesco, Giorgio Chiovelli, and Sebastian Hohmann (2022). "The employment effects of ethnic politics". Available at SSRN 3504599.
- Anderson, Siwan (2018). "Legal origins and female HIV". *American Economic Review* 108.6, pp. 1407–1439.
- De Luca, Giacomo et al. (2018). "Ethnic favoritism: An axiom of politics?". *Journal of Development Economics* 132, pp. 115–129.
- Dickens, Andrew (2018). "Ethnolinguistic favoritism in African politics". *American Economic Journal: Applied Economics* 10.3, pp. 370–402.
- Easterly, William and Ross Levine (1997). "Africa's growth tragedy: policies and ethnic divisions". *The quarterly journal of economics*, pp. 1203–1250.

References

- Elvidge, Christopher D et al. (2021). "Annual time series of global VIIRS nighttime lights derived from monthly averages: 2012 to 2019". *Remote Sensing* 13.5, p. 922.
- Finseraas, Henning et al. (2019). "Trust, ethnic diversity, and personal contact: A field experiment". *Journal of Public Economics* 173, pp. 72–84.
- Hjort, Jonas (2014). "Ethnic divisions and production in firms". *The Quarterly Journal of Economics* 129.4, pp. 1899–1946.
- Laitin, David D and Rajesh Ramachandran (2022). "Linguistic diversity, official language choice and human capital". *Journal of Development Economics* 156, p. 102811.
- Michalopoulos, Stelios and Elias Papaioannou (2016). "The long-run effects of the scramble for Africa". *American Economic Review* 106.7, pp. 1802–48.
- Murdock, George Peter (1959). "Africa, its peoples and their culture history".

References

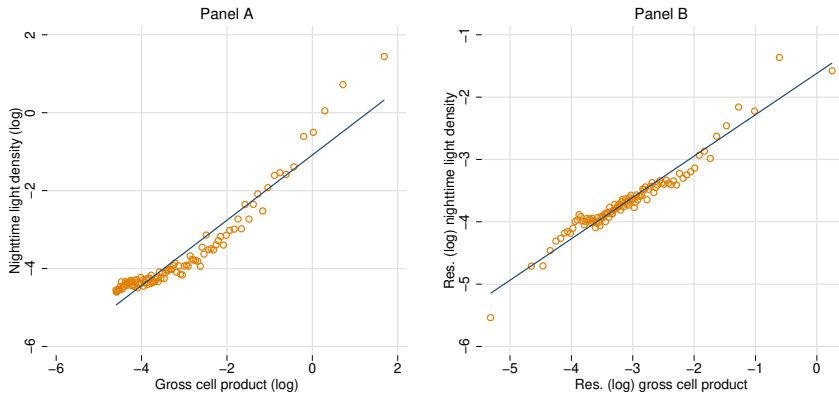
- Nordhaus, William et al. (2006). "The G-Econ database on gridded output: methods and data". *Yale University, New Haven* 6.
- Nunn, Nathan and Diego Puga (2012). "Ruggedness: The blessing of bad geography in Africa". *Review of Economics and Statistics* 94.1, pp. 20–36.
- Peters, Michael (2022). "Market Size and Spatial Growth—Evidence From Germany's Post-War Population Expulsions". *Econometrica* 90.5, pp. 2357–2396.
- Sequeira, Sandra, Nathan Nunn, and Nancy Qian (2020). "Immigrants and the Making of America". *The Review of Economic Studies* 87.1, pp. 382–419.

Table 9: Description of baseline controls

Geography		
Name	Source	Note
Distance to the coast	NASA	Raster file
Indicator for landlocked		Based on the map of Africa
Altitude in log	FAO-Network/GAEZ	Raster file, 30arc sec
Soil suitability	FAO-Network/GAEZ	Raster file, 30arc sec
Terrain ruggedness	Nunn and Puga, 2012	
Slope (%)	Nunn and Puga, 2012	
Malaria suitability	Malaria Project	Raster file
Share of land underwater	SEDAC	Raster file of water surface, 30arc sec
Ground water depth (m)	AQUASTAT/FAO	Latest estimate
Precipitation	FAO, CRU CL 2.0	Average per month in mm during 1961-1990
Aridity index	Resource Watch	1970-2000
Soil characteristics		
Name	Source	Note
Share irrigated cropland	HYDE	Raster file for 1880
Share rain-fed cropland	HYDE	Raster file for 1880
Share pasture grazing land	HYDE	Raster file for 1880
Share rangeland grazin land	HYDE	Raster file for 1880
Share grazing land converted to rangeland	HYDE	Raster file for 1880

back

Figure 4: Plots of light density versus gross cell product at the grid-cell level for Africa



[back](#)

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 $\ln(- + 0.01)$.

Figure 5: Average light density in pairs tribe-country

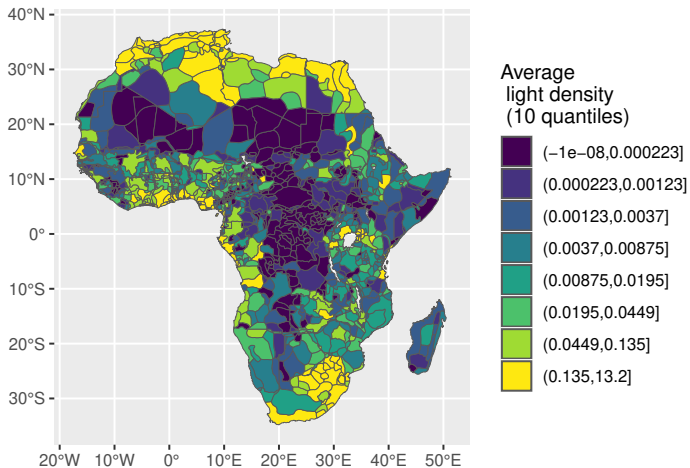


Table 10: Geography [back](#)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Log(distance to coast)	Landlocked indicator	Log(mean altitude)	Soil suitability index	Ruggedness index	Average slope in %	Malaria suitability	Share of land underwater	Ground water depth (m)	Precipitation in mm/month for 1961-1990	Aridity index for 1970-2000
Population in 1880 (log)	0.019* (0.010)	-0.040*** (0.012)	0.037*** (0.012)	60.836 (84.437)	0.000 (0.000)	0.000 (0.000)	-0.001 (0.002)	0.000 (0.001)	2.244* (1.130)	-2.751 (5.391)	-24.400 (36.891)
Ethnic FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
# Observations	535	535	535	535	535	535	528	535	535	535	535
Adj. R^2	0.972	0.711	0.946	0.881	0.782	0.783	0.939	0.413	0.922	0.984	0.982
Mean Dep. Var.	5.946	0.882	6.064	4770.193	0.000	0.006	0.203	0.007	91.097	1037.027	5896.364
S.D. Dep. Var.	1.206	0.323	0.975	3598.435	0.000	0.007	0.158	0.019	59.286	615.912	4531.451

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

Table 11: Estimated land use in 1880, shares

	Cropland		Grazing land		
	(1)	(2)	(3)	(4)	(5)
	Irrigated	Rain-fed	Pasture	Rangeland	Converted rangeland
Population in 1880 (log)	0.006 (0.006)	-0.096 (0.152)	0.369** (0.172)	-0.029 (0.308)	-0.038 (0.092)
Ethnic FE	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓
# Observations	535	535	535	535	535
Adj. R^2	0.324	0.788	0.666	0.764	0.750
Mean Dep. Var.	0.028	3.157	3.267	9.647	1.369
S.D. Dep. Var.	0.185	4.437	5.674	11.715	4.138

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

Table 12: Alternative light measures [back](#)

	Share of lit pixels			Log of luminosity per capita		
	(1)	(2)	(3)	(4)	(5)	(6)
Population in 1880 (log)	0.003* (0.001)	0.003** (0.001)	0.003** (0.001)	0.064* (0.033)	0.046** (0.019)	0.043** (0.018)
Ethnic FE	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓
Geography		✓	✓		✓	✓
Land use			✓			✓
# Observations	535	535	535	535	535	535
Adj. R^2	0.737	0.742	0.750	0.600	0.700	0.698
Mean Dep. Var.	0.024	0.024	0.024	-4.262	-4.262	-4.262
S.D. Dep. Var.	0.052	0.052	0.052	0.757	0.757	0.757

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

Table 13: 2SLS model for historical population size of ethnic groups [back](#)

	2nd stage estimates			1st stage estimates		
	(1)	(2)	(3)	(4)	(5)	(6)
Population in 1880 (log)	0.113** (0.045)	0.119*** (0.045)	0.114** (0.047)			
<i>Instruments</i>						
Area in km^2 (log)				0.096** (0.039)	0.085** (0.040)	0.102** (0.045)
Suitability index					0.000* (0.000)	
Slope					19.279 (12.102)	
Distance to lake/river (km)					0.001 (0.001)	
Ethnic FE	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓
Geography			✓			✓
Land use			✓			✓
# Observations	535	535	535	535	535	535
K-P F-statistic	255.901	74.351	284.664			
Mean Dep. Var.	-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

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

Table 14: Effects of population size on nighttime light density, controlling for population density [back](#)

	Dep. var.: Light density (log)						
	Test: mechanic effect?				Test: path dependence?		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Population in 1880 (log)	0.137*** (0.029)	0.141*** (0.028)	0.136*** (0.028)	0.137*** (0.028)	0.147*** (0.033)	0.168*** (0.031)	0.167*** (0.031)
Log of population density today	0.345*** (0.063)	0.348*** (0.059)	0.318*** (0.065)	0.325*** (0.063)			
Log of population density in 1880					0.282** (0.137)		
At least one urban center in sec. XVIII						0.770** (0.356)	
Log(Pop urban centers in sec. XVIII + 0.01)							0.056** (0.027)
Ethnic FE	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓
Geography		✓		✓	✓	✓	✓
Land use			✓	✓	✓	✓	✓
# Observations	535	535	535	535	535	535	535
Adj. R ²	0.758	0.762	0.756	0.761	0.747	0.736	0.736
Mean Dep. Var.	-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

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

- Pop. density today: NASA Socioeconomic Data and Applications Center (SEDAC)

- Urban centers in sec. XVIII: Historical Urban Population: 3700 BC - AD 2000, NASA Socioeconomic Data and Applications Center (SEDAC)

Validation of main outcome using the DHS

Table 15: Ethnic group size and living conditions rob

	Wealth index						
	Validation of sample	Overall	Private assets			Public assets	Urban indicator
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Log(light density)		Overall	Dwelling quality	Appliances and devices	Overall	
Population in 1880 (log)	0.599*** (0.140)	0.190*** (0.060)	0.164*** (0.059)	0.225*** (0.059)	0.109** (0.052)	0.280*** (0.065)	0.154*** (0.049)
Ethnic FE	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓
Log(area), geography and Land use	✓	✓	✓	✓	✓	✓	✓
# Observations	264	196,099	196,099	196,099	196,099	196,099	196,099
R ²	0.743	0.192	0.183	0.382	0.160	0.344	0.317
Mean Dep. Var.	-3.644	0.008	0.011	-0.088	-0.046	-0.161	0.335
Sd Dep. Var.	1.029	0.963	0.955	0.922	0.970	0.956	0.472

* p<0.10 ** p<0.05 *** p<0.01. Weighted regressions. The table presents the estimates for equation 1 in the sample of partitioned ethnic groups whose population shares across borders are above 5%. Standard errors double-clustered for ethnicity and country.

Sample of ethnic groups split across different colonizers

Table 16: Effects of population size on nighttime light density [back](#)

	Light density (log)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Population in 1880 (log)	0.178*** (0.057)	0.108* (0.061)	0.279*** (0.038)	0.134*** (0.042)	0.150*** (0.049)	0.131*** (0.039)	0.143*** (0.047)
Ethnic FE		✓		✓	✓	✓	✓
Country FE			✓	✓	✓	✓	✓
Geography					✓		✓
Land use						✓	✓
# Observations	335	335	335	335	335	335	335
Adj. R^2	0.059	0.592	0.456	0.718	0.726	0.733	0.739
Mean Dep. Var.	-3.800	-3.800	-3.800	-3.800	-3.800	-3.800	-3.800
S.D. Dep. Var.	1.108	1.108	1.108	1.108	1.108	1.108	1.108

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

Influence of natural endowments

- Point estimates are stable to the inclusion of geography and soil quality controls
- Yet, larger groups ended up in larger areas: higher probability of coming across economically exploitable natural endowments?

Table 17: Correlation of population size and economically meaningful natural endowments

	Water availability			Mineral reserves			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	River extension per km ² of land	Lakes	Major inland water bodies	Onshore oil fields	Diamond mines	Other gems mines	Gold mines
Population in 1880 (log)	0.000 (0.002)	0.083*** (0.024)	0.063* (0.037)	-0.037 (0.025)	0.013 (0.033)	0.008 (0.024)	0.042 (0.029)
Ethnic FE	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓
Geography and land use	✓	✓	✓	✓	✓	✓	✓
# Observations	535	535	535	535	535	535	535
Adj. R ²	0.439	0.466	0.608	0.513	0.321	0.403	0.205
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%. Standard errors double-clustered for ethnicity and country. Dependent variables of columns (2) through (7) are indicators.

Influence of natural endowments

Table 18: Main table, controlling for natural endowments rob

	Dep. var.: Light density in log							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Population in 1880 (log)	0.175*** (0.031)	0.156*** (0.035)	0.178*** (0.031)	0.177*** (0.032)	0.187*** (0.031)	0.192*** (0.036)	0.445*** (0.121)	0.429*** (0.127)
River length	2.869 (2.041)					2.975 (2.111)	2.365 (1.951)	
Lake		-0.232 (0.175)				-0.287* (0.164)	-0.362** (0.174)	
Major water bodies		0.229 (0.146)				0.245 (0.148)	0.295** (0.144)	
Onshore oil fields			-0.200 (0.255)			-0.231 (0.250)	-0.082 (0.227)	
Diamond mines				-0.125 (0.128)		-0.084 (0.125)	-0.060 (0.116)	
Other gems mines					-0.500*** (0.163)	-0.555*** (0.159)	-0.513*** (0.154)	
Gold mines					-0.029 (0.186)	-0.055 (0.186)	-0.090 (0.180)	
Log(area excluding water bodies)							-0.287** (0.131)	-0.282** (0.137)
Ethnic FE	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Geography and land use	✓	✓	✓	✓	✓	✓	✓	✓
Observations	535	535	535	535	535	535	535	535
Adj. R ²	0.734	0.735	0.733	0.733	0.739	0.742	0.755	0.747
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%. Standard errors double-clustered for ethnicity and country. Dependent variables of columns (2) through (7) are indicators.

Table 19: Ethnic group size and urbanization

	Quantity of urban clusters according to population size					
	(1)	(2)	(3)	(4)	(5)	(6)
	10k to 20k	20k to 50k	50k to 100k	100k to 500k	500k to 2.5mi	More than 2.5mi
Population in 1880 (log)	0.100** (0.046)	0.124*** (0.045)	0.091** (0.042)	0.077* (0.040)	0.072** (0.032)	0.029** (0.014)
Log(area)	✓	✓	✓	✓	✓	✓
Ethnic FE	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓
Geography and soil controls	✓	✓	✓	✓	✓	✓
# Observations	535	535	535	535	535	535
Adj. R^2	0.417	0.397	0.339	0.405	0.407	0.359
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*

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.

Table 20: Consistent results across surveys

	Panel A: Demographic and Health Surveys (DHS)					
	Validation	Education		Working conditions		
	(1)	(2)	(3)	(4)	(5)	(6)
	Log(light density)	Post-sec education	Secondary or higher education)	Currently working	In-kind payment for work	Permanent working position
Population in 1880 (log)	0.599*** (0.140)	0.028*** (0.008)	0.027** (0.012)	0.018 (0.032)	-0.088*** (0.032)	0.131** (0.051)
# Observations	264	176,710	176,710	172,761	128,086	128,086
Adj. R^2	0.743	0.060	0.130	0.182	0.119	0.237
Mean Dep. Var.	-3.644	0.062	0.125	0.766	0.167	0.563
S.D. Dep. Var.	1.029	0.241	0.330	0.424	0.373	0.496
Panel B: Afrobarometer round 7						
	(1)	(2)	(3)	(4)	(5)	(6)
Population in 1880 (log)	0.526** (0.187)	0.080** (0.035)	0.115*** (0.036)	0.106** (0.040)		
Ethnic FE	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓
Log(area) + controls	✓	✓	✓	✓	✓	✓
Individual controls		✓	✓	✓	✓	✓
# Observations	146	14,113	14,113	14,067		
R^2	0.913	0.109	0.203	0.125		
Mean Dep. Var.	1.000	0.139	0.279	0.318		
S.D. Dep. Var.	0.000	0.345	0.449	0.466		

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$. Standard errors double-clustered for ethnicity and country. Men and women samples, controlling for age, age squared, and gender.

Table 21: Consistent results across surveys

Panel A: Demographic and Health Surveys (DHS)						
	Occupation					
	(1)	(2)	(3)	(4)	(5)	(6)
	Agriculture	Sales and services	Unskilled manual	Skilled manual)	Professional	Domestic work
Population in 1880 (log)	-0.106* (0.058)	0.078*** (0.026)	0.010 (0.011)	0.023*** (0.007)	0.006 (0.007)	0.005 (0.006)
# Observations	169,294	169,294	169,294	169,294	169,294	169,294
Adj. R^2	0.252	0.135	0.084	0.088	0.041	0.062
Mean Dep. Var.	0.362	0.245	0.052	0.071	0.067	0.017
S.D. Dep. Var.	0.481	0.430	0.223	0.257	0.249	0.129
Panel B: Afrobarometer round 7						
	Occupation					
	(1)	(2)	(3)	(4)	(5)	(6)
Population in 1880 (log)	-0.167*** (0.039)	0.072*** (0.020)	0.014 (0.023)	0.053*** (0.016)	0.043 (0.028)	-0.015 (0.022)
Ethnic FE	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓
Log(area) + controls	✓	✓	✓	✓	✓	✓
Individual controls	✓	✓	✓	✓	✓	✓
# Observations	14,149	14,149	14,149	14,149	14,149	14,149
R^2	0.241	0.065	0.067	0.066	0.044	0.164
Mean Dep. Var.	0.280	0.144	0.088	0.074	0.094	0.092
S.D. Dep. Var.	0.449	0.351	0.284	0.262	0.292	0.289

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$. Standard errors double-clustered for ethnicity and country. Men and women samples, controlling for age, age squared, and gender.

Table 22: Ethnic group size and public infrastructure, according to Afrobarometer

	Available at the primary sampling area									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Electricity grid	Piped water system	Sewage system	Cell phone service	School	Police station	Health clinic	Market stalls	Bank	Paid transport
Population in 1880 (log)	0.246*** (0.072)	0.359*** (0.083)	0.183*** (0.053)	0.103** (0.045)	0.153** (0.069)	0.220*** (0.060)	0.027 (0.093)	0.192* (0.109)	0.156* (0.083)	0.234*** (0.080)
Log(area)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Ethnic FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Geography and soil resources	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Survey controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
# Observations	17987	17855	17726	17912	17940	17809	17772	17885	17887	17951
Adj. R ²	0.424	0.350	0.352	0.237	0.185	0.193	0.205	0.252	0.199	0.280
Mean Dep. Var.	0.596	0.483	0.206	0.917	0.841	0.296	0.532	0.659	0.188	0.741
S.D. Dep. Var.	0.491	0.500	0.404	0.275	0.366	0.457	0.499	0.474	0.391	0.438

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$. Respondent-level regressions weighted by cross-country sample weights provided by Afrobarometer. Standard errors double-clustered for ethnicity and country.

Table 23: Health outcomes in the DHS samples of women aged 25 and above

	Health indicators				Last pregnancy				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	z-score of weight for height (full sample)	Ln(body mass index)	Indicator for anemia	Infant mortality	Months until 1st check-up	Place of birth: public facility	Place of birth: private facility	Place of birth: home	Log(child's weight)
Population in 1880 (log)	0.131** (0.059)	0.036*** (0.013)	-0.009 (0.039)	-0.051*** (0.010)	-0.159 (0.101)	-0.019 (0.043)	0.038 (0.031)	0.012 (0.033)	-0.012 (0.012)
Ethnic FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Log(area), geography and soil, individual's characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓
Birth history					✓	✓	✓	✓	✓
# Observations	49,995	49,995	49,995	47,097	65,645	65,645	65,645	65,645	47,278
R ²	0.184	0.170	0.069	0.050	0.148	0.304	0.143	0.214	0.032
Mean Dep. Var.	-0.123	7.769	0.431	0.898	4.070	0.587	0.067	0.189	1.145
Sd Dep. Var.	0.892	0.194	0.495	0.181	1.563	0.492	0.249	0.391	0.220

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

Table 24: Representation in national politics [Executive Power] [back](#)

Panel A: Log of population count [benchmark]				
	Political representation			
	(1)	(2)	(3)	(4)
	Politically relevant in country	Occupies highest political positions	Powerless or discriminated	Share of years in highest positions since independence
Population in 1880 (log)	0.128** (0.053)	0.160*** (0.049)	-0.197*** (0.060)	0.148*** (0.039)
# Observations	530	530	530	530
Adj. R^2	0.380	0.308	0.269	0.385
Panel B: Population share				
	(1)	(2)	(3)	(4)
Population share in 1880	0.259 (0.413)	0.579 (0.447)	-0.489 (0.373)	0.733* (0.371)
# Observations	530	530	530	530
Adj. R^2	0.368	0.282	0.239	0.361
Panel C: Comparison				
	(1)	(2)	(3)	(4)
Population in 1880 (log)	0.126** (0.055)	0.151*** (0.051)	-0.192*** (0.059)	0.134*** (0.040)
Population share in 1880	0.058 (0.420)	0.339 (0.449)	-0.183 (0.362)	0.520 (0.370)
Ethnic FE	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
Log(area) + baseline controls	✓	✓	✓	✓
# Observations	530	530	530	530
Adj. R^2	0.377	0.308	0.266	0.390
Mean Dep. Var.	0.506	0.191	0.611	0.168
S.D. Dep. Var.	0.500	0.393	0.488	0.331

Extras

Heterogeneity

Table 25: Effects of ethnic group size on the distribution of light density

	<i>N</i>-th quantile of light density (log)										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	<i>N</i> = 10	<i>N</i> = 20	<i>N</i> = 30	<i>N</i> = 40	<i>N</i> = 50	<i>N</i> = 60	<i>N</i> = 70	<i>N</i> = 80	<i>N</i> = 90	<i>N</i> = 95	<i>N</i> = 99
Population in 1880 (log)	0.065 (0.055)	0.073 (0.074)	0.076 (0.087)	0.077 (0.096)	0.078 (0.104)	0.078 (0.110)	0.087 (0.117)	0.115 (0.126)	0.206 (0.140)	0.530*** (0.193)	0.317** (0.154)
Log(area)	0.055 (0.052)	0.110 (0.072)	0.151* (0.085)	0.184* (0.095)	0.211** (0.103)	0.234** (0.109)	0.243** (0.116)	0.231* (0.128)	0.174 (0.145)	-0.021 (0.205)	0.098 (0.159)
Ethnic FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Geography and soil	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
# Observations	535	535	535	535	535	535	535	535	535	535	535
Adj. R^2	0.232	0.262	0.275	0.284	0.289	0.294	0.302	0.315	0.370	0.585	0.449
Mean Dep. Var.	-3.841	-3.464	-3.209	-3.017	-2.861	-2.732	-2.614	-2.503	-2.366	-1.701	-2.228
S.D. Dep. Var.	0.455	0.639	0.762	0.855	0.929	0.992	1.054	1.118	1.227	1.837	1.369

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

Heterogeneity

Table 26: 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)	0.455** (0.178)	0.179*** (0.051)	0.467*** (0.137)	0.031*** (0.009)
Log(area)	0.021 (0.190)	-0.118** (0.056)	-0.277* (0.148)	-0.018* (0.010)
Ethnic FE	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓
# Observations	535	535	535	535
Adj. R^2	0.685	0.785	0.754	0.811
Mean Dep. Var.	-2.337	0.429	-3.519	0.098
S.D. Dep. Var.	1.779	0.489	1.270	0.105

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$. Standard errors double-clustered for ethnicity and country. Dependent variables of columns (2) through (7) are binary indicators.

Heterogeneity

Table 27: 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)	0.386*** (0.116)	0.305 (0.211)	0.700*** (0.173)
Population in 1880 (log) × Control's size	0.004 (0.002)		
Log(area)	✓	✓	✓
Ethnic FE	✓	✓	✓
Country FE	✓	✓	✓
Geography and soil	✓	✓	✓
# Observations	535	278	248
Adj. R ²	0.750	0.680	0.782
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	3.91%	[1.55%, 6.27%]	
50th percentile	4.20%	[1.76%, 6.65%]	
90th percentile	6.16%	[2.20%, 10.1%]	

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

Private versus Public gains of population (DHS)

Table 28: Correlations after conditioning on individual and household covariates

	Wealth index				Index of private assets		Index of public assets	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Population in 1880 (log)	0.197*** (0.063)	0.117** (0.056)	0.161*** (0.058)	0.197*** (0.063)	0.064 (0.055)	0.051 (0.055)	0.208*** (0.075)	0.106** (0.048)
Urban household		0.586*** (0.048)				0.172*** (0.041)		0.854*** (0.074)
Ln(Years of educ + 1)			0.273*** (0.022)					
Currently working				0.083* (0.043)				
Index of public assets					0.414*** (0.033)	0.372*** (0.033)		
Index of private assets							0.324*** (0.025)	0.237*** (0.023)
Female=1	-0.064*** (0.013)	-0.059*** (0.012)	0.032** (0.012)	-0.047** (0.020)	-0.059*** (0.011)	-0.058*** (0.011)	0.009 (0.010)	0.010 (0.009)
Ethnic FE	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Log(area), geography and soil, individul's characteristics	✓	✓	✓	✓	✓	✓	✓	✓
# Observations	172,724	172,724	172,724	172,724	172,724	172,724	172,724	172,724
R ²	0.200	0.252	0.253	0.201	0.298	0.301	0.449	0.555
Mean Dep. Var.	0.162	0.162	0.162	0.162	0.165	0.165	-0.083	-0.083
Sd Dep. Var.	1.000	1.000	1.000	1.000	0.992	0.992	0.991	0.991

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

Survey: social capital

Table 29: Ethnic group size and attitudes reported in Afrobarometer

	Social and cultural values					Access to information and mobilization			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Score: Trust in formal institutions	Score: Trust in informal institutions	Score: Difference (2) - (1)	Score: State capacity perception	Score: Gender equality	Score: Liberal leaning views	Binary: Consumes news everyday	Score: Access to public sector information	Score: Reporting corrupt behavior
Population in 1880 (log)	0.071 (0.103)	-0.139 (0.091)	-0.210*** (0.067)	0.267*** (0.072)	0.212* (0.108)	-0.155 (0.117)	0.143** (0.067)	0.220*** (0.075)	0.188*** (0.052)
Female	0.001 (0.019)	0.039*** (0.013)	0.038** (0.018)	-0.046** (0.020)	0.026 (0.020)	-0.075*** (0.021)	-0.119*** (0.018)	-0.090*** (0.029)	-0.048** (0.021)
Ethnic FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Log(area), geography and soil, survey controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
# Observations	16526	16526	16526	16526	16526	16526	16526	16526	16526
Adj. R^2	0.184	0.170	0.085	0.133	0.109	0.122	0.157	0.116	0.132
Mean Dep. Var.	1.624	1.988	0.364	3.790	4.096	3.039	0.581	1.362	1.707
S.D. Dep. Var.	0.881	0.975	0.848	0.993	0.925	0.922	0.493	0.940	0.924

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$. Respondent-level regressions weighted by cross-country sample weights provided by Afrobarometer. Standard errors double-clustered for ethnicity and country.

Afrobarometer: Current location versus ancestral homeland

Table 30: Ethnic group size and attitudes reported in Afrobarometer

	Social and cultural values					Access to information and mobilization		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Score: Trust in formal institutions	Score: Trust in informal institutions	Score: State capacity perception	Score: Gender equality	Score: Liberal leaning views	Binary: Consumes news everyday	Score: Access to public sector information	Score: Reporting corrupt behavior
Population in 1880 (log)	0.111 (0.163)	-0.178* (0.092)	0.288*** (0.051)	0.212* (0.111)	-0.027 (0.120)	0.174** (0.067)	0.284** (0.130)	0.220** (0.086)
Ancestral population in 1880 (log)	-0.003 (0.010)	-0.009 (0.016)	0.015 (0.012)	0.007 (0.014)	-0.011 (0.011)	0.000 (0.005)	-0.011 (0.013)	-0.007 (0.011)
Log(area)	✓	✓	✓	✓	✓	✓	✓	✓
Ethnic FE	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Geography and soil resources	✓	✓	✓	✓	✓	✓	✓	✓
Survey controls	✓	✓	✓	✓	✓	✓	✓	✓
# Observations	12133	12133	12133	12133	12133	12133	12133	12133
Adj. R ²	0.206	0.177	0.146	0.116	0.135	0.151	0.121	0.127
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. Standard errors double-clustered for ethnicity and country.

Afrobarometer: Current location versus ancestral homeland

Table 31: Ethnic group size and public infrastructure, according to Afrobarometer

	Available at the primary sampling area									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Electricity grid	Piped water system	Sewage system	Cell phone service	School	Police station	Health clinic	Market stalls	Bank	Paid transport
Population in 1880 (log)	0.261** (0.100)	0.356*** (0.082)	0.204*** (0.062)	0.070 (0.095)	0.132 (0.093)	0.145* (0.074)	-0.006 (0.093)	0.256** (0.107)	0.183** (0.080)	0.338*** (0.095)
Ancestral population in 1880 (log)	0.012* (0.007)	0.003 (0.007)	0.004 (0.006)	-0.003 (0.005)	0.007 (0.006)	0.023** (0.009)	0.016* (0.008)	0.010 (0.008)	0.011 (0.008)	0.009 (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
Adj. R^2	0.367	0.340	0.305	0.232	0.211	0.213	0.222	0.262	0.214	0.302
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. Standard errors double-clustered for ethnicity and country.

Afrobarometer: Current location versus ancestral homeland

Table 32: Ethnic group size and living conditions reported in Afrobarometer

	Cross-validation: living conditions									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Index of asset holdings	Urban household	Employed	Occupation: services	Occupation: agriculture	Occupation: unskilled manual	Occupation: skilled manual	Occupation: upper or mid level	Education: no education	Education: university or more
Population in 1880 (log)	0.296*** (0.097)	0.069 (0.059)	0.074* (0.036)	0.015 (0.029)	-0.155*** (0.038)	0.026 (0.016)	0.032** (0.013)	0.039 (0.031)	-0.131*** (0.035)	-0.001 (0.014)
Ancestral population in 1880 (log)	0.022* (0.013)	0.009 (0.007)	-0.001 (0.005)	0.005 (0.004)	-0.001 (0.005)	-0.006 (0.003)	0.007*** (0.002)	-0.001 (0.003)	-0.005 (0.004)	0.002 (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
Adj. R^2	0.262	0.297	0.112	0.054	0.221	0.052	0.055	0.032	0.327	0.057
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. Standard errors double-clustered for ethnicity and country.