SUPPLEMENT TO "PRE-COLONIAL ETHNIC INSTITUTIONS AND CONTEMPORARY AFRICAN DEVELOPMENT"

By Stelios Michalopoulos and Elias Papaioannou[1]

This supplement reports (i) descriptive statistics, (ii) further cross-validation tests of the luminosity data, (iii) numerous sensitivity checks, and (iv) the detailed Data Appendix.

A.1. DESCRIPTIVE ANALYSIS

Figure A.1(A) reports the distribution of satellite light density at night across ethnic homelands. Figure A.1(B) plots the distribution of the log of light density, adding a small number ($y_{i,c} \equiv \ln(0.01 + \text{Light Density}_{i,c})$). The latter is the dependent variable used in Tables II, III(A), and IV. This transformation assures that (i) we use all observations, and (ii) we minimize the problem of outliers. Figure A.1(C) graphs the distribution of log light density across lit ethnic–country areas ($y_{i,c} \equiv \ln(\text{Light Density}_{i,c})$). We use log luminosity as the dependent variable in Table III(B) so as to show that our results at the ethnic homeland level are not sensitive to the transformation of the luminosity data.

Table A.I, Panel A reports summary statistics for all control variables at the ethnic-country homeland level. Table A.I, Panel B reports summary statistics for the control variables at the pixel level.

A.2. FURTHER CROSS-VALIDATION CHECKS

We performed several cross-validation exercises investigating the relationship between luminosity and various economic indicators across African countries and regions within countries.

Cross-Country Analysis

We start by examining whether luminosity correlates with development across African countries. Figure A.2(A) illustrates the unconditional correlation between log light density and log GDP per capita in 2000. The R^2 is 0.35 and the estimate is more than 6 standard errors larger than zero. Besides economic performance, light density also reflects urbanization. Figure A.2(B) shows the relationship between log GDP per capita and log light density, partialling out log population density. The relationship between log light density and log GDP p.c. is now stronger (the coefficient increases from 0.31 to 0.47 and the *t*-stat jumps to 10).
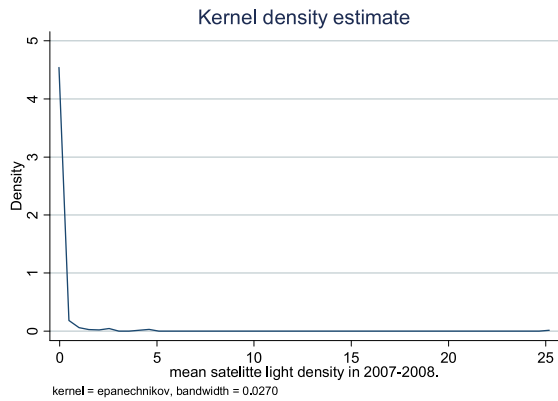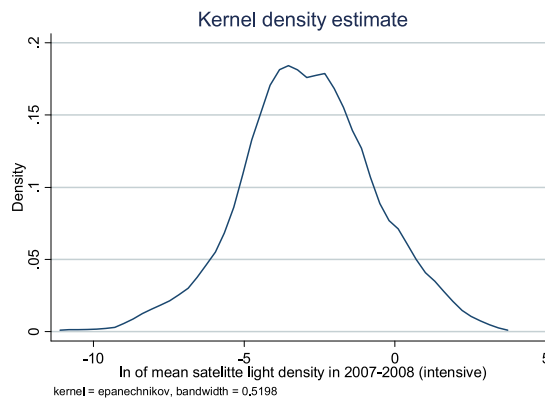

(A)
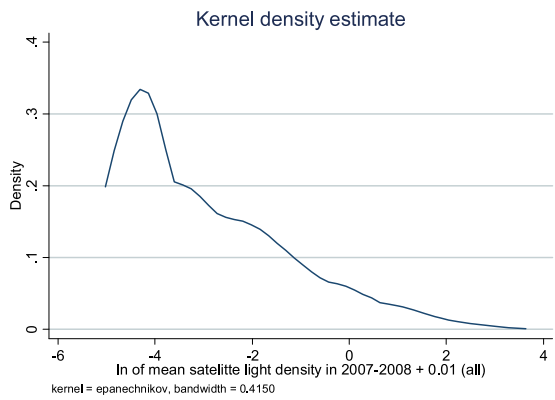
(B)

(C)

FIGURE A.1.—The distribution of luminosity across ethnic homelands.

TABLE A.I

SUMMARY STATISTICS FOR THE CONTROL VARIABLES[a]

Variable	Obs.	Mean	St. Dev.	p25	Median	p75	Min	Max
Panel A: Ethnic Homeland Sample								
ln(0.01 + Population Density)	683	2.90	2.02	2.18	3.24	4.20	−4.61	7.43
ln(1 + Water Area)	683	0.29	0.44	0.00	0.12	0.36	0.00	3.12
ln(Surface Area)	683	2.16	1.60	1.14	2.23	3.25	−2.21	6.20
Mean Elevation	683	0.63	0.47	0.28	0.49	1.00	0.00	2.18
Land Suitability for Agriculture	683	0.43	0.24	0.28	0.45	0.59	0.00	0.98
Malaria Stability Index	683	0.72	0.33	0.50	0.87	1.00	0.00	1.00
Oil Deposit Indicator	683	0.08	0.27	0.00	0.00	0.00	0.00	1.00
Diamond Mine Indicator	683	0.10	0.30	0.00	0.00	0.00	0.00	1.00
Distance to the Capital City	683	0.50	0.39	0.24	0.40	0.63	0.01	1.88
Distance to the Sea Coast	683	0.59	0.42	0.20	0.56	0.92	0.00	1.70
Distance to the National Border	683	0.10	0.11	0.02	0.06	0.15	0.00	0.62
Rule of Law in 2007	680	−0.88	0.53	−1.35	−0.96	−0.42	−2.63	0.65
Log GDP p.c. in 2007	680	7.12	0.90	6.67	7.10	7.57	5.07	9.89
Jurisdictional Hierarchy	683	1.22	0.97	0.00	1.00	2.00	0.00	4.00
Political Centralization	683	0.36	0.48	0.00	0.00	1.00	0.00	1.00
Panel B: Pixel Sample								
ln(Population Density)	66,570	2.37	1.85	1.02	2.48	3.72	−2.20	10.10
Water Body Indicator	66,570	0.14	0.34	0.00	0.00	0.00	0.00	1.00
ln(Area)	66,570	5.11	0.27	5.12	5.23	5.25	3.91	5.26
Mean Elevation	66,570	690.65	523.95	293.89	547.61	1064.89	−219.48	4314.56
Land Suitability for Agriculture	66,314	0.38	0.28	0.16	0.35	0.57	0.00	1.00
Malaria Stability Index	66,428	0.70	0.41	0.30	1.00	1.00	0.00	1.00
Oil Deposit Indicator	66,570	0.03	0.18	0.00	0.00	0.00	0.00	1.00
Diamond Mine Indicator	66,570	0.00	0.06	0.00	0.00	0.00	0.00	1.00
Distance to the Capital City	66,570	507.02	372.43	236.65	403.13	667.76	1.46	1923.01
Distance to the Sea Coast	66,570	576.12	423.34	200.35	509.78	905.03	0.55	1701.14
Distance to the National Border	66,570	141.82	124.59	44.65	106.35	206.35	0.67	755.60
Jurisdictional Hierarchy	66,570	1.49	1.03	1.00	1.00	2.00	0.00	4.00
Political Centralization	66,570	0.50	0.50	0.00	0.00	1.00	0.00	1.00

[a]The table reports descriptive statistics for all the control variables employed in the empirical analysis. Panel A reports summary statistics for all control variables in the country-ethnicity sample. Panel B reports summary statistics for all control variables in the pixel sample. The Data Appendix gives detailed variable definitions and data sources.

Cross-Region Analysis

Figures A.3(A) and A.3(B) plot luminosity and infant mortality across African administrative regions, using data from the Center for International Earth Science Information Network (CIESIN) of Columbia University's Earth Institute, which are comparable across African countries.[2] The figures illustrate the significantly negative correlation between log light density and infant mortality across 264 African regions. The estimate is −9.44 with a *t*-stat of 9;

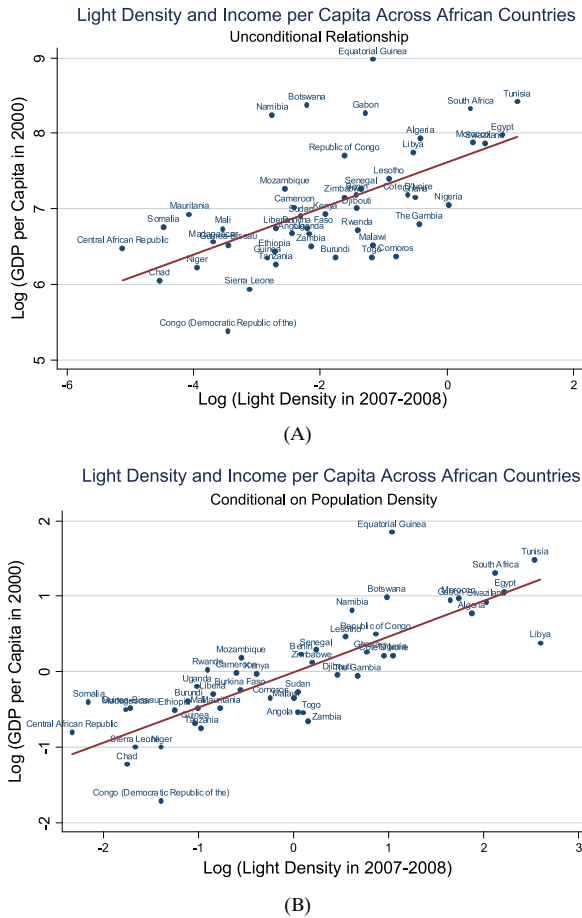[2]The data are available at: http://sedac.ciesin.columbia.edu/povmap/.

(A)

(B)

FIGURE A.2.—Luminosity and income per capita across African countries.

when we condition on log population density, the estimate increases in absolute value (−14.89), retaining its significance at the 99% confidence level.

Analysis across DHS Enumeration Areas

Table A.II reports the unconditional and conditional on population density correlations between light density and the DHS household composite wealth index, presence of electricity, and average schooling, for four large African countries across all African regions. Across all countries studied, Nigeria, Tanzania, Zimbabwe, and the Democratic Republic of Congo, there is highly significant (within-country) correlation between luminosity and these proxy measures of development (see also Figures 2(A)–2(D) in the main body of the paper for a graphical illustration).
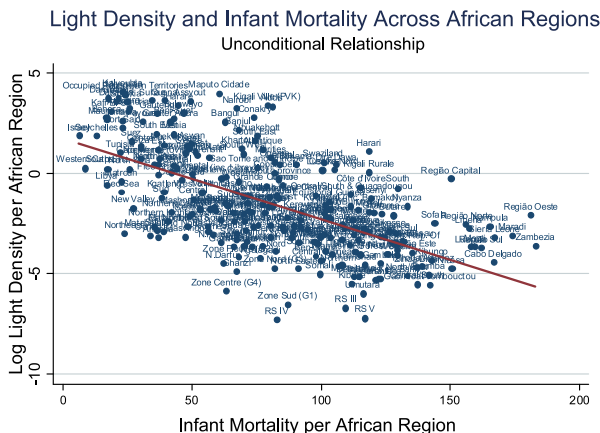
(A)
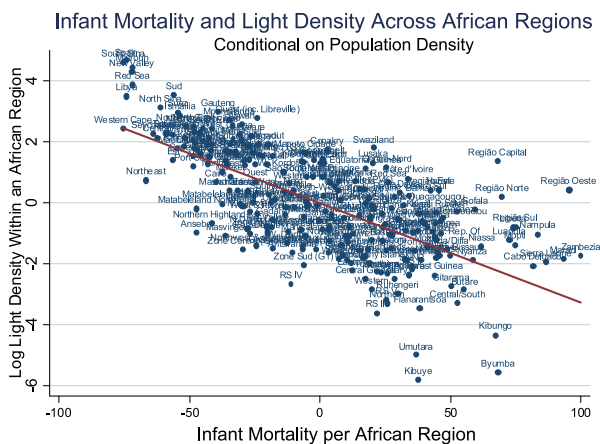
(B)

FIGURE A.3.—Luminosity and infant mortality across African regions.

A.3. FURTHER SENSITIVITY CHECKS

We performed various sensitivity checks to explore the robustness of our results.

Outliers

In column (1) of Table A.III, we report estimates dropping observations where the dependent variable exceeds the 99th percentile.[3] This is useful be-

---

[3]In the previous drafts of the paper, we also estimated specifications excluding also the top 2% or the top 5% of the dependent variable. If anything, the coefficient on the jurisdictional hierarchy index increases and becomes more accurately estimated.

TABLE A.II

SATELLITE LIGHT DENSITY AT NIGHT AND DHS MEASURES OF HOUSEHOLD WEALTH, ELECTRIFICATION, AND SCHOOLING[a]

	Tanzania 2007		Zimbabwe 2005		Nigeria 2008		Congo, DRC 2007	
	Unconditional Correlation	Conditional on PD	Unconditional Correlation	Conditional on PD	Unconditional Correlation	Conditional on PD	Unconditional Correlation	Conditional on PD
DHS Composite Wealth Index	0.7828	0.7104	0.8809	0.7540	0.7219	0.5496	0.8599	0.7424
Electrification	0.7045	0.6296	0.8423	0.7012	0.6064	0.4188	0.7733	0.6461
Years of Schooling	0.5575	0.5035	0.7171	0.5191	0.5603	0.5385	0.6344	0.5378
Enumeration Areas	466	466	398	398	886	886	293	293

[a]The table reports unconditional and conditional on log population density correlations between satellite light density at night ($\log(0.01 + \text{light density})$) and three proxy measures of development from the Demographic and Health Surveys (DHS) for Tanzania, Zimbabwe, Nigeria, and the Democratic Republic of Congo. The unit of analysis is the enumeration area of the Demographic and Health Surveys. As within each enumeration area there are several households, we average household responses across the enumeration area and use the respective mean values. Light density at night is the average light density in a radius of 10 km from the centroid of the DHS enumeration area. The DHS composite wealth index is derived using principal component analysis taking into account access of households to electricity, toilet facilities, clean drinking water, dwelling characteristics, etc. Schooling years denotes average years of schooling of the household's head. Electrification denotes whether households have access to electricity. All correlations are significant at the 1% confidence level.

TABLE A.III

SENSITIVITY ANALYSIS AT THE ETHNIC-COUNTRY LEVEL[a]

	Excluding						
	Outliers	Capitals	North	South	Central	East	West
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Jurisdictional Hierarchy	0.1825***	0.1754***	0.1834***	0.1676***	0.1327**	0.2218***	0.1655**
	(0.0529)	(0.0576)	(0.0526)	(0.0525)	(0.0560)	(0.0703)	(0.0652)
Adjusted R-squared	0.64	0.657	0.609	0.656	0.697	0.689	0.669
Observations	675	651	642	619	523	495	449
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Population Density	Yes	Yes	Yes	Yes	Yes	Yes	Yes

[a]The table reports within-country OLS estimates associating regional development with pre-colonial ethnic institutions, as reflected in Murdock's (1967) index of jurisdictional hierarchy beyond the local community index. The dependent variable is the $\log(0.01 + \text{light density at night from satellite})$ at the ethnicity-country level. In column (1) we exclude from the estimation observations where the dependent variable exceeds the 99th percentile. In column (2) we exclude observations where capital cities fall. In column (3) we exclude observations that fall in North Africa. In column (4) we exclude observations that fall in Southern Africa. In column (5) we exclude observations that fall in Central Africa. In column (6) we exclude observations that fall in East Africa. In column (7) we exclude observations that fall in West Africa. The regional classification follows Nunn (2008).

All specifications include a set of country fixed effects (constants not reported). In all specifications we control for population density, location, and geography. Specifically, the set of control variables includes the $\log(0.01 + \text{population density})$, the distance of the centroid of each ethnicity-country area from the respective capital city, the distance from the sea coast, the distance from the national border, $\log(1 + \text{area under water (lakes, rivers, and other streams)})$, $\log(\text{surface area})$ land suitability for agriculture, elevation, a malaria stability index, a diamond mine indicator, and an oil field indicator. The Data Appendix gives detailed variable definitions and data sources. Below the estimates, we report in parentheses double-clustered standard errors at the country and the ethnolinguistic family dimensions. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

cause there are some extreme values in the luminosity data (see discussion in Section 2 and the descriptive statistics in Table I(A)). The coefficient on the jurisdictional hierarchy index (0.183) is unchanged compared to the analogous specification in column (4) of Table III(A) (0.177). We also repeated estimation dropping ethnic areas where capital cities fall. This is an appropriate check because capitals are likely to be populated by people from several ethnic groups and thus the ethnic-specific index of jurisdictional hierarchy index may be inappropriate. Moreover, by excluding areas where capitals fall, we further account for outliers in the dependent variable. The positive and highly significant correlation between pre-colonial political centralization and regional development remains intact (column (2)).

Regional Heterogeneity

We also repeated our analysis excluding each time a different African region (using the classification of Nunn (2008). Doing so allows us to account for the fact that colonization took different forms across Africa, which may have affected the role of local institutions during the colonial era and after independence. Table A.III, columns (3)–(7) report the results. The estimates show that the strong positive association between pre-colonial ethnic political centralization and regional development is not driven by a particular African region.

Population Density

In Table A.IVA, we report cross–sectional and within-country estimates associating pre-colonial ethnic institutions with contemporary population density at the ethnic-country level. Table A.IVB reports analogous specifications at the pixel level. While urbanization has been linked to the emergence of complex political institutions before colonization in Africa and thus these estimates may suffer from endogeneity, it is useful to see whether the complexity of pre-colonial ethnic institutions correlates with contemporary population density, since the latter may be capturing (to some extent at least) the level of development. There is a significant and positive correlation between the jurisdictional hierarchy beyond the local community index and contemporary population density. However, the estimates indicate that ethnic institutions are less strongly related to population density as compared to luminosity; for example, the standardized "beta" coefficient for the models in columns (1) and (3) is 0.08 and 0.10; the analogous specifications with log luminosity as the dependent variable (column (1), Table II and column (1), Table III(A)) yield beta coefficients of 0.22 and 0.19, respectively. These comparisons suggest that while political centralization does interact with population density, its relationship with economic development is stronger.

TABLE A.IVA

PRE-COLONIAL ETHNIC INSTITUTIONS AND REGIONAL POPULATION DENSITY ETHNIC HOMELAND ANALYSIS[a]

	Jurisdictional Hierarchy (0–4)				Political Centralization (0–1)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ethnic Institutions	0.1687	0.2854***	0.2455**	0.2348**	0.2903	0.5689***	0.3890*	0.4461**
	(0.1104)	(0.0769)	(0.1173)	(0.1114)	(0.2729)	(0.1880)	(0.2351)	(0.2215)
Adjusted R-squared	0.008	0.391	0.403	0.557	0.006	0.391	0.400	0.557
Observations	683	683	682	682	683	683	682	682
Country Fixed Effects	No	No	Yes	Yes	No	No	Yes	Yes
Location Controls	No	Yes	No	Yes	No	Yes	No	Yes
Geographic Controls	No	Yes	No	Yes	No	Yes	No	Yes

[a]The table reports cross-sectional and within-country OLS estimates associating regional development, as reflected in log population density, with pre-colonial ethnic institutions. In columns (1)–(4) we measure pre-colonial ethnic institutions using Murdock's (1967) jurisdictional hierarchy beyond the local community index. In columns (5)–(8) we use a binary political centralization index that is based on Murdock's (1967) jurisdictional hierarchy beyond the local community variable. Following Gennaioli and Rainer (2007), this index takes on the value of zero for stateless societies and ethnic groups that are part of petty chiefdoms and 1 otherwise (for ethnicities that are organized as large paramount chiefdoms and ethnicities that are part of large states).

The specifications in columns (3), (4), (7), and (8) include a set of country fixed effects (constants not reported). In even-numbered columns we control for location and geography. The set of control variables includes the distance of the centroid of each ethnicity-country area from the respective capital city, the distance from the sea coast, the distance from the national border, log(1 + area under water (lakes, rivers, and other streams)), log(surface area), land suitability for agriculture, elevation, a malaria stability index, a diamond mine indicator, and an oil field indicator. The Data Appendix gives detailed variable definitions and data sources. Below the estimates, we report in parentheses double-clustered standard errors at the country and the ethnolinguistic family dimensions. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

TABLE A.IVB

PRE-COLONIAL ETHNIC INSTITUTIONS AND REGIONAL POPULATION DENSITY PIXEL-LEVEL ANALYSIS[a]

	Jurisdictional Hierarchy (0–4)				Political Centralization (0–1)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ethnic Institutions	0.3183***	0.2425*	0.2107**	0.2105***	0.4487**	0.3315	0.3957**	0.3985**
	(0.1095)	(0.1325)	(0.1005)	(0.0811)	(0.2275)	(0.2426)	(0.1939)	(0.1612)
Adjusted R-squared	0.031	0.26	0.335	0.345	0.015	0.254	0.334	0.344
Observations	66,570	66,570	66,173	66,173	66,570	66,570	66,173	66,173
Country Fixed Effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Controls Pixel Level	No	No	Yes	Yes	No	No	Yes	Yes
Controls Ethnic-Country Level	No	No	No	Yes	No	No	No	Yes

[a]The table reports OLS estimates associating regional development, as reflected in log population density, with pre-colonial ethnic institutions. In columns (1)–(4) we measure pre-colonial ethnic institutions using Murdock's (1967) jurisdictional hierarchy beyond the local community index. In columns (5)–(8) we use a binary political centralization index that is based on Murdock's (1967) jurisdictional hierarchy beyond the local community variable. Following Gennaioli and Rainer (2007), this index takes on the value of zero for stateless societies and ethnic groups that are part of petty chiefdoms and 1 otherwise (for ethnicities that are organized as large paramount chiefdoms and ethnicities that are part of large states). The unit of analysis is a pixel of 0.125×0.125 decimal degrees (around 12×12 kilometers).

Specifications in columns (2)–(4) and (6)–(8) include a set of country fixed effects (constants not reported). In columns (3), (4), (7), and (8) we control for a set of geographic and location variables at the pixel level. The set of controls includes the distance of the centroid of each pixel from the respective capital city, its distance from the sea coast, its distance from the national border, an indicator for pixels that have water (lakes, rivers, and other streams), an indicator for pixels with diamond mines, an indicator for pixels with oil fields, the pixel's land suitability for agriculture, pixel's mean elevation, pixel's average value of a malaria stability index, and the log of the pixel's area. In columns (4) and (8) we also control for location and geographic features at the ethnic-country level. Specifically, the set of control variables includes the distance of the centroid of each ethnicity-country area from the respective capital city, the distance from the sea coast, the distance from the national border, $\log(1 + \text{area under water (lakes, rivers, and other streams)})$, $\log(\text{surface area})$, land suitability for agriculture, elevation, a malaria stability index, a diamond mine indicator, and an oil field indicator. The Data Appendix gives detailed variable definitions and data sources. Below the estimates, we report in parentheses double-clustered standard errors at the country and the ethnolinguistic family dimensions. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Results With Afrobarometer's Living Condition Index and Education

While we consider the use of the satellite images at night as a proxy for development part of our contribution, we also estimated specifications with alternative proxies of regional development using the Afrobarometer Surveys. In Table A.V, we report specifications associating pre-colonial ethnic institutions with two measures from the 2005 Afrobarometer Surveys that capture economic well-being, a living conditions index and years of schooling. The Afrobarometer Surveys are based on interviews conducted in a random sample of either 1200 or 2400 individuals of voting age in 17 Sub-Saharan African countries.[4] As a result, the sample drops significantly not only because the number of countries and ethnicities falls, but also because, in many countries, we do not have adequate spatial coverage across all ethnic areas. We assigned each respondent's current residence area (Afrobarometer's enumeration area) to one of the 834 ethnic homelands in Murdock's (1959) ethnolinguistic map and then aggregated the living conditions index and schooling at the ethnic homeland level in each country. We then estimated our baseline specification (equation (1)) at the ethnic-country homeland level using the log of the living conditions index and the log of average years of schooling as the dependent variable.[5] The results with these alternative proxies of development are in line with our estimates with luminosity. On average, respondents residing in ethnic homelands with centralized pre-colonial political institutions report better living conditions. Similarly, the average level of education is higher in ethnic areas of politically centralized societies. Yet it should be noted that due to the limited number of respondents within each ethnic area, these results should be cautiously interpreted. Moreover, in some reasonable permutations (e.g., when we use the level of the living conditions index or the level of education as the dependent variable), the coefficient on the political centralization index loses significance at standard confidence levels.

A.4. DATA APPENDIX

A.4.1. *Variables at the Ethnicity-Country Level/ Pixel Level*

Light Density at Night: Light Density is the average luminosity across pixels that fall within the unit of analysis. We use the average of the val-

---

[4]These countries are Benin, Botswana, Ghana, Kenya, Lesotho, Madagascar, Malawi, Mali, Mozambique, Namibia, Nigeria, Senegal, South Africa, Tanzania, Uganda, Zambia, and Zimbabwe. We thank Nathan Nunn for providing us the data.

[5]The living conditions index reflects respondents' view of their present living conditions, which can be: (i) very bad, (ii) fairly bad, (iii) neither good nor bad, (iv) fairly good, or (v) very good.

The schooling data are constructed aggregating respondents' answers on the following categories: (i) no formal schooling, (ii) informal schooling only, (iii) some primary schooling, (iv) primary school completed, (v) some secondary school/high school, (vi) secondary school completed/high school, (vii) post-secondary qualifications, but no university, (viii) some university, (ix) university completed, and (x) post-graduate.

TABLE A.V

PRE-COLONIAL ETHNIC INSTITUTIONS AND REGIONAL DEVELOPMENT WITHIN AFRICAN COUNTRIES RESULTS WITH AFROBAROMETER
SURVEY DATA ON LIVING CONDITIONS AND SCHOOLING[a]

	Living Conditions Index				Years of Schooling			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Binary Political Centralization	0.0693**	0.0692**	0.0696**	0.0592**	0.1180*	0.1134*	0.1359**	0.0949*
Double-clustered s.e.	(0.0282)	(0.0286)	(0.0338)	(0.0292)	(0.0605)	(0.0665)	(0.0621)	(0.0527)
Adjusted R-squared	0.358	0.358	0.372	0.395	0.496	0.497	0.523	0.549
Observations	194	194	194	194	193	193	193	193
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Population Density	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Location Controls	No	No	Yes	Yes	No	No	Yes	Yes
Geographic Controls	No	No	No	Yes	No	No	No	Yes

[a]The table reports within-country OLS estimates associating regional development, as reflected in a living conditions index and years of schooling, with pre-colonial ethnic institutions. We measure pre-colonial ethnic institutions using a binary political centralization index that is based on Murdock's (1967) jurisdictional hierarchy beyond the local community variable. Following Gennaioli and Rainer (2007), this index takes on the value of zero for stateless societies and ethnic groups that are part of petty chiefdoms and 1 otherwise (for ethnicities that are organized as large paramount chiefdoms and ethnicities that are part of large states). All specifications include a set of country fixed effects (constants not reported). In columns (1)–(4) the dependent variable is the log of a living conditions index, as reported in the 2005 Afrobarometer Surveys. In columns (5)–(8) the dependent variable is the log of years of schooling, as reported in the 2005 Afrobarometer Surveys.

In columns (2), (3), (4), (6), (7), and (8) we control for $\log(0.01 + \text{population density})$. In columns (3), (4), (7), and (8) we control for location augmenting the specification with distance of the centroid of each ethnicity-country area from the respective capital city, distance from the closest sea coast, and distance from the national border. The set of geographic controls in columns (4) and (8) includes $\log(1 + \text{area under water (lakes, rivers, and other streams)})$, log(surface area), land suitability for agriculture, elevation, a malaria stability index, a diamond mine indicator, and an oil field indicator. The Data Appendix gives detailed variable definitions and data sources. Below the estimates, we report in parentheses double-clustered standard errors at the country and the ethnolinguistic family dimensions. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

TABLE A.VI

PRE-COLONIAL ETHNIC INSTITUTIONS AND REGIONAL DEVELOPMENT PIXEL-LEVEL ANALYSIS[a]

	Lit/Unlit Pixels					ln(0.01 + Luminosity)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Binary Political Centralization	0.1000**	0.0690**	0.0460***	0.0565***	0.0486***	0.5183*	0.3524**	0.2369***	0.3094***	0.2806***
Double-clustered s.e.	(0.0489)	(0.0317)	(0.0149)	(0.0150)	(0.0122)	(0.2723)	(0.1785)	(0.0913)	(0.0903)	(0.0808)
Adjusted R-squared	0.018	0.269	0.357	0.375	0.379	0.022	0.315	0.416	0.447	0.456
Country Fixed Effects	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Population Density	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Controls Pixel Level	No	No	No	Yes	Yes	No	No	No	Yes	Yes
Controls Ethnic-Country Level	No	No	No	No	Yes	No	No	No	No	Yes
Observations	66,570	66,570	66,570	66,173	66,173	66,570	66,570	66,570	66,173	66,173

[a]The table reports OLS estimates associating regional development, as reflected in satellite light density at night, with pre-colonial ethnic institutions. We measure pre-colonial ethnic institutions using a binary political centralization index that is based on Murdock's (1967) jurisdictional hierarchy beyond the local community variable. Following Gennaioli and Rainer (2007), this index takes on the value of zero for stateless societies and ethnic groups that are part of petty chiefdoms and 1 otherwise (for ethnicities that are organized as large paramount chiefdoms and ethnicities that are part of large states). The unit of analysis is a pixel of 0.125×0.125 decimal degrees. In columns (1)–(5) the dependent variable is a dummy variable that takes on the value of 1 if the pixel is lit and zero otherwise. In columns (6)–(10) the dependent variable is the log(0.01 + light density at night from satellite).

In columns (3)–(5) and (8)–(10) we control for ln(pixel population density). In columns (4), (5), (9), and (10) we control for a set of geographic and location variables at the pixel level. The set of controls includes the distance of the centroid of each pixel from the respective capital city, its distance from the sea coast, its distance from the national border, an indicator for pixels that have water (lakes, rivers, and other streams), an indicator for pixels with diamond mines, an indicator for pixels with oil fields, the pixel's land suitability for agriculture, pixel's mean elevation, pixel's average value of a malaria stability index, and the log of the pixel's area. In columns (5) and (10) we also control for location and geographic features at the ethnic-country level. Specifically, the set of control variables includes the distance of the centroid of each ethnicity-country area from the respective capital city, the distance from the sea coast, the distance from the national border, log(1 + area under water (lakes, rivers, and other streams)), log(surface area), land suitability for agriculture, elevation, a malaria stability index, a diamond mine indicator, and an oil field indicator. The Data Appendix gives detailed variable definitions and data sources. Below the estimates, we report in parentheses double-clustered standard errors at the country and the ethnolinguistic family dimensions. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

ues in 2007 and 2008. In the regressions at the ethnic homeland we use log(0.01 + Average Luminosity) or log(Average Luminosity). In the regressions at the pixel level we use as the dependent variable a dummy variable that takes on the value of 1 if the area is lit and zero otherwise; we also use the log(0.01 + Average Luminosity). *Source: Available at http://www.ngdc.noaa.gov/ dmsp/downloadV4composites.html.*

Population Density: log(0.01 + population density per sq. km in 2000). For the pixel-level analysis we use log(population density per sq. km in 2000). *Source: Nelson, Andy, 2004. African Population Database Documentation, UNEP GRID Sioux Falls. Available at http://na.unep.net/siouxfalls/datasets/ datalist.php.*

Water Area: log(1 + total area covered by rivers or lakes in sq. km). For the pixel-level analysis we use an indicator that equals 1 if there is some water body at the pixel. *Source: Constructed using the "Inland water area features" dataset from Global Mapping International, Colorado Springs, CO, USA. Global Ministry Mapping System.*

Elevation: Average elevation in km in each ethnicity-country or in each pixel. *Source: National Oceanic and Atmospheric Administration (NOAA) and U.S. National Geophysical Data Center, TerrainBase, release 1.0 (CD-ROM), Boulder, Colorado. Available at http://www.sage.wisc.edu/atlas/data.php?incdataset= Topography.*

Land Suitability for Agriculture: Average land quality for cultivation within each ethnicity-country or within each pixel. The index is the product of two components capturing the climatic and soil suitability for farming. *Source: Michalopoulos (2012); Original Source: Atlas of the Biosphere. Available at http://www.sage.wisc.edu/iamdata/grid_data_sel.php.*

Malaria Stability Index: The index takes into account the prevalence and type of mosquitoes indigenous to a region, their human biting rate, their daily survival rate, and their incubation period. We use the average value within each ethnicity-country or within each pixel. *Source: Kiszewski, Mellinger, Spielman, Malaney, Sachs, and Sachs (2004).*

Distance to the Capital City: The geodesic distance from the centroid of each ethnic group in a country or of each pixel to the capital city of the country it belongs to, measured in 1000s of km's. *Source: Calculated using the Haversine formula.*

Distance to the Sea Coast: The geodesic distance from the centroid of each ethnic group in a country or of each pixel to the nearest coastline, measured in 1000s of km's. *Source: Global Mapping International, Colorado Springs, Colorado, USA. Series name: Global Ministry Mapping System. Series issue: Version 3.0.*

Distance to the Border: The geodesic distance from the centroid of each ethnic group in a country or from each pixel to the nearest border, measured in 1000s of km's. *Source: Calculated using ArcGis.*

Petroleum: Indicator variable that equals 1 if there is an oil field in the homeland or pixel of ethnic group i in country c. *Source: The Petroleum*

Dataset v.1.1 contains information on all known on-shore oil and gas deposits throughout the world. Available at http://www.prio.no/CSCW/Datasets/ Geographical-and-Resource/Petroleum-Dataset/Petroleum-Dataset-v11.

Diamond: Indicator variable that equals 1 if there is a diamond mine in the homeland or pixel of ethnic group *i* in country *c*. *Source: Map of Diamond Resources. Available at www.prio.no/CSCW/Datasets/Geographical-and-Resource/Diamond-Resources/*.

A.4.2. *Country-Level Variables*

Rule of Law: The index is "capturing perceptions of the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence." The standardized index ranges from -2.5 to $+2.5$ with higher values indicating better functioning institutions. We use the value of the index in 2007. *Source: World Bank Governance Matters Indicators Database (Kaufman, Kraay, and Mastruzzi (2005)). Available at http://info.worldbank.org/governance/wgi/index.asp*.

Income: Log of per capita GDP in PPP in 2007. *Source: Penn World Tables*.

A.4.3. *Pre-Colonial Ethnicity-Level Variables*

Jurisdictional Hierarchy Beyond Local Community: Ordered variable ranging from 0 to 4 indicating the number of jurisdictional levels (political complexity) in each society above the local level. A 0 indicates stateless societies, 1 and 2 indicate petty and large paramount chiefdoms (or their equivalent), 3 and 4 indicate large states. *Source: Murdock (1967); variable code in the Ethnolinguistic Atlas v33; a revised version of Murdock's Atlas has been made available by J. Patrick Gray at http://eclectic.ss.uci.edu/~drwhite/worldcul/ EthnographicAtlasWCRevisedByWorldCultures.sav*.

Centralization Indicator: This binary index takes the value 0 if the Jurisdictional Hierarchy Beyond Local Community variable equals 0 or 1. The index takes on the value 1 if the Jurisdictional Hierarchy Beyond Local Community variable equals 2, 3, and 4. This aggregation follows Gennaioli and Rainer (2006, 2007). *Source: Murdock (1967)*.

Gathering: Binary index that indicates the reliance of the economy on "*the collection of wild plants and small land fauna.*" The index equals zero when the dependence is between 0% and 5%; the index equals 1 when dependence is greater than 5% dependence. *Source: Murdock (1967); variable code in the Ethnographic Atlas v1*.

Hunting: Binary index that indicates the intensity in hunting (including trapping and fowling). The index equals zero when the dependence is between 0% and 5%; the index equals 1 when dependence is greater than 5%. *Source: Murdock (1967); variable code in the Ethnographic Atlas v2*.

Fishing: Binary index that indicates the intensity in fishing (including shell fishing and the pursuit of large aquatic animals). The index equals zero when the dependence is between 0% and 5%; the index equals 1 when dependence is greater than 5%. *Source: Murdock (1967); variable code in the Ethnographic Atlas v3*.

Animal Husbandry: 0–9 scale index reflecting the percentage of subsistence coming from animal husbandry for each ethnicity at the time of colonization. The index equals 0 when there is 0%–5% dependence; 1 when there is 6%–15% dependence; 2 when there is 16%–25% dependence; 3 when there is 26%–35% dependence; 4 when there is 36%–45% dependence; 5 when there is 46%–55% dependence; 6 when there is 56%–65% dependence; 7 when there is 66%–75% dependence; 8 when there is 76%–85% dependence; and 9 when there is 86%–100% dependence. *Source: Murdock (1967); variable code in the Ethnographic Atlas v4*.

Milking: Binary index that equals 1 when "*domestic animals are milked more often that sporadically*" and zero when "*little or no milking*." *Source: Murdock (1967); variable code in the Ethnographic Atlas v41*.

Agriculture Dependence: 0–9 scale index reflecting the intensity of agriculture. "*It includes penetration of the soil, planting, tending the growing crops, and harvesting but not subsequent food preparation*." The index equals 0 when there is 0%–5% dependence; 1 when there is 6%–15% dependence; 2 when there is 16%–25% dependence; 3 when there is 26%–35% dependence; 4 when there is 36%–45% dependence; 5 when there is 46%–55% dependence; 6 when there is 56%–65% dependence; 7 when there is 66%–75% dependence; 8 when there is 76%–85% dependence; and 9 when there is 86%–100% dependence. *Source: Murdock (1967); variable code in the Ethnographic Atlas v5*.

Agriculture Type: 0–4 scale index reflecting the type of agriculture. The index equals 0 when there is "*no agriculture*"; 1 when there is "*causal agriculture*"; 2 when there is "*extensive or shifting agriculture*"; 3 when there is "*intensive agriculture*"; and 4 when there is "*intensive irrigated agriculture*." *Source: Murdock (1967); variable code in the Ethnographic Atlas v28*.

Polygyny: Indicator that equals 1 when polygyny is practiced and zero otherwise. *Source: Murdock (1967); variable code in the Ethnographic Atlas v8*.

Alternative Polygyny: Alternative indicator that equals 1 when polygyny is practiced and zero otherwise, based on Murdock's domestic organization variable. The indicator equals zero when "*independent nuclear, monogamous families*" are the norm and 1 when polygyny is common. *Source: Murdock (1967); variable code in the Ethnographic Atlas v9; the construction of the variable follows Fenske (2009)*.

Clan Communities: Indicator that equals 1 when Murdock's community marriage organization variable indicates that "*clan communities or clan barrios*" are present and zero otherwise. *Source: Murdock (1967); variable code in the Ethnographic Atlas v15*.

Settlement Pattern: Ordered variable ranging from 1 to 8 quantifying "*settlement pattern of each group*." 1 indicates fully nomadic (migratory) groups, 2

indicates semi-nomadic groups, 3 indicates semi-sedentary groups, 4 identifies groups that live in compact and impermanent settlements, 5 indicates societies living in neighborhoods of dispersed family homes, 6 indicates groups in separated hamlets forming a single community, 7 indicates societies living in compact and relatively permanent settlements, and 8 denotes groups residing in complex settlements. *Source: Murdock (1967); variable code in the Ethnographic Atlas v30.*

Complex Settlements: Indicator that equals 1 for ethnicities living in compact and relatively permanent settlements (v30 = 7) or in complex settlements (v30 = 8), and zero otherwise. *Source: Murdock (1967); variable code in the Ethnographic Atlas v30.*

Jurisdictional Hierarchy of Local Community: Ordered variable ranging from 0 to 2 reflecting the hierarchy of local community organization. A zero score indicates the theoretical minimum of two (e.g., family and band), while a score of 2 indicates the theoretical maximum of four levels (e.g., nuclear family, extended family, clan barrio, village levels). *Source: Murdock (1967); variable code in the Ethnographic Atlas v32.*

Patrilineal Descent Type: Indicator that equals 1 if the society is characterized by patrilineal descent and zero otherwise. *Source: Murdock (1967); variable code in the Ethnographic Atlas v43.*

Class Stratification: Ordered variable ranging from 0 to 4 quantifying "*the degree of class differentiation, excluding purely political and religious statuses.*" A zero score indicates "*absence of significant class distinctions among freemen, ignoring variations in individual repute achieved through skill, valor, piety, or wisdom.*" A score of 1 indicates "*the presence of wealth distinctions, based on possession or distribution of property, which however have not crystallized into distinct and hereditary social classes.*" A score of 2 indicates "*elite stratification in which an elite class derives its superior status from control over scarce resources, particularly land, and is thereby differentiated from a property-less proletariat or serf class.*" A score of 3 indicates a "*dual stratification into a hereditary aristocracy and a lower class of ordinary commoners or freemen, where traditionally ascribed noble status is at least as decisive as control over scarce resources.*" A score of 4 indicates "*complex stratification into social classes correlated in large measure with extensive differentiation of occupational statuses.*" *Source: Murdock (1967); variable code in the Ethnographic Atlas v67.*

Class Stratification Indicator: Following Gennaioli and Rainer (2007), we define a dummy stratification index that equals zero when Murdock's variable equals zero indicating "*absence of significant class distinctions among freemen, ignoring variations in individual repute achieved through skill, valor, piety, or wisdom,*" and 1 when Murdock's class stratification measure equals 1, 2, 3, or 4. *Source: Murdock (1967); variable code in the Ethnographic Atlas v67.*

Elections: Indicator that equals 1 when succession to the office of the local headman is conducted via "*election or other formal consensus, nonhereditary*" and zero otherwise. *Source: Murdock (1967); variable code in the Ethnographic Atlas v72.*

Slavery: Indicator that equals 1 when some type of slavery (hereditary, incipient, or significant) is present and zero when it is absent or near absent. *Source: Murdock (1967); variable code in the Ethnographic Atlas v70; the construction of the index follows Fenske (2009)*.

Inheritance Rules for Property: Indicator that equals 1 when some form of inheritance rule of real property (land) is present; the binary indicator equals zero when there is "*absence of individual property rights.*" *Source: Murdock (1967); variable code in the Ethnographic Atlas v74*.

A.4.4. *Data for the Cross-Validation Analysis*

DHS Composite Wealth Index: The DHS wealth index is composed taking into account consumer durables, electricity, toilet facilities, source of drinking water, dwelling characteristics, and some country-specific attributes such as whether there is a domestic servant, for example. The measure is derived by the DHS using principal component analysis to assign indicator weights resulting in a composite standardized index for each country. *Source: DHS; available at http://www.measuredhs.com*.

REFERENCES

FENSKE, J. (2009): "Does Land Abundance Explain African Institutions," Report, Yale University. [16,18]

GENNAIOLI, N., AND I. RAINER (2006): "Precolonial Centralization and Institutional Quality in Africa," in *Institutions and Norms in Economic Development*, ed. by M. Gradstein and K. Konrad. Boston, MA: MIT Press. [15]

——— (2007): "The Modern Impact of Precolonial Centralization in Africa," *Journal of Economic Growth*, 12 (3), 185–234. [9,10,12,13,15,17]

KAUFMANN, D., A. KRAAY, AND M. MASTRUZZI (2005): "Governance Matters IV: Governance Indicators for 1996–2004," Policy Research Working Paper Series, The World Bank. [15]

KISZEWSKI, A., A. MELLINGER, A. SPIELMAN, P. MALANEY, S. E. SACHS, AND J. SACHS (2004): "A Global Index of the Stability of Malaria Transmission," *American Journal of Tropical Medicine and Hygiene*, 70 (5), 486–498. [14]

MICHALOPOULOS, S. (2012): "The Origins of Ethnolinguistic Diversity," *American Economic Review*, 102 (4), 1508–1539. [14]

MURDOCK, G. P. (1959): *Africa: Its Peoples and Their Culture History*. New York: McGraw-Hill Book Company. [11]

——— (1967): *Ethnographic Atlas*. Pittsburgh, PA: University of Pittsburgh Press. [7,9,10,12,13, 15-18]

NUNN, N. (2008): "The Long Term Effects of Africa's Slave Trades," *Quarterly Journal of Economics*, 123 (1), 139–176. [7,8]

Dept. of Economics, Brown University, Providence, RI 02912, U.S.A.; smichalo@brown.edu

and

Dept. of Economics, London Business School, London NW1 4SA, United Kingdom; eliaspapaioannou@london.edu.