Appendix for
"Endogenous Colonial Borders: Precolonial States and Geography in the Partition of Africa"

We organize the supplementary material into four distinct sections.

- Appendix A is the main appendix (24 pages).
  - Appendix A.1: Supporting information to establish the conventional wisdom about border formation in Africa and summarize our data on major border revisions.
  - Appendix A.2: Supporting information for the grid-cell regressions.
  - Appendix A.3: Supporting information for the ethnic-group regressions.
- Appendix B is the first supplemental appendix (20 pages). It provides extensive coding notes for our polygons of precolonial states.
- Appendix C is the second supplemental appendix (138 pages). It presents a case study for all 107 bilateral borders in Africa. It also provides details on our coding standards and lists every bilateral border that we code as directly affected by a historical political frontier.
- Appendix D is the third supplemental appendix (7 pages). It presents additional regression tables that accompany the coefficient plots in Appendix A.

A MAIN APPENDIX

A.1 SUPPORTING INFORMATION FOR HISTORICAL BACKGROUND SECTION

A.1.1 Conventional Wisdom on Arbitrary African Borders

- Encyclopedia of Africa (Appiah and Gates 2010): "Rivalry between Great Britain and France led Bismarck to intervene, and in late 1884 he called a meeting of European powers in Berlin. In the subsequent meetings, Great Britain, France, Germany, Portugal, and King Leopold II negotiated their claims to African territory, which were then formalized and mapped."
- Michalopoulos and Papaioannou (2016, 1803) consider the "Scramble for Africa as a 'quasi-natural' experiment." "During the 'Scramble for Africa,' that starts with the Berlin Conference of 1884–1885 and is completed by the turn of the twentieth century, Europeans partitioned Africa into spheres of influence, protectorates, and colonies. The borders were designed in European capitals at a time when Europeans had barely settled in Africa and had limited knowledge of local conditions. Despite their arbitrariness, boundaries outlived the colonial era" (p. 1802). On the basis of their statistical analysis of ethnic groups, they conclude, "[w]ith the exceptions of the land mass of the historical ethnic homeland and the presence of lakes, there are no significant differences between split and non-split homelands along a comprehensive set of covariates ... These results offer support to a long-standing assertion within the African historiography regarding the largely arbitrary nature of African borders, at least with respect to ethnic partitioning" (p. 1803).
- Christensen and Laitin (2019): "The infamous Berlin Conference of 1884–85 set administrative boundaries in Africa and granted vast territories to the leading European powers ... Berlin set the colonial boundaries and determined, in large stretches, the borders of contemporary African states" (p. 167–68, 174). They also cite Michalopoulos and Papaioannou's evidence as establishing "the arbitrariness—statisticians would say as-if randomness—with which borders were drawn in Berlin ... " (p. 173).
- Herbst (1989) and Herbst (2000, Ch. 3): "[t]he overwhelming importance of imperial military and geopolitical interests in the scramble for Africa meant that the Europeans necessarily ignored factors that are generally considered relevant to the partitioning of land." He also supports the view that "[t]he arbitrary division of the continent by the European powers [exhibited] little or no respect for preexisting social and political groupings, or even, sometimes, for 'natural' geographical features" (Herbst 1989, 675).
- Scholars commonly cite an estimate by Barbour (1961, 305) that 44% of African borders are parallel/meridian lines, 30% are mathematical (i.e., non-astronomical) lines, and 26% are geographical features (Herbst 2000, 75; Englebert 2002, 88; Abraham 2007). Similarly, Alesina, Easterly and Matuszeski (2011, 246, 251) assert, "[e]ighty percent of African borders follow latitudinal and longitudinal lines ... Africa is the region most notorious for arbitrary borders"; and Yakemtchouk (1971) claims, "Some eight-tenths of African borders are unrelated to traditional and ethnic boundaries" (p. 70).
- Englebert (2002, 84–88): "With borders inherited from the colonial scramble for Africa ... they usually lack geographical congruence with the institutions of the precolonial era." In

the Democratic Republic of the Congo, he mentions that "several precolonial kingdoms and states ... [were] partitioned with neighboring colonies ... These are not exceptional cases ... Colonial partition seemed to be the norm rather than the exception. In many cases, the existence of an integrated precolonial system did not prevent partition by colonials."

- Abraham (2007): "A 'tea and macaroon' approach to boundary delimitation during the process of colonisation—culminating in the Berlin Conference of 1884-1885—rendered [territorial disputes] inevitable" (p. 62).
- Examples from popular press: "In 1885 European leaders met at the infamous Berlin Conference to divide Africa and arbitrarily draw up borders that exist to this day."[1] "The Partition of Africa began in earnest with the Berlin Conference of 1884-1885, and was the cause of most of Africa's borders today."[2] "The Berlin Conference spanned almost four months of deliberations, from 15 November 1884 to 26 February 1885. By the end of the Conference the European powers had neatly divided Africa up amongst themselves, drawing the boundaries of Africa much as we know them today."[3] "At the Berlin Conference, the European colonial powers scrambled to gain control over the interior of the continent. The conference lasted until February 26, 1885 — a three-month period where colonial powers haggled over geometric boundaries in the interior of the continent, disregarding the cultural and linguistic boundaries already established by the indigenous African population."[4]

A.1.2 Did the Berlin Conference Matter in Other Ways?

Our evidence on the timing of border formation rules out conventional claims (see Appendix A.1.1) that the Berlin Conference played an important role in determining specific borders. But the Berlin Conference may have affected later borders by affecting macro-level claims or by determining rules for claiming territory. Neither appears well supported, which further raises the need for a new model of African border formation.

The Berlin Conference undoubtedly influenced some macro-level claims. For example, to bolster their positions at the Conference, Britain accelerated its process of gaining treaties along the Niger river and Germany claimed territories in various parts of the continent in 1884 (Craven 2015, 40–41). However, many macro-level claims in place as of the mid-1880s cannot be attributed to Berlin, such as Britain's in southern Africa and France's in Algeria. More important, micro-level borders are not mere derivatives of macro-level claims, and most later borders did not exhibit an obvious path dependence with earlier ones.[5] Europeans followed a rough notion of a hinterland doctrine: a power with claims to the coast had a right to its hinterland.[6] However, this principle was too

[1] Fischer 2015.

[2] Faal 2009.

[3] South African History Online 2019.

[4] Rosenberg 2019.

[5] An exception was West Africa, where many later borders extended initially short rays that emanated from the coast. However, even in these cases, most initial borders were later revised to replace straight lines with water bodies and roads.

[6] In intra-French partitions, "[t]he French postulate that the inland regions of Sudan have different outlets depending on their proximity to the coast. Each of the four French colonies bordering the Atlantic is therefore assigned the hinterland for which it is the logical outlet" (Sandouno 2015, 20–21).

imprecise to determine even rough spheres of influence at the meso level, let alone specific borders at the micro level (Wesseling 1996, 127). For example, in the late 1880s in East Africa, Britain and Germany agreed not to annex territory located in the "rear" of the other's coastal territory (Hertslet 1909, 888–89), but the vagueness of the idea "left considerable room for misunderstanding in the future" and was explicitly rejected by statesmen such as Prime Minister Lord Salisbury (Louis 1963, 9–10). The ensuing Anglo–German Agreement of 1890 yielded, among other concessions, British control over Uganda. This agreement reflected Germany's desire to gain the small island of Heligoland in the North Sea, as opposed to an inevitable extension of its coastal possessions (some of which, such as Witu, it relinquished). As another example, in disputes in 1898 along what became the Benin–Nigeria border, French diplomats argued that "[t]he hinterland theory was useless, for the coast was curved and every hinterland point must lie in several hinterlands" (Flint 1960, 287). This argument was convenient for their bargaining posture, as a hinterland-doctrine interpretation of territorial control would cost them Bussa, located at the terminus of the navigable part of the Niger.

In its concluding General Act, the Conference decreed rules of effective occupation for claiming territory (Hertslet 1909, 484–85). Such rules, even if successful, would not predict micro-level border features. Nonetheless, the formal rules appear to have simply acknowledged that Europeans were claiming territory without discernibly altering this behavior. The standards for effective occupation were vague, a product of British resistance to this principle, and applied only to coastal settlements—many of which were already occupied (Crowe 1942, 190–91; Wesseling 1996, 124–30). In practice, as we highlight, "effective occupation" came to mean treaties with local rulers. This created scope for African participation, despite their lack of representation at Berlin.

Despite minimal impact on specific borders, the Berlin Conference may have influenced the eventual annexation of African territory. Whereas treaties with local rulers created protectorates (i.e., they granted Europeans control over external but not internal affairs), Europeans later ignored these limitations and imposed local governance institutions. Alexandrowicz (1973, 148) interprets the stipulations of the Berlin Conference as an agreement among the powers to permit such rights of annexation (see also Craven 2015, 42–49).

A.1.3 Major Revisions to Colonial Borders

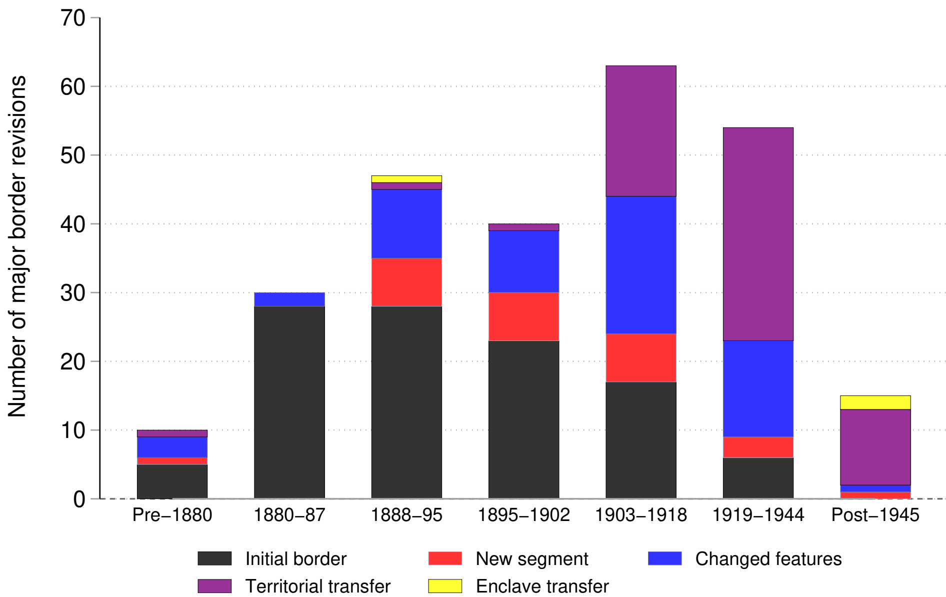
In the article, we presented two figures based on our original data that, for all 107 bilateral borders, codes the initial year of border formation and all subsequent years with a major revision (including the type of revision). We describe these data in the article.[7] The following tables and figures supplement Figure 2. In Figure A.1, we plot the frequency of different types of revisions over time. In Table A.1, we provide more details on the subcategories for types of major revisions. In Table A.2, we list every large territorial transfer since 1900. Unlike the preceding figure and

[7]We do not code a major revision if a colony was temporarily merged with others but the bilateral borders were unchanged (e.g., the federation of Southern Rhodesia, Northern Rhodesia, and Nyasaland in the Central African Federation from 1953–63), or if a merger occurred entirely within the boundaries of the final colony (e.g., amalgamation of Northern and Southern Nigeria in 1914). We code major revisions for the relevant bilateral borders when Upper Volta was merged (1932) and re-created (1947) because its territory was divided among three other French colonies, which altered a series of bilateral borders.

table, the last table is organized by distinct instances of territorial transfers rather than by bilateral borders, and some of the large territorial transfers affected multiple bilateral borders.

To produce Figure 3 we digitized colonial maps of 1887, 1895, and 1902 from Sanderson (1985). We combined these maps with our detailed notes on each bilateral border to code which border segments in each year corresponded with the final colonial borders. This process allows us to correct inevitable inaccuracies in historical maps. We then calculated two sets of figures for each map to quantify how colonial claims and borders evolved over time. First, we computed (using polygons of claimed territories) the percentage of all African territory claimed by Europeans, disaggregated by coastal and interior (300 km from the coast). Second, we computed (using polylines of borders) the total length of borders in their final form as a percentage of the total length of borders in 1960.

Figure A.1: Major Border Revisions Over Time



Notes: The unit of analysis for the underlying data is bilateral border-year. The dependent variable equals 1 in any year in which the specified type of major border revision occurred, and 0 otherwise. For each time period (x-axis), the bar segments represent the total number of revisions of each type that occurred, and the sum of these segments is the total number of major border revisions (y-axis).

Table A.1: Types of Major Border Revisions

Type	Total	20th century
Territorial transfer*		
Large	47	45
Small	17	17
Enclave	3	2
Changed features		
Switch lines to local features	22	14
Clarify local features	28	19
Change local features	5	2
Change location of straight line	4	3
New segment	26	13

Notes: The unit of analysis for the underlying data is bilateral border-year. The dependent variable equals 1 in any year in which the specified type of major border revision occurred, and 0 otherwise. Each cell sums up the number of 1's for the specified category and time period.

* This is the number of *bilateral borders* that experienced a major revision due to a territorial transfer, not the number of *unique territorial transfers*; a single transfer can affect multiple bilateral borders. The 45 bilateral borders affected by a large territorial transfer since 1900 (listed in the present table) correspond with the 21 distinct instances in which a large territory was transferred in the twentieth century (listed in Table A.2).

Table A.2: Large Territorial Transfers Since 1900

Gaining state	Losing state	Year	Territory	Approx. sq.km.
Kenya	Uganda	1902	Eastern Province	84,000
Zambia	Angola	1905	Lozi territory	88,000
Ethiopia	Kenya	1907	Menelik's claims	225,000
Sudan	DRC	1910	Lado Enclave	39,000
Cameroon	AEF	1911	Neukamerun	295,000
Western Sahara	Morocco	1912	Cape Juby	33,000
Uganda	Sudan	1914	Part of Lado Enclave*	47,000
Ghana	Togo	1919	British Togoland	34,000
Nigeria	Cameroon	1919	British Cameroons	53,000
AEF	Cameroon	1919	Neukamerun	295,000
Rwanda	Tanzania	1924	Gisaka district	7,000
Somalia	Kenya	1925	Jubaland	110,000
Kenya	Uganda	1926	Rudolf Province	37,000
Chad	Niger	1931	Tibesti mountains	134,000
AOF**	Burkina Faso	1932	Upper Volta	274,000
Libya	Sudan	1934	Sarra Triangle	72,000
Somalia	Ethiopia	1936	Ogaden	327,000
Burkina Faso	AOF**	1947	Upper Volta	274,000
Ethiopia	Somalia	1954	Ogaden	327,000
Morocco	Western Sahara	1958	Cape Juby	33,000
Cameroon	Nigeria	1961	Southern Cameroons	43,000

* Other parts of northern Uganda were transferred to Sudan (see Sudan–Uganda).

** Upper Volta was split among three AOF colonies in 1932 (Niger, Soudan/Mali, and Cote d'Ivoire) and reconstituted in 1947 (see Burkina Faso–Ivory Coast and Burkina Faso–Mali).

A.1.4 Comparison to Goemans and Schultz Data

Our coding exercise for the years of border formation is similar to that in Goemans and Schultz (2017), although the two datasets provide somewhat different findings. In their article, they examine "the age of each border segment by determining the year in which the line inherited upon independence was created. Since borders were often created over a period of years and successively defined in greater detail, we do not use a precise age; instead, we employ a broad distinction between old and new borders, using 1919 as a dividing line between borders that were created during the Scramble for Africa or the World War I settlement, and those that were not created until later." However, they graciously shared their replication data, which in fact contains more detailed information on the year in which a border was formed and the last year in which a substantial revision occurred. Their primary source for coding this variable is Brownlie (1979), which is also one of our primary sources.

The main difference in findings is that we tend to code earlier dates than Goemans and Schultz. Their median year for an initial border is 1899, compared to our median year of 1891. Of the 101 bilateral borders,[8] we code the same initial year for 40 borders, an earlier year for 57, and a later year for 4. Many of the differences are small, but some are not; in 31, the difference is more than a decade (in all but one, ours is earlier). Exhibiting similar discrepancies, their median year for a final border is 1912, compared to our median year of 1908. We code the same final year for 39 borders, an earlier year for 43, and a later year for 19. Of these, the discrepancy is more than a decade for 33 (23 earlier, 10 later).

Various factors account for these discrepancies. First, for initial borders, we require only that Europeans agreed upon an outline for a border, even if imprecise. We always count the first inter-European treaty as constituting an initial border, whereas Goemans and Schultz do not. Second, we appear to have a higher threshold for what constitutes a major revision. They count some cases in which the revisions were fairly minor, such as clarifying technical details related to alignment or altering minor features (e.g., which country controls certain islands located within a river that constitutes the border). Third, we consulted a wider range of sources. For intra-French borders in particular, we consulted additional (mainly French language) sources because Brownlie lacks detailed information for many of these borders. These additional sources provide evidence of earlier agreements to establish certain borders and present more precise information about subsequent revisions.

Although the third difference is a clear advantage of our coding effort, neither choice for the first two differences is unambiguously better. Nonetheless, given our overarching claim that Africa's borders tended to be formed later than scholars and commentators commonly proclaim, a coding procedure that biases toward earlier dates creates a harder test for our argument. Another advantage of our coding procedure is that we code all instances of major revisions and their type, accompanied by extensive coding notes that enable researchers to scrutinize and verify all our coding decisions.

[8]Following Brownlie, they do not distinguish Eritrea from Ethiopia nor British Somaliland from Somalia, yielding six fewer bilateral borders than in our dataset.

A.2 Supporting Information for Grid-Cell Regressions

Section A.2.1 presents data sources for the variables we analyze. Section A.2.2 presents the corresponding regression tables for Figure 6 and related robustness checks. Section A.2.3 changes the reference category in the PCS regressions by including both PCS variables (both PCS border cells and PCS interior cells) in the same model. Section A.2.4 discusses issues related to spatial dependence, assesses robustness checks for Conley SEs, and performs an alternative procedure for calculating SEs using the wild bootstrap. Section A.2.5 computes Oster bounds to assess the sensitivity of our estimates to unobservables using information from observables.

Table of Contents

Appendix A.2.1: Data Sources for Variables

Figure A.2: Contested Coastal Areas (Trading Posts and Natural Harbors)

Appendix A.2.2: Regression Tables for Figure 6 and Robustness Checks

Table A.3: Correlates of Precolonial States
Table A.4: Regression Table for Figure 6: Geography (Top Panel)
Table A.5: Regression Table for Figure 6: Precolonial States (Bottom Panel)
Figure A.3: Precolonial States and African Borders without Controls
Figure A.4: Correlates of African Borders with PCS FE

Appendix A.2.3: Changing the PCS Reference Category

Figure A.5: Correlates of African Borders with both PCS Variables
Figure A.6: Correlates of African Borders with PCS Variables and Ethnic Borders

Appendix A.2.4: Spatial Dependence and Conley Standard Errors

Figure A.7: Correlates of African Borders with Various Distance Cutoffs
Figure A.8: Confidence Curves Using Wild Cluster Bootstrapped Standard Errors

Appendix A.2.5: Oster Bounds

Table A.6: Assessing Possible Bias from Unobservables

A.2.1 Data Sources for Variables

1. **Top 10 River**: Equals 1 for grid cells with any of the 10 longest rivers in Africa; 0 otherwise. The top 10 rivers in our data are the Nile, Congo, Niger, Zambezi, Ubangi, Kasai, Orange, Limpopo, Senegal, and Blue Nile; although note that different sources offer somewhat different lists because there is no standard procedure for which tributaries count. *Source: "Rivers and lake centerlines" shapefile from Natural Earth (2023).*

2. **Any River**: Equals 1 for grid cells with a river; 0 otherwise. *Source: "Rivers and lake centerlines" shapefile from Natural Earth (2023).*

3. **Minor River**: Equals 1 for grid cells with a river but not a top 10 river; 0 otherwise. *Source: "Rivers and lake centerlines" shapefile from Natural Earth (2023).*

4. **Major Watershed**: Equals 1 for grid cells that contain a major watershed divide; 0 otherwise.[9] The major watersheds in our data are Chad (Lake), Congo, Etosha, Gambia, Kunene, Limpopo, Niger, Nile, Okavango, Orange, Rift Valley, Ruvuma, Senegal, Shebelli - Juba, Volta, and Zambezi. *Source: Constructed by authors using FAO maps of Hydrological basins in Africa from Food and Agricultural Organization of the United Nations (2022), which we cross-referenced with maps from Vivid Maps (2001).*

5. **Top 10 Lake**: Equals 1 for grid cells with any of the 10 largest lakes in Africa; 0 otherwise. Top 10 lakes: Lake Victoria, Tanganyika, Malawi, Chad, Turkana, Albert, Mweru, Tana, Kivu, and Edward. *Source: "Rivers and lake centerlines" shapefile from Natural Earth (2023).*

6. **Any Lake**: Equals 1 for grid cells with a lake; 0 otherwise. *Source: "Rivers and lake centerlines" shapefile from Natural Earth (2023).*

7. **Minor Lake**: Equals 1 for grid cells with a lake but not a top 10 lake; 0 otherwise. *Source: "Rivers and lake centerlines" shapefile from Natural Earth (2023).*

8. **Cell in Desert**: Equals 1 for grid cells that contain any non-vegetated or sparsely vegetated area. *Source: UNESCO Vegetation Map of Africa by White (1983).*

9. **Area of Ethnic Group**: Measured for Murdock ethnic homelands; equals the logged surface area (in 1000s of km^2) of the ethnic homeland located within a grid cell. If multiple groups are located within a cell, we compute the average weighted by the land area of each group in that cell. *Source: Michalopoulos and Papaioannou (2016). Original Source: Global Mapping International, Colorado Springs, Colorado, USA.*

10. **Distance to the Coast**: The shortest geodesic distance of the centroid of each grid cell from the coast, measured in 1000s of km. Calculated by authors in ArcGIS.

11. **Suitability for European Settlement**: Equals 1 for grid cells that contain any amount of an area that was suitable for large-scale European agricultural settlements; 0 otherwise. Suitable areas contain either temperate climate (northern and southern tips of continent), or each of

[9] Watersheds are land ridges that separate water flowing into different rivers. They are sometimes called watershed boundaries or drainage divides.

high rainfall, high elevation, and low tsetse fly prevalence (various parts of the interior of the continent). *Source: Paine (2019).*

12. **Agricultural Intensity**: Measured for Murdock ethnic homelands. Index for the intensity of agriculture practiced by members of the ethnic group; 1: "complete absence of agriculture"; 2: "casual agriculture"; 3: "extensive or shifting cultivation"; 4: "horticulture"; 5: "intensive agriculture on permanent fields"; 6: "intensive cultivation where it is largely dependent upon irrigation." For each grid cell, we use the value for the ethnic group containing the cell. If multiple groups are located within a cell, we compute the average weighted by the land area of each group in that cell. *Source: Murdock (1967); variable* `v28`.

13. **Population Density in 1850**: Population density in 1850 within each grid cell, measured in 1000s/km^2. *Source: Utrecht University (2022).*

14. **Ecological Diversity**: Measured for Murdock ethnic homelands; ecological diversity index ranging between 0 and 1. For each grid cell, we use the value for the ethnic group containing the cell. If multiple groups are located within a cell, we compute the average weighted by the land area of each group in that cell. To fill in missing data points, we compute the index for major lakes not included in Fenske, following his method using White's vegetation data. *Sources: Fenske (2014); White (1983).*

15. **TseTse Suitability Index (TSI)**: The standardized Z-score of the potential steady-state TseTse population, which takes into account temperature and humidity requirements for TseTse viability. The underlying spatial data are a collection of points. We compute the average TSI for the points in each grid cell. Some coastal cells do not contain any points, and we take the value of the nearest point for those cells. *Source: Alsan (2015).*

16. **Contested Coastal Areas**: Equals 1 for grid cells containing any part of a contested coastal area; 0 otherwise. To identify contested coastal areas, we (1) coded colonial claims over natural harbors and precolonial trading posts as of 1887, (2) identified "competing ports," located wherever two neighboring harbors/posts were claimed by different powers, and (3) created rectangular polygons by extending inland at 90°(until reaching 300km from the coast) the line connecting each pair of competing ports. Figure A.2 depicts this variable. *Source: Natural harbors and precolonial trading posts from Ricart-Huguet (2022), which we extended to the entire continent.*

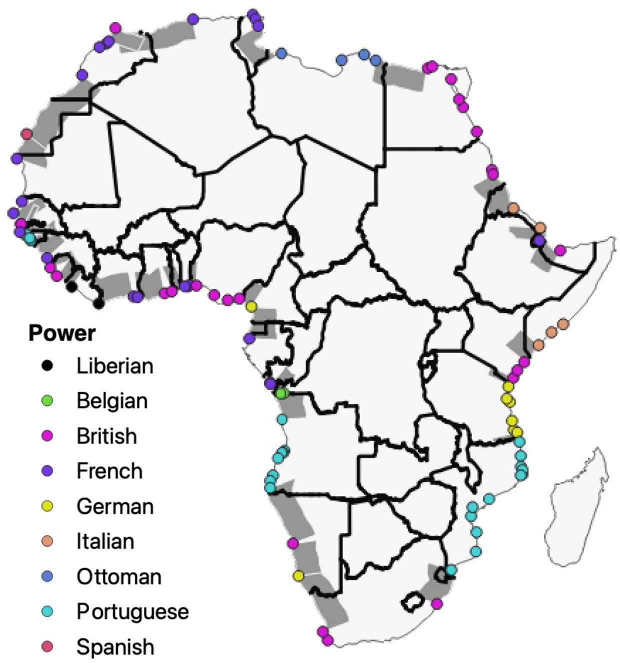
17. **Slave Exports**: Measured for Murdock ethnic homelands; the logged number of slave exports scaled by land area of the ethnic group (log(1+ exports/km^2)). For each grid cell, we use the value for the ethnic group containing the cell. If multiple groups are located within a cell, we compute the average weighted by the land area of each group in that cell. *Source: Nunn (2008).*

18. **Historical Natural Resources**: Equals 1 for grid cells that contain a historical natural resource site; 0 otherwise. *Source: Ricart-Huguet (2022), which we extend to the whole continent.*

19. **Regions**: We constructed five regions based on latitudes and longitudes. North: cells north of 18° N, roughly all areas including and north of the Sahara desert (excludes Sahel); South: cells south of 15° S, roughly all areas south of Lake Malawi; West: cells between 18° N

and 14.5° S and west of 14° E, roughly all areas west of Lake Chad and south of North Africa; East: cells between 18° N and 15° S and east of 14° E, roughly all areas east of Lake Tanganyika that are neither North Africa nor Southern Africa; Central: all remaining cells.

Figure A.2: Contested Coastal Areas (Trading Posts and Natural Harbors)



Note: The colored dots represent precolonial trading posts and natural harbors, and the black lines represent country borders. The gray polygons are the contested coastal areas, as described above. Appendix Table A.5 shows that cells in contested coastal areas are more likely to contain colonial borders.

A.2.2 Regression Tables for Figure 6 and Robustness Checks

Table A.3: Correlates of Precolonial States

	DV: PCS FRONTIER				DV: PCS INTERIOR			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Major (top 10) river	0.04*	0.06**	0.03+	0.05**	0.04*	-0.02	0.02	-0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
Minor river	0.02**	0.03**	0.03**	0.04**	-0.01	-0.02**	0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Major watershed	0.03**	0.03**	0.02*	0.02+	0.04**	0.01	0.04**	0.02*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Major (top 10) lake	0.12**	0.13**	0.17**	0.18**	-0.07**	-0.07**	0.01	0.02
	(0.03)	(0.03)	(0.04)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)
Minor lake	0.15**	0.13**	0.13**	0.12**	-0.05**	-0.07**	-0.05**	-0.04**
	(0.02)	(0.02)	(0.02)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)
Cell in desert	-0.07**	-0.05**	-0.06**	-0.05**	-0.03**	-0.03**	-0.01	-0.05**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Latitude			0.00**	0.00**			0.01**	0.01**
			(0.00)	(0.00)			(0.00)	(0.00)
Longitude			0.00**	0.01**			0.01**	0.01**
			(0.00)	(0.00)			(0.00)	(0.00)
Contested coastal areas			0.06**	0.08**			-0.00	-0.01
			(0.01)	(0.01)			(0.01)	(0.01)
Area of ethnic group			-0.03**	-0.03**			-0.03**	-0.04**
			(0.00)	(0.00)			(0.00)	(0.00)
Distance to the coast			0.01	0.02+			0.01	0.02*
			(0.01)	(0.01)			(0.01)	(0.01)
Historical natural resources			0.01	0.02			0.03+	0.02
			(0.02)	(0.02)			(0.02)	(0.02)
Slave exports			0.00+	0.01**			0.01**	0.01**
			(0.00)	(0.00)			(0.00)	(0.00)
Suitability for			0.01	-0.01			-0.04**	-0.07**
European settlement			(0.01)	(0.01)			(0.01)	(0.01)
Agricultural intensity			-0.00	0.01*			-0.01*	-0.01*
			(0.00)	(0.00)			(0.00)	(0.00)
Population density in 1850			1.36**	1.90*			0.57	0.47
			(0.41)	(0.74)			(0.40)	(0.46)
Ecological diversity			0.02	0.02			0.03+	0.05**
			(0.02)	(0.02)			(0.01)	(0.02)
Tsetse suitability index			-0.02**	-0.03**			-0.03**	-0.05**
			(0.00)	(0.01)			(0.00)	(0.01)
Constant	0.10**	0.10**	0.11**	0.04	0.10**	0.11**	0.10**	0.04+
	(0.00)	(0.00)	(0.02)	(0.03)	(0.00)	(0.00)	(0.02)	(0.02)
Grid cells	10341	7228	9913	6816	10341	7228	9913	6816
Adjusted R^2	0.03	0.02	0.10	0.08	0.01	0.00	0.13	0.14
Sample	Full	SSA	Full	SSA	Full	SSA	Full	SSA
Region FE			✓	✓			✓	✓

Notes: The table reports regression results for correlates of precolonial state formation in Africa. The dependent variables are PCS FRONTIER (Columns 1–4) and PCS INTERIOR (Columns 5–8). "Full" models include all grid cells; Sub-Saharan ("SSA") models are limited to grid cells south of the 18°N parallel. All models are estimated using OLS with robust standard errors in parentheses; we report less conservative standard error estimates than in our main tables to avoid Type II errors, given the goal of identifying correlates of PCS borders that might also affect colonial border formation. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table A.4: Regression Table for Figure 6: Geography (Top Panel)

	DV: COUNTRY BORDER IN CELL							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Any river	0.11**							
	(0.03)							
Major (top 10) river		0.19**					0.18**	0.19**
		(0.06)					(0.06)	(0.06)
Minor river		0.09**					0.08**	0.07**
		(0.03)					(0.03)	(0.03)
Any lake			0.13**					
			(0.04)					
Major (top 10) lake				0.35**			0.34**	0.33**
				(0.09)			(0.09)	(0.09)
Minor lake				0.03			0.01	0.01
				(0.03)			(0.03)	(0.03)
Major watershed					0.10**		0.10**	0.10**
					(0.03)		(0.03)	(0.02)
Cell in desert						-0.07**	-0.04+	-0.03
						(0.02)	(0.02)	(0.02)
Distance to the coast								0.00
								(0.02)
Suitability for European settlement								0.05+
								(0.03)
Tsetse suitability index								0.04**
								(0.01)
Constant	0.13**	0.13**	0.14**	0.14**	0.13**	0.18**	0.12**	0.13**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.04)
Grid cells	10341	10341	10341	10341	10341	10341	10341	10341
Adjusted R^2	0.02	0.02	0.01	0.01	0.01	0.01	0.04	0.06
Lat & Lon								✓

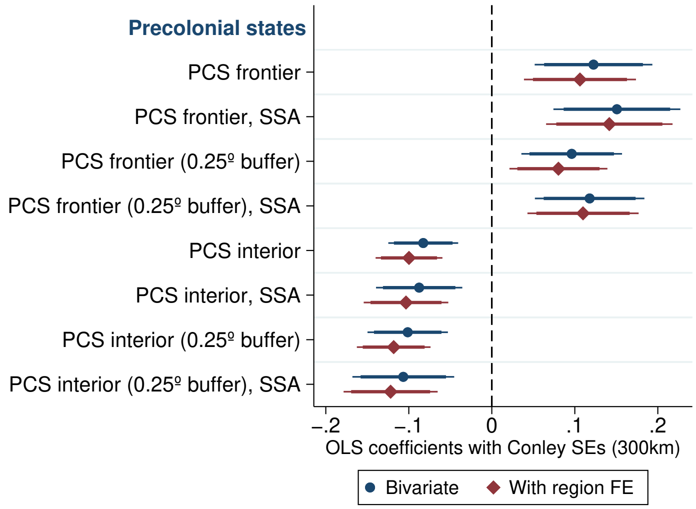
Notes: This regression table accompanies the top panel of Figure 6. All models are estimated using OLS with Conley standard errors in parentheses (distance cutoff $= 300$ km). $^+ \, p < 0.10$, $^* \, p < 0.05$, $^{**} \, p < 0.01$

Table A.5: Regression Table for Figure 6: Precolonial States (Bottom Panel)

	DV: COUNTRY BORDER IN CELL							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PCS frontier	0.10**				0.12**			
	(0.03)				(0.04)			
PCS frontier (0.25° buffer)		0.07*				0.09**		
		(0.03)				(0.03)		
PCS interior			-0.09**				-0.09**	
			(0.02)				(0.03)	
PCS interior (0.25° buffer)				-0.11**				-0.11**
				(0.02)				(0.03)
Major (top 10) river	0.20**	0.20**	0.20**	0.20**	0.22**	0.22**	0.22**	0.22**
	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)
Minor river	0.07*	0.07*	0.07**	0.07**	0.07**	0.07**	0.08**	0.08**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Major (top 10) lake	0.29**	0.28**	0.30**	0.30**	0.29**	0.28**	0.31**	0.31**
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
Minor lake	0.00	0.01	0.01	0.01	-0.00	0.00	0.01	0.01
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Major watershed	0.09**	0.09**	0.10**	0.10**	0.10**	0.10**	0.10**	0.10**
	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Cell in desert	-0.02	-0.02	-0.03	-0.03	-0.04	-0.04	-0.05+	-0.05+
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Contested coastal areas	0.09**	0.09**	0.09**	0.09**	0.08*	0.08*	0.08**	0.08**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Ethnic group border in cell	0.01	0.01	0.01	0.02+	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Area of ethnic group	-0.01	-0.01	-0.01	-0.01	-0.01	-0.00	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Distance to the coast	0.03	0.03	0.03	0.03	0.01	0.01	0.02	0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Historical natural resources	-0.05*	-0.06*	-0.05*	-0.05*	-0.06+	-0.06*	-0.05+	-0.06+
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Slave exports	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)
Suitability for European settlement	0.05	0.05	0.04	0.04	0.05+	0.06+	0.05	0.05
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Agricultural intensity	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Population density in 1850	-0.84*	-0.81*	-0.66*	-0.66*	-1.26*	-1.23+	-0.98	-0.99
	(0.34)	(0.34)	(0.31)	(0.31)	(0.63)	(0.65)	(0.65)	(0.65)
Ecological diversity	-0.07	-0.07	-0.06	-0.07	-0.10*	-0.10*	-0.10+	-0.10+
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Tsetse suitability index	0.04**	0.04**	0.04**	0.04**	0.04**	0.04**	0.03**	0.03**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	0.14*	0.13*	0.15*	0.15*	0.18*	0.18*	0.19**	0.19**
	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)
Grid cells	9913	9913	9913	9913	6816	6816	6816	6816
Adjusted R^2	0.07	0.07	0.07	0.07	0.09	0.08	0.08	0.08
Sample	Full	Full	Full	Full	SSA	SSA	SSA	SSA
Region FE	✓	✓	✓	✓	✓	✓	✓	✓
Lat & Lon	✓	✓	✓	✓	✓	✓	✓	✓

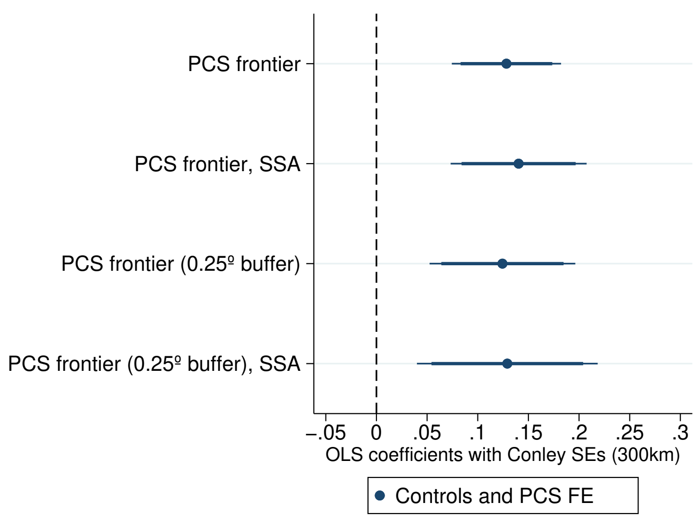
Notes: This regression table accompanies the bottom panel of Figure 6. All models are estimated using OLS with Conley standard errors in parentheses (distance cutoff = 300 km). "Full" models include all grid cells; Sub-Saharan ("SSA") models are limited to grid cells south of the 18°N parallel. This table contains fewer observations than Table A.4 because Agricultural intensity contains missing values. $^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Figure A.3: Precolonial States and African Borders without Controls



Notes: This figure presents a series of coefficient plots similar to those in the bottom part of Figure 6, but without the battery of control variables. See Table D.1 for the corresponding regression table.

Figure A.4: Correlates of African Borders with PCS FE



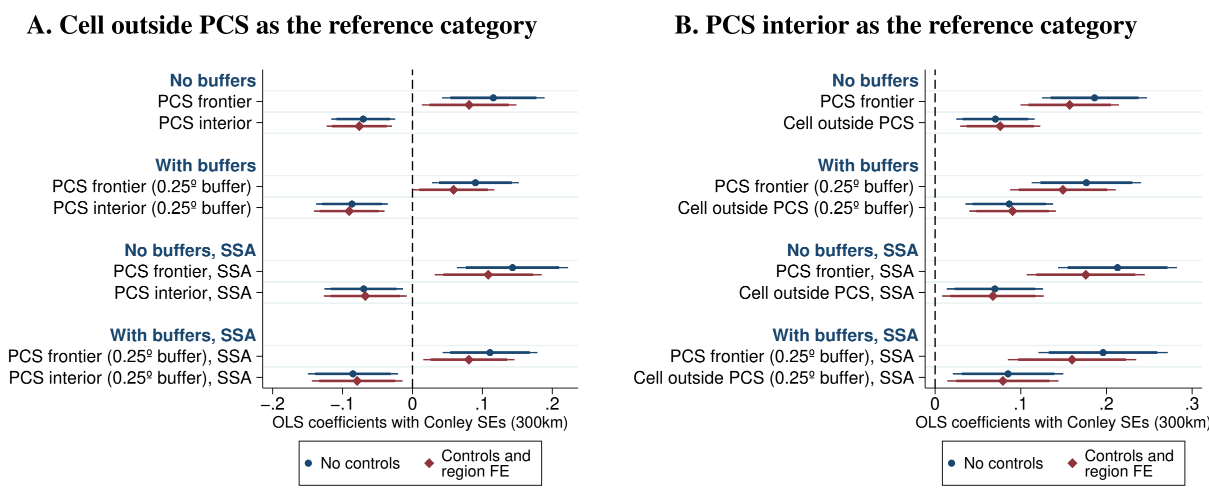
Notes: This figure presents coefficients from models that add fixed effects for PCS to the specification. Using PCS FE causes cells outside PCS to drop and essentially compares cells with PCS borders to cells inside PCS while stratifying on the PCS, which guards against any source of omitted variable bias common to each PCS. Table D.2 provides the accompanying regression table.

A.2.3 Changing the PCS reference category

Our theoretical expectations are that PCS frontier cells should more frequently contain a colonial border compared to any other cells, and that PCS interior cells should less frequently contain a colonial border compared to any other cells. In the article, we discuss specifications that include one PCS indicator at a time, therefore leaving as the reference category cells with either the other PCS category or without a PCS.

In Figure A.5, we present results with more homogeneous basis categories. In Panel A, the model includes both PCS FRONTIER and PCS INTERIOR as regressors, which leaves CELL OUTSIDE PCS as the reference category. In Panel B, the model includes both PCS FRONTIER and CELL OUTSIDE PCS as regressors, which leaves PCS INTERIOR as the reference category.[10] This robustness check demonstrates that the significance of the PCS coefficients in Figure 6 is not driven by a particular way of specifying the reference category. Panel A demonstrates that PCS frontier cells are significantly more likely to contain a colonial border, and PCS interior cells significantly less likely, compared to non-PCS cells. Panel B demonstrates that both PCS frontier and non-PCS cells are significantly more likely to contain a colonial border, compared to PCS interior cells.

Figure A.5: Correlates of African Borders with both PCS Variables



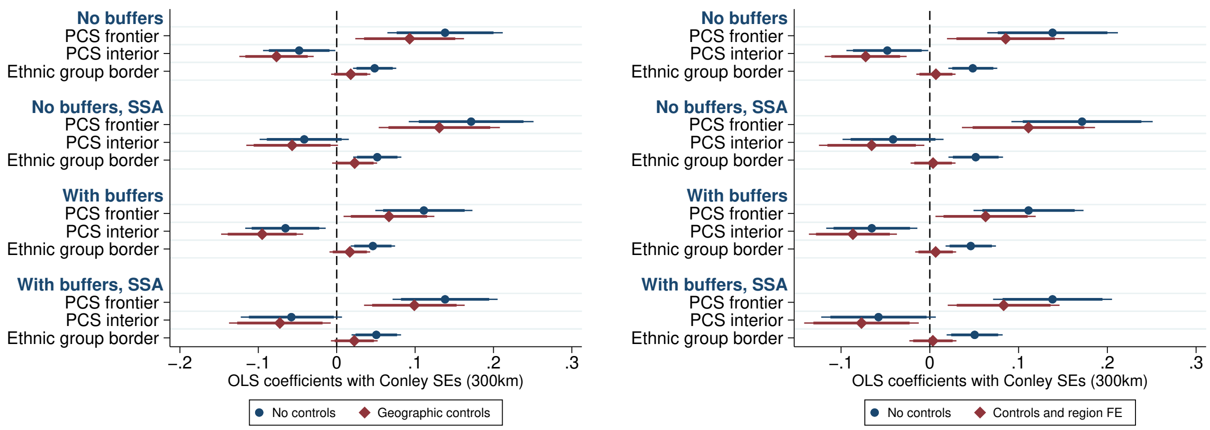
Notes: This figure presents a series of coefficient plots similar to Figure 6, but each model contains two PCS regressors. This yields a reference category of CELL OUTSIDE PCS in Panel A and PCS INTERIOR in Panel B. Tables D.3 and D.4 provide the accompanying regression tables.

We directly incorporate ethnic homeland boundaries in another set of robustness checks. We distinguish cells outside PCS into (a) those that contain the boundaries of a Murdock ethnic homeland (ETHNIC GROUP BORDER = 1), and (b) those that do not (PCS FRONTIER = PCS INTERIOR = ETHNIC GROUP BORDER = 0). The models in Figure A.6 control for all of PCS FRONTIER, PCS INTERIOR, and ETHNIC GROUP BORDER, which leaves as the reference category cells that contain neither a PCS nor a Murdock ethnic boundary. The left panel reports models with either (a) no controls or (b) geographical covariates only. The right panel reports models with either (a) no controls

[10] A third panel with PCS FRONTIER as the reference category would yield identical information as in these two plots.

or (b) the full set of covariates and region FE. Cells with ETHNIC GROUP BORDER $= 1$ are positively associated with international borders in the bivariate specifications, but this correlation is not robust to adding either geographic or all covariates. This suggests that the frontiers of decentralized groups did not systematically affect colonial borders, although we caution that this specification is difficult to interpret because of the incongruity between our polygons and Murdock's.

Figure A.6: Correlates of African Borders with PCS Variables and Ethnic Borders



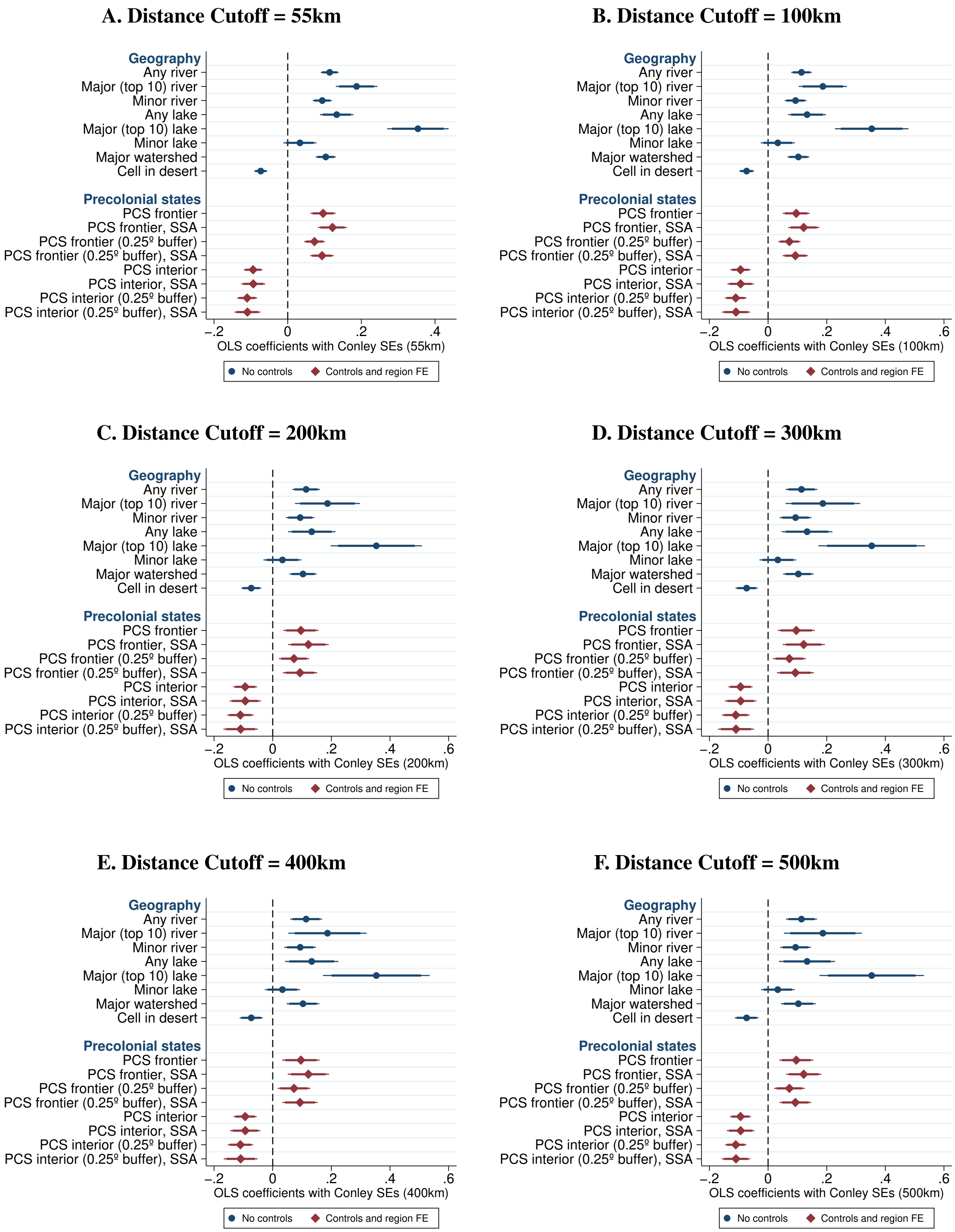
Notes: This figure presents a series of coefficient plots similar to Figure 6, but with three indicator variables: PCS FRONTIER, PCS INTERIOR, and ETHNIC GROUP BORDER. This leaves as the reference category cells that contain neither any portion of a PCS nor a Murdock ethnic boundary. Table D.5 provides the accompanying regression table.

A.2.4 Spatial Dependence and Conley Standard Errors

Spatially proximate units may be highly correlated in their unobservables; but, as the distance increases, the correlation gradually dissipates. In our analysis, any cell proximate to a cell containing a country border is itself highly likely to contain a country border. Conley standard errors, which we use in all specifications, account for such spatial dependence. This method adjusts the variance-covariance matrix by incorporating information about the spatial distance between observations. To compute the variance-covariance matrix, the method uses a uniform kernel function to weight pairs of observations such that the weight equals 1 if two observations are within a specified distance threshold, and 0 otherwise. The kernel function thus distinguishes between observations that are near and those that are far. The choice of the cutoff distance affects the standard error estimates; in Figure A.7, we verify that our results are robust to various distance thresholds.

One assumption inherent to calculating Conley SEs that cannot, formally, be relaxed is uniform spatial dependence. That is, the covariance measure depends on distance but not direction. This assumption might be violated because our outcome and main explanatory variables are lines. For example, if a country border is horizontal in a cell, then the cells north and south of it are less likely to contain a border than cells west and east of it. We address this possibility in two ways.

Figure A.7: Correlates of African Borders with Various Distance Cutoffs

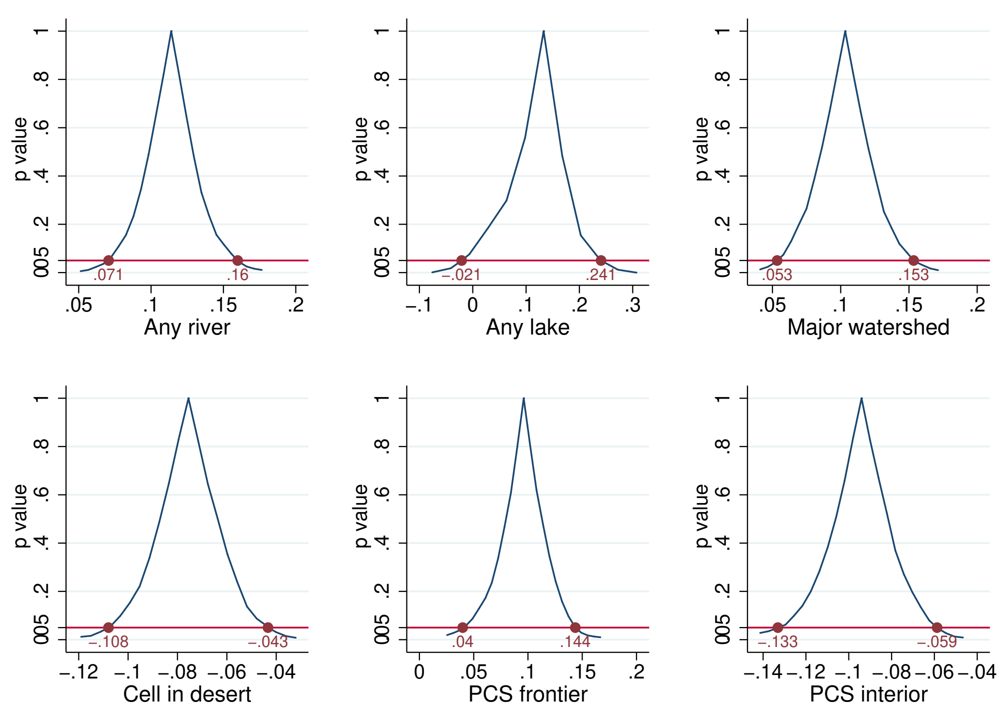


Notes: This figure presents a series of coefficient plots similar to Figure 6, but with varying distance cutoffs for the Conley standard errors. The coefficient estimates are unchanged from Tables A.4 and A.5.

First, if violations of uniform spatial dependence within a given radius have a significant impact, then standard error estimates should vary drastically upon varying the radius. However, as shown in Appendix Figure A.7, this is not the case. Our results are qualitatively unchanged for any distance cutoff ranging from 55km, capturing a single neighboring cell, up to 500km, which corresponds with large tracts of territory.

Second, as an alternative to Conley standard errors, we perform a robustness check in which we cluster our observations by artificially constructed rectangular regions, as neighboring cells may be related in many ways. Technically, the off-diagonals in the variance-covariance matrix are unlikely to be 0 for spatially proximate grid cells. We are conservative and create large clusters of roughly 10ºx10º (roughly 550 km at the equator). We compute standard errors using the wild bootstrap, a method designed "for regression models with heteroskedasticity of unknown form" with a small number of large clusters (Roodman et al. 2019, 1), which applies to our case. This allows us to relax the uniform spatial dependence assumption and account for flexible forms of spatial correlations among neighboring cells.

Figure A.8: Confidence Curves Using Wild Cluster Bootstrapped Standard Errors



Notes: This figure presents wild cluster bootstrapped standard errors at the 95% level for geographical and PCS variables of theoretical interest. The coefficient estimates are unchanged from Tables A.4 and A.5.

Figure A.8 presents the 95% confidence set for each of the main explanatory variables in Figure 6.[11] For example, our main model in Table A.5 estimates a coefficient of 0.10 for PCS FRONTIER. Figure A.8 shows that the 95% confidence set for PCS FRONTIER encompasses a range of 0.07

[11]The confidence set consists of the range of coefficient estimates for which the bootstrapped p-value for the test of the null is equal to or greater than 0.05.

to 0.16, and hence all values have the same sign. For five of the six variables, the coefficient estimates contained within the 95% confidence set have the same sign; this is true also for ANY RIVER, WATERSHED, CELL IN DESERT, and PCS INTERIOR. Only for ANY LAKE does the edge of the confidence contain negative values (for a positively estimated coefficient). Overall, our main results are robust to alternative ways of modeling spatial dependence.

A.2.5 Oster Bounds

We assess the likelihood that unobserved confounding variables account for the effect of precolonial states. Oster's (2019) test computes the share of variation that unobservables would need to explain, relative to the observables included in the model, in order to reduce the coefficient of interest to zero. This share is denoted by δ. For instance, $\delta = 2$ indicates that unobservables would need to be twice as important as observables for the coefficient to be zero (Oster 2019, 195).

The implementation of the Oster (2019) test requires specifying a value of R^2_{max}, which denotes the R^2 from a hypothetical regression that included both observed and unobserved controls. For example, $R^2_{max} = 1.5R^2$ means that including unobservables would increase the observed R^2 by 50%.

To bias against our results, and because our setting is observational, we use very large values of R^2_{max}: 1.5, 2, and 3 (Oster 2019 uses 1.3 in her article). That is, we assume that our R^2 could be up to three times as large due to unobserved confounders even though all models in Table A.6, just as in Figure 6, already include a battery of controls and region fixed effects.

Table A.6 shows that our main explanatory variables in Figure 6 (PCS frontier and PCS interior) are very robust to unobservables. We observe that $\delta > 1$ even when $R^2_{max} = 3R^2$.

We also calculate the bounds on the effect of each variable (β) on the likelihood of having a country border in that cell assuming $\delta = 1$ (that is, assuming that unobservables explain as much variation as observables). The range excludes 0 for all values of R^2_{max}. The two results convey the same idea: unobservables would need to be more than three times as important as observables for the effect of our main explanatory variables to become zero.

Table A.6: Assessing Possible Bias from Unobservables

PCS frontier	$R^2_{max} = 1.5R^2$ $= 0.11$	$R^2_{max} = 2R^2$ $= 0.15$	$R^2_{max} = 3R^2$ $= 0.23$
δ (unobservables/observables)	4.15	2.22	1.15
Bounds on β (for $\delta = 1$)	(0.10, 0.08)	(0.10, 0.06)	(0.10, 0.02)

PCS interior	$R^2_{max} = 1.5R^2$ $= 0.11$	$R^2_{max} = 2R^2$ $= 0.15$	$R^2_{max} = 3R^2$ $= 0.23$
δ (unobservables/observables)	49.31	26.16	13.49
Bounds on β (for $\delta = 1$)	(-0.09, -0.10)	(-0.09, -0.11)	(-0.09, -0.13)

Notes: The bounds are (β, β'), where β is the effect estimated from the main regression model and β' is the effect with $\delta = 1$ and the value of R^2_{max} specified in the column.

A.3 Supporting Information for Ethnic Partition Regressions

We conducted supplementary regressions using ethnic groups as the unit of analysis. We also discuss the important shortcomings of using the Murdock data for assessing the relationship between precolonial states and ethnic partition.

A.3.1 Data and Results

Data and models. We largely follow Michalopoulos and Papaioannou's (2016) setup for assessing the correlates of ethnic partition. They identify partitioned groups using Murdock's Ethnolinguistic Map (1959), digitized by Nunn (2008), which describes and geolocates ethnic groups in Africa at the time of European colonization. The sample consists of 825 ethnic homelands, after dropping uninhabited areas and small islands. Given inevitable error in the Murdock-drawn ethnic homeland boundaries, Michalopoulos and Papaioannou code as partitioned any group for which at least 10% of its territory falls into more than one country. We additionally coded, for each partitioned group, whether the border segment that split the group was primarily squiggly or a straight line (following the conceptual distinction in Alesina, Easterly and Matuszeski 2011).

Our measure of precolonial states is PCS MURDOCK. This variable equals 1 for any ethnic group that scores three levels of authority beyond the village or higher on Murdock's jurisdictional hierarchy variable, which he refers to as "states." Given our theoretical expectation that only the homelands of large and centralized groups created focal zones for Europeans to draw borders, a binary variable for precolonial states is easier to interpret than the original ordinal measure commonly used in the literature. However, we verified in unreported regressions that the results are qualitatively similar when using the same ordinal measure as Michalopoulos and Papaioannou (2016).

For geography, we measure whether each ethnic homeland contains ANY RIVER (the same measure used in Michalopoulos and Papaioannou 2016);[12] and, following the scheme used in our main analysis, we disaggregate rivers into TOP 10 RIVER and MINOR RIVER. Similarly, we measure whether an ethnic homeland contains ANY LAKE, TOP 10 LAKE, and MINOR LAKE; or MAJOR WATERSHED. Finally, we include the percentage of desert area for each ethnic homeland, SHARE OF DESERT.

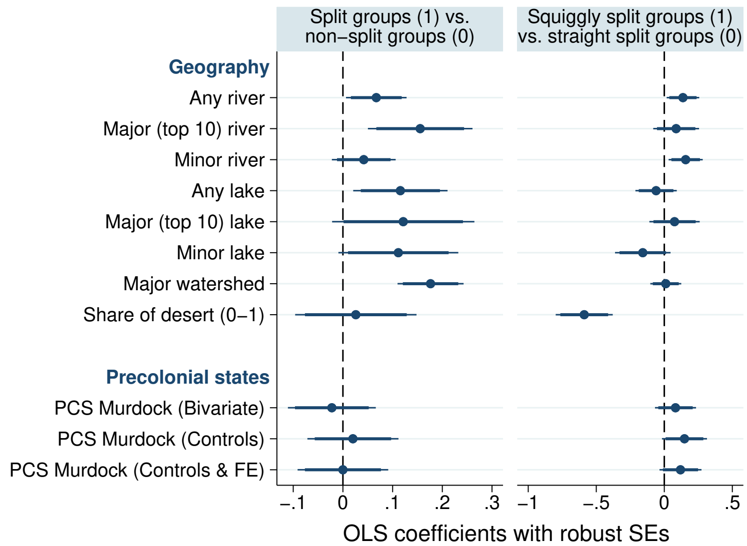
We use the same set of covariates as for our main grid-cell regressions. In Appendix A.2.1, we provide descriptions for how we measure the variables for each grid cell, and we perform the same procedure for every ethnic group. In some cases, the variable is already measured at the level of ethnic homelands, in which case we use that value verbatim.

In Figure A.9, we present coefficient plots for a series of models estimated using OLS, which examine geographical and precolonial-state correlates of ethnic group partition.[13] The left panel compares ethnic groups split across international borders with non-split groups. The right panel compares groups split by a squiggly border with those partitioned by a straight line. Across the entire sample, 229 of the 825 ethnic groups (28%) are partitioned across multiple countries. In 78% of the 229 split groups, a majority of the border is squiggly.

[12] Note that we use the same spatial data for water bodies as in Michalopoulos and Papaioannou (2016).

[13] The estimating equations are similar to Equations 1 and 2 presented in the article.

Figure A.9: Correlates of Ethnic Partition



Notes: This figure summarizes a series of OLS estimates with explanatory variables in rows and the dependent variables in columns. It presents point estimates and both 95% and 90% confidence intervals calculated with robust standard errors. Left panel: 229 split groups and 596 non-split groups. Right panel: 178 squiggly-split groups and 51 straight-split groups. The top panel shows estimates for geographic variables. The disaggregated rivers and lakes regressions include controls for both major/minor. Other models in the top panel are bivariate. The bottom panel presents three estimates for precolonial states: bivariate, the same set of control variables from the grid cell analysis, and those plus region FE. PCS MURDOCK, which is coded from Murdock's jurisdictional hierarchy variable, has 69 missing values, causing 69 observations to drop in the left panel and 9 observations to drop in the right panel. Tables D.6 and D.7 provide accompanying regression tables.

Results for geography. The top panel of Figure A.9 presents estimates for geographic variables. Most specifications are binary, although we include (a) major and minor rivers in the same model, and (b) major and minor lakes in the same model. Visible geographic focal points—rivers, lakes, and major watersheds—covary with an elevated likelihood of ethnic group partition, consistent with our theoretical expectations. Ethnic homelands containing a river or a lake are more likely to be partitioned: 31% of groups with any river in their territory were partitioned compared to 24% among groups lacking this feature, and the figures are similar for lakes (38% vs. 26%). The effect of river on partition is primarily driven by major rivers as opposed to minor ones: 40% of groups with a top 10 river were partitioned compared to 28% among groups with only minor rivers in their territory. Major watersheds affect the likelihood of partition: 39% of groups with a major watershed in their territory were partitioned compared to 22% among groups without a major watershed. Rivers also affect the *type* of partition. The presence of any river increases the likelihood of a squiggly split (80% versus 66% otherwise). Lakes, on the other hand, do not affect the type of split. Unlike inherently squiggly river borders, some international borders involving lakes follow the squiggly median line between shores (e.g., Lake Tanganyika) whereas others cut across the lakes with straight lines (e.g., Lake Victoria), leading to a null aggregate effect. Overall, the statistical results suggest that water bodies influenced border formation.

As expected, an ethnic group's percentage of desert area does not affect the likelihood of partition.

However, a larger desert area increases the likelihood of ethnic partition via a straight-line border. This supports our expectation that Europeans drew more haphazard borders where population density and strategic interests were low.

Results for precolonial states. The bottom panel of Figure A.9 shows results for PCS MURDOCK. We first present the bivariate result. Since PCS MURDOCK is endogenous, next we control for the same set of geographic and other covariates used in the grid cell analysis. Finally, we control for region fixed effects to compare groups within similar regions of Africa.

In our main analysis with grid cells, we demonstrate that precolonial states are less likely to be partitioned. We do not replicate this finding with Murdock ethnic groups. The coefficients for PCS MURDOCK on the left panel are close to 0 and insignificant. Furthermore, the raw magnitudes are small: 27% of groups with PCS MURDOCK=1 were partitioned compared to 29% with PCS MURDOCK=0. The coefficients on the right panel are positive but statistically insignificant, failing to confirm that PCS MURDOCK affects the type of split.

A.3.2 Shortcomings of the Murdock Data

The Murdock data are too noisy to provide a valid assessment of the relationship between precolonial states and partition. This helps to account for why the correlation is strong when using our original data but null correlations when using Murdock. We critique the Murdock jurisdictional hierarchy variable on two grounds: (1) Murdock's jurisdictional hierarchy variable exhibits substantial *measurement error*, and (2) ethnic groups exhibit a *conceptual mismatch* with the spatial reach of historical states.

Table A.7: Partitioned Ethnic Groups with Precolonial States: Murdock

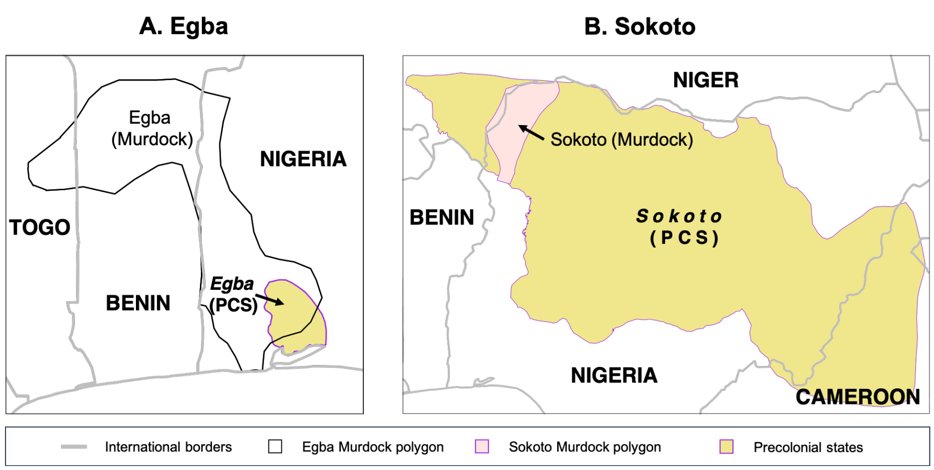
Murdock group	Country	Our assessment	Murdock group	Country	Our assessment
Delim	Western Sahara	Not a state	Regeibat	Mauritania	Not a state
Esa	Somalia	Not a state	Ronga	Mozambique	Not a state
Fon	Benin	Not partitioned (Dahomey)	Ruanda	Rwanda	Not partitioned
Gil	Morocco	Not a state	Rundi	Burundi	Not partitioned
Hamama	Tunisia	Not a state	Runga	Chad	Not a state
Hiechware	Botswana	Not a state	Songhai	Mali	Not a state
Imragen	Western Sahara	Not a state	Sotho	South Africa	Agree
Ishaak	Somalia	Not a state	Subia	Namibia	Not a state
Jerid	Tunisia	Not a state	Swazi	Swaziland	Agree
Kgatla	South Africa	Not a state	Tabwa	DRC	Not a state
Mandara	Nigeria	Not a state	Tama	Sudan	Not a state
Manga	Niger	Not a state	Tienga	Nigeria	Not a state
Masalit	Sudan	Not a state	Tlokwa	South Africa	Not a state
Mashi	Zambia	Not a state	Tripolitanians	Libya	Not a state
Mpezeni	Zambia	Not a state	Tunisians	Tunisia	Not partitioned
Popp	Benin	Not a state	Wakura	Nigeria	Not a state

Notes: This table lists every ethnic group for which (a) Murdock codes a jurisdictional hierarchy score of 3 or above, and (b) Michalopoulos and Papaioannou (2016) code the group as partitioned. We report the primary country to which the group belonged using the information from Michalopoulos and Papaioannou (2016).

Measurement error. The null relationship shown above between Murdock states and ethnic partition *almost entirely reflects measurement error in the data*. To show this, in Table A.7, we sample every "positive-positive" case from the regressions presented above, that is, every case with PCS MURDOCK=1 and the ethnic group is partitioned according to the criterion in Michalopoulos and Papaioannou (2016). For only two of the 32 cases do we agree that (a) members of the ethnic group created centralized political institutions and (b) the core area of the historical state was partitioned across international borders. To make this assessment, we compared the Murdock groups with high jurisdictional hierarchy scores to the PCS list used for our main analysis, concluding that 26 of these ethnic groups did not belong to historical states. Among the few remaining groups that met our PCS criteria, we assessed that only two of the six corresponding states were partitioned in the sense that core areas of the state were divided across colonial borders (Appendix B provides details on the frontiers of every PCS in our data set).

Conceptual mismatch. Ethnic group is not a valid unit of analysis for measuring precolonial African states. Typically, even where states existed, the ethnic homeland was geographically larger (and thus subject to partition even if the core state(s) were not); and some states comprised territory settled by multiple ethnic groups. We illustrate these points with two examples. Panel A of Figure A.10 presents the Murdock polygon for Egba in white and ours in yellow. As we discuss in Appendix B, we incorporate the historical state governed by the Alake of Egba; as we note, if anything, our polygon is too big. But, assuming Murdock's assessment of the Egba homeland is correct, by measuring the location of ethnic groups rather than states, using his polygon yields a false positive when assessing whether the historical state was partitioned. This case highlights two additional problems with the Murdock. First, Murdock codes Egba as two levels of political hierarchy above the village level, that is, a paramount chieftaincy rather than a state. However, historical sources argue that Egba was the most powerful state to emerge in Yorubaland following the collapse of the Oyo Empire early in the nineteenth century (see Appendix B). Second, Murdock's Egba polygon is undoubtedly too large even given the goal of measuring ethnic groups (see the map in Forde 1951).

Figure A.10: Comparing Murdock Polygons



Panel B examines the Sokoto Caliphate. Our polygon is much bigger and corresponds with the extent of the historical state. This is an odd entry in Murdock. The Sokoto Caliphate was governed by ethnic Fulani, and many of the new emirates in the north displaced historical Hausa states. Sokoto was a state, not an ethnic group, and thus should not appear in his data set at all. Furthermore, his Sokoto polygon corresponds roughly with the Sokoto emirate only, not the entire empire. Finally, and strangely, Murdock incorrectly codes Sokoto as exhibiting only one level of hierarchy above the village level. In sum, in both this and the Egba case, even if we correct the jurisdictional hierarchy score, the Murdock polygon is too inaccurate to use for empirical analysis.

References

Abraham, Garth. 2007. ""Lines upon Maps": Africa and the Sanctity of African Boundaries." *African Journal of International and Comparative Law* 15:61–84.

Alesina, Alberto, William Easterly and Janina Matuszeski. 2011. "Artificial States." *Journal of the European Economic Association* 9(2):246–277.

Alexandrowicz, C.H. 1973. The Partition of Africa by Treaty. In *Foreign Relations of African States*, ed. Kenneth Ingham. Butterworths pp. 129–155.

Alsan, Marcella. 2015. "The Effect of the Tsetse Fly on African Development." *American Economic Review* 105(1):382–410.

Appiah, Anthony and Henry Louis Gates. 2010. *Encyclopedia of Africa*. 1 ed. Oxford University Press.

Barbour, K.M. 1961. A Geographical Analysis of Boundaries in Inter-tropical Africa. In *Essays on African Population*, ed. K.M. Barbour and R.M. Prothero. pp. 303–323.

Brownlie, Ian. 1979. *African Boundaries: A Legal and Diplomatic Encyclopaedia*. C. Hurst & Co. Publishers.

Christensen, Darin and David D Laitin. 2019. *African States since Independence: Order, Development, and Democracy*. Yale University Press.

Craven, Matthew. 2015. "Between Law and History: The Berlin Conference of 1884–1885 and the Logic of Free Trade." *London Review of International Law* 3(1):31–59.

Crowe, S.E. 1942. *The Berlin West African Conference 1884–1885*. Negro Universities Press.

Englebert, Pierre. 2002. *State Legitimacy and Development in Africa*. Lynne Rienner Publishers.

Faal, Courtney. 2009. "The Partition of Africa.".
URL: *https://www.blackpast.org/global-african-history/partition-africa/*

Fenske, James. 2014. "Ecology, Trade, and States in Pre-Colonial Africa." *Journal of the European Economic Association* 12(3):612–640.

Fischer, Hilke. 2015. "130 years ago: carving up Africa in Berlin.".
URL: *https://www.dw.com/en/130-years-ago-carving-up-africa-in-berlin/a-18278894*

Flint, John E. 1960. *Sir George Goldie and the Making of Nigeria*. Oxford University Press.

Food and Agricultural Organization of the United Nations. 2022. "Hydrological Basins in Africa." Available at https://data.apps.fao.org/catalog/iso/e54e2014-d23b-402b-8e73-c827628d17f4.

Forde, Daryll. 1951. *Ethnographic Survey of Africa: Western Africa Part IV. The Yoruba-Speaking Peoples of South-western Nigeria*. International African Institute.

Goemans, Hein E. and Kenneth A. Schultz. 2017. "The Politics of Territorial Claims: A Geospatial Approach Applied to Africa." *International Organization* 71(1):31–64.

Herbst, Jeffrey. 1989. "The Creation and Maintenance of National Boundaries in Africa." *International Organization* 43(4):673–692.

Herbst, Jeffrey. 2000. *States and Power in Africa: Comparative Lessons in Authority and Control*. Princeton University Press.

Hertslet, E. 1909. *The Map of Africa by Treaty*. His Majesty's Stationery Office.

Louis, William Roger. 1963. *Ruanda-Urundi, 1884-1919*. Clarendon Press.

Michalopoulos, Stelios and Elias Papaioannou. 2016. "The Long-Run Effects of the Scramble for Africa." *American Economic Review* 106(7):1802–1848.

Murdock, George Peter. 1959. *Africa: Its Peoples and their Culture History*. McGraw-Hill.

Murdock, George Peter. 1967. *Ethnographic Atlas*. University of Pittsburgh Press.

Natural Earth. 2023. "Rivers + lake centerlines." Available at https://www.naturalearthdata.com/downloads/10m-physical-vectors/10m-rivers-lake-centerlines.

Nunn, Nathan. 2008. "The Long-Term Effects of Africa's Slave Trades." *Quarterly Journal of Economics* 123(1):139–176.

Oster, Emily. 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics* 37(2):187–204.

Paine, Jack. 2019. "Redistributive Political Transitions: Minority Rule and Liberation Wars in Colonial Africa." *Journal of Politics* 81(2):505–523.

Ricart-Huguet, Joan. 2022. "The Origins of Colonial Investments in Former British and French Africa." *British Journal of Political Science* 52(2):736–757.

Roodman, David, Morten Ørregaard Nielsen, James G MacKinnon and Matthew D Webb. 2019. "Fast and Wild: Bootstrap Inference in Stata using boottest." *The Stata Journal* 19(1):4–60.

Rosenberg, Matt. 2019. "The Berlin Conference to Divide Africa.".
URL: *https://www.thoughtco.com/berlin-conference-1884-1885-divide-africa-1433556*

Sanderson, G.N. 1985. The European Partition of Africa: Origins and Dynamics. In *The Cambridge History of Africa Vol. 6: from 1870 to 1905*, ed. Roland Oliver and G. N. Sanderson. Cambridge University Press pp. 96–158.

Sandouno, Moïse. 2015. Une histoire des frontières guinéennes (années 1880-2010): héritage colonial, négociation et conflictualité PhD thesis Université Toulouse le Mirail.
URL: *https://theses.hal.science/tel-01140352*

South African History Online. 2019. "The Berlin Conference.".
URL: *https://www.sahistory.org.za/article/berlin-conference*

Utrecht University. 2022. "History Database of the Global Environment 3.2." Available at https://public.yoda.uu.nl/geo/UU01/MO2FF3.html.

Vivid Maps. 2001. "The Major River Basins of Africa." Available at https://vividmaps.com/the-major-river-basins-of-africa.

Wesseling, H.L. 1996. *Divide and Rule: The Partition of Africa, 1880–1914*. Praeger.

White, Frank. 1983. *The Vegetation of Africa*. Paris, France: UNESCO.

Yakemtchouk, Romain. 1971. *L'Afrique en Droit International*. Librairie generale de droit et de jurisprudence.