Online Appendix for
Why Does Ethnic Partition Foster Violence? Unpacking the Deep Historical Roots of Civil Conflicts

Gaku Ito*
Hiroshima University
gkit@hiroshima-u.ac.jp

June 19, 2020

Contents

	Overview	**2**
A	**Descriptive Statistics and Regression Results**	**2**
A.1	Covariate Balance	2
A.2	Determinants of Ethnic Partition	5
A.3	Ethnic Partition and Political Exclusion	5
A.4	Territorial and Governmental Conflicts	7
A.5	Summary Statistics of Posttreatment Variables	8
B	**Confounding**	**9**
B.1	Model Dependence	9
B.2	Sensitivity Analysis	10
C	**Alternative Coding of Ethnic Partition**	**11**
D	**Measurement Error and Geographically Small-Sized Groups**	**13**
E	**Ethnic Partition, Political Discrimination, and Conflict Onset**	**14**

*Assistant Professor, Graduate School of Humanities and Social Sciences, Hiroshima University. 1–5–1 Kagamiyama, Higashihiroshima, Hiroshima 739–8529, Japan.

Overview

This Online Appendix reports a series of robustness checks and additional results. Section A presents supplementary summary statistics and regression estimates that are briefly reported in the main text. Section B addresses the issue of possible confounding, both observed and unobserved. Section B.1 examines possible model dependence in the baseline total effect estimate by subsequently estimating the average treatment effects (ATEs) for every possible model specification where different combinations of the ten pretreatment covariates listed in the main text enter the model with *Partition*, country fixed effects, and spatial and temporal polynomials. To address the related but another issue of unobserved confounding, Section B.2 utilizes the E-value approach to examine the minimum strength of omitted variable bias to negate the conflict-escalating effect of ethnic partitioning (Ding & VanderWeele, 2016; VanderWeele & Ding, 2017). The following two sections address the sensitivity concerns for the coding of ethnic partitioning. Specifically, Section C replicates the ATE estimates and sequential g-estimation with an alternative threshold value of 95% for the coding of *Partition*. Section D addresses the concern for measurement error in coding of ethnic partitioning by a series of subsample estimates dropping geographically small-sized and potentially 'noisy' observations. Finally, Section E reports further regression results focusing on the political discrimination-conflict link.

The empirical analysis and estimations reported in this article were conducted in R 3.6.1. Note that the geoprocessing primarily relied on sf and sp packages in R (Bivand et al., 2013; Pebesma & Bivand, 2005), a series of original R implementations, and the Universal Transverse Mercator (UTM) coordinate system. I employed the CShapes dataset to detect of ethnic partition by contemporary political boundaries (Weidmann et al., 2010), and sandwich package in R to compute the standard errors robust to multiway clustering reported in the regression tables (Berger et al., 2017; Cameron et al., 2011; Zeileis, 2004).

A Descriptive Statistics and Regression Results

A.1 Covariate Balance

Table A.I reports the balance statistics between the treated (split) and control (non-split) ethnic groups. These 239 ethnic groups are included in the empirical analysis reported in the main text. Table A.II and Figure A.1 report the analogous balance statistics for the full sample containing 825 ethnic groups listed in Murdock (1959).

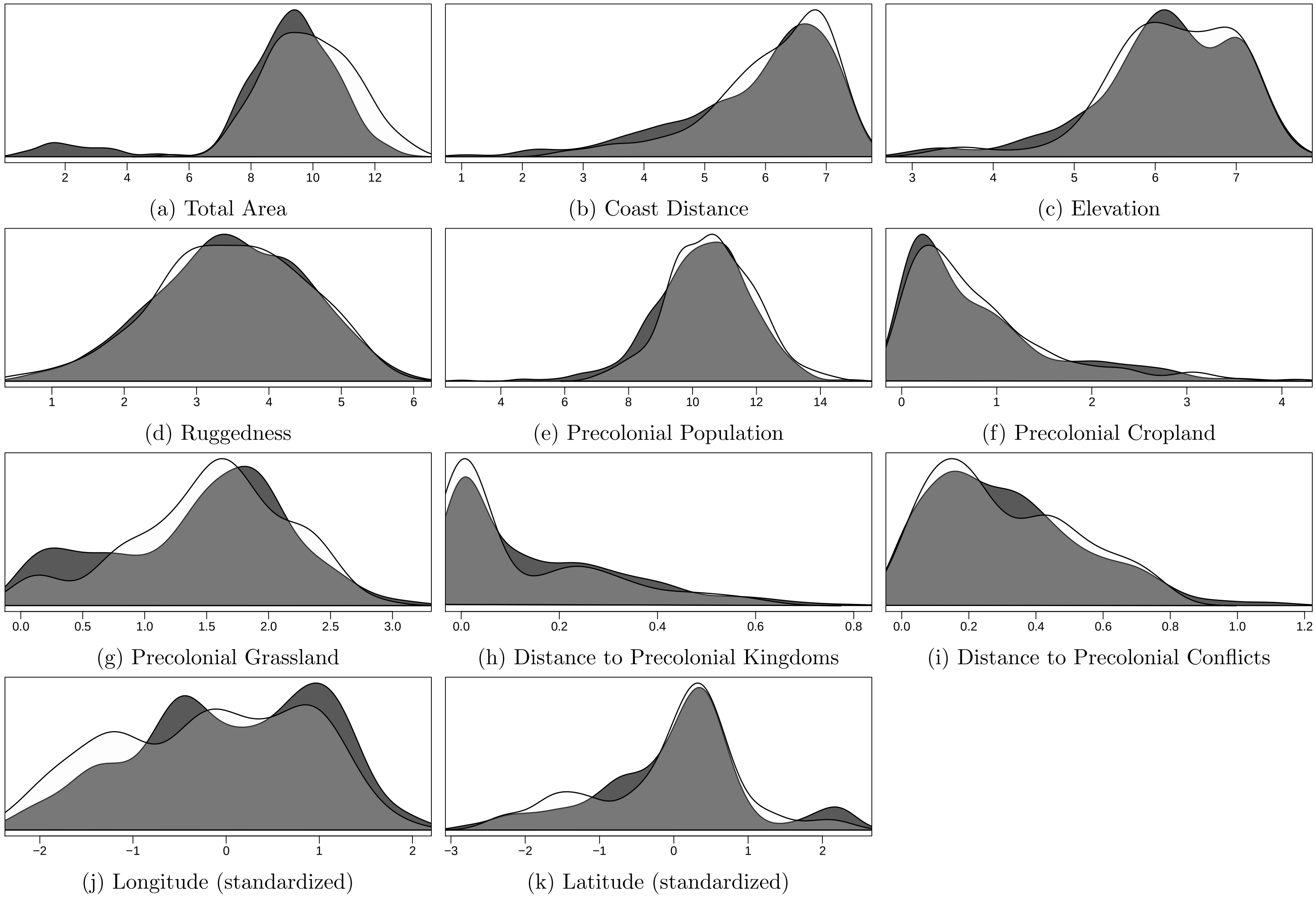


Figure A.1. Density Plots for the Pretreatment Covariates for All Ethnic Groups Listed in Murdock (1959)

Note: The light gray density curves represent the distributions of covariates for the treated (split) ethnic groups, whereas the dark gray curves indicate the distributions of control (non-split) groups.

Table A.I. Covariate Balance: Ethnic Groups in the Main Analysis

	Mean T	Variance T	Mean C	Variance C	Mean Difference	Standardized Bias (in %)
	($N_{\text{Treated}} = 89$)		($N_{\text{Control}} = 150$)			
Geographic						
Total Area	10.344	1.995	9.352	5.601	0.993	50.937
Coast Distance	6.185	1.175	6.166	1.371	0.018	1.633
Elevation	6.124	0.734	6.179	0.446	−0.055	−7.142
Ruggedness	3.235	1.069	3.33	1.053	−0.095	−9.192
Water Body (dummy)	0.674	0.222	0.567	0.247	0.107	22.189
Longitude	−0.364	0.987	0.007	0.902	−0.371	−38.195
Latitude	0.047	0.931	0.248	0.892	−0.201	−21.083
Socioeconomic						
Precolonial Population	10.961	1.959	10.817	2.286	0.144	9.858
Precolonial Cropland	0.827	0.7	0.712	0.576	0.115	14.396
Precolonial Grassland	1.559	0.419	1.439	0.525	0.12	17.444
Distance to Precolonial Kingdom	0.119	0.03	0.162	0.04	−0.043	−22.829
Distance to Precolonial Conflict	0.322	0.053	0.345	0.043	−0.023	−10.672

Note: Standardized bias $= 100 \times (\bar{x}_T - \bar{x}_C)/\sqrt{(s^2_T(x) + s^2_C(x))/2}$, where $\bar{x}_T$ ($\bar{x}_C$) is the sample mean in the treatment (control) group, and $s^2_T(x)$ ($s^2_C(x)$) denotes sample variance.

Table A.II. Covariate Balance: All Ethnic Groups in Murdock (1959)

	Mean T	Variance T	Mean C	Variance C	Mean Difference	Standardized Bias (in %)
	($N_{\text{Treated}} = 213$)		($N_{\text{Control}} = 612$)			
Geographic						
Total Area	9.901	2.01	8.925	4.976	0.976	52.222
Coast Distance	6.103	1.017	5.891	1.528	0.212	18.785
Elevation	6.202	0.669	6.144	0.801	0.058	6.761
Ruggedness	3.474	1.05	3.49	1.056	−0.015	−1.488
Water Body (dummy)	0.624	0.236	0.534	0.249	0.09	18.299
Longitude	−0.171	1.041	0.059	0.974	−0.23	−22.91
Latitude	−0.049	0.926	0.017	1.026	−0.066	−6.712
Socioeconomic						
Precolonial Population	10.604	1.716	10.323	2.224	0.281	20.034
Precolonial Cropland	0.799	0.591	0.792	0.613	0.007	0.94
Precolonial Grassland	1.517	0.407	1.437	0.517	0.079	11.666
Distance to Precolonial Kingdom	0.119	0.027	0.151	0.031	−0.032	−18.801
Distance to Precolonial Conflict	0.304	0.045	0.315	0.052	−0.01	−4.612

Note: Standardized bias $= 100 \times (\bar{x}_T - \bar{x}_C)/\sqrt{(s^2_T(x) + s^2_C(x))/2}$, where $\bar{x}_T$ ($\bar{x}_C$) is the sample mean in the treatment (control) group, and $s^2_T(x)$ ($s^2_C(x)$) denotes sample variance.

A.2 Determinants of Ethnic Partition

Table A.III presents a series of regressions to evaluate the effects of the pretreatment variables on ethnic partition, briefly reported in the main text. As discussed in the main text, the estimation results show that several pretreatment covariates are systematically associated with the treatment assignment (ethnic partition). Reflecting the earlier findings in Michalopoulos & Papaioannou (2016) that geographical size of ethnic homelands is a key determinant of ethnic partition. In addition, while the statistical significance varies, *Precolonial Cropland*, *Elevation*, and *Ruggedness* are consistently and positively associated with the probability of ethnic partition, while *Precolonial Population* yields a negative effect. While the relatively large standard errors do not allow us to draw a definite conclusion, these results suggest that several geographic and socioeconomic factors shaped African border design by affecting precolonial state formation and colonizers' knowledge of localities (see also, Green, 2012; Osafo-Kwaako & Robinson, 2013).

Omitting these local-level factors in the total treatment effect estimation can lead to spurious findings if these factors systematically influence ethnic partition *and* postcolonial civil conflicts. To address this potential concern for confounding, the empirical analysis in the main text controls for these covariates.

A.3 Ethnic Partition and Political Exclusion

The sample of the main analysis includes the ethnic groups that are politically excluded from the central state power of host countries for the years of observations. As briefly noted in the main text, one may wonder if political exclusion is affected by ethnic partition, and thus this sample restriction induces a form of sample selection (collider) bias. This concern would be, at least partly, alleviated if ethnic partitioning is not systematically associated with the probability of political inclusion/exclusion during the postcolonial period.

To explicitly address this concern, Table A.IV reports the regression estimates of political exclusion on ethnic partition as well as the pretreatment covariates using the full sample in panel and cross-section formats (with both politically included and excluded groups, $N_{\text{obs}} = 30,843$), with and without country fixed effects and polynomial terms. Models 1 to 3 utilize the full panel dataset and regress *Exclusion* on *Partition* and different sets of covariates. Model 1 includes the treatment, pretreatment covariates, and calender year and spatial polynomials, whereas Model 2 further adjusts for country fixed effects and group-level peace-year polynomial. Model 3 replicates the specification of Model 2 with the weighted least squares with the number of group ties between Murdock's (1959) ethnolinguistic maps and

Table A.III. Determinants of Ethnic Partition

	Dependent variable: Partition			
	(1) OLS LPM	(2) GAM LPM	(3) GLM logit	(4) GAM logit
Geographic				
Total Area	0.069*** (0.016)	0.065*** (0.020)	0.754*** (0.175)	0.649*** (0.222)
Coast Distance	−0.060 (0.067)	−0.083 (0.086)	−0.451 (0.359)	−0.576 (0.501)
Elevation	0.129 (0.080)	0.152 (0.110)	0.782** (0.332)	0.905 (0.657)
Ruggedness	0.013 (0.010)	0.003 (0.051)	0.101 (0.108)	0.014 (0.278)
Water Body	0.054 (0.096)	0.085 (0.080)	0.220 (0.457)	0.358 (0.440)
Socioeconomic				
Precolonial Population	−0.023 (0.025)	−0.017 (0.030)	−0.343 (0.268)	−0.261 (0.195)
Precolonial Cropland	0.088 (0.063)	0.113 (0.069)	0.857** (0.411)	0.915** (0.416)
Precolonial Grassland	0.084 (0.066)	0.067 (0.067)	0.518 (0.436)	0.414 (0.372)
Distance to Precolonial Kingdom	−0.107 (0.267)	0.019 (0.317)	−0.470 (1.445)	0.011 (1.846)
Distance to Precolonial Conflict	0.132 (0.402)	0.162 (0.301)	1.365 (2.316)	1.372 (1.714)
Region FEs	✓	✓	✓	✓
Spatial polynomial (lon, lat)	✓		✓	
Spatial spline (lon, lat)		✓		✓
Observations	239	239	239	239
Adjusted R^2	0.127			
AIC		314.453	296.512	294.146
UBRE		0.220		0.231
Residual Std. Error	0.453 (df = 219)			

Note: *$p<0.1$; **$p<0.05$; ***$p<0.01$. Constants not reported for brevity.
Cluster-robust standard errors with multiway clustering on regions and countries in parentheses (OLS and GLM). Standard errors in parentheses (GAMs). UBRE: Unbiased Risk Estimator.

the EPR dataset (Michalopoulos & Papaioannou, 2016). Model 4 collapses the panel data into a cross-section format and regresses an alternative dependent variable *Excluded Years*, or logged total years in which each group (nested by host countries) has been excluded from central political power, on all pretreatment covariates and country fixed effects.[1]

Across model specifications, the association between ethnic partition and postcolonial political exclusion fails to retain substantial and statistical significance. *Partition* is never retain statistical significance at the conventional 5% level regardless of the dependent variables. Another key result emerging from the estimations is that the size of its coefficients remain substantially small while remaining stable across two OLS specifications (Models 1 and 2) despite the noticeable increase in the R^2 statistics, suggesting that the estimates are not likely to be driven by omitted variable bias (Oster, 2019). These estimation results

[1] 1 is added before taking logarithm for all observations to code *Excluded Years* such that *Excluded Years* = ln(total years excluded + 1).

Table A.IV. Ethnic Partition and Political Exclusion

	Dependent variable:			
		Exclusion		*Excluded Years*
	(1) OLS	(2) OLS	(3) WLS	(4) OLS
Partition	0.040 (0.045)	0.045 (0.030)	0.011 (0.021)	0.145 (0.190)
Pretreatment Covariates	✓	✓	✓	✓
Country FEs		✓	✓	✓
Calender year polynomial	✓	✓	✓	
Peace year polynomial		✓	✓	
Spatial polynomial (lon, lat)	✓	✓	✓	✓
Data	Panel	Panel	Panel	Cross-section
Observations	30,843	30,843	30,843	643
Adjusted R^2	0.166	0.521	0.703	0.112
Residual Std. Error	0.450	0.341	0.812	1.682
	(df = 30,823)	(df = 30,782)	(df = 30,782)	(df = 626)

Note: *p <0.1; **p <0.05; ***p <0.01. Constants not reported for brevity.
Robust standard errors with multiway clustering on ethnic groups, countries, and years (Models 1 to 3) and ethnic groups and countries in parentheses (Model 4).

suggest that the sample restriction in the main analysis would not likely lead a serious concern sample selection issues while allowing for clearly setting the scope of the analysis to the (failure of) conflict bargaining between the central government and politically excluded groups illustrated by the theoretical accounts in the main text.

A.4 Territorial and Governmental Conflicts

The analysis in the main text does not distinguish conflict issues when examining the long-run impact of ethnic partition. Yet, one might reasonably suspect whether ethnic partition leads to civil conflict depends on the disputed issues such that the conflict escalating effect is stronger for the instances of secessionist or territorial conflicts. To address this concern, Table A.V disaggregates the dependent variable into two binary variables each indicating the onset of conflicts that involve territorial or governmental incompatibilities. The coding of incompatibilities follows the ACD2EPR dataset.

As the estimation results suggest, we see little heterogeneity in the conflict-escalating effect of ethnic partition across different types of conflict. *Partition* is positively signed in all model specifications, indicating that ethnic partition is followed by increased risks of armed conflict regardless of the disputed issues. If anything, and somewhat counterinsuitively, the coefficient on *Partition* in Model 2 in Table A.V retains only marginal statistical significance at the conventional 5% level (t value = 1.8222).

Table A.V. Total Effect of Ethnic Partition across Conflict Issues

	Dependent variable: Onset		
	(1) All	(2) Territorial	(3) Governmental
Partition	0.0082*** (0.0028)	0.0035* (0.0019)	0.0047*** (0.0017)
Pretreatment Covariates	✓	✓	✓
Country FEs	✓	✓	✓
Calender year polynomial	✓	✓	✓
Peace year polynomial	✓	✓	✓
Spatial polynomial (lon, lat)	✓	✓	✓
Observations	10,246	10,246	10,246
Adjusted R^2	0.0174	0.0211	0.0187
Residual Std. Error (df = 10,193)	0.1284	0.0786	0.1021

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$. Constants not reported for brevity. Robust standard errors with multiway clustering on ethnic groups, countries, and years in parentheses.

A.5 Summary Statistics of Posttreatment Variables

Table A.VI presents the summary statistics of mediator and intermediate confounders employed in the sequential g-estimation reported in the main text. *Discrimination* enters the two-stage estimation as an intermediate confounder (of the first-stage model) in the specification with *Group Size* or *Demographic Balance* as the mediator. Analogously, *Group Size* is included as an intermediate confounder in the specification where *Discrimination* is set as the mediator.

Table A.VI. Summary Statistics: Posttreatment Variables

	Mean	SD	Min	25th Percentile	75th Percentile	Max
Mediator						
Group Size	7.253	1.503	2.536	6.418	8.415	10.118
Demographic Balance	−2.172	1.102	−5.298	−2.996	−1.273	−0.163
Discrimination	0.092	0.289	0	0	0	1
Group-level variables						
Area	9.688	1.528	3.728	8.913	10.671	13.110
Border Distance	4.434	1.173	1.212	3.639	5.365	6.398
Capital Distance	6.218	0.785	2.531	5.759	6.824	7.564
Population 1940	10.884	2.323	0.000	10.101	12.261	15.392
Cropland 1940	1.331	0.939	0.000	0.622	1.753	4.309
Grassland 1940	2.721	1.124	0.000	2.173	3.591	4.379
Country-level variables						
Total Population	9.425	1.057	5.966	8.880	10.154	11.998
per capita GDP	6.540	0.970	4.395	5.642	7.255	9.468
State Age	3.127	1.077	0	2.6	3.7	6

B Confounding

B.1 Model Dependence

As noted in the main text, an important concern arising from the imbalance between the treated (split) and control (non-split) observations across pretreatment covariates is model dependence, or the sensitivity of the estimates to alternative model specifications. Although the relative stability of the coefficients on *Partition* reported in Table I in the main text implies that covariate imbalance may not be a serious concern in the current context, this methodological concern is worth a focused evaluation.

I follow the approach in Ho et al. (2007) to investigate model dependence. Specifically, I estimate the ATEs for every possible model specification where different combinations of 10 pretreatment covariates in Model 2 in Table I enter the model with *Partition*, country fixed effects, and spatial and temporal polynomials. The combinations amount to $2^{10} = 1,024$ model specifications.

Figure B.1 uses a kernel density plot to summarize the empirical distribution of the ATE estimates obtained from the 1,024 model specifications. The results broadly suggest that the covariate imbalance has limited impacts on the treatment effect estimates. The mean of the point estimates is 0.0083 (95% CI: 0.0069, 0.0098) and lies close to the estimate from Model 2 of 0.0082. If anything, the baseline result from Model 2 slightly underestimates the ATE. Moreover, as represented in Figure B.1, the empirical distribution of the ATE estimates is close to the normal distribution, indicating that covariates included are having little systematic impact on the ATE estimates.

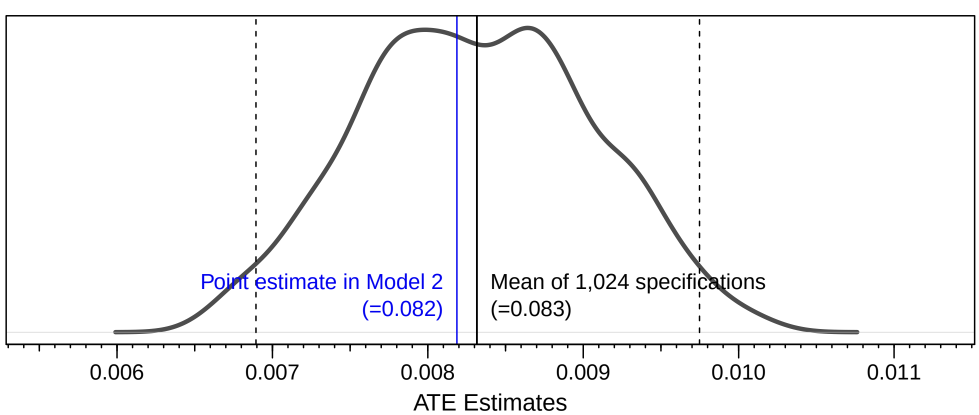


Figure B.1. Kernel Density Plot of the ATE Estimates across 1,024 Specifications

Note: Blue vertical segment indicates the ATE estimate in Model 2 in Table I in the main text. Black vertical segments represent the mean of the ATE estimates across 1,024 specifications (solid) and the corresponding 95% confidence intervals (dashed).

B.2 Sensitivity Analysis

A remaining methodological concern is unobserved confounding. If the total effect estimate of ethnic partition suffer from omitted variable bias in the first place, the validity of the subsequent mediation analysis also remain inconclusive at best. Insights from previous studies suggest that several precolonial features may affect both ethnic partition (treatment) and contemporary conflicts (outcome), including precolonial polities (Besley & Reynal-Querol, 2014) and group-level traditional political institutions inherited since the pre-colonial period (Michalopoulos & Papaioannou, 2013; Ray, 2019; Wig, 2016). Although the main analysis adjusts for the major observed, group-level confounders, the concern for unobserved confounding is worth a further analysis.

While the assumption of no omitted variables can hardly be tested against observational data, the sensitivity (bias) analysis techniques allow us to shed light on the potential effect of unobserved confounding on the treatment effect estimates. Here, I employ a bounding factor, the E-value, recently proposed by Ding & VanderWeele (2016) and VanderWeele & Ding (2017). Let RR_τ denote an observed treatment effect in the risk ratio scale and a pair of RR_{UT} ($U \rightarrow$ treatment) and RR_{UO} ($U \rightarrow$ outcome) associations be possible unobserved confounding 'treatment $\leftarrow U \rightarrow$ outcome.' The E-value measures, in the risk ratio scale, the minimum size of RR_{UT} and RR_{UO} to eliminate RR_τ when $RR_{UT} = RR_{UO}$. Specifically, the E-value is defined as $E = RR^+_\tau + \sqrt{RR^+_\tau(RR^+_\tau - 1)}$, where RR_τ denotes the treatment-outcome association in the risk ratio scale adjusting for observed covariates, and $RR^+_\tau = RR_\tau$ if $RR_\tau > 1$ and $1/RR_\tau$ if $RR_\tau < 1$. A large E-value implies that strong unobserved confounding is needed to attenuate away RR_τ, whereas a small (close to the minimum value of 1) E-value implies that weak confounding is sufficient.An E-value of 1 indicates that no more confounding is needed to eliminate a treatment-outcome association. When the prevalence of an outcome is low as is the case in the current analysis, the odds ratios approximate well to the relative risk so that we can obtain the E-value by simply replacing RR_τ with an odds-ratio scale treatment effect OR_τ (VanderWeele & Ding, 2017).

The baseline estimates reported in the main text suggest that comparatively strong unobserved confounding is needed to move the treatment effect to the null value of 1. Specifically, the estimate of Model 4 (full logit) in Table I, or an odds ratio of $e^{0.3512} = 1.4208$ translates into $E^{\text{Full}} = 2.1939$. An E-value of 2.1939 indicates that in a special case with $OR_{UT} = OR_{UO}$, unobserved confounding needs to be associated with both ethnic partitioning and conflict onset by an odds ratio of roughly 2.2-fold each to induce the omitted variable bias that can explain away the treatment effect of 1.4207. Yet weaker confounding remains insufficient to attenuate away the treatment effect. Indeed, a quick comparison of the E-values

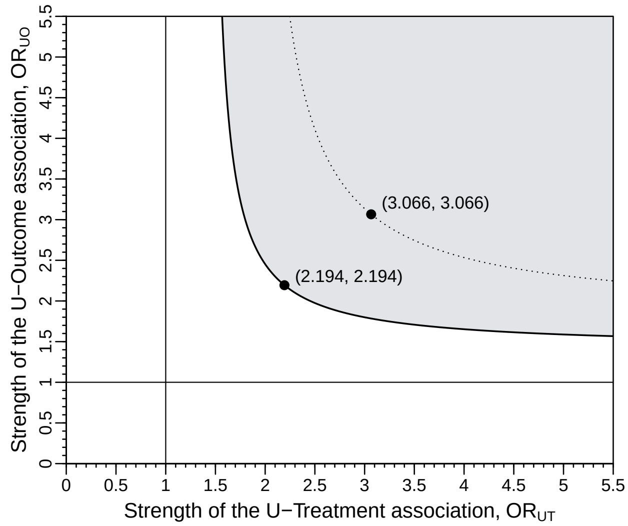


Figure B.2. Joint values of OR_{UT} and OR_{UO} to negate the treatment effect

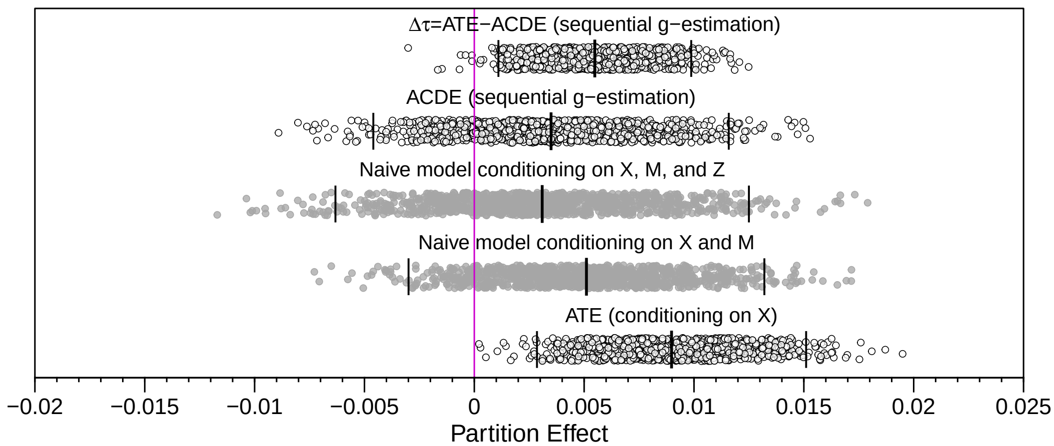
Note: The solid curve illustrates the joint values of OR_{UT} and OR_{UO} that are sufficiently large to satisfy $\frac{OR_{UT}OR_{UO}}{OR_{UT}+OR_{UO}-1} \geq OR_{\tau}^{+} = 2.1939$, whereas the dotted curve represents the corresponding estimate for the restricted model. The dots represent the special cases with $OR_{UT} = OR_{UO}$, summarized by the E-values of the full and restricted models.

derived from Model 4 (full model) and a restricted model (a logit estimate of Model 1 in Table I) indicates that the E-value drops from $E^{\text{Restricted}} = 3.0658$ ($OR_{\tau} = e^{0.6052} = 1.8316$) to 2.1939 by conditioning on 10 pretreatment covariates. The difference in the E-values of $\Delta E = E^{\text{Restricted}} - E^{\text{Full}} = 3.0658 - 2.1939 = 0.8719$ measures the size of bias invited by omitting observed confounders. The ΔE of 0.8719 suggests that the effect of unobserved confounding at least would have to be larger than that of observed pretreatment covariates (ΔE) to eliminate the effect of ethnic partition reported in the main text.

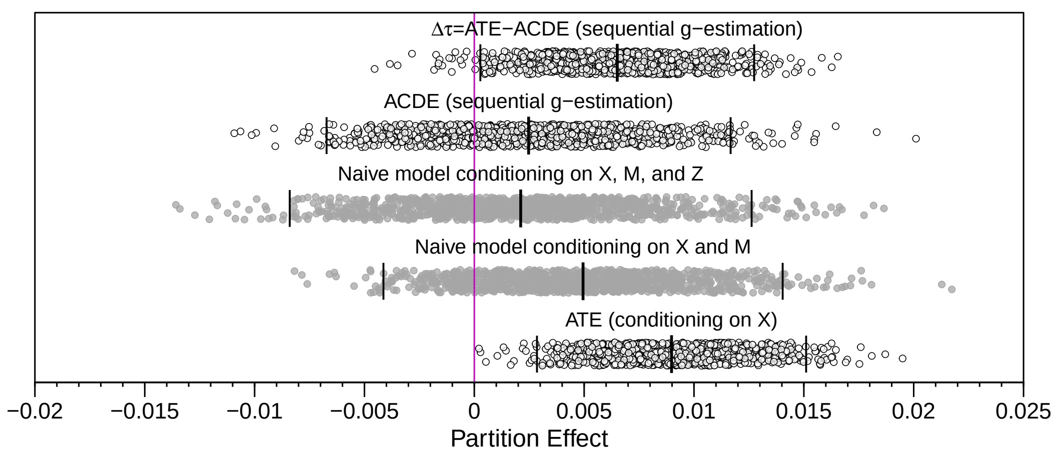
The E-value gives a summary statistic of the minimum strength of unobserved confounding sufficient to negate a treatment effect under the assumption of $RR_{UT} = RR_{UO}$, but different combinations of (OR_{UT}, OR_{UO}) can induce the same level of confounding bias. The general condition follows that the joint values of OR_{UT} and OR_{UO} must be sufficiently large to satisfy $\frac{OR_{UT}OR_{UO}}{OR_{UT}+OR_{UO}-1} \geq OR_{\tau}^{+}$ to fully wipe out OR_{τ} (Ding & VanderWeele, 2016: 370–372). To give a better illustration, Figure B.2 plots the joint values of (OR_{UT}, OR_{UO}) that suffice the condition, including the combination of $OR_{UT} = OR_{UO}$ as a special case.

C Alternative Coding of Ethnic Partition

The baseline analysis relies on a specific coding rule of ethnic partition such that *Partition* is coded 1 for the groups with their traditional settlements split into more than one country



(a) *Group Size*



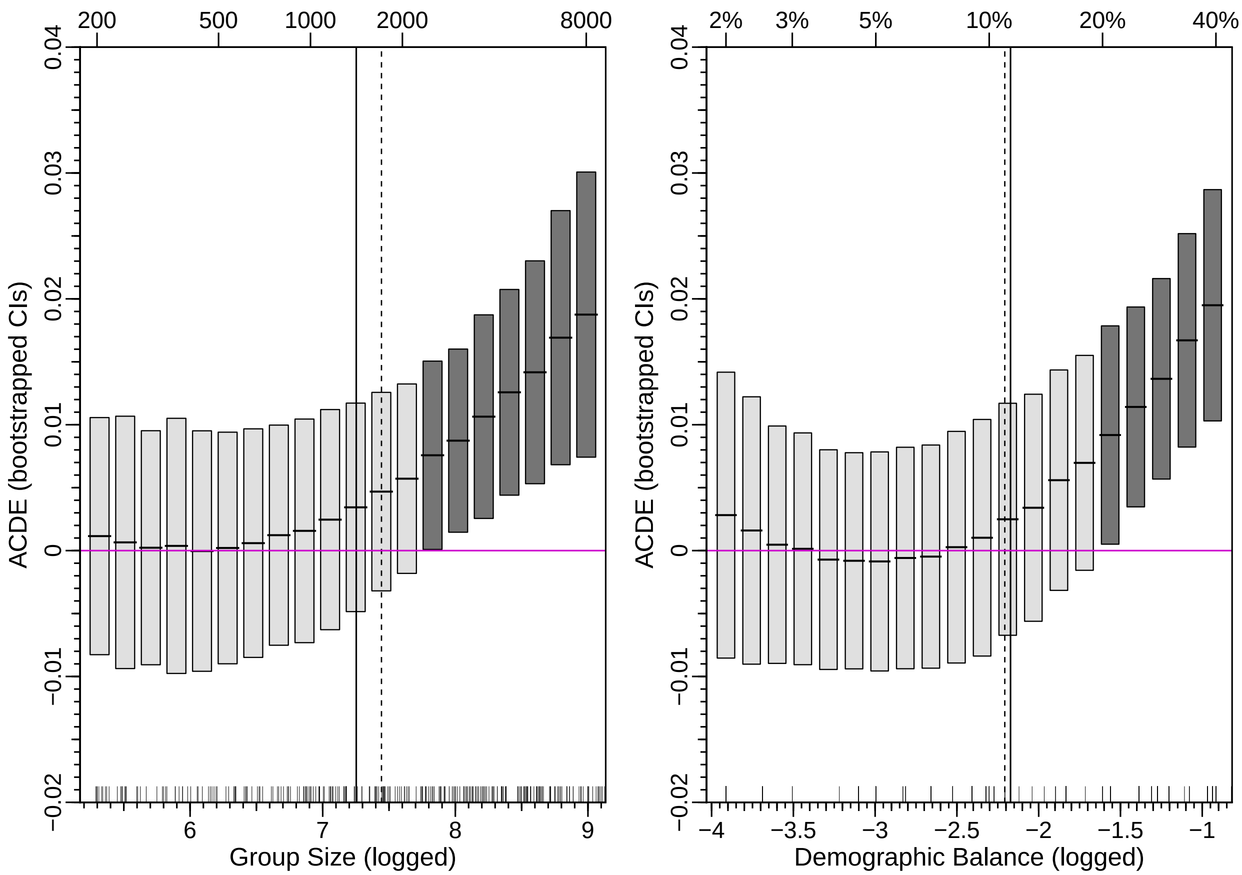
(b) *Demographic Balance*

Figure C.1. Effect of *Partition* on Conflict Onset with the Mediator Recentered to Mean, 95% threshold for *Partition*

Note: Direct effect estimates with (a) *Group Size* and (b) *Demographic Balance* as the mediator variable, along with the ATE estimate. Each dot indicates a bootstrap estimate, whereas vertical stripes represent the point estimates (mean) and 95% confidence intervals obtained via 1,000 bootstrap repetitions.

and 0 otherwise. *Partition* is coded 0 for a given ethnic groups if more than 90% of its settlement area falls within a single country. A possible concern arising from the coding procedure is the sensitivity of the estimation results to the specific threshold value of 90%.

To address this concern, Figures C.1 and C.2 employ an alternative threshold of 95% and replicates the mediation analysis. As represented in these figures, the main findings remain qualitatively unchanged with the alternative threshold. Contemporary demographic size shapes the long-run association between ethnic partitioning and postcolonial conflicts regardless of the choice of the threshold value.



(a) ACDE across *Group Size*　　　(b) ACDE across *Demographic Balance*

Figure C.2. Controlled Direct Effect of *Partition* across Different Demographic Sizes, with 95% threshold for *Partition*

Note: ACDE across the 10th to 90th percentile values of (a) *Group Size* (logged group population in 1,000) and (b) *Demographic Balance* (logged fraction of group population relative to host country's total population). Vertical rugs at the bottom indicate the distribution of mediator (randomly sampled 100 observations for visibility). Vertical segments spanning the panels represent mean (solid) and median (dashed) values. Horizontal stripes indicate the point estimates (mean), and rectangle edges represent the corresponding 95% bootstrap CIs obtained via 1,000 bootstrap repetitions. The estimates statistically significant at the 5% level are plotted in darker colors.

D Measurement Error and Geographically Small-Sized Groups

One may reasonably wonder if the results could simply reflect measurement error such that the coding of ethnic partitioning becomes 'noisier' for geographically (and demographically) smaller groups. If (classical) measurement error in ethnic partitioning is larger for geographically and demographically smaller groups than larger groups, it would attenuate the treatment effect for smaller-sized groups and thereby generates the mediating and conditioning role of contemporary group size for non-causal reasons.

Recall that the coding of ethnic partitioning relies on the geospatial processing based on the settlement areas of individual ethnic groups compiled by Murdock (1959). A quick evaluation of this inferential threat, therefore, is to examine whether the main findings hold when we restrict the sample such that geographically small-sized groups are excluded from the analysis. The subsample-based robustness check provides additional credibility to the main findings if the results remain stable with the restricted subsamples, while deviating results would undermine the empirical claims.

Following the intuition, I replicate the main two-step analysis with subsamples dropping geographically small-sized groups from the dataset. First, Table D.I presents the total treatment effect estimates using subsamples excluding geographically small-sized ethnic groups defined by several threshold values for settlement areas (within host countries). Specifically, Models 1 to 4 subsequently drop the geographically small-sized observations from the full sample ($N_{\text{obs}} = 10,246$) varying the threshold values from 100 km^2 to 1,000 km^2 and then rerun the treatment effect estimate with the model specification of Model 2 (full model) in Table I in the main text. Second, Figures D.1 and D.2 replicate the mediation analysis dropping the ethnic groups with geographic sizes of settlement areas smaller than 1,000 km^2 (i.e., using the subsample in Model 4 in Table D.I). The subsample estimates discard roughly 5% (geographically small) group-year observations.

The treatment effect estimates in the subsample settings generally remain stable, suggesting that potential measurement error is not likely to invite serious bias to our empirical analysis. The total treatment effect or ATE estimates with different subsamples in Table D.I yielded coefficient estimates close to the baseline estimate of 0.0082 reported in the main text. Figures D.1 and D.2 also indicate that the subsample mediation analysis produced a similar size-dependent effect of ethnic partitioning on contemporary conflicts. The main finding, or the conditioning and causal interaction effects induced by contemporary groups size, remains almost intact in the subsample estimations.

E Ethnic Partition, Political Discrimination, and Conflict Onset

The main results highlight the role of contemporary groups size as the major mediator in the causal channel linking ethnic partition and postcolonial civil conflict, while yielding little support for the discrimination mechanism. Given the increasing scholarly interests in group-level political status and its consequences (e.g., Cederman et al., 2015, 2010; Francois

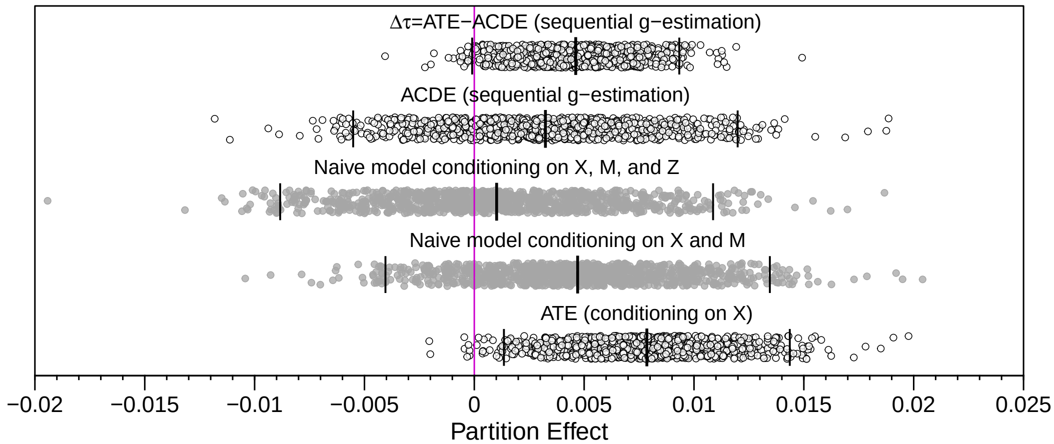
Table D.I. Total Effect of Ethnic Partition for Subsamples

	Dependent variable: Onset			
	Subsample:			
	(1) > 100 km^2	(2) > 300 km^2	(3) > 500 km^2	(4) > 1000 km^2
Partition	0.0081***	0.0080***	0.0085***	0.0082***
	(0.0028)	(0.0028)	(0.0028)	(0.0029)
Pretreatment Covariates	✓	✓	✓	✓
Country FEs	✓	✓	✓	✓
Calender year polynomial	✓	✓	✓	✓
Peace year polynomial	✓	✓	✓	✓
Spatial polynomial (lon, lat)	✓	✓	✓	✓
Observations	10,157	10,085	9,956	9,784
Adjusted R^2	0.0172	0.0171	0.0172	0.0165
Residual Std. Error	0.1283	0.1284	0.1292	0.1288
	(df = 10,104)	(df = 10,032)	(df = 9,904)	(df = 9,732)

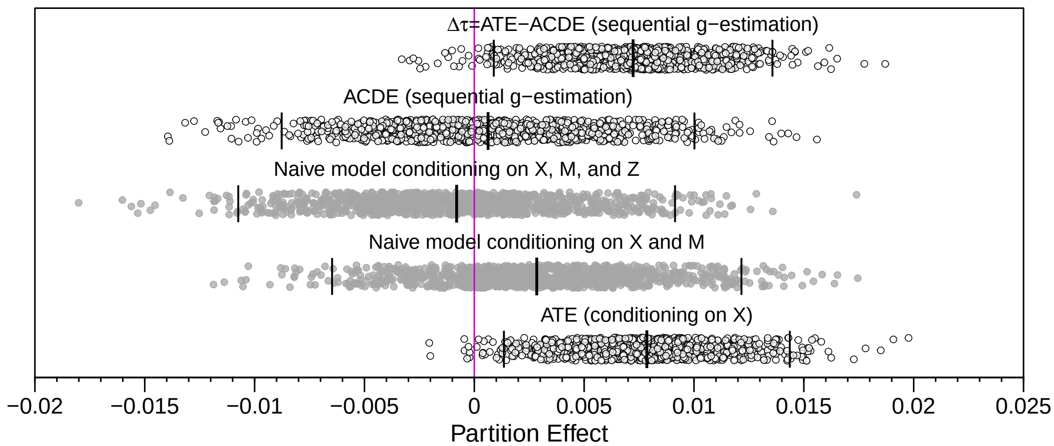
Note: $^{*}p$ <0.1; $^{**}p$ <0.05; $^{***}p$ <0.01. Constants not reported for brevity. Robust standard errors with multiway clustering on ethnic groups, countries, and years in parentheses.

et al., 2015; Wucherpfennig et al., 2016), the discrimination-conflict nexus warrants further investigation. Such an analysis also serves as additional investigation into the discrimination mechanism in Hypothesis 2. For this purpose, Table E.I reports a series of linear probability models with group-level conflict onset as the dependent variable. Models 1 to 4 each regress *Onset* on *Discrimination*, with and without *Partition* as well as potential confounders.

The regression results provide further support for the empirical claims by highlighting that the discrimination channel is not likely to drive the partition-conflict association. The results can be summarized as follows. First, consistent with previous studies, Models 1 to 4 underscore the conflict-escalating effect of *Discrimination*. Across model specifications, *Discrimination* is positively signed and retains the statistical significance at the conventional 5% level. The estimates broadly confirm the earlier findings of Cederman et al. (2010) and underline the conflict-escalating effect of ethnic discrimination. Second, the coefficient sizes of *Discrimination* reported in Table E.I suggest that pretreatment covariates such as geographic factors, rather than ethnic partition, are of more importance in determining the effect of *Discrimination* on conflict initiation. Indeed, the coefficient on *Discrimination* increases by roughly 19.5%, or from 0.041 (Model 1) to 0.049 (Model 2), when adjusting for the pretreatment covariates, while the point estimate remains stable regardless of the inclusion of *Partition* (Models 2 and 3). Finally, Model 4 reveals that the effect of *Discrimination* also remains stable regardless of the further adjustment for posttreatment variables potentially affected by *Partition*. Collectively, these results provide little empirical support for



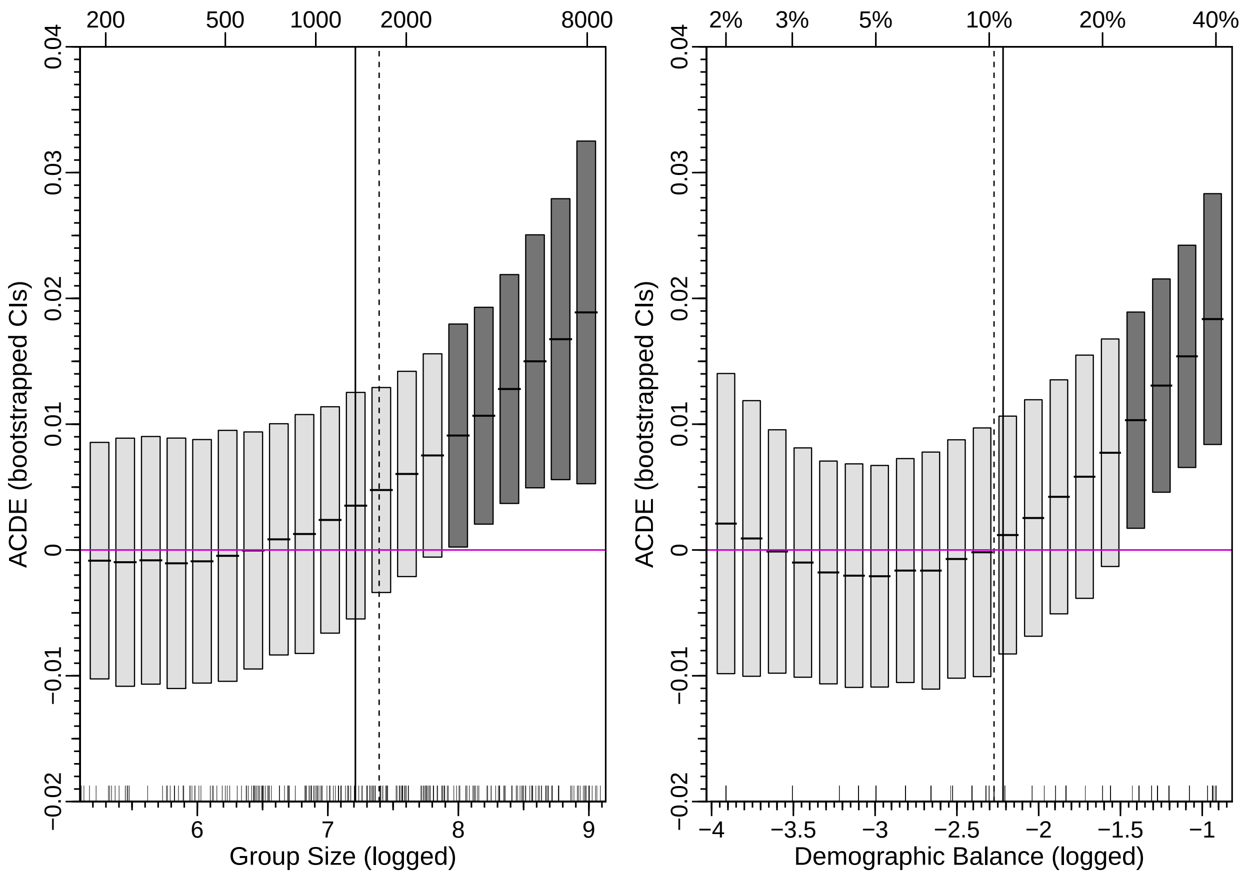
(a) *Group Size*



(b) *Demographic Balance*

Figure D.1. Effect of *Partition* on Conflict Onset with the Mediator Recentered to Mean with the Restricted Subsample ($1,000\,\mathrm{km}^2$ threshold)
Note: See notes in Figure C.1

the 'Ethnic partition → Political discrimination → Conflict' pathway, or the discrimination mechanism while illuminating that contemporary political discrimination is an important determinant of postcolonial conflicts.



(a) ACDE across *Group Size* (b) ACDE across *Demographic Balance*

Figure D.2. Controlled Direct Effect of *Partition* across Different Demographic Sizes with the Restricted Subsample ($1,000\,\mathrm{km}^2$ threshold)

See notes in Figure C.2.

Table E.I. Political Discrimination and Conflict Onset

	Dependent variable: Onset			
	(1) Restricted LPM	(2) LPM w/ X	(3) LPM w/ D and X	(4) LPM w/ D, X, and Z
Discrimination	0.0408*	0.0490**	0.0489**	0.0475**
	(0.0228)	(0.0215)	(0.0211)	(0.0211)
Partition			✓	✓
Pretreatment Covariates		✓	✓	✓
Posttreatment Variables				✓
Country FEs	✓	✓	✓	✓
Calender year polynomial	✓	✓	✓	✓
Peace year polynomial	✓	✓	✓	✓
Spatial polynomial (lon, lat)	✓	✓	✓	✓
Observations	10,246	10,246	10,246	10,246
Adjusted R^2	0.0182	0.0222	0.0226	0.0233
Residual Std. Error	0.1284	0.1281	0.1281	0.1281
	(df = 10,203)	(df = 10,193)	(df = 10,192)	(df = 10,182)

Note: *p <0.1; **p <0.05; ***p <0.01. Constants not reported for brevity. Robust standard errors with multiway clustering on ethnic groups, countries, and years in parentheses.

References

Berger, Susanne; Nathaniel Graham & Achim Zeileis (2017) Various versatile variances: An object-oriented implementation of clustered covariances in R. Working Paper 2017-12 Working Papers in Economics and Statistics, Research Platform Empirical and Experimental Economics, Universität Innsbruck.

Besley, Timothy & Marta Reynal-Querol (2014) The Legacy of Historical Conflict: Evidence from Africa. *American Political Science Review* 108(2): 319–336.

Bivand, Roger S; Edzer Pebesma & Virgilio Gomez-Rubio (2013) *Applied spatial data analysis with R* volume 2nd edition. New York: Springer.

Cameron, Colin A; Jonah B Gelbach & Douglas L Miller (2011) Robust inference with multiway clustering. *Journal of Business and Economic Statistics* 29(2): 238–249.

Cederman, L.-E; Simon Hug; Andreas Schädel & Julian Wucherpfennig (2015) Territorial Autonomy in the Shadow of Conflict: Too Little, Too Late? *American Political Science Review* 109(2): 354–370.

Cederman, L.-E; Andreas Wimmer & Brian Min (2010) Why Do Ethnic Groups Rebel? New Data and Analysis. *World Politics* 62(1): 87–119.

Ding, Peng & Tyler J VanderWeele (2016) Sensitivity analysis without assumptions. *Epidemiology* 27(3): 368–377.

Francois, Patrick; Ilia Rainer & Francesco Trebbi (2015) How Is Power Shared in Africa? *Econometrica* 83(2): 465–503.

Green, Elliott (2012) On the Size and Shape of African States. *International Studies Quarterly* 56(2): 229–244.

Ho, Daniel E; Kosuke Imai; Gary King & Elizabeth A Stuart (2007) Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis* 15(3): 199–236.

Michalopoulos, Stelios & Elias Papaioannou (2013) Pre-Colonial Ethnic Institutions and Contemporary African Development. *Econometrica* 81(1): 113–152.

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

Murdock, George P (1959) *Africa: Its Peoples and their Culture History*. New York: McGraw-Hill.

Osafo-Kwaako, Philip & James A Robinson (2013) Political centralization in pre-colonial Africa. *Journal of Comparative Economics* 41(1): 6–21.

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

Pebesma, Edzer J & Roger S Bivand (2005) Classes and methods for spatial data in R. *R News* 5(2): 9–13.

Ray, Subhasish (2019) History and Ethnic Conflict: Does Precolonial Centralization Matter? *International Studies Quarterly* 63(2): 417–431.

VanderWeele, Tyler J & Peng Ding (2017) Sensitivity analysis in observational research: Introducing the E-Value. *Annals of Internal Medicine* 167(4): 268–274.

Weidmann, Nils B; Doreen Kuse & Kristian S Gleditsch (2010) The geography of the international system: The CShapes dataset. *International Interactions* 36(1): 86–106.

Wig, Tore (2016) Peace from the Past: Pre-colonial Political Institutions and Contemporary Ethnic Civil Wars in Africa. *Journal of Peace Research* 53(4): 509–524.

Wucherpfennig, Julian; Philipp Hunziker & L.-E Cederman (2016) Who Inherits the State? Colonial Rule and Postcolonial Conflict. *American Journal of Political Science* 60(4): 882–898.

Zeileis, Achim (2004) Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software* 11(10).