Supplement to Diversity and Conflict

Supplement A Supplement to the Country-Level Analyses

A.1 Robustness Checks for the Analysis of Civil Conflict in Cross-Country Data

In this appendix section, we present several robustness checks for our cross-country analysis of the influence of contemporary population diversity on the temporal frequency of civil conflict outbreaks in the post-1960 time horizon.

Robustness to Accounting for Ecological/Climatic Covariates A nascent interdisciplinary literature (e.g., Burke et al., 2009; Hsiang et al., 2013; Burke et al., 2015) has emphasized the role of climatic factors, like temperature and precipitation, as important correlates of the risk of civil conflict. Further, Fenske (2014) shows that ecological diversity facilitated state centralization in pre-colonial Africa. To prevent our main specifications from becoming too unwieldy, we chose to exclude the aforementioned climatic and ecological variables from our baseline set of covariates, especially because this set already included a sizable vector of geographical factors that are known to be correlated with the former. In Table SA.I, however, we establish that population diversity remains a significant predictor of civil conflicts when we augment our baseline set of covariates in Table I with controls for (i) time-invariant fractionalization and polarization measures of the ecological diversity of land (e.g., Fenske, 2014); and (ii) the temporal mean and volatility of climatic experience (e.g., Burke et al., 2015) with respect to annual temperature and annual precipitation over the post-1960 time period.

Robustness to Accounting for Deep-Rooted Determinants of Economic Development In Table SA.II, we establish the robustness of our baseline cross-country analysis of civil conflict to *additionally* accounting for the potentially confounding influence of other deep-rooted determinants of comparative economic development. Specifically, we augment the analysis in Table I with controls for (i) the time elapsed since the onset of the Neolithic Revolution (e.g., Ashraf and Galor, 2013a); (ii) an index of experience with institutionalized statehood since antiquity (e.g., Bockstette et al., 2002); (iii) the time elapsed since initial human settlement in prehistory (e.g., Ahlerup and Olsson, 2012); and (iv) the great-circle distance to the closest regional technological frontier in the year 1500 (e.g., Ashraf and Galor, 2013a). The results indicate that regardless of the estimation sample or the specification, contemporary population diversity remains a significant predictor of the annual frequency of civil conflict onsets.

Robustness to Accounting for Ethnic and Spatial Inequality In Table SA.III, we check the robustness of our findings from Table I to *additionally* accounting for intra-country economic inequality (e.g., Alesina et al., 2016), as captured by the subnational spatial distribution of per-capita adjusted nighttime luminosity in the year 2000 across either (i) the georeferenced homelands of ethnic groups (ethnic inequality); or (ii) 2.5×2.5-degree geospatial grid cells (spatial inequality). The two inequality measures enter these regressions with a positive coefficient, and in at least one case, the coefficient on ethnic inequality is statistically significant. Nonetheless, our results indicate that the positive and significant influence of population diversity on the annual frequency of civil conflicts cannot be attributed to the potentially confounding influence of these inequality measures.

Robustness to Using Alternative Measures of Ethnolinguistic Fragmentation Due to the sizable cross-country correlation between the ethnic and linguistic fractionalization measures of Alesina et al. (2003), rather than exploiting both variables simultaneously, we chose to employ the more widely used of the two indices – namely, ethnic fractionalization – as one of the many

covariates in our baseline analysis of the influence of population diversity on civil conflict frequency. In Table SA.IV, we examine the sensitivity of our baseline findings from Table I to employing the *linguistic* fractionalization index of Alesina et al. (2003) in lieu of our baseline control for the *ethnic* fractionalization index from the same source. Furthermore, in Table SA.V, we examine the robustness of our baseline findings to employing the country-level counterparts of our measures of linguistic fractionalization and polarization from our analysis of conflicts at the ethnic homelands level. Specifically, these measures are constructed using georeferenced information on the spatial distribution of language homelands (from the World Language Mapping System [WLMS]) in combination with gridded population data, and they enter our regressions in Table SA.V in liue of our baseline controls for ethnic fractionalization from Alesina et al. (2003) and ethnolinguistic polarization from Desmet et al. (2012). Reassuringly, the results in Tables SA.IV–SA.V confirm that all our baseline findings regarding the significant influence of population diversity on the temporal frequency of civil conflict onsets remain qualitatively intact under these alternative controls for ethnolinguistic fragmentation.

Robustness to Using Initial Values of Time-Varying Covariates In Table SA.VI, we exploit the initial or year-1960 values of the time-dependent baseline controls employed by our analysis in Table I (i.e., the degree of executive constraints, indicators for democracy and autocracy, total population, and GDP per capita), rather than their respective temporal averages over the 1960–2017 time period. This robustness check is intended to examine whether our baseline estimates of the influence of population diversity in Table I could be explained away by the fact that the temporal averages of our time-varying controls over the entire sample period are likely to be more endogenous to the frequency of civil conflict onsets over the same period. Reassuringly, population diversity continues to remain a significant predictor of conflict frequency in these alternative specifications.

Robustness to Accounting for Spatial Autocorrelation in Errors As with any analysis that exploits spatial variations in cross-sectional data, autocorrelation in disturbance terms across observations could be biasing our estimates of the standard errors in our baseline cross-country regressions of conflict frequency. Table SA.VII therefore reports, for our key specifications from Table I, standard errors that are corrected for cross-sectional spatial dependence, using the methodology proposed by Conley (1999). To perform this robustness check, the spatial distribution of observations is specified on the Euclidean plane using the full set of pairwise geodesic distances between country centroids, and the spatial autoregressive process across residuals is modeled as varying inversely with distance from each observation up to a maximum threshold of 25,000 kilometers, thus admitting the possibility of spatial dependence at a global scale. The GMM specifications in this table correspond to the 2SLS specifications from Table I. Reassuringly, depending on the specification, the corrected standard errors of the estimated coefficient on population diversity are either similar in magnitude or noticeably smaller when compared to their heteroskedasticity robust counterparts from our baseline analysis.

Robustness to the Elimination of Regions from the Estimation Sample Following the norm in cross-country empirical studies of civil conflict, we investigate whether our main findings are driven by potentially influential world regions. The analysis in Table SA.VIII checks the qualitative robustness of the results associated with our fully specified empirical models in Columns 8 and 12 of Table I, eliminating one-at-a-time the following world regions from our global sample of countries: Sub-Saharan Africa (SSA), Middle East and North Africa (MENA), East Asia and Pacific (EAP), and Latin America and the Caribbean (LAC). Due to the lower degrees of freedom afforded by the regression samples with eliminated regions, the analysis omits continent dummies from the empirical models in order to preserve as much of the cross-country variation in conflict frequency

as possible. The findings reassuringly reveal that the significant influence of population diversity on conflict frequency is not qualitatively sensitive to the exclusion of any one of these potentially influential world region from our full estimation sample.

TABLE SA.I: Population Diversity and the Frequency of Civil Conflict Onset across Countries – Robustness to Accounting for Ecological/Climatic Covariates

Cross-country sample:	Global					Old World		Global	
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) 2SLS	(9) 2SLS
	Log number of new PRIO25 civil conflict onsets per year, 1960–2017								
Population diversity (ancestry adjusted)	0.209***	0.409***	0.306**	0.313**	0.290**	0.558**	0.636**	0.577***	0.703***
	(0.066)	(0.104)	(0.119)	(0.126)	(0.132)	(0.247)	(0.248)	(0.206)	(0.217)
Ecological fractionalization		−0.004	−0.001	−0.003	−0.003	0.001	0.003	−0.004	−0.010
		(0.016)	(0.017)	(0.017)	(0.020)	(0.021)	(0.024)	(0.016)	(0.018)
Ecological polarization		0.028	0.027	0.028	0.005	0.028	−0.002	0.030*	0.007
		(0.017)	(0.018)	(0.018)	(0.020)	(0.021)	(0.023)	(0.017)	(0.017)
Annual temperature, 1960–2016 average		0.002*	0.001	0.001	0.000	0.002	−0.001	0.002*	0.000
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Annual precipitation, 1960–2016 average		0.010	0.006	0.005	−0.001	0.018**	0.006	0.011*	0.004
		(0.006)	(0.006)	(0.006)	(0.006)	(0.009)	(0.009)	(0.006)	(0.006)
Volatility of annual temperature, 1960–2016		0.029	0.016	0.010	−0.003	0.007	−0.019	0.012	−0.013
		(0.024)	(0.024)	(0.022)	(0.023)	(0.029)	(0.026)	(0.023)	(0.021)
Volatility of annual precipitation, 1960–2016		−0.081*	−0.057	−0.054	−0.021	−0.143*	−0.067	−0.053	−0.011
		(0.043)	(0.042)	(0.041)	(0.046)	(0.085)	(0.089)	(0.045)	(0.052)
Continent dummies			×	×	×	×	×	×	×
Controls for geography		×	×	×	×	×	×	×	×
Controls for ethnic diversity				×	×		×		×
Controls for institutions					×		×		×
Controls for oil, population, and income					×		×		×
Observations	150	150	150	150	147	123	121	150	147
Partial R^2 of population diversity		0.090	0.038	0.039	0.038	0.049	0.062		
Adjusted R^2	0.029	0.208	0.213	0.210	0.327	0.221	0.360		
Effect of 10th–90th %ile move in diversity	0.014***	0.027***	0.020**	0.021**	0.020**	0.027**	0.027**	0.038***	0.048***
	(0.004)	(0.007)	(0.008)	(0.008)	(0.009)	(0.012)	(0.011)	(0.014)	(0.015)
First-stage F statistic								93.172	63.364

Notes: This table conducts a robustness check on the results from the baseline cross-country analysis of the reduced-form impact of contemporary population diversity on the annual frequency of civil conflict onsets, as shown in Table I. Specifically, it establishes robustness to *additionally* accounting for the potentially confounding influence of (i) time-invariant fractionalization and polarization measures of the ecological diversity of land (e.g., Fenske, 2014); and (ii) the temporal mean and volatility of climatic experience (e.g., Burke et al., 2015) with respect to annual temperature and annual precipitation over the post-1960 time period. The specifications examined in this table are otherwise identical to corresponding ones reported in Table I. The reader is therefore referred to Table I and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis as well as the identification strategy employed by the 2SLS regressions. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the number of new conflict onsets per year. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.II: Population Diversity and the Frequency of Civil Conflict Onset across Countries – Robustness to Accounting for Deep-Rooted Determinants of Economic Development

Cross-country sample:	Global					Old World		Global	
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) 2SLS	(9) 2SLS
	Log number of new PRIO25 civil conflict onsets per year, 1960–2017								
Population diversity (ancestry adjusted)	0.228***	0.378***	0.315***	0.316***	0.325**	0.547**	0.664**	0.498***	0.603***
	(0.070)	(0.103)	(0.112)	(0.116)	(0.140)	(0.266)	(0.275)	(0.192)	(0.203)
Log years since Neolithic Revolution		0.008*	0.011**	0.010*	0.008	0.004	−0.001	0.010*	0.008
		(0.004)	(0.005)	(0.005)	(0.006)	(0.010)	(0.011)	(0.005)	(0.006)
Log index of state antiquity		0.007**	0.008**	0.008**	0.004	0.008*	0.001	0.008**	0.005
		(0.003)	(0.004)	(0.004)	(0.005)	(0.004)	(0.006)	(0.003)	(0.005)
Log duration of human settlement		0.005**	0.001	0.001	0.003	0.003	0.009*	0.000	0.002
		(0.002)	(0.003)	(0.003)	(0.003)	(0.004)	(0.005)	(0.003)	(0.003)
Log distance from regional frontier in 1500		0.002	0.002	0.002	0.001	0.003	0.002	0.002	0.001
		(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Continent dummies			×	×	×	×	×	×	×
Controls for geography		×	×	×	×	×	×	×	×
Controls for ethnic diversity				×	×		×		×
Controls for institutions					×		×		×
Controls for oil, population, and income					×		×		×
Observations	136	136	136	136	135	110	109	136	135
Partial R^2 of population diversity		0.085	0.046	0.044	0.054	0.044	0.077		
Adjusted R^2	0.034	0.228	0.220	0.218	0.350	0.215	0.401		
Effect of 10th–90th %ile move in diversity	0.016***	0.026***	0.022***	0.022***	0.022**	0.026**	0.033**	0.034***	0.041***
	(0.005)	(0.007)	(0.008)	(0.008)	(0.010)	(0.013)	(0.014)	(0.013)	(0.014)
First-stage F statistic								69.283	52.108

Notes: This table conducts a robustness check on the results from the baseline cross-country analysis of the reduced-form impact of contemporary population diversity on the annual frequency of civil conflict onsets, as shown in Table I. Specifically, it establishes robustness to *additionally* accounting for the potentially confounding influence of other deep-rooted determinants of comparative economic development, including (i) the time elapsed since the onset of the Neolithic Revolution (e.g., Ashraf and Galor, 2013a); (ii) an index of experience with institutionalized statehood since antiquity (e.g., Bockstette et al., 2002); (iii) the time elapsed since initial human settlement in prehistory (e.g., Ahlerup and Olsson, 2012); and (iv) the great-circle distance to the closest regional technological frontier in the year 1500 (e.g., Ashraf and Galor, 2013a). The specifications examined in this table are otherwise identical to corresponding ones reported in Table I. The reader is therefore referred to Table I and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis as well as the identification strategy employed by the 2SLS regressions. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the number of new conflict onsets per year. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.III: Population Diversity and the Frequency of Civil Conflict Onset across Countries – Robustness to Accounting for Ethnic and Spatial Inequality

Cross-country sample:	Global					Old World		Global	
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) 2SLS	(9) 2SLS
	Log number of new PRIO25 civil conflict onsets per year, 1960–2017								
Population diversity (ancestry adjusted)	0.214***	0.443***	0.338***	0.353***	0.337**	0.665***	0.760***	0.674***	0.747***
	(0.066)	(0.108)	(0.123)	(0.127)	(0.132)	(0.211)	(0.213)	(0.197)	(0.188)
Ethnic inequality in luminosity		0.021	0.020	0.018	0.013	0.023	0.022	0.024*	0.018
		(0.014)	(0.014)	(0.015)	(0.017)	(0.017)	(0.018)	(0.014)	(0.015)
Spatial inequality in luminosity		0.004	0.014	0.015	0.013	0.021	0.019	0.018	0.014
		(0.017)	(0.017)	(0.018)	(0.015)	(0.021)	(0.018)	(0.016)	(0.014)
Continent dummies			×	×	×	×	×	×	×
Controls for geography		×	×	×	×	×	×	×	×
Controls for ethnic diversity				×	×		×		×
Controls for institutions					×		×		×
Controls for oil, population, and income					×		×		×
Observations	147	147	147	147	145	120	119	147	145
Partial R^2 of population diversity		0.132	0.054	0.056	0.062	0.094	0.139		
Adjusted R^2	0.032	0.181	0.211	0.209	0.359	0.235	0.424		
Effect of 10th–90th %ile move in diversity	0.015***	0.030***	0.023***	0.024***	0.023**	0.028***	0.033***	0.046***	0.051***
	(0.004)	(0.007)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.013)	(0.013)
First-stage F statistic								133.897	80.495

Notes: This table conducts a robustness check on the results from the baseline cross-country analysis of the reduced-form impact of contemporary population diversity on the annual frequency of civil conflict onsets, as shown in Table I. Specifically, it establishes robustness to *additionally* accounting for the potentially confounding influence of measures of intra-country economic inequality (e.g., Alesina et al., 2016), as captured by the subnational spatial distribution of per-capita adjusted nighttime luminosity in the year 2000 across either (i) the georeferenced homelands of ethnic groups (ethnic inequality); or (ii) 2.5×2.5-degree geospatial grid cells (spatial inequality). The specifications examined in this table are otherwise identical to corresponding ones reported in Table I. The reader is therefore referred to Table I and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis as well as the identification strategy employed by the 2SLS regressions. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the number of new conflict onsets per year. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.IV: Population Diversity and the Frequency of Civil Conflict Onset across Countries – The Analysis under Linguistic Fractionalization

Cross-country sample:	Global					Old World		Global	
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) 2SLS	(9) 2SLS
	Log number of new PRIO25 civil conflict onsets per year, 1960–2017								
Population diversity (ancestry adjusted)	0.218***	0.470***	0.338***	0.357***	0.332**	0.545***	0.605***	0.554***	0.603***
	(0.069)	(0.109)	(0.125)	(0.125)	(0.136)	(0.193)	(0.211)	(0.182)	(0.190)
Linguistic fractionalization				0.011	0.005		0.010		0.005
				(0.012)	(0.009)		(0.011)		(0.009)
Ethnolinguistic polarization				0.014	0.012		0.013		0.016
				(0.013)	(0.012)		(0.014)		(0.012)
Continent dummies			×	×	×	×	×	×	×
Controls for geography		×	×	×	×	×	×	×	×
Controls for institutions					×		×		×
Controls for oil, population, and income					×		×		×
Observations	146	146	146	146	143	122	120	146	143
Partial R^2 of population diversity		0.138	0.049	0.056	0.057	0.068	0.092		
Adjusted R^2	0.031	0.196	0.217	0.227	0.372	0.226	0.407		
Effect of 10th–90th %ile move in diversity	0.014***	0.031***	0.022***	0.023***	0.022**	0.025***	0.027***	0.036***	0.039***
	(0.004)	(0.007)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.012)	(0.012)
First-stage F statistic								163.933	100.133

Notes: This table conducts a robustness check on the results from the baseline cross-country analysis of the reduced-form impact of contemporary population diversity on the annual frequency of civil conflict onsets, as shown in Table I. Specifically, it establishes robustness to accounting for the potentially confounding influence of linguistic rather than ethnic fractionalization (e.g., Alesina et al., 2003), as a baseline control for subnational intergroup cultural fragmentation. The specifications examined in this table are otherwise identical to corresponding ones reported in Table I. The reader is therefore referred to Table I and the corresponding table notes for additional details on the other baseline covariates considered by the current analysis as well as the identification strategy employed by the 2SLS regressions. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the number of new conflict onsets per year. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.V: Population Diversity and the Frequency of Civil Conflict Onset across Countries – The Analysis under Georeferenced Linguistic Fractionalization and Polarization

Cross-country sample:	Global					Old World		Global	
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) 2SLS	(9) 2SLS
	Log number of new PRIO25 civil conflict onsets per year, 1960–2017								
Population diversity (ancestry adjusted)	0.212***	0.443***	0.315***	0.326***	0.285**	0.556***	0.578***	0.543***	0.556***
	(0.066)	(0.103)	(0.115)	(0.118)	(0.123)	(0.191)	(0.210)	(0.176)	(0.182)
Linguistic fractionalization (georeferenced)				0.002	−0.006		−0.008		−0.002
				(0.011)	(0.010)		(0.012)		(0.010)
Linguistic polarization (georeferenced)				0.006	0.008		0.010		0.009
				(0.012)	(0.010)		(0.010)		(0.009)
Continent dummies			×	×	×	×	×	×	×
Controls for geography		×	×	×	×	×	×	×	×
Controls for institutions					×		×		×
Controls for oil, population, and income					×		×		×
Observations	151	151	151	151	148	124	122	151	148
Partial R^2 of population diversity		0.129	0.047	0.049	0.046	0.070	0.083		
Adjusted R^2	0.030	0.188	0.214	0.206	0.359	0.226	0.389		
Effect of 10th–90th %ile move in diversity	0.014***	0.029***	0.021***	0.021***	0.019**	0.027***	0.025***	0.035***	0.038***
	(0.004)	(0.007)	(0.007)	(0.008)	(0.008)	(0.009)	(0.009)	(0.011)	(0.012)
First-stage F statistic								157.089	98.473

Notes: This table conducts a robustness check on the results from the baseline cross-country analysis of the reduced-form impact of contemporary population diversity on the annual frequency of civil conflict onsets, as shown in Table I. Specifically, it establishes robustness to accounting for the potentially confounding influence of linguistic fractionalization and polarization, constructed using georeferenced information on the spatial distribution of language homelands (from the World Language Mapping System [WLMS]) in combination with gridded population data, rather than ethnic fractionalization (e.g., Alesina et al., 2003) and ethnolinguistic polarization (e.g., Desmet et al., 2012), as baseline controls for subnational intergroup cultural fragmentation. The specifications examined in this table are otherwise identical to corresponding ones reported in Table I. The reader is therefore referred to Table I and the corresponding table notes for additional details on the other baseline covariates considered by the current analysis as well as the identification strategy employed by the 2SLS regressions. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the number of new conflict onsets per year. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.VI: Population Diversity and the Frequency of Civil Conflict Onset across Countries – The Analysis under Initial Values of Time-Varying Covariates

Cross-country sample:	Global					Old World		Global	
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) 2SLS	(9) 2SLS
	Log number of new PRIO25 civil conflict onsets per year, 1960–2017								
Population diversity (ancestry adjusted)	0.209***	0.439***	0.306***	0.318***	0.366***	0.548***	0.734***	0.537***	0.693***
	(0.066)	(0.104)	(0.115)	(0.119)	(0.136)	(0.191)	(0.215)	(0.176)	(0.192)
Executive constraints in initial year					0.004		0.003		0.005**
					(0.002)		(0.003)		(0.002)
Democracy score in initial year					−0.002		−0.002		−0.003**
					(0.002)		(0.002)		(0.002)
Autocracy score in initial year					−0.001		−0.000		−0.001
					(0.001)		(0.002)		(0.001)
Log population in initial year					0.005*		0.007**		0.004*
					(0.003)		(0.003)		(0.002)
Log GDP per capita in initial year					−0.004*		−0.004*		−0.005**
					(0.002)		(0.002)		(0.002)
Continent dummies			×	×	×	×	×	×	×
Controls for geography		×	×	×	×	×	×	×	×
Controls for ethnic diversity				×	×		×		×
Controls for legal origin and colonial history					×		×		×
Control for oil or gas reserve discovery					×		×		×
Observations	150	150	150	150	145	123	119	150	145
Partial R^2 of population diversity		0.128	0.044	0.046	0.063	0.068	0.118		
Adjusted R^2	0.029	0.189	0.213	0.215	0.276	0.225	0.339		
Effect of 10th–90th %ile move in diversity	0.014***	0.029***	0.020***	0.021***	0.025***	0.026***	0.031***	0.036***	0.047***
	(0.004)	(0.007)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.012)	(0.013)
First-stage F statistic								153.543	81.221

Notes: This table conducts a robustness check on the results from the baseline cross-country analysis of the reduced-form impact of contemporary population diversity on the annual frequency of civil conflict onsets, as shown in Table I. Specifically, it establishes robustness to considering the initial or year-1960 values of the time-dependent baseline controls for institutions (i.e., the degree of executive constraints and indicators for democracy and autocracy), total population, and GDP per capita, rather than their respective temporal averages over the 1960–2017 time period. The methodology exploited by the current analysis aims to reduce any ex ante bias in the baseline estimates of the influence of population diversity, arising from the fact that the temporal averages of the aforementioned time-varying controls may well vary more endogenously across countries with the contemporaneous measure of civil conflict onsets. In order to maintain a cross-country sample that as consistent as possible with the baseline analysis, observations of the time-dependent covariates from the earliest available year after 1960 are used for the subset of countries with missing 1960 data. The specifications examined in this table are otherwise identical to corresponding ones reported in Table I. The reader is therefore referred to Table I and the corresponding table notes for additional details on the other baseline covariates considered by the current analysis as well as the identification strategy employed by the 2SLS regressions. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the number of new conflict onsets per year. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.VII: Population Diversity and the Frequency of Civil Conflict Onset across Countries – Robustness to Accounting for Spatial Autocorrelation in Errors

Cross-country sample:	Global					Old World		Global	
	(1) Conley OLS	(2) Conley OLS	(3) Conley OLS	(4) Conley OLS	(5) Conley OLS	(6) Conley OLS	(7) Conley OLS	(8) Conley GMM	(9) Conley GMM
	Log number of new PRIO25 civil conflict onsets per year, 1960–2017								
Population diversity (ancestry adjusted)	0.209***	0.439***	0.306***	0.318***	0.309***	0.548***	0.597***	0.537***	0.602***
	(0.036)	(0.068)	(0.117)	(0.110)	(0.111)	(0.076)	(0.076)	(0.084)	(0.085)
Continent dummies			×	×	×	×	×	×	×
Controls for geography		×	×	×	×	×	×	×	×
Controls for ethnic diversity				×	×		×		×
Controls for institutions					×		×		×
Controls for oil, population, and income					×		×		×
Observations	150	150	150	150	147	123	121	150	147
Adjusted R^2	0.364	0.468	0.484	0.485	0.582	0.512	0.619		

Notes: This table conducts a robustness check on the results from the baseline cross-country analysis of the reduced-form impact of contemporary population diversity on the annual frequency of civil conflict onsets, as shown in Table I. Specifically, it establishes robustness of the standard-error estimates to accounting for spatial dependence across observations, following the methodology of Conley (1999). To perform this robustness check, the spatial distribution of observations is specified on the Euclidean plane using the full set of pairwise geodesic distances between country centroids, and the spatial autoregressive process across residuals is modeled as varying inversely with distance from each observation up to a maximum threshold of 25,000 kilometers, thus admitting the possibility of spatial dependence at a global scale. The GMM specifications in this table correspond to the 2SLS specifications from Table I, exploiting prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary population diversity. The specifications examined in this table are otherwise identical to corresponding ones reported in Table I. The reader is therefore referred to Table I and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis. Standard errors, corrected for spatial autocorrelation, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.VIII: Population Diversity and the Frequency of Civil Conflict Onset across Countries – Robustness to the Elimination of Regions from the Global Sample

Omitted region:	None		SSA		MENA		EAP		LAC	
	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS	(5) OLS	(6) 2SLS	(7) OLS	(8) 2SLS	(9) OLS	(10) 2SLS
	Log number of new PRIO25 civil conflict onsets per year, 1960–2017									
Population diversity (ancestry adjusted)	0.344***	0.587***	0.411***	1.243***	0.368***	0.604***	0.310**	0.561***	0.385**	0.558***
	(0.115)	(0.178)	(0.139)	(0.379)	(0.128)	(0.187)	(0.124)	(0.193)	(0.161)	(0.204)
Controls for geography	×	×	×	×	×	×	×	×	×	×
Controls for ethnic diversity	×	×	×	×	×	×	×	×	×	×
Controls for institutions	×	×	×	×	×	×	×	×	×	×
Controls for oil, population, and income	×	×	×	×	×	×	×	×	×	×
Observations	147	147	105	105	131	131	132	132	126	126
Partial R^2 of population diversity	0.051		0.058		0.039		0.011		0.087	
Adjusted R^2	0.342		0.343		0.359		0.334		0.357	
Effect of 10th–90th %ile move in diversity	0.023***	0.040***	0.026***	0.077***	0.025***	0.041***	0.018**	0.033***	0.019**	0.027***
	(0.008)	(0.012)	(0.009)	(0.024)	(0.009)	(0.013)	(0.007)	(0.011)	(0.008)	(0.010)
First-stage F statistic		59.534		17.579		57.894		50.576		73.441

Notes: This table conducts a robustness check on the results associated with the fully specified empirical models in the baseline cross-country analysis of the reduced-form impact of contemporary population diversity on the annual frequency of civil conflict onsets, as shown in Columns 8 and 12 of Table I. Specifically, it establishes robustness to the one-at-a-time elimination of world regions from the global sample, including Sub-Saharan Africa (SSA), Middle East and North Africa (MENA), East Asia and Pacific (EAP), and Latin America and the Caribbean (LAC). Due to the lower degrees of freedom afforded by the regression samples with eliminated regions, the current analysis omits continent dummies from the empirical models in order to preserve as much of the cross-country variation in conflict as possible. The regressions in Columns 1–2 should therefore be viewed as the relevant baselines for assessing the robustness results presented in the remaining columns. The set of covariates, however, is otherwise identical to those reported in Columns 8 and 12 of Table I. The reader is therefore referred to Table I and the corresponding table notes for additional details on the set of covariates considered by the current analysis as well as the identification strategy employed by the 2SLS regressions. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the number of new conflict onsets per year. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.IX: Ethnic Fractionalization, Polarization, and the Frequency of Civil Conflict Onset across Countries

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) OLS	(9) OLS
	Log number of new PRIO25 civil conflict onsets per year, 1960–2017								
Ethnic fractionalization	0.024***	0.021*	0.016				0.022***	0.015	0.012
	(0.007)	(0.012)	(0.012)				(0.007)	(0.012)	(0.012)
Ethnolinguistic polarization				0.014	0.019*	0.012	0.007	0.014	0.008
				(0.008)	(0.010)	(0.010)	(0.009)	(0.010)	(0.010)
Continent dummies			×			×			×
Controls for geography		×	×		×	×		×	×
Observations	154	154	154	154	154	154	154	154	154
Adjusted R^2	0.037	0.095	0.182	0.006	0.096	0.180	0.034	0.098	0.179

Notes: This table examines the sensitivity of the association between ethnic fractionalization and ethnolinguistic polarization, on the one hand, and the annual frequency of new civil conflict onsets during the 1960–2017 time period, on the other, to controls for potentially confounding geographical characteristics and continent fixed effects. The controls for geography include absolute latitude, ruggedness, distance to the nearest waterway, the mean and range of agricultural suitability, the mean and range of elevation, and an indicator for small island nations. The set of continent dummies includes five indicators for Africa, Asia, North America, South America, and Oceania. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

A.2 Robustness Checks for the Analysis of Civil Conflict in Repeated Cross-Country Data

In this appendix section, we present several robustness checks for our analysis of the influence of contemporary population diversity on the quinquennial incidence or annual onset of civil conflict in repeated cross-country data for the post-1960 time horizon.

Robustness to Accounting for Ecological/Climatic Covariates A nascent interdisciplinary literature (e.g., Burke et al., 2009; Hsiang et al., 2013; Burke et al., 2015) has emphasized the role of climatic factors, like temperature and precipitation, as important correlates of the risk of civil conflict. Further, Fenske (2014) shows that ecological diversity facilitated state centralization in pre-colonial Africa. To prevent our main specifications from becoming too unwieldy, we chose to exclude the aforementioned climatic and ecological variables from our baseline set of covariates, especially because this set already included a sizable vector of geographical factors that are known to be correlated with the former. In Table SA.X, however, we establish that population diversity remains a significant predictor of both the quinquennial incidence (Columns 1–4) and the annual onset (Columns 5–8) of civil conflict when we augment our baseline set of covariates in Table II with controls for (i) time-invariant fractionalization and polarization measures of the ecological diversity of land (e.g., Fenske, 2014); and (ii) climatic experience in the recent past (e.g., Burke et al., 2015), as captured by either (a) the temporal mean and volatility of annual temperature and annual precipitation over the previous 5-year interval for the quinquennial incidence regressions; or (b) the lagged values of annual temperature and annual precipitation as well as their temporal volatility over the previous 5 years for the annual onset regressions.

Robustness to Accounting for Deep-Rooted Determinants of Economic Development The analysis in Table SA.XI establishes the robustness of our baseline results for the quinquennial incidence and annual onset of civil conflict in repeated cross-country data to *additionally* accounting for the potentially confounding influence of other deep-rooted determinants of comparative economic development. Specifically, we augment the analysis in Table II with controls for (i) the time elapsed since the onset of the Neolithic Revolution (e.g., Ashraf and Galor, 2013a); (ii) an index of experience with institutionalized statehood since antiquity (e.g., Bockstette et al., 2002); (iii) the time elapsed since initial human settlement in prehistory (e.g., Ahlerup and Olsson, 2012); and (iv) the great-circle distance to the closest regional technological frontier in the year 1500 (e.g., Ashraf and Galor, 2013a). The results indicate that regardless of the estimation sample or the specification, contemporary population diversity remains a significant predictor of both the quinquennial likelihood of a conflict incidence (Columns 1–4) and the annual likelihood of a conflict onset (Columns 5–8).

Robustness to Accounting for Ethnic and Spatial Inequality In Table SA.XII, we check the robustness of our findings from Table II to *additionally* accounting for intra-country economic inequality (e.g., Alesina et al., 2016), as captured by the subnational spatial distribution of per-capita adjusted nighttime luminosity in the year 2000 across either (i) the georeferenced homelands of ethnic groups (ethnic inequality); or (ii) 2.5×2.5-degree geospatial grid cells (spatial inequality). The two inequality measures enter these regressions with mostly positive but invariably insignificant coefficients. Thus, unsurprisingly, the positive and significant influence of population diversity on either the quinquennial incidence or the annual onset of civil conflict remains qualitatively unaffected.

Robustness to Accounting for Alternative Correlates of Conflict Incidence The analysis in Table SA.XIII checks the robustness of our baseline results for conflict incidence to controlling for the potentially confounding influence of *alternative* distributional indices of intergroup diversity

(e.g., Fearon, 2003; Alesina et al., 2003; Esteban et al., 2012) as well as *additional* geographical correlates of conflict (e.g., Fearon and Laitin, 2003; Cervellati et al., 2017). The specifications examined by this robustness analysis are identical to the fully specified baseline models reported in Columns 2 and 4 of Table II, with the exception that in Columns 1–3 and 6–8 of the current analysis, each of the reported control variables is employed in lieu of the baseline control for ethnic fractionalization (Alesina et al., 2003), whereas in Columns 4 and 9, the set of reported control variables replaces the baseline controls for both ethnic fractionalization and ethnolinguistic polarization (Desmet et al., 2012), in the interest of mitigating multicollinearity. Further, in Columns 5 and 10, the set of reported geographical controls augment our fully specified baseline models of conflict incidence. Among the additional controls considered, ethnolinguistic polarization (Esteban et al., 2012) and the geographical variables that capture the percentage of mountainous terrain and the presence of noncontiguous territories (Fearon and Laitin, 2003) enter the IV Probit regressions in the global sample of countries with positive and significant coefficients. Nevertheless, our baseline findings regarding the significant impact of population diversity on the quinquennial incidence of civil conflict remain qualitatively unaltered across all specifications.

Robustness to Employing the Classical Logit and Rare-Events Logit Estimators The analysis in Table SA.XIV establishes the robustness of our baseline results for the quinquennial incidence and annual onset of civil conflict in repeated cross-sectional data on countries from the Old World, as shown in Columns 1–2 and 5–6 of Table II, to employing the classical logit and rare-events logit (King and Zeng, 2001) estimators, rather than the standard probit estimator. Given the absence of readily available ordinary logit and rare-events logit estimators that permit instrumentation, the current analysis is unable to implement our global-sample identification strategy of exploiting prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary population diversity. As expected, the rare-events logit estimates in Table SA.XIV are somewhat smaller in absolute value than their counterparts under the classical logit estimator, due to bias arising in the latter estimates from ignoring the fact that civil conflict events (involving at least 25 battle-related deaths in a year) are generally rare occurrences in repeated cross-country data. Nonetheless, the findings attest to the robustness of the reduced-form influence of population diversity on either the quinquennial incidence or the annual onset of civil conflict under these alternative estimators.

Robustness to Accounting for Spatiotemporal Dependence using Two-Way Clustering of Standard Errors In Table SA.XV, we check the robustness of the results from our baseline probit and logit analyses of the quinquennial incidence or annual onset of civil conflict in repeated cross-sectional data on countries from the Old World, as shown in Columns 1–2 and 5–6 of Table II and in odd-numbered columns of Table SA.XIV, to accounting for spatiotemporal dependence across country-time observations. Specifically, we probe the statistical precision of our coefficient estimates by implementing multi-dimensional clustering of standard errors, following the methodology of Cameron et al. (2011). To implement this robustness check, the standard errors across country-time observations are clustered in two dimensions: (i) the country level, which allows for temporal dependence within a country over time (i.e., across either 5-year intervals or years); and (ii) the time level, which allows for spatial dependence across countries within a given time period (i.e., either a 5-year interval or a year). Given the absence of readily available probit and logit estimators that not only allow for multi-dimensional clustering of standard errors but also permit instrumentation, the current analysis is unable to implement the global-sample identification strategy of exploiting prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary population diversity. Reassuringly, the bi-dimensionally clustered standard errors of our coefficient of interest are either similar to

or, in the specifications for conflict incidence, noticeably smaller in magnitude than their classically estimated counterparts in Tables II and SA.XIV that do not admit spatiotemporal dependence across country-time observations.

Robustness to Accounting for Alternative Correlates of Conflict Onset In Table SA.XVI, we check the robustness of the results from our baseline analysis of the annual onset of civil conflict in repeated cross-country data, as shown in Columns 5–8 of Table II, to accounting for the potentially confounding influence of an *additional* time-invariant distributional index of intergroup diversity, capturing the degree of "ethnic dominance" (e.g., Collier and Hoeffler, 2004), and *additional* time-varying institutional correlates of conflict onset, capturing the lagged annual values of an index of political instability and an indicator for the emergence of a newly independent state from colonial powers (e.g., Fearon and Laitin, 2003). In light of constraints imposed by the availability of data on these additional control variables, the analysis is restricted to a smaller sample of countries and to the 1960–1999 (as opposed to the 1960–2017) time period. Therefore, the specification presented in each odd-numbered column of the table is intended to provide a relevant baseline for the robustness check in the subsequent even-numbered column (i.e., by holding fixed the regression sample). Turning to the results in Table SA.XVI, the lagged index of political instability does appear to enter some of our specifications with a positive and statistically significant coefficient, although the other additional controls considered by the analysis do not seem to be significantly correlated with conflict onset. However, despite the substantial reduction in both the sample time-frame and the number of countries in the cross-section, our coefficient of interest reassuringly remains positive and precisely estimated, regardless of the inclusion of these additional controls to the specifications.

Robustness to Accounting for Commodity Export Price Shocks The analysis in Table SA.XVII checks the robustness of our baseline results for the annual onset of civil conflict in repeated cross-country data, as shown in Columns 5–8 of Table II, to *additionally* accounting for the potentially confounding "income effect" of commodity export price shocks (e.g., Bazzi and Blattman, 2014), as captured by the contemporaneous, lagged, and twice lagged values of either an annual price shock that has been aggregated across commodity export types (Columns 1–2 and 5–6) or annual price shocks disaggregated by type of commodity export, including export price shocks associated with annual crops, perennial crops, and extractive crops (Columns 3–4 and 7–8). These export price shock variables are all obtained from the data set of Bazzi and Blattman (2014), so the reader is referred to that work for additional details on these variables. In light of constraints imposed by the availability of data on these additional covariates, the analysis is restricted to a smaller sample of countries and to the 1960–2007 (as opposed to the 1960–2017) time period. As is evident from the results in Table SA.XVII, there is indeed a significant mitigating "income effect" on the annual likelihood of a conflict onset associated with the contemporaneous and twice lagged values of commodity export price shocks (for both aggregated and disaggregated variants of these shocks). Nonetheless, despite the reduction in both the number of countries in the cross-section and the sample time-frame, our coefficient of interest reassuringly remains positive and statistically significant when subjected to these additional covariates in the specifications.

TABLE SA.X: Population Diversity and the Incidence or Onset of Civil Conflict in Repeated Cross-Country Data – Robustness to Accounting for Ecological/Climatic Covariates

Cross-country sample:	Old World		Global		Old World		Global	
	(1) Probit	(2) Probit	(3) IV Probit	(4) IV Probit	(5) Probit	(6) Probit	(7) IV Probit	(8) IV Probit
	Quinquennial PRIO25 civil conflict incidence, 1960–2017				Annual PRIO25 civil conflict onset, 1960–2017			
Population diversity (ancestry adjusted)	14.367***	10.178**	17.325***	15.651***	6.172*	6.001*	7.063**	9.482**
	(4.264)	(4.488)	(4.387)	(5.167)	(3.306)	(3.538)	(3.425)	(4.282)
Ecological fractionalization	−0.368	−0.080	−0.503	−0.394	0.018	−0.401	−0.027	−0.432
	(0.456)	(0.524)	(0.432)	(0.494)	(0.274)	(0.371)	(0.275)	(0.376)
Ecological polarization	0.865**	0.327	1.086***	0.927**	0.238	0.330	0.406	0.529
	(0.417)	(0.504)	(0.398)	(0.471)	(0.301)	(0.419)	(0.303)	(0.420)
Lagged temperature	0.078***	0.002	0.067***	0.023	0.033*	−0.004	0.032*	0.009
	(0.027)	(0.034)	(0.021)	(0.025)	(0.019)	(0.024)	(0.016)	(0.020)
Lagged precipitation	0.177	−0.042	0.248	0.148	0.096	−0.002	0.110	0.086
	(0.178)	(0.166)	(0.167)	(0.176)	(0.124)	(0.138)	(0.122)	(0.140)
Lagged temperature volatility	−0.576*	−0.416	−0.356	−0.274	0.307	0.249	0.218	0.239
	(0.342)	(0.382)	(0.307)	(0.332)	(0.287)	(0.281)	(0.272)	(0.263)
Lagged precipitation volatility	−1.326	−1.363	−0.504	−0.439	−0.282	−0.152	−0.566	−0.221
	(0.814)	(1.096)	(0.603)	(0.742)	(0.592)	(0.708)	(0.595)	(0.647)
Continent dummies	×	×	×	×	×	×	×	×
Time dummies	×	×	×	×	×	×	×	×
Controls for temporal spillovers	×	×	×	×	×	×	×	×
Controls for geography	×	×	×	×	×	×	×	×
Controls for ethnic diversity		×		×		×		×
Controls for institutions		×		×		×		×
Controls for oil, population, and income		×		×		×		×
Observations	1,270	1,045	1,583	1,311	5,452	4,377	6,996	5,757
Countries	123	121	150	147	123	121	150	147
Pseudo R^2	0.431	0.443			0.135	0.163		
Marginal effect of diversity	2.675***	1.873**	3.364***	2.981***	0.322*	0.312*	0.333*	0.454*
	(0.796)	(0.833)	(0.908)	(1.046)	(0.177)	(0.186)	(0.173)	(0.233)
First-stage F statistic			83.318	70.585			94.679	77.102

Notes: This table conducts a robustness check on the results from the baseline analysis of the reduced-form impact of contemporary population diversity on either the quinquennial incidence or the annual onset of civil conflict in repeated cross-country data, as shown in Table II. Specifically, it establishes robustness to *additionally* accounting for the potentially confounding influence of (i) time-invariant fractionalization and polarization measures of the ecological diversity of land (e.g., Fenske, 2014); and (ii) climatic experience in the recent past (e.g., Burke et al., 2015), as captured by either (a) the temporal mean and volatility of annual temperature and annual precipitation over the previous 5-year interval for the quinquennial incidence regressions; or (b) the lagged values of annual temperature and annual precipitation as well as their temporal volatility over the previous 5 years for the annual onset regressions. The specifications examined in this table are otherwise identical to corresponding ones reported in Table II. The reader is therefore referred to Table II and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis, the identification strategy employed by the IV probit regressions, and the estimation and interpretation of the marginal effect of population diversity on the incidence or onset of conflict. Heteroskedasticity-robust standard errors, clustered at the country level, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.XI: Population Diversity and the Incidence or Onset of Civil Conflict in Repeated Cross-Country Data – Robustness to Accounting for Deep-Rooted Determinants of Economic Development

Cross-country sample:	Old World		Global		Old World		Global	
	(1) Probit	(2) Probit	(3) IV Probit	(4) IV Probit	(5) Probit	(6) Probit	(7) IV Probit	(8) IV Probit
	Quinquennial PRIO25 civil conflict incidence, 1960–2017				Annual PRIO25 civil conflict onset, 1960–2017			
Population diversity (ancestry adjusted)	15.404***	9.821**	19.297***	15.653**	5.222*	4.777*	8.565**	11.664***
	(4.670)	(4.781)	(5.404)	(6.386)	(2.939)	(2.784)	(3.657)	(4.255)
Log years since Neolithic Revolution	0.085	0.187	−0.290	−0.243	0.333**	0.324*	0.029	−0.160
	(0.270)	(0.296)	(0.285)	(0.334)	(0.147)	(0.174)	(0.194)	(0.232)
Log index of state antiquity	0.244***	0.076	0.286***	0.143	0.093**	0.035	0.125**	0.096
	(0.088)	(0.103)	(0.101)	(0.116)	(0.041)	(0.057)	(0.051)	(0.070)
Log duration of human settlement	0.000	0.070	−0.024	−0.009	0.039	0.044	0.004	0.019
	(0.131)	(0.131)	(0.097)	(0.118)	(0.066)	(0.071)	(0.059)	(0.069)
Log distance from regional frontier in 1500	−0.031	0.001	−0.057	−0.025	0.049	0.050	−0.004	−0.018
	(0.052)	(0.051)	(0.040)	(0.047)	(0.032)	(0.038)	(0.026)	(0.031)
Continent dummies	×	×	×	×	×	×	×	×
Time dummies	×	×	×	×	×	×	×	×
Controls for temporal spillovers	×	×	×	×	×	×	×	×
Controls for geography	×	×	×	×	×	×	×	×
Controls for ethnic diversity		×		×		×		×
Controls for institutions		×		×		×		×
Controls for oil, population, and income		×		×		×		×
Observations	1,141	953	1,447	1,219	4,810	4,481	6,280	5,886
Countries	110	109	136	135	110	109	136	135
Pseudo R^2	0.425	0.432			0.143	0.151		
Marginal effect of diversity	2.992***	1.901**	3.885***	3.105**	0.293*	0.263*	0.437**	0.604**
	(0.896)	(0.936)	(1.140)	(1.333)	(0.165)	(0.154)	(0.203)	(0.257)
First-stage F statistic			41.126	39.893			48.227	44.985

Notes: This table conducts a robustness check on the results from the baseline analysis of the reduced-form impact of contemporary population diversity on either the quinquennial incidence or the annual onset of civil conflict in repeated cross-country data, as shown in Table II. Specifically, it establishes robustness to *additionally* accounting for the potentially confounding influence of other deep-rooted determinants of comparative economic development, including (i) the time elapsed since the onset of the Neolithic Revolution (e.g., Ashraf and Galor, 2013a); (ii) an index of experience with institutionalized statehood since antiquity (e.g., Bockstette et al., 2002); (iii) the time elapsed since initial human settlement in prehistory (e.g., Ahlerup and Olsson, 2012); and (iv) the great-circle distance to the closest regional technological frontier in the year 1500 (e.g., Ashraf and Galor, 2013a). The specifications examined in this table are otherwise identical to corresponding ones reported in Table II. The reader is therefore referred to Table II and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis, the identification strategy employed by the IV probit regressions, and the estimation and interpretation of the marginal effect of population diversity on the incidence or onset of conflict. Heteroskedasticity-robust standard errors, clustered at the country level, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.XII: Population Diversity and the Incidence or Onset of Civil Conflict in Repeated Cross-Country Data – Robustness to Accounting for Ethnic and Spatial Inequality

Cross-country sample:	Old World		Global		Old World		Global	
	(1) Probit	(2) Probit	(3) IV Probit	(4) IV Probit	(5) Probit	(6) Probit	(7) IV Probit	(8) IV Probit
	Quinquennial PRIO25 civil conflict incidence, 1960–2017				Annual PRIO25 civil conflict onset, 1960–2017			
Population diversity (ancestry adjusted)	14.732***	14.259***	16.367***	16.080***	6.687**	6.812**	7.892***	9.098***
	(3.867)	(3.801)	(3.782)	(4.046)	(2.862)	(2.952)	(2.971)	(3.367)
Ethnic inequality in luminosity	0.593	0.675	0.331	0.277	0.330	0.330	0.263	0.142
	(0.372)	(0.451)	(0.376)	(0.445)	(0.261)	(0.262)	(0.257)	(0.255)
Spatial inequality in luminosity	−0.035	0.150	0.294	0.519	−0.053	−0.017	0.070	0.086
	(0.409)	(0.425)	(0.392)	(0.410)	(0.256)	(0.259)	(0.247)	(0.279)
Continent dummies	×	×	×	×	×	×	×	×
Time dummies	×	×	×	×	×	×	×	×
Controls for temporal spillovers	×	×	×	×	×	×	×	×
Controls for geography	×	×	×	×	×	×	×	×
Controls for ethnic diversity		×		×		×		×
Controls for institutions		×		×		×		×
Controls for oil, population, and income		×		×		×		×
Observations	1,234	1,038	1,547	1,304	5,206	4,342	6,840	5,722
Countries	120	119	147	145	120	119	147	145
Pseudo R^2	0.408	0.442			0.133	0.172		
Marginal effect of diversity	2.838***	2.626***	3.272***	3.094***	0.348**	0.347**	0.370**	0.431**
	(0.717)	(0.702)	(0.787)	(0.843)	(0.154)	(0.153)	(0.153)	(0.182)
First-stage F statistic			125.548	93.701			133.266	99.940

Notes: This table conducts a robustness check on the results from the baseline analysis of the reduced-form impact of contemporary population diversity on either the quinquennial incidence or the annual onset of civil conflict in repeated cross-country data, as shown in Table II. Specifically, it establishes robustness to *additionally* accounting for the potentially confounding influence of measures of intrastate economic inequality (e.g., Alesina et al., 2016), as captured by the subnational spatial distribution of per-capita adjusted nighttime luminosity in the year 2000 across either (i) the georeferenced homelands of ethnic groups (ethnic inequality); or (ii) 2.5×2.5-degree geospatial grid cells (spatial inequality). The specifications examined in this table are otherwise identical to corresponding ones reported in Table II. The reader is therefore referred to Table II and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis, the identification strategy employed by the IV probit regressions, and the estimation and interpretation of the marginal effect of population diversity on the incidence or onset of conflict. Heteroskedasticity-robust standard errors, clustered at the country level, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.XIII: Population Diversity and the Incidence of Civil Conflict in Repeated Cross-Country Data – Robustness to Accounting for Alternative Correlates of Conflict Incidence

Cross-country sample:	Old World					Global				
	(1) Probit	(2) Probit	(3) Probit	(4) Probit	(5) Probit	(6) IV Probit	(7) IV Probit	(8) IV Probit	(9) IV Probit	(10) IV Probit
	Quinquennial PRIO25 civil conflict incidence, 1960–2017									
Population diversity (ancestry adjusted)	12.439***	12.412***	13.672***	9.587**	13.200***	13.115***	13.929***	14.428***	10.985**	14.758***
	(3.718)	(3.745)	(4.027)	(4.202)	(4.052)	(4.107)	(4.149)	(4.427)	(4.442)	(4.774)
Ethnic fractionalization (Fearon, 2003)	−0.266					−0.147				
	(0.332)					(0.329)				
Linguistic fractionalization (Alesina et al., 2003)		0.348					0.276			
		(0.354)					(0.317)			
Religious fractionalization (Alesina et al., 2003)			−0.463*					−0.705**		
			(0.280)					(0.276)		
Ethnolinguistic fractionalization (Esteban et al., 2012)				0.106					0.179	
				(0.365)					(0.346)	
Ethnolinguistic polarization (Esteban et al., 2012)				0.717					3.225**	
				(1.488)					(1.374)	
Gini index of ethnolinguistic diversity (Esteban et al., 2012)				−0.519					−1.358	
				(0.716)					(1.053)	
Log percentage mountainous terrain					0.099					0.112*
					(0.063)					(0.062)
Noncontiguous state dummy					0.371*					0.560***
					(0.214)					(0.182)
Disease richness					0.000					−0.007
					(0.010)					(0.010)
Controls for all baseline covariates	×	×	×	×	×	×	×	×	×	×
Observations	1,020	1,035	1,046	950	1,015	1,286	1,278	1,312	1,177	1,281
Countries	119	120	121	106	118	145	143	147	128	144
Pseudo R^2	0.429	0.436	0.438	0.451	0.436					
Marginal effect of diversity	2.387***	2.309***	2.547***	1.779**	2.499***	2.577***	2.664***	2.759***	2.124**	2.853***
	(0.722)	(0.700)	(0.762)	(0.789)	(0.784)	(0.852)	(0.833)	(0.894)	(0.891)	(0.978)
First-stage F statistic						100.578	104.976	98.705	68.499	70.482

Notes: This table conducts a robustness check on the results from the baseline analysis of the reduced-form impact of contemporary population diversity on the quinquennial incidence of civil conflict in repeated cross-country data, as shown in Columns 2 and 4 of Table II. Specifically, it establishes robustness to accounting for the potentially confounding influence of *alternative* distributional indices of intergroup diversity (e.g., Fearon, 2003; Alesina et al., 2003; Esteban et al., 2012) and *additional* geographical correlates of conflict (e.g., Fearon and Laitin, 2003; Cervellati et al., 2017). The specifications examined in this table are identical to the fully specified baseline models of conflict incidence, as reported in Columns 2 and 4 of Table II, with the exception that in Columns 1–3 and 6–8 of the current analysis, each of the reported control variables is employed in lieu of the baseline control for ethnic fractionalization (Alesina et al., 2003), whereas in Columns 4 and 9, the set of reported control variables replaces the baseline controls for both ethnic fractionalization and ethnolinguistic polarization (Desmet et al., 2012), in the interest of mitigating multicollinearity. Further, in Columns 5 and 10 of the current analysis, the set of reported geographical controls augment the fully specified baseline models from Columns 2 and 4 of Table II. The reader is therefore referred to Table II and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis, the identification strategy employed by the IV probit regressions, and the estimation and interpretation of the marginal effect of population diversity on the incidence of conflict. Heteroskedasticity-robust standard errors, clustered at the country level, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.XIV: Population Diversity and the Incidence or Onset of Civil Conflict in Repeated Cross-Country Data – Robustness to Employing the Classical Logit and Rare-Events Logit Estimators

	(1) Classical Logit	(2) Rare-Events Logit	(3) Classical Logit	(4) Rare-Events Logit	(5) Classical Logit	(6) Rare-Events Logit	(7) Classical Logit	(8) Rare-Events Logit
	Quinquennial PRIO25 civil conflict incidence, 1960–2017				Annual PRIO25 civil conflict onset, 1960–2017			
Population diversity (ancestry adjusted)	24.420***	23.755***	22.262***	20.941***	13.857**	13.409**	13.175**	12.442*
	(6.653)	(6.529)	(6.703)	(6.479)	(6.266)	(6.177)	(6.584)	(6.517)
Continent dummies	×	×	×	×	×	×	×	×
Time dummies	×	×	×	×	×	×	×	×
Controls for temporal spillovers	×	×	×	×	×	×	×	×
Controls for geography	×	×	×	×	×	×	×	×
Controls for ethnic diversity			×	×			×	×
Controls for institutions			×	×			×	×
Controls for oil, population, and income			×	×			×	×
Observations	1,270	1,270	1,045	1,045	5,452	6,280	4,377	5,221
Countries	123	123	121	121	123	123	121	121
Pseudo R^2	0.414		0.441		0.133		0.164	
Marginal effect of diversity	3.733***	3.964***	2.992***	3.230***	0.191**	0.194**	0.156*	0.171*
	(1.009)	(1.128)	(0.937)	(1.088)	(0.086)	(0.097)	(0.081)	(0.095)

Notes: This table conducts a robustness check on the results from the baseline analysis of the reduced-form impact of contemporary population diversity on either the quinquennial incidence or the annual onset of civil conflict in repeated cross-sectional data for the Old World sample of countries, as shown in Columns 1–2 and 5–6 of Table II. Specifically, it establishes robustness to employing the ordinary logit and rare-events logit (King and Zeng, 2001) estimators, rather than the probit estimator, for estimating the relevant empirical models of conflict incidence and onset. The specifications examined in this table are otherwise identical to corresponding ones reported in Columns 1–2 and 5–6 of Table II. The reader is therefore referred to Table II and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis. Given the absence of readily available ordinary logit and rare-events logit estimators that permit instrumentation, the current analysis is unable to implement the global-sample identification strategy of exploiting prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary population diversity. The estimated marginal effect of a 1 percentage point increase in population diversity is the marginal effect at the mean value of diversity in the cross-section, and it reflects the increase in either the quinquennial likelihood of a conflict incidence (Columns 1–4) or the annual likelihood of a conflict onset (Columns 5–8), both expressed in percentage points. Heteroskedasticity-robust standard errors, clustered at the country level, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.XV: Population Diversity and the Incidence or Onset of Civil Conflict in Repeated Cross-Country Data – Robustness to Accounting for Spatiotemporal Dependence using Two-Way Clustering of Standard Errors

	(1) Probit	(2) Logit	(3) Probit	(4) Logit	(5) Probit	(6) Logit	(7) Probit	(8) Logit
	Quinquennial PRIO25 civil conflict incidence, 1960–2017				Annual PRIO25 civil conflict onset, 1960–2017			
Population diversity (ancestry adjusted)	13.366***	24.420***	12.203***	22.262***	6.172**	13.857**	6.356*	13.175*
	(2.616)	(4.261)	(3.381)	(6.025)	(2.906)	(6.528)	(3.478)	(7.368)
Continent dummies	×	×	×	×	×	×	×	×
Time dummies	×	×	×	×	×	×	×	×
Controls for temporal spillovers	×	×	×	×	×	×	×	×
Controls for geography	×	×	×	×	×	×	×	×
Controls for ethnic diversity			×	×			×	×
Controls for institutions			×	×			×	×
Controls for oil, population, and income			×	×			×	×
Observations	1,270	1,270	1,045	1,045	5,452	5,452	4,377	4,377
Countries	123	123	121	121	123	123	121	121
Pseudo R^2	0.416	0.414	0.440	0.441	0.131	0.133	0.161	0.164

Notes: This table conducts a robustness check on the results from the baseline probit and logit analyses of the reduced-form impact of contemporary population diversity on either the quinquennial incidence or the annual onset of civil conflict in repeated cross-sectional data for the Old World sample of countries, as shown in Columns 1–2 and 5–6 of Table II and in odd-numbered columns of Table SA.XIV. Specifically, it establishes robustness of the standard-error estimates to accounting for spatiotemporal dependence across country-time observations by implementing multi-dimensional clustering of standard errors, following the methodology of Cameron et al. (2011). To implement this robustness check, the standard errors across country-time observations are clustered in two dimensions: (i) the country level, which allows for temporal dependence within a country over time (i.e., across either 5-year intervals or years); and (ii) the time level, which allows for spatial dependence across countries within a given time period (i.e., either a 5-year interval or a year). The specifications examined in this table are otherwise identical to corresponding ones reported in Columns 1–2 and 5–6 of Table II and in odd-numbered columns of Table SA.XIV. The reader is therefore referred to Table II and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis. Given the absence of readily available probit and logit estimators that not only allow for multi-dimensional clustering of standard errors but also permit instrumentation, the current analysis is unable to implement the global-sample identification strategy of exploiting prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary population diversity. Heteroskedasticity-robust standard errors, clustered multi-dimensionally at both the country and time levels, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.XVI: Population Diversity and the Onset of Civil Conflict in Repeated Cross-Country Data – Robustness to Accounting for Alternative Correlates of Conflict Onset

Cross-country sample:	Old World				Global			
	(1) Probit	(2) Probit	(3) Probit	(4) Probit	(5) IV Probit	(6) IV Probit	(7) IV Probit	(8) IV Probit
	Annual PRIO25 civil conflict onset, 1960–1999							
Population diversity (ancestry adjusted)	7.791**	6.872**	8.267**	8.330*	8.808**	8.111**	11.955**	11.507**
	(3.657)	(3.469)	(4.181)	(4.342)	(3.516)	(3.417)	(4.838)	(4.975)
Ethnic dominance		0.147		−0.002		0.147		0.040
		(0.115)		(0.135)		(0.103)		(0.129)
Political instability, lagged		0.264**		0.165		0.245**		0.056
		(0.106)		(0.136)		(0.098)		(0.128)
New state dummy, lagged		0.125				−0.149		
		(0.527)				(0.494)		
Continent dummies	×	×	×	×	×	×	×	×
Time dummies	×	×	×	×	×	×	×	×
Controls for temporal spillovers	×	×	×	×	×	×	×	×
Controls for geography	×	×	×	×	×	×	×	×
Controls for ethnic diversity			×	×			×	×
Controls for institutions			×	×			×	×
Controls for oil, population, and income			×	×			×	×
Observations	2,761	2,761	2,139	2,139	3,728	3,728	3,031	3,031
Countries	96	96	94	94	121	121	119	119
Pseudo R^2	0.137	0.145	0.155	0.157				
Marginal effect of diversity	0.472**	0.413*	0.516*	0.519*	0.495**	0.448**	0.706**	0.672*
	(0.231)	(0.216)	(0.267)	(0.277)	(0.224)	(0.210)	(0.349)	(0.350)
First-stage F statistic					132.831	132.602	78.279	73.849

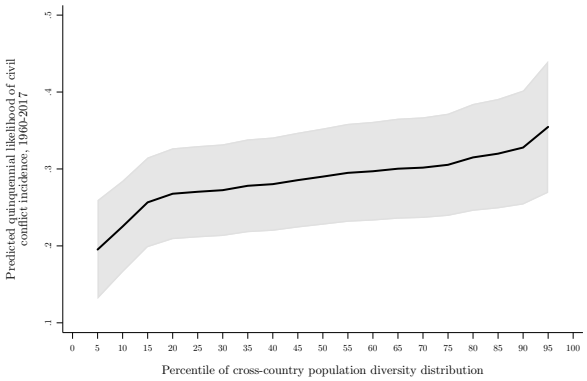
Notes: This table conducts a robustness check on the results from the baseline analysis of the reduced-form impact of contemporary population diversity on the annual onset of civil conflict in repeated cross-country data, as shown in Columns 5–8 of Table II. Specifically, it establishes robustness to accounting for the potentially confounding influence of an *additional* distributional index of intergroup diversity (e.g., Collier and Hoeffler, 2004) and *additional* time-varying institutional correlates of conflict (e.g., Fearon and Laitin, 2003). The lagged indicator for the emergence of a newly independent state from colonial powers is dropped from the specifications in Columns 4 and 8 due to multicollinearity. In light of constraints imposed by the availability of data on the additional control variables in this table, the analysis is restricted to the 1960–1999 as opposed to the 1960–2017 time period. Therefore, the specification presented in each odd-numbered column of the table is intended to provide a relevant baseline for the robustness check in the subsequent even-numbered column (i.e., by holding fixed the regression sample). The specifications examined in this table are otherwise identical to the baseline models of conflict onset, as reported in Columns 5–8 of Table II. The reader is therefore referred to Table II and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis, the identification strategy employed by the IV probit regressions, and the estimation and interpretation of the marginal effect of population diversity on the onset of conflict. Heteroskedasticity-robust standard errors, clustered at the country level, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.XVII: Population Diversity and the Onset of Civil Conflict in Repeated Cross-Country Data – Robustness to Accounting for Commodity Export Price Shocks

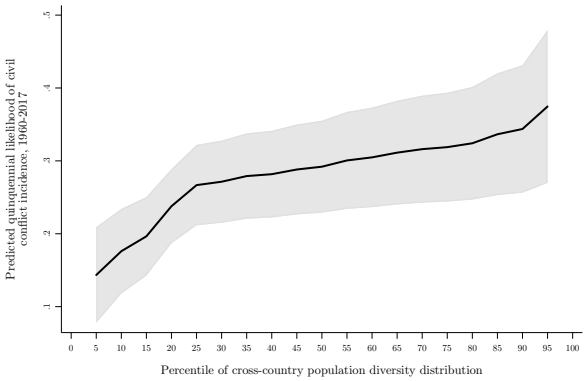
Cross-country sample:	Old World				Global			
	(1) Probit	(2) Probit	(3) Probit	(4) Probit	(5) IV Probit	(6) IV Probit	(7) IV Probit	(8) IV Probit
	Annual PRIO25 civil conflict onset, 1960–2007							
Population diversity (ancestry adjusted)	8.596**	8.946**	8.632**	8.734**	9.007***	10.656**	9.086***	10.592**
	(3.665)	(3.894)	(3.622)	(3.899)	(3.401)	(4.537)	(3.388)	(4.570)
Aggregate price shock	−0.128**	−0.159***			−0.137***	−0.190***		
	(0.052)	(0.059)			(0.053)	(0.056)		
Aggregate price shock, lagged	0.026	0.021			0.014	0.017		
	(0.060)	(0.069)			(0.058)	(0.062)		
Aggregate price shock, twice lagged	−0.172***	−0.179***			−0.113*	−0.121*		
	(0.060)	(0.066)			(0.058)	(0.064)		
Annual crop price shock			−0.161**	−0.191**			−0.156**	−0.223***
			(0.071)	(0.083)			(0.071)	(0.075)
Annual crop price shock, lagged			−0.039	−0.048			−0.049	−0.045
			(0.083)	(0.093)			(0.082)	(0.088)
Annual crop price shock, twice lagged			−0.176**	−0.178*			−0.101	−0.112
			(0.084)	(0.094)			(0.084)	(0.095)
Perennial crop price shock			−0.127*	−0.144**			−0.127**	−0.154***
			(0.066)	(0.070)			(0.058)	(0.059)
Perennial crop price shock, lagged			0.116***	0.120**			0.094**	0.089*
			(0.045)	(0.054)			(0.046)	(0.051)
Perennial crop price shock, twice lagged			−0.130***	−0.145***			−0.076	−0.083*
			(0.050)	(0.053)			(0.046)	(0.049)
Extractive crop price shock			−0.187**	−0.247***			−0.185**	−0.275***
			(0.081)	(0.092)			(0.081)	(0.086)
Extractive crop price shock, lagged			0.051	0.055			0.031	0.041
			(0.088)	(0.098)			(0.088)	(0.094)
Extractive crop price shock, twice lagged			−0.330***	−0.332***			−0.256***	−0.264**
			(0.103)	(0.111)			(0.096)	(0.104)
Continent dummies	×	×	×	×	×	×	×	×
Time dummies	×	×	×	×	×	×	×	×
Controls for temporal spillovers	×	×	×	×	×	×	×	×
Controls for geography	×	×	×	×	×	×	×	×
Controls for ethnic diversity		×		×		×		×
Controls for institutions		×		×		×		×
Observations	2,876	2,626	2,876	2,626	3,906	3,599	3,906	3,599
Countries	82	81	82	81	105	103	105	103
Pseudo R^2	0.122	0.150	0.133	0.162				
Marginal effect of diversity	0.531**	0.535**	0.528**	0.516**	0.501**	0.577**	0.500**	0.568**
	(0.237)	(0.242)	(0.232)	(0.240)	(0.213)	(0.281)	(0.211)	(0.280)
First-stage F statistic					102.975	51.265	102.702	51.169

Notes: This table conducts a robustness check on the results from the baseline analysis of the reduced-form impact of contemporary population diversity on the annual onset of civil conflict in repeated cross-country data, as shown in Columns 5–8 of Table II. Specifically, it establishes robustness to *additionally* accounting for the potentially confounding "income effect" of commodity export price shocks (e.g., Bazzi and Blattman, 2014), as captured by the contemporaneous, lagged, and twice lagged values of either an annual price shock that has been aggregated across commodity export types (Columns 1–2 and 5–6) or annual price shocks disaggregated by type of commodity export, including export price shocks associated with annual crops, perennial crops, and extractive crops (Columns 3–4 and 7–8). These export price shock variables are all obtained from the data set of Bazzi and Blattman (2014), so the reader is referred to that work for additional details on these variables. In light of constraints imposed by the availability of data on these export price shock variables, the analysis is restricted to the 1960–2007 as opposed to the 1960–2017 time period. The specifications examined in this table are otherwise identical to those reported in Columns 5–8 of Table II, with the exception that the fully specified models in the current analysis omit the controls for oil presence, total population, and GDP per capita, in the interest of minimizing endogeneity with the export price shock variables and maximizing degrees of freedom. The reader is therefore referred to Table II and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis, the identification strategy employed by the IV probit regressions, and the estimation and interpretation of the marginal effect of population diversity on the onset of conflict. Heteroskedasticity-robust standard errors, clustered at the country level, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

A.3 Supplementary Figures



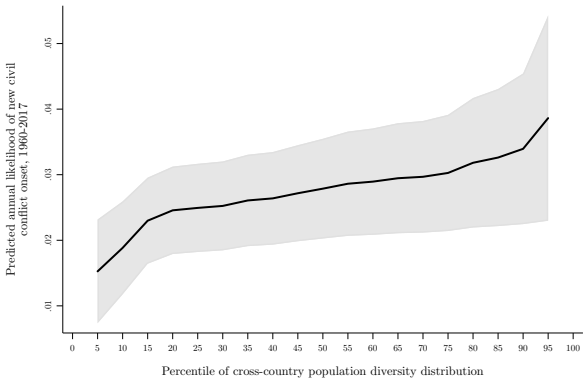
(a) Old-World sample



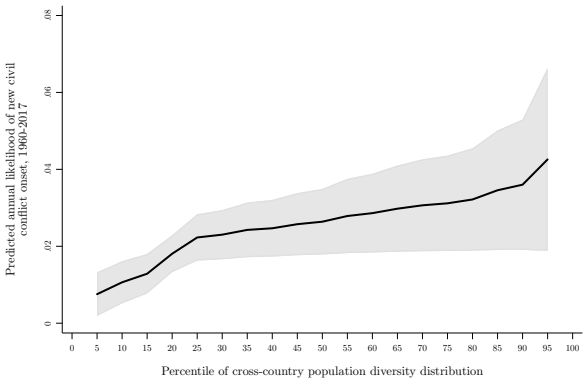
(b) Global sample

FIGURE SA.1: Population Diversity and the Incidence of Civil Conflict

Notes: This figure depicts the influence of contemporary population diversity on the *predicted* likelihood of observing the incidence of a PRIO25 civil conflict in any given 5-year interval during the 1960–2017 time period, conditional on the full set of control variables, as considered by the specifications in Columns 2 and 4 of Table II. In each panel, the predicted likelihood of civil conflict incidence is illustrated as a function of the percentile of the cross-country diversity distribution in the relevant estimation sample, and the shaded area reflects the 95-percent confidence-interval region of the depicted relationship.



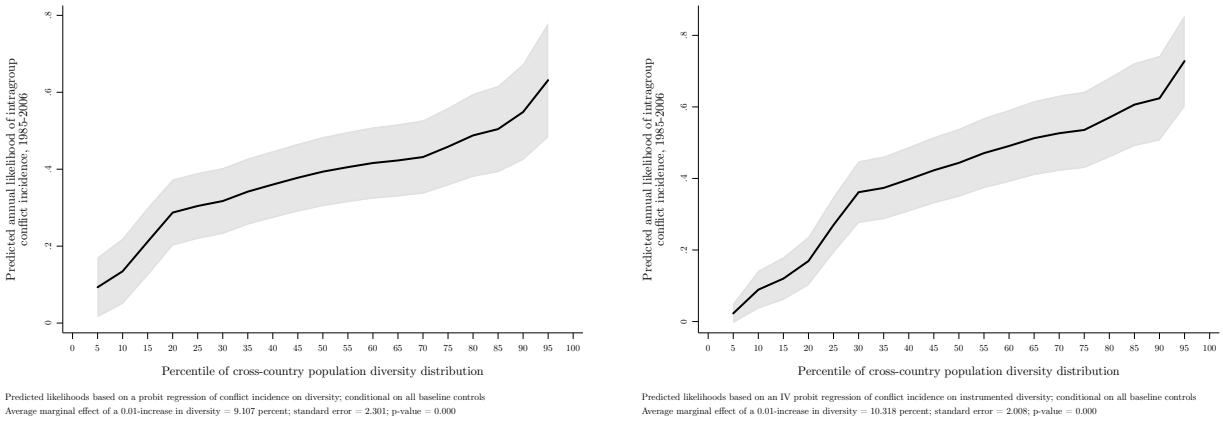
(a) Old-World sample



(b) Global sample

FIGURE SA.2: Population Diversity and the Onset of Civil Conflict

Notes: This figure depicts the influence of contemporary population diversity on the *predicted* likelihood of observing the onset of a new PRIO25 civil conflict in any given year during the 1960–2017 time period, conditional on the full set of control variables, as considered by the specifications in Columns 6 and 8 of Table II. In each panel, the predicted likelihood of civil conflict onset is illustrated as a function of the percentile of the cross-country diversity distribution in the relevant estimation sample, and the shaded area reflects the 95-percent confidence-interval region of the depicted relationship.



(a) Old-World sample

(b) Global sample

FIGURE SA.3: Population Diversity and the Incidence of Intragroup Conflict

Notes: This figure depicts the influence of contemporary population diversity on the *predicted* likelihood of observing the incidence of one or more intragroup conflicts in any given year during the 1985–2006 time period, conditional on the full set of control variables, as considered by the specifications in Columns 2 and 5 in Panel B of Table III. In each panel, the predicted likelihood of intragroup conflict incidence is illustrated as a function of the percentile of the cross-country diversity distribution in the estimation relevant sample, and the shaded area reflects the 95-percent confidence-interval region of the depicted relationship.

A.4 Variable Definitions for the Country-level Analyses

Migratory Distance and Population Diversity

1. **Migratory distance from East Africa:** The great circle distance from Addis Ababa, Ethiopia to a country's capital city along a land-restricted path forced through one or more of five intercontinental waypoints, including Cairo, Egypt; Istanbul, Turkey; Phnom Penh, Cambodia; Anadyr, Russia; and Prince Rupert, Canada. Distances are calculated using the Haversine formula and are measured in units of ten thousand kilometers. The methodology underlying the construction of this measure is adopted from Ramachandran et al. (2005). The geographical coordinates of the waypoints are obtained from Ramachandran et al. (2005) and those of the capital cities are obtained from the Central Intelligence Agency's (CIA) World Factbook. See Ashraf and Galor (2013a) for additional details.

2. **Population diversity (precolonial):** The expected heterozygosity (neutral genetic diversity) of a country's precolonial population as predicted by migratory distance from East Africa (i.e., Addis Ababa, Ethiopia) to the country's capital city. This measure is calculated by applying the regression coefficients obtained from regressing expected heterozygosity on migratory distance at the ethnic group level, using a worldwide sample of 53 ethnic groups from the HGDP-CEPH Human Genome Diversity Cell Line Panel. The expected heterozygosities and geographical coordinates of the ethnic groups are from Ramachandran et al. (2005). See Ashraf and Galor (2013a) for additional details.

3. **Population diversity (ancestry adjusted):** The expected heterozygosity (neutral genetic diversity) of a country's contemporary national population, as developed by Ashraf and Galor (2013a). This measure is based on migratory distances from East Africa to the year 1500 locations of the ancestral populations of the country's component ethnic groups in 2000 and on the pairwise migratory distances among these ancestral populations. The source countries of the ancestral populations are identified from the World Migration Matrix, 1500–2000 (Putterman and Weil, 2010), and the capital cities of these countries are used to compute the aforementioned migratory distances. The measure of population diversity is then computed by applying (i) the coefficients obtained from regressing expected heterozygosity on migratory distance from East Africa at the ethnic group level, using a worldwide sample of 53 ethnic groups from the HGDP-CEPH Human Genome Diversity Cell Line Panel; (ii) the coefficients obtained from regressing pairwise genetic distance on pairwise migratory distance in a sample of 1,378 HGDP-CEPH ethnic group pairs, and (iii) the ancestry weights representing the fractions of the year 2000 national population (i.e., of the country for which the measure is being computed) that can trace their ancestral origins to different source countries in the year 1500. The data at the ethnic-group (or group-pair) level on expected heterozygosities, geographical coordinates, and pairwise genetic distances are obtained from Ramachandran et al. (2005), and the country-level data on ancestry weights are obtained from the World Migration Matrix, 1500–2000. See Ashraf and Galor (2013a) for a detailed discussion of the methodology underlying the construction of this measure.

Conflict outcomes

1. **PRIO civil conflict and civil war outcomes:** Our primary measures of civil conflict are based on Version 18.1 of the UCDP/PRIO Armed Conflict Dataset (ACD), covering the 1946–2017 time period (Gleditsch et al., 2002; Pettersson and Eck, 2018). In this dataset, an armed

conflict is defined as "a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths in a calendar year." In our study, the term *PRIO25 civil conflict* indicates an internal armed conflict between the government of a state and one or more internal opposition group(s), without any intervention from other states as independent actors or intervention from other states to support either side of the conflict. Thus, the measures of civil conflict in our study exclude internationalized internal armed conflicts. In addition, extrasytemic and interstate conflicts are also excluded from the analysis, following the standard definition of civil conflict. For further information on the data underlying our various civil conflict measures (discussed below), the interested reader is referred to the codebook for Version 18.1 of the UCDP/PRIO ACD.

The main conflict variable examined in our cross-sectional analyses of civil conflict is the ***log number of new PRIO25 civil conflict onsets per year*** during the 1960–2017 time period. This measure is obtained by first computing the total count of *new* civil conflicts that took place on the territory of a country in our sample during this period. Then, this count is divided by the number of years over the same time period in which the territory was home to one or more entities included in the Gleditsch and Ward list of independent states, as employed by the UCDP/PRIO ACD. Finally, the resulting average annual conflict frequency is scaled up by 1 and log-transformed. Each *new* conflict is identified by a unique conflict identifier provided by the UCDP/PRIO ACD. In this definition, two or more conflict episodes involving the same actors fighting over the same incompatibility are not treated as separate (new) conflicts. Instead, they are assigned the same conflict identifier.

The main outcome examined by our regressions using annually repeated cross-country data is ***annual PRIO25 civil conflict onset***. It is equal to 1 for each year when at least one new PRIO25 conflict broke out and zero otherwise. The date of a *new* conflict outbreak (or onset) is the starting year of the first conflict episode for a given conflict, and it reflects the first year in which the conflict reached or surpassed the annual fatality threshold of 25 battle-related deaths. Subsequent years of a given conflict episode or outbreaks of subsequent conflict episodes of the same conflict are not considered *new* conflict onsets.

Quinquennial PRIO25 civil conflict incidence is the main outcome examined by our regressions using quinquennially repeated cross-country data over the 1960–2017 time period. It is equal to 1 for a given 5-year interval for a country if there was an active (ongoing) PRIO25 civil conflict in at least one year during that time interval and zero otherwise. A conflict is deemed active in a given calendar year if it resulted in at least 25 battle-related deaths during that year. ***Annual PRIO25 civil conflict incidence*** is defined in a similar manner except that the incidence is coded for each country-year observation instead of a 5-year time interval for a country.

Quinquennial PRIO1000 civil war incidence is an alternative outcome examined by our robustness checks in regressions using quinquennially repeated cross-country data. This variable is constructed in a manner similar to ***quinquennial PRIO25 civil conflict incidence***. The only difference is that for civil wars, a conflict is deemed as active (ongoing) in a given year only if a much higher fatality threshold of 1,000 (instead of 25) battle-related deaths is exceeded in that year.

2. **Intragroup (intracommunal) factional conflict:** The outcome variables employed by the analysis of intragroup conflict are based on the All Minorities At Risk (AMAR) Sample

Data of the AMAR Phase I Project (Birnir et al., 2018). The AMAR sample contains longitudinal data on 365 AMAR ethnic groups. Of these groups, 291 were included in the original Minorities At Risk (MAR) Project (Phases I–V), and the remaining 74 were selected randomly from the sample frame of socially relevant groups outlined by Birnir et al. (2015), according to the new AMAR criteria summarized in the AMAR codebook.

The measures of intragroup factional conflict we employ are constructed using the *INTRACON* variable in the AMAR Sample Data. This is a dummy variable, coded for each group in the AMAR sample, indicating the presence of an intracommunal conflict within that group in a given year. Specifically, the variable is coded for each year during the 1980–2006 time period. However, since the coverage of AMAR groups for the 1980–1984 time period is rather limited, our measures of intragroup conflict are based on information for the 1985–2006 time frame. Thus, the outcome variable in our cross-country analysis of intragroup conflict is the ***share of AMAR group-years with at least one intracommunal conflict*** within a country during this time period. Further, the outcome variable in our analysis of intragroup conflict using annually repeated cross-country data is ***annual intracommunal conflict incidence***, coded 1 for each country-year in which there was at least one AMAR sample group with an active intracommunal conflict and zero otherwise. For further information on the data underlying our measures of intragroup conflict, the reader is referred to Version 1 of the codebook for the AMAR Phase I Project.

3. **Historical conflict outcomes:** To construct historical conflict outcomes between the 15th and 19th centuries, we make use of information on the locations of violent conflicts during the 1400–1799 time period, as compiled by Brecke (1999) and georeferenced by Dincecco et al. (2015). The georeferenced conflict locations are used to map historical conflicts to territories, as defined by their contemporary national borders. It may be noted that in the catalog of conflicts from Dincecco et al. (2015), there were a small number of instances where the country assignment did not match the country implied by the georeferenced location of the conflict in ArcGIS. In such cases, supplementary information from the catalog (e.g., the actors in the conflict or the place where the conflict occurred) was consulted to first determine if the mismatch was due to an error in the original country assignment or an error in the supplied coordinates. Then, either the country assignment or the coordinates were altered to match our understanding of the true location of the conflict. In addition, for naval conflicts or for conflicts between actors that took place on lands to which neither actor was native, these specific conflicts were assigned to either one of the actors' countries (rather than the country implied by the location of the conflict) but only if the actors possessed comparable levels of diversity (e.g., if the actors were both European colonial powers engaged in a conflict on a colonized territory).

   As for the underlying conflict data, the definition of a violent conflict in Brecke's dataset is based on Cioffi-Revilla (1996): "An occurrence of purposive and lethal violence among 2+ social groups pursuing conflicting political goals that results in fatalities, with at least one belligerent group organized under the command of authoritative leadership. The state does not have to be an actor. Data can include massacres of unarmed civilians or territorial conflicts between warlords." The list is comprised of conflicts that resulted in at least 32 fatalities. This fatality level corresponds to a magnitude of 1.5 or higher on Richardson's (1960) base-10 log conflict scale. Although the dataset does not systematically distinguish between intrastate and interstate conflicts, the latter appear to form the basis of the recorded conflicts, and while the recorded conflicts do not necessarily represent the whole universe

of conflict events during the sample period, the list contains almost all major conflicts that have been documented by historians. The conflict catalog is also considered to be fairly comprehensive in terms of its broad regional coverage, including five regions of the world: Western Europe, Eastern Europe, North Africa, West & Central Africa, East & Southern Africa, as well as Central Asia & Siberia.

Based on these conflict data, our study employs two distinct categories of country-level outcome measures: (1) the number of distinct conflicts, occurring in each century of the 1400–1799 time period or across this entire time frame; and (2) the likelihood of observing one or more conflicts, either during the entire 1400–1799 time period or in each century therein.

4. **MEPV civil conflict severity:** This variable is constructed using information provided by the Major Episodes of Political Violence (MEPV) War List (1946–2017), maintained by the Center for Systemic Peace. This list is a regularly updated version of Appendix C from Marshall (1999) and further detailed in Marshall (2002).

   A major episode of political violence is defined as the systematic and sustained use of lethal violence by one or more organized groups, resulting in at least 500 directly-related deaths over the course of the episode. Episodes are coded for both time span and a general magnitude of societal-systemic impact (an eleven-point scale, 0-10). These magnitude scores are considered to be consistent and comparable across categories and cases. Further, each episode is assigned to one of seven categories of armed conflict: international violence (IV), international war (IW), international independence war (IN), civil violence (CV), civil war (CW), ethnic violence (EV), and ethnic war (EW). Episodes belonging to the last four of these categories constitute the universe of intrastate episodes that are of interest to our analysis. The magnitude scores for these episodes are aggregated into the *CIVTOT* variable in the MEPV dataset. *CIVTOT* is an annual ordinal index of civil conflict intensity at the country level that underlies the particular measure of ***quinquennial MEPV civil conflict severity*** we employ – namely, the maximum value of *CIVTOT* across all years in any given 5-year interval during the 1960–2017 time period. For further information on the data underlying our measure of civil conflict severity, the reader is referred to the codebook for the MEPV dataset.

5. **CNTS social conflict index:** This variable is based on the Domestic Conflict Event Data from the Cross-National Time Series (CNTS) Data Archive 2018 Edition (Banks and Wilson, 2018), which covers the 1815–2017 time period.

   Specifically, the basis of our CNTS social conflict index is the variable *Domestic9* from the CNTS Data Archive. *Domestic9* is an annual continuous index of the degree of social unrest, computed by first taking the weighted sum of the counts of different unrest/conflict events (given by the variables *domestic1-8*) in a country-year. As of October 2007, the weights employed were as follows: Assassinations (25), Strikes (20), Guerrilla Warfare (100), Government Crises (20), Purges (20), Riots (25), Revolutions (150), and Anti-Government Demonstrations (10). In a second step, the weighted sum is multiplied by 100/8 to obtain *Domestic9*. The specific measure used in our study is a ***quinquennial CNTS social conflict index***, calculated for each country as the maximum value of *Domestic9* across all years in any given 5-year interval during the 1960–2017 time period. For further information on the source data for our social conflict index, the reader is referred to the website of the CNTS Data Archive.

6. **UCDP nonstate conflict incidence:** This measure is based on information from Version 18.1 of the UCDP Non-State Conflict Dataset, covering the 1989–2017 time period (Sundberg et al., 2012).

   A non-state conflict is defined by the Uppsala Conflict Data Program (UCDP) as "the use of armed force between two organized armed groups, neither of which is the government of a state, which results in at least 25 battle-related deaths in a year." An *organized group* can be either (i) a *formally organized group*, i.e., any non-governmental group of people having announced a name for their group and using armed force against another similarly organized group; or (ii) an *informally organized group*. The latter type of group does not have an announced name, but it uses armed force against another similarly organized group such that there is a clear pattern of violent incidents that are connected and in which both groups use armed force against the other. ***Quinquennial UCDP nonstate conflict incidence*** is coded 1 for any 5-year interval for a country if in any year during this interval there was at least one active (ongoing) non-state conflict in the country. A conflict is deemed active in a given calendar year if it resulted in at least 25 battle-related deaths during that year. For further information on the source data for our measure of non-state conflict incidence, the reader is referred to the codebook for Version 18.1 of the UCDP Non-State Conflict Dataset.

Other outcomes

1. **Number of ethnic groups:** The total number of distinct ethnic groups in a country's population, as compiled by Fearon (2003). The specific variable employed by our analysis is the natural logarithm of one plus the number of ethnic groups. See Fearon (2003) for additional details on primary data sources and methodological assumptions.

2. **Prevalence of interpersonal trust:** This variable is constructed using information from the World Values Survey (2006, 2009) (henceforth, WVS) on the prevalence of generalized interpersonal trust in a country's population. In particular, this well-known measure of social capital at the country level reflects the proportion of all respondents (from across five different waves of the WVS, conducted over the 1981–2009 time period) that opted for the answer "Most people can be trusted" (as opposed to "Can't be too careful") when responding to the survey question "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?" For additional details, the reader is referred to documentation available on the WVS website.

3. **Variation in political attitudes:** The intra-country dispersion in self-reported individual political positions on a "left"–"right" categorical scale, based on data from the WVS. Specifically, this measure of heterogeneity in political attitudes at the country level is calculated as the intra-country standard deviation across all respondents (sampled over five different waves of the WVS during the 1981–2009 time period) of their self-reported positions on a categorical scale from 1 (politically "left") to 10 (politically "right") when answering the survey question "In political matters, people talk of 'the left' and 'the right.' How would you place your views on this scale, generally speaking?" Given that this variable's unit of measurement does not possess a natural interpretation, we standardize the cross-country distribution of this variable prior to conducting our regressions. For additional details, the reader is referred to documentation available on the WVS website.

Main Control Variables

1. **Ethnic fractionalization:** This is the well-known ethnic fractionalization index of a country, reflecting the probability that two individuals, randomly selected from the country's population, will belong to different ethnic groups. Formally, a country's ethnic fractionalization index is calculated as follows:

$$FRAC = 1 - \sum_{i=1}^{n} p_i^2,$$

where p_i is the proportional representation of ethnic group i in the national population; and n is the total number of ethnic groups in the country. The specific variable we employ is based on the list of ethnic groups (and their national population shares) by country as compiled by Alesina et al. (2003). See Alesina et al. (2003) for additional details on primary data sources and methodological assumptions.

2. **Ethnolinguistic polarization:** An ethnolinguistic polarization index at the country level, calculated by applying the following definition of polarization due to Reynal-Querol (2002) and Montalvo and Reynal-Querol (2005):

$$POL = 4 \sum_{i=1}^{n} p_i^2 \left[1 - p_i\right],$$

where p_i is the proportional representation of linguistic group i in the national population; and n is the total number of linguistic groups in the country. The employed ethnolinguistic polarization index is sourced from the replication dataset of Desmet et al. (2012). The authors provide measures of several such polarization indices, constructed at different levels of aggregation of linguistic groups in a country's population (based on hierarchical linguistic trees). The specific polarization measure we use corresponds to the most disaggregated level of the linguistic tree, and it reflects the extent of polarization across subnational groups classified according to modern-day languages. See Desmet et al. (2012) for additional details on primary data sources and methodological assumptions.

3. **Absolute latitude:** The absolute value of the latitude of a country's geodesic centroid, as reported by the *At These Coordinates* resource repository, based on metadata from (i) the National Geospatial-Intelligence Agency's (NGA) GEOnet Names Server (GNS); and (ii) the United States Geological Survey's (USGS) Geographic Names Information System (GNIS).

4. **Ruggedness:** A measure of the degree of terrain ruggedness of a country's territory. Based on Riley et al. (1999), the ruggedness of a grid cell, i, is defined as

$$RIX(i) = \sqrt{\sum_{k=1}^{8} (h_i - h_{j_k})^2},$$

where h_l is the elevation (in meters above sea level) of cell $l = i, j_1, j_2, ..., j_8$, and the cells indexed by j are the eight neighboring cells of i. The country-level measure of ruggedness used by our study is the mean value of $RIX(i)$ across all 1 km $\times$ 1 km grid cells of a country. The cell-level ruggedness index is computed by Özak (2010), based on topographical data from

the Global Land One-Kilometer Base Elevation (GLOBE) digital elevation model (Hastings et al., 1999).

5. **Mean and range of elevation:** The country-level mean and range of elevation (in thousands of kilometers above sea level), calculated using geospatial elevation data at a 1-degree resolution from the Geographically based Economic data (G-ECON) project (Nordhaus, 2006), based on similar data at a 10-minute resolution from New et al. (2002). The mean of elevation at the country level reflects the average value across the grid cells that are located within a country's national borders, whereas the range of elevation reflects the difference between the maximum and minimum values across the same set of grid cells. See the G-ECON project website for additional details.

6. **Mean and range of land suitability:** The country-level mean and range of a geospatial index of the suitability of land for agriculture, based on ecological indicators of climate suitability for cultivation, such as growing degree days and the ratio of actual to potential evapotranspiration, as well as on ecological indicators of soil suitability for cultivation, such as soil carbon density and soil pH. This index was initially developed at a half-degree resolution by Ramankutty et al. (2002), and it has been aggregated to the country level by Michalopoulos (2012), with the mean at the country level reflecting the average value of the index across the grid cells that are located within a country's national borders, and the range reflecting the difference between the maximum and minimum values of the index across the same set of grid cells. See Michalopoulos (2012) for additional details.

7. **Island nation:** An indicator for whether a country shares a land border with any other country, as reported by the CIA's World Factbook. Of the 147 countries in our baseline sample, the following 7 are coded as island nations: Australia, Cuba, Japan, Sri Lanka, Madagascar, New Zealand, and Philippines.

8. **Distance to nearest waterway:** The distance (in thousands of kilometers) from a grid cell to the nearest ice-free coastline or sea-navigable river, averaged across the grid cells of a country. This variable was originally constructed by Gallup et al. (1999) and is available from the Research Datasets online repository maintained by Harvard University's Center for International Development.

9. **Colonial history:** A set of three indicators reflecting a country's experience of colonial rule by (i) the U.K., (ii) France, or (iii) any other major colonizing power, respectively. Therefore, the omitted category is the absence of colonial rule. These variables are constructed based on information from various sources, including the CIA's World Factbook, the Encyclopaedia Brittanica, Country Studies of the Library of Congress, and rulers.org amongst others. Additional details are available from the authors upon request.

   In cross-sectional regressions at the country level, the relevant measures comprise time-invariant indicators for the historical presence of colonial rule – i.e., whether the country has ever been ruled by the colonizing power in question. In regressions using repeated cross-country data, the relevant measures comprise time-varying indicators of the lagged prevalence of colonial rule – i.e., whether the country was ruled by the colonizing power in question at any point in the preceding 5-year time interval or in the preceding year, depending on the temporal dimension of the repeated cross-section.

10. **Legal origins:** A set of two time-invariant indicators for British and French legal origins, as reported by La Porta et al. (1999). Specifically, these indicators identify whether the legal origin of country's Company Law or Commercial Code is (i) the English Common Law or (ii) the French Commercial Code, respectively. The omitted category is German, Scandinavian, or Socialist legal origins, as recognized by La Porta et al. (1999).

11. **Executive constraints:** An index, reported at an annual frequency as a 7-point categorical variable (from 1 to 7) by the Polity IV Project (Version 2017), quantifying the extent of institutionalized constraints on the decision-making power of chief executives (Marshall et al., 2017). The specific version of the Polity IV Project dataset employed by our study covers the 1800–2017 time period. For further information on the index of executive constraints, the reader is referred to the codebook for Version 2017 of the Polity IV Project dataset.

    In cross-sectional regressions at the country level, the relevant measure is the temporal average of the index across all years in the 1960–2017 time period. In regressions using quinquennially repeated cross-country data, the relevant measure is the temporal average of the index across all years in the preceding 5-year time interval. Finally, in regressions based on annually repeated cross-country data, the relevant measure is the value of the index from the preceding year.

12. **Type of political regime:** Our measures of the type of political regime are based on two indicators reflecting whether a country is classified as a democracy (or not) and as an autocracy (or not) in a given year. The omitted category is anocracy, a hybrid regime that constitutes the middle range of the autocracy-democracy political spectrum. This regime classification is based on the *POLITY2* index (the Revised Combined Polity Score), as reported at an annual frequency by the Polity IV Project (Version 2017) for the 1800–2017 time period (Marshall et al., 2017). *POLITY2* is a discrete index that ranges from -10 (strongly autocratic) to +10 (strongly democratic). Following the norm in the literature, a country-year is coded as a *democracy* if the *POLITY2* score is above 5 or as an *autocracy* if the score is below -5. The prevalence of *anocracy*, occurring when the *POLITY2* score is between -5 and 5 for a country-year, therefore serves as the omitted political regime category. For further information on the *POLITY2* index, the reader is referred to the codebook for Version 2017 of the Polity IV Project dataset.

    In cross-sectional regressions at the country level, the relevant measures of regime type are the fractions of years during the 1960–2017 time period that a country spent as a democracy and as an autocracy, respectively. In regressions using quinquennially repeated cross-country data, the relevant measures are the fractions of years during the preceding 5-year time interval that a country spent as a democracy and as an autocracy, respectively. Finally, in regressions based on annually repeated cross-country data, the relevant measures are the indicators for democracy and autocracy for the preceding year.

13. **Oil or gas reserve discovery:** A time-invariant indicator of at least one petroleum (oil or gas) reserve on the land territory of a country. This variable is based on information provided in the Petroleum Dataset (Version 1.2), covering the 1946–2003 time period (Lujala et al., 2007). Therefore, the available data does not provide information about any petroleum deposit discovered after 2003. The dataset is compiled for the main purpose of investigating the relationship between armed civil conflict and natural resources. Each on-shore petroleum (oil or gas) reserve – identified as polygons in the shapefile accompanying the dataset – is

assigned to a modern-day country based on the coordinates of the centroids of the deposit polygons. For additional information, the reader is referred to the codebook for Version 1.2 of the Petroleum Dataset, available from the Geographical and Resource Datasets online repository maintained by PRIO.

14. **Log population size:** The log-transformed size of a country's population, as reported by the World Bank's World Development Indicators (WDI) online data catalog.

    In cross-sectional regressions at the country level, the relevant measure is the log-transformed temporal average of annual population observations across all years in the 1960–2017 time period. In regressions using quinquennially repeated cross-country data, the relevant measure is the log-transformed temporal average of observations across all years in the preceding 5-year time interval. Finally, in regressions based on annually repeated cross-country data, the relevant measure is the log-transformed observation from the preceding year.

15. **Log GDP per capita:** The log-transformed per-capita GDP (in current US$) of a country, as reported by the World Bank's World Development Indicators (WDI) online data catalog.

    In cross-sectional regressions at the country level, the relevant measure is the log-transformed temporal average of annual per-capita GDP observations across all years in the 1960–2017 time period. In regressions using quinquennially repeated cross-country data, the relevant measure is the log-transformed temporal average of observations across all years in the preceding 5-year time interval. Finally, in regressions based on annually repeated cross-country data, the relevant measure is the log-transformed observation from the preceding year.

Other Control Variables (for Robustness Checks)

1. **Ecological fractionalization and polarization:** These measures of ecological diversity are motivated by Fenske (2014). The measure of *ecological fractionalization* is a Herfindahl index, constructed as

$$\text{Ecological fractionalization}_i = 1 - \sum_{t=1}^{t=18} (s_i^t)^2;$$

   and *ecological polarization* index is given by

$$\text{Ecological polarization}_i = 1 - \sum_{t=1}^{t=18} \left(\frac{0.5 - s_i^t}{0.5} \right)^2 s_i^t,$$

   where s_i^t is the share of the area of country i that is occupied by ecological type t. The polarization index measures the degree to which a country's area approximates a territory in which two ecological types each occupy half the total area. The relevant information on the spatial distribution of ecological types across the land surface of the earth is derived from global maps of agro-ecological zones from the Food and Agriculture Organization (FAO) of the United Nations.

2. **Mean and volatility of temperature and precipitation:** These four variables are constructed using information on mean temperature (in degree Celcius) per annum and total precipitation (in mm) per annum as reported by the Climate Research Unit (CRU) (Harris et al., 2014). Specifically, we employ the country-level spatial aggregates of annual mean

temperature and annual total precipitation, provided the CRU CY Version 4.01 dataset, which spans the 1901–2016 time period.

In cross-sectional regressions at the country level, the relevant measures of mean temperature and total precipitation reflect the temporal averages of the annual observations of these variables across all years in the 1960–2017 time period, whereas the corresponding volatility measures capture their respective temporal standard deviations during the same time span. In regressions using quinquennially repeated cross-country data, the relevant mean and volatility measures are similarly defined, except that the temporal averages and standard deviations are calculated across the years of the preceding 5-year time interval (rather than the full sample period). Finally, in regressions based on annually repeated cross-country data, the relevant measures are the one-year lags of annual mean temperature and annual total precipitation as well as the interannual standard deviations of temperature and precipitation over a 5-year rolling window that ends in the preceding year.

3. **Log years since Neolithic Revolution:** The log-transformed number of thousand years elapsed (as of the year 2000) since the majority of the population residing in a territory defined by a country's modern national borders began practicing sedentary agriculture as the primary mode of subsistence. This measure, initially reported by Putterman (2008), is compiled using a host of both region- and country-specific archaeological studies as well as more general encyclopedic works on the transition from hunting and gathering to agriculture during the Neolithic Revolution. The reader is referred to Putterman's website for a detailed description of the primary and secondary data sources employed in the construction of this variable.

4. **Log index of state antiquity:** The log-transformation of an index reflecting a country's cumulative experience with institutionalized statehood since antiquity. Specifically, we employ the State Antiquity Index (version 3.1), first introduced by Bockstette et al. (2002). The underlying index quantifies the exposure of a territory – as defined by a country's modern national borders – to formal statehood (i.e., being an independent nation-state or part of a larger kingdom or an empire) since the year 1 CE and until 1950. In particular, for each 50-year time interval, information on a territory's status with respect to the following 3 questions (each with specific weights applied) is employed: (i) is there a government above the tribal level?; (ii) is this government foreign or locally based?; and (iii) how much of the territory of the modern country was ruled by this government? These information are then aggregated over time to produce an index that ranges between 0 and 1. The reader is referred to Putterman's website for a detailed description of the methodology and data sources employed in the construction of this index.

5. **Log duration of human settlement:** The natural logarithm of the maximum duration (in tens of thousands of years) of uninterrupted settlement by anatomically modern humans across locations in a territory defined by a country's modern national borders. The underlying measure is obtained from the dataset of Ahlerup and Olsson (2012). The reader is therefore referred to that work for additional details on data sources and methodological assumptions.

6. **Log distance from regional frontier in 1500:** The great circle distance from a country's capital city to the closest regional technological frontier around the year 1500. The variable is obtained from the dataset of Ashraf and Galor (2013a). The set of regional frontiers

comprises the two most populous cities, reported for the year 1500 and belonging to different civilizations or sociopolitical entities, from each of Africa, Europe, Asia, and the Americas. Distances are calculated using the Haversine formula and are measured in kilometers. The historical urban population data used to identify the frontiers are sourced from Chandler (1987) and Modelski (2003), and the geographical coordinates of ancient urban centers are sourced from online resources such as Wikipedia.

7. **Ethnic inequality in luminosity:** A measure of intra-country economic inequality as captured by the subnational spatial distribution of per-capita adjusted nighttime luminosity in the year 2000 across the georeferenced homelands of ethnic groups. This measure is sourced from the replication dataset of Alesina et al. (2016). The reader is therefore referred to that work for additional details on data sources and methodological assumptions.

8. **Spatial inequality in luminosity:** A measure of intra-country economic inequality as captured by the subnational spatial distribution of per-capita adjusted nighttime luminosity in the year 2000 across 2.5×2.5-degree geospatial grid cells. This measure is sourced from the replication dataset of Alesina et al. (2016). The reader is therefore referred to that work for additional details on data sources and methodological assumptions.

9. **Linguistic fractionalization and polarization (georeferenced):** These are the country-level counterparts of the measures of linguistic fractionalization and polarization that are used in our analysis of conflicts at the ethnic homelands level. Specifically, these measures are constructed using georeferenced information on the spatial distribution of language homelands from the World Language Mapping System (WLMS) along with gridded population data from the Gridded Population of the World (GPW) dataset.

10. **Ethnic fractionalization (Fearon, 2003):** The ethnic fractionalization index compiled by Fearon (2003). The index reflects the probability that two individuals, randomly selected from a country's population, will belong to different ethnic groups.

11. **Linguistic fractionalization (Alesina et al., 2003):** The linguistic fractionalization index compiled by Alesina et al. (2003). The index reflects the probability that two individuals, randomly selected from a country's population, will belong to different linguistic groups.

12. **Religious fractionalization (Alesina et al., 2003):** The religious fractionalization index compiled by Alesina et al. (2003). The index reflects the probability that two individuals, randomly selected from a country's population, will belong to different religions.

13. **Ethnolinguistic fractionalization (Esteban et al., 2012):** An index of ethnolinguistic fractionalization, as represented by the *frac_fear* variable in the replication dataset of Esteban et al. (2012). The underlying ethnolinguistic population shares are sourced from Fearon (2003).

14. **Ethnolinguistic polarization (Esteban et al., 2012):** The Esteban-Ray index of ethnolinguistic polarization with $\delta = 0.05$, as represented by the *er_fear_delta005* variable in the replication dataset of Esteban et al. (2012). The underlying ethnolinguistic population shares are sourced from Fearon (2003).

15. **Gini index of ethnolinguistic diversity (Esteban et al., 2012):** The gini index of ethnolinguistic diversity per capita with $\delta = 0.05$, as represented by the variable named *gini_fear_delta005_PERCAPTA* in the replication dataset of Esteban et al. (2012). It is obtained after dividing the gini index of ethnolinguistic diversity by population size. The underlying ethnolinguistic population shares are sourced from Fearon (2003).

16. **Log percentage mountainous terrain:** The log-transformation of the proportion (in percentage) of a country's territory that is "mountainous" according to the codings of the geographer A.J. Gerard. This variable is sourced from the replication dataset of Fearon and Laitin (2003), where it is used to test the hypothesis that "rough terrain, poorly served by roads, at a distance from the centers of state power should favor insurgency and civil war."

17. **Noncontiguous state dummy:** A time-invariant indicator of whether a country possesses a territory with a population of at least 10,000 that is separated from the region containing its capital city either by land or 100 kilometers of water. This variable is sourced from the replication dataset of Fearon and Laitin (2003), where it is used to test the hypothesis that "the presence of a territory that is separated from the center of national governance by water or distance can help rebels more easily sustain insurgent activity and, thereby, make civil war more likely."

18. **Disease richness:** The total number of different types of infectious diseases in a country as reported by Fincher and Thornhill (2008), based on the Global Infectious Disease and Epidemiology Network (GIDEON; www.gideononline.com).

19. **Ethnic dominance:** A time-invariant indicator of whether the largest ethnic group in a country constitutes 45-90% of the national population. This variable is sourced from the replication dataset of Hegre and Sambanis (2006), but the primary source of the measure is Collier and Hoeffler (2004).

20. **Political instability:** A time-varying indicator at the country-year level of whether there was a change in the Polity IV regime index by 3 or more points in any of the three years prior to the country-year in question. Periods of regime transition (-88) and "interruptions" (indicating a complete collapse of central authority) are also coded as cases of political instability. Episodes of foreign occupation, however, are treated as missing observations. In robustness checks of our civil conflict onset regressions, the one-year lagged value of this variable is employed. This variable is sourced from the replication dataset of Hegre and Sambanis (2006), but the primary source is Fearon and Laitin (2003).

21. **New state dummy:** A time-varying indicator at the country-year level for whether the current year is the first year of the country's existence (e.g., as a newly independent state from colonial rule). In robustness checks of our civil conflict onset regressions, the one-year lagged value of this variable is employed. This variable is sourced from the replication dataset of Hegre and Sambanis (2006).

22. **Commodity export price shocks:** A set of four variables capturing different types of commodity export price shocks on an annual basis, sourced from the replication dataset of Bazzi and Blattman (2014). The first variable reflects *aggregate price shocks* and is computed

as the annual change in a country's log commodity export price index (a geometric average of all commodity export prices weighted by lagged export shares). The remaining variables reflect three types of disaggregated price shocks. The first of these reflects *annual crop price shocks*, i.e., price shocks to annual agricultural goods, such as oilseeds, food crops, and livestock, that are more likely to accrue to households. The second reflects *perennial crop price shocks*, i.e., price shocks to perennial tree crops like cocoa, coffee, rubber, or lumber. Finally, the third type of disaggregated price shocks captures *extractive crop price shocks*, i.e., price shocks to extractive products, namely, minerals, oil, and gas, that are more likely accrue to states. By construction, the sum of the three disaggregated types of shocks yields the *aggregate price shock* variable. In robustness checks of our civil conflict onset regressions, we employ the contemporaneous as well as the one- and two-year lagged values of these various commodity export price shock variables. For additional details, the reader is referred to Bazzi and Blattman (2014).

Supplement B Supplement to the Ethnicity-Level Analyses

B.1 Construction of the Georeferenced Ethnicity-Level Dataset

This research constructs a novel geo-referenced data set of population diversity for a large number of ethnic groups across the globe. Two measures are constructed: (i) a measure of genetic diversity for 207 ethnic homelands for all individuals covered in the Pemberton et al. (2013) dataset that can be mapped to an ethnic homeland, and (ii) a measure of predicted population diversity for 901 ethnic homelands covered in the Geo-Referencing of Ethnic Groups (GREG) map of Weidmann et al. (2010).

The geo-referenced dataset for observed genetic diversity maps all 10,386 linkable individuals in the Pemberton et al. (2013) dataset into their ethnic homelands. This mapping results in a sample of 207 ethnic homelands for which, in addition to the measure of genetic diversity, spatial characteristics (e.g., geographic, climatic, and societal attributes) are available. Furthermore, using data on the spatial distribution of language areas in conjunction with data on the spatial distribution of population sizes, the study generates measures of linguistic fractionalization and polarization for each ethnic homeland. Finally, using gridded PRIO data (PRIO-GRID version 1.01) as reported by Tollefsen et al. (2012) based on the UCDP/PRIO Armed Conflict Dataset (Gleditsch et al., 2002) as well as data on UCDP Georeferenced conflict events (Sundberg et al., 2012; Croicu and Sundberg, 2015) the study generates a range of measures of conflict within each ethnic homeland.

The mapping of the 10,386 linkable individuals in the Pemberton et al. (2013) dataset into their ethnic homelands was based on the individual's ethnic identity, location, and geographical coordinates, where the polygons for the ethnic homelands were based on (i) polygons found in Murdock (1959) and digitized by Nunn (2008); Nunn and Wantchekon (2011), (ii) the Handbook of North American Indians (Heizer, 1978), (iii) Global Mapping International's World Language Mapping System (WLMS) (see http://worldgeodatasets.com/language), (iv) the Geo-Referencing of Ethnic Groups (GREG) map of Weidmann et al. (2010), and (v) the Database of Global Administrative Areas (GADM) map version 3.6 (gadm.org).

The geo-referenced dataset for predicted predicted population diversity for 901 ethnic homelands covered in the Geo-Referencing of Ethnic Groups (GREG) map of Weidmann et al. (2010) is constructed based on the migratory distance from Addis Ababa in East Africa to the centroid of the homeland.[1]

B.2 Variable Definitions for the Ethnic-level Analyses

Conflict measures

1. **Conflict prevalence:** The average yearly share of the area of each ethnic homeland, over the period 1989–2008, that was within the boundaries of internal armed conflict event (between the government of a state and internal opposition groups). This measure is calculated using the gridded PRIO data (PRIO-GRID version 1.01) as reported by Tollefsen et al. (2012) based on the UCDP/PRIO Armed Conflict Dataset (Gleditsch et al., 2002).

2. **Number of conflict events:** The number of conflict events within each ethnic homeland in the UCDP Georeferenced Event Dataset covering the period 1989–2017 (Sundberg et al., 2012; Croicu and Sundberg, 2015).

[1] One homeland spanning territories in South America and Mauritius labeled "Indians of India and Pakistan" is excluded from the sample. The qualitative results would not be affected by the inclusion of this territory.

3. **Number of deaths:** The best (i.e., most likely) estimate of total fatalities resulting from a conflict event within each ethnic homeland in the UCDP Georeferenced Event Dataset covering the period 1989–2017 (Sundberg et al., 2012; Croicu and Sundberg, 2015).

4. **Number of deaths per event:** The number of deaths per event within each ethnic homeland in the UCDP Georeferenced Event Dataset covering the period 1989–2017 (Sundberg et al., 2012; Croicu and Sundberg, 2015).

Trust-related measures

1. **Intra-group trust (Africa):** The measure of an individual's trust in individuals from the same ethnic group in the 2005 Afrobarometer survey (3rd wave), as linked by Nunn and Wantchekon (2011) to the ethnicity names used in the Ethnographic Atlas. The measure takes the value 0 if the response to the question "How much do you trust each of the following types of people: People from your own ethnic group?" is "not at all", 1 if the response is "just a little", 2 if the value is "I trust them somewhat" and 3 if the value is "I trust them a lot".

2. **Slave exports (Africa):** A measure of the number of slaves taken from each ethnicity in transatlantic and Indian Ocean slave trades. The measure comes from Nunn and Wantchekon (2011) and is based on data from Nunn (2008).

3. **Other control variables (Africa):** The measures come from Nunn and Wantchekon (2011) and are based on data from 2005 Afrobarometer survey (3rd wave).

4. **Trust (US):** A measure of an individual's trust in people in general based on data from the General Social Survey 1972–2014 Release 6b Smith et al. (2018). The measure takes the value 1 if the response to the question "Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?" is "cannot trust", 2 if the response is "depends", and 3 if the value is "can trust".

Migratory distance and interpersonal population diversity

1. **Observed population diversity:** The expected heterozygosity (genetic diversity) of individuals in each of the 207 ethnic homelands, as calculated using Nei's formula (Nei, 1973), based on the individual-level data from Pemberton et al. (2013).

2. **Predicted population diversity:** The predicted level of population diversity of an ethnic homeland based on the migratory distance from East Africa to the centroid of the homeland, using the linear regression fit between observed population diversity and migratory distance from Addis Ababa obtained in sample of 207 ethnic homelands for which observed genetic diversity is available. The migratory distance from Addis is defined as the shortest traversable paths from Addis Ababa to the centroid of each ethnic group was computed. Given the limited ability of humans to travel across large bodies of water, the traversable area included bodies of water at a distance of 100km from land mass (excluding migration from Africa into Europe via Italy or Spain).[2]

[2]For the computation of predicted population diversity, distances to islands, where travel on water exceeds 100kms, are ignored since the Serial Founder Effect requires the serial foundation of populations along the migratory path and this was not feasible on water.

Control variables

1. **Linguistic fractionalization and polarization:** The degree of fractionalization in the ethnic homeland, using the formula $1 - \sum_i s_i^2$, and the degree of polarization in the ethnic homeland, using the formula $4 \sum_i s_i^2(1 - s_i)$, where s_i is an estimate of the population share of language group i in the homeland. Using the WLMS map of the spatial distribution of language areas in conjunction with the Gridded Population of the World dataset, the study estimates the number of individuals living in each intersection between ethnic homelands and language areas, assuming that population counts in overlapping language areas are equally split between these languages.

2. **Absolute latitude:** The absolute value of the latitude of an ethnic homeland's geodesic centroid, or, when the centroid is outside of the homeland, a representative interior point.

3. **Ruggedness:** The average level of the Terrain Ruggedness Index measure of Nunn and Puga (2012) across the grid cells that are located within a homeland.

4. **Mean and range of elevation:** The mean and range of elevation above sea level of an ethnic homeland, calculated using geospatial data from the *Atlas of the Biosphere* project (nelson.wisc.edu/sage/data-and-models/atlas/), across the grid cells that are located within a homeland.[3]

5. **Mean and range of land suitability:** The mean and range of the post-1500 optimal Caloric Suitability Index, measured by Galor and Özak (2016), across the grid cells that are located within a homeland.

6. **Island location:** A dummy variable indicating if the land type of an ethnic homeland's geodesic centroid (or a representative interior point) is a "small island" or a "very small island" as reported in the *World Countries* geographical dataset provided by ESRI (arcgis.com/home/item.html?id=ac80670eb213440ea5899bbf92a04998).

7. **Distance to nearest waterway:** The mean of the geodesic distance to the nearest coast or river, across the grid cells that are located within a homeland. Coastline locations are reported in the *Global Self-consistent, Hierarchical, High-resolution Geography Database* (http://soest.hawaii.edu/pwessel/gshhg). River locations are reported in the 1:10m *Natural Earth River + Lake Centerlines* dataset version 4 (http://naturalearthdata.com/downloads/10m-physical-vectors/10m-rivers-lake-centerlines).

8. **Temperature:** The mean of the daily average temperature (in degree Celcius), across the grid cells that are located within a homeland, based on data from the CRU TS dataset version 3.21 for the period 1901–2012, as reported by Climate Research Unit (CRU) (Harris et al., 2014).

9. **Precipitation:** The mean of the annual total precipitation (in mm), across the grid cells that are located within a homeland, based on data from the CRU TS dataset version 3.21 for the period 1901–2012, as reported by Climate Research Unit (CRU) (Harris et al., 2014).

[3]The mean elevation can be negative in some cases due to the existence of places on land with elevation below sea level or the inclusion of territories at sea in the homeland polygon, for which the elevation is negative.

10. **Time since settlement:** The earliest year with a positive population count estimate in the ethnic homeland. Specifically, the study employs the population count data from the *History Database of the Global Environment* dataset version 3.1 (themasites.pbl.nl/tridion/en/themasites/hyde/download/index-2.html), described in Klein Goldewijk et al. (2010, 2011).

11. **Malaria:** The mean level of plasmodium falciparum malaria endemicity in 2010, across the grid cells that are located within a homeland. Specifically, the current study employs the data on the age-standardised plasmodium falciparum Parasite Rate from Gething et al. (2011). It represents the estimated proportion of 2–10 year olds in the general population that are infected with plasmodium falciparum, averaged over the months of 2010. The estimates are based on data from parasite rate surveys and a geostatistical model that produces a range of predicted endemicities for each location. The model includes environmental covariates which improves the accuracy of the prediction. The environmental covariates include rainfall, temperature, land cover and urban/rural status. The endemicity data reports the mean value for the probability distribution at each location (approx. $1km^2$).

12. **Oil or gas reserve discovery:** A time-constant dummy for the presence of at least one petroleum (oil or gas) reserve on the territory of an ethnic homeland. The variable is based on information provided in the Petroleum Dataset (version 1.2, dated 2009) covering the period 1946–2003 (Lujala et al., 2007). The dataset is compiled for the main purpose of investigating the relationship between armed civil conflict and natural resources. Each on-shore petroleum reserve (oil or gas) – indicated as polygons in the shapefile accompanying the dataset – is assigned to an ethnic homeland using the coordinates of the centerpoints of the deposit polygons.

13. **Luminosity:** The mean level of cloud-free nighttime light intensity for the years 1992–2013, accross the grid cells that are located within a homeland. Specifically, the current study employs all available data in version 4 of the Defense Meteorological Satellite Program – Operational Linescan System (DMSP-OLS) Nighttime Lights Time Series (ngdc.noaa.gov/eog/dmsp/downloadV4composites.html). Since the log of zero is undefined, log luminosity is defined as the log of the sum of 0.001 and the luminosity measure.

B.3 Robustness Checks

TABLE SB.I: Population Diversity and Conflict across Ethnic Homelands – Robustness to Accounting for Alternative Distances

	Log conflict prevalence					
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS
Observed population diversity	28.338***	31.342***	30.591***			
	(9.622)	(9.692)	(9.735)			
Predicted population diversity				73.828***	70.194***	75.334***
				(7.390)	(7.313)	(7.305)
Distance to Technological Frontier in Year 1 (in 1000 kms)	-0.045			-0.172***		
	(0.163)			(0.066)		
Distance to Technological Frontier in Year 1000 (in 1000 kms)		-0.324*			-0.268***	
		(0.168)			(0.062)	
Distance to Technological Frontier in Year 1500 (in 1000 kms)			-0.210			-0.124**
			(0.148)			(0.061)
Ethnolinguistic fractionalization	1.633	1.446	1.474	0.279	0.340	0.330
	(1.219)	(1.171)	(1.196)	(0.383)	(0.381)	(0.381)
Ethnolinguistic polarization	-0.353	-0.213	-0.237	0.332	0.315	0.296
	(1.029)	(0.990)	(1.010)	(0.348)	(0.344)	(0.347)
Regional dummies	Yes	Yes	Yes	Yes	Yes	Yes
Geographical controls	Yes	Yes	Yes	Yes	Yes	Yes
Climatic controls	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Observed	Observed	Observed	Predicted	Predicted	Predicted
Observations	207	207	207	901	901	901
Effect of 10th90th %ile move in diversity	0.443***	0.490***	0.478***	1.639***	1.558***	1.672***
	(0.150)	(0.152)	(0.152)	(0.164)	(0.162)	(0.162)
First-stage F statistic						
Adjusted R^2	0.304	0.316	0.310	0.367	0.375	0.365
β^*	26.359	28.224	29.899	80.379	77.719	77.280

Notes: This table exploits variations across ethnic homelands to establish a significant positive impact of observed and predicted population diversity on the log conflict prevalence during the 1989–2008 period, conditional on migratory distances from historical technological frontiers as well as the baseline geographical characteristics. The set of continent and regional dummies includes indicators for Europe, Asia, North America, South America, Oceania, North Africa, and Sub-Saharan Africa. Additional climatic covariates refer to the average diurnal temperature range, average cloud cover, and average temperature range in the homeland. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its distribution is expressed in terms of the change in the prevalence of conflicts within the territory of a homeland over the years 1989–2008. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SB.II: Observed Population Diversity and Conflict across Ethnic Homelands – Robustness to Accounting for Measures of Ecological Diversity

	Log conflict prevalence						
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS
Observed population diversity	27.700***	32.958***	24.748***	25.591***	24.996***	26.869**	26.325**
	(10.372)	(10.482)	(9.315)	(9.313)	(9.287)	(10.427)	(10.425)
Ecological diversity	-0.838	-0.637	1.029	0.748	0.909	0.733	0.843
	(1.430)	(1.595)	(1.429)	(1.418)	(1.414)	(1.384)	(1.379)
Ecological polarization	0.942	1.103	0.675	0.702	0.687	1.006	1.009
	(1.141)	(1.228)	(1.065)	(1.045)	(1.054)	(1.024)	(1.025)
Ethnolinguistic fractionalization				1.140*		0.893	
				(0.636)		(0.652)	
Ethnolinguistic polarization					0.734		0.641
					(0.527)		(0.530)
Regional dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographical controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Climatic controls	No	No	Yes	Yes	Yes	Yes	Yes
Development outcomes	No	No	No	No	No	Yes	Yes
Disease environment controls	No	No	No	No	No	Yes	Yes
Sample	Observed	Observed	Observed	Observed	Observed	Observed	Observed
Observations	205	205	205	205	205	205	205
Effect of 10th-90th %ile move in diversity	0.433***	0.515***	0.387***	0.400***	0.391***	0.420***	0.411**
	(0.162)	(0.164)	(0.146)	(0.146)	(0.145)	(0.163)	(0.163)
Adjusted R^2	0.106	0.168	0.308	0.317	0.312	0.330	0.328
β^*		37.005	23.299	24.574	23.683	26.483	25.685

Notes: This table exploits cross-ethnicity variations to establish a significant positive impact of contemporary population diversity on the log spatio-temporal prevalence of UCDP/PRIO conflicts during the 1989–2008 period, conditional on ecological diversity and ecological polarization as well as the baseline control variables. The set of continent and regional dummies includes indicators for Europe, Asia, North America, South America, Oceania, North Africa, and Sub-Saharan Africa. Additional climatic covariates refer to the average diurnal temperature range, average cloud cover, and average temperature range in the homeland. The 2SLS regressions exploit prehistoric migratory distance from East Africa to each ethnic homeland as an excluded instrument for the observed population diversity of this ethnic group. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the change in the average yearly share of the area of each ethnic homeland that was within the boundaries of internal armed conflict over the period 1989–2008. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SB.III: Predicted Population Diversity and the Spatiotemporal Prevalence of Conflict across Ethnic Homelands – Robustness to Accounting for Measures of Ecological Diversity

	Log conflict prevalence						
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) 2SLS
Predicted population diversity	77.597***	79.803***	76.148***	75.668***	77.910***	77.646***	
	(6.245)	(7.314)	(7.425)	(7.458)	(9.700)	(9.807)	
Observed population diversity							130.105***
							(33.284)
Ecological diversity	0.711	0.808	1.064*	1.070*	1.565**	1.496**	-0.078
	(0.631)	(0.638)	(0.629)	(0.634)	(0.714)	(0.719)	(1.722)
Ecological polarization	0.396	0.466	0.317	0.299	-0.455	-0.435	0.263
	(0.587)	(0.541)	(0.533)	(0.536)	(0.596)	(0.599)	(1.233)
Ethnolinguistic fractionalization			0.341		0.174		
			(0.300)		(0.354)		
Ethnolinguistic polarization				0.450*		0.565*	
				(0.267)		(0.315)	
Regional dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographical controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Climatic controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Development outcomes	No	No	Yes	Yes	Yes	Yes	No
Disease environment controls	No	No	Yes	Yes	Yes	Yes	No
Sample	Predicted	Predicted	Predicted	Predicted	Old World	Old World	Observed
Observations	891	891	891	891	697	697	205
Effect of 10th-90th %ile move in diversity	1.748***	1.797***	1.715***	1.704***	0.976***	0.972***	2.034***
	(0.141)	(0.165)	(0.167)	(0.168)	(0.121)	(0.123)	(0.520)
Adjusted R^2	0.207	0.365	0.381	0.382	0.406	0.409	
β^*		81.333	75.203	74.414	69.099	68.719	
Migratory distance from East Africa (in 10,000 km)							-0.043***
							(0.009)
First-stage F-statistic							23.605

Notes: This table exploits cross-ethnicity variations to establish a significant positive impact of predicted population diversity on the log spatio-temporal prevalence of UCDP/PRIO conflicts during the 1989–2008 period, conditional on ecological diversity and ecological polarization as well as the baseline control variables. The set of continent and regional dummies includes indicators for Europe, Asia, North America, South America, Oceania, North Africa, and Sub-Saharan Africa. Additional climatic covariates refer to the average diurnal temperature range, average cloud cover, and average temperature range in the homeland. The 2SLS regressions exploit prehistoric migratory distance from East Africa to each ethnic homeland as an excluded instrument for the observed population diversity of this ethnic group. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the change in the average yearly share of the area of each ethnic homeland that was within the boundaries of internal armed conflict over the period 1989–2008. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

B.4 Descriptive Statistics for the Trust Analyses

TABLE SB.IV: Summary Statistics

			Percentile		
	Mean	SD	10th	90th	N
PANEL A	African sample				
Intra-group trust	1.52	1.00	0.00	3.00	3,212
Population diversity (observed)	0.76	0.00	0.76	0.77	3,212
Age	35.82	14.54	20.00	58.00	3,212
Male	0.49	0.50	0.00	1.00	3,212
Ethnic fractionalization	0.27	0.28	0.00	0.72	3,212
Ethnolinguistic polarization	0.53	0.13	0.30	0.62	3,212
Proportion of ethnic group in district	0.73	0.33	0.12	1.00	3,212
School present	0.84	0.37	0.00	1.00	3,208
Electricity present	0.65	0.48	0.00	1.00	3,210
Piped water present	0.44	0.50	0.00	1.00	3,157
Sewage present	0.23	0.42	0.00	1.00	3,054
Health clinic present	0.58	0.49	0.00	1.00	3,060
Living in an urban area	0.44	0.50	0.00	1.00	3,212
Living condition categories	2.65	1.25	1.00	4.00	3,206
Education categories	3.51	2.10	0.00	6.00	3,207
Occupation categories	18.92	92.10	1.00	23.00	3,201
Religion categories	10.52	51.36	2.00	12.00	3,204
Slave exports (Atlantic and Indian)	277.44	262.45	0.17	665.97	3,212
PANEL B	US sample				
Trust	1.88	0.97	1.00	3.00	2,294
Population diversity (predicted)	0.72	0.02	0.67	0.74	2,294
GSS year	1993.94	10.59	1980.00	2010.00	2,294
Age	54.37	19.46	27.00	80.00	2,284
Sex	1.55	0.50	1.00	2.00	2,294
Family income categories	2.73	0.89	2.00	4.00	1,803
Religion categories	2.02	1.29	1.00	3.00	2,283
Highest educational degree categories	1.30	1.20	0.00	3.00	2,290
Ethnic fractionalization (ancestral)	0.23	0.18	0.11	0.54	2,294
Ethnolinguistic polarization (ancestral)	0.41	0.21	0.12	0.67	2,294
Absolute latitude (ancestral)	46.07	11.82	23.00	60.00	2,294
Ruggedness (ancestral)	131.80	94.05	30.64	237.76	2,294
Mean elevation (ancestral)	436.42	339.34	105.77	1015.28	2,294
Mean land suitability (ancestral)	0.48	0.21	0.10	0.75	2,294
Range of land suitability (ancestral)	0.92	0.12	0.82	1.00	2,294
Distance to nearest waterway (ancestral)	223.00	496.37	29.43	332.58	2,294

Supplemental References

Ahlerup, P. and O. Olsson (2012): "The Roots of Ethnic Diversity," *Journal of Economic Growth*, 17, 71–102.

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg (2003): "Fractionalization," *Journal of Economic Growth*, 8, 155–194.

Alesina, A., S. Michalopoulos, and E. Papaioannou (2016): "Ethnic Inequality," *Journal of Political Economy*, 124, 428–488.

Ashraf, Q. and O. Galor (2013a): "The "Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic Development," *American Economic Review*, 103, 1–48.

Banks, A. S. and K. A. Wilson (2018): "Cross-National Time-Series Data Archive [Data file]," Databanks International, Jerusalem, Israel. https://www.cntsdata.com/.

Bazzi, S. and C. Blattman (2014): "Economic Shocks and Conflict: Evidence from Commodity Prices," *American Economic Journal: Macroeconomics*, 6, 1–38.

Birnir, J. K., D. D. Laitin, J. Wilkenfeld, D. M. Waguespack, A. S. Hultquist, and T. R. Gurr (2018): "Introducing the AMAR (All Minorities at Risk) Data," *Journal of Conflict Resolution*, 62, 203–226.

Birnir, J. K., J. Wilkenfeld, J. D. Fearon, D. D. Laitin, T. R. Gurr, D. Brancati, S. M. Saideman, A. Pate, and A. S. Hultquist (2015): "Socially Relevant Ethnic Groups, Ethnic Structure, and AMAR," *Journal of Peace Research*, 52, 110–115.

Bockstette, V., A. Chanda, and L. Putterman (2002): "States and Markets: The Advantage of an Early Start," *Journal of Economic Growth*, 7, 347–369.

Brecke, P. (1999): "Violent Conflicts 1400 A.D. to the Present in Different Regions of the World," Paper presented at the 1999 Annual Meeting of the Peace Science Society, October 8–10.

Burke, M., S. M. Hsiang, and E. Miguel (2015): "Climate and Conflict," *Annual Review of Economics*, 7, 577–617.

Burke, M. B., E. Miguel, S. Satyanath, J. A. Dykema, and D. B. Lobell (2009): "Warming Increases the Risk of Civil War in Africa," *Proceedings of the National Academy of Sciences*, 106, 20670–20674.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011): "Robust Inference With Multiway Clustering," *Journal of Business & Economic Statistics*, 29, 238–249.

Central Intelligence Agency (2018): "The World Factbook," The Central Intelligence Agency, Washington, DC. Data retrieved at https://www.cia.gov/library/publications/the-world-factbook/.

Cervellati, M., U. Sunde, and S. Valmori (2017): "Pathogens, Weather Shocks and Civil Conflicts," *Economic Journal*, 127, 2581–2616.

Chandler, T. (1987): *Four Thousand Years of Urban Growth: An Historical Census*, Lewiston, NY: The Edwin Mellen Press.

Cioffi-Revilla, C. (1996): "Origins and Evolution of War and Politics," *International Studies Quarterly*, 40, 1–22.

Collier, P. and A. Hoeffler (2004): "Greed and Grievance in Civil War," *Oxford Economic Papers*, 56, 563–595.

Conley, T. G. (1999): "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, 92, 1–45.

Croicu, M. and R. Sundberg (2015): "UCDP Georeferenced Event Dataset Codebook version 4.0," Department of Peace and Conflict Research, Uppsala University. http://ucdp.uu.se/downloads/ged/ucdp-ged-40-codebook.pdf.

Desmet, K., I. Ortuño-Ortín, and R. Wacziarg (2012): "The Political Economy of Linguistic Cleavages," *Journal of Development Economics*, 97, 322–338.

Dincecco, M., J. Fenske, and M. G. Onorato (2015): "Is Africa Different? Historical Conflict and State Development," IMT Lucca EIC Working Paper No. 08/2015, IMT Institute for Advance Studies Lucca.

Esteban, J., L. Mayoral, and D. Ray (2012): "Ethnicity and Conflict: An Empirical Study," *American Economic Review*, 102, 1310–1342.

Fearon, J. D. (2003): "Ethnic and Cultural Diversity by Country," *Journal of Economic Growth*, 8, 195–222.

Fearon, J. D. and D. D. Laitin (2003): "Ethnicity, Insurgency, and Civil War," *American Political Science Review*, 97, 75–90.

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

Fincher, C. L. and R. Thornhill (2008): "Assortative Sociality, Limited Dispersal, Infectious Disease and the Genesis of the Global Pattern of Religion Diversity," *Proceedings of the Royal Society B: Biological Sciences*, 275, 2587–2594.

Gallup, J. L., J. D. Sachs, and A. D. Mellinger (1999): "Geography and Economic Development," *International Regional Science Review*, 22, 179–232.

Galor, O. and Ö. Özak (2016): "The Agricultural Origins of Time Preference," *American Economic Review*, 106, 3064–3103.

Gething, P. W., A. P. Patil, D. L. Smith, C. A. Guerra, I. R. Elyazar, G. L. Johnston, A. J. Tatem, and S. I. Hay (2011): "A New World Malaria Map: *Plasmodium falciparum* Endemicity in 2010," *Malaria Journal*, 10.

Gleditsch, N. P., P. Wallensteen, M. Eriksson, M. Sollenberg, and H. Strand (2002): "Armed Conflict 1946-2001: A New Dataset," *Journal of Peace Research*, 39, 615–637.

Harris, I., P. D. Jones, T. J. Osborn, and D. H. Lister (2014): "Updated High-Resolution Grids of Monthly Climatic Observations – The CRU TS3.10 Dataset," *International Journal of Climatology*, 34, 623–642.

Harris, I. C. and P. D. Jones (2013): "CRU TS3.21: Climatic Research Unit (CRU) Time-Series (TS) Version 3.21 of High Resolution Gridded Data of Month-by-Month Variation in Climate (Jan. 1901 - Dec. 2012) [Data file]," University of East Anglia Climatic Research Unit, NCAS British Atmospheric Data Centre, 24 September 2013. doi:10.5285/D0E1585D-3417-485F-87AE-4FCECF10A992.

——— (2017): "CRU CY4.01: Climatic Research Unit (CRU) Year-by-Year Variation of Selected Climate Variables by Country (CY) version 4.01 (Jan. 1901 - Dec. 2016) [Data file]," University of East Anglia Climatic Research Unit, Centre for Environmental Data Analysis, 4 December 2017. doi:10.5285/d4e823f0172947c5ae6e6b265656c273.

Hastings, D. A., P. K. Dunbar, G. M. Elphingstone, M. Bootz, H. Murakami, H. Maruyama, H. Masaharu, P. Holland, J. Payne, N. A. Bryant, et al. (1999): "The Global Land One-kilometer Base Elevation (GLOBE) Digital Elevation Model, Version 1.0," National Oceanic and Atmospheric Administration, National Geophysical Data Center, Boulder, CO. Data retrieved at https://www.ngdc.noaa.gov/mgg/topo/globe.html.

Hegre, H. and N. Sambanis (2006): "Sensitivity Analysis of Empirical Results on Civil War Onset," *Journal of Conflict Resolution*, 50, 508–535.

Heizer, R. F. (1978): *Handbook of North American Indians, Vol. 8: California*, Washington, DC: Smithsonian Institution.

Hsiang, S. M., M. Burke, and E. Miguel (2013): "Quantifying the Influence of Climate on Human Conflict," *Science*, 341, 1235367/1–14.

King, G. and L. Zeng (2001): "Logistic Regression in Rare Events Data," *Political Analysis*, 9, 137–163.

Klein Goldewijk, K., A. Beusen, and P. Janssen (2010): "Long-Term Dynamic Modeling of Global Population and Built-Up Area in a Spatially Explicit Way: HYDE 3.1," *The Holocene*, 20, 565–573.

Klein Goldewijk, K., A. Beusen, G. van Drecht, and M. de Vos (2011): "The HYDE 3.1 Spatially Explicit Database of Human-Induced Global Land-Use Change Over the Past 12,000 Years," *Global Ecology and Biogeography*, 20, 73–86.

La Porta, R., F. Lopez-de-Silanes, A. Shleifer, and R. Vishny (1999): "The Quality of Government," *Journal of Law, Economics, and Organization*, 15, 222–279.

Lujala, P., J. Ketil Rod, and N. Thieme (2007): "Fighting over Oil: Introducing a New Dataset," *Conflict Management and Peace Science*, 24, 239–256.

Marshall, M. G. (1999): *Third World War*, Lanham, MD: Rowman & Littlefield Publishers.

——— (2002): "Measuring the Societal Impact of War," in *From Reaction to Conflict Prevention: Opportunities for the UN System*, ed. by F. O. Hampson and D. M. Malone, Boulder, CO: Lynne Rienner, 63–105.

——— (2017): "Major Episodes of Political Violence (MEPV) and Conflict Regions, 1946–2017," Center for Systemic Peace, Vienna, VA. Data retrieved at http://www.systemicpeace.org/inscrdata.html.

Marshall, M. G., T. R. Gurr, and K. Jaggers (2017): "Polity IV Project: Political Regime Characteristics and Transitions, 1800–2017," Center for Systemic Peace, Vienna, VA. Data retrieved at http://www.systemicpeace.org/inscrdata.html.

Michalopoulos, S. (2012): "The Origins of Ethnolinguistic Diversity," *American Economic Review*, 102, 1508–1539.

Modelski, G. (2003): *World Cities: -3000 to 2000*, Washington, DC: FAROS 2000.

Montalvo, J. G. and M. Reynal-Querol (2005): "Ethnic Polarization, Potential Conflict, and Civil Wars," *American Economic Review*, 95, 796–816.

Murdock, G. P. (1959): *Africa: Its Peoples and Their Culture History*, New York, NY: McGraw-Hill Book Co., Inc.

Nei, M. (1973): "Analysis of Gene Diversity in Subdivided Populations," *Proceedings of the National Academy of Sciences*, 70, 3321–3323.

New, M., D. Lister, M. Hulme, and I. Makin (2002): "A High-Resolution Data Set of Surface Climate Over Global Land Areas," *Climate Research*, 21, 1–25.

Nordhaus, W. D. (2006): "Geography and Macroeconomics: New Data and New Findings," *Proceedings of the National Academy of Sciences*, 103, 3510–3517.

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

Nunn, N. and D. Puga (2012): "Ruggedness: The Blessing of Bad Geography in Africa," *Review of Economics and Statistics*, 94, 20–36.

Nunn, N. and L. Wantchekon (2011): "The Slave Trade and the Origins of Mistrust in Africa," *American Economic Review*, 101, 3221–3252.

Özak, Ö. (2010): "The Voyage of Homo-Economicus: Some Economic Measures of Distance," Unpublished manuscript. Department of Economics, Southern Methodist University.

Pemberton, T. J., M. DeGiorgio, and N. A. Rosenberg (2013): "Population Structure in a Comprehensive Genomic Data Set on Human Microsatellite Variation," *G3: Genes, Genomes, and Genetics*, 3, 891–907.

Pettersson, T. and K. Eck (2018): "Organized Violence, 1989–2017," *Journal of Peace Research*, 55, 535–547.

Putterman, L. (2008): "Agriculture, Diffusion, and Development: Ripple Effects of the Neolithic Revolution," *Economica*, 75, 729–748.

Putterman, L. and D. N. Weil (2010): "Post-1500 Population Flows and The Long-Run Determinants of Economic Growth and Inequality," *Quarterly Journal of Economics*, 125, 1627–1682.

Ramachandran, S., O. Deshpande, C. C. Roseman, N. A. Rosenberg, M. W. Feldman, and L. L. Cavalli-Sforza (2005): "Support from the Relationship of Genetic and Geographic Distance in Human Populations for a Serial Founder Effect Originating in Africa," *Proceedings of the National Academy of Sciences*, 102, 15942–15947.

Ramankutty, N., J. A. Foley, J. Norman, and K. McSweeney (2002): "The Global Distribution of Cultivable Lands: Current Patterns and Sensitivity to Possible Climate Change," *Global Ecology and Biogeography*, 11, 377–392.

Reynal-Querol, M. (2002): "Ethnicity, Political Systems, and Civil Wars," *Journal of Conflict Resolution*, 46, 29–54.

Riley, S. J., S. D. DeGloria, and R. Elliot (1999): "A Terrain Ruggedness Index that Quantifies Topographic Heterogeneity," *Intermountain Journal of Sciences*, 5, 23–27.

Smith, T. W., M. Davern, J. Freese, and S. L. Morgan (2018): "General Social Surveys, 1972–2018 [Data file]," National Data Program for the Social Sciences, Chicago, IL. Data retrieved at gss.norc.org.

Sundberg, R., K. Eck, and J. Kreutz (2012): "Introducing the UCDP Non-State Conflict Dataset," *Journal of Peace Research*, 49, 351–362.

Tollefsen, A. F., H. Strand, and H. Buhaug (2012): "PRIO-GRID: A Unified Spatial Data Structure," *Journal of Peace Research*, 49, 363–374.

Weidmann, N. B., J. K. Rød, and L.-E. Cederman (2010): "Representing Ethnic Groups in Space: A New Dataset," *Journal of Peace Research*, 47, 491–499.

World Bank (2018): "World Development Indicators," The World Bank, Washington, DC. Data retrieved at https://datacatalog.worldbank.org/dataset/world-development-indicators.

World Values Survey (2006): "European and World Values Surveys, Four-Wave Integrated Data File, 1981–2004, version 20060423," The World Values Survey Association, Stockholm, Sweden. Data retrieved at http://www.worldvaluessurvey.org.

——— (2009): "World Values Survey, 1981–2008 Official Aggregate, version 20090914," The World Values Survey Association, Stockholm, Sweden. Data retrieved at http://www.worldvaluessurvey.org.