

Supplementary Material

Census Enumeration and Group Conflict: A Global Analysis of the
Consequences of Counting

World Politics, Vol. 69, No. 1 January 2017

doi: 10.1017/S0043887116000198

Evan S. Lieberman*

Prerna Singh†

*Professor, Department of Political Science, MIT.

†Assistant Professor, Department of Political Science, Brown University.

Contents

1	Descriptive statistics	2
2	Ordered Logit estimates of Vanhanen and MAR violence data	3
3	Model estimates using alternate measures of Politically relevant ethnic groups	4
4	Estimates conditional on various measures of ethnic diversity	8
5	Sensitivity to removal of country cases, regions	13
6	Rare events logit estimates	17
7	Cross-cuttingness of ethnic cleavags	18
8	Missing census questionnaires	21

1 Descriptive statistics

Table 1: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Annual (1946-2005)					
Ethnic enumeration, dummy (authors)	0.736	0.441	0	1	6638
Ethnic enumeration, total (authors)	1.462	1.216	0	5	6638
Race enumeration, dummy (authors)	0.194	0.395	0	1	6638
Tribe/ethnic enumeration, dummy (authors)	0.265	0.441	0	1	6638
Religion enumeration, dummy (authors)	0.476	0.499	0	1	6638
Language enumeration, dummy (authors)	0.399	0.49	0	1	6638
Caste enumeration, dummy (authors)	0.023	0.151	0	1	6638
Indigenous enumeration, dummy (authors)	0.105	0.307	0	1	6638
L.GDP/capita (PWT/EPR)	6.194	7.371	0.134	110.315	6534
L.Ln population (WB)	9.217	1.393	5.581	14.076	6565
Years since independence	56.399	49.581	0	189	6638
Political instability (Polity/EPR)	0.124	0.329	0	1	6638
L.Anocracy (Polity/EPR)	0.223	0.416	0	1	6494
Polity regime score (Polity/EPR)	0.352	7.514	-10	10	6420
Cross-cutting Lang x Relig (Sel)	0.772	0.179	0.054	1	4939
Enumeration of Lang and Relig (authors)	0.875	0.781	0	2	6638
L.Oil per capita (EPR)	2.12	13.51	0	272.403	6565
Ln ethnic excluded population share (EPR)	1.855	1.58	0	4.595	6621
N ethnopolitically relevant groups (EPR)	4.199	6.509	0	57	6621
Year of new ethnic war (EPR)	0.016	0.125	0	1	6638
N Peace years – since last war or 1946 (EPR)	15.916	15.531	0	59	6638
Decade (1940s - 1990s)					
Communal conflict dummy (MAR)	0.138	0.345	0	1	509
Communal conflict index (MAR)	1.194	1.967	0	6	509
Time invariant					
Violent ethnic conflict index (Van)	23.079	25.026	0	100	151
Violent ethnic conflict dummy (Van)	0.265	0.443	0	1	151
Institutionalized ethnic conflict index (Van)	29.795	24.065	0	100	151
Ethnic political conflict dummy (Van)	0.775	0.419	0	1	151
Ethnolinguistic Fractionalization (ANM/EPR)	0.412	0.278	0.004	0.925	145
Religious Fractionalization (EPR)	0.378	0.218	0	0.783	145
Ethnic Fractionalization (Al/Qog)	0.468	0.252	0.002	0.93	152
Linguistic Fractionalization (Al/Qog)	0.413	0.287	0.002	0.923	148
Religious Fractionalization (Al/Qog)	0.436	0.233	0.002	0.86	152
Politically relevant eth groups, binary (Pos)	0.846	0.366	0	1	39

Al: Alesina et al. (2003); ANM: Bruk and Apenchecko (1964); EPR: Wimmer, Cederman and Min (2009); MAR: Center for International Development and Conflict Management (2009); Pol: Marshall and Jagers (2009); Pos: Posner (2004); PWT: Penn World Tables; Sel: Selway (2010); Qog: Teorell et al. (2011); Van: Vanhanen (1999). Population data originally in millions. GDP/Capita data in constant 2000 dollars. L. indicates lagged.

2 Ordered Logit estimates of Vanhanen and MAR violence data

As discussed in the paper, we convert the Vanhanen and MAR scales of ethnic violence into binary variables for clarity of interpretation, and also because of concerns about the reliability of more subtle classifications. However, those decisions and the meaningful interpretation of an investigator-imposed scale are arbitrary. Here, we replicate the model 1 estimates of these two violence indicators using ordered logit estimates. Both analyses confirm the positive relationship between Ethnic enumeration and the likelihood of experiencing increasingly more severe manifestations of ethnic violence following the enumeration of ethnic categories.

Table 2: Ordered Logit Estimates of the effect of Ethnic Enumeration on Ethnic Violence

	Violent ethnic conflict index (Van)	Communal conflict index (MAR)
L1.Ethnic enumeration	1.253* (0.360)	0.608* (0.290)
L10.Logged GDP/capita	-0.0882* (0.0430)	-0.0145 (0.0226)
L1.Ln population	0.315* (0.141)	0.625* (0.155)
Years since independence	0.00432 (0.00426)	0.00123 (0.00486)
Observations	142	492

Standard errors in parentheses

L. indicates lagged one year

+ $p < 0.10$, * $p < 0.05$

3 Model estimates using alternate measures of Politically relevant ethnic groups

In the main text, we estimate the effect of enumeration on the number of politically relevant ethnic groups. Here we consider two other strategies in order to test for robustness of specification and approach to measurement:

First, in table 3, we replicate the analyses from table 1 in the main text, but here, we convert the outcome variable – *number* of politically relevant ethnic groups – to a simple binary variable which takes a value of 1 if there are *any* politically relevant ethnic groups, and a value of 0 otherwise.

Next, in table 4, we use the (Posner 2004, 854-6) classification of politically relevant ethnic group fractionalization in Africa. He calculated fractionalization scores for all “politically relevant” ethnic groups in sub-Saharan Africa based on the appearance of consistent descriptions of ethnic groups competing over economic resources within a wide range of newspapers, periodicals, and other printed materials. We simply recode as 1 all countries with positive values on that measure because it indicates that there is at least one relevant ethnic group. Although this is only a cross-section of data for just one region, we still find that Ethnic enumeration is a good predictor of which countries will (not) have politically relevant ethnic groups.

Finally, in table 5, we use the Vanhanen measure of Ethnic political conflict, which classifies countries in terms of the extent to which ethnicity is relevant in political parties, government and everyday discrimination. Our OLS estimates are generally consistent with all previous results, though again, once we include a fuller set of covariates, including a measure of ethnic fractionalization, the size of the estimated coefficient is attenuated.

Table 3: Effect of Ethnic Enumeration on Existence of Politically Relevant Ethnic Groups (EPR), Logit Estimates (1946-2005)

	Any politically relevant ethnic groups (EPR)?				
Ethnic enumeration (any) max10	1.207** (0.355)	1.284** (0.387)	1.260** (0.391)	1.105** (0.382)	1.030* (0.477)
Log GDP/cap lag10	-0.028 (0.029)	-0.012 (0.026)	0.0001 (0.027)	0.012 (0.031)	-0.009 (0.029)
Number languages			0.088 (0.081)		
Log pop lag1	0.412* (0.195)	0.411* (0.190)	0.371 ⁺ (0.191)	1.211 ⁺ (0.735)	0.496* (0.225)
British col.		-0.348 (0.603)	-0.573 (0.572)	-0.844 (1.071)	-1.483 ⁺ (0.771)
French col.		0.491 (0.860)	0.251 (0.818)	0.780 (0.834)	-0.666 (0.976)
Ethnic Fractionalization (AL)					2.404 (1.748)
Religious Fractionalization (AL)					1.381 (1.260)
Linguistic Fractionalization (AL)					2.114 (1.791)
Anocracy		0.232 (0.325)	0.203 (0.324)	-0.064 (0.428)	0.230 (0.393)
Instability		0.869* (0.355)	0.869* (0.367)	0.349 (0.405)	0.795 (0.499)
Polity		-0.020 (0.031)	-0.013 (0.031)	0.027 (0.040)	-0.011 (0.031)
Absolute latitude				-1.071 (1.508)	
Variation in elevation				0.466 (0.762)	
Variation in land quality				-0.144 (0.359)	
Mean elevation				0.046 (0.699)	
Mean land quality				-0.128 (0.405)	
Mean precipitation				0.427 (0.630)	
Mean temperature				-0.614 (1.366)	
Log area				-1.482 (1.100)	
Distance from Sea				1.667 (1.376)	
Migratory distance from E.Africa				-0.051 (0.505)	
Log Population density in 1995				-0.484 (0.457)	
Log Population density in 1500				-0.325 (0.802)	
Timing transition to agriculture				0.797* (0.383)	
Years since independence	-0.004 (0.005)	-0.003 (0.006)	-0.002 (0.006)	-0.011 (0.008)	-0.005 (0.007)
Constant	-2.654 (1.691)	-2.911 ⁺ (1.689)	-2.913 ⁺ (1.762)	-8.163 (5.615)	-4.610* (2.203)
N	635	627	627	580	590
Log Likelihood	-263.247	-255.163	-249.137	-183.898	-192.944
AIC	546.494	540.326	530.273	423.796	421.888

** p < .01; * p < .05; ⁺ p < .1

Country clustered standard errors in parentheses. Decade controls not shown.

Table 4: Effect of Ethnic Enumeration on Political Relevance Using Posner Data: Logit Estimates

	Any politically relevant ethnic groups 1990 decade (Posner)?			
Ethnic enumeration (any)	4.711** (1.776)	3.917* (1.768)		
Ethnic enumeration (total)			4.466* (2.148)	3.452+ (2.058)
Log GDP/cap (lag10)	0.436 (0.584)	0.237 (0.635)	1.022 (0.994)	0.556 (1.076)
Ethnic Fractionalization		4.162+ (2.485)		3.126 (2.725)
Log pop	1.298 (0.798)	0.991 (0.956)	2.313 (1.505)	1.561 (1.577)
Time since independence	0.003 (0.079)	-0.001 (0.106)	0.021 (0.212)	0.022 (0.234)
Constant	-12.981+ (7.711)	-11.824 (9.119)	-24.052 (15.253)	-17.923 (15.763)
N	38	38	38	38
Log Likelihood	-7.816	-6.290	-5.957	-5.269
AIC	25.631	24.579	21.913	22.539

**p < .01; *p < .05; +p < .1
Standard errors in parentheses.

Table 5: Effect of Ethnic Enumeration on Ethnic Political Conflict Using Vanhanen Data: OLS Estimates

	Ethnic Political Conflict Index 1990-6 (Vanhanen)	
Ethnic enumeration, any (lag1)	0.137** (0.041)	0.083* (0.040)
Log GDP/cap (lag10)	-0.005+ (0.003)	-0.001 (0.003)
Log population (lag1)	0.005 (0.015)	-0.007 (0.014)
British col.		0.112* (0.043)
French col.		0.015 (0.051)
Anocracy		0.108* (0.042)
Instability		-0.039 (0.045)
Polity		-0.001 (0.003)
Ln ethnic excluded population share		0.045** (0.013)
Years since independence	-0.001 (0.0004)	-0.0003 (0.0004)
Constant	0.203 (0.141)	0.173 (0.138)
N	142	139
R-squared	0.138	0.300

**p < .01; *p < .05; +p < .1
Standard errors in parentheses.

4 Estimates conditional on various measures of ethnic diversity

As discussed in the paper, a fundamental challenge to the ignorability assumption embedded in our inferences about the causal effects of Ethnic enumeration is a *primordial* one, that a pre-existing landscape of ethnic attributes and salient ethnic categories across the African continent were largely determinative of different choices about Ethnic enumeration, which in turn were correlated with ethnic conflict, such that the role of enumeration was endogenous or even possibly irrelevant. In the paper, we highlight some of our concerns about including measures of ethnic diversity in models estimating the effects of Ethnic enumeration, because we suggest that the causal model is different – that institutions in fact structure which ethnic categories become socially and politically relevant, and thus recognizable to citizens as well as outside observers.

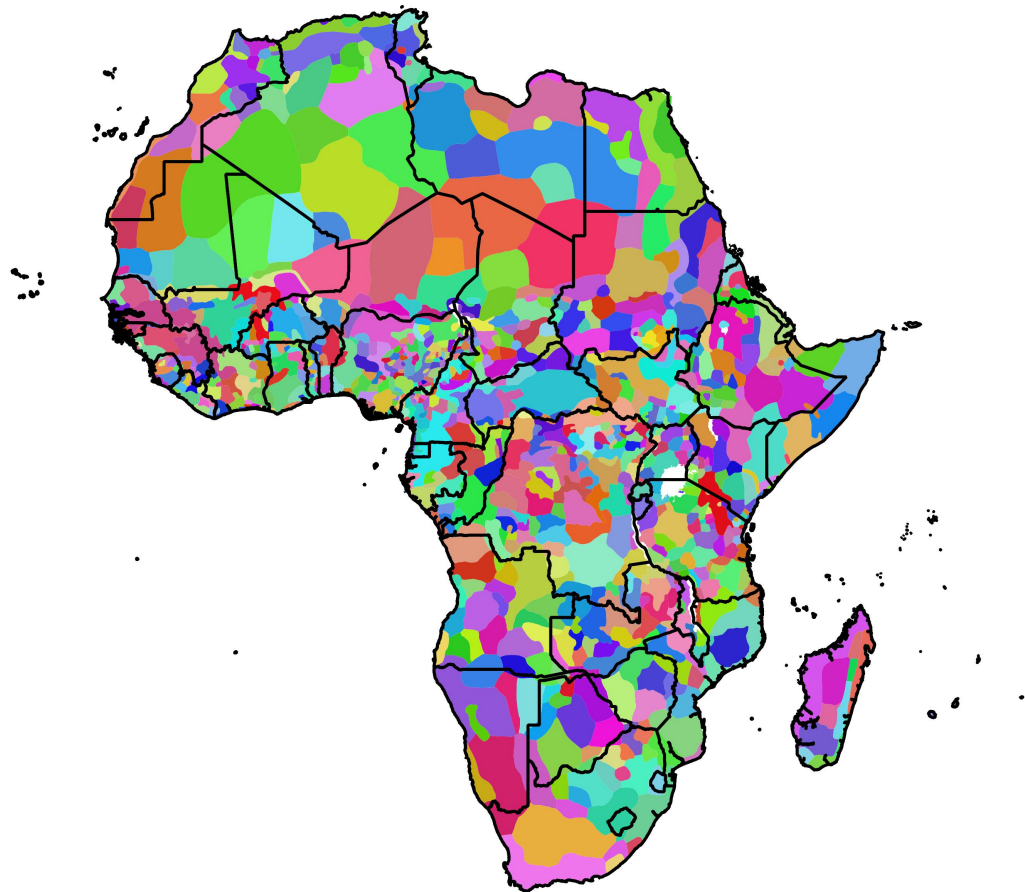
But of course, there is a quintessential “chicken-and-egg” problem here because, as we point out in the paper, states cannot simply enumerate categories out of thin air. They can, however, choose to enumerate some cleavages and categories and not others, or simply to avoid enumeration altogether.

For example, consider the important work of Murdock et al. (1959), who used various anthropological records to describe every existing ethnic group in sub-Saharan Africa in the late 19th century. In figure 1, we plot a map of contemporary Africa state boundaries against a map of the over 1000 identified groups. What is particularly relevant for our analyses here is that with only the exception of the countries of Swaziland and Lesotho, *every single country was characterized by some degree of ethnic diversity*. And yet, when we consider datasets such as Posner (2004) and the EPR dataset, we find several countries with no politically relevant ethnic groups. Moreover, one can easily find substantial differences across datasets in terms of which ethnic groups are relevant and the extent of diversity within countries. These types of discrepancies are the lifeblood of the constructivist paradigm!

Nonetheless, in order to more seriously consider the potential challenge to our findings that diversity or demographics drove institutional choices, rather than the reverse, we re-estimated our “base” models of ethnic conflict from Table 1 in the paper, conditioning on a range of different measures of ethnic diversity. In these estimates, presented in appendix Tables 6, 7, and 8, we find only minor attenuation of our results. In each table, we control for ethnolinguistic fractionalization (as analyzed in the paper), a measure of religious fractionalization, and Alesina et al’s measures of ethnic, language, and religious diversity. We also condition on the two ethno-political variables from the EPR dataset, which we have analyzed in the paper and above: the log of the share of the population that are associated with politically relevant ethnic groups who have been excluded from power; and on the number of ethnic groups in power.

In a total of 28 estimates, we do not find that the estimated relationship with Ethnic enumeration falls below conventional levels of statistical significance (at the .90 level). Overall, we interpret these findings to suggest that even if we were to take as given the patterns of ethnic diversity along various dimensions, including ethnic configurations of power, the institutionalization of ethnic categories still remains a strong additional predictor of the likelihood of ethnic conflict.

Figure 1: Murdock Map of Ethnic Groups in Africa in the Late 19th Century with Contemporary Africa State Borders Overlaid



Shape file from Nathaniel Nunn, replication data for “The Slave Trade and the Origins of Mistrust in Africa,” at <http://scholar.harvard.edu/nunn/pages/data-0>.

Table 6: Effect of ethnic enumeration on violent ethnic conflict (Vanhanen), Sensitivity to different ethnic diversity controls

	Ethnic Conflict 1990-6						
Ethnic enumeration (any) lag1	2.207** (0.807)	2.165** (0.749)	2.044** (0.779)	1.920* (0.802)	2.105** (0.760)	2.076* (0.828)	2.129** (0.773)
Log GDP/cap lag10	-0.072 (0.077)	-0.071 (0.077)	-0.073 (0.076)	-0.064 (0.074)	-0.055 (0.077)	-0.046 (0.082)	-0.075 (0.076)
Log pop lag1	0.304 (0.233)	0.298 (0.231)	0.301 (0.230)	0.355 (0.243)	0.412 [†] (0.248)	0.230 (0.239)	0.253 (0.244)
British col.	-0.028 (0.694)	-0.029 (0.696)	-0.025 (0.698)	0.076 (0.726)	-0.069 (0.704)	0.550 (0.769)	0.081 (0.724)
French col.	-1.975* (0.825)	-1.976* (0.829)	-1.967* (0.829)	-1.869* (0.855)	-2.150* (0.846)	-1.769* (0.892)	-1.944* (0.837)
Anocracy	1.093* (0.533)	1.082* (0.540)	1.012 [†] (0.544)	1.263* (0.552)	1.299* (0.554)	0.731 (0.562)	0.961 [†] (0.535)
Instability	0.483 (0.584)	0.482 (0.586)	0.482 (0.584)	0.572 (0.583)	0.480 (0.582)	0.756 (0.631)	0.504 (0.591)
Polity	-0.030 (0.042)	-0.030 (0.042)	-0.027 (0.043)	-0.011 (0.045)	-0.008 (0.046)	-0.021 (0.045)	-0.044 (0.046)
Years since independence	0.005 (0.009)	0.005 (0.009)	0.005 (0.009)	-0.0004 (0.009)	-0.002 (0.010)	0.005 (0.009)	0.006 (0.009)
Eth Frac.	-0.169 (1.162)						
Rel Frac.		0.015 (1.385)					
Ethnic Fractionalization (AL)			0.768 (1.424)				
Linguistic Fractionalization (AL)				0.599 (1.263)			
Religious Fractionalization (AL)					-1.063 (1.332)		
Log Percent Eth Excl Pop						0.482* (0.200)	
N Eth Grps in Power							0.059 (0.138)
Constant	-22.000 (1329.751)	-21.971 (1330.954)	-22.069 (1333.019)	-22.180 (1361.510)	-22.145 (1368.708)	-22.816 (1266.513)	-21.472 (1326.603)
N	141	141	141	136	140	139	139
Log Likelihood	-55.501	-55.511	-55.365	-53.474	-54.082	-51.116	-54.271
AIC	143.002	143.022	142.730	138.948	140.163	134.233	140.542

**p < .01; *p < .05; [†]p < .1
Regional dummies not shown

Table 7: Effect of ethnic enumeration on ethnic violence (MAR), Sensitivity to different ethnic diversity controls

	Ethnic Violence, 1950-2000						
Ethnic enumeration (any) max10	1.298*	1.325*	1.339*	1.245*	1.469*	1.099 ⁺	1.287*
	(0.587)	(0.560)	(0.563)	(0.612)	(0.574)	(0.570)	(0.523)
Log GDP/cap lag10	-0.096 ⁺	-0.110 ⁺	-0.091	-0.100 ⁺	-0.134 ⁺	-0.113 ⁺	-0.082
	(0.058)	(0.065)	(0.055)	(0.056)	(0.069)	(0.061)	(0.055)
Log pop lag1	0.486**	0.484**	0.498**	0.493**	0.465**	0.429**	0.521**
	(0.149)	(0.160)	(0.142)	(0.150)	(0.162)	(0.160)	(0.155)
British col.	1.036 ⁺	0.854	0.892	1.079	0.587	0.869	0.843
	(0.593)	(0.590)	(0.578)	(0.704)	(0.643)	(0.538)	(0.594)
French col.	0.876	0.773	0.767	0.975	0.595	0.731	0.669
	(0.808)	(0.773)	(0.783)	(0.896)	(0.779)	(0.789)	(0.773)
Anocracy	0.400	0.453	0.325	0.340	0.238	0.368	0.304
	(0.520)	(0.467)	(0.554)	(0.556)	(0.478)	(0.508)	(0.487)
Instability	-0.736	-0.588	-0.779	-0.777	-0.611	-0.879 ⁺	-0.646
	(0.490)	(0.460)	(0.546)	(0.555)	(0.528)	(0.506)	(0.463)
Polity	0.056	0.052	0.061	0.055	0.063	0.075 ⁺	0.056
	(0.039)	(0.041)	(0.040)	(0.040)	(0.041)	(0.040)	(0.041)
Years since independence	0.009	0.010	0.008	0.009	0.009	0.006	0.009
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)	(0.006)
Eth Frac.	0.831						
	(1.139)						
Rel Frac.		2.198 ⁺					
		(1.145)					
Ethnic Fractionalization (AL)			0.990				
			(1.150)				
Linguistic Fractionalization (AL)				1.043			
				(1.129)			
Religious Fractionalization (AL)					2.076 ⁺		
					(1.096)		
Log Percent Eth Excl Pop						0.479**	
						(0.122)	
N Eth Grps in Power							0.266*
							(0.121)
Constant	-9.672**	-10.286**	-9.965**	-9.859**	-10.189**	-9.203**	-10.126**
	(1.578)	(1.888)	(1.578)	(1.582)	(1.848)	(1.696)	(1.761)
N	486	486	472	455	471	485	485
Log Likelihood	-133.080	-130.500	-130.948	-128.456	-127.677	-125.918	-129.263
AIC	306.160	301.001	301.896	296.912	295.354	291.837	298.526

*** p < .01; ** p < .05; * p < .1

Decade and Regional dummies not shown

Country-clustered standard errors in parentheses.

Table 8: Effect of ethnic enumeration on ethnic armed conflict onset (epr), Sensitivity to different ethnic diversity controls

Ethnic Armed Conflict Onset, 1946-2005							
Ethnic enumeration (any) lag1	0.818*	1.072**	0.934**	0.762*	1.054**	0.941**	1.003**
	(0.320)	(0.315)	(0.310)	(0.307)	(0.314)	(0.308)	(0.305)
Log GDP/cap lag10	-0.083	-0.068	-0.053	-0.090	-0.058	-0.079	-0.058
	(0.060)	(0.049)	(0.054)	(0.058)	(0.049)	(0.053)	(0.049)
Log pop lag1	0.061	0.166	0.183	0.108	0.184	0.186	0.096
	(0.117)	(0.129)	(0.125)	(0.120)	(0.129)	(0.147)	(0.121)
British col.	-0.172	-0.142	-0.201	0.044	-0.141	-0.107	-0.273
	(0.320)	(0.312)	(0.289)	(0.324)	(0.315)	(0.287)	(0.362)
French col.	0.101	0.086	0.057	0.346	0.093	0.184	-0.103
	(0.328)	(0.362)	(0.342)	(0.340)	(0.367)	(0.375)	(0.406)
Anocracy	0.370 ⁺	0.442*	0.361	0.454 ⁺	0.467*	0.352	0.480*
	(0.222)	(0.224)	(0.220)	(0.232)	(0.226)	(0.230)	(0.230)
Instability	-0.118	-0.069	-0.111	-0.029	-0.022	-0.060	-0.025
	(0.274)	(0.281)	(0.286)	(0.283)	(0.291)	(0.288)	(0.282)
Polity	0.026	0.033 ⁺	0.038 ⁺	0.027	0.035 ⁺	0.044*	0.020
	(0.019)	(0.020)	(0.020)	(0.019)	(0.020)	(0.022)	(0.022)
Oil per capita lag	-0.0001	0.004	-0.017	0.004	0.003	0.010	0.001
	(0.018)	(0.013)	(0.025)	(0.012)	(0.014)	(0.011)	(0.015)
Log mountainous	0.254*	0.230*	0.236*	0.201*	0.234*	0.198*	0.219*
	(0.108)	(0.101)	(0.096)	(0.102)	(0.103)	(0.091)	(0.096)
N Peace Years	0.200	0.175	0.198	0.171	0.170	0.177	0.183
	(0.130)	(0.127)	(0.130)	(0.131)	(0.131)	(0.127)	(0.131)
History ethnic war	1.086**	1.031**	1.110**	1.037**	1.037**	0.833**	1.038**
	(0.344)	(0.324)	(0.325)	(0.332)	(0.332)	(0.320)	(0.324)
Years since independence	-0.002	-0.004	-0.005	-0.002	-0.005	-0.004	-0.002
	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
Eth Frac.	-0.003	-0.003	-0.005	-0.006	-0.005	-0.0001	-0.002
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Rel Frac.	2.438**						
	(0.625)						
Ethnic Fractionalization (AL)		0.111					
		(0.713)					
Linguistic Fractionalization (AL)			2.067**				
			(0.560)				
Religious Fractionalization (AL)				2.634**			
				(0.605)			
Log Percent Eth Excl Pop					-0.246		
					(0.601)		
N Eth Grps in Power						0.239**	
						(0.088)	
egipgrps							0.076
							(0.048)
Constant	-1.735	-3.443	-0.551	3.995	0.756	-8.939	-4.292
	(13.667)	(13.648)	(13.164)	(13.974)	(13.706)	(12.869)	(14.542)
N	6415	6415	6261	6051	6250	6398	6398
Log Likelihood	-428.918	-438.848	-427.450	-409.723	-426.745	-434.291	-436.927
AIC	905.836	925.696	902.901	867.446	901.490	916.583	921.854

**p < .01; *p < .05; ⁺p < .1

Regional dummies and splines not shown

5 Sensitivity to removal of country cases, regions

First, as discussed in the main text, when considering analyses of the number of politically relevant ethnic groups, we drop the cases of Russia, the Soviet Union, and China because they are severe outliers (see figure 2 – all observations with greater than 20 ethnopolitically relevant groups are from these country cases). In table 9, we reproduce Table 1 from the main text, but *keep* those cases in the analysis. Looking at the first five columns, the magnitude of the estimates of the effect of ethnic enumeration are generally larger than what is presented in the main text, though the size of the standard errors are also somewhat larger (while still being within conventional bounds of statistical significance). The one substantial diversion is the fixed effects estimates in column 6, in which the estimated coefficient sign has flipped and is no longer significant. This finding is clearly an artifact of the outlying cases, and as the estimates are not sensitive to the exclusion of other country cases, we believe that the estimates presented in the main text are more reliable.

Second, with respect to our analyses of conflict, of course, there is always the possibility that the results are being driven by a particularly influential country or region. We verify this is not the case by re-estimating each of the “base” models of ethnic conflict, iteratively removing one country at a time, and then one world region at a time. In figure 3, we show in the top panel that for all three models, the p-value for the estimated relationship between Ethnic enumeration and the associated outcome always fall well below the .05 level. Similarly, in the case of the removal of entire regions from the dataset (bottom panel), for 18 of 18 estimates, we also find that the p-value stays below the .05 level. Overall, we conclude that there are no highly influential outliers and our findings are generally robust in the global dataset.

Table 9: Effect of Ethnic Enumeration on Number of Politically Relevant Ethnic Groups (EPR), OLS Estimates (1946-2005), INCLUDING SEVERE OUTLIER CASES

	Number politically relevant ethnic groups					
Ethnic enumeration (any) lagged	2.147*	1.808*	1.628*	1.475*	1.693 ⁺	-0.036
	(0.900)	(0.750)	(0.788)	(0.714)	(0.894)	(0.099)
Log GDP/cap lag10	-0.073	-0.026	0.019	-0.025	-0.037	
	(0.055)	(0.044)	(0.034)	(0.067)	(0.043)	
Number languages			0.196**			
			(0.065)			
Log GDP/cap lag1						0.015*
						(0.007)
Log population	2.078**	2.090**	1.803**	1.482*	1.984**	0.403**
	(0.664)	(0.676)	(0.647)	(0.669)	(0.619)	(0.087)
British col.		0.590	-0.056	2.314 ⁺	-0.200	
		(0.732)	(0.714)	(1.318)	(0.871)	
French col.		-0.083	-0.719	0.955	-0.561	
		(0.822)	(0.867)	(1.121)	(0.823)	
Ethnic Fractionalization (AL)					-1.992	
					(2.657)	
Religious Fractionalization (AL)					3.659	
					(2.380)	
Linguistic Fractionalization (AL)					2.139	
					(1.916)	
Anocracy		0.519	0.492	0.843	0.371	0.025
		(0.634)	(0.657)	(0.867)	(0.651)	(0.076)
Instability		-0.700	-0.644	-0.258	-0.474	-0.170*
		(0.872)	(0.864)	(0.654)	(0.712)	(0.085)
Polity		-0.172 ⁺	-0.162 ⁺	-0.221 ⁺	-0.197 ⁺	-0.017**
		(0.091)	(0.090)	(0.127)	(0.113)	(0.006)
Absolute latitude				-3.209		
				(2.492)		
Variation in elevation				1.058		
				(0.776)		
Variation in land quality				-0.040		
				(0.510)		
Mean elevation				-2.534		
				(1.767)		
Mean land quality				-0.256		
				(0.560)		
Mean precipitation				0.870		
				(0.563)		
Mean temperature				-5.245		
				(3.728)		
Log area				0.346		
				(0.848)		
Distance from Sea				1.911*		
				(0.849)		
Migratory distance from E.Africa				-0.159		
				(0.586)		
Log Population density in 1995				-0.784		
				(0.603)		
Log Population density in 1500				0.868		
				(0.868)		
Timing transition to agriculture				0.493		
				(0.439)		
Years since independence	0.003	0.010	0.015	0.008	0.014	
	(0.015)	(0.016)	(0.015)	(0.016)	(0.018)	
Constant	-16.355*	-16.812**	-15.234*	-11.576 ⁺	-16.884**	
	(6.432)	(6.432)	(6.219)	(6.074)	(6.070)	
Fixed effects?	N	N	N	N	N	Y
Panel	Decade	Decade	Decade	Decade	Decade	Annual
N	635	627	627	580	590	6384
R-squared	0.219	0.247	0.279	0.403	0.271	0.933

**p < .01; *p < .05; ⁺p < .1

Country-clustered standard errors in parentheses, except in final column. Decade controls not shown. INCLUDES (outlier) Russia and China cases.

Figure 2: Number, Politically Relevant Ethnic Groups (EPR)

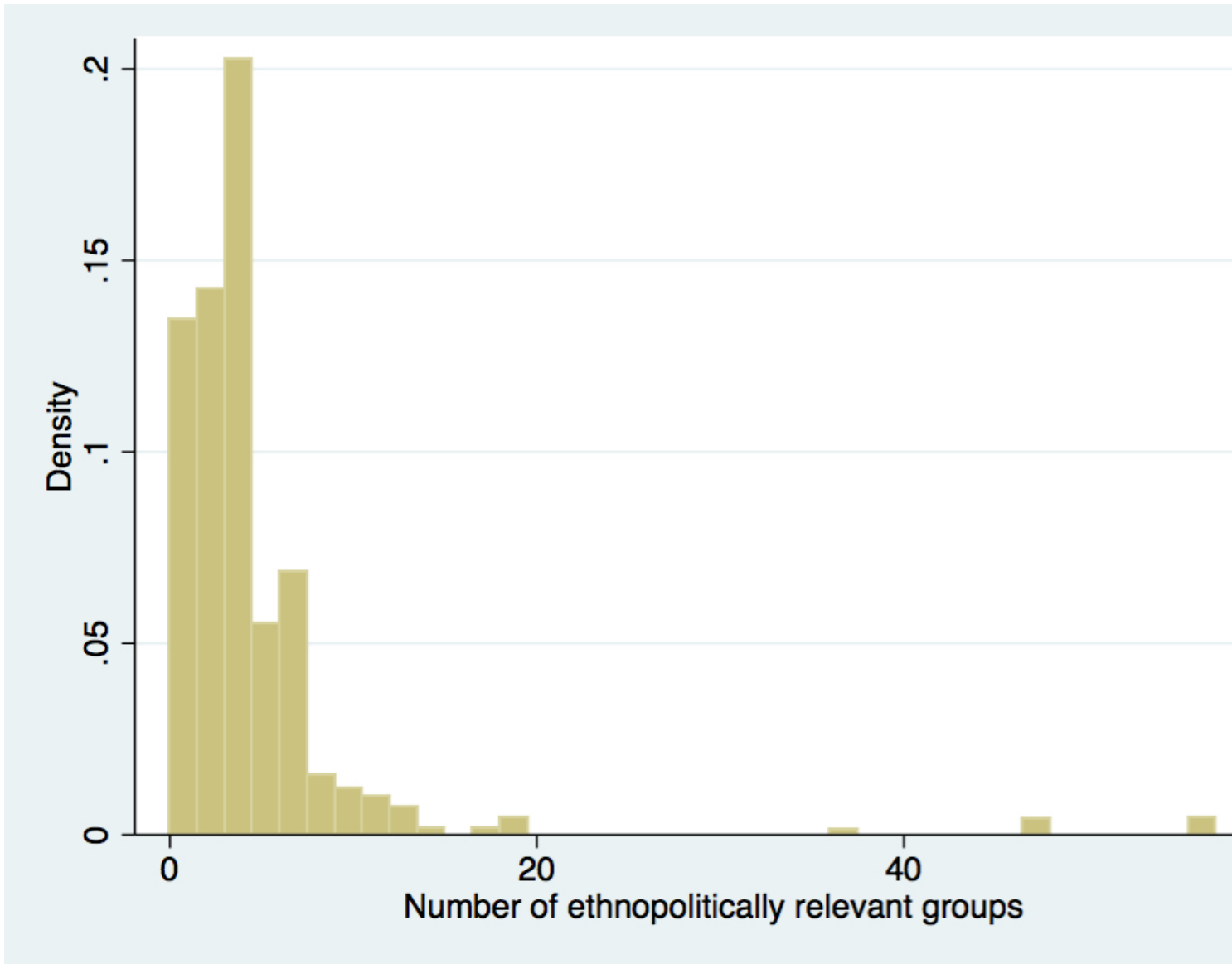
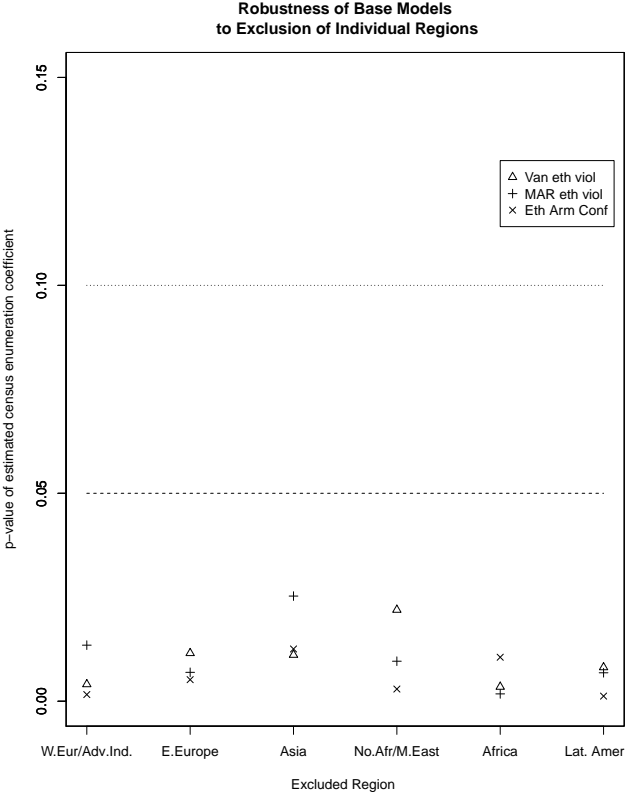
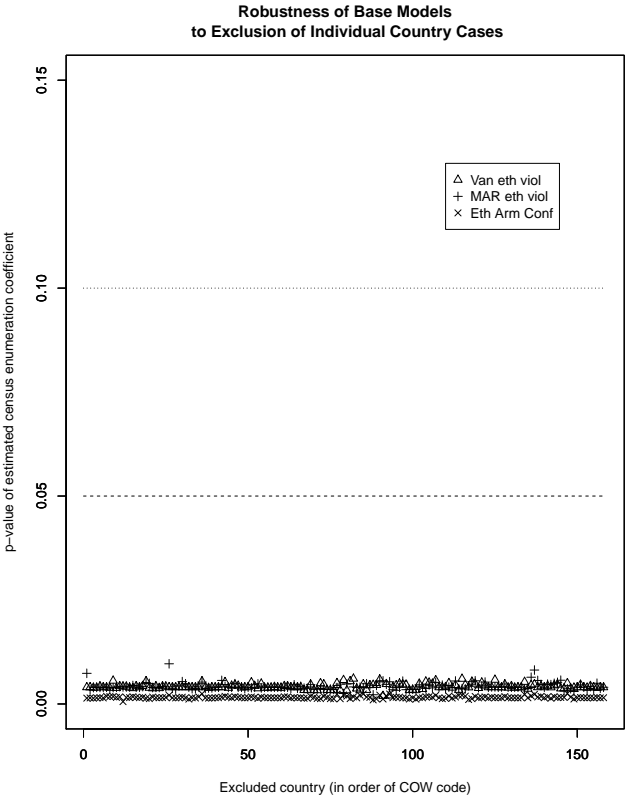


Figure 3: Sensitivity analyses



6 Rare events logit estimates

King and Zeng (2001) highlight the possibility of making biased inferences when analyzing the occurrence of rare events, such as our ethnic armed conflict outcome, when using standard logistic regression. When we implement their relogit package in Stata, we find a very slight attenuation of the estimated relationship with census enumeration, but we obtain similar results as with our regular logit estimates, as shown in table 10.

Table 10: Rare Events Logit Estimates of Ethnic enumeration as Predictor of Ethnic Armed Conflict (1946-2005)

	Year of new ethnic war (EPR)	Year of new ethnic war (EPR)
Ethnic enum	1.073* (0.338)	1.015* (0.337)
GDP/cap lag10	-0.0901* (0.0445)	-0.0806+ (0.0444)
Population logged	0.250* (0.107)	0.249* (0.106)
year	0.00897 (0.00684)	0.00858 (0.00683)
Years since independence	-0.0000272 (0.00391)	0.000116 (0.00391)
Observations	6486	6486

Country-clustered standard errors in parentheses

L. indicates lagged one year

+ $p < 0.10$, * $p < 0.05$

7 Cross-cuttingness of ethnic cleavags

We recognize that the proposition that the enumeration of multiple ethnic cleavages ought to increase the likelihood of conflict is potentially at odds with a theory emphasizing the role of cross-cutting cleavages – the degree to which people who share an ethnic category along one cleavage, are distributed among different categories on other cleavages – are likely to reduce the prospects of conflict by mitigating the strength of attachments to competing loyalties (Lipset and Rokkan 1967, Dahl 1982.). In particular, Selway (2010) has found that at low levels of ethnic fractionalization, cross-cuttingness reduces the likelihood of civil war onset. In the wake of such theory and empirical support, we hypothesize that the enumeration of multiple cleavages would only exacerbate the mollifying cross-pressures associated with cross-cuttingness.

We consider the phenomenon of cross-cutting ethnic cleavages using the Selway (2010) measure of the cross-cuttingness of language and religion, and an index of the enumeration of language and religion cleavages on the census, which can take on values of 0, 1, or 2. In analysis presented in table 11, we confirm that ethnic enumeration increases and cross-cuttingness reduces the likelihood of ethnic violence. We also estimate a model with the inclusion of an interaction term, the product of Selway’s cross-cuttingness indicator and a dummy variable coded as 1 for cases in which the census enumerated both language and religion. The estimated coefficient is negative, but with a large standard error of equal size. However, as (Selway 2010, 130-3) points out, our central concern should be with the marginal association of institutionalization for a range of values of cross-cuttingness. We follow their lead by plotting the marginal relationship with institutionalization along a range of values of cross-cuttingness in figure 4, and find that the relationship between the joint enumeration of language and religion and violence is significant only at low levels of cross-cuttingness with respect to those cleavages. Thus, our results are consistent with even this more nuanced consideration of multiple cleavage structures.

Table 11: Logit Estimates of the Interaction of Ethnic Enumeration and Cross Cutting Ethnic Cleavages on Outbreak of Ethnic Civil War (1946-2005)

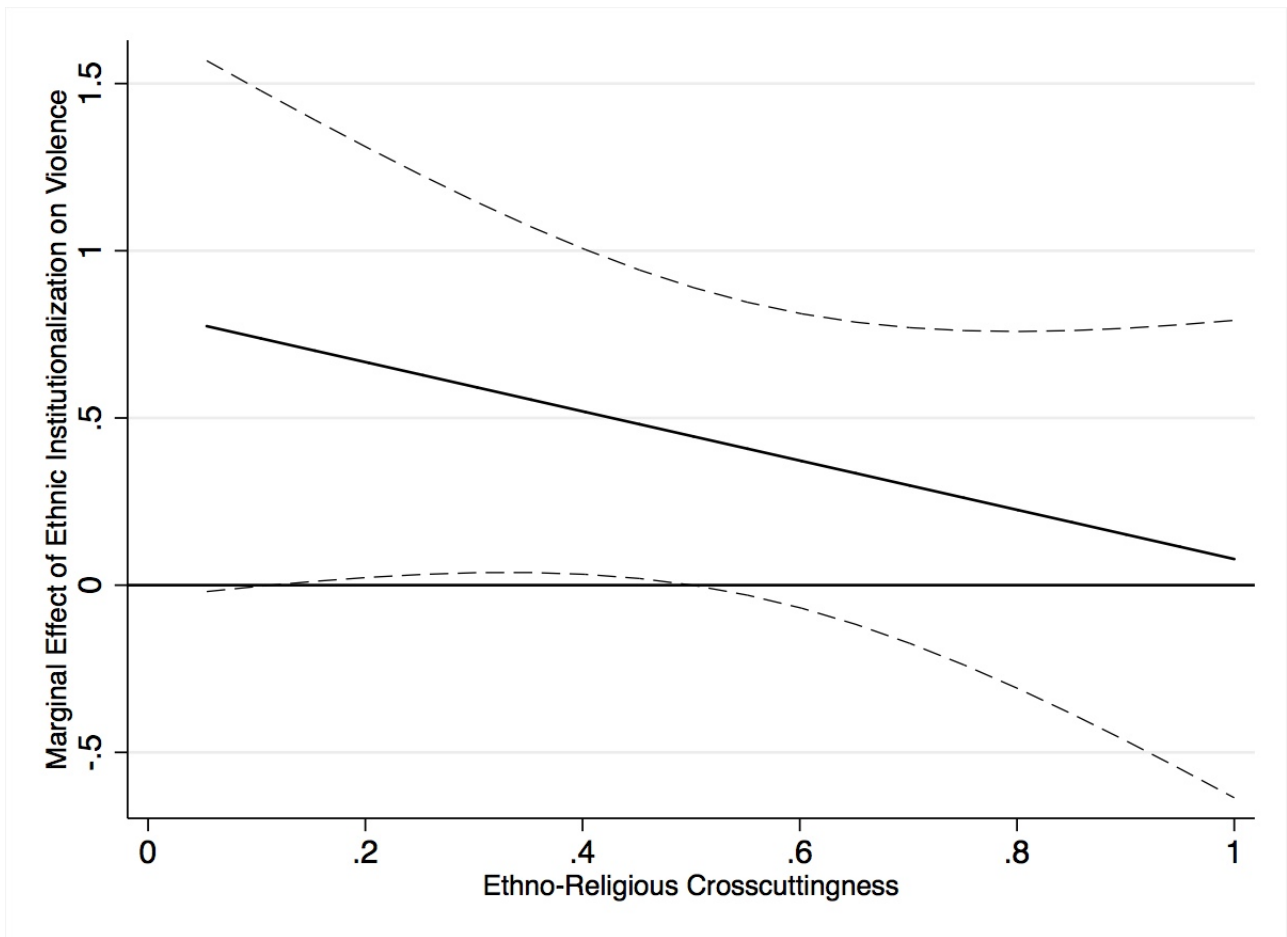
	Year of new ethnic war (EPR)
L.Enumeration of Lang and Relig (authors)	0.822 ⁺ (0.434)
Cross-cutting Lang x Relig (Sel)	-2.131* (0.877)
Cross-cutting * Enum Lang and Relig	-0.750 (0.661)
L.GDP/capita (PWT/EPR)	-0.166* (0.0465)
L.Ln population (WB)	0.363* (0.111)
ln (Mountainous terrain)	0.226 ⁺ (0.119)
Political instability (Polity/EPR)	0.114 (0.350)
L.Anocracy (Polity/EPR)	0.668* (0.309)
L.Oil per capita (EPR)	0.0806* (0.0328)
Observations	4801

Standard errors in parentheses

L. indicates lagged one year

⁺ $p < 0.10$, * $p < 0.05$

Figure 4: Enumeration of Religion and Language Categories as Predictors of Likelihood of New Ethnic Armed Conflict Onset, Conditional on Degree of Cross-Cuttingness



8 Missing census questionnaires

As reported in the text, we believe that for the period 1800-2005, we are missing information for about just 45 of 1378 large-scale censuses conducted in all countries that would have over 500,000 people by the year 2000. This represents 36 of 811 censuses during the 1946-2005 period. The missing census questionnaires are listed below. (We hope that scholars will forward information about missing questionnaires to us.) In our analyses we treat these years as we do other years when no census was enumerated.

BHUTAN 1991
BOSNIA 2005
BULGARIA 1992
BURMA 1941
CONGO 1984, 1996
CYPRUS 1992
DEM. REP. CONGO 1975
DENMARK 1916
IRAQ 1977
JAMAICA 1982
JORDAN 1994
KUWAIT 2005
LESOTHO 1891
LIBYA 1984, 1995, 2004
MALI 2000
MONGOLIA 1956, 1963, 1969
N. KOREA 1993
NETHERLANDS 1991, 2001
NICARAGUA 1940, 1986
OMAN 1993
PARAGUAY 1899, 1936
PERU 1828
SAUDI ARABIA 1992, 2004
SIERRA LEONE 1985
SOMALIA 1987
TRINIDAD AND TOBAGO 1990
TUNISIA 1994
U. ARAB EMIRATES 1980, 1985, 1990, 1995
URUGUAY 1985
VENEZUELA 1873
ZIMBABWE 1931

References

- Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat and Romain Wacziarg. 2003. "Fractionalization." *Journal of Economic growth* 8(2):155–194.
- Bruk, Solomon I. and V. S. Apenchecko. 1964. *Atlas Narodov Mira*. Glavnoe upravlenie geodezii i kartografii gosudarstvennogo geologicheskogo komiteta SSSR and Institut ethnografii im. H.H. Miklukho-Maklaia Akademii nauk SSSR.
- Center for International Development and Conflict Management. 2009. "Minorities at Risk Dataset."
- Dahl, Robert Alan. 1982. *Dilemmas of Pluralist Democracy: Autonomy vs. Control*. New Haven, CT: Yale University Press.
- Hlavac, Marek. 2015. "Stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2. <http://CRAN.R-project.org/package=stargazer>."
- King, Gary and Langche Zeng. 2001. "Explaining rare events in international relations." *International Organization* 55(3):693–715.
- Lipset, Seymour Martin and Stein Rokkan. 1967. *Cleavage Structures, Party Systems, and Voter Alignments*. The Free Press.
- Marshall, Monty G and Keith Jagers. 2009. "Polity IV dataset version 2007 and Dataset users' manual." *Fairfax, VA: Center for Systemic Peace and Center for Global Policy, George Mason University*. <http://www.systemic-peace.org>. Accessed 7(1):2009.
- Murdock, George Peter et al. 1959. *Africa: its peoples and their culture history*. McGraw-Hill New York.
- Posner, Daniel. 2004. "Measuring Ethnic Fractionalization in Africa." *American Journal of Political Science* 48(4):849–863.
- Selway, Joel Sawat. 2010. "Cross-cuttingness, cleavage structures and civil war onset." *British Journal of Political Science* 41(1):111–138.
- Teorell, Jan, Nicholas Charron, Marcus Samanni, Sören Holmberg and Bo Rothstein. 2011. "The Quality of Government Dataset, version 6Apr11. University of Gothenburg: The Quality of Government Institute."
- Vanhanen, Tatu. 1999. "Domestic ethnic conflict and ethnic nepotism: A comparative analysis." *Journal of Peace Research* 36(1):55–73.
- Wimmer, Andreas, Lars-Erik Cederman and Brian Min. 2009. "Ethnic Politics and Armed Conflict: A Configurational Analysis of a New Global Data Set." *American Sociological Review* 74(April):316–337.

Note: Regression output created with Hlavac 2015.