

Auxiliary Materials

Zachary Elkins and John Sides. 2007. "Can Institutions Build Unity in Multiethnic Societies?"

American Political Science Review 101(4): 693-708.

Appendix 1: Representativeness of MARs Sampled in WVS

Of the universe of 314 MARs, there are 90 ("coded groups") that we have identified in our 51-country sample, 65 ("uncoded groups") that we were unable to identify in the WVS sample, and 159 ("excluded groups") who reside in countries that are not part of the WVS. (Both coded and uncoded groups are listed in Table A-1.) That our analysis includes only a subset of groups raises questions about the representativeness of these groups. How do the coded groups compare to the uncoded groups and to the excluded groups? We compared these three different sets of groups on a host of attributes, including their size, geographic concentration, strength of identity, experience with discrimination, severity of grievances, and amount of rebellion or protest activity. We found very few statistically significant differences between the coded and excluded groups. The handful of differences that did exist suggests that the 159 excluded groups had fewer grievances and protest activity.

When we compared the 90 coded groups to the 65 uncoded groups, we found several significant differences. On average, the uncoded groups were smaller and more geographically concentrated, with a stronger sense of group identity and higher levels of both cultural distinctiveness and economic discrimination. That many of these smaller groups—some of which constitute much less than 1 percent of a country's population—would not be captured by a survey sample is understandable. That these groups are also different in terms of strength of identity, level of distinctiveness, and degree of discrimination suggests that our sample of 93 misses some of the most "at risk" groups. However, these groups are not statistically different in other ways—such as

their level of political and cultural discrimination, the strength of their political, economic, and cultural grievances, and their level of rebellion and protest activity. This suggests that our sample is reasonably representative in other respects. Thus, while we must be cautious in generalizing from our particular findings, we can have some confidence that our results are not based on too exotic a sample of communal groups. Further details about the attributes of these three different sets of groups are available from the authors.

Table A-1. Countries and Groups Represented in Dataset

Country	Yr of survey	Cntry N	Coded groups	Group N	Sample %	Pop %	Uncoded groups	Pop %
Albania	2002	1000	Greeks	16	1.6	3.7		
Algeria	2002	1282	Berbers	165	12.9	24.2		
Argentina	1999	1280	Jews	10	0.7	1.0	Indigenous Peoples	0.010
Australia	1995	2048	Aborigines	8	0.4	1.4		
Azerbaijan	1996	2002	Lezgins	155	7.7	2.5	Armenians	0.023
			Russians	140	7.0	25.0		
Bangladesh	2002	1499	Hindus	106	7.1	10.8	Chittagong Hill Tribes	0.006
							Biharis	0.002
Belarus	1996	2092	Russians	285	13.6	13.2		
			Poles	127	6.1	4.1		
Bosnia	2001	1200	Serbs	524	43.7	40.0		
			Croats	186	15.5	22.0		
			Muslims	485	40.4	38.0		
Brazil	1997	1149	Afro-Brazilians	227	19.8	39.9		
			Amazonian Indians	26	2.3	0.2		
Bulgaria	1997	1072	Turks	99	9.2	8.5		
			Roma	40	3.7	8.9		
Canada	2000	1931	Quebecois	454	22.0	19.7		
			French Canadians	78	2.3	3.8		
			Indigenous Peoples	18	1.1	2.9		
Chile	2000	1200	Indigenous Peoples	20	1.6	5.0		
Colombia	1997	6025	Blacks	318	5.3	4.5		
			Indigenous Peoples	253	4.2	1.0		
Croatia	1995	1196	Serbs	14	1.1	5.3	Roma	0.008
Dominican Rep.	1996	417	Haitian Blacks	50	12.0	9.7		
Egypt	2000	3000	Copts	168	5.6	8.9		
Estonia	1999	1005	Russians	304	30.3	29.8		
Georgia	1996	2008	Ossetians (South)	2	0.1	3.2	Abkhazians	0.018
			Russians	234	11.7	4.8	Adzhars	0.058
Germany	1999	2036	Turks	23	2.1	2.4		
India	2001	2002	Kashmiris	3	0.2	0.6	Nagas	0.3
			Muslims	248	12.4	11.4	Santals	0.7
			Scheduled Tribes	99	5.0	7.5	Mizos	0.1
			Sikhs	28	1.4	2.0	Tripuras	0.1
			Assamese	10	0.5	1.5	Bodos	0.5
Indonesia	2001	1004	Chinese	18	1.8	2.7	East Timorese	0.4
							Papuans	0.5
							Acehnese	1.7
Iran	2001	2532	Azerbaijanis	288	11.4	24.4	Baha'Is	0.9
			Christians	24	1.0	0.4	Bakhtiari	1.1
							Baluchis	1.9
							Kurds	7.4
							Turkmen	1.9
							Arabs	2.7
Israel	2001	1199	Arabs	151	12.6	17.0	Palestinians	28.8

Country	Cntry N		Coded groups	Group N	Sample %	Pop %	Uncoded groups	Pop %
Italy	1999	2000	Sardinians	57	2.9	3.0	South Tyrolians	0.5
Jordan	2001	1223	Palestinians	484	47.3	47.0	Roma	0.2
Latvia	1999	1013	Russians	396	39.1	34.0		
Lithuania	1996	1009	Russians	102	10.2	8.7	Poles	7.0
Macedonia	2001	1055	Albanians	236	22.4	22.9		
			Serbs	20	1.9	2.4		
			Roma	16	1.5	12.0		
Mexico	2000	1535	indigenous	25	1.9	10.8		
Moldova	2000	1008	Gagauz	16	1.7	3.5		
			Slavs	208	21.2	26.8		
Morocco	2001	2264	Berbers	294	14.2	37.0	Saharawis	0.7
New Zealand	1998	1201	Maori	1	0.1	9.8		
Nigeria	2000	2022	Ibo	424	21.0	17.0	Hausa-Fulani	29.0
			Yoruba	515	25.5	20.0	Ijaw	3.6
							Ogoni	0.5
Pakistan	2001	2000	Baluchis	64	3.2	4.0	Ahmadis	3.5
			Pashtuns (Pushtuns)	312	15.6	9.0	Hindus	1.9
			Sindhis	189	9.5	11.6	Mohajirs	8.0
Peru	2001	1501	Afro-Peruvians	122	8.1	0.9	Lowland indigenous	1.2
			Indigenous highlanders	12	0.8	38.4		
Philippines	2001	1200	Igorots	1	0.1	1.4		
			Moros	37	2.8	5.3		
Romania	1999	1146	Magyars (Hungarians)	108	9.4	8.7	Germans	1.6
							Roma	8.6
Russia	1995	2040	Tatars	93	4.7	3.8	Chechens	0.6
							Karachay	0.1
							Roma	0.2
							Avars	0.4
							Ingush	0.2
							Lezgins	1.7
							Buryat	0.3
							Kumyks	0.2
							Tuvinians	0.1
							Yakut	0.3
Serbia & Montenegro	2001	2260	Kosovo Albanians	88	1.1	12.7	Slovenes	8.0
			Hungarians	52	4.0	4.8		
			Sandzak Muslims	182	4.0	1.8		
			Roma	17	1.3	3.5		
			Croats	26	1.3	1.3		
Singapore	2002	1512	Malays	511	14.4	14.4		
South Africa	2001	3000	Asians	299	3.6	2.6		
			Coloreds	499	10.7	8.7		
			Europeans	899	14.2	13.2		
			Xhosa	318	10.9	17.0		
			Zulus	400	22.0	15.9		
Spain	2000	2409	Basques	11	0.5	5.2	Roma	1.9
			Catalans	73	3.0	14.8		
Switzerland	1996	1212	Foreign Workers	51	5.0	18.4	Jurassians	2.2

Country		Cntry N	Coded groups	Group N	Sample %	Pop %	Uncoded groups	Pop %
Taiwan	1994	780	Aboriginal Taiwanese	66	8.5	1.7		
			Mainland Chinese	181	23.2	13.9		
			Taiwanese	490	62.8	84.2		
Turkey	1996	1907	Kurds	40	2.3	20.0	Roma	1.1
Uganda	2001	1002	Baganda	280	23.4	16.0	Acholi	4.0
							Ankole	8.0
							Kakwa	3.0
							Karamojong	2.0
							Konjo/Amba	3.0
							Langi	6.0
							Lugbara/Madi	4.9
							Nyarwanda	5.9
Ukraine	1996	2811	Russians	1061	37.7	22.0	Crimean Tartars	0.5
			Crimean Russians	121	4.3	3.3		
United Kingdom	1999	2000	Catholics In N. Ireland	366	1.1	1.2	Afro-Caribbeans	1.9
			Scots	84	8.5	9.5		
			Asians	19	2.3	2.9		
United States	1999	1200	African-Americans	191	16.0	12.4	Native Americans	
			Hispanics	102	8.5	9.7	Native Hawaiians	
Venezuela	2000	1200	Blacks	51	4.3	10.0		
			Indigenous Peoples	6	0.5	1.4		
Zimbabwe	2001	1002	Ndebele	150	14.0	16.8	Europeans	1.4

All country samples come either the 1995-97 or 1999-2001 waves of the World Value Survey. In cases where the survey was in the field in two different years, the earlier year was entered. The WVS documentation for Moldova had no information as to survey year, so 2000 was assigned. Group names and their percentage of the population are taken directly from the MAR dataset (variables GROUP and GPRO). The sample percentages are weighted. In Mexico, the WVS identified respondents only as “indigenous” even though there are three indigenous MARs in the country: Mayans, Zapotecs, and “other indigenous peoples.” Thus, the WVS sample could include one or more of these groups, so the population percent presented above is the combined figure for all three groups.

Appendix 2. Concepts and Primary Measures

Concept	Measure/Variable	Data Source
Individual level		
Minority-at-risk	dummy variable based on ethnicity (v233/x051), language (v209/g016), language of interview (n.a./s016), religion (v179/f025), and/or region (v234/x048)	WVS (1995-97/1999-2001)
Age	age in years (v216/x003)	WVS
Education	education in country-specific deciles (v217/x015)	WVS
Economic insecurity	income in country-specific deciles (v227/x047)	WVS
Life satisfaction	“All things considered, how satisfied are you with your life as a whole these days?” (1-10 scale) (v65/a170)	WVS
Cosmopolitanism	number of postmaterial priorities (v104-109/e001-e006)	WVS
In-party	dummy variable if respondent’s self-reported party preference is same as current executive’s (v210/e179)	WVS; Beck et al. (2001)
Group level		
Proportion of population	GPRO	MAR
Geographical concentration	4-category measure from GC2, GC7 0 (0-25 percent of the group does not live in a spatially contiguous area (other than an urban area), 1 (25-49 percent), 2 (50-74 percent), or 3 (75 percent or more).	MAR
Strength of identity	CATNESS	MAR
National minority	dummy variable from TRADITN	MAR
New immigrant	dummy variable from TRADITN	MAR
Autonomy grievances	scale from AUTGR2- AUTGR4	MAR
Political grievances	scale from POLGR2- POLGR4	MAR
Economic grievances	scale from ECONGR2- ECONGR5	MAR
Cultural grievances	scale from CULGR2-CULGR4	MAR
Ethnic party	Dummy variable if group has at least one party representing it (ORNAME1-ORNAME3)	MAR
Country level		
Economic development	GDP per capita (10-year avg)	World Development Indicators (World Bank)
Level of democracy	procedural democracy (10-year avg)	Polity IV
Accumulated democracy	cumulative sum of level of democracy	Polity IV
Age of state	calculated as the difference between the year of the survey and the year of independence	ICOW
Proportionality of electoral system	median district magnitude in the lower house (logged)	Golder (2004); Beck et al. (2001)
Federalism	(1) Federalism dummy; (2) index of decentralization	(1) Polity III; (2) Arzaghi and Henderson (2005)
Former axis power	dummy variable for Italy and Germany	-
Former Communist country	dummy variable for Russia, former Soviet Republics, and Eastern bloc countries	-

For the WVS, see Inglehart *et al.* (2004). For MAR, see CIDCM/MAR (2003). For Polity IV, see Marshall, Jaggers, and Gurr (2004).

Appendix 3. First-Stage Equations for Instrumented Variables

Because a country's selection of institutions (federalism and electoral systems) is very likely to be dependent upon the national attitudes of its minorities, we follow an instrumental variables approach. We use the instruments listed in Table A-2 below to predict values of our three principal institutional variables. The instruments capture important aspects of the political, cultural, and economic tradition of countries. The other exogenous country-level variables from the models of state attachment are also included as instruments.

Our selection of instruments is based on extant theory on the origins of electoral systems and federalism as well as our own intuition. For example, prior theory suggests that ethnic heterogeneity (Boix 1999), a large geographic area (Boix 1999), economic openness (Rogowski 1987), geographic diffusion (Elkins 2003), are associated with electoral system choice. Theory regarding the origins of federalism suggests that external threat (Riker 1964), a British colonial ancestry (Riker 1964), a large geographic area, and infrastructural capacity (Ziblatt 2006) lead to federal bargains. We also speculate that countries with a similar language, religious tradition, vintage of their first democratic constitution, ecology (urban/rural distribution), and geographic position will tend to have developed similar institutions. Provided that the instruments are clean and display predictive power (properties we test formally below), our inclination is to err on the side of inclusion in order to realize efficiency gains in the prediction of the endogenous variables.

The predictions are derived from a cross-sectional model using data from 1995. All states with non-missing values on the dependent variable are included in the regression. Missing values on several independent variables were replaced with the variable's mean. Equations were estimated with OLS for the decentralization index and median district magnitude and with logit for the federal dummy. Some variables (e.g., region) were excluded from the equation for federalism because they predicted the dependent variable perfectly.

The selection of instruments is, of course, of crucial concern. Based on the overall fit of the models (r-squared ranging from 0.39 to 0.95, across the three models), we are encouraged that we have identified a set of instruments that predict the institutional variables adequately well. A standard summary measure of the predictive power of instruments is the joint test that the coefficients on the instruments are zero. Not surprisingly, given the high r-squared statistics, we can strongly reject the null hypothesis for both federalism and district magnitude ($F(33, 120)=2.92$, $p<.01$ and $F(33, 182)=3.47$, $p<.01$, respectively).

We also believe that the instruments are “clean,” in that they are not themselves highly correlated with the error term in the second stage equations. We tested this possibility with the Sargan test, in which we regress the residuals from the second-stage equation on all of the instruments used in the first stage. The null hypothesis is that all instruments are uncorrelated with the error and the test has chi-squared distribution with degrees of freedom equaling the number of overidentifying restrictions (the exclusion restrictions).¹ In Table A-3, we report these results for each dependent variable, for both minorities and majorities. Note that all of these tests, like the first-stage regressions themselves, were conducted at the level of the state. Accordingly, the tests are limited to the state-level variables with the exception of the dependent variables, which are state-level means of the survey items, for both minorities and majorities. As the results in Table A-3 attest, none of the chi-square statistics is significant, suggesting that we cannot reject the null hypothesis that all the instruments are clean.

¹ These tests were run in Stata using the “overid” command following an instrumental variables regression of each of the dependent variables (aggregated to the state level) on the state-level variables using the “ivreg2” command.

Table A-2. First-Stage Regression Results

	Federalism Dummy	Centralization Index	Median District Magnitude
Accumulated Democracy	0.003* (0.001)	0.000 (0.000)	0.002 (0.007)
Level of Democracy	0.284 (0.197)	-0.335*** (0.061)	-0.530 (0.954)
Age of State	-0.037** (0.016)	-0.000 (0.002)	-0.036 (0.058)
Civil Liberties	0.292 (0.860)	-0.294 (0.230)	3.037 (4.200)
Ethnic Fractionalization	-0.095** (0.037)	0.013** (0.006)	-0.358** (0.169)
Gini Coefficient of Inequality	0.083* (0.043)	-0.030** (0.011)	-0.185 (0.286)
Percent Catholic	0.040 (0.028)	-0.001 (0.005)	-0.127 (0.152)
Percent Muslim	-0.047 (0.054)	-0.003 (0.007)	-0.119 (0.172)
British Colony	-4.937 (3.419)	1.288** (0.466)	-18.125 (11.682)
Former Great Power	2.132 (2.335)	-1.147** (0.503)	14.126 (11.368)
Political Liberties	-0.033 (1.031)	0.760*** (0.220)	0.112 (4.886)
GDP per capita	1.589 (0.996)	0.491* (0.241)	1.385 (5.325)
Land Area	0.649** (0.279)	0.096** (0.039)	2.782* (1.448)
Population	3.533 (3.150)	4.101*** (0.901)	5.404 (35.805)
Industrial Society	5.384 (4.244)	-2.684*** (0.871)	6.003 (18.009)
Agrarian Society	4.420 (4.988)	-4.713*** (1.058)	12.435 (21.118)
Newer democratic Tradition	-1.241 (1.862)	2.943*** (0.919)	17.772 (12.517)
Semi-democratic Tradition	0.720 (2.551)	3.823*** (0.899)	20.748 (15.310)
Non-democratic Tradition	5.295 (4.140)	0.203 (1.180)	10.267 (22.599)
Asia-Pacific		-1.202** (0.428)	1.859 (9.744)
Central or Eastern Europe		-2.278*** (0.549)	14.766 (9.632)
Middle East		-3.055*** (0.625)	11.240 (11.412)
North America		-2.036*** (0.686)	-15.775 (22.051)
South America		-1.700** (0.653)	-3.850 (10.252)

	Federalism Dummy	Centralization Index	Median District Magnitude
Scandinavia		0.000 (0.000)	21.194 (19.424)
Western Europe		-1.338** (0.568)	18.809 (15.498)
Predominantly Protestant		-0.511 (0.312)	-12.968 (8.858)
Predominantly Orthodox		-0.967* (0.542)	-1.308 (11.350)
Predominantly Jewish		0.000 (0.000)	116.593*** (31.994)
Predominantly Muslim		-0.055 (0.577)	-15.967 (10.448)
Predominantly Hindu		-0.396 (0.714)	-12.628 (21.274)
Predominantly Buddhist		-1.847*** (0.577)	-8.957 (14.635)
Predominantly Other Religion		-3.721*** (0.625)	-17.088 (12.481)
Constant	-4.483 (8.689)	2.899 (1.668)	18.669 (43.865)
Observations	157	48	146
R2 (or pseudo R2)	0.59	0.95	0.39

Table entries are unstandardized logit or OLS coefficients, with estimated standard errors in parentheses.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table A-3 Sargan Tests of Exogeneity

Dependent Variable(sample)	χ^2	d.f.	p-value
Pride (majorities)	33.42	27	.18
Pride (minorities)	29.43	25	.24
Identification (majorities)	32.43	25	.15
Identification (minorities)	30.15	25	.22

References

- Boix, Carles. 1999. "Setting the Rules of the Game: The Choice of Electoral Systems in Advanced Democracies." *American Political Science Review* 93 (3):609-624.
- Elkins, Zachary. 2003. *Designed by Diffusion: International Networks and the Spread of Democracy*. Doctoral Dissertation, University of California, Berkeley.
- Riker, William H. 1964. *Federalism: Origin, Operation, Significance*. New York: Little, Brown, and Company.
- Rogowski, R. 1987. "Trade and the Variety of Democratic Institutions." *International Organization* 41 (2):203-223.
- Ziblatt, Daniel. 2006. *Structuring the State: The Formation of Italy, Germany, and the Puzzle of Federalism*. Princeton: Princeton University Press.