Ethnic Politics and Armed Conflict: Regional Differences

John Snyder, Carl Gershenson, Burak Eskici

Department of Sociology

April 29, 2010

In this paper, we replicate and comment Wimmer, Cederman, and Min's (2009) "Ethnic Politics and Armed Conflicts." First we discuss structural issues with the data that may interfere with valid causal inference. Then we estimate a number of first differences in order to question the theoretical conclusions the authors drew from their own models. Finally, we discuss the role of ethnic exclusion across regions. We estimate first differences in Eastern Europe and the Middle East, showing that region and ethnic exclusion relate to each other differently in these regions. [1]

In their paper "Ethnic Politics and Armed Conflicts" Wimmer, Cederman, and Min (2009) argue that violent conflicts within nations result from the relation of ethnic groups to state power, rather than from ethnic diversity itself. States that exclude large portions of the population from state participation on ethnic grounds will be more susceptible to rebellion, while states characterized by competing ethnic power elites will be more susceptible to violent infighting. We argue that Wimmer et al may be correct, but that they overplay the substantive importance of their results. Furthermore, they ignore important structural characteristics of their dataset. At best, Wimmer et al must settle for a pared down version of their theory of ethnic conflict. At worst, their inferences are completely model dependent and must be treated with the requisite skepticism.

To understand why we believe Wimmer et al to have overstated their hand, we must investigate the theories that the authors see themselves as refuting. First and foremost, they hold their model against Fearon and Laitin's (2003) insurgency model, which in part predicts that conflict erupts

[1] The data, code and other documentation needed for replication are available through Dataverse. See the bibliography for details.

when the state is weak. Laitin (2007) later argued that while over the 20^{th} century both scholars and insurgents became more likely to label civil strife as "ethnic conflicts," the structural factors underlying civil strife have remained the same. Wimmer et al call this the "greed and opportunity" model.

The authors also see themselves as opposing what they call the "diversity breeds conflict" school. These theories argue that ethnically diverse states will inherently experience more conflict because the nation-state derives its legitimacy from its cultural homogeneity (Gellner 1991; Nairn 1993). These studies generally represent ethnic diversity through a "linguistic fractionalization index." [2]

The first part of the authors' "institutional configurational theory" closely follows the "diversity breeds conflict" model. The authors agree that the modern nation-state relies on ethno-national principles for political legitimacy, but they argue that the linguistic fractionalization index does not account for how ethnic groups are arranged within the state. In particular, the authors propose three mechanisms for ethnic conflict. Conflict between multiple power sharing groups will lead to violent infighting - for example, in Iraq after the fall of Saddam Hussein. Conflict involving a group excluded from the political process but sharing the same geographic territory leads to ethnic rebellion - as in the United Kingdom during the Troubles. Finally, when the excluded ethnic group is geographically isolated from the rest of the nation, secessionist movements may emerge - as in the former Yugoslavian republics.

In this paper, we do not dispute that the mechanisms identified by Wimmer et al affect the probability of ethnic conflicts. However, we believe that the structure of the data leaves the authors open to criticisms of model dependency. Furthermore, we argue that the authors' exaggerate the importance of the ethnic variables while ignoring the substantively larger effect played by the level of development of a state as measured by per capita GDP. We therefore argue that the authors' findings are better interpreted as an addendum to the "greed and opportunity" model, rather than as a refutation thereof. Finally, based on theoretical considerations, we question the universality of the ethnic mechanisms in play as claimed by the authors. We find that these mechanisms may

[2] We ran models using the linguistic fractionalization index included in the Ethnic Power Relations dataset. However, it was never significant alongside the authors' preferred ethnic power variables. (Tables not shown; code included in replication file.)

operate differently depending on cultural region.

Discussion of the Data Set

The Ethnic Power Relations (EPR) dataset is one of the most comprehensive datasets on ethnic conflict, including 7155 observations over a period of about 60 years. This data set measures access among minority groups to executive level state power between 1946 and 2005. In these data, a conflict is considered "any armed and organized confrontation between government troops and rebel organizations, or between army factions, that reaches an annual battle-death threshold of 25 people." This definition does not include genocides, or ethnic conflicts that do not involve the government. The dependent variable used identifies the onset year of such a conflict, but not the subsequent years over which the conflict takes place. Ethnic groups considered relevant to the data include ethnolinguistic, ethnosomatic, and ethnoreligious groups. These groups are considered politically relevant within the country if at least one significant national political actor claims to represent the interests of the group. The data only contain information about executive level actors, such as the presidency, cabinet, and senior administration posts. Any ethnic group not fitting this description is counted as an excluded ethnic group. Additionally, any ethnic group that has been subject to widespread and systematic discrimination is counted as an excluded ethnic group. The data do not code for the degree of representation for ethnic groups.

By collecting this data and making it publicly available, Wimmer et al have contributed greatly to the study of ethnic conflict. However, we have several concerns with the EPR. We will deal with each of our concerns below.

First, the dataset is missing about 2 out of every 1,000 values. Because values seemed most likely to be missing when a less-developed country experiences civil conflict (two variables included in our dataset), we believe that missingness is a function of observed values. As such, we used multiple imputations to fill in missing values.

More importantly, the data is unbalanced to a problematic. The original data have an L1 value of 1, which implies that any causal inferences will be model dependent. We tried coarsened exact matching (CEM), which gave us L1 = 99.2, but at the cost of the majority of our observations. We

also tried nearest neighbor matching, which cut 50% of our treatment cases without any improvement in balance (L1 = 1). We felt that lack of balance is simply an intractable issue with data such as these - after all, only the passage of time will give us more observations - and we therefore acknowledge the model dependence of our analyses. [3]

Thirdly, we noticed surprising omissions in their data. For example, the French - Algerian war is not coded as a civil conflict for either France or Algeria. As far as we understand this battle should have been coded according to the rules of data collection process provided by Wimmer et al, as it is represents an ethnic secessionist movement inside of the former boundaries of the French state. In general, we noted that the often violent de-colonization wave of the post-war era is not well captured by this data set.

We were also surprised by coding decisions regarding what counts as the "onset" of a "new" civil conflict. When a conflict starts, the o dichotomous "war-onset" variable for that year is coded as 1; after that year "war-onset" is coded as 0 and the control variable "ongoing-war" is coded as 1 until the conflict ends. For example, Israel is observed from 1948 to 2005, but it is only coded as having a civil conflict onset once, in 1949. From 1950 to 2005, Israel is coded as 0 for the onset variables and 1 for ongoing-war. This coding scheme assumes that the conflict was static throughout the years, and masks significant events like the First and Second Intifadas, which have claimed the lives of over 8,500 Israelis and Palestinians. We feel, then, that Israeli civil conflict is undercounted in the EPR.

Another example is Kurdish rebellion in Turkey, which started in 1984. War-onset was coded as 1 in 1984 and 0 up to 2005. According to the EPR, the war is still ongoing. However, the conflict has not been constant; it has seen peaks and caesuras. For example, the PKK (according to the EU, a terrorist organization of Kurdish rebels) stopped violence for 5 years starting in 1999. When the Turkish courts ordered the closure of the Kurdish political party in 2004, the PKK began violent activities again. Turkey has three positive onset variables, two unrelated to the Kurdish

[3] All values displayed were calculated on the multiply-imputed, unmatched data. However, we did run several regressions on the CEM-treated and nearest-neighbor treated data. The significance of GDP per capita survived both, while the significance of excluded population survived only the latter. (Tables not shown, but code is included in the replication file.)

issue. But the Kurdish issue is by far the most important conflict and deserves more salience in any analysis of ethnic relations and civil conflict in Turkey. In a way, we would prefer using a dependent variable that recorded new "campaigns" or waves of ethnic conflict, rather than simply the start of an ethnic war. We believe that the dependent variable as is undercounts the long term conflicts that are precisely the most consequential form of ethnic conflict.

Lastly, we caution that the data do not code for major historical events that substantially change the underlying probabilities of civil conflict. The figure below shows the frequency of years in which there is an onset in Eastern Europe and Latin America. As seen, all but one conflict in Eastern Europe occurred after 1989, whereas the distribution for Latin America is considerably more uniform.

Quantities of Interest in Wimmer et al.'s Original Results

In their conclusion, Wimmer et al. write that their results "represent a major challenge to the greed-and-opportunity school, which discounts ethnicity as a relevant factor in explaining civil war" (334). While they are right that ethnicity should not be discounted in models of civil war, their dismissal of greed-and-opportunity is premature. While the statistical significance of "log of the percentage of the population belonging to an excluded ethnic group" and "number of ethnic groups with access to state power" may have held up to many controls (and matching, in the case of the former), the authors' investigation of the substantive importance of their explanatory variables relative to GDP per capita is brief. In the appendix, they show that GDP per capita and excluded population have similar first differences when moving up a standard deviation from the mean. However, we believe this comparison to overstate the importance of excluded population to their model.

In the following section, we exactly replicate their first model, performing a logistic regression of the "civil conflict onset" indicator onto the following covariates: Log of percent of ethnic population excluded from state power, Number of power sharing groups, Imperial Past, Lagged GDP per capita, Lagged log of population size, Lagged ongoing war indicator, Year, Number of years since onset of conflict, and three spline terms. According to Beck, Katz, and Tucker (1998), the inclusion of these terms will account for the dependence of observations on the same country over time.

Variable	Observations	Mean	Std. Dev.	Min	Max
Exclusion (ln)	7138	1.864	1.589	0	4.595
Power Sharing Groups	7138	1.638	1.856	0	14
Imperial Past	7155	0.475	0.314	0	1
GDP per capita (lagged and in 1000 USD)	6990	5.968	7.292	0.028	110.315
Population (ln and lagged)	7060	9.188	1.390	5.581	14.076
Conflict Onset	7155	0.030		0	1
Ethnic Conflict Onset	7155	0.015		0	1

Table 1: *Descriptive statistics of explanatory variables*

We then estimated the first difference associated with the interquartile ranges of "excluded population" and GDP per capita. We also estimated the first difference associated with going from the minimum to the maximum value of excluded population, and for going from the 75^{th} to the 95^{th} to the 100^{th} quantile for GDP per capita. While the max-to-min first difference and the 95^{th}-to-100^{th} are not valid counterfactuals, here we are simply evaluating the author's inferences from their own model.

	First Differences (95 % CI)	Risk Ratio (95 % CI)	$E(Y\|X_0)$	X_1	X_0
Excluded Population (1^{st} to 3^{rd} Quartile)	0.010 (0.005, 0.016)	1.79 (1.30, 2.37)	0.014	3.3	0
Excluded Population (Max versus Min)	0.016 (0.007, 0.027)	2.25 (1.46, 3.36)	0.014	4.6	0
GDP per Capita (3^{rd} to 1^{st} Quartile)	0.018 (0.011, 0.025)	2.10 (1.51, 2.79)	0.017	7.6	1.29
GDP per Capita: (95^{th} percentile vs 3^{rd} Quartile)	0.013 (0.011, 0.016)	5.43 (2.69, 9.64)	0.004	21.6	7.6
Max Excluded Pop, and GDP: (95^{th} quantile vs 3^{rd} quartile)	0.021 (0.016, 0.027)	5.34 (2.62, 9.83)	0.006	22	8

Table 2: *Relationship between GDP per capita and Civil Conflict in Wimmer's et al. original model*

As the above table shows, moving from the 1^{st} to 3^{rd} quartile of logged excluded population is associated with moving from a 1.4 % chance of civil conflict to a 2.4% chance. Equivalently, the risk ratio shows that this move toward greater ethnic exclusion makes civil conflict 1.8 times more likely.

Moving from the 3^{rd} to the 1st quartile for GDP per capital is almost twice as important substantively than the equivalent move in excluded population: in fact, it is most comparable to moving from 0% excluded population to 99.5% excluded population. But even this understates the importance of GDP per capita, because this variable is heavily skewed. The 3rd quartile gets us to states like Romania, Gabon, and Kazakhstan, while we are yet to estimate the stability of states like Portugal , Malta, and Estonia. In row four of the above table, we see that the estimate probability of civil conflict in a state with Portugal's GDP per capita is less than half of a percent. By the time we get to a state with Luxembourg's GDP per capita, the probability of civil conflict is effectively zero. [4]

The fifth row, while admittedly an exercise in extrapolation, drives home the implications of Wimmer et al's model. We estimated the probability of civil conflict in a states with GDPs equivalent to Portugal and Kazakhstan, given the maximum value on ethnic exclusion and mean values on all other covariates. While Kazakhstan would be unstable (probability of conflict around 2.7%), Portugal would be remarkably stable (probability of conflict around 0.6%). Referring back to row four, this shows that an entirely exclusive Portugal would be about as stable as a Kazakhstan with average exclusiveness.

Because of model dependency and extrapolation bias, we do not recommend taking the above numbers too seriously. Nonetheless, these comparisons are nonetheless telling. The authors' conclusions require considerable scope conditions: ethnic exclusion is substantively important only in countries with low to modest levels of economic development. [5]

Number of Included Groups

We also estimated first differences for "number of groups in power sharing coalition," and found a similar story. Over three quarters of our observations have values from 0-2 on the "included groups" category. But as the first row shows, including one extra ethnic group in a coalition has statistically significant but ultimately negligible effect on state stability: an additional 0.1 percentage point

[4] Results not displayed. See replication file for relevant code.

[5] Of course, this would mean that ethnic mechanisms are still important in three-quarters of countries. We attach further scope conditions in the following section.

chance of civil conflict. We suspect that this variable's significance may derive from unusual states like India (14 ethnic groups in the power sharing coalition; see row 2). While power sharing may well cause instability in such a complex situation, we conclude that the vast majority of states are not affected by this variable.

	First Differences (95 % CI)	Risk Ratio (95 % CI)	X_1	X_0
Included Groups (1^{st} to 3^{rd} quartile)	0.001 (0.000, 0.003)	1.08 (1.02, 1.13)	2	1
Included Groups (Max versus Min)	0.034 (0.006, 0.082)	3.14 (1.34, 6.33)	14	0

Table 3: *Relationship of Group power sharing to Civil Conflict in Wimmer's et al. original model*

Regional Differences in Ethnic Power Relations

Inherent in the choice to use a variable in social science analysis is the assumption that the values of that variable have the same meaning across observations. This ideal is perhaps never met in cross-national research, as attested to by anecdotes about back-of-the-envelope calculations of GDP in the least developed nations. While most analysts nonetheless accept the use of estimates of GDP and population size, the case is much harder to make for the ethnic variables central to Wimmer et al's models.

The authors write: *"Qualitative comparative work shows the importance of taking different ethnopolitical constellations into account and of acknowledging the causal heterogeneity of the processes that lead to ethnic conflict"* (319). However, the authors take this to mean that different ethnopolitical constellations predict different kinds of conflict, rather than asking whether the same ethnopolitical constellations have different effects depending on each country's ethnic history.

For example, one can reasonably argue that being an excluded ethnic group in a weak state like Afghanistan is a fundamentally different experience than being an excluded ethnic group in a stronger state like China. In the New World, excluded ethnic groups will often be of different ethnosomatic background than the dominant ethnic group, while sharing a linguistic and religious heritage (for example, African-Americans in the United States). In Nigeria, exclusion will be ex-

perienced along tribal and linguistic lines. In the Muslim world, much of exclusion is religious or sectarian in nature. And of course, from the Balkans to the Caucasus to India, ethnic exclusion comes in every permutation of the three. Furthermore, each ethnic relationship is idiosyncratic, steeped in long historical experience.

With this last thought in mind, it is worth noting that when we ran a fixed effects model, not a single covariate was significant at the 0.05 level, although Imperial Past came close (p = 0.055).[6] This is a reminder that a study such as this will always be subject to criticisms of omitted variable bias, as many of the relevant covariates - intensity of ethnic hatred, degree of perceived oppression and exploitation, and the aforementioned "idiosyncratic ethnopolitical constellations" - resist easy quantification. Nonetheless, we will proceed by interacting regional dummies with excluded population to represent each region's unique ethno-cultural inheritance.

We predict that excluded ethnic population will be most consequential in Eastern Europe and sub-Saharan Africa. We believe that these regions had similar post-colonial opportunity structures. For significant portions of the 20^{th} century, different ethnic groups were forced to live under a single state by a foreign power. Upon independence, multiple ethnic actors had realistic chances at seizing power or forming their own state. In the 20^{th} century, these opportunities were many fewer in the West, Latin America, Asia, and the Muslim world.

The Model

We use the dummy variable indicating onset of a new ethnic conflict for this portion of the paper, because the hypothesis is specifically concerned with how ethnic power relations affect civil conflict. We regress ethnic conflict onset on the same covariates as Wimmer et al, except we have added dummies for each region and interactions between these dummies and the percentage of the population excluded from state power. "North Africa and the Middle East" was used as the reference category.

Unlike Wimmer et al, we display results from the rare-events correction for logistic estimators. Wimmer et al. note that they tried the rare-events correction but found no substantive differences.

[6]No results are shown for the fixed effects model. However, the relevant code is included in our replication files.

This is true if one looks only at the regression output. However, we found that the rare-events estimator gave us somewhat different estimates for certain quantities of interest.

Results

Our hypothesis was partially confirmed. The interaction for eastern Europe is significant (0.938, p = 0.047), although the interaction for sub-Saharan Africa was not (p = 0.305). The number of power sharing groups, lagged GDP, and lagged population all remained significant.

To interpret these results, we will compare hypothetical Eastern European countries Middle Eastern countries at various levels of exclusion.

EE ∼ Eastern Europe ME ∼ Middle East	First Differences (95 % CI)	Risk Ratio (95 % CI)	$E(Y\|X_0)$	X_1	X_0
Eastern Europe	0.007	11.74	0.002	1.96	0.00
(Median vs No Exclusion)	(0.003, 0.014)	(1.49, 40.47)			
Eastern Europe	0.026	4.85	0.009	3.30	1.96
(3^{rd} vs 2^{nd} Quartile)	(0.006, 0.056)	(1.36, 11.74)			
Eastern Europe	0.106	3.91	0.035	4.60	3.30
(Max to 3^{rd} Quartile)	(0.008, 0.317)	(1.42, 7.78)			
Middle East	0.002	1.42	0.011	1.96	0.00
(Median vs No Exclusion)	(-0.011, 0.008)	(0.60, 2.90)			
Middle East	0.003	1.27	0.013	3.30	1.96
(3^{rd} vs 2^{nd} Quartile)	(-0.005, 0.012)	(0.76, 2.07)			
Middle East	0.004	1.24	0.016	4.60	3.30
(Max to 3^{rd} Quartile)	(-0.004, 0.021)	(0.732, 2.01)			
No Exclusion:	0.009	25.21	0.002	ME	EE
(ME vs. EE)	(-0.003, 0.033)	(0.64, 145.9)			
Median Exclusion:	0.004	1.82	0.035	ME	EE
(ME vs. EE)	(-0.009, 0.017)	(0.49, 5.14)			
3^{rd} Quartile Exclusion:	-0.020	0.48	-1.71	ME	EE
(ME vs. EE)	(-0.05, 0.000)	(0.19, 1.03)			
Max Exclusion:	-0.128	0.21	0.148	ME	EE
(ME vs. EE)	(-0.376, -0.010)	(0.03, 0.70)			

Table 4: *Relationship between Ethnic Exclusion and Ethnic Conflict in Middle East and Eastern Europe*

Rows 1 through 3 show that Eastern European nations are relatively stable, except in the presence of extreme levels of ethnic exclusion. Row 3 shows that an Eastern European country with 27%

of its population excluded from state power has a 3.5% chance of experiencing civil conflict. Rows 4-6 show that our hypothetical Middle Eastern countries are moderately stable regardless of levels of ethnic exclusion. Rows 7-10 compare Eastern European countries to Middle Eastern countries directly. While Eastern Europe is 25 times more stable than the Middle East with no ethnic exclusion, the Middle East is twice as stable as Eastern Europe with 27% exclusion, and 5 times as stable at 99.5% exclusion (although this final estimate is certainly an extrapolation and should be treated with caution).

Conclusion

We acknowledge the difficulties inherent in quantitative studies of this type: the world has not given us a well-balanced dataset. We therefore urge authors to be modest about their findings and to be upfront about model dependency - and for readers to do the same for our results. We believe we are on strong theoretical ground in adding regional interaction terms to the model. Our conclusion - that ethnic mechanisms operate differently depending on the regional context - deserves further analysis.

As we have shown above, a significant portion of the relationship between excluded population and ethnic conflict is driven by the Eastern European nations. Throughout the Soviet era, Eastern Europe was the most stable region on the planet, experiencing only a single civil conflict in 1948.[7] Yet the fall of the Soviet Union saw Eastern Europe turn into one of the most civil violence prone regions, especially in the ethnically diverse countries of former Yugoslavia.

So what social force was responsible for this sudden wave of violence in Eastern Europe? The relative stability of ethnically homogenous nations like the Baltic states - which provide stark contrasts to the experiences of the Balkan states - does point us toward Wimmer et al's ethnic mechanisms. While the ethnic composition of the region has not changed much from 1948 to the present, the relations of these ethnic groups to their state did change drastically. But should we then attribute causal primacy to the reconfiguration of ethnic-institutional relations, or to the sudden disappearance of the regional hegemonic power that made this ethnic reconfiguration possible?

[7] according to the Ethnic Power Relation dataset, in any case. This characterization will certainly be doubted by some.

Both explanations have worth, and it is not obvious we need to choose between the two. Yet bringing the Soviet Union back in (or another destabilizing influence for a different region) alters the theoretical conclusions reached by Wimmer et al. The fall of the Soviet Union opened enormous opportunities for mobilization by all varieties of local actors. In the Balkan states in particular, these opportunities were seized by ethnically-motivated groups. Therefore we cannot agree with Wimmer et al that their results are a challenge to the "greed-and-opportunity" model of civil conflict: state strength (as represented imperfectly by GDP per capita in our models) is still the predominant predictor of civil conflict. However, given low state capacity - whether because of perennially low levels of development or because of a sudden exogenous shock like the fall of the Soviet Union - ethnic mechanisms do have additional power in predicting where civil conflict will erupt. We therefore urge future scholars to consider Wimmer et al's findings as an addendum to - not a refutation of - the "greed-and-opportunity" model of civil conflict.

Bibliography

Beck, Nathaniel, Jonathan N. Katz, and Richard Tucker. 1998. "Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable." Journal of Political Science 42:1260-88.

Fearon, James D., Kimuli Kasara, and David D. Laitin. 2007. "Ethnic Minority Rule and Civil War Onset." American Political Science Review 101:187-93.

Fearon, James D. and David D. Laitin. 2003. "Ethnicity, Insurgency, and Civil War." American Political Science Review 97:1-16.

Gellner, Ernest. 1991. "Nationalism and Politics in Eastern Europe." New Left Review 189:127-43.

Ho, Daniel, Kosuke Imai, Gary King, Elizabeth Stewart. 2009. "MatchIt: Nonparametric Preprocessing for Parametric Causal Inference."

Honaker, James, Gary King, Matthew Blackwell. 2010. "Amelia II: A Program for Missing Data." http://gking.harvard.edu/amelia/

Iacus, Stefano M, Gary King, Giuseppe Porro. 2010. "Coarsened Exact Matching." http://gking.harvard.edu/cem/

Imai, Kosuke, Gary King, and Oliva Lau. 2008. "logit: Logistic Regression for Dichotomous Dependent Variables" in Kosuke Imai, Gary King, and Olivia Lau, "Zelig: Everyone's Statistical Software," http://gking.harvard.edu/zelig

Nairn, Tim. 1993. "All Bosnians Now?" Dissent Fall:403-10.

Snyder, John; Carl Gershenson; Burak Eskici, 2010, "Replication data for: Ethnic Power Rela-

tions", hdl:1902.1/14617 UNF:5:zy5BK1SWfJxsxGjEC6LRKA==

Stoll, Heather, Gary King, Langche Zeng. 2009. "WhatIf: Software for Evaluating Counterfactuals." http://gking.harvard.edu/whatif/

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 Vol. 74:316-337.