Disaggregating Ethno-Nationalist Civil Wars: A Dyadic Test of Exclusion Theory

Halvard Buhaug, Lars-Erik Cederman, and Jan Ketil Rød

Word count: 7,999

(Includes everything except front page, acknowledgements, and abstract)

Acknowledgements

Halvard Buhaug

Centre for the Study of Civil War (CSCW), PRIO

0186 Oslo, Norway

halvardb@prio.no

Lars-Erik Cederman

International Conflict Research

ETH Zurich

8092 Zurich, Switzerland

lcederman@ethz.ch

Jan Ketil Rød

Department of Geography

Norwegian University of Science and Technology (NTNU)

7491 Trondheim, Norway,

and CSCW, PRIO

jan.rod@svt.ntnu.no

Disaggregating Ethno-Nationalist Civil Wars: A Dyadic Test of Exclusion Theory

Halvard Buhaug, Lars-Erik Cederman, and Jan Ketil Rød

Abstract

Contemporary conflict research usually measures the influence of ethnicity on conflict by capturing ethnic constellations as country-based indices, such as ethnic fractionalization or polarization. However, such aggregated measures are likely to conceal the actual operation of actor-specific mechanisms. In this paper, therefore, we introduce a disaggregated model that measures ethnic groups' access to power. We do so by disaggregating both ethnicity and conflict to the level of explicitly geo-coded center-periphery dyads. This procedure allows us to measure the power balance between politically excluded ethnic groups and dominant actors in terms of group sizes, distances between the center and the periphery, and the roughness of the latter's terrain. We rely on geographic information systems (GIS) to compute demographic and ethno-geographic variables. The dyadic analysis enables us to show that exclusion of powerful ethnic minorities increases the likelihood of conflict considerably. In addition, we show that the risk of conflict is positively associated with the extent of rough terrain in the peripheral group's home region and its distance from the political center.

Introduction

Recent quantitative studies of civil wars have questioned the impact of ethnic grievances on the onset of civil wars. As with so many other findings in the political-economy literature, however, this "non-result" is based on highly aggregated proxies.[1] In most cases, country-level indicators, such as GDP per capita or ethnic fractionalization indices, do not allow us to distinguish one explanation from another and are rarely supported by evidence at lower levels of aggregation.[2] Indeed, the debate over the role of ethnicity in internal conflict is far from resolved.

As a way to overcome these shortcomings in the literature, this paper disaggregates civil wars to the level of dyads pitting states against their ethno-nationalist challengers. This approach enables explicit empirical evaluation of actor-centered theories that expect violence to erupt where strong peripheral actors are excluded from state power. Thus, in this study, we do not ask whether ethnic power struggles generate violent conflict, but how.

Focusing specifically on dyadic ethno-nationalist conflicts, our analysis enables us to measure the center-periphery power distribution in terms of demographic and geographic variables. This is an important contribution to the literature, because unlike interstate warfare, civil wars do not lend themselves to straight-forward measurement of power

---

[1] E.g. Fearon and Laitin 2003; Collier and Hoeffler 2004.

[2] Sambanis 2004; Ross 2006; Tarrow 2007.

balances. In order to compute the logistical variables and population estimates, we make extensive and innovative use of geographic information systems (GIS).[3]

Our findings suggest that the dyadic power balance, measured as the peripheral group's demographic size compared to the center, is indeed a strong predictor of ethno-nationalist conflict. We also find evidence supporting the hypothesis that dyads with groups far from the capital and those located in rough terrain are more likely to experience conflict. Standard measures of wealth – measured at the national level – appear to be at best only weakly related to ethnic conflict at the dyadic level of analysis.

Postulating dyadic conflict mechanisms

In the following, we propose a group-based model of ethno-nationalist conflict that is based on a star-shaped actor constellation and on corresponding conflict mechanisms connecting the state's governmental center with its periphery.[4] Drawing directly on Wimmer's exclusion theory, we interpret ethno-nationalist conflict as the consequence of one particular path of nation-state formation.[5] It occurs where the elites of nationalizing states were not resourceful enough to include and integrate the vast majority of the population into the imagined community of the nation. Especially in societies where no networks of civil society organizations were available, they relied on ethnic clientelism to

---

[3] For an earlier application of GIS in quantitative analysis of civil war, see Buhaug and Rød 2006.

[4] Cederman and Girardin 2007.

[5] Wimmer 2002.

mobilize a political following, thus excluding ethnic others from access to power and public goods.

Whereas nation-building in strong and resourceful states with well developed civil societies incorporated large parts of the population, sometimes with the help of widespread ethnic cleansing, the situation in weaker states, mostly in the previously colonized parts of the world, was quite different. Exclusion along ethnic lines often led to political mobilization by counter-elites, which denounced their under-representation at the center of power and demanded inclusion (e.g. the ANC in South Africa), or the creation of a separate state in which their own ethnic group would represent the dominant nation (e.g. the Bengalis of former East Pakistan, now Bangladesh).

If this exclusion perspective is correct, we would expect powerful ethnic groups excluded from power to be most likely to mobilize around an ethno-nationalist program and to initiate conflict against the government, and conversely, the government to engage in repression to curb the power of such threatening contenders. Thus, we postulate that collective ethno-nationalist action requires both opportunities and willingness to challenge central state power.[6] Accordingly, civil wars occur when peripheral contenders to the government are powerful enough to challenge the center and sufficiently motivated to do so.[7] In the following, we elaborate on how the proposed mechanisms are constituted and how they can be measured.

---

[6] Most and Starr 1989.

[7] Gurr 2000.

For operational reasons, and without any claim to completeness, we focus on two crucial aspects of ethnic center-periphery relationships which we expect to systematically influence both the peripheral group's opportunities and willingness to rebel: (1) the demographic balance between the government and the peripheral ethnic group, and (2) the ethno-geographic constellation – especially with respect to the distance between the capital and the peripheral group – and the roughness of the terrain within the group's settlement area.

Other things being equal, we expect that larger groups will be able to stage successful collective action thanks to their superior numbers. While the former aspect of the hypothesis can be justified both in terms of resource mobilization, the latter one hinges on legitimacy.[8] We would expect the frustrations of being excluded and, correspondingly, motivations to engage in rebellious collective action to increase with the size of the excluded population.[9] This reasoning can be summarized as follows:

> H1. The probability of conflict increases with the relative demographic size of the excluded group.

However, the size-related hypothesis needs to be complemented, because in large states with difficult terrain, even relatively small groups may be able to wage surprisingly

[8] DeNardo 1985; McCarthy and Zald 1977.

[9] Horowitz 1985.

effective campaigns against powerful political centers, as the insurgencies in Aceh and Chechnya have demonstrated. On average, then, we would expect the prospects of peripheral challenges to the central government to be the most successful in the cases where the latter's reach is least developed. This logic yields two straight-forward hypotheses:

> H2a. The probability of conflict increases with the distance between the excluded group and the capital.

> H2b. The probability of conflict increases with the roughness of the terrain within the settlement area of the excluded group.

We argue that remoteness in these two senses supports excluded groups' efforts to challenge the center's power both in terms of opportunities and willingness to stage insurgencies. The former argument is directly related to the geographic reach of state power, as suggested by Boulding's notion of the "loss-of-strength gradient" (LSG).[10] This notion postulates a declining function of projected power over distance, often expressed as exponential decay.[11] In addition to this materialist mechanism, it can be expected that the cultural penetration of the center declines with distance and geographic obstacles.[12]

---

[10] Boulding 1962.

[11] See Buhaug 2007.

[12] Rokkan 1999.

Operationalizing ethno-nationalist dyads

As has been argued above, the most obvious way of evaluating the exclusion perspective is to model it as a center-periphery dyad. By measuring ethnic dyads directly and introducing an explicitly geographic dimension, the current study provides a fine-grained picture of the mechanisms driving ethno-nationalist wars. Our goal is to study conflict between ethnically defined state authorities, i.e. the "ethnic group(s) in power" and their challengers.[13] To succeed in our endeavor, we need information on:

> (1) the identity and location of ethnic groups,
>
> (2) demographic group sizes,
>
> (3) ethnic group(s) in power,
>
> (4) geo-coded data on distances and terrain, and
>
> (5) ethnic dyadic conflicts.

In the following, we describe our data-collection efforts before turning to a presentation of the results in the subsequent section.

(1) Geocoding ethnic groups

To our knowledge, there is no ready-to-use dataset that systematically pins down the location of ethnic groups in a large number of comparable cases. To create such a dataset, we chose to focus on the well-known *Atlas Narodov Mira* (*ANM*).[14] The *ANM* stems

---

[13] Cederman and Girardin 2007.

[14] Bruk and Apenchenko 1964.

from a major project of charting ethnic groups undertaken by Soviet ethnographers in the early 1960s. Their efforts bore fruit in the extensive but still un-translated atlas, containing 57 ethno-geographic maps at various scales that jointly cover all inhabited parts of the world.

The *ANM* has several strengths: it is complete and carefully researched, it relies on a consistent classification of ethnicity, it offers a uniform group list that is valid across state borders, and it provides high-quality maps. Among its weaknesses, it should be stressed that the *ANM* adopts an ethno-geographic rather than a political perspective, and is based on data available in the 1960s. In most cases, however, ethnic settlement patterns exhibit considerable inertia, so it seems reasonable to use this dataset as a starting point.

In order to make the data of the *ANM* available for statistical analysis, a number of preparatory steps had to be conducted. First, all of the *Atlas*' maps had to be converted to a digital format through scanning. Then, the scanned maps were geo-referenced, meaning that they were linked to a particular coordinate system specifying locations on the earth's surface. Once referenced, group polygons were replicated from a computer screen displaying the geo-referenced scanned map using the mouse as digitizing cursor. Finally, all legends and group names were translated from Russian into English and entered into the dataset's attribute table with links to the corresponding group polygons. The resulting dataset, labeled GREG (Geo-Referencing of Ethnic Groups), contains information on the

10

geographical location of more than 1,600 ethnic groups identified in the *ANM* allowing computation of spatial measurements such as area and distance.[15]

(2) Estimating group sizes as share of state populations

Having determined the geographic location of all ethnic groups, our next task is to construct a suitable measure of the power balance in the center-periphery dyads. Because it is difficult to find direct indicators of group strength in such settings, we decided to use demographic estimates of group size. One possible source of such information would be the *ANM*. In addition to the map material, the *Atlas* contains group-size estimates, which have often served as a basis for the calculation of ELF values.[16] However, these statistics may not be ideal for our purpose because a number of states did not exist at the time of publication, such as the post-Soviet republics of Russia, Ukraine, Belarus, etc. How could one derive population statistics of ethnic groups in these states based on the Soviet data?

Fortunately, GIS provides a convenient solution to this dilemma by enabling complex computational operations involving several spatially arranged data layers. Assuming the ethnic map to be constant, we used a spatial estimation method based on an intersection of territorial country masks, the group polygons, and population density maps. Using the boundaries in a country layer (shapefile) representing the Post-Cold War period as "cookie cutters," we singled out the group polygons (or parts thereof) that fell within the borders of each state. We then intersected the ethnic group layer with a gridded

[15] See Cederman et al. 2006.

[16] E.g. Taylor and Hudson 1972.

11

population density layer from Columbia University.[17] This allowed us to measure the size of the population that fell within a given ethnic group's "state-cropped" polygon(s). In cases of multiple groups per polygon, we divided the population figures evenly between the groups. By summarizing the population for all polygons belonging to an ethnic group and repeating this procedure for all groups we obtained population estimates for all groups in all countries in the *ANM*. In a further step, we divided these numbers by the total population for the corresponding country, which yielded a table of the relative population share for each ethnic group.

To limit the computational effort, we decided to generate the spatially derived population shares for two years only; 1964, the *ANM*'s year of publication, and 1994, which captures the post-Cold War situation. We then merged the two datasets, using the estimates from 1964 for countries that existed in that year and data for 1994 for younger states. Finally, by way of validating the spatially computed data, we compared our group size estimates from 1964 to the population data provided by the *ANM*. This simple test indicated a very high correlation, at $r = 0.99$. Validation of the 1994-based estimates posed more problems due to the difficulty of finding matching group data that can serve as a reference. Nevertheless, a comparison with the population statistics for the corresponding ethnic groups in Fearon's dataset yielded a reasonably high correlation coefficient ($r = 0.93$).[18]

---

[17] CIESIN 2005.

[18] Fearon 2003.

(3) Identifying the ethnic group(s) in power

So far, our discussion has focused on groups and polygons, but this study concerns *dyadic* dynamics. The notion of "ethnic group(s) in power" (EGIP) that we introduced above provides the crucial piece of information that makes this conceptual step possible. In our empirical analysis, we follow Cederman and Girardin in considering a group, or a coalition of groups, to be in power if their leaders serve (at least intermittently) in senior governmental positions, especially within the cabinet.[19] Thus we focus on ethnic groups' influence over the executive at the national level rather than on their legislative or local power. In addition to the ethnic background of senior cabinet members, specific institutional arrangements, such as different types of power-sharing and consociationalism, may also be indicators of power inclusion. By power sharing, we mean any arrangement that divides the access to power among the groups making up the governing coalition. Accordingly, EGIPs can consist of more than one group. For example, we code all the four language groups of Switzerland as constituting the EGIP (again, see our online supplement).

We build directly on the coding of Cederman and Girardin, who rely on Fearon's list of ethnic groups.[20] Principal sources of information for identifying the EGIPs are Gurr's Minorities at Risk (MAR) database, the CIA Factbook, and a new dataset on ethnic affiliation of political leaders, collected by Goemans et al.[21] Drawing on expert input

---

[19] Cederman and Girardin 2007.

[20] Fearon 2003.

[21] Goemans et al. 2004.

from an ongoing online survey, we have made an effort to improve the EGIP coding in several country cases. Wherever deemed appropriate, we introduced period-dependent EGIPs to capture major shifts in the political constellation of power access (e.g. Afghanistan and Yugoslavia).

By definition, any group not coded as an EGIP is a marginalized ethnic group (MEG). We can now form center-periphery dyads as pairwise constellations of a country's EGIP and each of its MEGs. Given the enormous difficulty of identifying EGIPs in Sub-Saharan Africa, we decided to limit our empirical sample to Eurasia and North Africa, thus covering roughly half of the world's states.

Our key independent variable, r, measures the power balance between the EGIP and the marginalized group. It is operationalized as the periphery's share of the dyadic population. To illustrate, consider the Basque minority in Spain. According to our GIS-generated estimates, the Basques comprise 5.5 % of Spain's population whereas the Spaniards, the EGIP, account for 67.7 %. The MEG's relative power is then calculated as $r = 5.5 / (5.5 + 67.7) = 0.075$. As one would expect, the random dyad is overwhelmingly dominated by the center (mean r score is 0.019, meaning that the MEG comprises a mere 1.9 % of the dyadic population). Only four marginalized ethnic groups in our sample are larger than the groups in power ($r > 0.5$).[22] In the statistical analysis below, we apply a logarithmic transformation of the power balance to moderate a highly right-skewed

---

[22] These are the Serbs in Bosnia and Herzegovina (1992-95), Arabs of Saudi Arabia in South Yemen (1967-90), Afghans in Afghanistan (1992-95) and Bengalis in Pakistan (1947-71).

distribution of values. In addition, it seems reasonable to assume that an increase in a group's relative size by one percentage point should make a much larger difference for small groups than for large ones.

 (4) Geocoding distances and terrain

It goes without saying that using raw demographic data as a proxy for relative power leaves much to be desired. As argued above, power decays as it is projected across distance, so politically excluded minority groups may offset some of their inferiority by residing in remote provinces and enjoying the protection of rough terrain. To account for the role of geographic mechanisms, we thus need data on the relative location of the MEGs and extent of mountainous terrain in their home regions.

The location of the ethnic groups can be readily determined by means of geo-referenced polygons. We used GIS to extract their precise mean center, in latitude and longitude coordinates. Groups belonging to the EGIP are by default coded with the capital city as their location. We then ran a script that measured the geodesic distances from each MEG's polygon centroid to the capital city (see Figure 1). For MEGs represented by two or more polygons, we generated a weighted distance measure that gives the average centroid-capital distance for all locations of the group, weighted by the population size in each polygon. The population weighting is necessary to prevent a distortion of distances due to small clusters far away from the group's core settlement area. To reduce outlier influence and to account for an expected declining effect of distance with higher values, we take the natural logarithm of the distance variable.

[FIGURE 2 ABOUT HERE]

The second geographic proxy, that measures the extent of mountainous terrain, was computed in a manner similar to that of group populations. By intersecting gridded mountain data with the boundaries of the ethnic groups, we were able to calculate the share of the two-dimensional area of each polygon covered by mountains.[23] For groups represented by multiple polygons, we weighted the average terrain values by the area of each polygon. The area-weighted mountain variable takes on values between zero (for groups in the plains) and one (for groups settled entirely in the mountains).

(5) Determining the onset of dyadic ethnic conflict.

The final step of data generation concerns the dependent variable, namely the outbreak of dyadic ethnic conflict. We relied on the UCDP/PRIO Armed Conflicts Dataset v. 4-2006, henceforth ACD,[24] which is arguably the most established country-level dataset on armed conflict. It applies a comparably low minimum casualty threshold of just 25 annual battle-related deaths and allows for and distinguishes between multiple simultaneous conflicts within single countries.

---

[23] Mountain data obtained from UNEP-WCMC 2002.

[24] Gleditsch et al. 2002.

16

Our operational definition of ethnic conflict is similar that used by Fearon and Laitin,[25] which focuses on all conflicts in which the parties are organized primarily along ethnic lines. We make no attempt to uncover whether the contested issues are truly linked to ethnicity. In order to separate out the ethnic rebellions from the total sample of intrastate conflicts in the ACD, we first identified those conflicts coded by other datasets as ethnic and dropped the ideological ones.[26] For the remaining conflicts, we conferred relevant sources (UCDP database; Keesing's Online; Wikipedia.org; Globalsecurity.org; the MAR project) on a case-by-case basis to determine whether mobilization was based on ethnic affiliation. Finally, we identified the MEG(s) that were involved in each ethnic conflict.[27] In cases where more than one group challenged the capital at the same time, an onset of conflict was recorded in each of the corresponding dyads. Ongoing years of conflict are coded as missing. In case of a lull in the fighting or a peace agreement that lasts for at least two consecutive calendar years, the next observation of conflict in the dyad is coded as a new onset.

---

[25] Fearon and Laitin 2003.

[26] Some conflicts identified as ethnic wars by other data sources are excluded from our sample. In most cases this is due to lack of differentiation between the warring parties in the *ANM* list, though there are also discrepancies in the way some conflicts are coded in various datasets.

[27] Incidentally, our candidate list also includes three cases of ethnic conflict within multiple-group EGIPs: Ukrainians vs. the USSR (outbreak 1946), Assamese vs. India (1990, 1994), and Panjabis vs. India (1983). Since these cases represent center-center rather than center-periphery dyads, they were automatically dropped from the analysis. Nevertheless, elite infighting is an interesting form of ethnic conflict that can also be studied with disaggregated methods.

To facilitate comparison with country-level studies, we also generated an aggregated version of the dependent variable. Here, country-years in which there are one or more onset(s) of dyadic conflict are coded as "1" while observations without onset (including ongoing years of conflict) are coded "0". In all, the dyadic dataset contains 110 onsets of ethnic conflict among the roughly 35,000 valid dyad-years (1946–2005) while the monadic dataset includes 55 onsets in 3,463 country-years.

In addition to our key dyadic variables as described in (3) and (4) above, the statistical analysis includes a number of country-level controls. These include Fearon and Laitin's comprehensive selection of explanatory variables (GDP per capita, log of population, log of mountainous terrain, non-contiguous territory, oil exports, new state, instability, democracy score, and ethnic and religious fractionalization), Gleditsch's logged GDP per capita,[28] Cederman and Girardin's ethnic exclusion index (N^*),[29] a variable measuring annual calendar time, and controls for duration dependence. All reported models are estimated by robust logit regression, clustering the standard errors on the countries in the dyadic analysis.

Analyzing country-level effects

We start by testing models of ethnic conflict at the country level. As our reference point, we apply Fearon and Laitin's insurgency model to our restricted sample of Eurasia and

---

[28] Gleditsch 2002.

[29] Cederman and Girardin 2007.

North Africa. Model 1 in Table 1 presents the findings of this replication analysis, with

Fearon and Laitin's coding of "ethnic war" regressed on their standard set of independent

variables. In order to exclude countries not at risk of ethnic conflict, Model 1 is limited to

country years with at least a 5% ethnic minority.[30]

As expected, this regression produces similar results to Fearon and Laitin's global model.

Even though the replicated estimates are slightly weaker due to the reduced sample size,

the variables that are significant in their model also perform well in our replication. The

overall correspondence to the original findings gives little reason to suspect substantial

systematic bias in the reduced spatial domain of this study.

Our next step is to replace Fearon and Laitin's ethnic wars with our country-level

measure of ethnic conflict (and an accompanying "prior war" dummy), derived from the

ACD (see Model 2). Given the latter dataset's more comprehensive definition of conflict,

which captures multiple onsets per country and numerous sporadic episodes of violence,

it is not surprising that Model 2 tells a somewhat different story. Most strikingly, the

more inclusive dependent variable completely obliterates the effect of GDP per capita.

While wealth is often widely acclaimed as an effective insurance against civil war, it does

not seem to be a guarantee against less severe ethnic unrest in our sample.

A second noteworthy deviation from Model 1 is that the population estimate has

increased considerably and the z score has almost tripled. This is consistent with the fact

---

[30] See Fearon and Laitin 2003, 84.

that populous countries, on average, not only contain a higher number of ethnic groups, but also tend to cover a larger geographical area, implying that some segments of the population reside at a considerable distance from the capital city. If these peripheral groups constitute marginalized ethnic minorities, the combination of cultural and geographic distance is likely to amplify their opportunities and desire for self-determination.

The dramatic increase in the impact of the ethnic fractionalization index in Model 2 also supports this logic, even if the coefficient estimate just misses the standard threshold of 95 % level of confidence. An overwhelming majority of the ethnic conflicts in our sample are violent separatist attempts, which are generally not perceived as a direct threat to the regime as a whole and thus require less drastic countermeasures. Besides, separatist rebellion often involves guerrilla warfare, in which the rebels take advantage of rough terrain and safe havens across the border, while avoiding large-scale encounters with militarily superior governmental forces. The increased size of the estimate for mountains in Model 2 may reflect a similar byproduct of a relatively higher share of guerrilla wars in the ACD dataset.

As a preliminary test of Hypothesis 1, we conclude the country-level analysis with an evaluation of Cederman and Girardin's N^* index. According to this conception, a state features an ethnic configuration comprising n groups with sizes $\{s_0, s_1, s_2 \ldots s_{n-1}\}$ where s_0 denotes the size of the EGIP. Assuming that only dyadic conflict between the

EGIP and the respective peripheral groups can happen, it is possible to compute the

probability of civil war as

$$N^* = \Pr(CivilConflict) = 1 - \prod_{i=1}^{n-1}(1 - p(i))$$

where $p(i)$ is the probability of dyadic conflict erupting between the EGIP and the

marginalized group i.

Furthermore, it is assumed that conflict happens if the power balance tips in favor of the

peripheral group in question. Using relative demographic group sizes as a proxy for

power, their study postulates that the probability of conflict in the dyad involving the

EGIP and group i can be written as

$$p(i) = \frac{1}{1 + \{r(i)/r\}^{-k}}$$

where $r(i) = s_i /(s_i + s_0)$ is group i's share of the total dyadic population, r is a threshold

value and k a slope parameter.

In line with their study, we computed the index with the default settings $r = 0.5$ and $k = 5$,

though our N^* measure is based on the ethnic groups' population shares provided by the

ANM rather than on Fearon's original group data.[31] Moreover, in our case, the *N** measure is extremely right-skewed. It could thus be argued that one unit change matters less for higher values of *N**, which would call for a log-transformation. In fact, the logged *N** in Model 3 is both positive and highly significant, thus confirming Cederman and Girardin's original findings. And unlike their findings, our results do not hinge on a limited number of high-risk cases in the Middle East.[32]

[TABLE 1 ABOUT HERE]

Estimating models of ethnic exclusion at the dyadic level

Do these findings imply that Hypothesis 1 should be accepted and that the exclusion perspective has thereby been definitively vindicated? We are inclined to think that such a conclusion would be premature, because all analysis thus far has been conducted at the country level whereas the causal mechanisms are located at the sub-state level. A dyadic analysis of ethic conflict is much better suited to uncover the true impact of ethno-nationalist exclusion.

The unit of observation in the dyadic analysis is the ethnic dyad, connecting an EGIP at the center with a marginalized ethnic group. All dyads are observed at annual intervals from 1946 or the year of independence for younger states. To control for non-independence among dyads belonging to the same country, we specify clustered standard

---

[31] Fearon 2003.

[32] Cf. Fearon, Kasara, and Laitin 2007.

errors in all models. Moreover, we include a peace-years count and three natural cubic splines to handle duration dependence.[33]

We start by considering a center-periphery model that, in addition to the logged dyadic power balance, r, includes as many independent variables as possible from Fearon and Laitin's reference model (see Model 4 in Table 2). However, we decided to drop the country-level fractionalization indices since ethnicity is measured by our proxy for relative power. We also dropped the insignificant "non-contiguous" dummy, as it carries little meaning at the dyadic level, as well as the "new state" indicator, whose influence is largely captured by the peace-years variable. Finally, due to the uneven temporal distribution of ethnic conflicts, with a relatively higher share occurring after the Cold War, we introduce a trend variable based on calendar years.

[TABLE 2 ABOUT HERE]

The results are striking. As we anticipate from the exclusion perspective (Hypothesis 1), the dyadic power balance has a positive and significant effect; the likelihood of ethnic conflict increases as the peripheral group's relative size goes up. The marginal impact is quite substantial as well. An increase in the MEG's power from the median ($r = 0.003$, or 0.3 % of the dyadic population) to the 95th percentile value ($r = 0.1$) corresponds to a non-trivial fivefold increase in the probability of conflict, all else held at median values. By contrast, the GDP variable performs very poorly, again failing to reach significance.

---

[33] See Beck, Katz, and Tucker 1998.

Of the other country-level measures, the population variable is the only one that exhibits any effect. Even after controlling for degree of marginalization, politically excluded groups in large countries have a higher baseline probability of hosting a rebellion. Beyond that, the increasing time trend in ethnic conflict onsets is strongly confirmed, as is the healing effect of time since the last dyadic conflict.

Let us now add the geographic variables: the peripheral group's distance to the capital (Hypothesis 2a) and extent of mountainous terrain in the excluded group's home region (Hypothesis 2b). Model 5 provides strong evidence in support of these hypotheses. It is clear that, *ceteris paribus*, remote groups and groups in rough terrain are significantly more likely to be involved in conflict than those closer to the capital and those located in the plains. The latter finding is particularly interesting when compared to the insignificance of the corresponding country-level measure in the previous model and nicely illustrates the power of disaggregation. Note also that the impact of country population is substantially reduced as much of its effect is captured by the correlated distance variable. Adding the ethno-geographic variables to Model 4 boosts the influence of the already highly significant dyadic power balance. Moreover, the democracy term is now strongly significant, implying a higher baseline risk of ethnic rebellion in democratic societies. While this finding intuitively might seem suspicious, a similar result is reported by Buhaug for separatist conflict.[34]

---

[34] Buhaug 2006.

So far, we have used most of Fearon and Laitin's relatively long list of variables as the

benchmark of our dyadic exploration. However, Achen warns against uncritically adding

explanatory variables and estimating large models, as this is likely to distort inference.[35]

In addition, several of the country-level controls are neither appropriate nor theoretically

interesting in this context. In Model 6 and subsequent models, we thus limit the selection

of country-level controls to per capita income. We further decided to drop Fearon and

Laitin's measure in favor of Gleditsch's GDP per capita data,[36] because the latter dataset

has fewer missing observations and covers two additional years (2000–01) in the most

relevant period (though it lacks data for years before 1950). Moreover, we take the

natural logarithm of the measure to compensate for a declining effect of one unit change

with higher income values.

Apart from making the overall model better specified, Model 6 shows that dropping

irrelevant controls further strengthens the empirical case for the geographically amended

ethnic exclusion perspective. Per capita income, in contrast, once again fails to reach

significance.

Figure 2 provides a graphic illustration of the association between r and the risk of dyadic

conflict. The lower, solid line shows the marginal effect of r, holding all other covariates

at their median values. Obviously, peripheral groups that face vastly superior EGIPs are

not likely to rebel, but the risk of conflict increases markedly with the relative size of the

---

[35] Achen 2002.

[36] Gleditsch 2002.

marginalized group. In the middle, dashed plot, we changed the distance from the capital

to the peripheral group from the 50th to the 95th percentile value. This is associated with a

considerably higher overall risk of conflict, as is also indicated by the positive coefficient

in Model 6. The upper, dotted line shows the likelihood of conflict across values of *r* for

both mountainous (95th percentile) and distant (95th percentile) groups. In line with our

hypotheses, relatively sizable marginalized groups residing in rough terrain and at a

considerable distance from the capital (such as Kurds in Iran and Bengalis in Pakistan

pre-1971) are those most likely to challenge the center by violent means.

[FIGURE 2 ABOUT HERE]

Checking the robustness of the dyadic analysis

The results provide strong and compelling evidence in favor of our ethnic exclusion

argument and simultaneously demonstrate the benefits of a disaggregated, dyad-level

design. Yet, certain aspects of the analysis deserve more consideration. We therefore

present three alternatively specified models that might help determine the robustness of

the findings presented above.

First, there is a potential problem with non-independence among time-series observations

of the same dyad as well as among dyads within the same country. The data are also

likely to exhibit some degree of unit heterogeneity not fully accounted for by the current

selection of regressors. While clustered standard errors address the former problem, unit

heterogeneity remains largely unmeasured. A conventional method to model cross-

26

sectional idiosyncrasies is to use a fixed-effects estimator. This procedure essentially

adds one dummy variable per unit that allows for a unique intercept, and thus an

individual baseline propensity for conflict for each country case. Model 7 in Table 3

presents the results from a fixed-effects logit analysis of the variables in the previous

model.

All dyad-specific explanatory variables confirm our preliminary conclusions. Political

exclusion of large ethnic cohorts poses a serious risk, and particularly so if the groups

reside in inaccessible terrain and at a considerable distance from the capital. Intriguingly,

the estimate for the income variable is now significantly positive, while both the trend

variable and the peace-year count replicate previous results.

The fixed-effects estimator comes with certain unfortunate side effects. One such effect is

that only units with variation on the dependent variable influences the regression

estimates. This explains the dramatic (⅓) decrease in the number of observations in

Model 7 compared to the previous model; all dyads in all countries without conflict in the

sample period are by design dropped from the analysis. Yet, some of the excluded cases

may have avoided conflict *because* of the very attributes on the explanatory variables.

Deeming these cases irrelevant makes little sense.[37]

Second, our ethno-geographic measures are in essence snapshots at one point in time, and

the power balance, too, is static in most cases. A lack of temporal variation on key

---

[37] Beck and Katz 2001.

explanatory variables means that the model is not well-suited to predict the timing of conflict onset. Thus we next conduct a strictly cross-sectional analysis, with only one observation per dyad. In this case, the binary dependent variable distinguishes between dyads that exhibited one or more conflicts during the sample period and those that remained at peace. The time-varying covariates, dyadic power balance and GDP, are fixed at the earliest available figures for each dyad.

Model 8 shows the results from the static analysis. By and large, our main results are once again confirmed. Above all, the dyadic power ratio retains its powerful influence on the likelihood of ethnic conflict. A center-periphery dyad with median values on all covariates has an estimated risk of experiencing at least one conflict between 1946 and 2003 of $p = 0.05$. An otherwise similar dyad for which the marginalized group is relatively large (r set at the 95th percentile value, corresponding to a still hefty 90–10 power preponderance in favor of the EGIP) is six times more likely to be engaged in conflict, with $p = 0.30$. The two ethno-geographic variables also hold up quite well, even though the parameter estimates have shrunk and the rough terrain proxy is only significant with a 90 % confidence threshold. The effect for per capita income is still not even remotely significant.

A final sensitivity test remains. Up to this point, our dyadic analysis has hinged entirely on the ethnic data of the *Atlas Narodov Mira* and the conflict data of the UCDP/PRIO Armed Conflicts Dataset. Since these datasets deviate from alternative sources in important ways, one might speculate whether our results are artifacts of specific sample

28

selection criteria and variable operationalizations. In particular, the *ANM* identifies a very large number of ethnic groups, many of which are clearly not politically relevant. It does not, however, capture demonstrably critical sectarian cleavages, e.g. in Arab countries. The GIS-based method of generating group population estimates could also introduce measurement errors. Similarly, the ACD's inclusive definition of intrastate conflict (at least 25 annual battle-deaths) implies that the sample of conflicts are dominated by low-level insurgencies not considered "civil wars" by competing providers of conflict data. Hence, in Model 9 (Table 3) we replicate the analysis on Model 6 with a dyadic version of Fearon and Laitin's ethnic civil war data, using Fearon's group list and its associated group size values in order to construct the dyadic power balance indicator. Unlike the *ANM,* Fearon's ethnicity data contain no information on the location of the ethnic groups, so we are not able to test the robustness of Hypotheses 2a and 2b here.

The results again provide strong and compelling support for the ethno-nationalist exclusion approach. Even with an entirely different group list and conflict data with a considerably higher minimum casualty threshold, Hypothesis 1 holds. We also note that by dropping the geographic factors and using a sample of major civil wars only, the GDP per capita measure regains its negative and significant effect. The temporal control variables in Model 9 exert little impact on the overall fit of the model.

We also performed a series of other sensitivity analyses (not reported here) that confirm the robustness of our main findings. The results for the dyadic power balance and ethno-geographic measures remain if we control for the number of groups in each country,

multiple simultaneous conflicts, various types of spatial and temporal patterns, or apply a different temporal threshold for coding recurring onsets. Applying a minimum *r*-score threshold to exclude "politically irrelevant" MEGs, successively removing the cases with the highest *r*-scores, or dropping the countries with the highest frequencies of dyadic conflict, made no significant difference either. Our results are also robust to exclusion of the Middle East region, which could be considered problematic due to the *ANM*'s lack of distinction between Shia and Sunni Islam.

[TABLE 3 ABOUT HERE]

Conclusion

By disaggregating civil wars to the level of ethnic center-periphery dyads, this study underlines the importance of politicized ethnicity as a major factor driving internal conflict. It thus casts further doubt on recent studies that reject ethnic grievances as determinants of civil wars. From a policy perspective, this is an important insight because of the prominence that such studies have gained through the sponsorship of the World Bank and other international organizations.

More specifically, we are able to say something about why ethnic configurations have an impact on conflict behavior. Expecting powerful disgruntled ethnic minorities to be a major source of conflict, the dyadic perspective adopted here vindicates, and extends,

Wimmer's theory of ethno-nationalist exclusion.[38] Center-periphery dyads characterized by such demographic asymmetries stand a much higher risk of civil war.

Our results also suggest that such an explanation interacts with ethno-geographic conditions. Other things being equal, ethnic groups far from their state capitals tend to be more prone to involvement in ethnic civil wars. The same can be said for those groups located in rough terrain. Geographic factors of this type can be seen as important refinements of highly aggregated measures of terrain[39] and cruder power estimations based merely on demographic group sizes.[40]

Obviously, the current study is far from the final word on the influence of ethnicity on conflict. It should once again be emphasized that because of data limitations, we have confined ourselves to the study of Eurasia and North Africa. Future work will reveal if our conclusions hold up for the remaining world.[41] There is also plenty of room for improvement as regards the geographic calculations. With better data, one might be able to account for the quality of local communication infrastructure. The power estimation could also profit by relying on additional empirical dimensions, including data on mobilization and organizational structure of the groups in question.

---

[38] Wimmer 2002.

[39] E.g. Fearon and Laitin 2003.

[40] Cederman and Girardin 2007.

[41] We do so in the ongoing Project "Expert Survey on Ethnic Groups (ESEG)", see Cederman, Girardin and Wimmer 2006.

For the time being, however, we advise against reducing civil wars to primarily greedy behavior, sheer criminality, or terrorist activities. The unresolved Israeli-Palestinian conflict, and mounting ethno-nationalist tensions within occupied Iraq, prove that ethnic nationalism is far from a spent force. If our findings are correct, this is a widespread problem that cries out for negotiated solutions. Military interventions that turn a blind eye to ethnic exclusion or try to manipulate the ethnic balance of power can be expected to foster future tension threatening to spill over international borders, as illustrated by those launched by the United States in Iraq in 2003 or by Israel in Lebanon in 2006. Although we have focused on civil wars here, it would be mistake to underestimate the border-transgressing and border-transforming influence of nationalism in today's world. Future research will show how trans-national ethnic links and international interventions interact with the exclusion mechanisms explored in the current study.

References

Achen, Christopher H. 2002. Toward a New Political Methodology: Microfoundations and ART. *Annual Review of Political Science* 5:423–50.

Beck, Nathaniel, and Jonathan N. Katz. 2001. Throwing Out the Baby with the Bath Water: A Comment on Green, Kim, and Yoon. *International Organization* 55 (2):487–95.

Beck, Nathaniel, Jonathan N. Katz, and Richard Tucker. 1998. Taking Time Seriously: Time-Series–Cross-Section Analysis with a Binary Dependent Variable. *American Journal of Political Science* 42 (4):1260–88.

Boulding, Kenneth. 1962. *Conflict and Defense*. New York: Harper & Row.

Bruk, Salomon I., and V. S. Apenchenko, eds. 1964. *Atlas Narodov Mira*. Moscow: Academy of Science USSR.

Buhaug, Halvard. 2006. Relative Capability and Rebel Objective in Civil War. *Journal of Peace Research* 43 (6):691–708.

Buhaug, Halvard. 2007. Dude, Where's My Conflict? LSG, Relative Strength, and the Location of Civil War. Unpublished manuscript, PRIO.

Buhaug, Halvard, and Jan Ketil Rød. 2006. Local Determinants of African Civil Wars, 1970–2001. *Political Geography* 25 (3):315–35.

Cederman, Lars-Erik. Forthcoming. Articulating the Geo-Cultural Logic of Nationalist Insurgency. In *Order, Conflict, and Violence*, edited by Stathis Kalyvas and Ian Shapiro. Cambridge: Cambridge University Press.

Cederman, Lars-Erik, and Luc Girardin. 2007. Beyond Fractionalization: Mapping Ethnicity onto Nationalist Insurgencies. *American Political Science Review* 101:173–85.

Cederman, Lars-Erik , Luc Girardin, Andreas Wimmer. 2006. Getting Ethnicity Right: An Expert Survey on Power Distributions. Paper presented at the Annual Convention of the American Political Science Association, August 30-September 3.

Cederman, Lars.Erik, Jan Ketil Rød, and Nils Weidmann. 2006. Geo-Referending of Ethnic Groups: Creating a New dataset. Unpublished manuscript, ETH Zurich.

CIESIN. 2005. Gridded Population of the World, Version 3 (GPWv3), available at http://sedac.ciesin.columbia.edu/gpw/.

Collier, Paul, and Anke Hoeffler. 2004. "Greed and Grievance in Civil Wars." *Oxford Economic Papers* 56 (4):563–95.

DeNardo, James. 1985. *Power in Numbers: The Political Strategy of Protest and Rebellion*. Princeton: Princeton University Press.

Fearon, James D. 2003. Ethnic and Cultural Diversity by Country. *Journal of Economic Growth* 8 (2):195–222.

Fearon, James D., and David D. Laitin. 2003. Ethnicity, Insurgency, and Civil War. *American Political Science Review* 97 (1):75–90.

Fearon, James D, Kimuli Kasara and David Laitin. 2007. "Ethnic Minority Rule and Civil War Onset." *American Political Science Review* 101 (1):187-93.

Gleditsch, Kristian S. 2002. Expanded Trade and GDP Data. *Journal of Conflict Resolution* 46 (5):712–24.

Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and

    Håvard Strand. 2002. Armed Conflict 1946–2001: A New Dataset. *Journal of*

    Peace Research 39 (5):615–37.

Goemans, Hein, Kristian Skrede Gleditsch, Giacomo Chiozza, and Jinhee L. Choung.

    2004. Archigos: A Database on Political Leaders. Unpublished manuscript,

    University of Rochester and University of California San Diego.

Gordon, Raymond G. JR, ed. 2005. *Ethnologue: Languages of the World* 15th edition.

    Dallas, TX: SIL International.

Gurr, Ted Robert. 2000. *Peoples Versus States: Minorities at Risk in the New Century*.

    Washington, DC: United States Institute of Peace Press.

Hegre, Håvard and Nicholas Sambanis. 2006. Sensitivity Analysis of Empirical Results

    on Civil War Onset. *Journal of Conflict Resolution* 50 (4):508–35.

Horowitz, Donald L. 1985. *Ethnic Groups in Conflict*. Berkeley: University of California

    Press.

McCarthy, John D., and Mayer Zald, N. 1977. Resource Mobilization and Social

    Movements: A Partial Theory. *American Journal of Sociology* 82 (6):1212–41.

Most, Benjamin A., and Harvey Starr. 1989. *Inquiry, Logic, and International Politic*s.

    Columbia: University of South Carolina Press.

Rokkan, Stein. 1999. *State Formation, Nation-Building, and Mass Politics in Europe:*

    The Theory of Stein Rokkan. Flora, Peter, ed. Oxford: Oxford University Press.

Ross. Michael. 2006. A Closer Look at Oil, Diamonds, and Civil War. *Annual Review of*

    Political Science 9:265–300.

Sambanis, Nicholas. 2004. Using Case Studies to Expand Economic Models of Civil

War. *Perspectives on Politics* 2 (2):259–79.

Tarrow, Sidney. 2007. Inside Insurgencies: Political Violence in an Age of Civil War.

Perspectives on Politics 5 (3):587–600.

Taylor, Charles L., and Michael C. Hudson. 1972. *World Handbook of Political and

Social Indicators*. New Haven, CT: Yale University Press.

UNEP-WCMC. 2002. *Mountain Watch*. UNEP World Conservation Monitoring Centre.

Wimmer, Andreas. 2002. *Nationalist Exclusion and Ethnic Conflict: Shadows of

Modernity*. Cambridge: Cambridge University Press.

Table 1. Logit Analysis of Onset of Ethnic Civil War, Country Level

	FL ethnic war	UCDP/PRIO ethnic conflict	
	(1)	(2)	(3)
Prior war	−0.917	0.435	0.387
	(1.45)	(0.99)	(0.87)
GDP capita [b]	−0.220	0.016	0.044
	(2.55)*	(0.23)	(0.64)
Population [a, b]	0.445	0.589	0.671
	(2.01)*	(5.44)***	(5.46)***
Mountains [a]	0.210	0.300	0.185
	(0.87)	(1.79)	(0.98)
Non-contiguous	0.271	−0.105	−0.142
	(0.50)	(0.27)	(0.39)
Oil	1.132	1.010	0.993
	(2.37)*	(2.10)*	(2.06)*
New state	1.963	2.523	2.509
	(3.25)***	(4.54)***	(4.43)***
Instability	0.086	−0.042	−0.117
	(0.17)	(0.09)	(0.25)
Polity score [b]	0.052	0.031	0.039
	(1.58)	(0.94)	(1.22)
Religious fractionalization	0.547	−0.378	
	(0.40)	(0.36)	
Ethnic fractionalization	0.084	1.406	
	(0.07)	(1.92)	
Ethnic exclusion index N^{*a}			0.094
			(3.03)**
Constant	−9.128	−11.936	−11.102
	(3.93)***	(9.96)***	(8.65)***
N	2,211	2,554	2,554

Note: Model 1 includes all countries in Eurasia and North Africa with at least a 5% ethnic minority, 1945–99; Models 2–3 include all countries in Eurasia and North Africa with at least one politically marginalized ethnic group, 1946–99. Robust absolute z scores are given in parenthesis.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

[a] Log-transformed

[b] Lagged one year

Table 2. Logit Analysis of Onset of Ethnic Civil War, Dyadic Level

| | UCDP/PRIO dyadic ethnic conflict | | |
	(4)	(5)	(6)
Group-level variables			
Demographic balance *r* [a]	0.491	0.595	0.540
	(4.12)***	(5.27)***	(5.05)***
Distance from capital [a]		0.589	0.817
		(2.03)*	(5.47)***
Mountains		1.066	1.328
		(3.75)***	(3.91)***
Country-level variables			
GDP capita [b]	−0.091	−0.096	−0.300 [a, b]
	(0.96)	(0.94)	(1.17)
Population [a, b]	0.580	0.369	
	(6.46)***	(2.37)*	
Mountains	−0.007		
	(0.04)		
Oil	−0.424	−0.584	
	(1.15)	(1.38)	
Instability	−0.087	−0.127	
	(0.10)	(0.14)	
Polity score [b]	0.046	0.065	
	(1.39)	(2.60)**	
Year	0.061	0.064	0.066
	(5.32)***	(5.49)***	(6.06)***
Peace years	−0.264	−0.254	−0.279
	(4.42)***	(4.25)***	(4.30)***
Constant	−126.799	−134.436	−135.095
	(5.64)***	(5.75)***	(6.52)***
N	32,720	32,720	33,882

Note: *r* represents the size of the marginalized ethnic group (MEG), relative to the joint population of the MEG and the ethnic group(s) in power (EGIP). Models 4–5 cover the 1946–99 period, Model 6 covers the years from 1950 to 2001. Robust absolute z scores, clustered on countries, are given in parenthesis. Estimates for three natural cubic splines not shown.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

[a] Log-transformed

[b] Lagged one year

Table 3. Alternative Specifications, Dyadic Level

	Country fixed effects (7)	Static analysis (8)	FL dyadic ethnic war (9)
Group-level variables			
Demographic balance r [a]	0.603	0.614	0.884
	(7.01)***	(5.09)***	(3.94)***
Distance from capital [a]	1.410	0.610	
	(4.84)***	(3.27)***	
Mountains	1.324	0.790	
	(2.90)**	(1.68)	
Country-level variables			
GDP capita [a, b]	1.266	−0.266	−0.436
	(2.14)*	(1.16)	(1.96)*
Year	0.046		0.014
	(2.84)**		(0.80)
Peace years	−0.101		−0.329
	(1.68)		(1.36)
Constant		−1.868	−25.823
		(0.88)	(0.79)
N	21,218	867	7,627

Note: Model 7 uses a fixed-effects logit estimator of dyadic ethnic conflict, 1950–2001. Model 8 is a cross-sectional model of dyadic conflict, 1950–2001, with earliest available data on power ratio and GDP. Model 9 uses a dyadic version of Fearon and Laitin's civil war data, 1950–99,[42] with dyadic power balance, r, calculated from group measures in Fearon.[43] Absolute z scores (robust in Model 8, clustered on countries in Model 9) are given in parenthesis. Estimates for three natural cubic splines (Models 7 and 9) not shown.
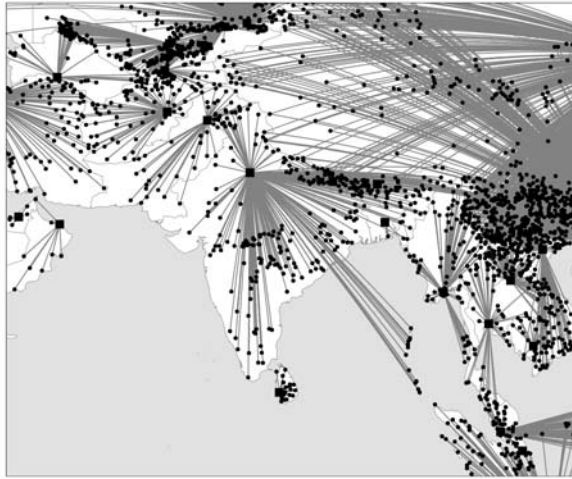
* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.
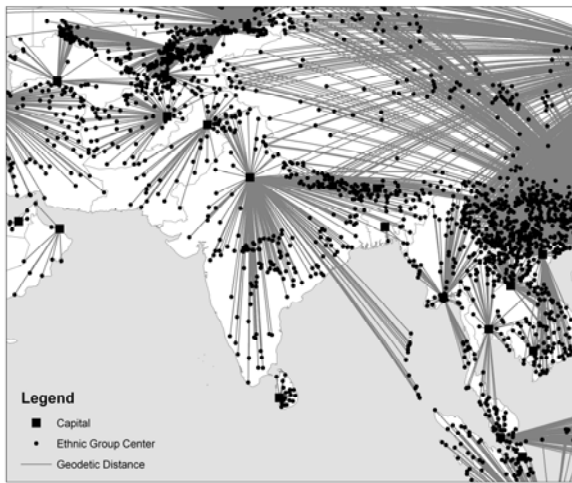
[a] Log-transformed

[b] Lagged one year

---

[42] Fearon and Laitin 2003.
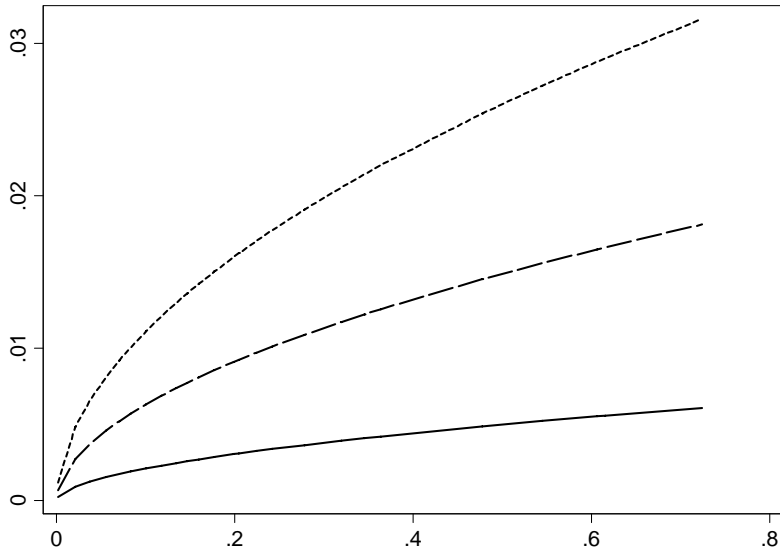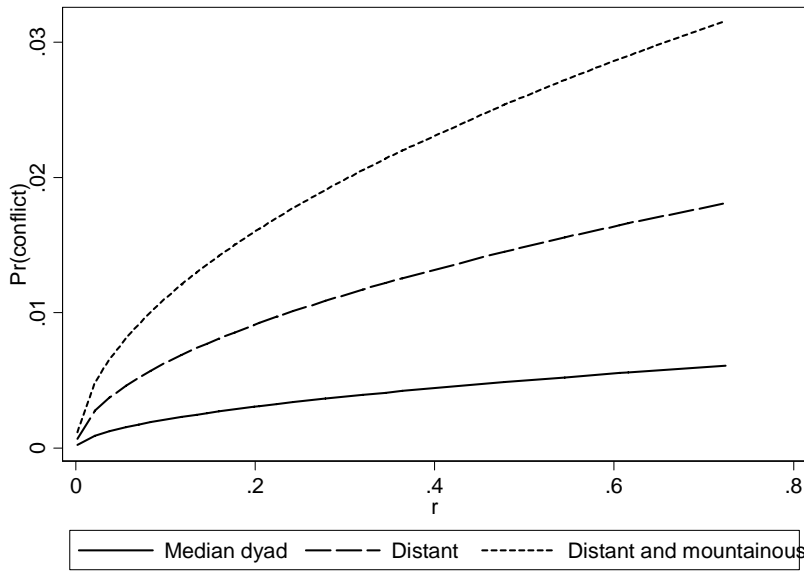
[43] Fearon 2003.

[without text]

[with text]

Figure 1. Geodesic Distances between Capitals and Ethnic Group Center Points

[without text]

[with text]

Figure 2. Estimated Risk of Conflict as a Function of the Periphery's Relative

Capability (*r*)