Online Appendix to "Horizontal Inequalities and Ethno-Nationalist Civil War: A Global Comparison," Forthcoming in the *American Political Science Review*.

Lars-Erik Cederman
Center for Comparative and International Studies
& Centre for the Study of Civil War, PRIO
ETH Zürich, IFW D 49.2
8092 Zürich, Switzerland
Phone: +41 44 632 67 59
Fax: +41 44 632 12 89
lcederman@ethz.ch

Nils B. Weidmann
Jackson Institute for Global Affairs
Yale University
P.O. Box 208206
New Haven CT 06520-8301, USA
& Centre for the Study of Civil War, PRIO
Phone: +1 203 436 5286
nils.weidmann@yale.edu

Kristian Skrede Gleditsch
Department of Government
University of Essex
Wivenhoe Park
Colchester, Essex CO4 3SQ, UK
& Centre for the Study of Civil War, PRIO
Tel: +44 1206 87 25 17
Fax: +44 1206 87 32 34
ksg@essex.ac.uk

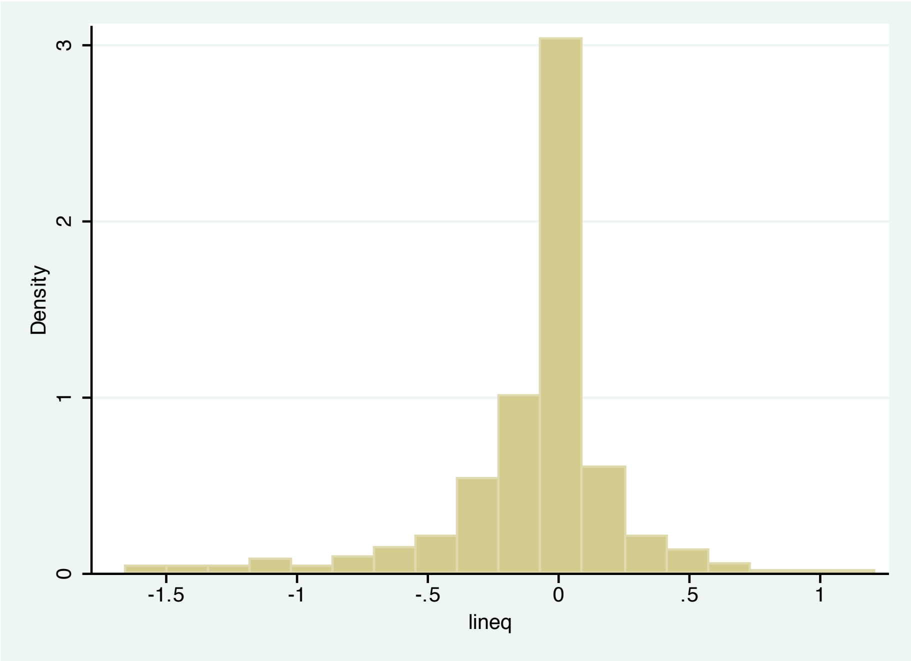
April 11, 2011

This online appendix provides supplementary description and analysis of the data presented in the article. We start by presenting descriptive statistics, followed by a section that provides supplementary sensitivity analysis.

1. Descriptive analysis

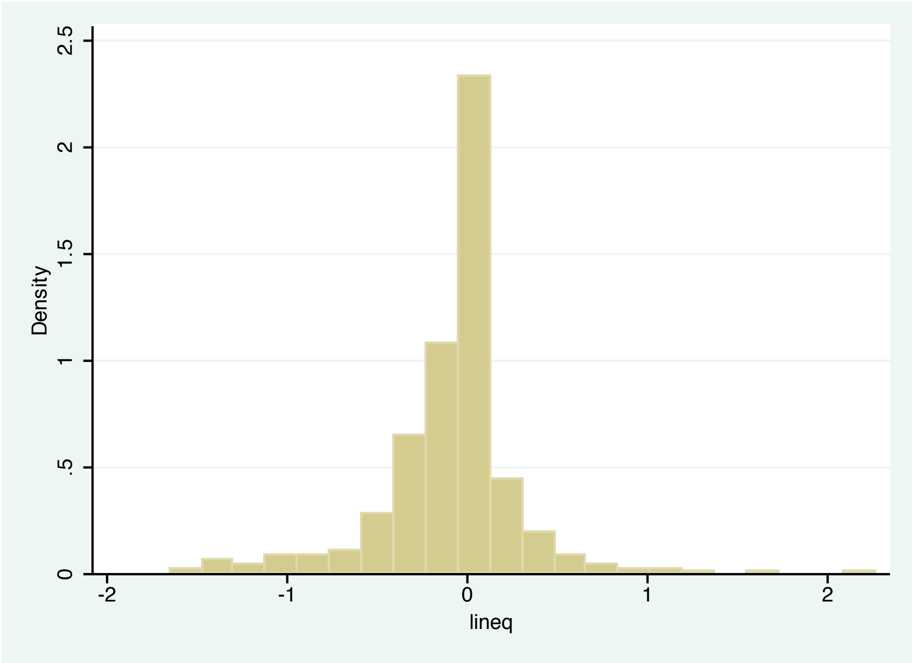
1.1. Economic horizontal inequalities

The main text describes our inequality estimates in some detail. In addition, we include a histogram of the logged ratio *lineq* for 1990, the year of the Nordhaus measurement, including groups with population below 500,000 (as specified in Models 2-5).



As expected, the distribution is centered around equality (lineq = 0) with thin tails stretching up to about 1.21 and -1.80. The former value corresponds to a *high_ratio* of 3.3 and the latter to a *low_ratio* of 6.0, i.e. inequalities amounting to a factor 3 and a factor 6 of the average income of the country in question.

A comparison with the full sample, including all group sizes for the year 1990, yields the following:



It can be concluded that the two distributions are quite similar (see Section 1.6 for detailed information on the differences).

1.2. Political horizontal inequalities

In supplement to the description of the EPR data in the main text, we display the absolute and relative frequencies broken up according to each power-status category, based on the restricted sample. Please note that the categories MONOPOLY and DOMINANT are excluded from the analysis since, by definition, those categories cannot engage in conflict with the state. The two first categories, SENIOR PARTNER and JUNIOR PARTNER correspond to included groups, and the remaining categories to excluded groups. The statistics show that the relative frequency of onsets is the highest for DISCRIMINATED and SEPARATIST AUTONOMY, and the lowest for the included groups.

```
. tab status onset_ra if sample==1, row

+----------------+
| Key            |
|----------------|
|   frequency    |
| row percentage |
+----------------+

                    |       onset.ra
             status |         0          1 |     Total
--------------------+----------------------+----------
     SENIOR PARTNER |       800          5 |       805
                    |     99.38       0.62 |    100.00
--------------------+----------------------+----------
     JUNIOR PARTNER |     1,062          7 |     1,069
                    |     99.35       0.65 |    100.00
--------------------+----------------------+----------
  REGIONAL AUTONOMY |       817          5 |       822
                    |     99.39       0.61 |    100.00
--------------------+----------------------+----------
          POWERLESS |       807          8 |       815
                    |     99.02       0.98 |    100.00
--------------------+----------------------+----------
      DISCRIMINATED |       343         12 |       355
                    |     96.62       3.38 |    100.00
--------------------+----------------------+----------
SEPARATIST AUTONOMY |        96          5 |       101
                    |     95.05       4.95 |    100.00
--------------------+----------------------+----------
              Total |     3,925         42 |     3,967
                    |     98.94       1.06 |    100.00
```

Further information about the EPR dataset can be found at http://dvn.iq.harvard.edu/dvn/dv/epr.

The correspondence between the EPR groups and the groups in the Minorities at Risk dataset is quite substantial. The Phase IV Minorities at Risk data lists 340 groups, of which 284 remain “at risk” and are actively tracked. EPR includes coverage of 310 of these groups. Of the 30 groups that EPR does not code, four are immigrant or non-citizen groups excluded from EPR (foreign workers in Switzerland, Turks in Germany, Muslims in France, and Vietnamese in Cambodia), and five are groups we do not consider ethnic (like the regionally-identified Honamese in Korea or clans like the Issaq in Somalia). The remaining 21 groups do not appear in EPR because they are not politically salient according to our scheme, including Native Hawaiians in the US, Chinese in Panama, and the Mossi-Dagomba in Ghana.

1.3. Group settlement areas

As specified in the paper, we draw the information on settlement areas from the GeoEPR dataset (see Wucherpfennig et al. 2010; and http://dvn.iq.harvard.edu/dvn/dv/epr). Explicitly introducing time periods ("geo-periods") reflecting major changes in settlement areas, The GeoEPR dataset provides information about whether each group in the non-spatial EPR dataset has a regional base of settlement. There are six possibilities according to which the group is either (1) regionally based, (2) urban, (3) regional and urban, (4) migrant, (5) dispersed or (6) aggregate. These categories are defined in the following way:

(1) regionally based: group members reside in a particular region/in particular regions that are easily distinguishable on a map. Note that it is explicitly possible to have overlays between different groups, i.e. co-habitation in the same region.
(2) urban: a group located primarily in urban areas and not in a particular region/in particular regions. A group is coded as urban when 60%+ of the group is located in cities.
(3) regional and urban: group members live both in a city/in cities and in a particular region/in particular regions.
(4) migrant: groups members change location often, as with nomadic or Roma groups.
(5) dispersed: group members do not inhabit any particular city/cities or region/regions and are not migrant.
(6) aggregate: a particular group which during a geo-period is aggregated from several smaller ones (provided that their ethnic geography remains constant).

The current dataset provides shapefiles for categories 1, 3, 5 and 6. As specified in the paper, we dropped the dispersed groups (geotype=5) for the purpose of the analysis (or about 9.38% of the groups). The following tabulation summarizes the composition of the data:

```
.  tab geotype pop500k if year==1992, col

+-------------------+
| Key               |
|-------------------|
|     frequency     |
| column percentage |
+-------------------+

           |        pop500k
   geotype |         0          1 |     Total
-----------+----------------------+----------
         1 |       127        183 |       310 
           |     75.60      47.66 |     56.16 
-----------+----------------------+----------
         3 |        37        157 |       194 
           |     22.02      40.89 |     35.14 
-----------+----------------------+----------
         5 |         0         36 |        36 
           |      0.00       9.38 |      6.52 
-----------+----------------------+----------
         6 |         4          8 |        12 
           |      2.38       2.08 |      2.17 
-----------+----------------------+----------
     Total |       168        384 |       552 
           |    100.00     100.00 |    100.00 
```

1.4. Comparison of our estimates of economic HIs with data from MAR

In order to compare our own spatial estimates drawing on Nordhaus' data, we matched as many MAR groups as possible with corresponding EPR groups. Basing the comparison on the year of 1992, which is the first time point in the datasets after the collapse of the Soviet Union, we were able to identify 259 of the 312 MAR groups, or 83%, wholly or at least partly with corresponding EPR groups (excluding those groups that GeoEPR flags as dispersed).

For the inequality evaluation, we used the most relevant variable in the MAR dataset, namely the indicator ECDIFX1, which measures income differentials on an ordinal scale as -1 or -2 for groups that are clearly advantaged, 0 for no difference, and 1 or 2 for groups that are disadvantaged (see MAR codebook 2001-2003).

A mean comparison of our *high_ratio* variable for the advantaged MAR groups to the remaining MAR groups shows that we can reject the hypotheses that the mean for the former (0.856) is the same as that of the latter (0.523) at the p=0.054 level. Likewise, the mean (0.954) of the *low_ratio* variable for disadvantaged groups is significantly (p=0.0235) higher than that of the other groups (0.702).

```
.  ttest high, by(rich2)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |     176    .5233628    .0708294    .9396576    .3835731    .6631525
       1 |      24    .8562237    .2053776     1.00614    .4313679     1.28108
---------+--------------------------------------------------------------------
combined |     200    .5633061    .0672767    .9514356    .4306395    .6959728
---------+--------------------------------------------------------------------
    diff |           -.3328609    .2061993               -.7394895    .0737677
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =  -1.6143
Ho: diff = 0                                     degrees of freedom =      198

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0540         Pr(|T| > |t|) = 0.1081          Pr(T > t) = 0.9460

.  ttest low, by(poor2)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |     102    .7020035    .0810793     .818861    .5411639     .862843
       1 |      98    .9544089     .097286    .9630825    .7613231    1.147495
---------+--------------------------------------------------------------------
combined |     200    .8256821     .063578    .8991292     .700309    .9510553
---------+--------------------------------------------------------------------
    diff |           -.2524054    .1262342               -.5013414   -.0034694
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =  -1.9995
Ho: diff = 0                                     degrees of freedom =      198

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0235         Pr(|T| > |t|) = 0.0469          Pr(T > t) = 0.9765
```

Still based on observations from 1992, we continue our comparison with lists of those cases where our estimates deviate from MAR's inequality coding. The first list shows the groups that MAR classifies as "advantaged" in terms of income, but where our coding indicates a below-average income. The second list enumerates the groups that were disadvantaged according to MAR, but that received an above-average ranking in our analysis. In both cases, we focus on groups that have a population of at least 500,000 and that deviated at least by 10% from the country average in our spatial estimation. Here the *high* and *low* variables refer to our ratio measures of HI, *high_ratio* and *low_ratio* respectively.

```
. list country group ecdifx1 low if ecdifx1<0 & high==0 & pop500k==1 & low>1.1
```

	country	group	ecdifx1	low
91.	RUSSIA	KUMYKS	-2	4.335216
135.	CAMEROON	BAMILEKE	-2	1.290527
153.	DEM. REP. CONGO	HUTUS	-2	1.154569
154.	DEM. REP. CONGO	TUTSIS	-2	1.148405

```
. list country group ecdifx1 high if ecdifx1>0 & low==0 & pop500k==1 & high>1.1
```

	country	group	ecdifx1	high
27.	ECUADOR	BLACKS	2	1.602747
96.	RUSSIA	YAKUT	2	2.400335
140.	NIGERIA	IJAW	2	3.344337
141.	NIGERIA	OGANI	2	2.02778
147.	ZAIRE	BAKONGO	2	1.125865
213.	IRAN	ARABS	2	1.991677
220.	IRAN	TURKMEN	2	1.285075
239.	AFGHANISTAN	HAZARAS	2	1.179323
276.	PAKISTAN	SINDHIS	2	1.175548
299.	MALAYSIA	EAST INDIANS	2	1.102906
307.	INDONESIA	PAPUANS	2	1.413777

We can make sense of many of the deviations by referring to the fact that the Nordhaus estimates include natural resources, including oil. Thus, some of the high values for the Nigerian groups, the Iranian Arabs, and possibly the also the Papuans in Indonesia and the Hazaras in Afghanistan, can be attributed to this factor. Other groups are not well captured by the spatial analysis due to difficulties of coding the settlement area in GeoEPR, such as the Ecuadoran Blacks, Bamileke in Cameroon, Hutus and Tutsis in the DRC, whose relative wealth is probably overestimated in our analysis. Some of the deviations may result from uncertainties in the MAR dataset. It seems difficult to understand why the Russian Kumyks were coded as advantaged according to ECDIFX1.

Despite these not too numerous deviations, we conclude that our inequality estimates seem to be strongly correlated with the MAR income differential values, thus lending validity to the new spatial data.

1.5. Stability of estimates of economic HIs over time

Since we extrapolate from the G-Econ estimates of GDP from 1990, it is important to show that economic HIs do not vary too much over time. In particular, there is the possibility that external shocks may drastically affect the economy in such a way that the relative wealth differences at the group level would be affected. Due to a general lack of dynamic data, it is difficult to examine this directly. It is a little bit easier to find time series that correspond to groups in federal countries whose territorial subunits correspond relatively closely to the groups in question. Based on official statistics, Lang (1975) offers such a comparison of Yugoslavia between the years of 1952 and 1971. During this period, the relative income per capita changes very little. As would be expected, the list was topped by the Slovenia and Croatia, both clearly above the country average. At the very bottom, we find Kosovo and Montenegro. See Figure A1. In fact, only one region, Vojvodina, moved from having been slightly below average in 1952 to being above in 1971, thus overtaking the Serbs. Otherwise, the relative inequality remained quite stable. In fact, the correlation coefficient between the two years is high as 0.9655, thus indicating that regional-level inequality is very stable over time.

Among developing countries, India provides a case of a federal country with relatively good disaggregated time series data. We draw on data from the Indian Census to study the stability of relative wealth among Indian regions. Since the 1956 States Reorganisation Act, India has been partitioned into states that often correspond to linguistic or ethnic communities (e.g. Assam, Jammu & Kashmir, Nagaland etc), although we note that there have been a number of border changes and new states created over the period since then, which obviously can influence state per capita figures. However, here too we find a very high persistence in terms of the relative state per capita income compared to the national average. The 1990 ratios are correlated with the 1970 and 1980 ratios at 0.973 and 0.977 and respectively, and the year to year ratios for individual states also have generally high year on year correlations, suggesting considerable stability in the relative ranks of the states.

Finally, we investigate whether MAR's income estimates are stable over time. Focusing on the period that corresponds to our analysis, our check looks at whether groups that are coded as disadvantaged in 1992 switched categories or became advantaged by 2004 and vice versa. This comparison shows that, according to MAR, only four groups out of 312 (1.28%), experienced this transformation:

```
. tab crosszero1 if year==1992

 crosszero1 |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        308       98.72       98.72
          1 |          4        1.28      100.00
------------+-----------------------------------
      Total |        312      100.00

. list country group ecdifx1 if crosszero==1 & year==1992

      +-----------------------------------+
      |      country      group   ecdifx1 |
      |-----------------------------------|
2531. |       RUSSIA     KUMYKS        -2 |
3443. |         TOGO        EWE         2 |
4932. | SOUTH AFRICA     ASIANS         2 |
4976. | SOUTH AFRICA   COLOREDS         2 |
      +-----------------------------------+

. list country group ecdifx1 if crosszero==1 & year==2003

      +-----------------------------------+
      |      country      group   ecdifx1 |
      |-----------------------------------|
2533. |       RUSSIA     KUMYKS         2 |
3447. |         TOGO        EWE        -2 |
4913. | SOUTH AFRICA     ASIANS        -2 |
4985. | SOUTH AFRICA   COLOREDS        -2 |
      +-----------------------------------+
```

It is unclear what caused the MAR team to move the Kumyks from having been advantaged to a situation of disadvantage. However, the changes involve the weakest, uncertain categories of -1 and 1. Moreover, the MAR's documentation indicates that no economic grievances have been recorded. See "Assessment for Kumyks in Russia," http://www.cidcm.umd.edu/mar/assessment.asp?groupId=36534

Similar conclusions can be drawn about the Ewe in Togo. Although the inequality coding in 2003 is as clear as -2, we have been unable to find any explicit justification for this coding. Quite on the contrary, the MAR documentation in this case states that: "Ewe are left to handle their own economic affairs. There also does not seem to be any cultural restrictions to speak of." See "Assessment of Ewe in Togo," http://www.cidcm.umd.edu/mar/assessment.asp?groupId=46101.

Presumably, the improved economic status of the Asians and Coloreds in South Africa is based on an assessment of economic affirmative action policies, although it is difficult to establish whether the coded shift is intended to reflect an actual improvement of the groups' wealth or changes in formal restrictions and opportunities. See e.g.
http://www.cidcm.umd.edu/mar/assessment.asp?groupId=56003

1.6. Groups below a population of 500,000

Table 1 in the paper offers descriptive statistics on the key variables, including our inequality indicators, for the restricted sample that excludes groups with a population below 500,000. Here we show the same statistics for the full sample, corresponding to Model 1:

```
sum lineq lineq2 low high excluded b b2 lgdpcapl exclgrps year pyrs_ra if sample1==1

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       lineq |      6438    -.092915    .3757218  -1.799328   2.265327
      lineq2 |      6438    .1497782    .4349172          0   5.131707
         low |      6438    .8664565     .825721          0   6.045586
        high |      6438    .4905804    .8017511          0   9.634275
    excluded |      6438    .6160298    .4863886          0          1
-------------+--------------------------------------------------------
           b |      6438    .1855156    .2394206   .0000206          1
          b2 |      6438    .0917293    .1931902   4.22e-10          1
    lgdpcapl |      6438    8.007631    1.061555   5.231002     10.494
    exclgrps |      6438    10.15502    15.08935          0         46
        year |      6438     1998.11    4.303287       1991       2005
-------------+--------------------------------------------------------
     pyrs_ra |      6438    34.35306    18.42051          0         59
```

The main difference in terms of inequality is that the maximum value of the *high_ratio* increases to 9.63 from 3.34 and the standard deviation of the inequality measure *lineq* is considerably higher (0.376 rather than 0.303), thus confirming that the sample restriction eliminates outliers, especially in the high range.

To get a firmer grip on the characteristics of the groups that were dropped from the restricted sample, we note that the inequality estimates differed quite substantially from those in the total population: the average estimate of the *low_ratio* within this subsample is 1.03 as opposed to 0.867 in the unrestricted sample. The average *high_ratio* differs less dramatically, but the standard deviation increases from 0.80 to 1.01 for the dropped groups.

```
. gen groupdrop = 1 if sample1==1 & sample!=1
(25643 missing values generated)

. sum high low if groupdrop==1 & year==1992

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         low |       161    1.026381    .8973729          0   4.554327
        high |       161    .4705167     1.01305          0   9.634275
```

For the sake of completeness, we list all excluded groups together with the spatial population estimate in 1992:

country	group	pldpop90
Albania	Greeks	160782.1
Albania	Macedonians	100759.2
Argentina	Quechua	251766.8
Armenia	Kurds	31093.97
Austria	Slovenes	96960.98
Azerbaijan	Lezgins	254256.8
Azerbaijan	Armenians	223525.7
Belgium	Germans	154577.7
Benin	Southeastern (Yoruba/Nago…	494306.9
Bolivia	Guaraní and other Eastern..	65314.95
Botswana	San (Bushmen, Basarwa)	332325.9
Cambodia	Khmer Loei (incl. Kui)	276912
Cambodia	Lao and Siamese	135541
Canada	Aboriginal peoples	262673.7
Central African Republic	Yakoma	89316.12
Central African Republic	Mbaka	177277.4
Chad	Toubou	324758.7
Chad	Zaghawa, Bideyat	89679.04
Chad	Hadjeraï	110722.9
Chile	Atacamenos	2038.683
China	Mulao	146136.3
China	Tu	484474
China	Achang	13934.42
China	Shui	356187.2
China	Daur	280788.3
China	Naxi	423868.6
China	Wa	338529.1
China	Lahu	217549
China	Dongxiang	189902.2
China	Pumi	40489.21
China	Maonan	40436.05
China	Blang	17019.29
China	Xibe	199723.2
China	Salar	71402.78
China	Yugur	82264.75
China	Tajiks	127594
China	Nu	13038.34
China	Bonan	145930.2
China	Jinuo	79618.24
China	Qiang	418606
China	Evenk	479509.7
Colombia	Afrocolumbians	290836.3
Congo	Kouyou	14575.63
Congo	Lari/Bakongo	313056.2
Congo	Mbochi (proper)	23792.54
Congo	Vili	372244.9
Congo	Nibolek (Bembe etc.)	29405.1
Costa Rica	Afrocosta Ricans	18976.84
Czechoslovakia	Hungarians	204554.9
Democratic Rep. of Congo	Ngbandi	384462.8
Democratic Rep. of Congo	Mbandja	333755.1
Finland	Swedes	476720.3
France	Corsicans	225670.6
France	Basques	110440.2
Gabon	Fang	468978.2
Gabon	Eshira/Bapounou	21756.39

Gabon	Myènè	70353.97
Gabon	Mbédé (Batéké, Obamba)	105876
Gambia	Fula	58492.45
Gambia	Diola	31446.45
Gambia	Wolof	81710.4
Georgia	Adzhars	265985.3
Georgia	Abkhazians	252642.5
Georgia	Azeri	235871.4
Georgia	Ossetians (South)	160735.9
Georgia	Armenians	176280
Greece	Muslims	213647.3
Guatemala	Garifunas	2511.196
Guatemala	Xinca	14710.66
Guinea-Bissau	Papel	97992.41
Guinea-Bissau	Manjaco	102134.1
Guinea-Bissau	Balanta	434140.7
Honduras	Tawahka	11946.25
Honduras	Chorti	18578.74
Honduras	Lenca	83890.75
Honduras	Garifs (Black Caribs)	161413.7
Honduras	Miskitos	27006.83
India	Mizo	87997.79
India	Indigenous Tripuri	349858.3
Indonesia	Chinese (Han)	468848.8
Indonesia	Amboinese	365258.3
Japan	Ainu	120499.2
Kuwait	Kuwaiti Sunni (Arab)	134003.3
Kyrgyzstan	Uzbeks	444163.9
Kyrgyzstan	Uyghur	7386.017
Laos	Tai L¸ and Tai Yuan	93096.47
Laos	Other Lao Thoeng	171216.7
Laos	Khmu	426240.3
Laos	Hmong	189418.1
Laos	Other Lao Sung	77174.65
Laos	Phuthai (incl. White, Bla..	461457.2
Laos	Yao	27525.35
Latvia	Byelorussians	21697.28
Lebanon	Greek Catholics	223471.2
Lebanon	Druze	216771.8
Lebanon	Palestinians (Arab)	411946.9
Lebanon	Sunnis (Arab)	450020.5
Lebanon	Eastern Orthodox Christians	301122.6
Malawi	Sena	200139.8
Malawi	Northerners (Nkonde-Tonga..	165386.8
Malawi	Lomwe (Nguru)	421517
Mauritania	Black Africans	306555.9
Moldova	Gagauz	141117
Moldova	Bulgarians	46933.64
Mongolia	Kazakh	30470.88
Mozambique	Makonde-Yao	164458
Myanmar	Chinese	12169.07
Namibia	Damara	17701.46
Namibia	Kavango	108267.9
Namibia	Nama	39759.27
Namibia	San	4802.278
Namibia	Basubia	13300.17
Namibia	Baster	28715.69
Namibia	Herero, Mbanderu	45736.44
Namibia	Mafwe	88868.68
New Zealand	Maori	91390.81
Nicaragua	Mayangnas	56627.16
Nicaragua	Afronicaraguans	5844.135
Nicaragua	Miskitos	40093.47
Niger	Toubou	15155.45

Niger	Kanouri	432221.3
Panama	Afropanamenos	30957.4
Poland	Byelorussians†††	193764.5
Russia	Balkars	280671.2
Russia	Circassians	35984.1
Russia	Buryats	470523.1
Russia	Altai	158071
Russia	Chukchi	66198.29
Russia	Kalmyks	154049.2
Russia	Karachai	175043.2
Russia	Laks	175058.1
Russia	Germans	298880.4
Russia	Adyghe	167380.1
Russia	Tuvinians	253947.5
Russia	Karelians	290858
Russia	Jews	10415.63
Russia	Tabasarans	41123.68
Russia	Lezgins	217660
Russia	Kabardins	305488.8
Russia	Khakass	87611.42
Russia	Nogai	40208.8
Russia	Avars	346839.3
Russia	Ingush	149935.9
Russia	Mari	409649.8
Russia	Dargins	202064.3
Saudi Arabia	Ja'afari Shia (Eastern Pr..	326104.8
Sudan	Bari	478897.7
Sudan	Latoka	97402.71
Sudan	Azande	270128.7
Sudan	Other Northern groups	489113.3
Switzerland	Swiss Italians	308508.1
Syria	Druze	198557.6
Tajikistan	Kyrgyz	19535.29
Thailand	Shan	80640.39
Ukraine	Hungarians	202375
Vietnam	Hoa (Chinese)	142164.6
Yugoslavia	Hungarians	247931.2
Zambia	Lozi (Barotse)	188543.7
Zambia	Luvale (NW Province)	182835.8
Zambia	Kaonde	166155.7
Zambia	Lunda (NW Province)	117749.2

2. Additional sensitivity analysis

In this part of the online appendix, we display results that could not be included in the paper for space reasons. Many of the models are referred to in the main text. All the models are numbered consecutively starting with Model 10.

The key to the variable names used below is as follows:

<u>Dependent variables:</u>

onset_ra = dichotomous onset of group-level conflict
acdonset_ra = onset of group-level territorial (1) or governmental (2) conflict
onset1991, onset1992, onset1993 = static onset for year 1991 (1992, 1993) until 2005

<u>Economic inequality variables (based on Nordhaus geo-coded data)</u>

lineq2 = symmetric inequality measure
low = asymmetric inequality low_ratio
high = asymmetric inequality high_ratio
dlow = asymmetric inequality low_ratio compared to wealth of EGIP
dhigh = asymmetric inequality high_ratio compared to wealth of EGIP
low*dum, high*dum = dummy variables for thresholds 1.3 and 1.5

<u>Political inequality variables (according to EPR)</u>

excluded = group excluded from central power
junior = group included in power sharing in JUNIOR role
autonomy = group excluded but enjoying REGIONAL AUTONOMY
powerless = group excluded and POWERLESS
discrim = group excluded and DISCRIMINATED
separatist = group excluded but enjoying SEPARATIST AUTONOMY
downgraded2 = group had its power status downgraded during previous 2 years

<u>Control variables:</u>

b = power balance
b2 = power balance (sq)
lgdpcapl = country-level log GDP/capita
exclgrps = number of excluded groups in the country
democl = country-level democracy dummy (Polity2 > 6)
anocl = country-level anocracy dummy (-6 ≤ Polity2 ≤ 6)
year = calendar year
pyrs_ra = peace years
spline*_ra = cubic splines
asia = Asian countries
nafrme = countries in North Africa and the Middle East
ssafrica = countries in sub-Saharan Africa
lameric = countries in Latin America

Model 5: Replication of the reference model

In this online appendix, we use the bi-directional Model 5 (or versions of it) as the starting point for our sensitivity analysis (if not otherwise stated). For purpose of direct comparison and clarification of the variable names, we display Model 5 as analyzed in the paper (see Table 2):

```
. // Model 5 testing asymmetry
. relogit onset_ra low high excluded b b2  lgdpcapl exclgrps year pyrs_ra spline*_ra
if ongoingdrop_ra==0 & monop==0 & dominant==0 & year>1990 & dispersed==0 &
pldpop90>$MINPOP, cluster(cowcode)
(2854 missing values generated)


Corrected logit estimates                            Number of obs =     3967

------------------------------------------------------------------------------
             |               Robust
    onset_ra |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         low |    1.12548   .1791567     6.28   0.000     .7743397    1.476621
        high |   1.087477   .2304977     4.72   0.000     .6357095    1.539244
    excluded |   1.206622   .3702018     3.26   0.001     .4810403    1.932205
           b |   5.298835   2.424425     2.19   0.029     .5470496    10.05062
          b2 |  -7.073564   3.570622    -1.98   0.048    -14.07185   -.0752727
    lgdpcapl |  -.8872904    .231266    -3.84   0.000    -1.340563   -.4340173
    exclgrps |  -.0511246   .0167811    -3.05   0.002    -.0840148   -.0182343
        year |  -.2111459    .059233    -3.56   0.000    -.3272405   -.0950513
     pyrs_ra |   .1540236    .103798     1.48   0.138    -.0494167    .3574639
  spline1_ra |   .0026229   .0010861     2.41   0.016     .0004942    .0047517
  spline2_ra |  -.0016319   .0007444    -2.19   0.028    -.0030909    -.000173
  spline3_ra |   .0003671   .0002762     1.33   0.184    -.0001742    .0009084
       _cons |    422.045   118.2983     3.57   0.000     190.1845    653.9055
------------------------------------------------------------------------------
```

Model 10: Multinomial model that separates conflict types

Here we split the dichotomous dependent variable into territorial (1) and governmental conflicts (2) depending on the conflict's incompatibility (see the definition in the Uppsala-PRIO Armed Conflicts Dataset (Gleditsch et al. 2002; Buhaug 2006). The results show that economic HIs apply in the cases of territorial conflict, but not at all for governmental conflict, where the effect is negative (and almost significant for wealthy groups). It could be that the latter category is partly constituted by incumbent groups that have little to gain from upsetting the status quo and already benefit from the current system, as opposed to secessionist groups that have a strong incentive to leave the polity. Model 10 also confirms that political HI operates for both kinds of conflict.

```
. // Model 10: Model 5 disaggregated according to conflict type
. legit acdonset_ra low high excluded b b2  lgdpcapl exclgrps year pyrs_ra spline*_ra
if ongoingdrop_ra==0 & monop==0 & dominant==0 & year>1990 & dispersed==0 &
pldpop90>$MINPOP, nolog cluster(cowcode)

Multinomial logistic regression                   Number of obs   =       3967
                                                  Wald chi2(24)   =     513.25
                                                  Prob > chi2     =     0.0000
Log pseudolikelihood = -204.69455                 Pseudo R2       =     0.2321

                               (Std. Err. adjusted for 95 clusters in cowcode)
------------------------------------------------------------------------------
             |               Robust
 acdonset_ra |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
0            |  (base outcome)
-------------+----------------------------------------------------------------
1            |
         low |   1.317663   .2229033     5.91   0.000     .8807805    1.754545
        high |     1.4847   .2748341     5.40   0.000     .9460351    2.023365
    excluded |   1.183665   .5265715     2.25   0.025     .1516038    2.215726
           b |   2.714804   4.358001     0.62   0.533    -5.826721    11.25633
          b2 |  -7.990996   7.544673    -1.06   0.290    -22.77828    6.796292
    lgdpcapl |  -1.042118   .3325834    -3.13   0.002     -1.69397   -.3902668
    exclgrps |  -.0584935   .0224325    -2.61   0.009    -.1024604   -.0145266
        year |  -.2492637   .0712278    -3.50   0.000    -.3888677   -.1096598
     pyrs_ra |   .2792417   .1353174     2.06   0.039     .0140245    .5444589
  spline1_ra |   .0036816   .0017834     2.06   0.039     .0001861    .0071771
  spline2_ra |  -.0021747   .0013022    -1.67   0.095    -.0047269    .0003776
  spline3_ra |   .0004279   .0004881     0.88   0.381    -.0005288    .0013846
       _cons |    498.304    141.484     3.52   0.000     221.0004    775.6076
-------------+----------------------------------------------------------------
2            |
         low |  -1.288374   .9294461    -1.39   0.166    -3.110055    .5333074
        high |  -1.887959   1.006522    -1.88   0.061    -3.860705    .0847868
    excluded |   1.565943   .4304251     3.64   0.000     .7223252    2.409561
           b |   11.28977   3.705536     3.05   0.002     4.027055    18.55249
          b2 |  -12.55571   5.135482    -2.44   0.014    -22.62107   -2.490345
    lgdpcapl |   -.904782    .274449    -3.30   0.001    -1.442692   -.3668718
    exclgrps |     -.1639   .1093827    -1.50   0.134     -.378286    .0504861
        year |  -.1387973   .1109073    -1.25   0.211    -.3561716    .0785769
     pyrs_ra |  -.0362819   .1696335    -0.21   0.831    -.3687575    .2961937
  spline1_ra |   .0013865   .0012901     1.07   0.282     -.001142     .003915
  spline2_ra |  -.0010928   .0007225    -1.51   0.130    -.0025089    .0003233
  spline3_ra |   .0003627   .0002544     1.43   0.154    -.0001359    .0008614
       _cons |   279.5823   220.6746     1.27   0.205    -152.9319    712.0966
------------------------------------------------------------------------------
```

Model 11: Extending the sample of Model 5

This model extends the sample to the whole post-WWII period as was done in Model 7. As opposed to that model, however, we do so by dividing the effect of poor and wealthy groups. As expected, this analysis confirms the previous result: both categories of groups experience conflict with a higher probability than those that are closer to the mean. Although the magnitude of the coefficients decreases quite significantly compared to Model 5, the results remain robustly significant.

```
. // Model 11: Model 5 with full sample
. relogit onset_ra low high excluded b b2  lgdpcapl exclgrps year pcw pcwyear pyrs_ra
spline*_ra if ongoingdrop_ra==0 & monop==0 & dominant==0 & year>0 & dispersed==0 &
pldpop90>$MINPOP, cluster(cowcode)
(2854 missing values generated)


Corrected logit estimates                             Number of obs =    13550

------------------------------------------------------------------------------
             |               Robust
    onset_ra |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         low |   .5002764    .129391     3.87   0.000     .2466748    .7538781
        high |   .3458384   .1072928     3.22   0.001     .1355484    .5561285
    excluded |   1.388678   .2550426     5.44   0.000     .8888031    1.888552
           b |     1.0247   1.885509     0.54   0.587    -2.670829    4.720229
          b2 |  -1.155083   2.139388    -0.54   0.589    -5.348206    3.038041
    lgdpcapl |  -.5067674    .146264    -3.46   0.001    -.7934396   -.2200953
    exclgrps |  -.0218526   .0137301    -1.59   0.111    -.0487631    .0050578
        year |   .0512168   .0174377     2.94   0.003     .0170395    .0853942
         pcw |   360.7415   86.57101     4.17   0.000     191.0655    530.4176
     pcwyear |  -.1810215   .0434545    -4.17   0.000    -.2661908   -.0958523
     pyrs_ra |  -.0822467   .0905569    -0.91   0.364    -.2597349    .0952416
  spline1_ra |   .0006261   .0006687     0.94   0.349    -.0006845    .0019366
  spline2_ra |  -.0007527     .00041    -1.84   0.066    -.0015564    .0000509
  spline3_ra |   .0004071   .0001603     2.54   0.011     .0000929    .0007213
       _cons |  -102.6105   34.28258    -2.99   0.003    -169.8032    -35.4179
------------------------------------------------------------------------------

.
```

Model 12: Estimating low and high ratios relative to the wealth of the EGIP

The relational logic of our theory should hold for comparisons with the group in power rather than with the country average as well. As indicated above, using Model 5 as the starting point, we introduce alternative ratio measures that refer to the GDP per capita of the governmental groups (EGIPs) as the comparison rather than to the average GDP per capita of the entire country. Unsurprisingly we find a strong, if marginally less pronounced (cf. Model 5), effect of the dyadic measures of economic HIs.

```
. // Model 12: Model 5 with relative low & high estimates
. relogit onset_ra dlow dhigh excluded b b2  lgdpcapl exclgrps year pyrs_ra spline*_ra
if ongoingdrop_ra==0 & monop==0 & dominant==0 & year>1990 & dispersed==0 &
pldpop90>$MINPOP, cluster(cowcode)
(3116 missing values generated)


Corrected logit estimates                              Number of obs =     3937

------------------------------------------------------------------------------
             |               Robust
    onset_ra |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        dlow |   .9200859   .1970493     4.67   0.000     .5338764    1.306295
       dhigh |   .7644951   .1714516     4.46   0.000      .428456    1.100534
    excluded |     1.1319   .3501855     3.23   0.001     .4455489    1.818251
           b |   5.306351    2.45152     2.16   0.030     .5014604    10.11124
          b2 |  -7.209645   3.628468    -1.99   0.047    -14.32131   -.0979787
    lgdpcapl |   -.827265   .2242247    -3.69   0.000    -1.266737   -.3877926
    exclgrps |  -.0429078   .0179814    -2.39   0.017    -.0781506    -.007665
        year |   -.209932    .059605    -3.52   0.000    -.3267558   -.0931083
     pyrs_ra |   .1374221   .1015695     1.35   0.176    -.0616505    .3364946
  spline1_ra |   .0024738   .0010798     2.29   0.022     .0003573    .0045902
  spline2_ra |  -.0015412   .0007483    -2.06   0.039    -.0030078   -.0000745
  spline3_ra |   .0003433   .0002819     1.22   0.223    -.0002093    .0008959
       _cons |   419.5683   119.0456     3.52   0.000     186.2433    652.8934
------------------------------------------------------------------------------
```

Models 13a-c: Testing the robustness by dropping extreme values

As we point out in the paper, it is important to check that our results are not driven by a very small set of extreme values. Therefore we introduce a series of checks that removes some of the most extreme values. Our analysis starts by removing the two most extreme conflict cases where the asymmetric inequality ratios are the highest. Among the poor groups, these cases correspond to the conflict onsets associated with the Chechens (*low_ratio* = 6.04). As regards the (resource) wealthy groups, the Nigerian Ijaws (*high_ratio* = 3.34) and the Iranian Arabs (*high_ratio* = 1.99). This modification of the sample hardly changes the size of the coefficients, but the precision of the estimate suffers considerably although it remains robustly significant according the usual standard.

```
. // Model 13a: Model 5 Without the Chechens, the Ijaw and the Iranian Arabs
. relogit onset_ra low high excluded b b2  lgdpcapl exclgrps year  pyrs_ra spline*_ra
if ongoingdrop_ra==0 & monop==0 & dominant==0 & year>1990 & dispersed==0 &
pldpop90>$MINPOP & group!="Chechens" & group!="Ijaw" & !(country=="Iran" &
group=="Arabs"), cluster(cowcode)
(3314 missing values generated)


Corrected logit estimates                              Number of obs =     3930

------------------------------------------------------------------------------
             |               Robust
    onset_ra |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         low |   .9923565   .3150963     3.15   0.002     .3747792    1.609934
        high |   1.107345   .5240424     2.11   0.035     .0802404    2.134449
    excluded |   1.225736   .3470308     3.53   0.000     .5455678    1.905903
           b |   5.560066   2.616472     2.13   0.034     .4318758    10.68826
          b2 |  -7.491057   3.937945    -1.90   0.057    -15.20929    .2271724
    lgdpcapl |  -.8518084   .2274569    -3.74   0.000    -1.297616    -.406001
    exclgrps |  -.0550715   .0220994    -2.49   0.013    -.0983855   -.0117575
        year |  -.2333603   .0652894    -3.57   0.000    -.3613251   -.1053954
     pyrs_ra |   .1566619   .1139146     1.38   0.169    -.0666066    .3799303
  spline1_ra |   .0025446   .0011187     2.27   0.023      .000352    .0047372
  spline2_ra |  -.0015696   .0007508    -2.09   0.037    -.0030411   -.0000982
  spline3_ra |   .0003481   .0002755     1.26   0.206    -.0001918    .0008881
       _cons |   466.1814   130.3732     3.58   0.000     210.6546    721.7082
------------------------------------------------------------------------------
```

Based on the multinomial Model 10, the robustness looks even more impressive for territorial conflicts:

```
. // Model 13b: Model 10 Without the Chechen, the Ijaw and the Iranian Arabs
. mlogit acdonset_ra low high excluded b b2  lgdpcapl exclgrps year  pyrs_ra
spline*_ra if ongoingdrop_ra==0 & monop==0 & dominant==0 & year>1990 & dispersed==0 &
pldpop90>$MINPOP & group!="Chechens" & group!="Ijaw" & !(country=="Iran" &
group=="Arabs"), nolog cluster(cowcode)

Multinomial logistic regression                   Number of obs   =       3930
                                                  Wald chi2(24)   =     424.98
                                                  Prob > chi2     =     0.0000
Log pseudolikelihood = -194.19713                 Pseudo R2       =     0.2258

                              (Std. Err. adjusted for 95 clusters in cowcode)
------------------------------------------------------------------------------
             |               Robust
  acdonset_ra |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
0            |  (base outcome)
-------------+----------------------------------------------------------------
1            |
         low |   1.389628   .4291458     3.24   0.001      .548518    2.230738
        high |   1.883261   .7006596     2.69   0.007     .5099938    3.256529
    excluded |   1.285646   .5185751     2.48   0.013     .2692577    2.302035
           b |   4.437058   5.700373     0.78   0.436    -6.735468    15.60958
          b2 |  -10.64252   9.705775    -1.10   0.273    -29.66549    8.380446
    lgdpcapl |  -1.066171   .3614753    -2.95   0.003     -1.77465   -.3576926
    exclgrps |  -.0674573   .0260542    -2.59   0.010    -.1185226   -.0163919
        year |  -.2983244   .0871125    -3.42   0.001    -.4690618    -.127587
     pyrs_ra |   .3089378   .1420976     2.17   0.030     .0304317    .5874439
  spline1_ra |   .0036834    .001683     2.19   0.029     .0003849     .006982
  spline2_ra |  -.0021301   .0012086    -1.76   0.078     -.004499    .0002387
  spline3_ra |   .0004019   .0004576     0.88   0.380    -.0004949    .0012987
       _cons |   595.7524   173.1625     3.44   0.001     256.3602    935.1446
-------------+----------------------------------------------------------------
2            |
         low |  -1.278281   .9363339    -1.37   0.172    -3.113462    .5568999
        high |  -1.870035   1.019149    -1.83   0.067     -3.86753    .1274599
    excluded |    1.57002   .4304858     3.65   0.000     .7262832    2.413756
           b |   11.30347   3.705038     3.05   0.002      4.04173    18.56521
          b2 |  -12.57842   5.140941    -2.45   0.014    -22.65448   -2.502359
    lgdpcapl |  -.9069968   .2749815    -3.30   0.001    -1.445951   -.3680431
    exclgrps |  -.1641087    .109329    -1.50   0.133    -.3783895    .0501721
        year |  -.1397177   .1112642    -1.26   0.209    -.3577916    .0783561
     pyrs_ra |  -.0351374   .1702219    -0.21   0.836    -.3687662    .2984914
  spline1_ra |    .001395   .0012938     1.08   0.281    -.0011408    .0039309
  spline2_ra |  -.0010969   .0007236    -1.52   0.130    -.0025151    .0003213
  spline3_ra |   .0003631   .0002543     1.43   0.153    -.0001352    .0008614
       _cons |   281.4167   221.3848     1.27   0.204    -152.4896     715.323
------------------------------------------------------------------------------
```

Finally, we drop all groups in the countries that are associated with the extreme values, namely Russia, Nigeria and Iran. As can be seen below, this does not significantly alter the results either:

```
. // Model 13c: Model 5 Without Russia, Nigeria and Iran
. relogit onset_ra low high excluded b b2  lgdpcapl exclgrps year  pyrs_ra spline*_ra
if ongoingdrop_ra==0 & monop==0 & dominant==0 & year>1990 & dispersed==0 &
pldpop90>$MINPOP & country!="Russia" & country!="Nigeria" & country!="Iran",
cluster(cowcode)
(3314 missing values generated)

Corrected logit estimates                              Number of obs =     3637

------------------------------------------------------------------------------
             |               Robust
    onset_ra |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         low |   1.110445   .3268305     3.40   0.001     .4698691    1.751021
        high |   1.247418   .5647772     2.21   0.027      .140475    2.354361
    excluded |    1.20947   .3493072     3.46   0.001     .5248402    1.894099
           b |   5.671244   2.563103     2.21   0.027     .6476536    10.69483
          b2 |  -7.612516   3.831917    -1.99   0.047    -15.12294   -.1020956
    lgdpcapl |  -.8336679   .2274341    -3.67   0.000    -1.279431   -.3879052
    exclgrps |  -.0385147   .0242277    -1.59   0.112    -.0860001    .0089708
        year |  -.2317681    .063971    -3.62   0.000    -.3571489   -.1063873
     pyrs_ra |    .158313   .1151212     1.38   0.169    -.0673204    .3839463
  spline1_ra |    .002541   .0011179     2.27   0.023       .00035     .004732
  spline2_ra |  -.0015776   .0007505    -2.10   0.036    -.0030487   -.0001066
  spline3_ra |   .0003615   .0002774     1.30   0.193    -.0001823    .0009053
       _cons |   462.6613   127.7048     3.62   0.000     212.3645    712.9581
------------------------------------------------------------------------------
```

Model 14: Disaggregating the EPR categories and introducing downgrading

Following the analysis of Cederman, Wimmer and Min (2010), we disaggregate the dichotomous coding of political HIs by relying on EPR categories. In this model, we use SENIOR PARTNER as the reference category while introducing separate dummy variables for each EPR category of power status. Apart from JUNIOR PARTNER, the remaining categories all correspond to excluded groups (see the description of the EPR categories above). In addition, we introduce a measure of whether the group in question was downgraded during the previous two years (cf. Cederman, Wimmer and Min 2010). Based on this disaggregation of political HIs, we find that all the main results hold. First, the *low_ratio* and *high_ratio* measures remain virtually unchanged. Of the separate dummy variables capturing political HIs, only SEPARATIST AUTONOMY reaches significant, but the effect grows roughly in order to more severe exclusion, as would be expected. In addition, the dummy measuring downgraded power status has a strong and clearly significant effect.

```
. // Model 14: Model 5 With all EPR categories & downgrading
. relogit onset_ra low high junior auto powerless discrim separatist downgraded2 b b2
lgdpcapl exclgrps year pyrs_ra spline*_ra if ongoingdrop_ra==0 & monop==0 &
dominant==0 & year>1990 & dispersed==0 & pldpop90>$MINPOP, cluster(cowcode)
(3314 missing values generated)

Corrected logit estimates                           Number of obs =     3967

------------------------------------------------------------------------------
             |               Robust
    onset_ra |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         low |   1.096688   .2274986     4.82   0.000     .6507993    1.542578
        high |   .9115534   .2838355     3.21   0.001     .3552461    1.467861
      junior |  -.1396038   .8505852    -0.16   0.870     -1.80672    1.527513
    autonomy |   1.103052   .8536449     1.29   0.196    -.5700616    2.776165
   powerless |   .5202338   .7189868     0.72   0.469    -.8889545    1.929422
     discrim |   1.060564   .7289997     1.45   0.146    -.3682496    2.489377
  separatist |   1.710564   .6876844     2.49   0.013      .362727      3.0584
 downgraded2 |   2.101366   .4548555     4.62   0.000     1.209865    2.992866
           b |   4.666654   2.565181     1.82   0.069    -.3610091    9.694317
          b2 |  -6.452856   3.763473    -1.71   0.086    -13.82913     .923415
    lgdpcapl |  -.7861268   .2299242    -3.42   0.001     -1.23677   -.3354836
    exclgrps |  -.0654124   .0270844    -2.42   0.016    -.1184969   -.0123279
        year |  -.1893262   .0663755    -2.85   0.004    -.3194198   -.0592326
     pyrs_ra |   .1853598   .1043873     1.78   0.076    -.0192356    .3899552
  spline1_ra |    .002523   .0010911     2.31   0.021     .0003845    .0046616
  spline2_ra |  -.0014644   .0007833    -1.87   0.062    -.0029996    .0000708
  spline3_ra |   .0002675   .0003177     0.84   0.400    -.0003551    .0008902
       _cons |    377.851   132.7721     2.85   0.004     117.6226    638.0795
------------------------------------------------------------------------------
```

Model 15a,b: Controlling for world regions

It could be that macro-historical processes undermine the assumption of causal homogeneity across world regions. We therefore introduce a series of dummy variables, defined according to the convention given by MAR (see codebook). Model 15a treats the western world and Eastern Europe as reference category and introduces a series of dummies that tests effects associated with Asia, North Africa and the Middle East (*nafrme*), Sub-Saharan Africa (*ssafrica*), and Latin America (*lamerica*). Again, we record no major changes in the main variables: the effect of economic and political HIs remains strong and significant. Dropping Eastern Europe and the Former Soviet Union makes no difference either (see Model 15b).

```
. // Model 15a: Controlling for world regions
. relogit onset_ra low high excluded b b2  lgdpcapl exclgrps year pyrs_ra spline*_ra
eeurop asia nafrme ssafrica lamerica if ongoingdrop_ra==0 & monop==0 & dominant==0 &
year>1990 & dispersed==0 & pldpop90>$MINPOP, cluster(cowcode)
(3314 missing values generated)

Corrected logit estimates                            Number of obs =     3967
```

onset_ra	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
low	1.266704	.2090516	6.06	0.000	.8569706	1.676438
high	1.21625	.2601556	4.68	0.000	.706354	1.726145
excluded	1.37706	.434196	3.17	0.002	.5260515	2.228069
b	8.832911	2.913473	3.03	0.002	3.122609	14.54321
b2	-10.2825	3.470599	-2.96	0.003	-17.08475	-3.480249
lgdpcapl	-1.170989	.3069489	-3.81	0.000	-1.772598	-.5693802
exclgrps	-.0609003	.0153569	-3.97	0.000	-.0909992	-.0308014
year	-.2061835	.0593612	-3.47	0.001	-.3225293	-.0898378
pyrs_ra	.1644511	.1132483	1.45	0.146	-.0575115	.3864138
spline1_ra	.0026975	.0011351	2.38	0.017	.0004727	.0049224
spline2_ra	-.0016986	.0007647	-2.22	0.026	-.0031975	-.0001997
spline3_ra	.000407	.0002779	1.46	0.143	-.0001376	.0009516
eeurop	-1.687747	1.045168	-1.61	0.106	-3.736238	.3607437
asia	-1.348251	1.173186	-1.15	0.250	-3.647654	.9511513
nafrme	-2.902783	1.205688	-2.41	0.016	-5.265889	-.5396772
ssafrica	-2.760396	1.288043	-2.14	0.032	-5.284914	-.2358785
lamerica	-2.524026	1.503845	-1.68	0.093	-5.471509	.4234562
_cons	415.6117	119.0845	3.49	0.000	182.2102	649.0131

```
. // Model 15b: Dropping Eastern Europe and the FSU
. relogit onset_ra low high excluded b b2  lgdpcapl exclgrps year pyrs_ra spline*_ra
if ongoingdrop_ra==0 & monop==0 & dominant==0 & year>1990 & dispersed==0 &
pldpop90>$MINPOP & eeurop==0, cluster(cowcode)
(2854 missing values generated)

Corrected logit estimates                            Number of obs =     3380
```

onset_ra	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
low	.8208961	.3004198	2.73	0.006	.2320841	1.409708
high	.8508218	.2506009	3.40	0.001	.359653	1.34199
excluded	1.405338	.3535726	3.97	0.000	.7123483	2.098327
b	3.681968	2.386158	1.54	0.123	-.9948161	8.358752
b2	-4.562759	3.073907	-1.48	0.138	-10.58751	1.461988
lgdpcapl	-.7271306	.2596935	-2.80	0.005	-1.236121	-.2181406
exclgrps	-.0425488	.0264989	-1.61	0.108	-.0944857	.0093882
year	-.1755738	.0593954	-2.96	0.003	-.2919867	-.059161
pyrs_ra	.1509673	.1114345	1.35	0.175	-.0674403	.369375
spline1_ra	.0025798	.0011681	2.21	0.027	.0002904	.0048692
spline2_ra	-.0016276	.0007881	-2.07	0.039	-.0031723	-.000083
spline3_ra	.0003957	.0002853	1.39	0.166	-.0001636	.000955
_cons	350.3298	118.7388	2.95	0.003	117.6061	583.0536

Model 16a-b: Replacing the continuous HI variables with thresholds

Our results so far have been dependent on assumptions about the functional form. Here we replace the *low_ratio* and *high_ratio* variables by simple dummy variables that are coded as one if the exceed a specific threshold and zero otherwise. The first test introduces a thresholds corresponding to a ratio of 1.3, which turn out to be positive but not at all significant.

```
. // Model 16a: Model 5 With dummies at 1.3
. relogit onset_ra low13dum high13dum excluded b b2  lgdpcapl exclgrps year  pyrs_ra
spline*_ra if ongoingdrop_ra==0 & monop==0 & dominant==0 & year>1990 & dispersed==0 &
pldpop90>$MINPOP, cluster(cowcode)
(790 missing values generated)


Corrected logit estimates                              Number of obs =     4799
```

onset_ra	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
low13dum	.8896108	.5325559	1.67	0.095	-.1541797	1.933401
high13dum	1.070043	.6066126	1.76	0.078	-.1188959	2.258982
excluded	1.200492	.3315633	3.62	0.000	.5506403	1.850344
b	6.010122	2.413707	2.49	0.013	1.279342	10.7409
b2	-7.660869	3.597533	-2.13	0.033	-14.7119	-.6098344
lgdpcapl	-.8730976	.1925284	-4.53	0.000	-1.250446	-.4957489
exclgrps	-.0220227	.0273943	-0.80	0.421	-.0757146	.0316691
year	-.209715	.0631326	-3.32	0.001	-.3334526	-.0859774
pyrs_ra	.0992047	.1034815	0.96	0.338	-.1036154	.3020248
spline1_ra	.0022565	.0010641	2.12	0.034	.000171	.0043421
spline2_ra	-.001474	.0007337	-2.01	0.045	-.0029119	-.000036
spline3_ra	.0003659	.0002759	1.33	0.185	-.0001748	.0009066
_cons	420.1591	126.1556	3.33	0.001	172.8987	667.4195

Moving the threshold to 1.5 improves the results significantly, but the choice of threshold location remains a difficult one, so we prefer the continuous functional form. There are good theoretical reasons to believe that the amount of HI should have an impact on the probability of conflict.

```
. // Model 16b: Model 5 With dummies at 1.5
. relogit onset_ra low15dum high15dum excluded b b2  lgdpcapl exclgrps year  pyrs_ra
spline*_ra if ongoingdrop_ra==0 & monop==0 & dominant==0 & year>1990 & dispersed==0 &
pldpop90>$MINPOP, cluster(cowcode)
(790 missing values generated)


Corrected logit estimates                              Number of obs =     4799
```

onset_ra	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
low15dum	1.359316	.5555949	2.45	0.014	.2703703	2.448262
high15dum	1.626444	.5365958	3.03	0.002	.5747355	2.678152
excluded	1.129497	.3481207	3.24	0.001	.4471929	1.811801
b	6.699521	2.69394	2.49	0.013	1.419496	11.97955
b2	-8.705619	4.13415	-2.11	0.035	-16.8084	-.602835
lgdpcapl	-.8813994	.1964597	-4.49	0.000	-1.266453	-.4963454
exclgrps	-.0184806	.02526	-0.73	0.464	-.0679893	.0310281
year	-.2099947	.0628813	-3.34	0.001	-.3332397	-.0867496
pyrs_ra	.1104146	.1029907	1.07	0.284	-.0914434	.3122725
spline1_ra	.0022997	.0010524	2.19	0.029	.0002371	.0043623
spline2_ra	-.0014756	.000726	-2.03	0.042	-.0028986	-.0000527
spline3_ra	.0003533	.000275	1.28	0.199	-.0001857	.0008923
_cons	420.7231	125.599	3.35	0.001	174.5536	666.8926

Model 17: Country-fixed effects analysis

Since we cannot exclude possible omitted-variable bias, we also run Model 3 with country fixed effect. As can be seen, this modification does not in any way weaken our main results (see H1 and H2). However, the GDP per capita measure becomes insignificant.

```
. // Model 17: Model 5 with country-fixed effects
. clogit onset_ra low high excluded b b2  lgdpcapl exclgrps year pyrs_ra spline1_ra
spline2_ra spline3_ra if ongoingdrop_ra==0 & monop==0 & dominant==0 & year>1990 &
dispersed==0 & pldpop90>$MINPOP, group(cowcode) cluster(cowcode)
note: multiple positive outcomes within groups encountered.
note: 71 groups (2558 obs) dropped because of all positive or
      all negative outcomes.

Conditional (fixed-effects) logistic regression   Number of obs   =       1409
                                                  Wald chi2(12)   =     125.68
                                                  Prob > chi2     =     0.0000
Log pseudolikelihood = -120.35959                 Pseudo R2       =     0.1743

                               (Std. Err. adjusted for 24 clusters in cowcode)
------------------------------------------------------------------------------
             |               Robust
    onset_ra |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         low |   1.512141   .4571472     3.31   0.001     .6161493    2.408133
        high |   1.784238   .6273846     2.84   0.004     .5545869    3.013889
    excluded |   1.370915    .679826     2.02   0.044     .0384807     2.70335
           b |   4.737713   5.664777     0.84   0.403    -6.365046    15.84047
          b2 |  -6.513885   6.634189    -0.98   0.326    -19.51666    6.488886
    lgdpcapl |   .1051605   .7903294     0.13   0.894    -1.443857    1.654178
    exclgrps |   .0178291   .3716362     0.05   0.962    -.7105645    .7462226
        year |  -.2138767   .0663929    -3.22   0.001    -.3440044   -.0837491
     pyrs_ra |   .2455128   .1594406     1.54   0.124    -.0669849    .5580106
  spline1_ra |   .0028842   .0023842     1.21   0.226    -.0017887    .0075572
  spline2_ra |  -.0017622   .0016828    -1.05   0.295    -.0050605    .0015361
  spline3_ra |    .000418   .0005653     0.74   0.460      -.00069    .0015261
------------------------------------------------------------------------------


.
```

Models 18a-c: Sensitivity analysis of Model 9

Model 9 uses 1992 as the base year for static analysis. Model 18a and 18b test the robustness of that assumption by using 1991 and 1993 as base years respectively. Model 18c changes the main IV by introducing asymmetric indicators of HI. These modifications make no major difference to the results reported in Model 9:

```
. // Model 18a: Static version based on groups in 1991 (see Model 9)
. relogit onset1991 lineq2 excluded b b2  lgdpcapl exclgrps if monop==0 & dominant==0
& year==1991 & dispersed==0 & pldpop90>$MINPOP, cluster(cowcode)
(2854 missing values generated)


Corrected logit estimates                              Number of obs =      276

------------------------------------------------------------------------------
             |               Robust
   onset1991 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lineq2 |   1.976879   .5072361     3.90   0.000     .9827143    2.971043
    excluded |   .7207285   .4701078     1.53   0.125     -.200666    1.642123
           b |  -1.308668   2.638273    -0.50   0.620    -6.479588    3.862252
          b2 |   1.150364   3.156767     0.36   0.716    -5.036786    7.337514
    lgdpcapl |  -.8060983   .3007333    -2.68   0.007    -1.395525   -.2166719
    exclgrps |  -.0847903   .0354168    -2.39   0.017     -.154206   -.0153746
       _cons |   4.162983   2.245876     1.85   0.064    -.2388536     8.56482
------------------------------------------------------------------------------

. // Model 18b: Static version based on groups in 1993 (see Model 9)
. relogit onset1993 lineq2 excluded b b2  lgdpcapl exclgrps if monop==0 & dominant==0
& year==1992 & dispersed==0 & pldpop90>$MINPOP, cluster(cowcode)
(2854 missing values generated)


Corrected logit estimates                              Number of obs =      286

------------------------------------------------------------------------------
             |               Robust
   onset1993 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lineq2 |    2.28353   .5870028     3.89   0.000     1.133026    3.434035
    excluded |    .545323   .4828046     1.13   0.259    -.4009566    1.491603
           b |  -1.044353   2.515811    -0.42   0.678    -5.975252    3.886545
          b2 |   .2562611   2.838991     0.09   0.928    -5.308058     5.82058
    lgdpcapl |  -1.014056   .3204531    -3.16   0.002    -1.642132   -.3859793
    exclgrps |  -.0858769   .0435568    -1.97   0.049    -.1712467   -.0005071
       _cons |    5.40961   2.263289     2.39   0.017     .9736456    9.845575
------------------------------------------------------------------------------

. // Model 18c: Model 9 Static version based on groups in 1992 but with low/high
ratios
. relogit onset1992 low high excluded b b2  lgdpcapl exclgrps if monop==0 &
dominant==0 & year==1992 & dispersed==0 & pldpop90>$MINPOP, cluster(cowcode)
(3314 missing values generated)


Corrected logit estimates                              Number of obs =      286

------------------------------------------------------------------------------
             |               Robust
   onset1992 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         low |   1.216225   .3458196     3.52   0.000     .5384312    1.894019
        high |   1.232954   .3460073     3.56   0.000     .5547919    1.911116
    excluded |    .645501   .4455879     1.45   0.147    -.2278353    1.518837
           b |  -1.180276   2.594524    -0.45   0.649    -6.265449    3.904897
          b2 |   .8118693   2.977512     0.27   0.785    -5.023946    6.647685
    lgdpcapl |  -.8922159   .2640618    -3.38   0.001    -1.409768   -.3746643
    exclgrps |  -.0748419   .0333416    -2.24   0.025    -.1401903   -.0094935
       _cons |   3.253353   1.842116     1.77   0.077    -.3571272    6.863834
------------------------------------------------------------------------------
```

Model 19. Model 5 based on full sample without size restriction

We also test the robustness of the group-size-dependent sample restriction by rerunning Model 5 for all group sizes. The coefficients for the asymmetric indicators for economic HIs become much smaller but remain robustly significant. There is no noticeable change as regards the *excluded* variable.

```
. // Model 19: Model 5 with all groups
. relogit onset_ra low high excluded b b2  lgdpcapl exclgrps year pyrs_ra spline*_ra
if ongoingdrop_ra==0 & monop==0 & dominant==0 & year>199
> 0 & dispersed==0, cluster(cowcode)
(2854 missing values generated)


Corrected logit estimates                             Number of obs =     6438

------------------------------------------------------------------------------
             |               Robust
    onset_ra |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         low |   .6378284   .2957289     2.16   0.031     .0582103    1.217446
        high |   .3700411   .1328615     2.79   0.005     .1096372    .6304449
    excluded |   1.100389   .3196524     3.44   0.001     .4738817    1.726896
           b |   5.693608   2.357507     2.42   0.016      1.07298    10.31424
          b2 |   -6.44521   2.990985    -2.15   0.031    -12.30743   -.5829882
    lgdpcapl |  -.5088808   .1943181    -2.62   0.009    -.8897373   -.1280243
    exclgrps |  -.0307882   .0162071    -1.90   0.057    -.0625536    .0009771
        year |  -.1814052   .0559526    -3.24   0.001    -.2910703   -.0717401
     pyrs_ra |   .0009199   .1052635     0.01   0.993    -.2053928    .2072325
  spline1_ra |   .0010286   .0009959     1.03   0.302    -.0009234    .0029806
  spline2_ra |   -.000695   .0006562    -1.06   0.290    -.0019812    .0005911
  spline3_ra |   .0001738   .0002459     0.71   0.480    -.0003083    .0006558
       _cons |   360.6668   111.8376     3.22   0.001     141.4692    579.8644
------------------------------------------------------------------------------
```

<u>Model 20: Model with interaction effects between political and economic HI</u>

In this model, we consider a possible interaction effect between political and economic HIs. The following interaction terms have been added:

*excl_low = excluded * low_ratio*

*excl_high = excluded * high_ratio*

As can be readily seen, this addition does not bring any interesting results. The straight effects remain relatively intact, but the sign of the interaction terms is negative but the effect is not at all significant. Thus it can be concluded that political and economic HIs operate mostly independently of each other and that the effect of one form of HI does not seem to depend on the level of the other.

```
. // Model 20: Model 5 with interaction effects
. relogit onset_ra low high excluded excl_low excl_high b b2  lgdpcapl exclgrps year
pyrs_ra spline*_ra if ongoingdrop_ra==0 & monop==0 & dom
> inant==0 & year>1990 & dispersed==0 & pldpop90>$MINPOP, cluster(cowcode)
(2854 missing values generated)

Corrected logit estimates                               Number of obs =     3967

------------------------------------------------------------------------------
             |               Robust
    onset_ra |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         low |   1.399138   .7180464     1.95   0.051    -.0082072    2.806483
        high |   2.279955   1.349953     1.69   0.091    -.3659038    4.925813
    excluded |   2.053642   1.232942     1.67   0.096    -.3628796    4.470163
    excl_low |  -.2054029   .7545026    -0.27   0.785    -1.684201    1.273395
   excl_high |  -1.338865   1.360487    -0.98   0.325     -4.00537     1.32764
           b |   5.880607   2.503134     2.35   0.019      .974555    10.78666
          b2 |  -7.946467    3.88245    -2.05   0.041    -15.55593   -.3370042
    lgdpcapl |  -.9000918     .23624    -3.81   0.000    -1.363114   -.4370698
    exclgrps |  -.0589938   .0188255    -3.13   0.002    -.0958911   -.0220964
        year |  -.2117731   .0583683    -3.63   0.000    -.3261729   -.0973734
     pyrs_ra |   .1776617   .1078056     1.65   0.099    -.0336334    .3889569
  spline1_ra |   .0028366   .0011086     2.56   0.011     .0006638    .0050094
  spline2_ra |  -.0017692   .0007546    -2.34   0.019    -.0032483   -.0002902
  spline3_ra |   .0004125   .0002764     1.49   0.136    -.0001293    .0009542
       _cons |   422.5021   116.1612     3.64   0.000     194.8304    650.1739
------------------------------------------------------------------------------

.
```

Model 21: Model with controls for regime type

Controlling for regime type at the country level with two dummies corresponding to lagged democracy and anocracy does little to alter the main results. It should be noted that these indicators, based on the Polity2 indicator, are subject to possible concerns over endogeneity to conflict (Vreeland 2008).

```
. // Model 21: Model 5 with regime type
. relogit onset_ra low high excluded b b2  lgdpcapl exclgrps democl anocl year pyrs_ra
spline*_ra if ongoingdrop_ra==0 & monop==0 & dominant==0
> & year>1990 & dispersed==0 & pldpop90>$MINPOP, cluster(cowcode)
(3203 missing values generated)


Corrected logit estimates                              Number of obs =     3860

------------------------------------------------------------------------------
             |               Robust
    onset_ra |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         low |   1.031156   .1905398     5.41   0.000     .6577054    1.404608
        high |   .9973124   .2284009     4.37   0.000     .5496548     1.44497
    excluded |   1.109284   .3909555     2.84   0.005     .3430254    1.875543
           b |   5.193843   2.590194     2.01   0.045      .117155    10.27053
          b2 |  -7.283756   3.821117    -1.91   0.057    -14.77301    .2054957
    lgdpcapl |  -.6143092   .2665107    -2.31   0.021    -1.136661   -.0919578
    exclgrps |   -.060739   .0177377    -3.42   0.001    -.0955043   -.0259737
      democl |  -1.241864   .6949375    -1.79   0.074    -2.603916    .1201884
       anocl |    .223378   .3552245     0.63   0.529    -.4728491    .9196051
        year |  -.1961169   .0574857    -3.41   0.001    -.3087868   -.0834469
     pyrs_ra |   .1463468   .1062937     1.38   0.169     -.061985    .3546786
  spline1_ra |   .0025324   .0011188     2.26   0.024     .0003396    .0047252
  spline2_ra |  -.0015573   .0007649    -2.04   0.042    -.0030565   -.0000581
  spline3_ra |     .00033   .0002796     1.18   0.238     -.000218    .0008779
       _cons |   390.4621   114.8954     3.40   0.001     165.2713     615.653
------------------------------------------------------------------------------
```

Model 22: Controlling for previous war

As indicated by Cederman, Wimmer and Min (2010) the effect of previous wars on conflict onset can be quite strong. We therefore introduce a count variable the counts the number of previous conflicts since 1946. Since our sample is restricted to cases after 1990, this variable exhibits a rather moderate effect that does not reach conventional levels of significance.

```
. // Model 22: Model 5 with count variable for previous war
. relogit onset_ra low high excluded b b2  lgdpcapl exclgrps year warhist_ra pyrs_ra
spline*_ra if ongoingdrop_ra==0 & monop==0 & dominant==0 &
> year>1990 & dispersed==0 & pldpop90>$MINPOP, cluster(cowcode)
(2854 missing values generated)

Corrected logit estimates                              Number of obs =     3967

------------------------------------------------------------------------------
             |               Robust
    onset_ra |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         low |    1.13029   .1793572     6.30   0.000     .7787566    1.481824
        high |   1.106311   .2401135     4.61   0.000     .6356975    1.576925
    excluded |   1.136888   .3717433     3.06   0.002     .4082845    1.865491
           b |   5.907219   2.479652     2.38   0.017      1.04719    10.76725
          b2 |  -7.874962   3.650894    -2.16   0.031    -15.03058    -.719341
    lgdpcapl |  -.8661335   .2272653    -3.81   0.000    -1.311565   -.4207018
    exclgrps |  -.0521687    .015781    -3.31   0.001    -.0830988   -.0212386
        year |  -.2190078    .059375    -3.69   0.000    -.3353806    -.102635
  warhist_ra |   .3565735   .2241686     1.59   0.112    -.0827889     .795936
     pyrs_ra |   .1699227    .102141     1.66   0.096    -.0302699    .3701154
  spline1_ra |   .0026363   .0011018     2.39   0.017     .0004768    .0047958
  spline2_ra |  -.0016241    .000766    -2.12   0.034    -.0031254   -.0001229
  spline3_ra |   .0003592   .0002875     1.25   0.211    -.0002042    .0009227
       _cons |   437.2264   118.4981     3.69   0.000     204.9744    669.4783
------------------------------------------------------------------------------
```

References

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

Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Håvard Strand. 2002. “Armed Conflict 1946-2001: A New Dataset.” *Journal of Peace Research* 39:615-637.

Lang, Nicholas R. 1975. “The Dialectics of Decentralization: Economic Reform and Regional Inequality in Yugoslavia.” *World Politics* 27 (3): 309-335.

Minorities at Risk Dataset Users Manual 030703 (2001-2003 Release). Downloaded from http://www.cidcm.umd.edu/mar/data.asp.

Vreeland, James Raymond. 2008. “The Effect of Political Regime on Civil War : Unpacking Anocracy.” *Journal of Conflict Resolution* 52:401-425.

Wucherpfennig et al. 2010. “Politically Relevant Ethnic Groups across Space and Time: Introducing the GeoEPR Dataset.” Unpublished ms., ETH Zurich.