Redemption Through Rebellion: Border Change, Lost Unity and Nationalist Conflict

Analysis datasets: Codebook

Definitions and sources of all variables used in the analysis

Lars-Erik Cederman
Seraina Rüegger
Guy Schvitz

February 10, 2021

1 epr_segment_level_analysis.dta

This dataset was used to generate the results shown in Tables 1, 2, 3 and Figures 7 and 8 in the main text and all Tables and Figures shown in the Appendix (except for Table A16 and Figure A4).

1.1 Sources

- EPR-AG (Schvitz, 2018; Vogt et al., 2015)
- GeoEPR-AG (Schvitz, 2018; Vogt et al., 2015)
- CShapes 2.0 (Schvitz et al., 2020)
- HYDE Historical Population Database (Klein Goldewijk et al., 2011)
- EPR-SDM Claim dataset (Schädel, 2018)
- EPR-EGIP-SDM Claim dataset (Rüegger, 2018)

Table 1: Codebook for analysis dataset 'epr_segment_level_analysis.dta'

varname	label	description	type	range	labels	sources
year	Year	Year of observation	Integer	$\mathbb{N}$		
countries_gwid	Country code (GW)	Unique identifier for each country in the dataset	Integer	$\mathbb{N}$		EPR-AG
ag_id	Aggregate group id	Unique identifier for each aggregate group in the dataset	Integer	$\mathbb{N}$		EPR-AG
gwgroupid	EPR group id	Unique identifier for each EPR group (segment) in the dataset	Integer	$\mathbb{N}$		EPR-AG
groupname	Group name	Name of the EPR group (segment)	String	NA		EPR-AG
onset_do_flag	Civil conflict onset	Binary variable indicating segment-level conflict onset (dropping ongoing conflict years)	Integer	[0;1]	0 = No 1 = Yes . = Missing	EPR-AG
ag_incidence_flag_lag	Conflict, lag	Binary variable indicating ongoing segment-level conflict among other aggregate group members (i.e. group segments in other countries that belong to the same aggregate group)	Integer	[0;1]	0 = No 1 = Yes . = Missing	EPR-AG
ag_all_excl	Exclusion (AG-level)	Binary variable indicating whether all members (segments) of an aggregate group are politically excluded in a given year	Integer	[0;1]	0 = No 1 = Yes . = Missing	EPR-AG
ag_incl	Inclusion (AG-level)	Binary variable indicating whether at least one member (segment) of an aggregate group has political power in a given year.	Integer	[0;1]	0 = No 1 = Yes . = Missing	EPR-AG
status_excl	Exclusion	Binary variable indicating whether group (segment) is excluded from government power in a given year	Integer	[0;1]	0 = No 1 = Yes	EPR-AG
downgraded2	Downgraded	Binary flag indicating whether this group segment has lost political power in the preceding 2 years	Integer	[0;1]	0 = No 1 = Yes	EPR-AG
rbal	Relative size	Group segment's population size compared to other ethnic groups in power	Double Precision	(0,1]		EPR-AG
warhist	War history	Count variable indicating number of conflict onsets this group segment has previously experienced	Integer	$\mathbb{N}$		EPR-AG

ln_rgdppc_lag	GDP, lag, log	Logged real GDP per capita in previous year (country-level)	Double Precision	$\mathbb{R} > 0$		EPR-AG
ln_pop_lag	Population, lag, log	Logged population estimate in previous year (country-level)	Double Precision	$\mathbb{R} > 0$		EPR-AG
ln_state_age	State age, log	Logged number of years since first a country was first recorded as an independent state in the period from 1816 to 2017	Double Precision	$\mathbb{R} >= 0$		EPR-AG
peaceyears	Peace years	Years since group segment first appeared in dataset or since the end of the last conflict episode	Integer	$\mathbb{N}$		EPR-AG
isrelevant	Political relevance	Binary variable indicating whether group (segment) is politically relevant according to EPR definition	Integer	[0;1]	0 = No 1 = Yes	EPR-AG
status_monopoly	Monopoly power status	Binary variable indicating whether group (segment) is the only power holder in its country in a given year	Integer	[0;1]	0 = No 1 = Yes	EPR-AG
status_dominant	Dominant power status (segment)	Binary variable indicating whether group (segment) has dominant power status in a given year	Integer	[0;1]	0 = No 1 = Yes	EPR-AG
seg_area_sqkm	Territory km^2, log (segment)	Logged size of group segment's settlement area in km^2	Double Precision	$\mathbb{R} > 0$		EPR-AG
ln_ag_area_sqkm	Territory km^2, log	Logged size of aggregate group's settlement area in km^2	Double Precision	$\mathbb{R} > 0$		GeoEPR-AG
X	Longitude	Longitude of group segment centroid	Double Precision	$\mathbb{R}$		GeoEPR-AG
Y	Latitude	Latitude of group segment centroid	Double Precision	$\mathbb{R}$		GeoEPR-AG
split	Divided group	Binary variable indicating whether an aggregate group's settlement area is currently split by international borders	Integer	[0;1]	0 = No 1 = Yes . = Missing	GeoEPR-AG, CShapes 2.0
tfrac	Fractionalization	Index that measures the degree of to which an aggregate group's settlement area is currently divided by international borders	Double Precision	[0,1)		GeoEPR-AG, CShapes 2.0

tfrac_incr	Frac. incr. since 1886	Difference between aggregate group's current level of territorial fractionalization and its lowest previous level since 1886	Double Precision	[0,1)		GeoEPR-AG, CShapes 2.0
tfrac_incr_post1946	Frac. incr. since 1946	Difference between aggregate group's current level of territorial fractionalization and its lowest previous level since 1946	Double Precision	[0,1)		GeoEPR-AG, CShapes 2.0
tfrac_incr_pre1946	Frac. incr. before 1946	Difference between aggregate group's level of territorial fractionalization in 1946 and its lowest previous level since 1886	Double Precision	[0,1)		GeoEPR-AG, CShapes 2.0
tfrac_incr_post1918	Frac. incr. since 1918	Difference between aggregate group's current level of territorial fractionalization and its lowest previous level since 1918	Double Precision	[0,1)		GeoEPR-AG, CShapes 2.0
ln_capdist	Distance to capital, log	Yearly distance of group segment centroid to country capital in km, logged	Double Precision	$\mathbb{R} >= 0$		GeoEPR-AG, CShapes 2.0
pfrac	Pop. fractionalization	Index that measures the degree of to which an aggregate group's population is currently divided by international borders	Double Precision	[0,1)		GeoEPR-AG, CShapes 2.0, HYDE Population Database 3.0
pfrac_incr_post1946	Pop. frac. incr. since 1946	Difference between aggregate group's current level of population fractionalization and its lowest previous level since 1946	Double Precision	[0,1)		GeoEPR-AG, CShapes 2.0, HYDE Population Database 3.0
pfrac_incr	Pop. frac. incr. since 1886	Difference between aggregate group's current level of population fractionalization and its lowest previous level since 1886	Double Precision	[0,1)		GeoEPR-AG, CShapes 2.0, HYDE Population Database 3.0
colonial_past	Colonial history	Binary variable indicating whether this group's host state was previously ruled as a colony	Integer	[0;1]	0 = No 1 = Yes	CShapes 2.0
allclaim2	EGIP Claim: Secession / Irredentism	Binary variable indicating whether a self-determination movement or a kin state government has made a claim for secession or irredentism on behalf of the group segment	Integer	[0;1]	0 = No 1 = Yes . = Missing	EPR-EGIP-SDM Claim dataset

allclaim3	EGIP Claim: Irredentism	Binary variable indicating whether a self-determination movement or a kin state government has made a claim for irredentism on behalf of the group segment	Integer	[0;1]	0 = No 1 = Yes . = Missing	EPR-EGIP-SDM Claim dataset

2 eur_map_1918_analysis.dta

This dataset was used to generate the results shown in Table 4 in the main text.

2.1 Sources

- Ethnographic Map of Europe (Gabrys, 1918)
- EPR-AG (Schvitz, 2018; Vogt et al., 2015)
- CShapes 2.0 (Schvitz et al., 2020)
- HYDE Population Database (Klein Goldewijk et al., 2011)

Table 2: Codebook for analysis dataset 'eur_map_1918_analysis.dta'

varname	label	description	type	range	labels	sources
year	Year	Year of observation	Integer	$\mathbb{N}$		
year	Year	Year of observation	Integer	$\mathbb{N}$		
countries_gwid	Country code (GW)	Unique identifier for each country in the dataset	Integer	$\mathbb{N}$		EPR-AG
ag_id	Aggregate group id	Unique identifier for each aggregate group in the dataset	Integer	$\mathbb{N}$		EPR-AG
gwgroupid	EPR group id	Unique identifier for each EPR group (segment) in the dataset	Integer	$\mathbb{N}$		EPR-AG
groupname	Group name	Name of the EPR group (segment)	String	NA		EPR-AG
ag_id	Aggregate group id	Unique identifier for each aggregate group in the dataset	Integer	$\mathbb{N}$		EPR-AG
onset_do_flag	Civil conflict onset	Binary variable indicating segment-level conflict onset (dropping ongoing conflict years)	Integer	[0;1]	0 = No 1 = Yes . = Missing	EPR-AG
ag_incidence_flag_lag	Conflict, lag	Binary variable indicating ongoing segment-level conflict among other aggregate group members (i.e. group segments in other countries that belong to the same aggregate group)	Integer	[0;1]	0 = No 1 = Yes . = Missing	EPR-AG
status_excl	Exclusion	Binary variable indicating whether group (segment) is excluded from government power in a given year	Integer	[0;1]	0 = No 1 = Yes	EPR-AG
status_monopoly	Monopoly power status	Binary variable indicating whether group (segment) is the only power holder in its country in a given year	Integer	[0;1]	0 = No 1 = Yes	EPR-AG
status_dominant	Dominant power status (segment)	Binary variable indicating whether group (segment) has dominant power status in a given year	Integer	[0;1]	0 = No 1 = Yes	EPR-AG
downgraded2	Downgraded	Binary flag indicating whether this group segment has lost political power in the preceding 2 years	Integer	[0;1]	0 = No 1 = Yes	EPR-AG
rbal	Relative size	Group segment's population size compared to other ethnic groups in power	Double Precision	(0,1]		EPR-AG

warhist	War history	Count variable indicating number of conflict onsets this group segment has previously experienced	Integer	$\mathbb{N}$		EPR-AG
ln_rgdppc_lag	GDP, lag, log	Logged real GDP per capita in previous year (country-level)	Double Precision	$\mathbb{R} > 0$		EPR-AG
ln_pop_lag	Population, lag, log	Logged population estimate in previous year (country-level)	Double Precision	$\mathbb{R} > 0$		EPR-AG
ln_state_age	State age, log	Logged number of years since first a country was first recorded as an independent state in the period from 1816 to 2017	Double Precision	$\mathbb{R} >= 0$		EPR-AG
peaceyears	Peace years	Years since group segment first appeared in dataset or since the end of the last conflict episode	Integer	$\mathbb{N}$		EPR-AG
isrelevant	Political relevance	Binary variable indicating whether group (segment) is politically relevant according to EPR definition	Integer	[0;1]	0 = No 1 = Yes	EPR-AG
ln_ag_area_sqkm	Territory km^2, log	Logged size of aggregate group's settlement area in km^2	Double Precision	$\mathbb{R} > 0$		GeoEPR-AG
ln_ag_area_sqkm	Territory km^2, log	Logged size of aggregate group's settlement area in km^2	Double Precision	$\mathbb{R} > 0$		GeoEPR-AG
tfrac	Fractionalization	Index that measures the degree of to which an aggregate group's settlement area is currently divided by international borders	Double Precision	[0,1)		GeoEPR-AG, CShapes 2.0
tfrac_incr	Frac. incr. since 1886	Difference between aggregate group's current level of territorial fractionalization and its lowest previous level since 1886	Double Precision	[0,1)		GeoEPR-AG, CShapes 2.0
ln_capdist	Distance to capital, log	Yearly distance of group segment centroid to country capital in km, logged	Double Precision	$\mathbb{R} >= 0$		GeoEPR-AG, CShapes 2.0
colonial_past	Colonial history	Binary variable indicating whether this group's host state was previously ruled as a colony	Integer	[0;1]	0 = No 1 = Yes	CShapes 2.0

tfrac_hist_incr	Frac. incr. since 1886 (1918 map)	Difference between aggregate group's current level of territorial fractionalization and its lowest previous level since 1886. For all years prior to 1946, we rely on the ethnographic map from 1918. For all years after 1946, we rely on GeoEPR-AG	Double Precision	[0,1)	Ethnographic map of Europe, GeoEPR-AG, Cshapes 2.0

3 murdock_analysis.dta

This dataset was used to generate Table 5 and Figure 9 in the main text.

3.1 Sources

- Murdock's tribal map of Africa (Nunn and Wantchekon, 2011)
- UCDP GED conflict event data (Sundberg and Melander, 2013)
- CShapes 2.0 (Schvitz et al., 2020)
- GTOPO30 Elevation Dataset (U.S. Geological Service, 1996)
- HYDE Historical Population Database (Klein Goldewijk et al., 2011)

Table 3: Codebook for analysis dataset 'murdock_analysis.dta'

varname	label	description	type	range	labels	sources
year	Year	Year of observation	Integer	$\mathbb{N}$		
split	Divided group	Binary variable indicating whether an aggregate group's settlement area is currently split by international borders	Integer	[0;1]	0 = No 1 = Yes	GeoEPR-AG, CShapes 2.0
tfrac	Fractionalization	Index that measures the degree of to which an aggregate group's settlement area is currently divided by international borders	Double Precision	[0,1)		GeoEPR-AG, CShapes 2.0
tfrac_incr	Frac. incr. since 1886	Difference between aggregate group's current level of territorial fractionalization and its lowest previous level since 1886	Double Precision	[0,1)		GeoEPR-AG, CShapes 2.0
name	Group name	Name of the Murdock group	String	NA		Murdock
tribe_code	Group id	Unique identifier for each Murdock group in the dataset	Integer	$\mathbb{N}$		Murdock
area_sqkm_log	Territory km^2, log	Logged size of group's settlement area in km^2	Double Precision	$\mathbb{R} > 0$		Murdock
statebased_os	Civil conflict onset	Binary variable indicating group-level conflict onset. Coded using the first GED event of each UCDP conflict episode in all conflicts that started after January 1st 1989	Integer	[0;1]	0 = No 1 = Yes	Murdock, UCDP GED
ln_capdist	Distance to capital, log	Yearly distance of group segment centroid to country capital in km, logged	Double Precision	$\mathbb{R} > 0$		Murdock, CShapes 2.0
french	French colony	Binary variable indicating whether group's settlement area predominantly falls within former French colony, based on 1989 borders	Integer	[0;1]	0 = No 1 = Yes	Murdock, CShapes 2.0
british	British colony	Binary variable indicating whether group's settlement area mostly lies within former British colony, based on 1989 borders	Integer	[0;1]	0 = No 1 = Yes	Murdock, CShapes 2.0
gwcode	Country code (GW)	Unique identifier for each country in the dataset	Integer	$\mathbb{N}$		Murdock, CShapes 2.0

pop1880_log	Population, log	Spatial estimate of aggregate group's population in 1880	Double Precision	$\mathbb{R} > 0$	Murdock, HYDE Population Database 3.0
elev_sd	Terrain ruggedness	Standard deviation of elevation within group's settlement area	Double Precision	$\mathbb{R} > 0$	Murdock, GTOPO30

4 epr_ag_level_analysis.dta

Used to generate the results shown in Table A16 in the Appendix.

4.1 Sources

- EPR-AG (Schvitz, 2018; Vogt et al., 2015)
- GeoEPR-AG (Schvitz, 2018; Vogt et al., 2015)
- CShapes 2.0 (Schvitz et al., 2020)

Table 4: Codebook for analysis dataset 'epr_ag_level_analysis.dta'

varname	label	description	type	range	labels	sources
year	Year	Year of observation	Integer	$\mathbb{N}$		
ag_id	Aggregate group id	Unique identifier for each aggregate group in the dataset	Integer	$\mathbb{N}$		EPR-AG
groupname	Group name	Name of the EPR group (segment)	String	NA		EPR-AG
ag_onset_ko_flag	Civil conflict onset (AG-level)	Binary variable indicating aggregate group-level conflict onset	Integer	[0;1]	0 = No 1 = Yes	EPR-AG
ag_irrelevant_flag	Irrelevance (AG-level)	Binary variable indicating whether all members (segments) of aggregate group are politically irrelevant in all host states	Integer	[0;1]	0 = No 1 = Yes	EPR-AG
ag_monop_flag	Monopoly power status (AG-level)	Binary variable indicating whether all members (segments) of aggregate group have monopoly power status in all host states	Integer	[0;1]	0 = No 1 = Yes	EPR-AG
ag_dominant_flag	Dominant power status (AG-level)	Binary variable indicating whether all members (segments) of aggregate group have dominant power status in all host states	Integer	[0;1]	0 = No 1 = Yes	EPR-AG
ag_excl_share	Exclusion (AG-level)	Population share of the aggregate group that is politically excluded in a given year	Double Precision	[0,1]		EPR-AG
ag_warhist	War history (AG-level)	Count variable indicating number of conflict onsets this aggregate group has previously experienced	Integer	$\mathbb{N}$		EPR-AG
ag_incidence_flag_lag	Conflict, lag	Binary variable indicating ongoing segment-level conflict among other aggregate group members (i.e. group segments in other countries that belong to the same aggregate group)	Integer	[0;1]	0 = No 1 = Yes	EPR-AG
ag_peaceyears	Peace years (AG-level)	Years since aggregate group first appeared in dataset or since the end of the last conflict episode	Integer	$\mathbb{N}$		EPR-AG
ln_ag_area_sqkm	Territory km^2, log	Logged size of aggregate group's settlement area in km^2	Double Precision	$\mathbb{R} > 0$		GeoEPR-AG

split	Divided group	Binary variable indicating whether an aggregate group's settlement area is currently split by international borders	Integer	[0;1]	0 = No 1 = Yes . = Missing	GeoEPR-AG, CShapes 2.0
tfrac	Fractionalization	Index that measures the degree of to which an aggregate group's settlement area is currently divided by international borders	Double Precision	[0,1)		GeoEPR-AG, CShapes 2.0
tfrac_incr	Frac. incr. since 1886	Difference between aggregate group's current level of territorial fractionalization and its lowest previous level since 1886	Double Precision	[0,1)		GeoEPR-AG, CShapes 2.0
tfrac_incr_post1946	Frac. incr. since 1946	Difference between aggregate group's current level of territorial fractionalization and its lowest previous level since 1946	Double Precision	[0,1)		GeoEPR-AG, CShapes 2.0
tfrac_incr_pre1946	Frac. incr. before 1946	Difference between aggregate group's level of territorial fractionalization in 1946 and its lowest previous level since 1886	Double Precision	[0,1)		GeoEPR-AG, CShapes 2.0

References

Gabrys, Joseph. 1918. *Carte ethnographique de l'Europe: Echelle moyenne 1: 500,000,000*. Librairie centrale des nationalités.

Klein Goldewijk, K, A Beusen, M De Vos and G Van Drecht. 2011. "The HYDE 3.1 Spatially Explicit Database of Human Induced Land Use Change Over the Past 12,000 Years." *Global Ecology and Biogeography* 20(1):73–86.

Nunn, Nathan and Leonard Wantchekon. 2011. "The slave trade and the origins of mistrust in Africa." *American Economic Review* 101(7):3221–52.

Rüegger, Seraina. 2018. Self-Determination Claims by EPR Groups in Power on Behalf of Co-Ethnics Abroad. Technical report International Conflict Research Group.

Schädel, A. 2018. Self-Determination Claims by EPR groups. Technical report International Conflict Research Group.

Schvitz, G, S Rüegger, L Girardin, L-E Cederman, N Weidmann and KS Gleditsch. 2020. "Mapping The International System, 1886-2017: The Cshapes 2.0 Dataset." Working Paper, Online: https://icr.ethz.ch/data/cshapes/. Accessed: 25.9.2020.

Schvitz, Guy. 2018. "Data on Aggregate Ethnic Groups, 1946-2017.". **URL:** *https://icr.ethz.ch/data/epr/ag/*

Sundberg, Ralph and Erik Melander. 2013. "Introducing the UCDP Georeferenced Event Dataset." *Journal of Peace Research* 50(4):523–532.

U.S. Geological Service. 1996. "Digital Elevation - Global 30 Arc-Second Elevation (GTOPO30).". **URL:** *https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-global-30-arc-second-elevation-gtopo30?qt-science_center_objects=0#qt-science_center_objects*

Vogt, Manuel, Nils-Christian Bormann, Seraina Rüegger, Lars-Erik Cederman, Philipp Hunziker and Luc Girardin. 2015. "Integrating Data on Ethnicity, Geography, and Conflict: The Ethnic Power Relations Dataset Family." *Journal of Conflict Resolution* 59(7):1327–1342.