GROW^{up} Research Front-End Documentation
RFE Release 1.1

Philipp Hunziker
ETH Zurich [1]
http://growup.ethz.ch/rfe

July 1, 2013

[1]For correspondence, please use growup@gess.ethz.ch

Contents

1 Group-Level Data **5**

1.1 The EPR Universe 5

1.1.1 Country Selection 5

1.1.2 Group Selection 6

1.2 Group Activity, Relevancy, and Hierarchies 6

1.2.1 Group Activity 6

1.2.2 Group Relevance 7

1.2.3 Group Hierarchies 7

1.3 Ethnic Conflict Data 8

1.3.1 ACD2EPR 9

1.3.2 NSA 9

1.3.3 ACD 10

1.3.4 Conflict Onset 11

1.3.5 KO and DO Options 11

1.3.6 Conflict Incidence 12

1.4 Geographical Data 13

1.5 Variables 13

1.5.1 CountryName 13

1.5.2 Countries_Cow 13

1.5.3 GroupName 14

1.5.4 CowGroupID 14

1.5.5 isActive 14

1.5.6 isRelevant 14

1.5.7 Onset_KO_Flag 15

1.5.8 Onset_DO_Flag 15

1.5.9 Onset_Count 15

1.5.10 Onset_UCDPID_a 15

1.5.11 Onset_KO_Terr_Flag 16

1.5.12 Onset_DO_Terr_Flag 16

1.5.13 Onset_Terr_UCDPID_a 16

1.5.14 Onset_KO_Gov_Flag . . . 17
1.5.15 Onset_DO_Gov_Flag . . . 17
1.5.16 Onset_Gov_UCDPID_a . . . 17
1.5.17 Onset_MaxIntens . . . 17
1.5.18 Onset_MinIntens . . . 18
1.5.19 Onset_Intens . . . 18
1.5.20 Onset_HighIntens_Flag . . . 18
1.5.21 Incidence_Flag . . . 18
1.5.22 Incidence_Count . . . 19
1.5.23 Incidence_UCDPID_a . . . 19
1.5.24 Incidence_Terr_Flag . . . 19
1.5.25 Incidence_Terr_UCDPID_a . . . 19
1.5.26 Incidence_Gov_Flag . . . 20
1.5.27 Incidence_Gov_UCDPID_a . . . 20
1.5.28 Incidence_MaxIntens . . . 20
1.5.29 Incidence_MinIntens . . . 20
1.5.30 Incidence_HighIntens_Flag . . . 21
1.5.31 PeaceYears . . . 21
1.5.32 WarHist . . . 21
1.5.33 WarHist_HighIntens . . . 21
1.5.34 HasActiveAnc . . . 22
1.5.35 ActiveAnc_GroupID . . . 22
1.5.36 HasActiveDesc . . . 22
1.5.37 ActiveDesc_GroupID_a . . . 22
1.5.38 StatusID . . . 23
1.5.39 StatusName . . . 23
1.5.40 Status_Egip . . . 23
1.5.41 Status_Excl . . . 24
1.5.42 Status_Monopoly - Status_Discrim . . . 24
1.5.43 Status_Pwrrank . . . 24
1.5.44 Egip_Groups_Count . . . 24
1.5.45 Excl_Groups_Count . . . 25
1.5.46 Rlvt_Groups_Count . . . 25
1.5.47 Actv_Groups_Count . . . 25
1.5.48 GroupSize . . . 25
1.5.49 lPop . . . 26
1.5.50 lSize . . . 26
1.5.51 EgipPop . . . 26
1.5.52 lEgipPop . . . 27
1.5.53 ExclPop . . . 27
1.5.54 lExclPop . . . 27

1.5.55 rbal . . . 28
1.5.56 rlbal . . . 28
1.5.57 Downgraded1 - Downgraded10 . . . 28
1.5.58 Downgraded_Excl1 - Downgraded_Excl10 . . . 29
1.5.59 Downgraded_Hist . . . 29
1.5.60 Downgraded_Excl_Hist . . . 29
1.5.61 isUrban . . . 29
1.5.62 isDispersed . . . 30
1.5.63 isMigrant . . . 30
1.5.64 hasSetArea . . . 30

2 Country-Level Data **31**
2.1 EPR-ETH Countries . . . 31
2.2 Country-Level Conflict Data . . . 32
2.2.1 ACD Aggregation . . . 32
2.2.2 Conflict Onset . . . 33
2.2.3 KO and DO Options . . . 33
2.2.4 Conflict Incidence . . . 34
2.3 Country-Level Information on Ethnic Groups . . . 34
2.4 Variables . . . 35
2.4.1 CountryName . . . 35
2.4.2 Countries_Cow . . . 35
2.4.3 All_Onset_KO_Flag . . . 35
2.4.4 All_Onset_DO_Flag . . . 35
2.4.5 All_Onset_UCDPID_a . . . 36
2.4.6 All_Onset_Count . . . 36
2.4.7 All_Onset_KO_Terr_Flag . . . 36
2.4.8 All_Onset_DO_Terr_Flag . . . 36
2.4.9 All_Onset_KO_Gov_Flag . . . 37
2.4.10 All_Onset_DO_Gov_Flag . . . 37
2.4.11 All_Onset_KO_Eth_Flag . . . 37
2.4.12 All_Onset_DO_Eth_Flag . . . 37
2.4.13 All_Onset_KO_NonEth_Flag . . . 38
2.4.14 All_Onset_DO_NonEth_Flag . . . 38
2.4.15 All_Onset_KO_Eth_Terr_Flag . . . 38
2.4.16 All_Onset_DO_Eth_Terr_Flag . . . 38
2.4.17 All_Onset_KO_Eth_Gov_Flag . . . 39
2.4.18 All_Onset_DO_Eth_Gov_Flag . . . 39
2.4.19 All_Onset_KO_NonEth_Terr_Flag . . . 39
2.4.20 All_Onset_DO_NonEth_Terr_Flag . . . 40
2.4.21 All_Onset_KO_NonEth_Gov_Flag . . . 40

2.4.22 All_Onset_DO_NonEth_Gov_Flag 40
2.4.23 All_Onset_Eth_Terr_Count 41
2.4.24 All_Onset_Eth_Gov_Count 41
2.4.25 All_Onset_NonEth_Terr_Count 41
2.4.26 All_Onset_NonEth_Gov_Count 41
2.4.27 All_Incidence_Flag 42
2.4.28 All_Incidence_UCDPID_a 42
2.4.29 All_Incidence_Count 42
2.4.30 All_Incidence_Terr_Flag 42
2.4.31 All_Incidence_Gov_Flag 42
2.4.32 All_Incidence_Eth_Flag 43
2.4.33 All_Incidence_NonEth_Flag 43
2.4.34 All_Incidence_Eth_Terr_Flag 43
2.4.35 All_Incidence_Eth_Gov_Flag 43
2.4.36 All_Incidence_NonEth_Terr_Flag 44
2.4.37 All_Incidence_NonEth_Gov_Flag 44
2.4.38 All_Incidence_Eth_Terr_Count 44
2.4.39 All_Incidence_Eth_Gov_Count 44
2.4.40 All_Incidence_NonEth_Terr_Count 45
2.4.41 All_Incidence_NonEth_Gov_Count 45
2.4.42 All_PeaceYears 45
2.4.43 All_WarHist 45
2.4.44 Egip_Groups_Count 46
2.4.45 Excl_Groups_Count 46
2.4.46 Rlvt_Groups_Count 46
2.4.47 Actv_Groups_Count 46
2.4.48 lPop 46
2.4.49 EgipPop 47
2.4.50 lEgipPop 47
2.4.51 ExclPop 47
2.4.52 lExclPop 47
2.4.53 maxExclPop 48
2.4.54 lmaxExclPop 48
2.4.55 Cntr_Relevance 48
2.4.56 Nstar 48

3 Appendices **49**
3.1 Terms and Conditions 49
3.2 Sources Release 1.1 49
3.3 Changelog 50
3.3.1 Version 1.0 to Version 1.1 50

Chapter 1

Group-Level Data

1.1 The EPR Universe

The sample universe of ethnic groups in the RFE group-level data is adopted from the EPR-ETH (*Ethnic Power Relations*) dataset (Cederman et al. 2010, for more detailed information concerning versions and sources see 3.2). Ethnic groups are included in the EPR-ETH dataset according to a two-step coding procedure:

1.1.1 Country Selection

Firstly, the EPR-ETH dataset defines a time-variant list of countries for which ethnic groups are coded. Specifically, the EPR-ETH dataset covers all countries in the period 1946 - 2009 that meet the following criteria:

(i) Administered by an intact sovereign state, i.e. colonies and failed states are not included [1].

(ii) Population is greater than or equal to 500'000 inhabitants.

Newly independent states are included in the dataset beginning with the year of independence, e.g. Croatia (independence on 25/06/1991) is included from 1991 onwards. Given these criteria, EPR-ETH covers a total of 7'988 country-years from 162 countries.

[1] In fact, countries considered "administered" by a failed state *are* included in the dataset if the period during which they are coded as failed states falls in between periods when the country meets the EPR-ETH inclusion criteria. During these periods, however, the coding on the ethnic group level (see below) is not continued.

1.1.2 Group Selection

Secondly, ethnic groups are coded on the basis of this list of country-years. For this purpose, EPR-ETH defines ethnicity as any subjectively experienced sense of commonality based on the belief in common ancestry and shared culture. Given this definition, an ethnic group (i.e. a group of individuals sharing a common ethnicity) is included in the EPR-ETH dataset if it is politically relevant at least once in the sample period. An ethnic group is classified as politically relevant if at least one political organization claims to represent it in national politics or if its members are subjected to state-led political discrimination.

With these two selection procedures in place, the EPR-ETH dataset identifies 790 politically relevant ethnic groups (before considering hierarchies, see 1.2.3) in 137 countries across the globe for the period 1946 - 2009. For the remaining 25 countries, no ethnic group is coded as politically relevant for the entire sample period. For these cases, so-called *placeholder groups* are defined which are coded as politically irrelevant for the entire sample period (e.g. Germans in Germany).

1.2 Group Activity, Relevancy, and Hierarchies

The unit of observation in the RFE group-level data is the ethnic-group-year. An EPR-ETH ethnic group is included in the RFE group-level data for all years during which the group's host state meets the EPR-ETH inclusion criteria (see 1.1). Consequently, researchers should note that ethnic groups are included in the RFE group-level data even if they are currently inactive or irrelevant, two concepts that are discussed subsequently.

1.2.1 Group Activity

An ethnic group is deemed active in a given group-year if it is currently physically present in a country *and* is not currently represented by an active ancestor or descendant (see below). Physical presence may be altered through border changes or large migration flows. In the RFE group-level data, whether a group is active during a given group-year is indicated by the *isActive* variable (see 1.5.5). The RFE group-level data covers a total of 39'642 active group-years.

1.2.2 Group Relevance

An ethnic groups is deemed relevant in a given group-year if at least one political organization claims to represent it in national politics or if its members are subjected to state-led political discrimination. Due to the EPR-ETH inclusion criteria (see above), all groups in the RFE group-level data except placeholder groups are coded as relevant at least once during the sample period. Group relevance and activity are nested concepts, i.e. a group can only be relevant if it is active. In the RFE group-level data, whether a group is relevant during a given group-year is indicated by the *isRelevant* variable (see 1.5.6). The RFE group-level data covers a total of 33'630 relevant group-years.

1.2.3 Group Hierarchies

Building on the EPR-ETH dataset, the RFE group-level data tracks hierarchical group relationships over time. There are two kinds of hierarchical transformations that an ethnic group may experience:

Split An ethnic group may split into n ($n \geq 2$) smaller groups. This is the case if the politically relevant concept of ethnicity is redefined over time and individuals that have formerly been considered part of the same ethnic group are now considered members of distinct groups.

Unification Several ethnic groups may unify into one larger group. This is the case if individuals that have formerly been considered part of different ethnic groups are now considered members of the same politically relevant ethnic group.

If either of these hierarchical transformations occur, the RFE group-level data defines a hierarchical relationship between the groups involved. The hierarchically superior group in these transformations (i.e. the group defined by the more inclusive definition of ethnicity) is called the *ancestor group*, whereas the hierarchically inferior groups are called *descendant groups*. Consequently, a split is a transformation from one ancestor group into several descendant groups, and a unification is a transformation from several descendant groups into one ancestor groups.

Hierarchical transformations are mirrored in the RFE group-level data in three different ways:

- Hierarchical transformations affect **group activity**: In the case of a split, the descendant groups are coded as inactive in the group-years prior to the split, whereas the ancestor group is coded as inactive in the group-years after the split. Analogously, in the case of a unification, the ancestor group is coded as inactive prior to the unification, and the descendant groups as inactive after the unification. This is to avoid that the same individuals in a country are double-counted as members of several different hierarchically related ethnic groups at the same time.

- Hierarchical transformations are tracked directly by at least two variables: **hasActiveAncestor** and **hasActiveDescendants**, which indicate whether a currently inactive group is represented by either a hierarchically superior group (an ancestor) or a hierarchically inferior group (a descendant). Furthermore, the variables ActiveAnc_GroupID and ActiveDesc_GroupID_a track the CowGroupIDs of a group's currently active ancestor or descendants. For more information on these variables see the respective subsections in 1.5.

- Hierarchical relationships are reflected in variables of the **family type**: The *family* prefix indicates that a variable applies to this group, its active ancestor group (if any), or its active descendant groups (if any). For example, the *Family_Incindence_Flag* variable indicates whether a group, its active ancestor group, or its active descendant group is involved in an ACD conflict episode in the given year. *family* type variables are useful for tracking phenomena that are a function of time while considering that members of a given ethnic group may have been members of another hierarchically related ethnic group in the past.

1.3 Ethnic Conflict Data

The information on ethnic conflicts in the RFE group-level data is compiled from three different sources: The ACD2EPR dataset (Wucherpfennig et al. 2012), the NSA dataset (Cunningham et al. 2009), and the ACD dataset (Gleditsch et al. 2002). For more information on these datasets and the exact versions used, please see 3.2. These sources allow the identification of ethnic conflicts and their mapping onto EPR-ETH groups in three steps:

- The ACD2EPR dataset links NSA rebel organizations to EPR-ETH groups.

- The NSA dataset provides information on rebel organizations involved in ACD conflicts.

- The ACD dataset provides information on the occurrence, duration, and intensity of intrastate conflicts.

The remainder of this section briefly describes the structure of these data sources and establishes how the latter are used to code the conflict related variables in the RFE group-level data.

1.3.1 ACD2EPR

The units of observation in the ACD2EPR dataset are rebel-organization/ethnic-group dyads. For every rebel organization included in the NSA dataset (see 1.3.2), ACD2EPR reports whether the organization can be linked to any EPR-ETH group in the same country through ethnic claims or significant ethnic recruitment[2]. More specifically, for the RFE group-level conflict data, a link between a rebel organization and an ethnic group is coded if the ACD2EPR *claim* variable indicates an explicit or implicit ethnic claim by the rebel organization ($claim >= 1$), and the *recruitment* variable indicates that the rebel organization has recruited from the respective ethnic group significantly ($recruitment = 1$). Note that the ACD2EPR coding is time invariant. Thus, links between rebel organizations and ethnic groups exists for the entire lifespan of a rebel organization[3]. The RFE coding rule is that for a definite link between an ethnic group and a rebel organization to be established, the two must be connected through both *claim and recruitment*. With this coding rule in place, we are able to produce a time invariant one-to-many table connecting EPR-ETH groups one or multiple rebel organizations (if any).

1.3.2 NSA

The NSA dataset reports detailed information on all rebel organizations involved in conflicts coded by the ACD dataset (see 1.3.3), and provides the initial list of rebel organizations for which the ACD2EPR dataset codes links to EPR-ETH groups. The NSA dataset is time variant and organized in

[2]The original ACD2EPR dataset also establishes links between rebel organizations and ethnic groups among other dimensions, but these are not relevant for the RFE group-level data conflict coding. Also, the ACD2EPR dataset does not code links between rebel organizations and EPR-ETH groups if the rebel organization is involved in a colonial conflict (as coded by the NSA dataset), since this type of conflict is not easily compatible with the inclusion criteria for countries into the EPR-ETH universe.

[3]"Lifespan" refers to the period(s) during which a rebel organization is included in the NSA dataset (see 1.3.2).

observations of dyad-periods: Dyads are unique combinations of a rebel-organization and its adversaries (one or more states) in ACD conflicts. The coding and identification of dyads is adopted directly from the ACD Actor Dataset (Uppsala Conflict Data Program 2010). Dyads enter the dataset on the date of the dyad's members first invovlement in an ACD conflict, and leave the dataset on the date of their last involvement in an ACD conflict. If attributes of the non-state side of the dyad (the rebel organization) change throughout this period, the NSA dataset captures these changes by defining a new dyad-period. Since by design the dyads in the NSA dataset are compatible with the ACD dataset, each dyad carries a *UCDP ID* that allows to link the rebel organization in the dyad to a ACD conflict. Consequently, combining the EPR-group-to-rebel-organization table derived above with the NSA dataset, we are able to create a new table indicating whether EPR-ETH groups are involved in one or more ACD conflicts through links with one or more rebel organizations.

1.3.3 ACD

The ACD dataset codes inter- and intrastate armed conflicts on a yearly basis. Conflicts in the ACD dataset follow the UCDP conflict definition (please see 3.2 for links to more information and references) and are assigned a unique *UCDP ID*. ACD conflicts are assigned to one of four categories: Extrasystemic, Interstate, Internal, and Internationalized Internal Conflicts. Of these types, only the latter two are relevant for the RFE group-level conflict coding, since we focus exclusively on ethnic civil wars[4].

ACD conflicts are distinguished along the incompatibility they originate from, not their temporal dimension. Hence, a conflict that ends and reoccurs after a substantial period of time is assigned the same UCDP ID, regardless of the time period separating the actual conflict episodes, as long as the incompatibility remains the same. The ACD dataset provides information on the temporal dimension of conflict termination and recurrence with a variable indicating the end of a conflict episode, whereas according to the ACD coding rules, a conflict episode ends in a given year if there is no conflict-related activity in the next calendar year[5]. For the purpose of the RFE group-level data conflict coding, this definition is altered: For all conflict onset and incidence variables in the RFE group-level dataset, a conflict episode is only

[4]Also, since the NSA dataset (see 1.3.2) only covers ACD dyads where at least one conflict party is a non-state actor, the mapping of EPR-ETH groups to ACD conflicts is only possible for Internal and Internationalized Internal Conflicts.

[5]"Conflict-related activity" refers to the UCDP threshold of at least 25 battledeaths per annum.

considered terminated if there is *no conflict-related activity in the following two calender years*. This recoding of ACD conflict episode is adopted in order to prevent an inflation of conflict episodes (and thus conflict onsets, see 1.3.4) in low-activity conflicts that reoccur repeatedly. With this 2-year episode coding rule in place, we are able to merge the table mapping EPR-ETH groups onto ACD conflicts discussed above (see 1.3.2) to the modified ACD conflict table, and thus create a time-variant dataset indicating whether and when EPR-ETH groups are involved in ACD conflict episodes through links with one or several rebel organizations.

1.3.4 Conflict Onset

The RFE group-level data follows a standard conflict onset coding based on the mapping of conflicts onto ethnic groups discussed above. Specifically, the coding rule imposed is that ethnic group G experiences conflict onset in year t if

(i) a rebel organization linked to G enters an ACD conflict episode active in year t,

and

(ii) no rebel organization linked to G has already entered the same ACD conflict episode prior to year t.

Element (ii) of this coding rule ensures that an ethnic group can only "join" an ACD conflict episode once. In effect, there are two scenarios where this addendum affects the RFE group-level onset variables: Firstly, element (ii) ensures that if a rebel organization linked to group G enters the same ACD conflict episode recurrently, while fighting between the government and another rebel organization over the same incompatibility continues, only one conflict onset for group G is coded. Secondly, element (ii) ensures that if multiple rebel organizations are linked to group G, and the latter enter the same ACD conflict episode in different years, only the temporally first entrance is coded as conflict onset for group G.

1.3.5 KO and DO Options

The RFE group-level data conflict onset variables come in two variants: The KO and the DO option. KO stands for *Keep Ongoing* and is the default option. Conflict onset variables with the *KO* option take the value of 1 for

group-years in which a group experiences conflict onset (see 1.3.4), and 0 in all other years.

DO stands for *Drop Ongoing* and represents a slightly more advanced coding. DO conflict onset variables indicate conflict onset in the same manner as the KO variables, but are censored in the years subsequent to conflict onset, during which the respective ACD conflict episode is ongoing. Note that this implies that DO onset variables are censored for the entire duration of an ACD conflict episode after group-level onset, even if the rebel organization linking the group to this conflict episode terminates fighting prior to the end of the episode. The intuition behind the DO coding is to remove group-years from conflict onset variables for which the RFE group-level onset coding rules (see 1.3.4) make positive values on the onset variable highly unlikely[6], as groups cannot join the same ACD conflict episode twice.

1.3.6 Conflict Incidence

The RFE group-level data includes a number of conflict incidence variables, which indicate whether in a given year a group is involved in an ACD conflict episode through links with one or several rebel organizations. Conflict incidence variables assume the value of 1 in years when at least one rebel organization linked to an ethnic group is involved in an ACD conflict episode, and 0 otherwise.

Please note that because of the 2-year episode coding rule imposed on the ACD data, conflict episodes may terminate for a single calendar year and then continue without a new group-level conflict onset occuring (see 1.3.3 and 1.3.4).

Please also note that the RFE group-level incidence variables do not fully match the censored values on the respective DO type onset variables, since conflict incidence variables represent actual involvement, whereas DO onset variables are censored for the entire duration of an ACD conflict episode. Specifically, if a rebel organization linked to group G leaves an ongoing ACD conflict episode in year t, the respective conflict incidence variable assumes the value 0 in year $t + 1$, whereas the according DO type onset variable remains censored until the entire ACD conflict episode terminates.

[6]We write "highly unlikely", and not "impossible", because there are configurations where the RFE group-level coding rules would allow onsets to occur, but these are censored by the DO coding. Specifically, this is the case if a group joins an ACD conflict episode while it is still involved in an entirely different ACD conflict (i.e. a conflict over a different incompatibility). However, this is an extremely rare event and occurs only twice in the entire RFE group-level dataset.

1.4 Geographical Data

In the current release, the RFE group-level data only provides settlement pattern information (see1.5.61 to 1.5.64). The respective variables are coded from the GeoEPR-ETH dataset (Wucherpfennig et al. 2011, for more information see 3.2), which provides information on the type of group settlement patterns and matching GIS polygons. Since the GeoEPR-ETH dataset only codes settlement patterns for relevant ethnic groups (see 1.2.2), all geographical variables offered in the RFE group-level data are coded as missing for irrelevant ethnic groups.

1.5 Variables

This section provides a full list of variables included in the RFE group-level data, together with information on substantial content, sources, and coding rules. Please note that knowledge of the underlying data structure as described in sections 1.1 to 1.4 is essential for fully understanding some of the information offered in this section.

1.5.1 CountryName

Type:	Text
Value Range:	NA
Description:	Country name, follows COW name wherever possible.
Family Type:	No
Sources:	COW State System Membership List

1.5.2 Countries_Cow

Type:	Integer
Value Range:	$[2, 920]$
Description:	Country's COW ID, see Correlates of War Project (2008).
Family Type:	No
Sources:	COW State System Membership List

1.5.3 GroupName

Type:	Text
Value Range:	NA
Description:	Ethnic group's name.
Family Type:	No
Sources:	EPR-ETH

1.5.4 CowGroupID

Type:	Integer
Value Range:	$[201000, 92004000]$
Description:	Unique ID for each ethnic group. Composed as $(CowID * 100000) + GroupID$, whereas GroupID is a 4-5 digit ID. GroupIDs represent hierarchies through nesting on decimal places (see 1.2.3).
Family Type:	No
Sources:	EPR-ETH

1.5.5 isActive

Type:	Integer
Value Range:	$[0; 1]$
Description:	Binary flag indicating whether group is physically present in a country and not represented by active ancestor or descendant groups (see 1.2.1).
Family Type:	No
Sources:	EPR-ETH

1.5.6 isRelevant

Type:	Integer
Value Range:	$[0; 1]$
Description:	Binary flag indicating whether group is politically relevant (see 1.2.2).
Family Type:	No
Sources:	EPR-ETH

1.5.7 Onset_KO_Flag

Type:	Integer
Value Range:	$[0; 1]$
Description:	Binary flag indicating group-level onset, KO option (see 1.3.4, 1.3.5).
Family Type:	Yes (see 1.2.3)
Sources:	ACD2EPR, NSA, ACD

1.5.8 Onset_DO_Flag

Type:	Integer
Value Range:	$[0; 1]$
Description:	Binary flag indicating group-level onset, DO option (see 1.3.4, 1.3.5).
Family Type:	Yes (see 1.2.3)
Sources:	ACD2EPR, NSA, ACD

1.5.9 Onset_Count

Type:	Integer
Value Range:	$\mathbb{N}$
Description:	Count variable indicating number of conflict onsets this group experiences in this year; based on KO coding rule (see 1.3.4, 1.3.5).
Family Type:	Yes (see 1.2.3)
Sources:	ACD2EPR, NSA, ACD

1.5.10 Onset_UCDPID_a

Type:	Integer Array {Int, Int, ...}
Value Range:	NA
Description:	Array variable indicating UCDP IDs of conflict onsets this group experiences in this year; based on KO coding rule (see 1.3.4, 1.3.5).
Family Type:	Yes (see 1.2.3)
Sources:	ACD2EPR, NSA, ACD

1.5.11 Onset_KO_Terr_Flag

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating group-level onset for ACD conflicts with territorial incompatibility (according to ACD's Incomp variable), KO option (see 1.3.4, 1.3.5).
Family Type:	Yes (see 1.2.3)
Sources:	ACD2EPR, NSA, ACD

1.5.12 Onset_DO_Terr_Flag

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating group-level onset for ACD conflicts with territorial incompatibility (according to ACD's Incomp variable), DO option (see 1.3.4, 1.3.5). DO coding rules only enforced for conflicts with appropriate incompatibility, i.e. variable censored only for ongoing territorial ACD episodes.
Family Type:	Yes (see 1.2.3)
Sources:	ACD2EPR, NSA, ACD

1.5.13 Onset_Terr_UCDPID_a

Type:	Integer Array {Int, Int, ...}
Value Range:	NA
Description:	Array variable indicating UCDP IDs of territorial conflict onsets (according to ACD's Incomp variable) this group experiences in this year; based on KO coding rule (see 1.3.4, 1.3.5).
Family Type:	Yes (see 1.2.3)
Sources:	ACD2EPR, NSA, ACD

1.5.14 Onset_KO_Gov_Flag

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating group-level onset for ACD conflicts with governmental incompatibility (according to ACD's Incomp variable), KO option (see 1.3.4, 1.3.5).
Family Type:	Yes (see 1.2.3)
Sources:	ACD2EPR, NSA, ACD

1.5.15 Onset_DO_Gov_Flag

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating group-level onset for ACD conflicts with governmental incompatibility (according to ACD's Incomp variable), DO option (see 1.3.4, 1.3.5). DO coding rules only enforced for conflicts with appropriate incompatibility, i.e. variable censored only for ongoing governmental ACD episodes.
Family Type:	Yes (see 1.2.3)
Sources:	ACD2EPR, NSA, ACD

1.5.16 Onset_Gov_UCDPID_a

Type:	Integer Array {Int, Int, ...}
Value Range:	NA
Description:	Array variable indicating UCDP IDs of governmental conflict onsets (according to ACD's Incomp variable) this group experiences in this year; based on KO coding rule (see 1.3.4, 1.3.5).
Family Type:	Yes (see 1.2.3)
Sources:	ACD2EPR, NSA, ACD

1.5.17 Onset_MaxIntens

Type:	Integer
Value Range:	[1; 2]
Description:	Maximum intensity (according to ACD's Intens variable) the ACD conflict episodes this group joins in this year ever reach during the period of this group's involvement.
Family Type:	Yes (see 1.2.3)
Sources:	ACD2EPR, NSA, ACD

1.5.18 Onset_MinIntens

Type:	Integer
Value Range:	[1; 2]
Description:	Minimum intensity (according to ACD's Intens variable) the ACD conflict episodes this group joins in this year ever reach during the period of this group's involvement.
Family Type:	Yes (see 1.2.3)
Sources:	ACD2EPR, NSA, ACD

1.5.19 Onset_Intens

Type:	Integer
Value Range:	[1; 2]
Description:	Current intensity (according to ACD's Intens variable) of the ACD conflict episodes this group joins in this year. Maximum value is used if this group joins several ACD conflict episodes in this year.
Family Type:	Yes (see 1.2.3)
Sources:	ACD2EPR, NSA, ACD

1.5.20 Onset_HighIntens_Flag

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating whether any of the onsets this group experiences in this year correspond to high intensity ACD conflict episodes (i.e. the ACD conflict episode reaches high intensity on ACD's Intens variable in at least one year).
Family Type:	Yes (see 1.2.3)
Sources:	ACD2EPR, NSA, ACD

1.5.21 Incidence_Flag

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating group involvement in an ACD conflict episode in the given year (see 1.3.6).
Family Type:	Yes (see 1.2.3)
Sources:	ACD2EPR, NSA, ACD

1.5.22 Incidence_Count

Type:	Integer
Value Range:	$\mathbb{N}$
Description:	Count variable indicating number of ACD conflict episodes this group is involved in in this year (see 1.3.6).
Family Type:	Yes (see 1.2.3)
Sources:	ACD2EPR, NSA, ACD

1.5.23 Incidence_UCDPID_a

Type:	Integer Array {Int, Int, ...}
Value Range:	NA
Description:	Array variable indicating UCDP IDs of ACD conflict episodes this group is involved in in this year (see 1.3.6).
Family Type:	Yes (see 1.2.3)
Sources:	ACD2EPR, NSA, ACD

1.5.24 Incidence_Terr_Flag

Type:	Integer
Value Range:	$[0; 1]$
Description:	Binary flag indicating group involvement in at least one ACD conflict with territorial incompatibility (according to ACD's Incomp variable) (see 1.3.6).
Family Type:	Yes (see 1.2.3)
Sources:	ACD2EPR, NSA, ACD

1.5.25 Incidence_Terr_UCDPID_a

Type:	Integer Array {Int, Int, ...}
Value Range:	NA
Description:	Array variable indicating UCDP IDs of territorial ACD conflict episodes (according to ACD's Incomp variable) this group is involved in in this year (see 1.3.6).
Family Type:	Yes (see 1.2.3)
Sources:	ACD2EPR, NSA, ACD

1.5.26 Incidence_Gov_Flag

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating group involvement in at least one ACD conflict with governmental incompatibility (according to ACD's Incomp variable) (see 1.3.6).
Family Type:	Yes (see 1.2.3)
Sources:	ACD2EPR, NSA, ACD

1.5.27 Incidence_Gov_UCDPID_a

Type:	Integer Array {Int, Int, ...}
Value Range:	NA
Description:	Array variable indicating UCDP IDs of governmental ACD conflict episodes (according to ACD's Incomp variable) this group is involved in in this year (see 1.3.6).
Family Type:	Yes (see 1.2.3)
Sources:	ACD2EPR, NSA, ACD

1.5.28 Incidence_MaxIntens

Type:	Integer
Value Range:	[1; 2]
Description:	Intensity (according to ACD's Intens variable) of the ACD conflict episodes this group is involved in in this year. Maximum value is chosen in case this group is involved in more than one ACD conflict episodes.
Family Type:	Yes (see 1.2.3)
Sources:	ACD2EPR, NSA, ACD

1.5.29 Incidence_MinIntens

Type:	Integer
Value Range:	[1; 2]
Description:	Intensity (according to ACD's Intens variable) of the ACD conflict episodes this group is involved in in this year. Minimum value is chosen in case this group is involved in more than one ACD conflict episodes.
Family Type:	Yes (see 1.2.3)
Sources:	ACD2EPR, NSA, ACD

1.5.30 Incidence_HighIntens_Flag

Type:	Integer
Value Range:	$[0; 1]$
Description:	Binary flag indicating whether any of the ACD conflict episodes this group is involved in in this year is of high intensity (according to ACD's Intens variable).
Family Type:	Yes (see 1.2.3)
Sources:	ACD2EPR, NSA, ACD

1.5.31 PeaceYears

Type:	Integer
Value Range:	$\mathbb{N}$
Description:	Count variable indicating the number of calendar years since this group's last involvement in an ACD conflict episode. In case a group has never been involved in an ACD conflict episode, the variable counts the number of calendar years since 1946 or the year this group's host state is first included in the EPR-ETH dataset. Based on Incidence_Flag (see1.5.21).
Family Type:	Yes (see 1.2.3)
Sources:	ACD2EPR, NSA, ACD

1.5.32 WarHist

Type:	Integer
Value Range:	$\mathbb{N}$
Description:	Count variable indicating the number of onsets this group has experienced in the years prior to this year. Based on Onset_KO_Flag (see 1.5.7).
Family Type:	Yes (see 1.2.3)
Sources:	ACD2EPR, NSA, ACD

1.5.33 WarHist_HighIntens

Type:	Integer
Value Range:	$\mathbb{N}$
Description:	Count variable indicating the number of high intensity onsets this group has experienced in the years prior to this year. Based on Onset_HighIntens_Flag (see 1.5.20).
Family Type:	Yes (see 1.2.3)
Sources:	ACD2EPR, NSA, ACD

1.5.34 HasActiveAnc

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating whether this group has a currently active ancestor group in this year (see 1.2.3). A positive value on this variable implies that this group is inactive (see 1.5.5).
Family Type:	No
Sources:	EPR-ETH

1.5.35 ActiveAnc_GroupID

Type:	Integer
Value Range:	[201000, 92004000]
Description:	CowGroupID (see 1.5.4) of this group's active ancestor ancestor group (if applicable). A non-missing value on this variable implies that this group is inactive (see 1.5.5).
Family Type:	No
Sources:	EPR-ETH

1.5.36 HasActiveDesc

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating whether this group has currently active descendant groups in this year (see 1.2.3). A positive value on this variable implies that this group is inactive (see 1.5.5).
Family Type:	No
Sources:	EPR-ETH

1.5.37 ActiveDesc_GroupID_a

Type:	Integer Array {Int, Int, ...}
Value Range:	NA
Description:	Array of CowGroupIDs (see 1.5.4) of this group's active descendant groups (if applicable). A non-missing value on this variable implies that this group is inactive (see 1.5.5).
Family Type:	No
Sources:	EPR-ETH

1.5.38 StatusID

Type:	Integer
Value Range:	$[1, 10]$
Description:	Unique ID indicating this group's political status (according to EPR-ETH) in this year: 1 Monopoly 2 Dominant 3 Senior Partner 4 Junior Partner 5 Regional Autonomy 6 Separatist Autonomy 7 Powerless 8 Discriminated 9 State Collapse 10 Irrelevant Note that this is a nominal variable, i.e. IDs do not imply ordinal ranking.
Family Type:	No
Sources:	EPR-ETH

1.5.39 StatusName

Type:	String
Value Range:	NA
Description:	Name of the corresponding value on StatusID (see 1.5.38).
Family Type:	No
Sources:	EPR-ETH

1.5.40 Status_Egip

Type:	Integer
Value Range:	$[0; 1]$
Description:	Binary flag indicating whether this group is an EGIP (Ethnic Group in Power) in this year, whereas groups are EGIP if StatusID $\in \{2, 3, 4, 9\}$ (see 1.5.38).
Family Type:	No
Sources:	EPR-ETH

1.5.41 Status_Excl

Type:	Integer
Value Range:	$[0; 1]$
Description:	Binary flag indicating whether this group is excluded (Ethnic Group in Power) in this year, whereas groups are excluded if StatusID $\in \{1, 5, 6, 7\}$ (see 1.5.38).
Family Type:	No
Sources:	EPR-ETH

1.5.42 Status_Monopoly - Status_Discrim

Type:	Integer
Value Range:	$[0; 1]$
Description:	Collection of binary flags indicating whether this group has respective political status (according to EPR-ETH) in this year.
Family Type:	No
Sources:	EPR-ETH

1.5.43 Status_Pwrrank

Type:	Integer
Value Range:	$[1, 7]$
Description:	Ordinal variable indicating this group's ranked political status in this year. Higher values correspond to more powerful groups, i.e. variable ranges from 1 (Discriminated) to 7 (Monopoly), whereas Regional Autonomy and Separatist Autonomy are both assigned the value 3.
Family Type:	No
Sources:	EPR-ETH

1.5.44 Egip_Groups_Count

Type:	Integer
Value Range:	$\mathbb{N}$
Description:	Number of EGIP (see 1.5.40) groups in this group's host country in this year.
Family Type:	No
Sources:	EPR-ETH

1.5.45 Excl_Groups_Count

Type:	Integer
Value Range:	$\mathbb{N}$
Description:	Number of excluded (see 1.5.41) groups in this group's host country in this year.
Family Type:	No
Sources:	EPR-ETH

1.5.46 Rlvt_Groups_Count

Type:	Integer
Value Range:	$\mathbb{N}$
Description:	Number of relevant groups in this group's host country in this year (see 1.5.6).
Family Type:	No
Sources:	EPR-ETH

1.5.47 Actv_Groups_Count

Type:	Integer
Value Range:	$\mathbb{N}$
Description:	Number of active groups in this group's host country in this year (see 1.5.5).
Family Type:	No
Sources:	EPR-ETH

1.5.48 GroupSize

Type:	Double Precision
Value Range:	$(0, 1]$
Description:	This group's population size relative to the host country's entire population (according to EPR-ETH).
Family Type:	No
Sources:	EPR-ETH

1.5.49 lPop

Type:	Double Precision
Value Range:	$(0, 1]$
Description:	Ethnically relevant population in this group's host country relative to total population in this group's host country (according to EPR-ETH). $lPop = \sum_i GroupSize_i, \forall i \in relevant$
Family Type:	No
Sources:	EPR-ETH

1.5.50 lSize

Type:	Double Precision
Value Range:	$(0, 1]$
Description:	This group's population size relative to the sum of ethnically relevant population in the host country (according to EPR-ETH). $lSize = \frac{GroupSize_i}{lPop}$
Family Type:	No
Sources:	EPR-ETH

1.5.51 EgipPop

Type:	Double Precision
Value Range:	$(0, 1]$
Description:	EGIP (see 1.5.40) population in this group's host country relative to total population in this group's host country (according to EPR-ETH). $EgipPop = \sum_i GroupSize_i, \forall i \in EGIP$
Family Type:	No
Sources:	EPR-ETH

1.5.52 lEgipPop

Type:	Double Precision
Value Range:	$(0, 1]$
Description:	Sum of EGIP population in this group's host country relative to the sum of ethnically relevant population in the host country (according to EPR-ETH). $lEgipPop = \frac{EgipPop}{lPop}$
Family Type:	No
Sources:	EPR-ETH

1.5.53 ExclPop

Type:	Double Precision
Value Range:	$(0, 1]$
Description:	Excluded (see 1.5.41) population in this group's host country relative to total population in this group's host country (according to EPR-ETH). $ExclPop = \sum_i GroupSize_i, \forall i \in excluded$
Family Type:	No
Sources:	EPR-ETH

1.5.54 lExclPop

Type:	Double Precision
Value Range:	$(0, 1]$
Description:	Sum of excluded population in this group's host country relative to the sum of ethnically relevant population in the host country (according to EPR-ETH). $lExclPop = \frac{ExclPop}{lPop}$
Family Type:	No
Sources:	EPR-ETH

1.5.55 rbal

Type: Double Precision
Value Range: $(0, 1]$
Description: Relative balance. If this group is excluded: This group's size relative to this group's size plus this group's host country's EGIP population. If this group is an EGIP: This group's size relative to this group's host state's EGIP population.

$$rbal = \begin{cases} GroupSize_i/(EgipPop + GroupSize_i) & \text{if } i \in excluded \\ GroupSize_i/EgipPop & \text{if } i \in EGIP \end{cases}$$

Family Type: No
Sources: EPR-ETH

1.5.56 rlbal

Type: Double Precision
Value Range: $(0, 1]$
Description: Relative balance. Same as rbal (see 1.5.55), but all group sizes are relative to the host country's relevant population.

$$rlbal = \begin{cases} lSize_i/(lEgipPop + lSize_i) & \text{if } i \in excluded \\ lSize_i/lEgipPop & \text{if } i \in EGIP \end{cases}$$

Family Type: No
Sources: EPR-ETH

1.5.57 Downgraded1 - Downgraded10

Type: Integer
Value Range: $[0; 1]$
Description: Collection of binary flags indicating whether this group has experienced a downgrade in its political status (i.e. a negative change on the status_pwrrank variable, see 1.5.43) in the previous N calendar years. N refers to the integer in the respective variable name: DowngradedN.
Family Type: Yes (see 1.2.3)
Sources: EPR-ETH

1.5.58 Downgraded_Excl1 - Downgraded_Excl10

Type:	Integer
Value Range:	[0; 1]
Description:	Collection of binary flags indicating whether this group has experienced a downgrade in its political status from EGIP (see 1.5.40) to excluded (see 1.5.41) in the previous N calendar years. N refers to the integer in the respective variable name: Downgraded_ExclN.
Family Type:	Yes (see 1.2.3)
Sources:	EPR-ETH

1.5.59 Downgraded_Hist

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating whether this group has ever experienced a downgrade in its political status (i.e. a negative change on the status_pwrrank variable, see 1.5.43) since its first inclusion in the dataset.
Family Type:	Yes (see 1.2.3)
Sources:	EPR-ETH

1.5.60 Downgraded_Excl_Hist

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating whether this group has ever experienced a downgrade in its political status from EGIP (see 1.5.40) to excluded (see 1.5.41) since its first inclusion in the dataset.
Family Type:	Yes (see 1.2.3)
Sources:	EPR-ETH

1.5.61 isUrban

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating whether this group's settlement pattern is coded as urban in the GeoEPR-ETH dataset.
Family Type:	No
Sources:	GeoEPR-ETH

1.5.62 isDispersed

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating whether this group's settlement pattern is coded as dispersed in the GeoEPR-ETH dataset.
Family Type:	No
Sources:	GeoEPR-ETH

1.5.63 isMigrant

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating whether this group's settlement pattern is coded as migrant in the GeoEPR-ETH dataset.
Family Type:	No
Sources:	GeoEPR-ETH

1.5.64 hasSetArea

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating whether this group's settlement pattern is coded as regionally based, regional and urban, or aggregate in the GeoEPR-ETH dataset. These categories imply that the group has a clearly defined settlement pattern, i.e. the group's settlement pattern is represented with a clearly defined polygon in the GeoEPR-ETH dataset.
Family Type:	No
Sources:	GeoEPR-ETH

Chapter 2

Country-Level Data

2.1 EPR-ETH Countries

The unit of observation in the RFE country-level data is the country-year. The sample universe of country-years in the RFE country-level data is adopted from the EPR-ETH (*Ethnic Power Relations*) dataset (citealtCederman2010, for more detailed information concerning versions and sources see 3.2). Consequently, the country-years included in the RFE country-level dataset are equivalent to the country-years for which the EPR-ETH dataset codes ethnic groups, and thus for which the RFE group-level dataset provides information on the ethnic-group-level (see 1.1).

The EPR-ETH dataset covers all countries in the period 1946 - 2009 that meet the following criteria:

(i) Administered by an intact sovereign state, i.e. colonies and failed states are not included [1].

(ii) Population is greater than or equal to 500'000 inhabitants.

Newly independent states are included in the dataset beginning with the year of independence, e.g. Croatia (independence on 25/06/1991) is included from 1991 onwards. Given these criteria, EPR-ETH covers a total of 7'988 country-years from 162 countries.

[1] In fact, countries considered "administered" by a failed state *are* included in the dataset if the period during which they are coded as failed states falls in between periods when the country meets the EPR-ETH inclusion criteria. During these periods, however, the coding on the ethnic group level (see below) is not continued.

2.2 Country-Level Conflict Data

The information on intrastate conflicts in the RFE country-level data originates primarily from the UCDP ACD dataset (Gleditsch et al. 2002). Information on whether intrastate conflicts are coded as ethnic is added using the RFE group-level conflict data, which is compiled from the ACD2EPR dataset (Wucherpfennig et al. 2012), the NSA dataset (Cunningham et al. 2009), and the ACD dataset (Gleditsch et al. 2002, (see 1.3). The remainder of this section briefly describes how the ACD data is aggregated into country-year format and elaborates the coding rules for country-level conflict onset and incidence.

2.2.1 ACD Aggregation

The ACD dataset codes inter- and intrastate armed conflicts on a yearly basis. Conflicts in the ACD dataset follow the UCDP conflict definition (please see 3.2 for links to more information and references) and are assigned a unique *UCDP ID*. ACD conflicts are assigned to one of four categories: Extrasystemic, Interstate, Internal, and Internationalized Internal Conflicts. Of these types, only the latter two are relevant for the RFE country-level conflict coding, since we focus exclusively on civil wars.

ACD conflicts are distinguished along the incompatibility they originate from, not their temporal dimension. Hence, a conflict that ends and reoccurs after a substantial period of time is assigned the same UCDP ID, regardless of the time period separating the actual conflict episodes, as long as the incompatibility remains the same. The ACD dataset provides information on the temporal dimension of conflict termination and recurrence with a variable indicating the end of a conflict episode, whereas according to the ACD coding rules, a conflict episode ends in a given year if there is no conflict-related activity in the next calendar year[2]. For the purpose of the RFE country-level data conflict coding, this definition is altered: For all conflict onset and incidence variables in the RFE country-level dataset, a conflict episode is only considered terminated if there is *no conflict-related activity in the following two calender years*. This recoding of ACD conflict episode is adopted in order to prevent an inflation of conflict episodes (and thus conflict onsets, see 1.3.4) in low-activity conflicts that reoccur repeatedly.

With these coding rules in place, the ACD conflict episodes are merged with the EPR-ETH country-level data (see 2.1) by assigning each EPR-ETH

[2]"Conflict-related activity" refers to the UCDP threshold of at least 25 battledeaths per annum.

observation one or several UCDP IDs if any the ACD dataset reports an ongoing conflict episode in the respective country (according to ACD's *GWNoLoc* variable) and year.

2.2.2 Conflict Onset

In the RFE country-level data a conflict onset occurs if a new ACD conflict episode starts in a given country-year (note that conflict episodes are redefined using a 2-year termination restriction, see 2.2.1).

The RFE country-level data also offeres conflict onset variables that distinguish between ethnic and non-ethnic conflicts. Conflict onsets are coded as ethnic if, according to the RFE group-level conflict data (see 1.3), at least one ethnic group is linked to the respective ACD conflict episode *in the onset year*. This implies that the number of ethnic conflict onsets defined in the RFE country-level data may differ from the aggregated number of onsets in the RFE group-level data, even though the two datasets are based on the same sources.

2.2.3 KO and DO Options

The RFE country-level data conflict onset variables come in two variants: The KO and the DO option. KO stands for *Keep Ongoing* and is the default option. Conflict onset variables with the *KO* option take the value of 1 for country-years in which a cizbtry experiences conflict onset (see 2.2.2), and 0 in all other years.

DO stands for *Drop Ongoing* and represents a slightly more advanced coding. DO conflict onset variables indicate conflict onset in the same manner as the KO variables, but are censored in the years subsequent to conflict onset, during which the respective ACD conflict episode is ongoing. The intuition behind the DO coding is to remove country-years from conflict onset variables for which the RFE country-level onset coding rules (see 2.2.2) make positive values on the onset variable unlikely[3], since conflict onsets are only coded for the *begianning* of ACD conflict episodes.

[3]We write "unlikely", and not "impossible", because there are configurations where the RFE country-level coding rules would allow onsets to occur, but these are censored by the DO coding. Specifically, this is the case if a country experiences the start of an entirely new ACD conflict in a period where another ACD conflict is still ongoing

2.2.4 Conflict Incidence

The RFE country-level data includes a number of conflict incidence variables, which indicate whether in a given year a country is involved in at least one ACD conflict episode. Conflict incidence variables assume the value of 1 in years when at least one ACD conflict episode is ongoing, and 0 in all other years.

Please note that because of the 2-year episode coding rule imposed on the ACD data, conflict episodes may terminate for a single calendar year and then continue without a new country-level conflict onset occuring (see 2.2.1 and 2.2.2).

Please also note that while conflict onsets are only coded as ethnic if at least one ethnic group is involved in the respective ACD conflict episode in the onset year, incidence variable distinguishing between ethnic and non-ethnic conflicts do so on a yearly basis. Consequently, ACD conflict episodes that do not entail an ethnic conflict onset may be coded as an ongoing ethnic conflict in subsequent years if an ethnic group becomes involed in the conflict episode at a later stage.

2.3 Country-Level Information on Ethnic Groups

The RFE country-level data includes a number of variables referring to the composition of a country's ethnic groups in terms of size and power status. This information is aggregated directly from the RFE group-level data, which is based on the EPR-ETH dataset (Cederman et al. 2010); for detailed information please see 1. Since the EPR-ETH dataset only provides information on group sizes and powerstati for groups considered politically relevant (see 1.2.2), the aggregated variables in the RFE country-year data also only refer to politically relevant groups.

Furthermore, the EPR-ETH dataset includes a set of countries for which no ethnic group is coded as politically relevant during the entire period covered by the dataset. As elaborated above (see 1.1.2), for these countries a so-called *placeholder-group* is coded which remains politically irrelevant throughout the entire period covered by the dataset. The variable *Cntr_Relevance* (see 2.4.55) indicates for which countries this is the case. For these countries, all variables referring to aggregate group–sizes or –powerstati are coded as missing.

2.4 Variables

This section provides a full list of variables included in the RFE country-level data, together with information on substantial content, sources, and coding rules. Please note that knowledge of the underlying data structure as described in sections 2.1 to 2.3 is essential for fully understanding some of the information offered in this section.

2.4.1 CountryName

Type:	Text
Value Range:	NA
Description:	Country name, follows COW name wherever possible.
Family Type:	No
Sources:	COW State System Membership List

2.4.2 Countries_Cow

Type:	Integer
Value Range:	$[2, 920]$
Description:	Country's COW ID.
Family Type:	No
Sources:	COW State System Membership List

2.4.3 All_Onset_KO_Flag

Type:	Integer
Value Range:	$[0; 1]$
Description:	Binary flag indicating country-level onset, KO option (see 2.2.2, 2.2.3).
Sources:	ACD

2.4.4 All_Onset_DO_Flag

Type:	Integer
Value Range:	$[0; 1]$
Description:	Binary flag indicating country-level onset, DO option (see 2.2.2, 2.2.3).
Sources:	ACD

2.4.5 All_Onset_UCDPID_a

Type:	Integer Array {Int, Int, ...}
Value Range:	NA
Description:	Array variable indicating UCDP IDs of conflict onsets this country experiences in this year; based on KO coding rule (see 2.2.2, 2.2.3).
Sources:	ACD

2.4.6 All_Onset_Count

Type:	Integer
Value Range:	$\mathbb{N}$
Description:	Count variable indicating number of conflict onsets this country experiences in this year; based on KO coding rule (see 2.2.2, 2.2.3).
Family Type:	Yes (see 1.2.3)
Sources:	ACD

2.4.7 All_Onset_KO_Terr_Flag

Type:	Integer
Value Range:	$[0; 1]$
Description:	Binary flag indicating country-level onset for ACD conflicts with territorial incompatibility (according to ACD's Incomp variable), KO option (see 2.2.2, 2.2.3).
Sources:	ACD

2.4.8 All_Onset_DO_Terr_Flag

Type:	Integer
Value Range:	$[0; 1]$
Description:	Binary flag indicating country-level onset for ACD conflicts with territorial incompatibility (according to ACD's Incomp variable), DO option (see 1.3.4, 1.3.5). DO coding rules only enforced for conflicts with appropriate incompatibility, i.e. variable censored only for ongoing territorial ACD episodes.
Sources:	ACD

2.4.9 All_Onset_KO_Gov_Flag

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating country-level onset for ACD conflicts with governmental incompatibility (according to ACD's Incomp variable), KO option (see 2.2.2, 2.2.3).
Sources:	ACD

2.4.10 All_Onset_DO_Gov_Flag

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating country-level onset for ACD conflicts with governmental incompatibility (according to ACD's Incomp variable), DO option (see 1.3.4, 1.3.5). DO coding rules only enforced for conflicts with appropriate incompatibility, i.e. variable censored only for ongoing governmental ACD episodes.
Sources:	ACD

2.4.11 All_Onset_KO_Eth_Flag

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating country-level ethnic conflict onset (according to the RFE group-level data), KO option (see 2.2.2, 2.2.3).
Sources:	ACD2EPR, NSA, ACD

2.4.12 All_Onset_DO_Eth_Flag

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating country-level ethnic conflict onset (according to the RFE group-level data), DO option (see 2.2.2, 2.2.3). DO coding rules only enforced for ethnic conflicts, i.e. variable censored only for ongoing ethnic conflicts.
Sources:	ACD2EPR, NSA, ACD

2.4.13 All_Onset_KO_NonEth_Flag

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating country-level non-ethnic conflict onset (according to the RFE group-level data), KO option (see 2.2.2, 2.2.3).
Sources:	ACD2EPR, NSA, ACD

2.4.14 All_Onset_DO_NonEth_Flag

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating country-level non-ethnic conflict onset (according to the RFE group-level data), DO option (see 2.2.2, 2.2.3). DO coding rules only enforced for non-ethnic conflicts, i.e. variable censored only for ongoing non-ethnic conflicts.
Sources:	ACD2EPR, NSA, ACD

2.4.15 All_Onset_KO_Eth_Terr_Flag

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating country-level ethnic onset for ACD conflicts with territorial incompatibility (according to the RFE group-level data and ACD's Incomp variable, respectively), KO option (see 2.2.2, 2.2.3).
Sources:	ACD2EPR, NSA, ACD

2.4.16 All_Onset_DO_Eth_Terr_Flag

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating country-level ethnic onset for ACD conflicts with territorial incompatibility (according to the RFE group-level data and ACD's Incomp variable, respectively), DO option (see 2.2.2, 2.2.3). DO coding rules only enforced for ethnic conflicts with appropriate incompatibility, i.e. variable censored only for ongoing territorial ethnic conflicts.
Sources:	ACD2EPR, NSA, ACD

2.4.17 All_Onset_KO_Eth_Gov_Flag

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating country-level ethnic onset for ACD conflicts with governmental incompatibility (according to the RFE group-level data and ACD's Incomp variable, respectively), KO option (see 2.2.2, 2.2.3).
Sources:	ACD2EPR, NSA, ACD

2.4.18 All_Onset_DO_Eth_Gov_Flag

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating ethnic country-level onset for ACD conflicts with governmental incompatibility (according to the RFE group-level data and ACD's Incomp variable, respectively), DO option (see 2.2.2, 2.2.3). DO coding rules only enforced for ethnic conflicts with appropriate incompatibility, i.e. variable censored only for ongoing governmental ethnic conflicts.
Sources:	ACD2EPR, NSA, ACD

2.4.19 All_Onset_KO_NonEth_Terr_Flag

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating country-level onset for non-ethnic ACD conflicts with territorial incompatibility (according to the RFE group-level data and ACD's Incomp variable, respectively), KO option (see 2.2.2, 2.2.3).
Sources:	ACD2EPR, NSA, ACD

2.4.20 All_Onset_DO_NonEth_Terr_Flag

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating country-level non-ethnic onset for ACD conflicts with territorial incompatibility (according to the RFE group-level data and ACD's Incomp variable, respectively), DO option (see 2.2.2, 2.2.3). DO coding rules only enforced for non-ethnic conflicts with appropriate incompatibility, i.e. variable censored only for ongoing territorial non-ethnic conflicts.
Sources:	ACD2EPR, NSA, ACD

2.4.21 All_Onset_KO_NonEth_Gov_Flag

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating country-level onset for non-ethnic ACD conflicts with governmental incompatibility (according to the RFE group-level data and ACD's Incomp variable, respectively), KO option (see 2.2.2, 2.2.3).
Sources:	ACD2EPR, NSA, ACD

2.4.22 All_Onset_DO_NonEth_Gov_Flag

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating country-level non-ethnic onset for ACD conflicts with governmental incompatibility (according to the RFE group-level data and ACD's Incomp variable, respectively), DO option (see 2.2.2, 2.2.3). DO coding rules only enforced for non-ethnic conflicts with appropriate incompatibility, i.e. variable censored only for ongoing governmental non-ethnic conflicts.
Sources:	ACD2EPR, NSA, ACD

2.4.23 All_Onset_Eth_Terr_Count

Type:	Integer
Value Range:	$\mathbb{N}$
Description:	Count variable indicating number of ethnic conflict onsets for ACD conflicts with territorial incompatability (according to ACD's Incomp variable) this country experiences in this year; based on KO coding rule (see 2.2.2, 2.2.3).
Sources:	ACD2EPR, NSA, ACD

2.4.24 All_Onset_Eth_Gov_Count

Type:	Integer
Value Range:	$\mathbb{N}$
Description:	Count variable indicating number of ethnic conflict onsets for ACD conflicts with governmental incompatability (according to ACD's Incomp variable) this country experiences in this year; based on KO coding rule (see 2.2.2, 2.2.3).
Sources:	ACD2EPR, NSA, ACD

2.4.25 All_Onset_NonEth_Terr_Count

Type:	Integer
Value Range:	$\mathbb{N}$
Description:	Count variable indicating number of non-ethnic conflict onsets for ACD conflicts with territorial incompatability (according to ACD's Incomp variable) this country experiences in this year; based on KO coding rule (see 2.2.2, 2.2.3).
Sources:	ACD2EPR, NSA, ACD

2.4.26 All_Onset_NonEth_Gov_Count

Type:	Integer
Value Range:	$\mathbb{N}$
Description:	Count variable indicating number of non-ethnic conflict onsets for ACD conflicts with governmental incompatability (according to ACD's Incomp variable) this country experiences in this year; based on KO coding rule (see 2.2.2, 2.2.3).
Sources:	ACD2EPR, NSA, ACD

2.4.27 All_Incidence_Flag

Type:	Integer
Value Range:	$[0; 1]$
Description:	Binary flag indicating ongoing ACD conflict episode in the given year (see 2.2.4).
Sources:	ACD

2.4.28 All_Incidence_UCDPID_a

Type:	Integer Array {Int, Int, ...}
Value Range:	NA
Description:	Array variable indicating UCDP IDs of ACD conflict episodes ongoing in this country in this year (see 2.2.4).
Sources:	ACD

2.4.29 All_Incidence_Count

Type:	Integer
Value Range:	$\mathbb{N}$
Description:	Count variable indicating number ongoing of ACD conflict episodes in this year (see 2.2.4).
Family Type:	Yes (see 1.2.3)
Sources:	ACD2EPR, NSA, ACD

2.4.30 All_Incidence_Terr_Flag

Type:	Integer
Value Range:	$[0; 1]$
Description:	Binary flag indicating ongoing ACD conflict episode with territorial incompatibility (according to ACD's Incomp variable) in given year (see 2.2.4).
Sources:	ACD

2.4.31 All_Incidence_Gov_Flag

Type:	Integer
Value Range:	$[0; 1]$
Description:	Binary flag indicating ongoing ACD conflict episode with governmental incompatibility (according to ACD's Incomp variable) in given year (see 2.2.4).
Sources:	ACD

2.4.32 All_Incidence_Eth_Flag

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating ongoing ethnic ACD conflict episode (according to the RFE group-level data) in given year (see 2.2.4).
Sources:	ACD2EPR, NSA, ACD

2.4.33 All_Incidence_NonEth_Flag

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating ongoing non-ethnic ACD conflict episode (according to the RFE group-level data) in given year (see 2.2.4).
Sources:	ACD2EPR, NSA, ACD

2.4.34 All_Incidence_Eth_Terr_Flag

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating ongoing ethnic ACD conflict episode with territorial incompatibility (according to the RFE group-level data and ACD's Incomp variable, respectively) in given year (see 2.2.4).
Sources:	ACD2EPR, NSA, ACD

2.4.35 All_Incidence_Eth_Gov_Flag

Type:	Integer
Value Range:	[0; 1]
Description:	Binary flag indicating ongoing ethnic ACD conflict episode with governmental incompatibility (according to the RFE group-level data and ACD's Incomp variable, respectively) in given year (see 2.2.4).
Sources:	ACD2EPR, NSA, ACD

2.4.36 All_Incidence_NonEth_Terr_Flag

Type:	Integer
Value Range:	$[0; 1]$
Description:	Binary flag indicating ongoing non-ethnic ACD conflict episode with territorial incompatibility (according to the RFE group-level data and ACD's Incomp variable, respectively) in given year (see 2.2.4).
Sources:	ACD2EPR, NSA, ACD

2.4.37 All_Incidence_NonEth_Gov_Flag

Type:	Integer
Value Range:	$[0; 1]$
Description:	Binary flag indicating ongoing non-ethnic ACD conflict episode with governmental incompatibility (according to the RFE group-level data and ACD's Incomp variable, respectively) in given year (see 2.2.4).
Sources:	ACD2EPR, NSA, ACD

2.4.38 All_Incidence_Eth_Terr_Count

Type:	Integer
Value Range:	$\mathbb{N}$
Description:	Count variable indicating number of ongoing ethnic ACD conflict episodes with territorial incompatibility (according to the RFE group-level data and ACD's Incomp variable, respectively) in this year (see 2.2.4).
Sources:	ACD2EPR, NSA, ACD

2.4.39 All_Incidence_Eth_Gov_Count

Type:	Integer
Value Range:	$\mathbb{N}$
Description:	Count variable indicating number of ongoing ethnic ACD conflict episodes with governmental incompatibility (according to the RFE group-level data and ACD's Incomp variable, respectively) in this year (see 2.2.4).
Sources:	ACD2EPR, NSA, ACD

2.4.40 All_Incidence_NonEth_Terr_Count

Type:	Integer
Value Range:	$\mathbb{N}$
Description:	Count variable indicating number of ongoing non-ethnic ACD conflict episodes with territorial incompatibility (according to the RFE group-level data and ACD's Incomp variable, respectively) in this year (see 2.2.4).
Sources:	ACD2EPR, NSA, ACD

2.4.41 All_Incidence_NonEth_Gov_Count

Type:	Integer
Value Range:	$\mathbb{N}$
Description:	Count variable indicating number of ongoing non-ethnic ACD conflict episodes with governmental incompatibility (according to the RFE group-level data and ACD's Incomp variable, respectively) in this year (see 2.2.4).
Sources:	ACD2EPR, NSA, ACD

2.4.42 All_PeaceYears

Type:	Integer
Value Range:	$\mathbb{N}$
Description:	Count variable indicating the number of calendar years since this country last experienced an ongoing ACD conflict episode. In case a country has never experienced an ACD conflict, the variable counts the number of calendar years since 1946 or the year this country is first included in the EPR-ETH dataset. Based on All_Incidence_Flag (see2.4.27).
Sources:	ACD

2.4.43 All_WarHist

Type:	Integer
Value Range:	$\mathbb{N}$
Description:	Count variable indicating the number of onsets this country has experienced in the years prior to this year. Based on All_Onset_KO_Flag (see 2.4.3).
Family Type:	Yes (see 1.2.3)
Sources:	ACD

2.4.44 Egip_Groups_Count

Type:	Integer
Value Range:	$\mathbb{N}$
Description:	Number of EGIP (see 1.5.40) groups in this country in this year.
Sources:	EPR-ETH

2.4.45 Excl_Groups_Count

Type:	Integer
Value Range:	$\mathbb{N}$
Description:	Number of excluded (see 1.5.41) groups in this country in this year.
Sources:	EPR-ETH

2.4.46 Rlvt_Groups_Count

Type:	Integer
Value Range:	$\mathbb{N}$
Description:	Number of relevant groups in this country in this year (see 1.5.6).
Sources:	EPR-ETH

2.4.47 Actv_Groups_Count

Type:	Integer
Value Range:	$\mathbb{N}$
Description:	Number of active groups in this country in this year (see 1.5.5).
Sources:	EPR-ETH

2.4.48 lPop

Type:	Double Precision
Value Range:	$(0, 1]$
Description:	Ethnically relevant population in this country relative to total population in this country (according to EPR-ETH). $lPop = \sum_i GroupSize_i, \forall i \in relevant$
Sources:	EPR-ETH

2.4.49 EgipPop

Type:	Double Precision
Value Range:	$(0, 1]$
Description:	EGIP (see 1.5.40) population in this group's host country relative to total population in this group's host country (according to EPR-ETH). $EgipPop = \sum_i GroupSize_i, \forall i \in EGIP$
Family Type:	No
Sources:	EPR-ETH

2.4.50 lEgipPop

Type:	Double Precision
Value Range:	$(0, 1]$
Description:	Sum of EGIP population in this country relative to the sum of ethnically relevant population in this country (according to EPR-ETH). $lEgipPop = \frac{EgipPop}{lPop}$
Sources:	EPR-ETH

2.4.51 ExclPop

Type:	Double Precision
Value Range:	$(0, 1]$
Description:	Excluded (see 1.5.41) population in this country relative to total population in this country (according to EPR-ETH). $ExclPop = \sum_i GroupSize_i, \forall i \in excluded$
Sources:	EPR-ETH

2.4.52 lExclPop

Type:	Double Precision
Value Range:	$(0, 1]$
Description:	Sum of EGIP population in this country relative to the sum of ethnically relevant population in this country (according to EPR-ETH). $lEgipPop = \frac{EgipPop}{lPop}$
Sources:	EPR-ETH

2.4.53 maxExclPop

Type:	Double Precision
Value Range:	$(0, 1]$
Description:	Size of the largest excluded group (see 1.5.41) in this country relative to total population in this country (according to EPR-ETH) $MaxExclPop = max(GroupSize_i), \forall i \in excluded$
Sources:	EPR-ETH

2.4.54 lmaxExclPop

Type:	Double Precision
Value Range:	$(0, 1]$
Description:	Size of the largest excluded group (see 1.5.41) in this country relative to the sum of ethnically relevant population in this country (according to EPR-ETH) $lMaxExclPop = max(lGroupSize_i), \forall i \in excluded$
Sources:	EPR-ETH

2.4.55 Cntr_Relevance

Type:	String
Value Range:	$\{R; P\}$
Description:	String indicating whether country hosts any politically relevant ethnic groups during coded period (according to EPR-ETH, see 2.3). 'R' indicates politically relevant groups are coded, 'P' indicates only placeholder-groups are coded.
Sources:	EPR-ETH

2.4.56 Nstar

Type:	Double Precision
Value Range:	$(0, 1)$
Description:	$N^*(0.5, 5)$, see Cederman and Girardin (2007)
Sources:	EPR-ETH

Chapter 3

Appendices

3.1 Terms and Conditions

By downloading data offered via the GROW^{up} Research Front-End (henceforth referred to as RFE), I agree to the following:

1. I agree that any books, articles, conference papers, theses, dissertations, reports, or other publications that I create using data distributed via the RFE reference the bibliographic citation accompanying the data. These citations include the data authors, data identifier, and other relevant information.

2. I understand that the data distributed via the RFE is in part compiled from external data sources and agree that the use of RFE data compiled from external data sources implies acceptance with the terms and conditions associated with these sources.

3. The distributor makes no warranties, expressed or implied, by operation of law or otherwise, regarding or relating to the data distributed via the RFE.

3.2 Sources Release 1.1

The currently productive RFE Release is version 1.1 and uses the sources listed in table 3.1. *Please note that according to the GROW^{up} RFE Terms and Conditions (see 3.1), you are required to cite all sources of a variable if you include it in your research.*

3.3 Changelog

3.3.1 Version 1.0 to Version 1.1

Due to a coding error in the underlying script, the following three variables measuring country-level non-ethnic conflict have been revised:

- *All_Onset_KO_NonEth_Flag* now indicates 7 additional non-ethnic conflict onsets.
- *All_Onset_DO_NonEth_Flag* now indicates 7 additional non-ethnic conflict onsets, and features 141 additional censored observations (due to previously unreported ongoing non-ethnic conflicts).
- *All_Incidence_NonEth_Flag* now indicates 148 additional non-ethnic conflict incidence years.

A detailed list of the affected observations is available here: http://www.icr.ethz.ch/data/growup/RFE_1.1_NonEth_ErrorCorrection.xlsx.

Table 3.1: RFE Sources

Dataset	Full Name	Version	Full Ref.	URL
COW	Correlates of War State System Membership List	v2008.1	Correlates of War Project (2008). State System Membership List, v2008.1. *Online:* ***http://correlatesofwar.org***.	`http://correlatesofwar.org`
EPR-ETH	Ethnic Power-Relations Dataset	2.0	Cederman, L.-E., A. Wimmer, and B. Min (2010). Why do ethnic groups rebel? New data and analysis. *World Politics 62* (1), pp. 87–119.	`http://www.icr.ethz.ch/data`
ACD2EPR	ACD2EPR Docking Dataset	1.2	Wucherpfennig, J., N. Metternich, L.-E. Cederman, and K. S. Gleditsch (2012). Ethnicity, the state and the duration of civil war. *World Politics 64* (1), pp. 79–115.	`http://www.icr.ethz.ch/data`
NSA	Non-State Actor Dataset	3.1	Cunningham, D. E., K. Skrede Gleditsch, and I. Salehyan (2009). It takes two. *Journal of Conflict Resolution 53* (4), pp. 570–597.	`NA`
ACD	UCDP Armed Conflict Dataset	v4-2010	Gleditsch, N. P., P. Wallensteen, M. Eriksson, M. Sollenberg, and H. Strand (2002). Armed conflict 1946-2001: A new dataset. *Journal of Peace Research 39* (5), pp. 615–637.	`http://www.prio.no/CSCW/Datasets/Armed-Conflict/UCDP-PRIO`
GeoEPR-ETH	GeoEPR-ETH Dataset	2.0	Wucherpfennig, J., N. B.Weidmann, L. Girardin, L.-E. Cederman, and A. Wimmer (2011). Politically relevant ethnic groups across space and time: Introducing the GeoEPR dataset. *Conflict Management and Peace Science 28* (5), pp. 423–437.	`http://www.icr.ethz.ch/data`

Bibliography

Cederman, L.-E. and L. Girardin (2007). Beyond fractionalization: Mapping ethnicity onto nationalist insurgencies. *The American Political Science Review 101*(1), pp. 173–185.

Cederman, L.-E., A. Wimmer, and B. Min (2010). Why do ethnic groups rebel? new data and analysis. *World Politics 62*(1), pp. 87–119.

Correlates of War Project (2008). State System Membership List, v2008.1. *Online:* `http://correlatesofwar.org`.

Cunningham, D. E., K. Skrede Gleditsch, and I. Salehyan (2009). It takes two. *Journal of Conflict Resolution 53*(4), pp. 570–597.

Gleditsch, N. P., P. Wallensteen, M. Eriksson, M. Sollenberg, and H. Strand (2002). Armed conflict 1946-2001: A new dataset. *Journal of Peace Research 39*(5), pp. 615–637.

Uppsala Conflict Data Program (2010). UCDP Actor Dataset 2.1-2010. *Uppsala University*.

Wucherpfennig, J., N. Metternich, L.-E. Cederman, and K. S. Gleditsch (2012). Ethnicity, the state and the duration of civil war. *Wolrd Politics 64*(1), pp. 79–115.

Wucherpfennig, J., N. B. Weidmann, L. Girardin, L.-E. Cederman, and A. Wimmer (2011). Politically relevant ethnic groups across space and time: Introducing the GeoEPR dataset. *Conflict Management and Peace Science*, Forthcoming.