

Can we 'measure' ethnic identity and conflict?

Andreas Juon

ECMI Summer School on National Minorities in Border Regions

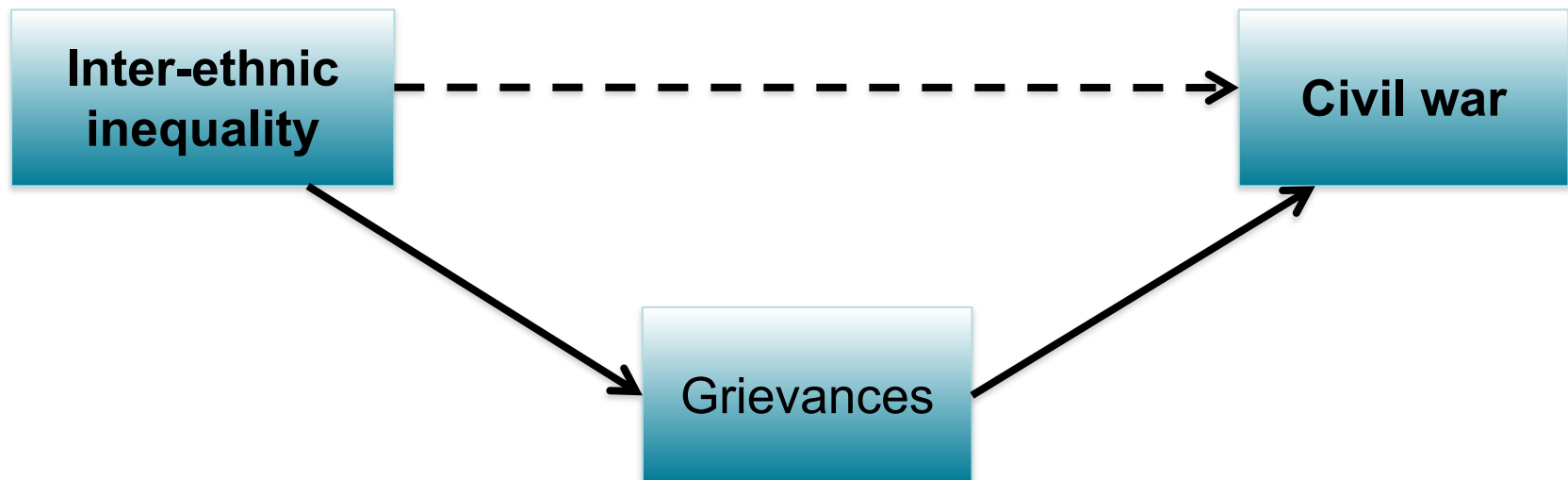
August 30, 2023

Contents

- Motivation
- **The Ethnic Power Relations Dataset**
- **Opportunities and challenges of quantitative research on ethnic conflict**
- Summary
- **Practical exercise**

Learning goals

- You can explain to a non-academic why the **measurement of ethnic identity and conflict** is consequential (for research *and* policy).
- You can explain how you could use **the EPR dataset** for your own research.
- If doing so in a quantitative analysis, you can explain **opportunities** and important **challenges**.



1. Motivation

Ethnic inequality as a driver of civil war?

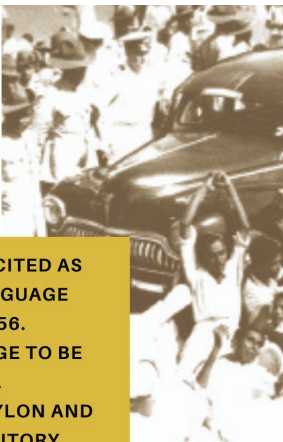
Ethnic inequality as a driver of civil war?

1. Grievance approach

- Case studies: ethnic inequality = key motivational factor in many civil wars
- E.g. Sri Lanka: Sinhala only official language, Buddhism state religion, citizenship restrictions + political exclusion against Tamils, resettlement programs → Tamil grievances → civil war (1983-2009)
- **Other examples?**

What did the
“Sinhala Only
Act” state?

THIS ACT MAY BE CITED AS
THE OFFICIAL LANGUAGE
ACT, NO. 33 OF 1956.
SINHALA LANGUAGE TO BE
THE ONE OFFICIAL
LANGUAGE OF CEYLON AND
TO ENABLE TRANSITORY
PROVISIONS TO BE MADE.



Ethnic inequality as a driver of civil war?

2. The skeptics

Ethnic identity and inequality are irrelevant (for explaining the outbreak of civil war).

So what causes civil war? Rebel movements themselves justify their actions in terms of a catalogue of grievances: repression, exploitation, exclusion. Politically motivated academics have piled in with their own hobbyhorses, which usually cast rebels as heroes. I have come to distrust this discourse of grievances as self-serving.

Bottom Billion, OUP 2007



Prof. Paul Collier, Oxford Univ.

Ethnic inequality as a driver of civil war?

2. The skeptics

- Omnipresence of grievances vs. rare civil wars
- Quantitative studies: Ethnic diversity and economic inequality do not predict civil war.

Ethnic fractionalization	0.010 (0.006)*	0.011 (0.007)*	0.012 (0.008)
Religious fractionalization	-0.003 (0.007)	-0.006 (0.008)	-0.004 (0.009)
Polarization $\alpha = 1.6$	-3.067 (7.021)	-4.682 (8.267)	-6.536 (8.579)
Ethnic dominance (45–90%)	0.414 (0.496)	0.575 (0.586)	1.084 (0.629)*
Democracy	-0.109 (0.044)***	-0.083 (0.051)*	-0.121 (0.053)**
Peace duration	-0.004 (0.001)***	-0.003 (0.001)***	-0.004 (0.001)***
Mountainous terrain	0.011 (0.007)	0.007 (0.009)	-0.0001 (0.009)
Geographic dispersion	-0.509 (0.856)	-0.763 (1.053)	-1.293 (1.102)
Ln population	0.221 (0.096)**	0.246 (0.119)**	0.300 (1.133)**
Income inequality		0.015 (0.018)	
Land inequality			0.461 (1.305)

The four proxies for ethnic and religious tension are surprisingly unimportant in view of the attention that the phenomenon attracts. Ethnic fractionalization is significant at 10% with the expected sign. Religious fractionalization and polarization are insignificant with the wrong sign, and ethnic dominance is insignificant. Nor are the three measures jointly significant.¹⁸ Democracy is highly significant

Ethnic inequality as a driver of civil war?

3. Real-world implications: how to prevent civil war?

- **Skeptics:** opportunities to rebel and greed as motivational factors
 - stronger states
 - alternative economic opportunities
 - preventing rebel financing
- **Grievance approach:**
 -
 -
 -

Exercise

In pairs, reflect on:

- How do Collier/Hoeffler **measure** ethnic inequality in their quantitative analysis (slides 7 and 10)?
 - What do these measurements (not) capture?
 - What would an ideal measurement capture?
- What **opportunities and challenges** of quantitatively examining ethnic conflict does this debate highlight?

Exercise

Ethnic / religious fractionalization

Ethnic and religious hatreds are widely perceived as a cause of civil conflict. Although such hatreds cannot be quantified, they can evidently only occur in societies that are multi-ethnic or multi-religious and so our proxies measure **various dimensions of diversity**. Our previously discussed measures of fractionalization are pertinent: inter-group hatreds must be greater in societies that are fractionalized than in those which are homogenous. However, arguably the source of inter-group tension is not diversity but polarization.¹⁰ Fortunately, the allowable class of measures of polarization is quite limited. We adopt a general measure due to Esteban and Ray (1994)

$$P = K \sum_{i=1}^n \sum_{j=1}^n \pi_i^{1+\alpha} \pi_j d \quad (1)$$

where π_i denotes the **percentage of people that belong to group i** in the total population. α and d are the parameters

Income / land inequality

redistribution.¹³ We measure income inequality by the **Gini coefficient** and by the ratio of the top-to-bottom quintiles of **income**. We measure asset inequality by the Gini coefficient of **land ownership**. The data are from Deininger and Squire (1996, 1998). Inequality is slightly higher prior to the conflict episodes.

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

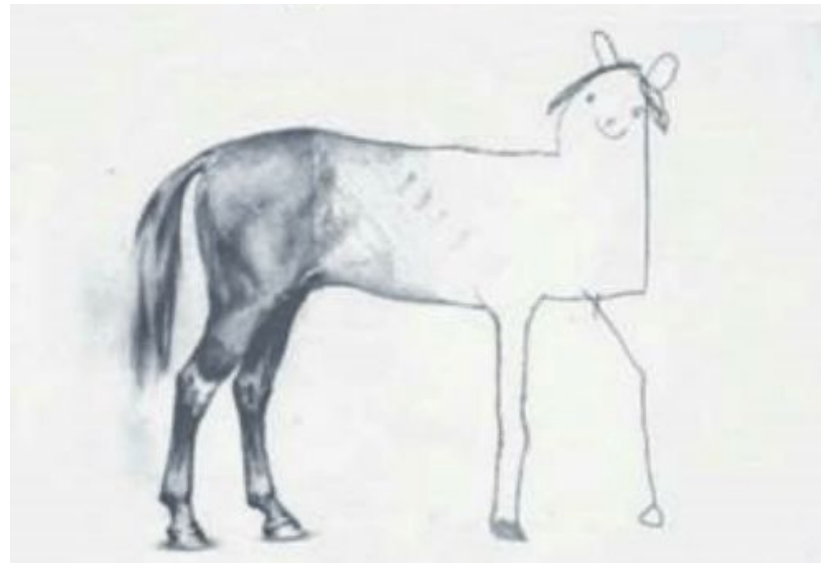


2. The Ethnic Power Relations Dataset

1. Conceptualization + measurement

Challenge

- Many important concepts are difficult to capture by numbers.
- Especially true for ethnic identity and ethnic inequality!



Concept ... vs. ... measurement

1. Conceptualization + measurement

1. Ethnic identity: Conceptualization

- 'Following the Weberian tradition, we define ethnicity as a **subjectively** experienced sense of commonality ... based on: common **language**, similar **phenotypical** features, adherence to the same **faith**, and so on.' (EPR Codebook)
- Identity basis can vary between contexts:
 - Israel: Ethnicity + religion
 - India: Religion + language
 - US / apartheid South Africa: race
 - ...

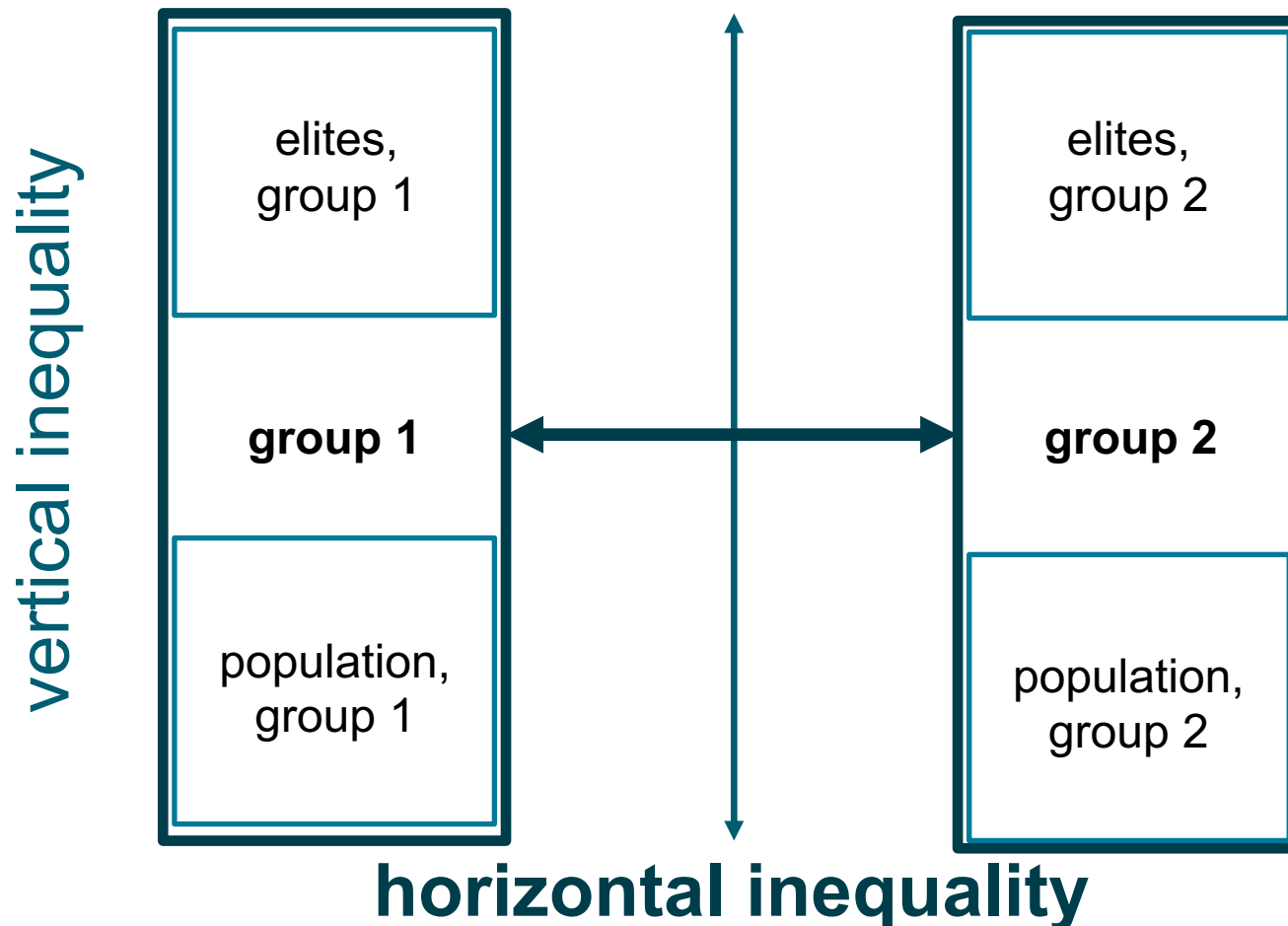
1. Conceptualization + measurement

1. Ethnic identity: Measurement

- List of **politically 'relevant' groups** (1 organization or systematic political discrimination)
- e.g. Sri Lanka:
 - Sinhalese (75%)
 - Sri Lankan Tamils (11%)
 - Moors (Muslims) (9%)
 - Indian Tamils (4%)

1. Conceptualization + measurement

■ 2. Ethnic inequality: conceptualization



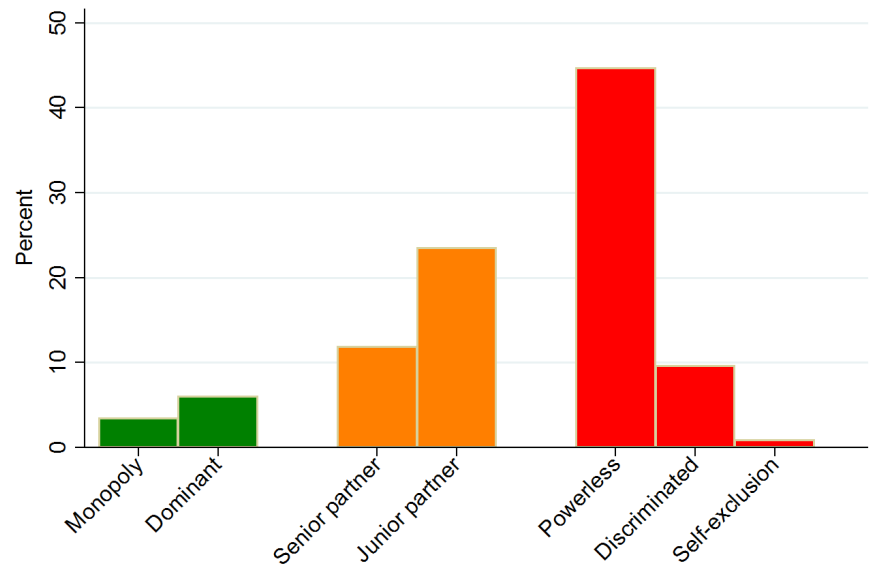
1. Conceptualization + measurement

■ 2. Ethnic inequality: measurement

- see <https://growup.ethz.ch/>
- EPR-Core: power relations in the central government + regional autonomy



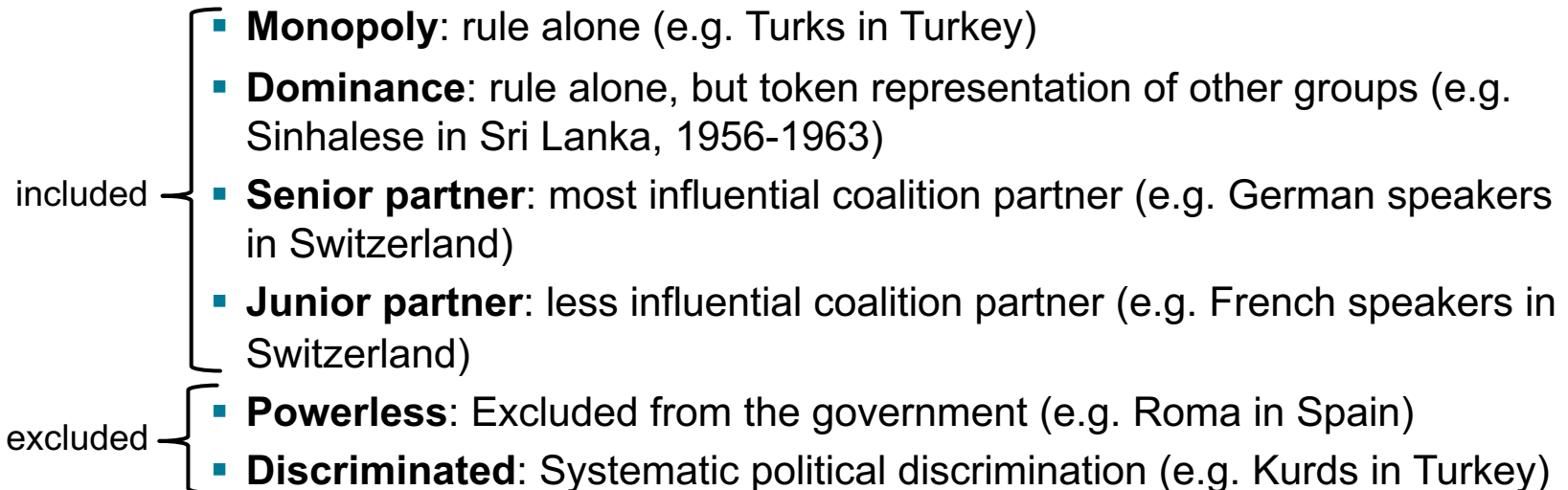
Figure 954: Political status of ethnic groups in Sri Lanka (Ceylon) during 1956-1963.



1. Conceptualization + measurement

■ 2. Ethnic inequality: measurement

- see <https://growup.ethz.ch/>
- EPR-core: power relations in the central government + regional autonomy



Reflection

- How is this conceptualization + measurement of inequality different from the one used by Collier & Hoeffler? (slides 7-10)



1. Conceptualization + measurement

■ 3. Ethnic conflict

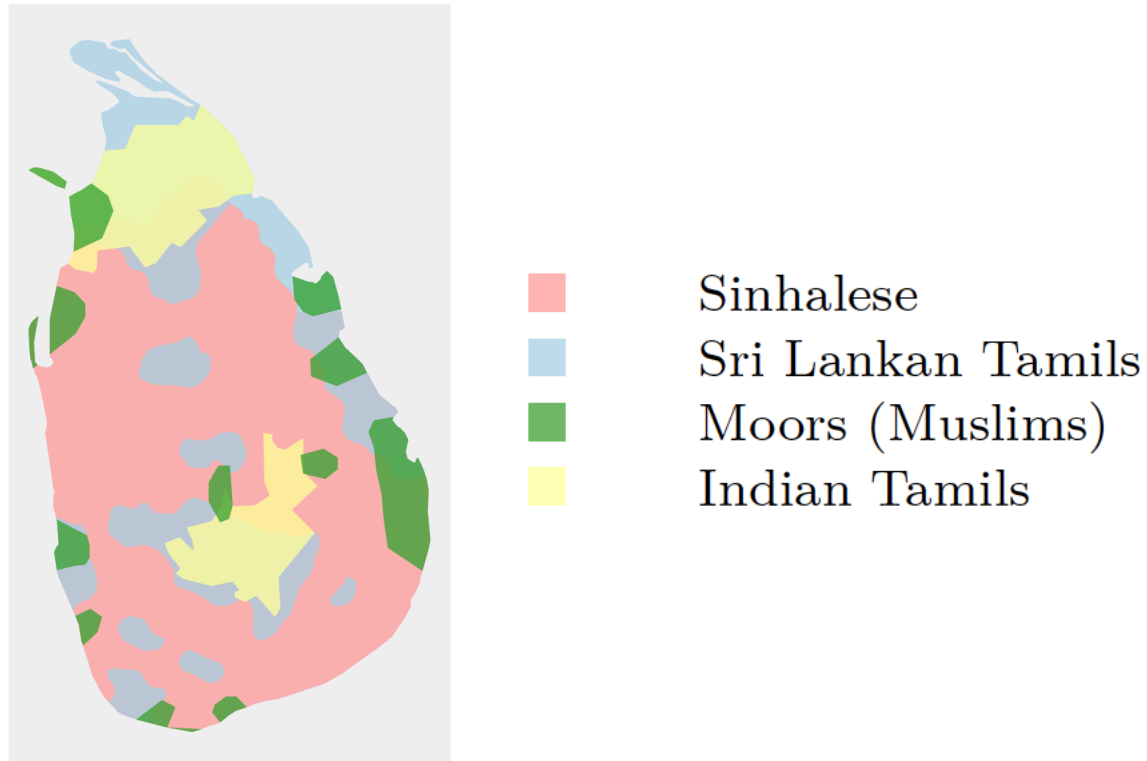
- ACD2EPR: civil war involvement of ethnic groups
- For each rebel organization (UCDP):
 - ethnic claims
 - ethnic recruitment
 - ethnic support

} ethnic civil war



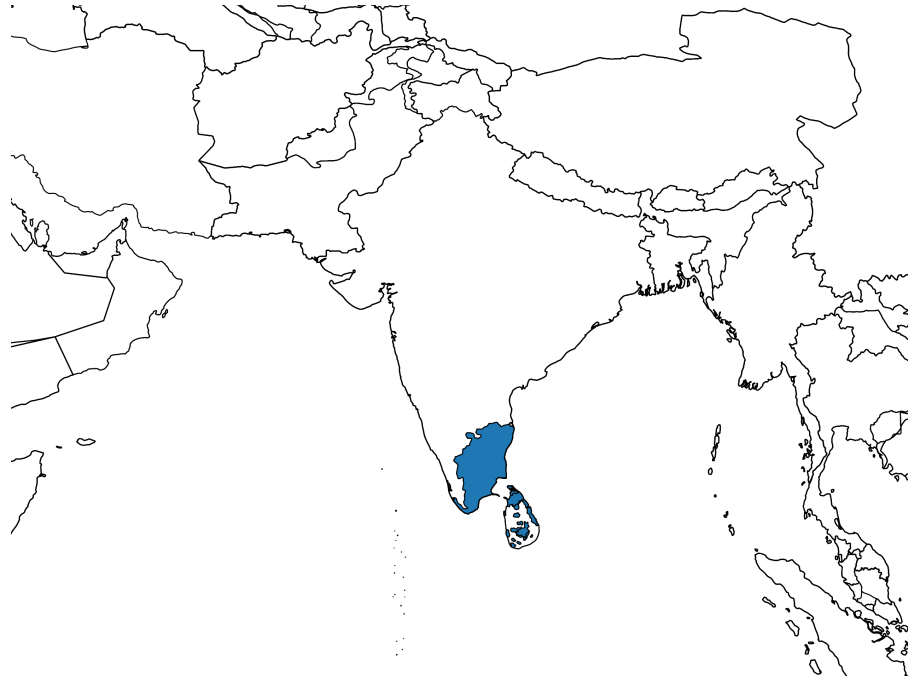
1. Conceptualization + measurement

- **4. Additional measures (1)**
 - GeoEPR: geo-coded settlement areas of all ethnic groups



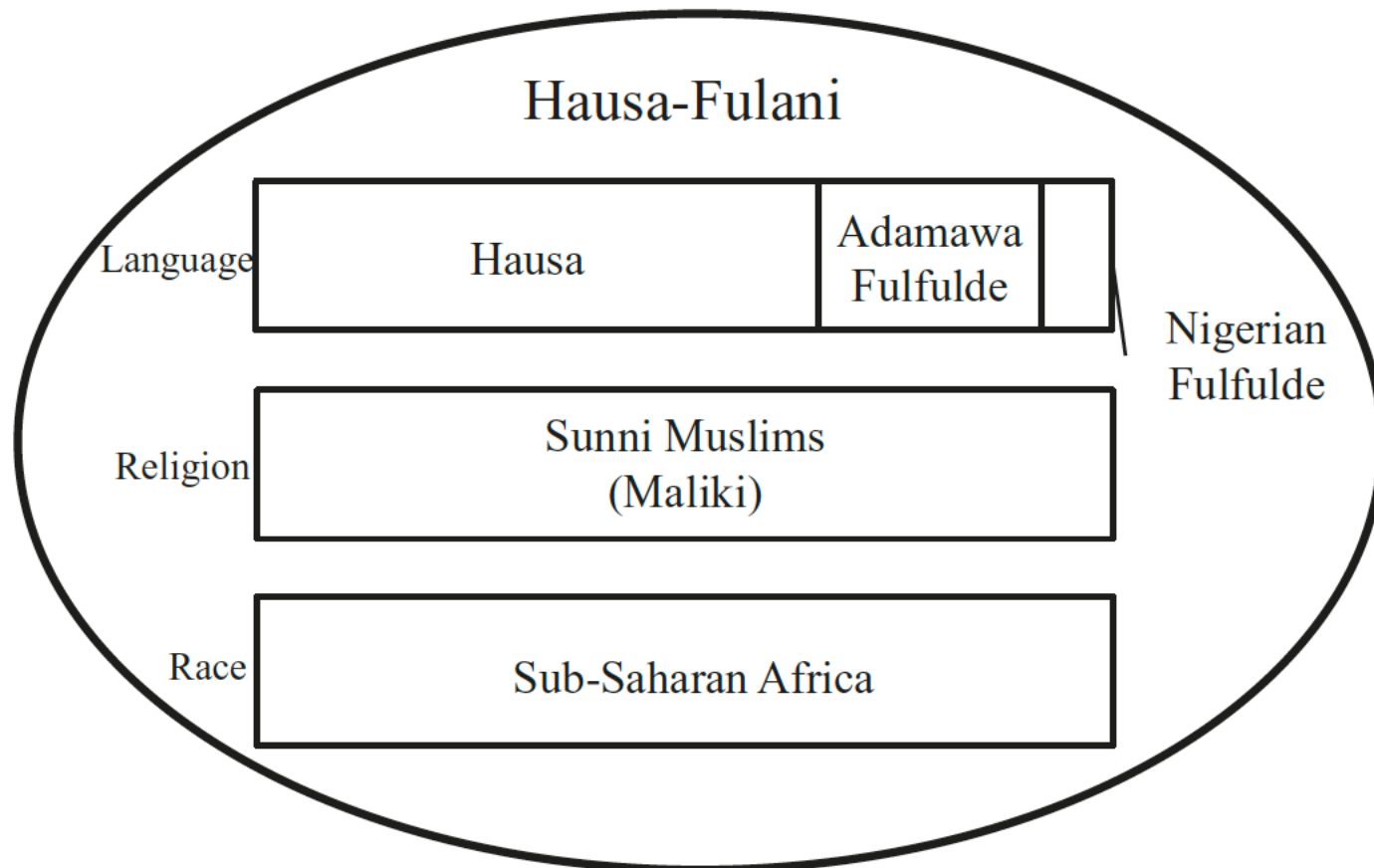
1. Conceptualization + measurement

- **4. Additional measures (2)**
 - EPR-TEK: Transnational ethnic kin groups



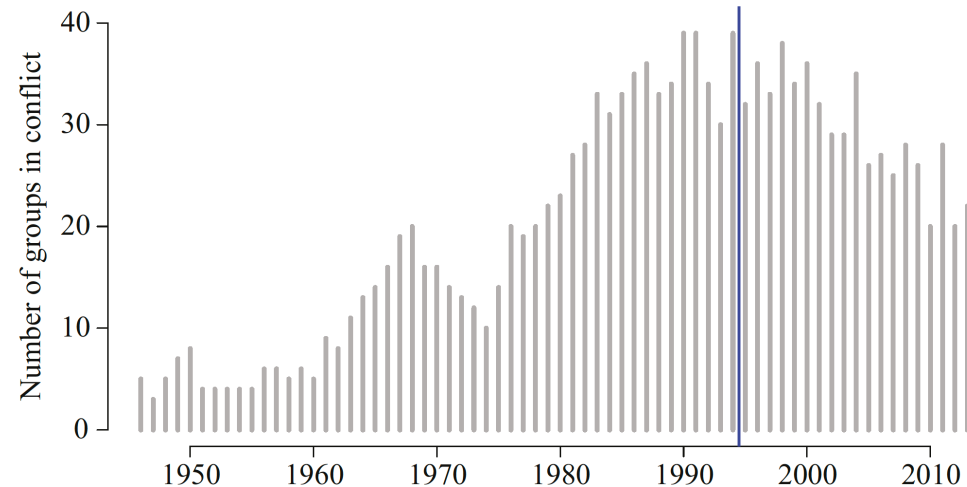
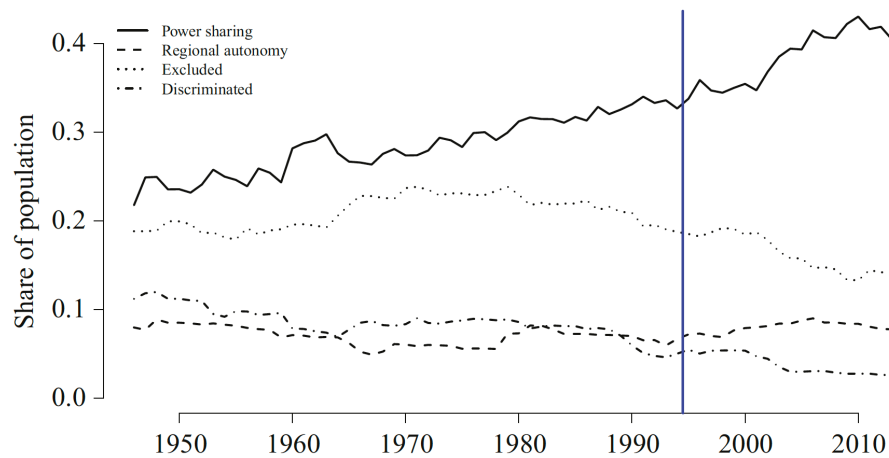
1. Conceptualization + measurement

- 4. Additional measures (3)
 - EPR-ED: Cultural cleavages



2. Descriptive analysis

- Summary of complex data with simple statistics (e.g., average power-sharing, number of ethnic civil wars)
- Clarifies general patterns which are difficult to see in raw data (and in single-case studies)



3. Quantitative analysis

Aim: make generalizable statements beyond analyzed cases

■ 1. Political inequality

- Political exclusion increases the risk of civil war onset (EPR-Core)

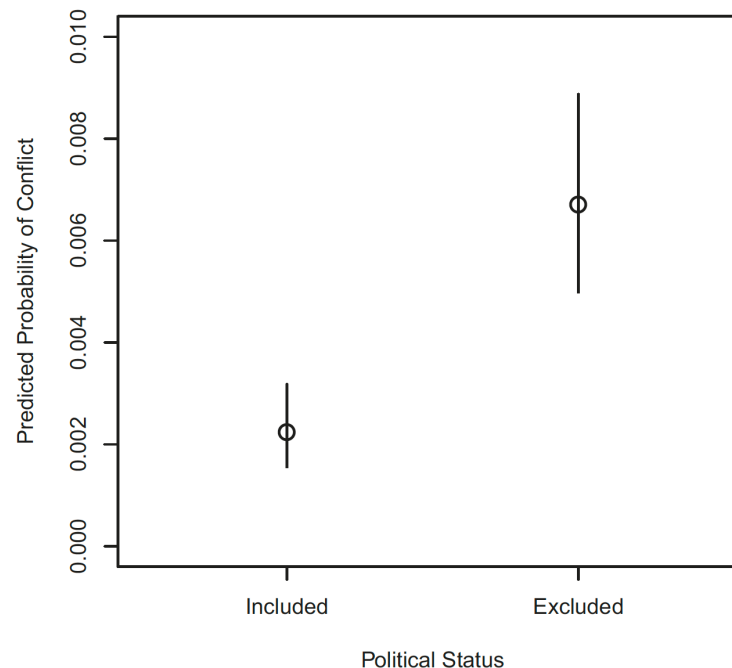
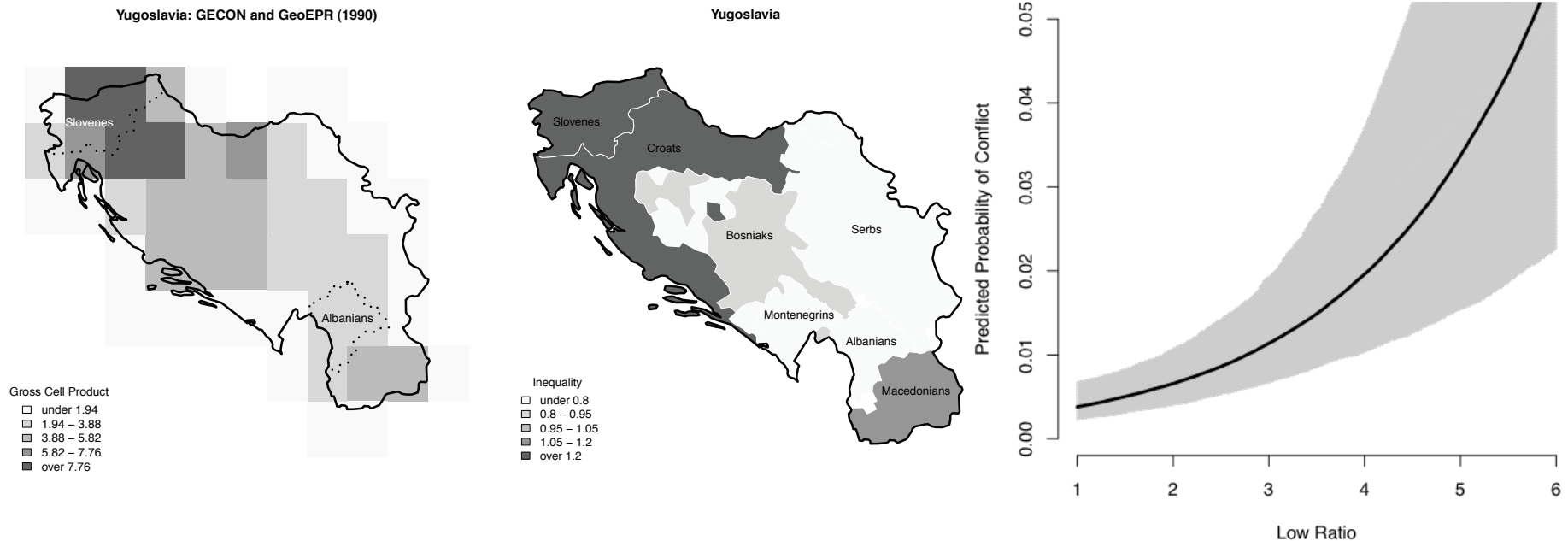


FIGURE 4.2. Comparing the effect of power access status on civil war onset.

3. Quantitative analysis

■ 2. Economic inequality

- Economic inequality increases the risk of civil war onset (Geo-EPR + regional GDP data)



Exercise

■ Using the EPR dataset

- Access the EPR dataset: <https://growup.ethz.ch/>.
- Find a case, where an ethnic group experiences a civil war onset. ★
- Access the description ('Read')
 - How are ethnic identities measured in the selected country and along which identity axes (language, religion, ...)?
 - What was selected group's power status when it became involved in civil war? (senior partner, junior partner, powerless, discriminated...)?
 - What did this mean concretely according to the description?
 - Was the group's power status connected to the civil war onset?



3. Opportunities and challenges of quantitative research on ethnic conflict

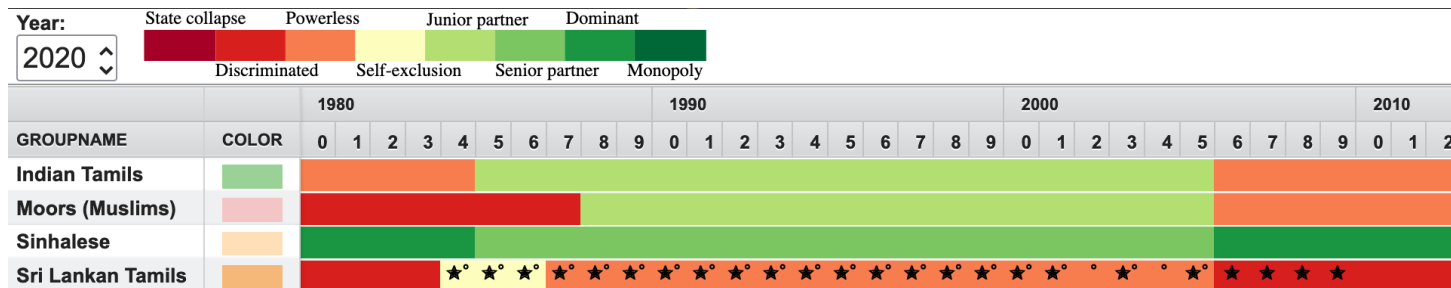
Own views and opinions?

- Opportunities/advantages of quantitatively studying conflict?
- Challenges/disadvantages?



1. Opportunities

- 1. Larger evidence base
 - Multiple, globally representative cases (states/groups, many years, ...)
 - Non-quantitative studies often have to focus on one or few select cases (e.g., only conflict cases).
 - Allows us to study slow-varying factors (e.g., ethnic diversity, ethnic power relations, institutions in general).



1. Opportunities

■ 2. Testing rival explanations

- Statistical analyses allow the use of **control** variables for **alternative explanations**.
- Often difficult or impossible in non-quantitative studies (too many factors for too few cases)
- E.g. analysis of ethnic inequality while controlling for other explanations (greed, opportunity)

TABLE 4.3. *Political Horizontal Inequality and Group-Level Onset of Civil War, 1946–2009*

	Model 4.1	Model 4.2	Model 4.3	Model 4.4
<i>Group-level variables:</i>				
Group Excluded	1.0958** (0.2020)		0.7086* (0.3042)	1.2182** (0.2446)
Rel. Group Size	1.1020** (0.3406)	1.3885** (0.3145)		1.3067** (0.3941)
Rel. Group Size (Excl.)			1.3578** (0.3978)	
Rel. Group Size (Incl.)			0.0860 (0.8046)	
Number of Previous Conflicts	0.6732** (0.0742)	0.5571** (0.0972)	0.6763** (0.0716)	0.7219** (0.0897)
<i>Country-level variables:</i>				
Ongoing Conflict, lag	0.6116* (0.2903)	0.3876 (0.2882)	0.6126* (0.2886)	0.6664* (0.2997)
GDP/capita, lag, log	−0.1982* (0.0809)	−0.1672* (0.0727)	−0.1885* (0.0822)	−0.2303* (0.0927)
Population, lag, log	0.0076 (0.0929)	0.0827 (0.0895)	0.0156 (0.0885)	0.0050 (0.0875)
Democracy, lag				0.6033 (0.4404)

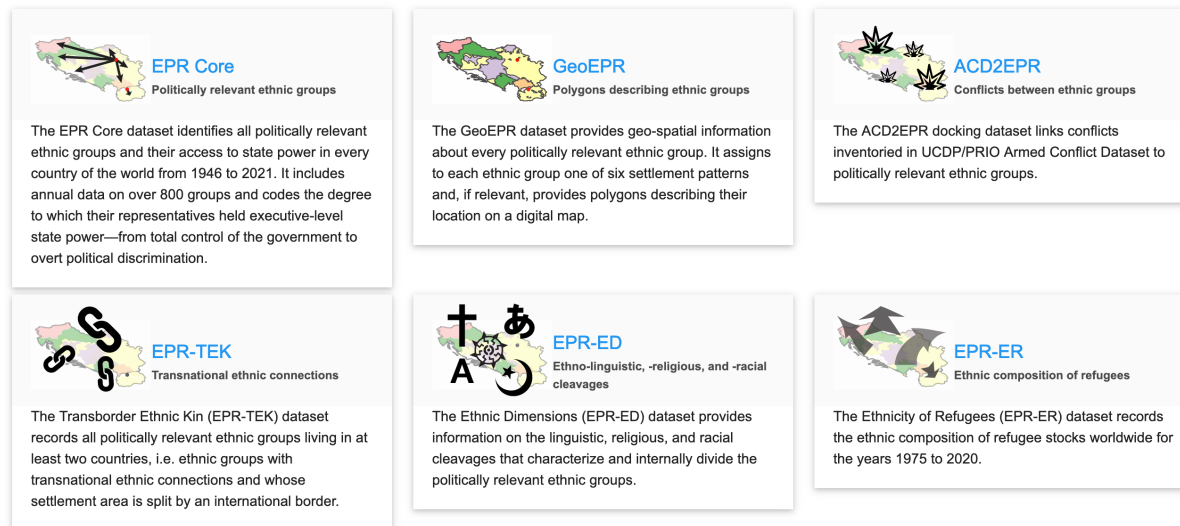
1. Opportunities

■ 3. Transparency and reproducibility

- Data **publicly** available
- Enables scrutiny, further use, and re-analysis by **others** (data criticism, different statistical models, other variables).

Ethnic Power Relations (EPR) Dataset Family 2021

The EPR Dataset Family provides data on ethnic groups' access to state power, their settlement patterns, links to rebel organizations, transborder ethnic kin relations, and intraethnic cleavages. The 2014 version has been introduced in [Vogt, Bormann, Rüegger, Cederman, Hunziker, Girardin \(2015\)](#) and has been updated in 2021 in a series of data sets on ethnicity that have stimulated civil war research in the past decade. It features a comprehensive system of tightly integrated data sets:



The EPR Dataset Family is complemented by the [EPR Aggregate Group \(EPR-AG\)](#) data.

Download

The Ethnic Power Relations Data: A Critique

Brenton D. Peterson^{*†}

March 8, 2016

Abstract

The Ethnic Power Relations (EPR) dataset (Cederman, Wimmer and Min (2010); Wimmer, Cederman and Min (2009a)) provides researchers broad cross-country data on ethnic groups' access to political power within their states. Since its inception, the data have been used extensively in the literature on intrastate wars, but they have also found use in studies of ethnic politics, elections and a variety of other topics. I argue that the EPR data is flawed for three primary reasons. First, ethnic groups in the data are often aggregated in ways that would be unrecognizable to country experts and group members themselves. Second, politically irrelevant ethnic groups are excluded from the data, but are often as politicized as those included. Third, groups' access to political power is often coded in ways that demonstrably contradict the coding rules themselves and do not match the empirical reality of ethnic power relations over time. I discuss these critiques in the context of the Kenyan data included in the EPR dataset, providing evidence for my claims from research in linguistics, history and anthropology, electoral results, new data on the ethnic background of Kenya's Cabinet ministers since independence, original survey data, and experimental behavioral measures of ethnic identification. The errors and lack of conceptual clarity in the EPR have significant consequences for studies using the data, at times introducing systematic biases and clouding our understanding of ethnic relations.

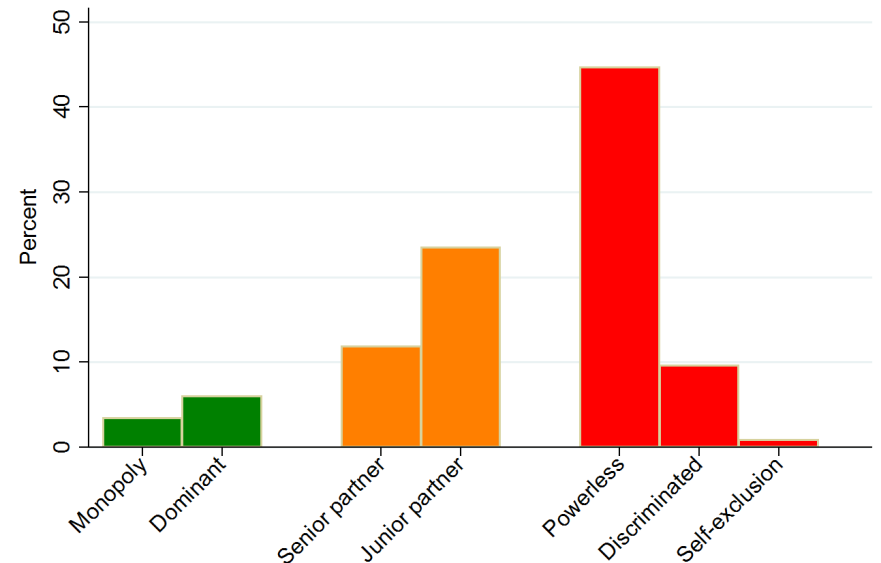
1. Opportunities

- **Ethnic inequality and civil war**
 - **Increasing consensus:** There is a relationship between (political + economic inequality) and civil war risk.
 - Visible thanks to measurement of ethnic **inequality** in political/economic terms at ethnic **group level** (as opposed to measures for demographic diversity and individual inequality)
 - Relationship can be shown thanks to global, representative sample that includes non-conflict cases.
 - Relationship remains, if we control for other explanations.
 - E.g. greed, opportunity (GDP per capita, oil revenue, group size...)
 - Often interplay between grievances and opportunity

2. Challenges

- 1. From concepts to numeric measurements
 - Often **crude** measurements, often contested
 - EPR: progress compared to demographic diversity measures, but...

Ethnic fractionalization	0.010 (0.006)*	0.011 (0.007)*	0.012 (0.008)
Religious fractionalization	-0.003 (0.007)	-0.006 (0.008)	-0.004 (0.009)
Polarization $\alpha = 1.6$	-3.067 (7.021)	-4.682 (8.267)	-6.536 (8.579)
Ethnic dominance (45–90%)	0.414 (0.496)	0.575 (0.586)	1.084 (0.629)*
Democracy	-0.109 (0.044)***	-0.083 (0.051)*	-0.121 (0.053)**
Peace duration	-0.004 (0.001)***	-0.003 (0.001)***	-0.004 (0.001)***
Mountainous terrain	0.011 (0.007)	0.007 (0.009)	-0.0001 (0.009)
Geographic dispersion	-0.509 (0.856)	-0.763 (1.053)	-1.293 (0.102)
Ln population	0.221 (0.096)**	0.246 (0.119)**	0.300 (1.133)**
Income inequality		0.015 (0.018)	
Land inequality			0.461 (1.305)



2. Challenges

■ 1. From concepts to numeric measurements

■ EPR limitations:

-
-
-

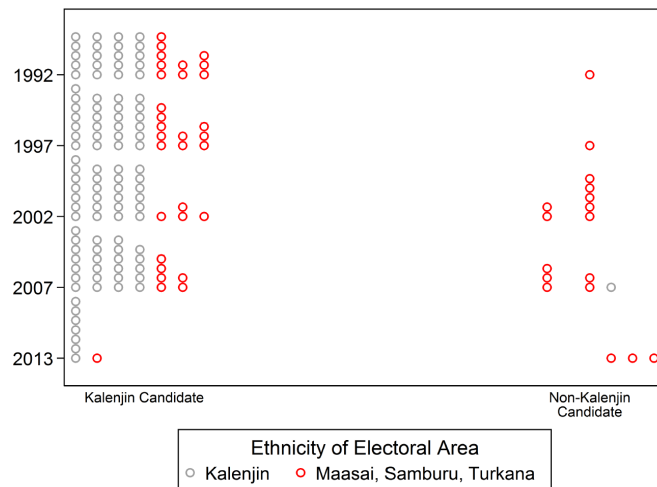


Figure 1: Candidates winning a majority, by the majority ethnic group of an electoral constituency (1992-2007) or county (2013). Kalenjin areas have always united behind a single candidate. Maasai, Samburu and Turkana have often gone their own way.
Source: The Kenya Election Database

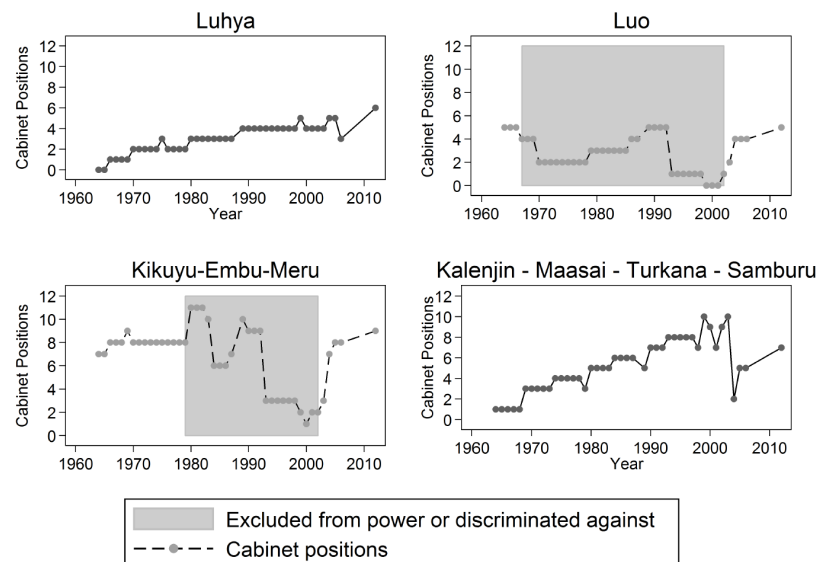
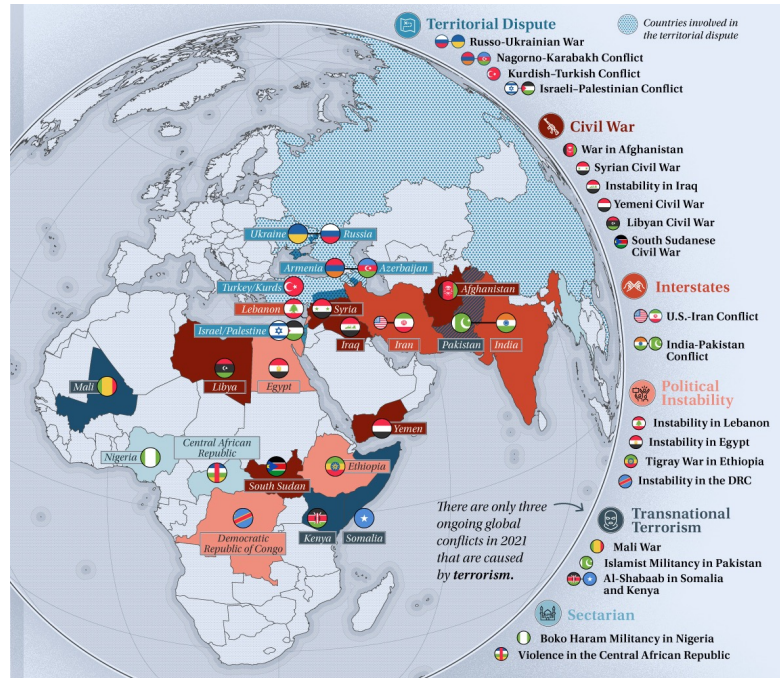


Figure 4: Cabinet posts by ethnic group over time.

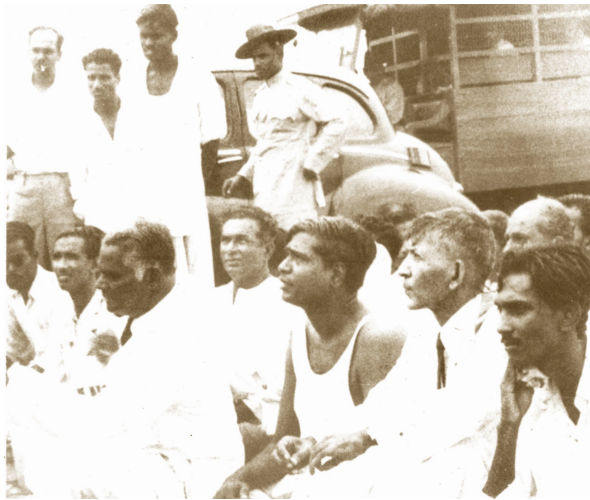
2. Challenges

- **2. Over-generalization of results**
 - Sometimes **unstable** results (measurement, model)
 - Sometimes results are driven by **single cases or regions** (greed, nationalism)
 - Temptation to generalize from such single cases



2. Challenges

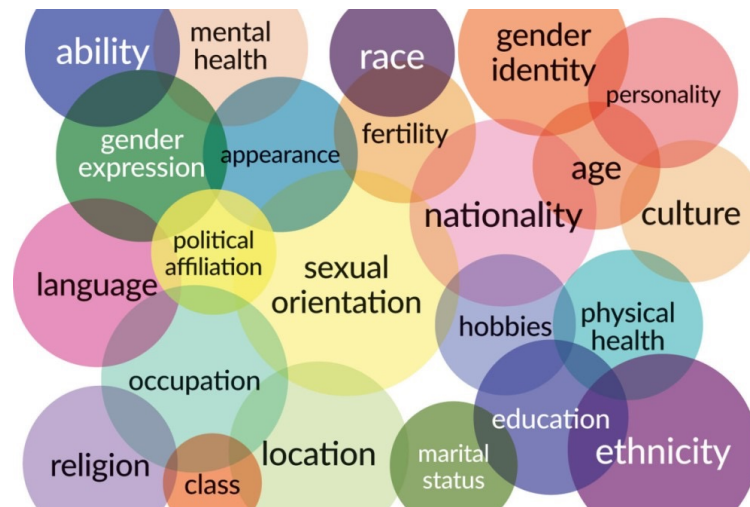
- **3. Difficulty to measure long-term escalation processes**
 - Beginning of conflict? Grievances, peaceful mobilization, low-level violence, civil war



Sri Lanka: Tamil protest (1950s), pogroms (1956), civil war (1983-2009)

2. Challenges

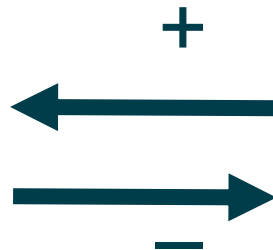
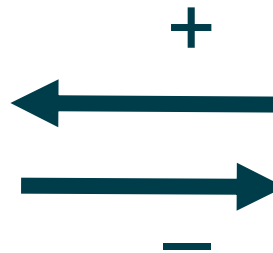
- **4. Difficulty measuring and controlling for central concepts**
 - E.g. limited information on political leaders or exploitation of specific social groups
 - Risk of systematically biased findings!



2. Challenges

■ 5. Endogeneity

- Civil wars (and the risk thereof) influence risk factors for civil war (e.g., economic growth, ethnic inequality)
- Risk of biased conclusions!



2. Challenges

- **5. Endogeneity**

- Civil war (risk) -> political representation of ethnic minority groups
- **Two possibilities:**
 - **Civil war risk -> Strategic inclusion of ethnic minorities:**
 - **Civil war risk -> Strategic exclusion of ethnic minorities:**



- **Debate:** Ethnic identity, inequality, and civil war?
- **EPR Dataset:** identification of 'politically relevant groups', representation in central government, autonomy, ...
- **Opportunities:** larger evidence base, testing alternative explanations, transparency
- **Challenges:** measurement, over-generalization, long-term escalation, no data for central concepts, endogeneity

4. Summary



5. Practical exercise

Instructions

In groups of 3-4 people:

- Think of a **quantitative application** (broadly) related to your research for which you could use the EPR dataset.
- Discuss:
 - **Which EPR data** you could use and how.
 - The **opportunities/advantages** of exploring this question quantitatively.
 - The **challenges/disadvantages** of doing so.
- Take notes and explain your thoughts to the plenum.