Power Mismatch and Civil Conflict: An Empirical Investigation

Massimo Morelli [1] Long Hong [2] Laura Ogliari [3]

[1] Università Bocconi & CEPR

[2] University of Wisconsin - Madison

[3] Università di Bergamo & CEPR

77th Economic Policy Panel Meeting – 20-21 April, 2023

Power Relations Introduction

- In International Relations longstanding debate on whether peace is more likely with **balance of power** or **preponderance of power** – always intended as *military* power.
- Recent literature (Cederman et al., 2010; Mueller and Rohner 2018) has also shown how exclusion from political power increases the chances of civil conflicts.
- Our research question: Does **power mismatch**, the asymmetry between different dimensions of power, matter for conflict (Herrera, Morelli and Nunnari, 2022)?

Introducing the key variables

- Consider a dominant group G and a group E considering whether to rebel, with the objective to appropriate the entire surplus S.
- Let p denote the political power of E, that is the share of surplus that E can enjoy in the status quo.
- Let m denote the probability that E has of winning a conflict against G and appropriating the whole surplus, S.
- Let c_E be the cost of war for the group.

Prediction

Let us start from the case $m > p$

- Expected utility for E from the war gamble: $U_E(war) = mS - c_E$
- Expected utility for E in the status quo: $U_E(s) = pS$
- We should expect war initiated by E if $U_R(war) > U_R(s)$ Given any ex ante uncertainty on c_E, represented by a distribution $F(\cdot)$ on the domain $[0, \infty)$, E rebels with probability $F((m-p)S)$

$\Rightarrow$ Incentive to rebel is increasing in $(m-p)$, which represents the *mismatch*.

Prediction

In the opposite case $m < p$

- G may have an incentive to start a (repression) conflict if $c_G < [(1-m)-(1-p)]S$
- Conflict exist with the corresponding probability that c_G is less than $G\left((p-m)S\right)$, where $G(\cdot)$ denotes the cumulative probability distribution of the possible realizations of c_G.

$\Rightarrow$ incentive to start a (repression) is increasing in $(p-m)$

Main prediction: Conflict is more likely to happen when $|m-p|$ is high.

This paper

We try to bring the theory of power war to the data and test it on civil (ethnic) conflict

- We construct a new data set that combines data on conflicts, military power and political power at ethnic group level.
- We show that the mismatch is always significant explanatory variable
- We test the relation between power mismatch and conflict type (big vs small, centrist vs secession).

Outline

- Data Construction
  - Actors
  - Dependent variable
  - Ethnicity of the government and rebel group(s)
  - Political and military power measures
- Empirical results
- Conclusions

The relevant players

- We exploit ethnicity to identify groups and we restrict our attention to Africa (and the Middle East).
- We use the list of Ethnic groups of Ethnic Power Relations (EPR) Dataset.

pros Ethnic groups are defined according to the ethnic categories most salient for national politics in each country $\Rightarrow$ politically relevant groups.

We have a measure of political power.

cons Groups are "big".

Relevant ethnic groups may change over time

Conflict Data

- As source of conflict data we use the UCDP GED dataset
  an event to be included needs to satisfy:
  1. use of armed force,
  2. organized actor (i.e., government, organized groups),
  3. result in at least 1 direct death in a specific location and date.

  Advantages:
  - it includes also "small" conflicts events,
  - info on the location (latitude and longitude) ,
  - estimates of fatalities borne **by each side** in a conflict.
- We Use conflict data for 2 purposes:
  - linking rebel to ethnic groups,
  - build the dependent variable

Ethnicity of the government groups

	Political power rank in the EPR Core dataset		
	Rules Alone	Share power	Excluded from power
Rank	7 and 6	5 and 4	1, 2, and 3

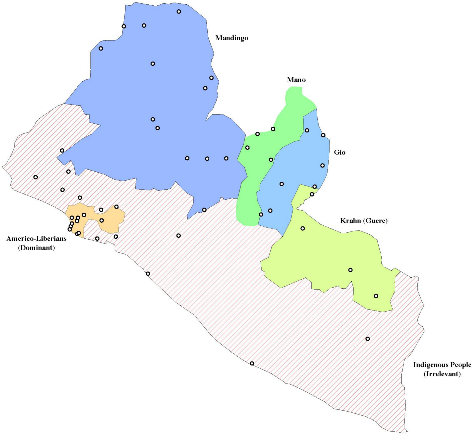
- if $\exists!$ 1 group with the highest power rank $\Rightarrow$ government group
- if $\exists$ more than one group with highest power rank (10% of the observations in Africa) $\Rightarrow$ manually check:
  - if allied$\Rightarrow$ we assign the government to all these groups and consider them as one entity.
  - if not allied $\Rightarrow$ try to determine which group as a larger advantage, if impossible drop the observation

Ethnicity of the Rebel groups

Similar to Michelapoulous and Papaioannu (2015), and Moscona, Nunn and Robinson (2018), we exploit the location of the conflicts events (in UCDP GED) and the location of the homeland of ethnic groups (in GEO-EPR).

1. Use the conversion table ACD2EPR developed by Vogt et al. (2015) which integrates UCDP/PRIO Armed Conflict Dataset with EPR (30.2% of conflicts).
2. For the rest of the sample:
    2.1 We keep all the conflicts against other (national) organized actors and exclude events occurring in the homeland of the government/irrelevant ethnicities.
    2.2 We count the number of times a rebel group has a conflict event in the homeland of a particular EPR ethnic group;
    2.3 We assign the ethnicity with the highest count,at least 3,to the rebel group (if ties, highest fatalities).

Ethnicity of the Rebel groups: Liberia example



Liberians United for Reconciliation and Democracy (LURD)

- 13 events in the Mandingo
- 4 events in the Mano
- 7 events in the Gio
- 4 events in the Krahn
- 10 events in the in the irrelevant group (exluded)
- 9 events in the dominant group (excluded)

$\Rightarrow$ We attribute LURD to the Mandingo Ethnicity.

Dependent Variable

Armed with the link between ethnic and rebel groups, we can assign conflicts a rebel group is involved in to the corresponding EPR ethnic group.

- We focus on conflicts between rebel groups and government forces,
- we consider a group involved in conflict in year t if it has at least one event in that year,
- we compute the number of fatalities borne by each side as the sum of the fatalities in all events in year t

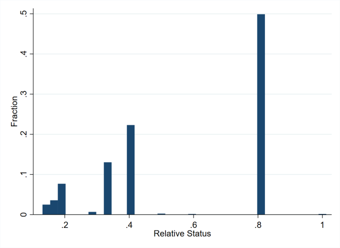
Political measures: Discrete index

We build two measures of political power of the ethnic group vis à vis the government group.

1. We use the EPR Power Rank and define

$$p_{eg}^{PR} = \frac{C_e^{PR}}{C_g^{PR}}$$

Government power rank	Rebel power rank: 1	2	3	4	5	Total
5	254	738	4	1,652	2	2,650
6	118	431	8	0	0	557
7	82	22	0	0	0	104
Total	454	1,191	12	1,652	2	3,311

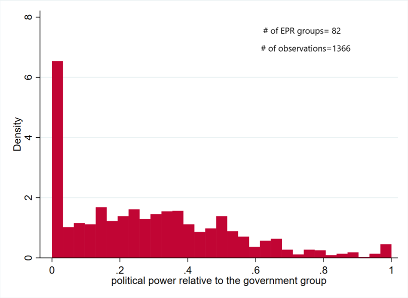


Political measures: Ethnicity of the Cabinet members

2 We follow Francois, Rainer and Trebbi (2015)and use participation in the government as a proxy of political power:

- We collect ethnicity of cabinet members for 14 countries in Sub-Saharan Africa in the period 1992-2012 (details);
- We define relative political power as

$$p^t_{r,c} := \frac{n^t_{r,c}}{n^t_{g,c}}$$



Power Rank	Relative political power (cabinet members)	Mean	Standard Dev.
	Dominant	0.532	0.188
	Senior partner	0.259	0.142
	Junior partner	0.143	0.125
	Powerless	0.092	0.098
	Discriminated	0.068	0.063

Military measure: overview

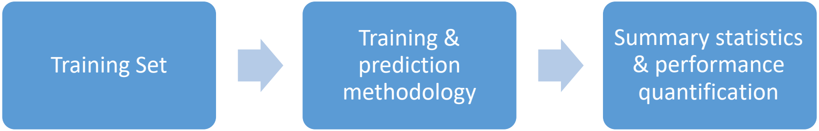
- We do not have direct information on the military power of each ethnic group.
- We approximate a group military power (relative to the country's group in power) with the predicted the probability of winning a conflict against the government.

BUT Estimating the probability of winning a conflict poses several challenges!

  - we do not have any conflict information for groups that never experienced a conflict;
  - there is little information (not military power, nor probability of winning, often not even the outcome of the conflict) for groups that experienced a conflicts in the past;
  - There is a wealth of information at the ethnic group level but relatively few conflicts.

Predicting military power via machine learning

- We rely on a machine learning technique inspired by Carroll and Kenkel (2020).
- We use an extended sample of observed conflicts to predict the probability of winning for all the ethnic rebel groups against government in our sample $\Rightarrow$ The probability of winning is defined dyadically.
- We use a rich set of observed ethnic group-level variables as predictors to infer the probabilities of victory for all potential conflicts between every ethnic (rebel) group and the government.



Training set

Outcome

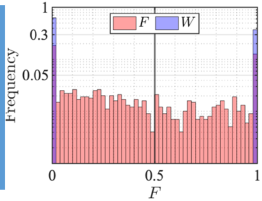
UCDP-GED

Africa and Asia

Use the number of fatalities to predict victory

$$Y := I\{F^t_{gr} \geq 0.5\}$$

$$F^t_{gr} = \frac{f^t_g}{f_{g^t} + f^t_r}$$



Predictors

Group

Land use, Geographic & Climate, Socio-demographic, Kinship, External support

Country

Latitude, Longitude, Regional Dummyes

Observations	N. of Predictors	Share of government wins
726	198	64%

The algorithm

Random Forest

- *Ensemble learning* method
- Bootstrap aggregating
- Each tree is a *weak learner* built independently on a subset of rows
- Intuition: averaging a set of observations reduces *variance*
- Majority vote
- The trees are further *decorrelated* by sampling the predictors.

Boosted Decision Trees

- *Ensemble learning* method
- Boosting
- Trees are built in a *forward* manner: *new* weak learner to overcome shortcomings of *existing* weak learners
- Trees are fitted on the residuals
- *Learning rate:* how strong will the residuals be updated
- Intuition: learning slowly is better than “once shot” learning.

Generalised Linear Models

- Classification models for outcomes following exponential distributions.
- To reduce the variance: use penalisation
- Two types of penalisation: ℓ_1 (LASSO) and ℓ_2 (RIDGE)
- Elastic net regularisation: combination of ℓ_1 and ℓ_2
- Intuition: ENT allows for variable selection but less “sharp” than ℓ_1.

Super learner

- Use Cross-validation to have a “testing” set $\Longrightarrow$ CV-error $\neq$ “training” error.
- Optimal weight for each learner that reduces the CV-error.

Details

Military measure: Performance

- Performance metric for a binary classification model:
  - Cross-Validated Log-Loss (perfect model Log-Loss=0)
  - Proportional reduction in CV Log-Loss

$$\mathrm{PRL}\left(\mathcal{M}^*_{(Y,X)}\right) = \frac{\mathrm{L}_{null} - \mathrm{L}\left(\mathcal{M}^*_{(Y,X)}\right)}{\mathrm{L}_{null}}.$$

Table: Algorithm's predictive power.

	CV Log-loss	PRL	Accuracy	Δ_{Null}(Accuracy)
Full model	0.554	15%	70.2%	9.1%
Population ratio	0.650	0.2%	64.4%	0%
Night light ratio	0.646	0.8%	65.1%	1.1%

Distribution

Robustness: outcome definition

Robustness: parameters

Variable relevance

Constructing mismatch

Using our estimates of $P(win)_{i,t}$ and of political power we can create two proxies of a group mismatch:

$$M^D_{e,g} = \begin{cases} 1 & if \ \ (p^{PR}_{eg} \leq \bar{p}^{PR}_{p50} \wedge m_{eg} > \bar{m}_{p66}) \vee (p^{PR}_{eg} > \bar{p}^{PR}_{p50} \wedge m_{eg} \leq \bar{m}_{p33}) \\ 0 & otherwise \end{cases}$$

$$M_{e,g} := |m_{e,g} - p_{e,g}|$$

Maps

Conflict incidence and mismatch

	Dep. Var.: Conflict incidence						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
M dummy	0.215***	0.231***	0.217***	0.230***	0.193***	0.141***	0.213**
	(0.0338)	(0.0362)	(0.0382)	(0.0359)	(0.0436)	(0.0385)	(0.100)
Observations	3,313	3,217	3,313	3,217	2,401	2,392	2,388
R-squared	0.622	0.648	0.530	0.555	0.574	0.754	0.777
Controls							
Peace years	✓					✓	✓
EPR family		✓				✓	✓
Natural resources			✓			✓	✓
Geographic controls x trend				✓		✓	✓
Pre-sample Economic Controls x trend					✓	✓	✓
Group Inequality					✓	✓	✓
Fixed effects							
Country-year FE	✓	✓	✓	✓	✓	✓	✓
Group FE							✓

Robustness on geomatching
Continuous Measure

Other Results

Furthermore we find that:

- The relationship between mismatch and conflict is convex (Non linearity results)
- Conflict participation is more likely when $m > p$ (One-sided mismatch results)
- Mismatch is correlated with "centrist" rather than territorial conflicts (Centrist vs territorial results)
- Mismatch is correlated with "big" rather than small conflicts (Big vs small results)

Discussion

Aim of the paper is mainly descriptive. Still, interesting to discuss three points.

- **Reverse causality** $\rightarrow$ show that mismatches rises before conflict start (Event study)
- **Forward looking behavior** $\rightarrow$ estimated military power is extremely persistent over time. We estimate the "structural military power".
- **Is mismatch different than the exclusion from political power?** $\rightarrow$ look at 30 cases of political "downgrading"

Table: Political power downgrading

	Frequency	Rank before Downgrade		Δ Rank		Military Power	
		mean	median	mean	median	mean	median
Conflict in the next 5 years	10	3.8	4	-2	-2	0.638	0.742
No Conflict in the next 5 years	20	3.95	4	-2	-2	0.393	0.275

Conclusions

- We build a new dataset at the ethnic-group level which combines information on conflicts and measures of political power and military power.
- We provide the first estimate of the military power of an ethnic group using machine learning techniques.
- We provide evidence that mismatch and conflict are positively related and that:
  - the relationship seems non-linear
  - power mismatch seems more relevant for centrist conflict and for "big" conflicts.

From a policy perspective

- We need to pay attention to the imbalance between different dimensions of (relative) power.
- Focusing just on military strength or economic or political power may be misleading: militarily strong groups may not be those who start a war if they have enough political power; groups that are discriminated against may not pose a threat if they are militarily weak.

Thank you!

Political measure: Ethnicity of the Cabinet members 2

- For years 1992-2004 we used the data from Francois, Rainer and Trebbi (2015) and converted everything in EPR groups.
- For years 2005-2012 we collected the data on cabinet membership from the C.I.A.'s "Chiefs of State and Cabinet Members of Foreign Government" and then assign to each minister an ethnic identity by using
  - direct information on the ministry ethnicity (or her parents ethnicity)
  - the location of the place of birth of the minister, when this was not possible we employed the location of the primary school, the district of election.
- There is some attrition: out of 2696 members of the cabinet we manage to attribute an ethnicity to 2537 (94.1%)

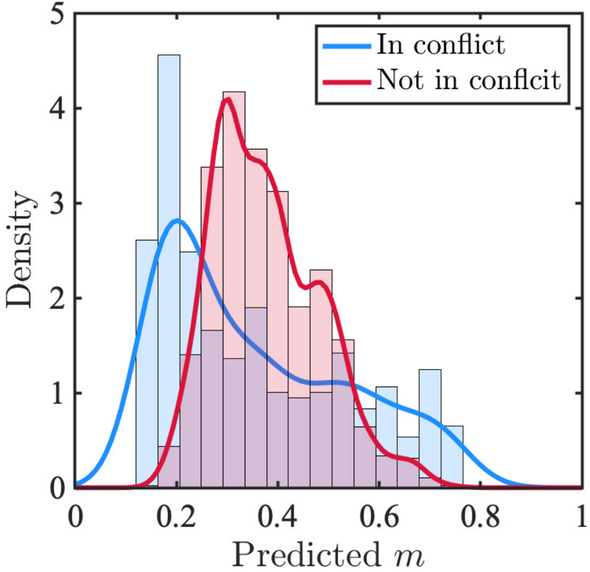
(Back)

The algorithm

- In order to *predict* the dyadic probability of winning, we use binary learning model based on the training data (X, Y).
  1. First, the *training* is the procedure by which we construct an optimal model $\mathcal{M}^*_{X,Y}$.
  2. Second, the *prediction* is the use of the model $\mathcal{M}^*_{X,Y}$ to determine the $m^t_{r,g}$ probability of victory of group r against the government g in year t.
- Main challenge: increasing the accuracy without over-fitting!
- Solution: use algorithms that perform variable selection through some *penalisation* technique.
- We rely on a *mix* of (1) tree-based models and (2) generalised linear models to determine $m^t_{r,g}$.
- The compounding of these models is handled by a *Super Learner* algorithm.

Back

Predicted military power distribution



Back

Robustness 1: Outcome definition

We validate our definition of victory using the intrarstate conflicts (163 out of 574) in the COW database.

- Fatalities ratio threshold at 0.5: accuracy = 77%

		True rebel win: Positive	True rebel win: Negative	
Fatalities' rebel win	Positive	34	30	
	Negative	8	90	
		81%	75%	77%

- Fatalities ratio threshold at 0.583 (optimal): accuracy=82%

		True rebel win: Positive	True rebel win: Negative	
Fatalities' rebel win	Positive	28	15	
	Negative	14	106	
		67%	88%	82%

Back

Robustness 2: algorithm parameters

- $\boldsymbol{S}_1$: # of cross-validation folds is changed from 6 to 10;
- $\boldsymbol{S}_2$: #(CV-folds) from 6 to 10; Random Forests' (RF) depth and span is increased; RF column sampling is also increased;
- $\boldsymbol{S}_3$: #(CV-folds) from 6 to 10, Boosted Decision Tree's (BDT) learning rate increased;
- $\boldsymbol{S}_4$: RF column sampling is decreased;
- $\boldsymbol{S}_5$: RF column sampling is increased; BDT learning rate increased; BDT total number of allowed trees decreased;
- $\boldsymbol{S}_6$: RF & BDT row sampling rate is increased.

Table: Robustness check: algorithm's parameters.

Models	$\boldsymbol{S}_1$	$\boldsymbol{S}_2$	$\boldsymbol{S}_3$	$\boldsymbol{S}_4$	$\boldsymbol{S}_5$	$\boldsymbol{S}_6$
$\mathrm{corr}(m_{\boldsymbol{S}_0}, m_{\boldsymbol{S}_i})$	> 0.99	> 0.99	> 0.99	> 0.99	0.98	0.97
$\Delta_{\mathrm{PRL}}^{\boldsymbol{S}_o,\boldsymbol{S}_i}$	0.01	0.02	0.03	0.03	0.04	0.05

Back

Unboxing the "black box"

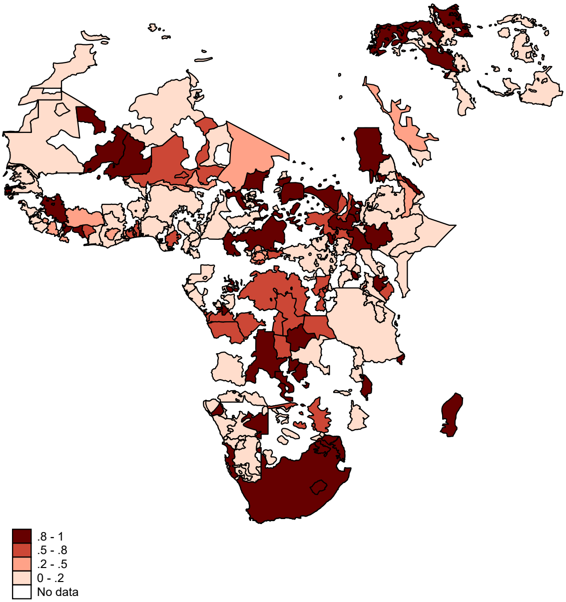
- ML has the disadvantage of not being able to know how important are the variables for the prediction
- Rerun our algorithm multiple times, each time removing a set of variables
- Intuitively, the higher the PRL loss, the higher the importance the variables for prediction in the original model

Table: Variables' predictive power.

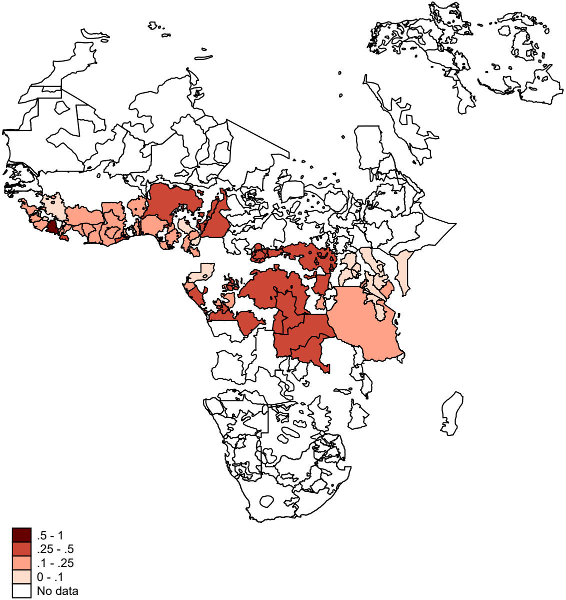
	Ext	Geo	Pop	PyWh	Tek	Land	Country
PRL loss (%)	3.1	1.9	2.3	5.6	1	1.6	0.5

Back

Mismatch Dummy



Mismatch Continuous Measure



Back

Conflict incidence and Mismatch: Robustness

Dependent variable: conflict incidence

	Mismatch Dummy (se)	R^2	Obs.	Mismatch Dummy (se)	R^2	Obs.
	(1)	(2)	(3)	(4)	(5)	(6)
Less than 50% of matches identifyed by 3 (or less) events	0.0727*** (0.0154)	0.601	3,740	0.0425* (0.0232)	0.734	3,740
Less than 50% of matches identifyed by 5 (or less) events	0.0835*** (0.0169)	0.612	3,544	0.0429* (0.0253)	0.744	3,544
No match identifyed by 5 (or less) events	0.0846*** (0.0179)	0.544	3,327	0.0541** (0.0273)	0.683	3,327
No geomatching (ACD2EPR only)	0.0392*** (0.0116)	0.530	3,612	0.0343* (0.0201)	0.693	3,612
No geomatching - EPR incidence	0.0410*** (0.0101)	0.500	3,453	0.0384** (0.0175)	0.726	3,453
Model fixed effects	Country $\times$ Year			Ethnic group & Country $\times$Year		

Back

Conflict incidence and Mismatch: Continuous Measure

Dependent variable: conflict incidence.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mismatch dummy	0.0794***	0.0553**	0.0534**	0.0705**				
	(0.0255)	(0.0223)	(0.0254)	(0.0313)				
Mismatch cont.					0.253***	0.158**	0.161**	0.166**
					(0.0687)	(0.0661)	(0.0771)	(0.079)
Observations	1,247	1,247	995	1,247	1,247	1,247	995	1,247
R-squared	0.232	0.481	0.571	0.615	0.234	0.481	0.571	0.615
Controls								
Peace years		✓	✓	✓		✓	✓	✓
Family		✓	✓	✓		✓	✓	✓
Natural resources		✓	✓	✓		✓	✓	✓
Geographic			✓				✓	
Socio-ecomic			✓				✓	
Fixed effects								
Country $\times$ year	✓	✓	✓	✓	✓	✓	✓	✓
Ethnic group				✓				✓

Back

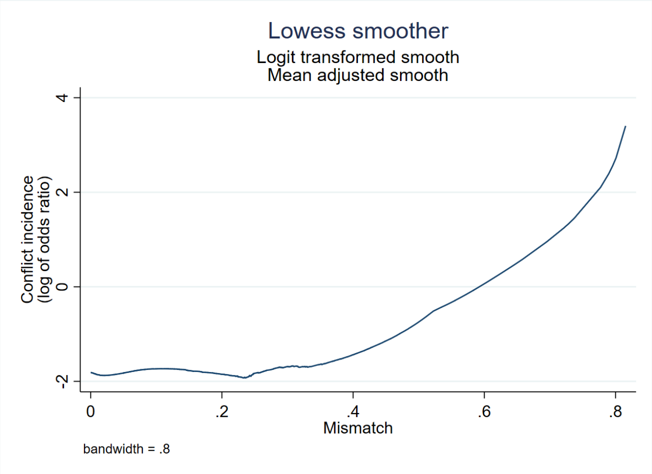
Non-linearity of the effect

Dependent variable: conflict incidence	Below median		Above median		Whole sample	
	(1)	(2)	(3)	(4)	(5)	(6)
Mismatch cont.	-0.487*	-0.460	0.434***	0.406**	-0.356**	-0.487**
	(0.262)	(0.281)	(0.128)	(0.180)	(0.147)	(0.191)
Mismatch squared					0.846***	1.137***
					(0.272)	(0.331)
Observations	541	530	564	550	1,226	1,225
R-squared	0.558	0.785	0.539	0.636	0.490	0.623
Fixed effects						
Country $\times$ year	✓	✓	✓	✓	✓	✓
Ethnic group		✓		✓		✓

Back

Non-linearity of the effect

Figure: Non-parametric local regression



Back

One-Sided Mismatch

Dependent variable: conflict incidence

	(1)	(2)	(3)	(4)	(5)	(6)
Mismatch dummy military	0.0733***	0.0787**	0.0787**	0.0952*		
	(0.0239)	(0.0337)	(0.0329)	(0.0511)		
Mismatch dummy political	0.0480***	0.00311	0.0334	0.0466		
	(0.0140)	(0.0260)	(0.0203)	(0.0355)		
Mismatch cont., m>p					0.248***	0.220**
					(0.0786)	(0.0872)
Mismatch cont., p>m					-0.0222	0.0663
					(0.0885)	(0.128)
Observations	4,260	4,260	1,247	1,247	1,247	1,247
R-squared	0.574	0.708	0.482	0.615	0.489	0.616
Fixed effects						
Country × year	✓	✓	✓	✓	✓	✓
Ethnic group		✓		✓		✓
Sample	Full	Full	Restricted	Restricted	Restricted	Restricted

Back

Centrist vs territorial conflicts

Conflict incidence:	Centrist (1)	Terrirorial (2)	Centrist (3)	Terrirorial (4)	Centrist (5)	Territorial (6)	Centrist (7)	Territorial (8)
Mismatch dummy	0.0583***	0.0151	0.0535**	-0.0136				
	(0.0139)	(0.00965)	(0.0209)	(0.0148)				
Mismatch cont.					0.151**	0.0210	0.147*	0.00147
					(0.0672)	(0.0222)	(0.0771)	(0.0247)
Observations	4,110	3,914	4,110	3,914	1,235	1,157	1,235	1,157
R-squared	0.546	0.449	0.700	0.752	0.451	0.351	0.614	0.433
Fixed effects								
Country $\times$ year	✓	✓	✓	✓	✓	✓	✓	✓
Ethnic group			✓	✓			✓	✓

Back

Big vs Small conflicts

Conflict incidence	Big (1)	Small (2)	Big (3)	Small (4)	Big (5)	Small (6)	Big (7)	Small (8)
Mismatch dummy	0.0516***	0.0175*	0.0444**	0.0107				
	(0.0143)	(0.00909)	(0.0222)	(0.0142)				
Mismatch cont.					0.131*	0.0527	0.0622	0.133
					(0.0676)	(0.0419)	(0.0578)	(0.0853)
Observations	4,150	3,883	4,150	3,883	1,220	1,174	1,174	1,220
R-squared	0.534	0.416	0.698	0.550	0.440	0.327	0.444	0.606
Fixed effects								
Country × year	✓	✓	✓	✓	✓	✓	✓	✓
Ethnic group			✓	✓			✓	✓

Back

Event Study

Figure: Power Mismatch Evolution

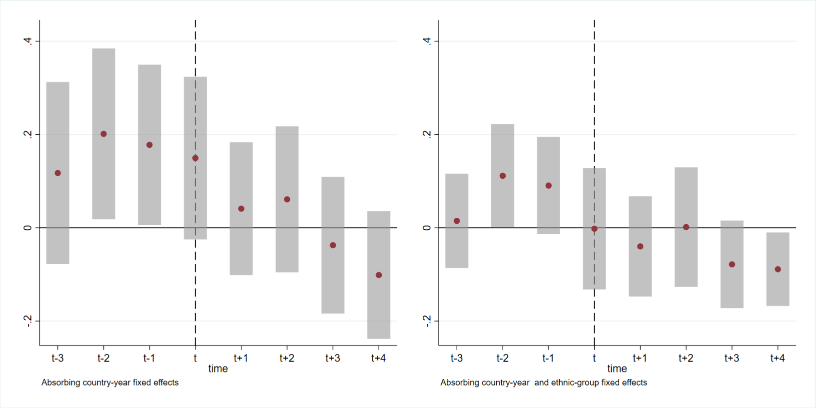


Figure: Mismatch dummy

Back