Subject to Change
Quantifying Transformation in Armed Conflict Actors

Margaret J. Foster

Prepared for TargetRWE

February 20, 2024

Duke University

Introduction

Computational social scientist who specializes in finding and solving tricky measurement problems

Technical skills:

- R, Python, Git, Markdown

Data and methods that I have worked with:

- Text (NLP, classification)
- Event Data (linear, survival models)
- Networks (inference, simulation)

Outline

Outline of talk

- Motivating problem
- Design and implementation of a solution
- Statistical applications: validation
- Statistical applications: research contribution
- Conclusion and additional projects

Substantive Problem

Problem: How to systematically quantify a changing state that we can not observe directly but for which we have theoretical expectations?

Constraints: The solution should be

- Dynamic
- Granular
- Interpretable
- Replicable
- Scalable
- Validatable

Strategy

- Accept that we won't know the internal state of the organization.
- Instead, model trends in event data, clustered at the group level.
- Given these constraints, what can we access and what can we do with it?

Data Goal and Starting Point

Starting point:

```
> print(head(ged))
      id              relid year active_year code_status type_of_violence conflict_dset_id conflict_new_id
1 244657 IRQ-2017-1-524-322 2017           1       Clear                1              259             259
2 132140   AFG-1989-1-411-2 1989           1       Clear                1              333             333
3 130364  AFG-1989-1-411-37 1989           1       Clear                1              333             333
4 130359   AFG-1989-1-411-4 1989           1       Clear                1              333             333
5 133883  AFG-1989-1-411-39 1989           1       Clear                1              333             333
6 133874   AFG-1989-1-411-6 1989           1       Clear                1              333             333
             conflict_name dyad_dset_id dyad_new_id
1       Iraq: Government          524         524
2 Afghanistan: Government          724         724
3 Afghanistan: Government          724         724
4 Afghanistan: Government          724         724
5 Afghanistan: Government          724         724
6 Afghanistan: Government          724         724
```

```
source_article
1                                                                "Agence France Presse,2017-08-01,Attackers targe
t Shiite mosque in Afghanistan's Herat";"Agence France Presse,2017-08-01,At least 20 killed in Shiite mosque attack in Afghani
stan's Herat";"Pajhwok News,2017-07-31,Assailants among 6 killed in Iraq Embassy attack"
2
The Times 13 Jan 1989 "Missiles and tea breaks for rebels;Afghanistan
3
R 18 Jan 1989 "KABUL REPORTS HUNDREDS OF REBEL CASUALTIES IN KUNDUZ PROVINCE.
4 Reuters 24 Jan 1989 "AFGHANS KILL NEARLY 400 REBELS, TASS SAYS." /The Washington Post 27 Jan 1989 "Rebels Send Food Convoy T
o Kabul;U.S. Orders Embassy Evacuated, Closed" /The Toronto Star 28 Jan 1989 "'Atrocity' blamed on rebels' refusal to spare Ka
bul" / The Toronto Star 30 Jan 1989 "Battle rages for key road as diplomats depart Kabul
5
The Sunday Times 5 Feb 1989 "The Russians move out...;Withdrawal from Afghanistan
6
St. Louis Post-Dispatch 1 Feb 1989 "JET CARRYING U.S. ENVOYS NEARLY HITS SOVIET PLANE
```

Objective:

	T1	T2	T3	TN
GroupA	1	0	0	...
GroupB	0	1	0	...
GroupC	0	0	0	...

▸▸ More detail

Transforming the Data: Design

High-level overview:

- For each group:
- Pull the source articles
- Estimate a dynamic topic model for each group
- Aggregate article topic proportions by group-year
- Create a threshold for topic proportion shifts
- Identify actor-years with topic proportion shifts as "change" years
- Validate via case knowledge or UCDP Encyclopedia

Project Design

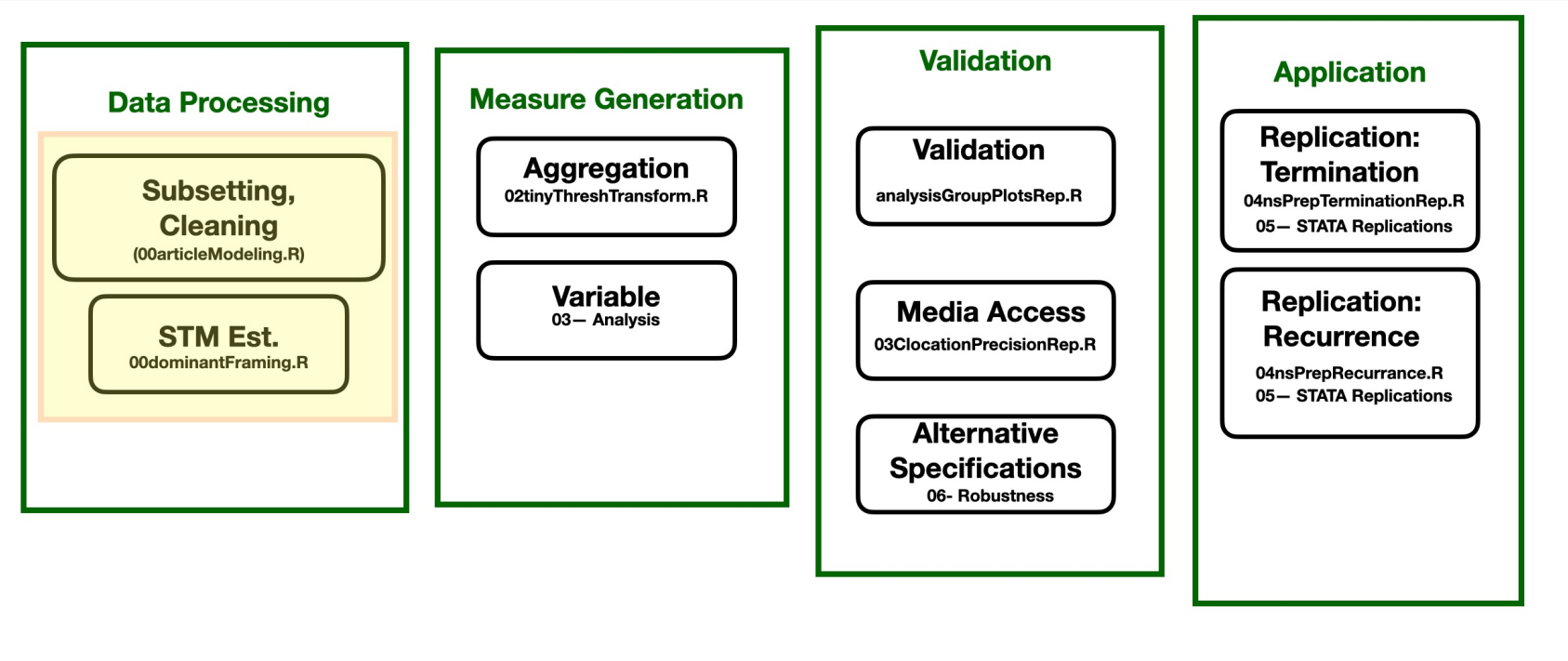


Figure 1: Project Outline

▸ Project Replication Script (GitHub)

Within one group

Oscillation: Abu Sayyaf

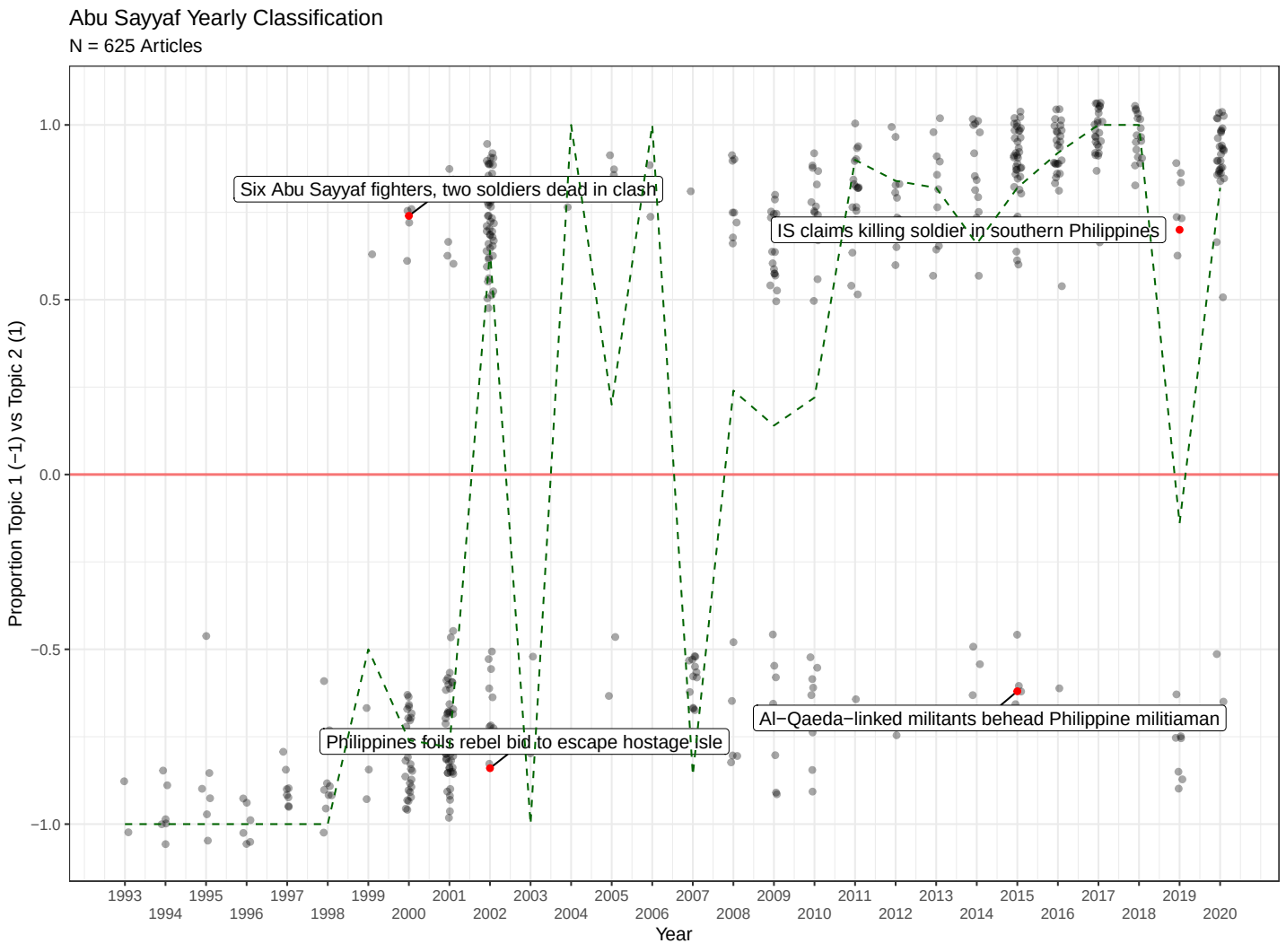
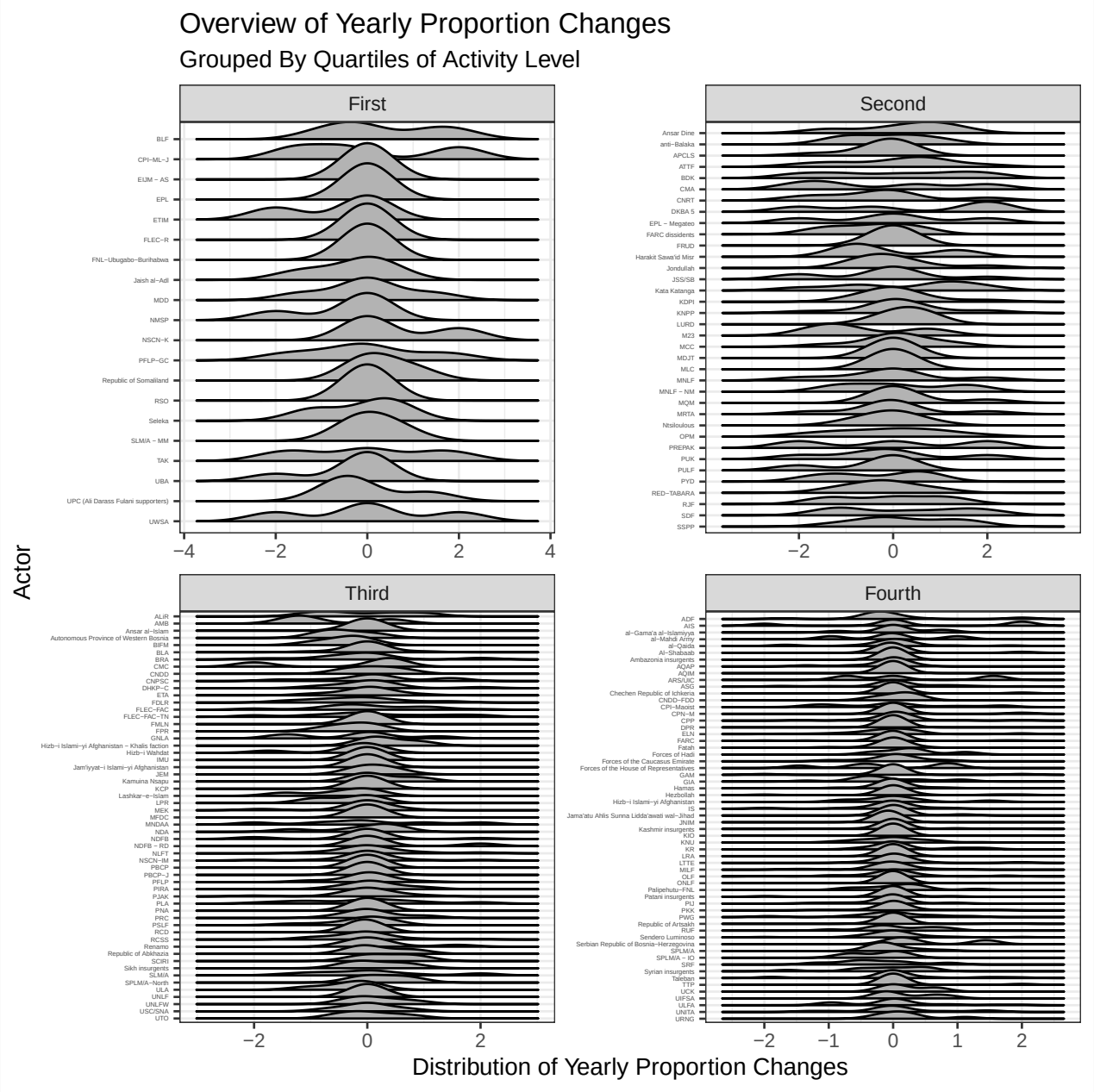


Figure 2: Change Model For Abu Sayyaf

Result


Overview of Yearly Proportion Changes
Grouped By Quartiles of Activity Level
First
Second
Third
Fourth
Actor
Distribution of Yearly Proportion Changes

Data Processing Code

Subsetting and cleaning

```
## takes a ucdp data subset,
## runs a two--topic STM on the full articles
article.analysis <- function(data){
    set.seed(6889)

    print(dim(data))

    tmp.articletext <- data$source_article

    ## Remove numbers
    data$source_article<- gsub(pattern='[[:digit:]]+',
                               replacement="",
                               x= data$source_article)

    data$source_article<- tolower(data$source_article)

    tmp.corpus <- quanteda::corpus(data,
                                   docid_field="id",
                                   text_field="source_article")


    ## pull out words that are very likely to refer to news agencies:
    ## and news reporting:
    news.agencies <- c("bbc*", "reuters","xinhua", "agenc*",
                       "post", "monitor*", "associ*",
                       "*service*", "afp", "france-press",
                       "*news*", "*source*","timeline*",
                       "*tv*","television", "*radio*",
                       "*voice*","audio", "dispatch",
                       "repor*", "track*", "media",
                       "ap", "*press*", "times",
                       "france","pti","international", ## international might drop content, but
                           probably mostly about international desks
                       "cnn", "nbc", "pna",  ## PNA is philippines news agency
                       "hrw",
                       "network", "*sharq*", "watan", ## al-shariqyah; al-Sharq Al Aswat
                       "*jazeera*", "*havat*",
```

Per-group model estimation

```
############################

## Text analysis:
## dominantFraming.R is a wrapper around
## (1) a call to topic model K=2 for a UCDP actor
## and then a series of summaries that produces yearly counts of the
## dominant topic.
## Returns group-year dominant discourse topic,
## topic proportions, and the first 7 REX words of the dominant topic


uids <- unique(ged.tiny$side_b_dset_id)

tiny.yearsum <- list()
tiny.basedata <- list()
i=1

for(u in uids){
    print(paste0("iteration: ", i))
    analysis<- dominantFraming(groupID= u,
                               ucdpdata=ged.tiny)

    tiny.yearsum[[i]] <- analysis$yearsum
    tiny.basedata[[i]] <- analysis$basedata
   i=i+1
}
```

Figure 4: Code Example

▸ Replication Script (GitHub) ▸ Subset, Clean, STM ▸ Aggregate To Correct Unit

Validation

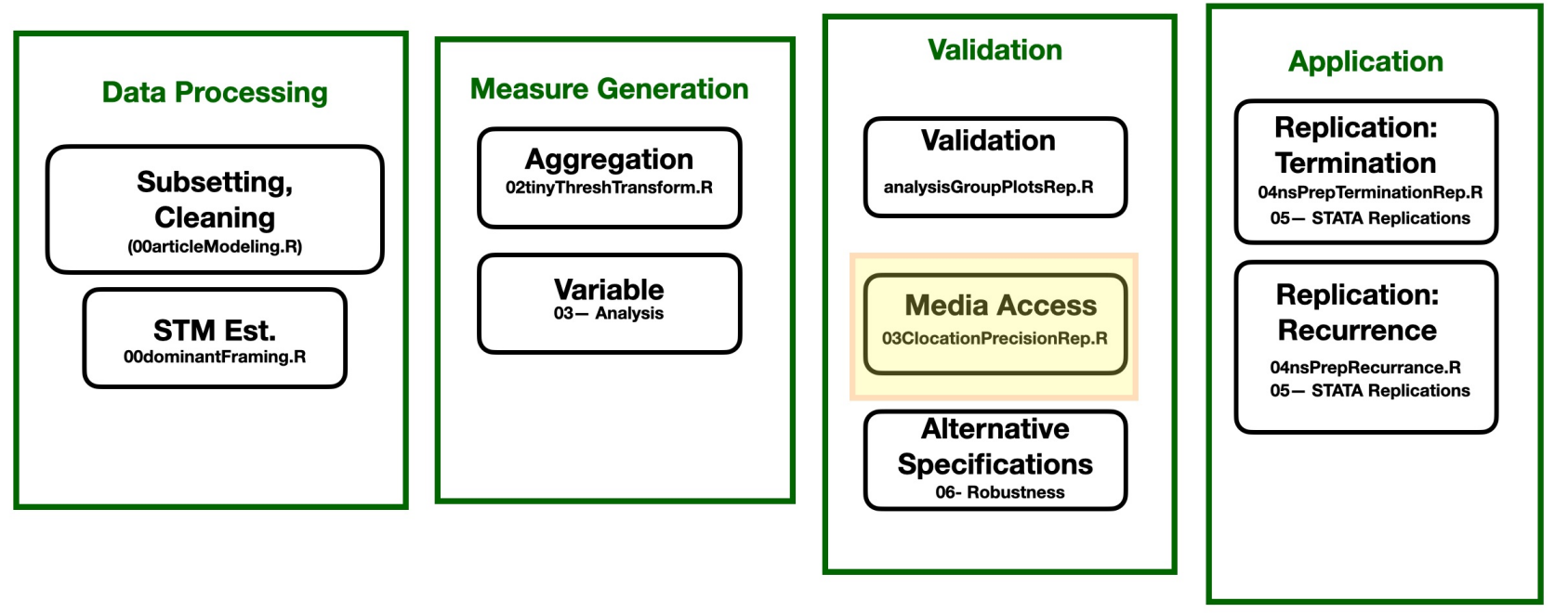


Figure 5: Project Outline

▸ Replication Script (GitHub)

Linear model, random effects, clustered S.E.

Code

```
## SE Grouped at group level:
mod2B<- plm(propdif.L1 ~ mean.loc.prec +
            mean.clarity + mean.time.prec +  MSF,
            model=c("random"),## random effects
            effect="time", ## focus on time
            index=c("groupID"),
            data=dat)

summary(mod2B)


df <- tidy(mod2B)
ci95 <- confint(mod2B, level=0.95) %>%
  data.frame() %>%
  rename("conf.low95" = "X2.5..",
         "conf.high95" = "X97.5..")

ci90 <- confint(mod2B, level=0.9) %>%
  data.frame() %>%
  rename("conf.low90" = "X5..",
         "conf.high90" = "X95..")

results <- bind_cols(df,
                     ci95,
                     ci90) %>%
    rename(Variable = term,
           Coefficient = estimate,
           SE = std.error) %>%
```

Results

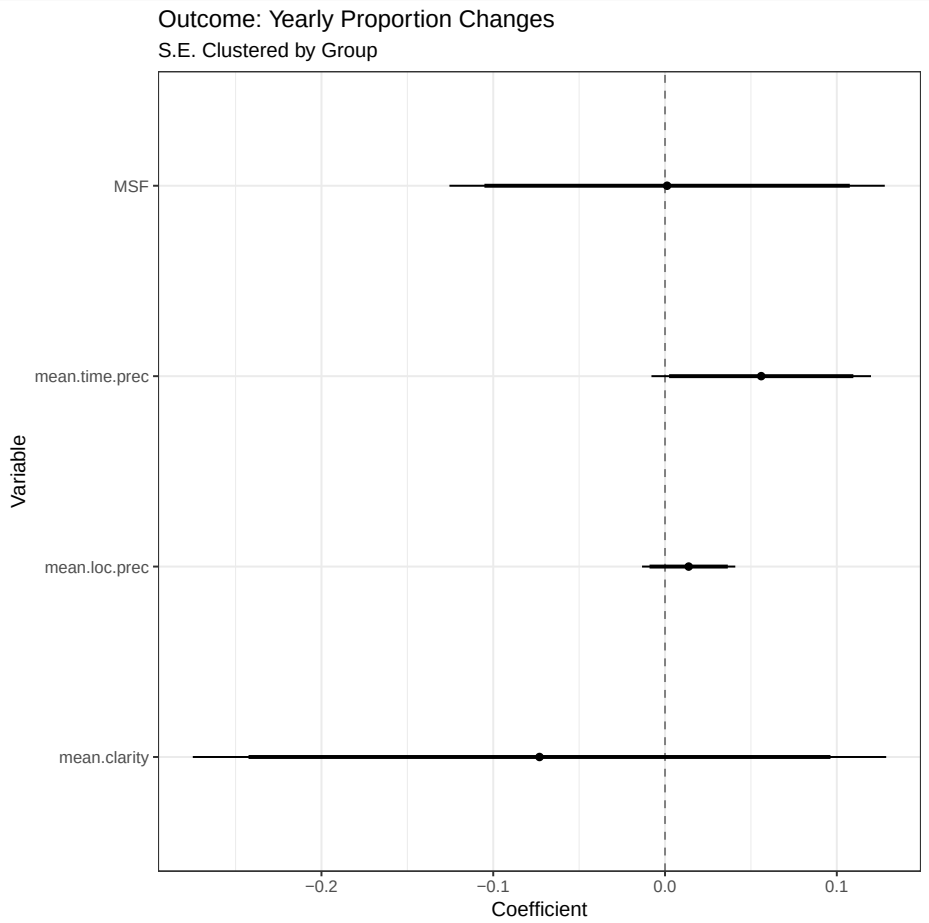


Figure 6: Relationship between change variable and media access, N = 2,298 group-country-years, 1989-2020

▸ Project Replication Script (GitHub) ▸ Precision Analysis

Zero-inflated negative binomial

Code

```
library("sandwich")
library(pscl)

## need to remove NA for clustered errors:

cols <- c("counter", "mean.loc.prec",
          "mean.clarity","mean.time.prec",
          "MSF", "groupID")

dat2 <- na.omit(dat[,cols]) ## 2298 x 6

mod3 <- zeroinfl(counter ~ mean.loc.prec +
            mean.clarity + mean.time.prec +  MSF,
                 data=dat2,
                 dist="negbin")

summary(mod3)

mod3ci <- coeftest(mod3,
                   vcov = vcovCL(mod3, cluster = dat2$groupID))

df <- tidy(mod3ci) ## then exoponentiate the coef:
```

Results

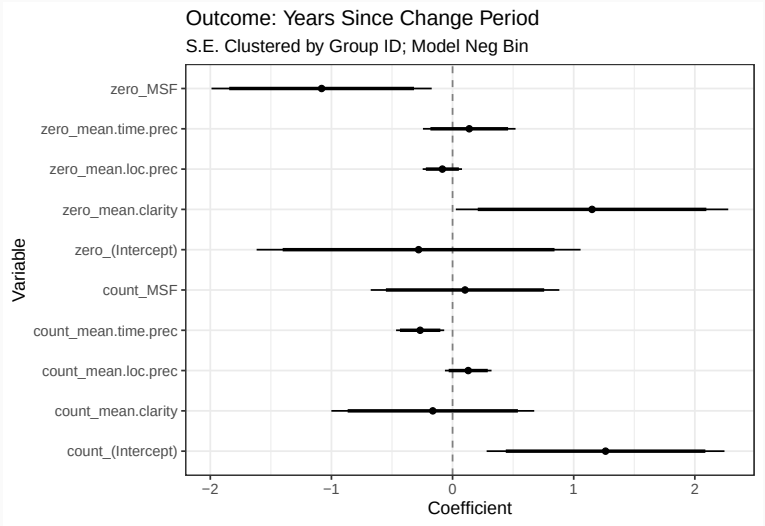


Figure 7: Relationship between media access and descriptive stability, N = 2,298 group-country-years, 1989-2020

▸ Project Replication Script (GitHub) ▸ Precision Analysis

Application

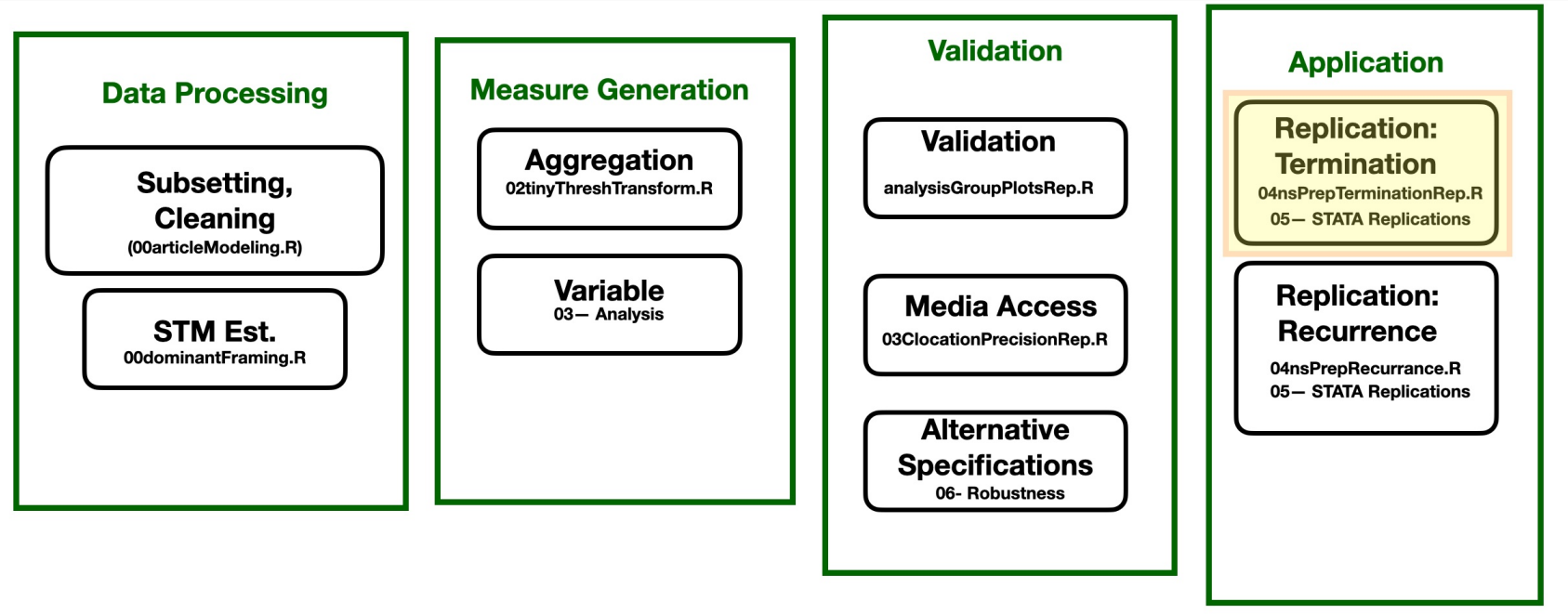


Figure 8: Project Outline

▸ Project Replication Script (GitHub)

Statistical Application

- Extended Nilsson and Svensson 2021: The Intractability of Islamist Insurgencies: Islamist Rebels and the Recurrence of Civil War
- Theory: Uncertainty about capabilities, pragmatism, and resources $\rightarrow$ longer conflicts that are more prone to recur
- Research design: "uncertainty" $\rightarrow$ "Islamist in first year"
- Statistical model: Cox proportional hazard models for Termination, Recurrence

⏩ Recurrence

Termination model

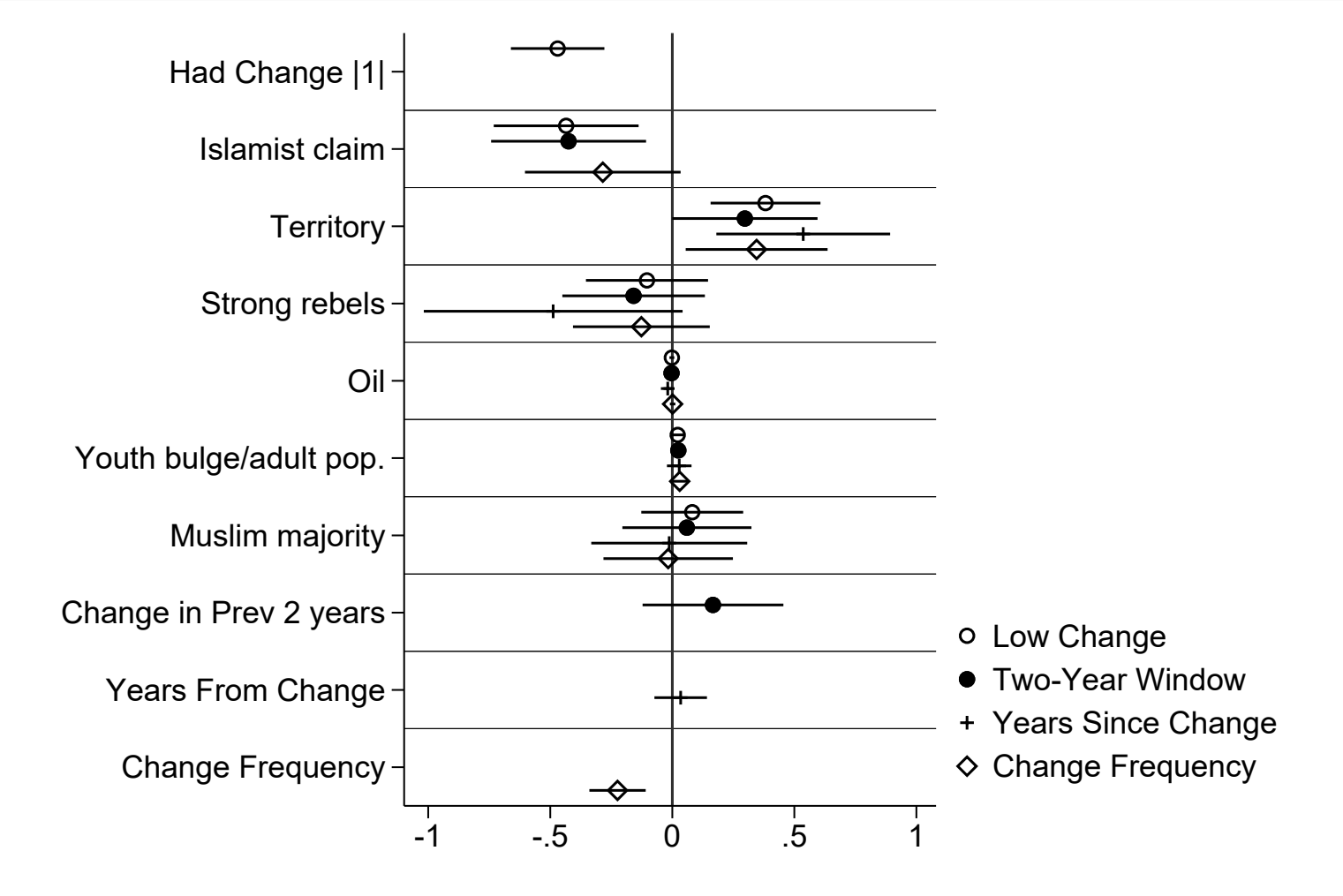


Figure 9: Cox Proportional Hazard Model "Risk" of Conflict Termination
N=1,104 conflict-years, 1989-2013

▸ Replication Script (GitHub) ▸ Data Prep ▸ STATA Analysis ▸ Variable Correlations

Conclusion

Goal: Contribution of a large-N, cross-group measure of changes in revealed behavior

- Strategy to identify transformation and change periods
- Built from an existing dataset, with interpretable group-level output
- Allows inclusion of group "transformation" in large-N studies of conflict

Selected Additional Projects

- Development of a bespoke theoretically identified IRT Model IRTM-Paper
- Application of IRT-M to Eurobarometer survey Eurobarometer
- Using IRT-M to extract sentiments from Afrobarometer surveys
- Identifying meaningful ties in network-derived variables SimDeg
- Designing and simulating network growth and fragmentation Dominos
- Measuring gridlock in diplomatic texts
- Using machine learning to measure growth-driven organizational changes

Thank you

Thank you!

Questions, Comments, Suggestions?

m.jenkins.foster@gmail.com

Extra Slides

Data and Strategy

- Data: UCDP Global Events Database
  - 261864 x 49 event dataset, derived from news
  - Subset to armed nonstate actors with more than 10 events
  - Use the source article field

- Strategy: Group-Specific Scale via Structural Topic Model
  - Remove words about reporting process
  - Model with K=2 and splined year covariates
  - Run independently for all groups

⏩ Why K=2 ⏩ K¿2 illustration ⏩ Back to overview

Correlations in Application (Termination)

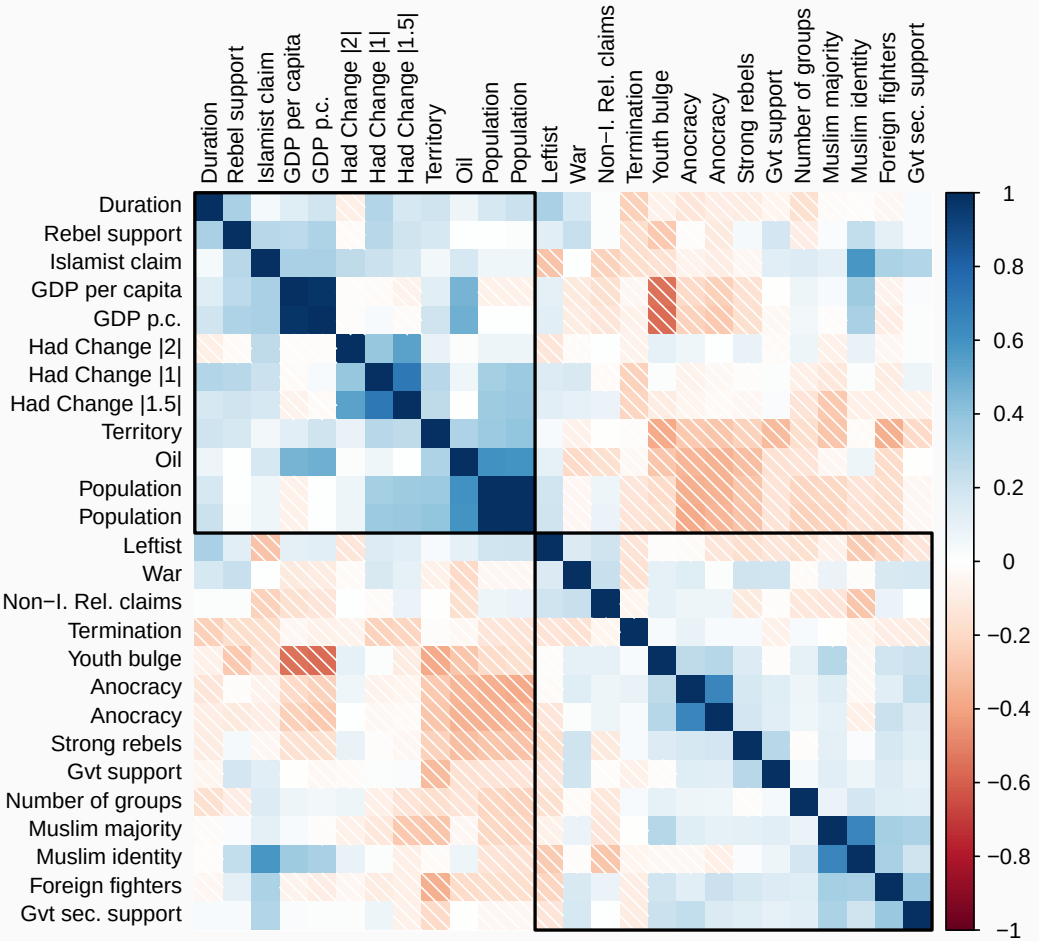


Figure 10: Correlation of change and ambiguity variables with variables in Nilsson and Svensson termination dataset.

⏩ Return to termination

Specification

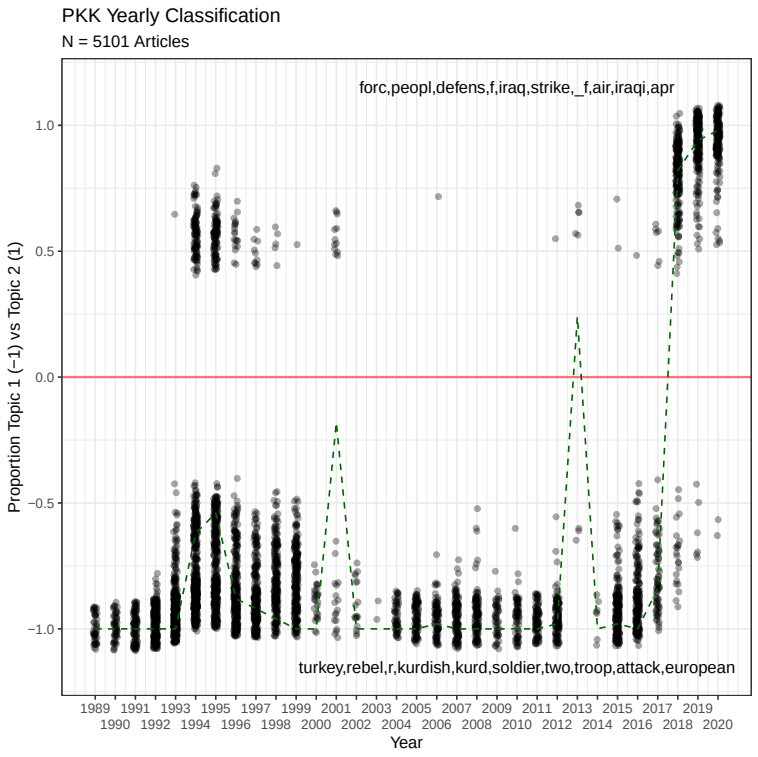
Why use an unsupervised unidimensional scale?

- Want to discover different end points for each group
- K=2 loses richness, but themes are not *that* expansive
- The approach is flexible
    - But: using more dimensions makes change points (1) subtle to identify and (2) resource intensive to analyze

⏩ Return to strategy

Example of Specification with K>2

Change Graph with K=2



Change Graph with K=5 (Clustered themes)

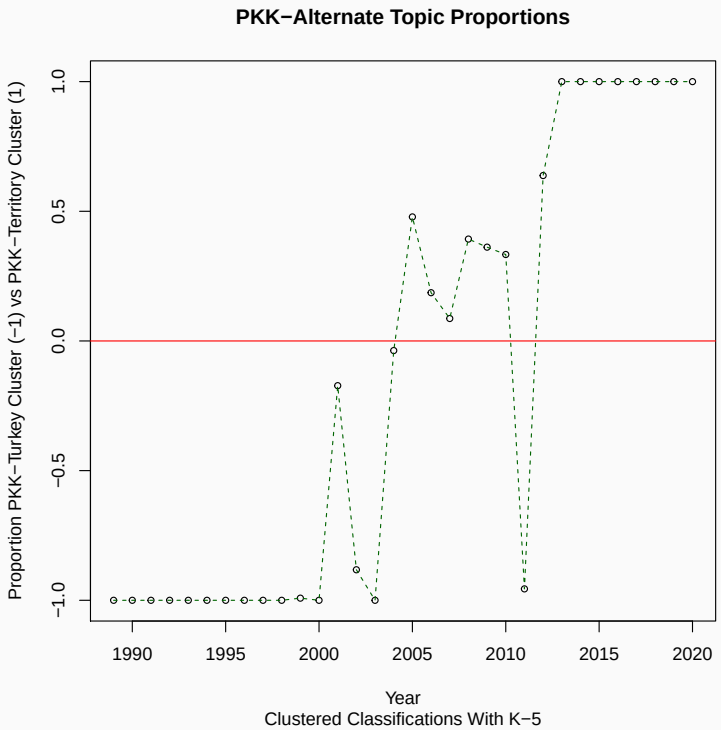


Figure 11: Comparison of PKK Change Periods Identified By Different Model Specifications

⏩ Return to strategy

Recurrence Replication

Base Model				Analysis time when record ends
Islamist claim	2.052	1.683	1.768	1.747
	(0.472)**	(0.368)*	(0.405)*	(0.399)*
Territory	2.481	2.441	2.500	2.586
	(0.648)**	(0.747)**	(0.760)**	(0.781)**
Strong rebels	0.774	0.682	0.673	0.656
	(0.395)	(0.342)	(0.334)	(0.330)
Oil	1.021	1.009	1.011	1.011
	(0.011)+	(0.012)	(0.012)	(0.011)
Youth bulge/adult pop.	1.055	1.065	1.071	1.070
	(0.022)*	(0.024)**	(0.024)**	(0.024)**
Muslim majority	0.661	0.698	0.682	0.688
	(0.183)	(0.204)	(0.201)	(0.208)
Had Change |1|		1.387		
		(0.317)		
Had Change |1.5|			1.403	
			(0.351)	
Had Change |2|				1.181
				(0.319)
N	5,554	2,427	2,427	2,427

$+\ p < 0.1$; * $p < 0.05$; ** $p < 0.01$

Robust standard errors in parentheses clustered on dyad.

⏩ Back to application

Figure 12: Results of Recurrence Model