Gov 2001: Extension of Original Paper

Sima Biondi, Priyanka Sethy, Natalie Ayers

2022-11-29

Introduction

We are extending "Redemption through Rebellion: Border Change, Lost Unity, and Nationalist Conflict." (2022), by Lars-Erik Cederman, Seraina Rüegger, and Guy Schvitz. The data and code for this paper are in the Harvard Dataverse.

Results

```
# Set working directory to own project folder
#setwd("./redemption-through-rebellion-dataverse_files/")

## Load dataset
#an.df <- read.dta("epr_segment_level_analysis_ext.dta")
#an.df.rugged <- read.csv(here("redemption-through-rebellion-dataverse_files","epr_segment_level_analys
exten.df = read.csv(here("redemption-through-rebellion-dataverse_files",
                         "epr_segment_level_analysis_extensions_rugged_claims.csv"))

## Code "peaceyears" variable: years since last conflict, squared, cubed
exten.df$pys <- exten.df$peaceyears
exten.df$pys2 <- exten.df$peaceyears^2
exten.df$pys3 <- exten.df$peaceyears^3

## Subset dataset: only politically relevant groups (EPR definition), exclude
## dominant and monopoly groups and groups without settlement area in GeoEPR,
exten.df.sub.allyears <- exten.df %>%
  filter(isrelevant==1,
         status_monopoly==0,
         status_dominant==0,
         !is.na(seg_area_sqkm),
         !is.na(onset_do_flag))
```

Table 1

Table 1 with Mean Terrain Ruggedness and AG violence/lost autonomy

The table below replicates Table 1 in Cederman et al. (2022), providing the results of 3 logit models with robust standard errors clustered at the AG level.

```r
# define variables for use in all models, including DV
vars_logit <- c("onset_do_flag", "ln_ag_area_sqkm", "ag_incidence_flag_lag",
                "status_excl", "downgraded2", "rugged_mean",
                "rbal", "warhist", "ln_capdist", "ln_rgdppc_lag",
                "ln_pop_lag",  "colonial_past", "ln_state_age",
                "pys", "pys2", "pys3", "mean_action", "lost_autonomy")
# identify AG "treatment" variables to be analyzed separately
treat.vars <- c("split", "tfrac", "tfrac_incr_post1946")

# for each "treatment" variable, run logit model
for(treat_var in treat.vars){
  full_lhs <- append(treat_var, vars_logit)
  onset_logit <- glm(onset_do_flag ~ .,
                     data=exten.df.sub.allyears[,full_lhs], family="binomial")
  assign(paste0("logit_",treat_var), onset_logit)
}

# include clustered standard errors for each generated model
logit_split_coefs <- coeftest(logit_split,
                              vcov. = vcovCL(logit_split,
                                             cluster = exten.df.sub.allyears$ag_id))
logit_tfrac_coefs <- coeftest(logit_tfrac,
                              vcov. = vcovCL(logit_tfrac,
                                             cluster = exten.df.sub.allyears$ag_id))
logit_tfrac_1946_coefs <- coeftest(logit_tfrac_incr_post1946,
                                   vcov. = vcovCL(logit_tfrac_incr_post1946,
                                                  cluster = exten.df.sub.allyears$ag_id))

# Display results of logit model
stargazer(logit_split_coefs, logit_tfrac_coefs, logit_tfrac_1946_coefs,
          type="text",
          dep.var.labels.include = FALSE,
          omit = c("pys", "pys2","pys3"),
          title = "Replication of Table 1",
          dep.var.caption = "Civil Conflict Onset",
          covariate.labels = c("Divided group", "Fractionalization",
                               "Frac. incr. since 1946", "Territory sq.km, log",
                               "Ongoing conflict, lag", "Exclusion", "Downgraded",
                               "Terrain Ruggedness (Mean)","Relative size", "War history",
                               "Distance to capital, log", "GDP, lag, log",
                               "Population, log","Colonial history",
                               "State age, log",
                               "Mean AG action", "Lost autonomy"))
```

Replication of Table 1

	Civil Conflict Onset		
	(1)	(2)	(3)

Divided group 0.745*** (0.207)

Divided group 0.686*** (0.266)
Fractionalization 1.807*** (0.574)
Frac. incr. since 1946 3.919*** (0.958)
Territory sq.km, log -0.211*** -0.239*** -0.118* (0.075) (0.065) (0.062)
Ongoing conflict, lag 0.250 0.017 0.231 (0.405) (0.379) (0.445)
Exclusion -0.132 -0.217 -0.127 (0.401) (0.388) (0.375)
Downgraded 1.376*** 1.488*** 1.469*** (0.449) (0.441) (0.446)
Terrain Ruggedness (Mean) -0.00003 -0.00002 -0.0001 (0.00005) (0.00005) (0.0001)
Relative size 2.570*** 2.571*** 2.606*** (0.935) (0.871) (0.822)
War history 0.300* 0.308* 0.254 (0.170) (0.166) (0.161)
Distance to capital, log -0.035 -0.046 -0.051 (0.141) (0.139) (0.129)
GDP, lag, log 0.089 0.196 0.011 (0.167) (0.175) (0.167)
Population, log -0.442* -0.437* -0.423* (0.247) (0.241) (0.246)
Colonial history 0.370 0.325 0.564* (0.392) (0.378) (0.323)
State age, log 0.189 0.172 0.335** (0.176) (0.176) (0.165)
AG level of violence 2.448*** 2.444*** 2.494*** (0.222) (0.218) (0.221)
Downgrade of autonomy (5 years) 0.500 0.583 0.544 (0.503) (0.486) (0.488)
Downgrade of inclusion (5 years) 0.606 0.611 0.655 (0.536) (0.540) (0.530)
Constant -3.933** -3.946** -5.016*** (1.657) (1.593) (1.621)
==========
==========
Note: *p<0.1;* ***p<0.05;*** p<0.01
Table 1 with Median Terrain Ruggedness
"'r # define variables for use in all models, including DV vars_logit <- c("onset_do_flag", "ln_ag_area_sqkm", "ag_incidence_flag_lag", "status_excl", "downgraded2", "rugged_med", "rbal", "warhist", "ln_capdist", "ln_rgdppc_lag", "ln_pop_lag", "colonial_past", "ln_state_age", "pys", "pys2", "pys3") # identify AG "treatment" variables to be analyzed separately treat.vars <- c("split", "tfrac", "tfrac_incr_post1946") # for each "treatment" variable, run logit model for(treat_var in treat.vars){ full_lhs <- append(treat_var, vars_logit) onset_logit <- glm(onset_do_flag ~ ., data=exten.df.sub.allyears[,full_lhs], family="binomial") assign(paste0("logit_",treat_var), onset_logit) } # include clustered standard errors for each generated model logit_split_coefs <- coeftest(logit_split, vcov. = vcovCL(logit_split, cluster = exten.df.sub.allyears$ag_id))$logit_tfrac_coefs <- -coeftest(logit_tfrac, vcov. = vcovCL(logit_tfrac, cluster = exten.df.sub.allyears$ag_id)) logit_tfrac_1946_coefs <- coeftest(logit_tfrac_incr_post1946, vcov. = vcovCL(logit_tfrac_incr_post1946, cluster = exten.df.sub.allyears$ag_id)) # Display results of logit model stargazer(logit_split_coefs, logit_tfrac_coefs, logit_tfrac_1946_coefs, type="text", dep.var.labels.include = FALSE, omit = c("pys", "pys2","pys3"), title = "Replication of Table 1", dep.var.caption = "Civil Conflict Onset", covariate.labels = c("Divided group", "Fractionalization", "Frac. incr. since 1946", "Territory sq.km, log", "Ongoing conflict, lag", "Exclusion", "Downgraded", "Terrain Ruggedness (Med)","Relative size", "War history", "Distance to capital, log", "GDP, lag, log", "Population, log","Colonial history", "State age, log")) "'
Replication of Table 1
==========
Civil Conflict Onset ———————— (1) (2) (3)

Divided group 0.603***
(0.182)

Fractionalization 1.565***
(0.360)

Frac. incr. since 1946 3.391*** (0.854)

Territory sq.km, log -0.066 -0.076 0.010
(0.052) (0.048) (0.045)

Ongoing conflict, lag 0.181 0.026 0.240
(0.419) (0.377) (0.438)

Exclusion 0.870*** 0.840*** 0.899*** (0.291) (0.274) (0.291)

Downgraded 1.253*** 1.298*** 1.317*** (0.271) (0.265) (0.270)

Terrain Ruggedness (Med) -0.0001* -0.0001* -0.0002*** (0.0001) (0.00005) (0.0001)

Relative size 1.078** 0.924** 0.970**
(0.459) (0.447) (0.410)

War history 0.655*** 0.654*** 0.640*** (0.117) (0.115) (0.103)

Distance to capital, log 0.125 0.128 0.120
(0.088) (0.094) (0.088)

GDP, lag, log 0.295** 0.322*** 0.255**
(0.119) (0.118) (0.119)

Population, log -0.715*** -0.719*** -0.712*** (0.189) (0.180) (0.183)

Colonial history 0.527 0.466 0.613**
(0.333) (0.320) (0.282)

State age, log 0.083 0.058 0.144
(0.133) (0.130) (0.118)

Constant -3.392*** -3.258*** -4.042*** (1.104) (1.040) (1.046)

==========

Note: *p<0.1;* ***p<0.05;*** p<0.01

Table 1 with Range (Max - Min) Terrain Ruggedness

```
# define variables for use in all models, including DV
vars_logit <- c("onset_do_flag", "ln_ag_area_sqkm", "ag_incidence_flag_lag",
                "status_excl", "downgraded2", "rugged_range",
                "rbal", "warhist", "ln_capdist", "ln_rgdppc_lag",
                "ln_pop_lag",  "colonial_past", "ln_state_age",
                "pys", "pys2", "pys3")
# identify AG "treatment" variables to be analyzed separately
treat.vars <- c("split", "tfrac", "tfrac_incr_post1946")

# for each "treatment" variable, run logit model
for(treat_var in treat.vars){
  full_lhs <- append(treat_var, vars_logit)
  onset_logit <- glm(onset_do_flag ~ .,
                     data=exten.df.sub.allyears[,full_lhs], family="binomial")
```

```
  assign(paste0("logit_",treat_var), onset_logit)
}

# include clustered standard errors for each generated model
logit_split_coefs <- coeftest(logit_split,
                              vcov. = vcovCL(logit_split,
                                             cluster = exten.df.sub.allyears$ag_id))
logit_tfrac_coefs <- coeftest(logit_tfrac,
                              vcov. = vcovCL(logit_tfrac,
                                             cluster = exten.df.sub.allyears$ag_id))
logit_tfrac_1946_coefs <- coeftest(logit_tfrac_incr_post1946,
                                   vcov. = vcovCL(logit_tfrac_incr_post1946,
                                                  cluster = exten.df.sub.allyears$ag_id))

# Display results of logit model
stargazer(logit_split_coefs, logit_tfrac_coefs, logit_tfrac_1946_coefs,
          type="text",
          dep.var.labels.include = FALSE,
          omit = c("pys", "pys2","pys3"),
          title = "Replication of Table 1",
          dep.var.caption = "Civil Conflict Onset",
          covariate.labels = c("Divided group", "Fractionalization",
                               "Frac. incr. since 1946", "Territory sq.km, log",
                               "Ongoing conflict, lag", "Exclusion", "Downgraded",
                               "Terrain Ruggedness (Range)","Relative size", "War history",
                               "Distance to capital, log", "GDP, lag, log",
                               "Population, log","Colonial history",
                               "State age, log"))
```

Replication of Table 1

	Civil Conflict Onset		
	(1)	(2)	(3)
Divided group	0.577*** (0.181)		
Fractionalization incr. since 1886		1.460*** (0.387)	
Frac. incr. before 1946	0.945**	0.838**	(0.385) (0.401)
Frac. incr. since 1946		3.346*** (0.812)	
Territory sq.km, log	-0.017	-0.003	0.0005 (0.049) (0.047) (0.045)
Ongoing conflict, lag	0.157	0.152	0.207 (0.438) (0.449) (0.440)
Exclusion	0.880***	0.896***	0.851*** (0.300) (0.294) (0.296)
Downgraded	1.263***	1.302***	1.259*** (0.268) (0.269) (0.269)
Terrain Ruggedness	-0.0001*	-0.0001**	-0.0001 (0.00004) (0.00004) (0.00004)
Relative size	0.901**	0.892**	0.931* (0.443) (0.398) (0.476)
War history	0.649***	0.625***	0.671*** (0.112) (0.104) (0.122)
Distance to capital, log	0.085	0.102	0.088 (0.094) (0.094) (0.092)
GDP, lag, log	0.291**	0.295**	0.256** (0.122) (0.122) (0.122)
Population, log	-0.703***	-0.684***	-0.720*** (0.176) (0.176) (0.184)
Colonial history	0.427	0.583**	0.396 (0.316) (0.282) (0.343)

Divided group 0.577*** (0.181)

State age, log 0.089 0.142 0.073 (0.128) (0.117) (0.136)
Constant -3.673*** -4.175*** -3.511*** (1.040) (1.034) (1.096)
==========
==========
Note: *p<0.1;* ***p<0.05;*** p<0.01 ## Table 3
Table 3 with Mean Terrain Ruggedness
The table below replicates Table 3 in Cederman et al. (2022), which displays the results of two fixed effects models to account for potential AG omitted effects and, they argue, provide a better test of their hypothesis that stronger fractionalization should lead to greater chance of conflict.
"'r # define variables for use in all models, including DV vars_logit <- c("onset_do_flag", "ln_ag_area_sqkm", "ag_incidence_flag_lag", "status_excl", "downgraded2", "rugged_mean", "rbal", "warhist", "ln_capdist", "ln_rgdppc_lag", "ln_pop_lag", "colonial_past", "ln_state_age", "pys", "pys2", "pys3")
# identify "treatment" variables treat.vars.3 <- c("tfrac", "tfrac_incr_post1946")
# run fixed effects models for both treatment variables # felm is very picky with equation inputs, so had to manually # paste each equation into each regression for(treat_var in treat.vars.3){ full_lhs <- append(treat_var, vars_logit) full_lhs <- append(full_lhs,"ag_id")
# use to create formula again if need to, copy and paste #felm_formula <- as.formula(paste("onset_do_flag ~", # paste(full_lhs[!full_lhs %in% c("onset_do_flag","ag_id")], # collapse="+")))
d <- exten.df.sub.allyears[,full_lhs] if(treat_var == "tfrac"){ # regression for `tfrac` variable onset_tfrac_fe_reg <- felm(onset_do_flag ~ tfrac + ln_ag_area_sqkm + ag_incidence_flag_lag + status_excl + downgraded2 + rugged_mean + rbal + warhist + ln_capdist + ln_rgdppc_lag + ln_pop_lag + colonial_past + ln_state_age + pys + pys2 + pys3 | ag_id | 0 | ag_id, data = d) } else{ # regression for `tfrac_incr_post1946` variable onset_tfrac_post1946_fe_reg <- felm(onset_do_flag ~ tfrac_incr_post1946 + ln_ag_area_sqkm + ag_incidence_flag_lag + status_excl + downgraded2 + rugged_mean + rbal + warhist + ln_capdist + ln_rgdppc_lag + ln_pop_lag + colonial_past + ln_state_age + pys + pys2 + pys3 | ag_id | 0 | ag_id, data = d) } }
# display results stargazer(onset_tfrac_fe_reg, onset_tfrac_post1946_fe_reg, type="text", dep.var.labels.include = FALSE, omit = c("pys", "pys2","pys3"), title = "Replication of Table 3", dep.var.caption = "Linear Probability Models: Civil Conflict Onset", covariate.labels = c("Fractionalization", "Frac. incr. since 1946", "Territory sq.km, log", "Ongoing conflict, lag", "Exclusion", "Downgraded", "Terrain Ruggedness (Mean)", "Relative size", "War history", "Distance to capital, log", "GDP, lag, log", "Population, log","Colonial history", "State age, log")) "'
Replication of Table 3
==========
Linear Probability Models: Civil Conflict Onset ———— (1) (2)

Fractionalization 0.046**
(0.021)

Frac. incr. since 1946 0.055**
(0.024)

Territory sq.km, log -0.014 -0.013
(0.011) (0.011)

Ongoing conflict, lag -0.015** -0.015**
(0.006) (0.006)

Exclusion 0.004* 0.004
(0.003) (0.003)

Downgraded 0.020*** 0.020***
(0.007) (0.007)

Terrain Ruggedness (Mean) -0.00000 -0.00000*
(0.00000) (0.00000)

Relative size 0.011 0.012
(0.007) (0.007)

War history 0.011 0.011
(0.007) (0.007)

Distance to capital, log 0.001 0.001
(0.001) (0.001)

GDP, lag, log 0.001 0.001
(0.002) (0.002)

Population, log -0.008*** -0.008***
(0.003) (0.003)

Colonial history 0.008 0.008
(0.005) (0.005)

State age, log 0.002 0.003
(0.002) (0.002)

Observations 27,105 27,105 R2 0.052 0.052 Adjusted R2 0.037 0.037 Residual Std. Error (df = 26676) 0.087 0.087

==========

Note: *p<0.1;* ***p<0.05;*** p<0.01

Fractionalization 0.045** (0.021)
Frac. incr. since 1946 0.055** (0.024)
Territory sq.km, log -0.009 -0.008 (0.009) (0.009)
Ongoing conflict, lag -0.015** -0.015** (0.006) (0.006)
Exclusion 0.004* 0.004 (0.003) (0.003)
Downgraded 0.020*** 0.020*** (0.007) (0.007)
Terrain Ruggedness (Med) 0.00000 0.00000 (0.00002) (0.00002)
Relative size 0.012 0.012 (0.007) (0.007)
War history 0.011 0.011 (0.007) (0.007)
Distance to capital, log 0.001 0.001 (0.001) (0.001)
GDP, lag, log 0.001 0.001 (0.002) (0.002)
Population, log -0.008*** -0.008*** (0.003) (0.003)
Colonial history 0.008 0.008 (0.005) (0.005)
State age, log 0.002 0.003 (0.002) (0.002)

Observations 27,105 27,105
R2 0.052 0.052
Adjusted R2 0.037 0.037
Residual Std. Error (df = 26676) 0.087 0.087

==========

Note: *p<0.1;* ***p<0.05;*** p<0.01

Table 3 with Range Terrain Ruggedness

```r
# define variables for use in all models, including DV
vars_logit <- c("onset_do_flag", "ln_ag_area_sqkm", "ag_incidence_flag_lag",
                "status_excl", "downgraded2", "rugged_range",
                "rbal", "warhist", "ln_capdist", "ln_rgdppc_lag",
                "ln_pop_lag",  "colonial_past", "ln_state_age",
                "pys", "pys2", "pys3")


# identify "treatment" variables
treat.vars.3 <- c("tfrac", "tfrac_incr_post1946")

# run fixed effects models for both treatment variables
# felm is very picky with equation inputs, so had to manually
# paste each equation into each regression
for(treat_var in treat.vars.3){
  full_lhs <- append(treat_var, vars_logit)
  full_lhs <- append(full_lhs,"ag_id")

  # use to create formula again if need to, copy and paste
  #felm_formula <- as.formula(paste("onset_do_flag ~ ",
  #                           paste(full_lhs[!full_lhs %in% c("onset_do_flag","ag_id")],
  #                           collapse="+")))

  d <- exten.df.sub.allyears[,full_lhs]
  if(treat_var == "tfrac"){
    # regression for `tfrac` variable
    onset_tfrac_fe_reg <- felm(onset_do_flag ~ tfrac + ln_ag_area_sqkm + ag_incidence_flag_lag +
                                 status_excl + downgraded2 + rugged_range + rbal + warhist + ln_capdist +
                                 ln_rgdppc_lag + ln_pop_lag + colonial_past + ln_state_age +
                                 pys + pys2 + pys3 | ag_id | 0 | ag_id, data = d)
  }
  else{
    # regression for `tfrac_incr_post1946` variable
    onset_tfrac_post1946_fe_reg <- felm(onset_do_flag ~ tfrac_incr_post1946 + ln_ag_area_sqkm +
                                          ag_incidence_flag_lag + status_excl + downgraded2 +
                                          rugged_range + rbal + warhist + ln_capdist + ln_rgdppc_lag +
                                          ln_pop_lag + colonial_past + ln_state_age +
                                          pys + pys2 + pys3 | ag_id | 0 | ag_id, data = d)
  }
}

# display results
stargazer(onset_tfrac_fe_reg, onset_tfrac_post1946_fe_reg,
          type="text",
          dep.var.labels.include = FALSE,
          omit = c("pys", "pys2","pys3"),
          title = "Replication of Table 3",
          dep.var.caption = "Linear Probability Models: Civil Conflict Onset",
          covariate.labels = c("Fractionalization",
                               "Frac. incr. since 1946",
                               "Territory sq.km, log",
                               "Ongoing conflict, lag", "Exclusion",
                               "Downgraded", "Terrain Ruggedness (Range)",
```

```
                                "Relative size", "War history",
                                "Distance to capital, log", "GDP, lag, log",
                                "Population, log","Colonial history",
                                "State age, log"))
```

Replication of Table 3

```
                                Linear Probability Models: Civil Conflict Onset
                                -----------------------------------------------
                                          (1)                     (2)
```

Fractionalization	0.046**	(0.021)
Frac. incr. since 1946	0.055**	(0.024)
Territory sq.km, log	-0.014	-0.013 (0.012) (0.012)
Ongoing conflict, lag	-0.015**	-0.015** (0.006) (0.006)
Exclusion	0.004*	0.004 (0.003) (0.003)
Downgraded	0.020***	0.020*** (0.007) (0.007)
Terrain Ruggedness (Range)	-0.00000	-0.00000 (0.00000) (0.00000)
Relative size	0.012	0.012 (0.007) (0.007)
War history	0.011	0.011 (0.007) (0.007)
Distance to capital, log	0.001	0.001 (0.001) (0.001)
GDP, lag, log	0.001	0.001 (0.002) (0.002)
Population, log	-0.008***	-0.008*** (0.003) (0.003)
Colonial history	0.008	0.008 (0.005) (0.005)
State age, log	0.002	0.003 (0.002) (0.002)
Observations	27,105	27,105
R2	0.052	0.052
Adjusted R2	0.037	0.037
Residual Std. Error (df = 26676)	0.087	0.087

==========

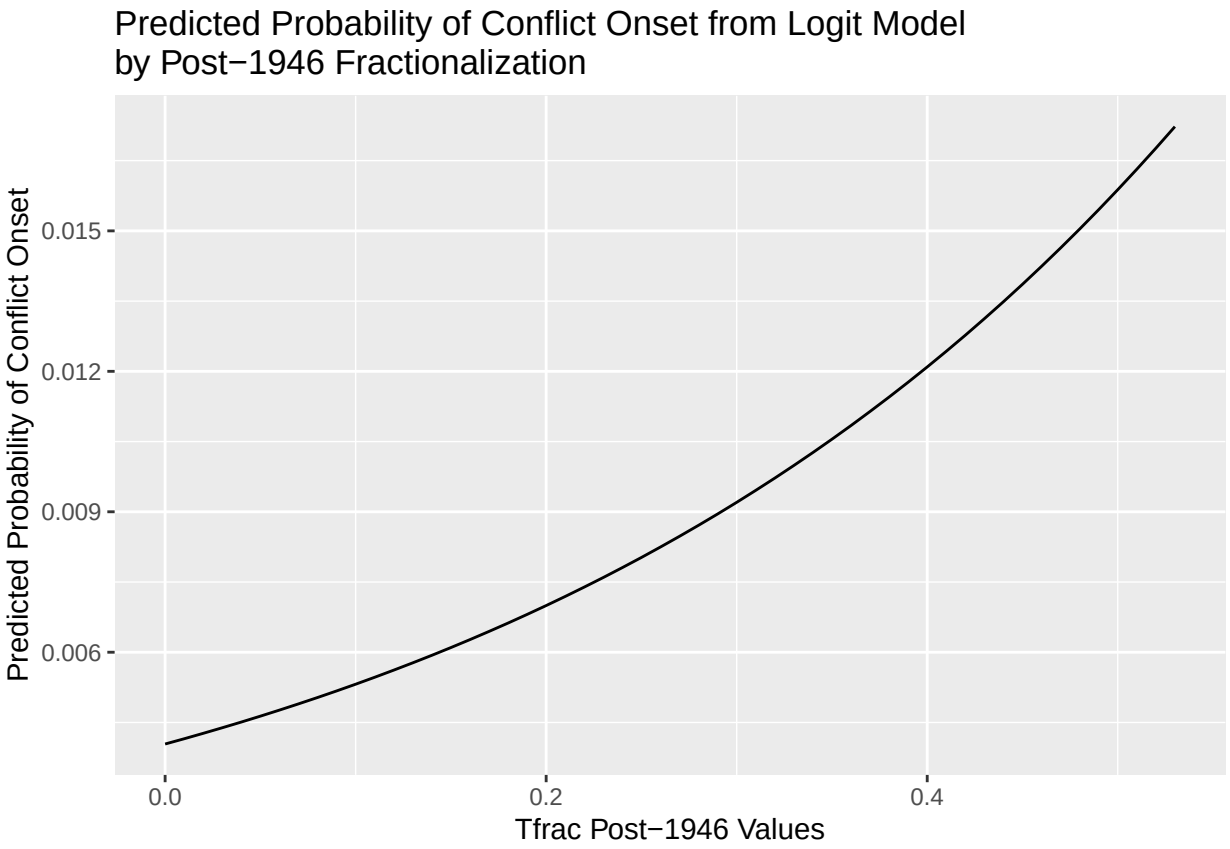
Note: *p<0.1;* ***p<0.05;*** p<0.01

Figure 7

The figure below replicates Figure 7 in Cederman et al. (2022), which provides predicted probabilities of conflict onset as the post-1946 fractionalization increases from 0 to its maximum value, according to the original logit model developed for use in Table 1. The authors use it to indicate that their results are substantively important, demonstrating the magnitude of the predicted increase in conflict onset with increasing territorial fractionalization. The results of our R code match closely with the figure generated by the authors in Stata. Given the difficulties with making precise estimates from empirical conflict models, we are treating this figure as fairly supplemental, rather than central, to our extension - it provides a visual that is more easily interpretable, but the actual predicted values should be taken with caution, as there are wide confidence intervals around these predictions.

```
pred_prob_lhs <- append("tfrac_incr_post1946", vars_logit)
pred_prob_post1946_onset_logit <- glm(onset_do_flag ~ .,
                    data=exten.df.sub.allyears[,pred_prob_lhs], family="binomial")
```

```r
predprob_logit_tfrac_post1946_coefs <- coeftest(pred_prob_post1946_onset_logit,
                                vcov. = vcovCL(pred_prob_post1946_onset_logit,
                                               cluster = exten.df.sub.allyears$ag_id))

# holding everything except tfrac_incr_post1946 at their means
mean_var_vals <- c(1, map(exten.df.sub.allyears[,pred_prob_lhs[3:17]], function(x) mean(x, na.rm=TRUE)))
names(mean_var_vals)[1] <- c("(Intercept)")
mean_var_vals <- unlist(mean_var_vals)

# get max val of tfrac_incr_post1946 to properly specify range:
max_tfrac_post1946 <- max(exten.df.sub.allyears$tfrac_incr_post1946, na.rm=TRUE)

# specify range of tfrac_incr_post1946 over which to predict
tfrac_post1946_range <- seq(0,max_tfrac_post1946, 0.01)

pred_prob_onset <- function(dyn_coef){
  full_coefs <- as.list(c(mean_var_vals[1],tfrac_incr_post1946=dyn_coef,mean_var_vals[2:16]))
  return(predict(pred_prob_post1946_onset_logit, full_coefs, type = "response"))
}

pred_probs_onset_tfrac_post1946 <- lapply(tfrac_post1946_range, pred_prob_onset)

pred_probs_onset_tfrac_post1946_df <- data.frame(pred_probs=unlist(pred_probs_onset_tfrac_post1946))
pred_probs_onset_tfrac_post1946_df$tfrac_incr_post1946 <- tfrac_post1946_range

# without Confidence Intervals
ggplot(data = pred_probs_onset_tfrac_post1946_df,
       aes(x = tfrac_post1946_range, y=pred_probs)) +
  geom_line() +
  xlab("Tfrac Post-1946 Values") +
  ylab("Predicted Probability of Conflict Onset") +
  labs(title = "Predicted Probability of Conflict Onset from Logit Model \nby Post-1946 Fractionalizati
```



Status of Extension Analysis

Instigator

The inclusion of an instigator field is not possible using the UCDP conflict data which Cederman et al. employ in their analysis. Therefore, we plan on joining the UCDP data with the Correlates of War Intra-State wars dataset (and potentially other datasets with instigator values, using xSub) to add this instigator field. We expect to lose a number of data points by performing this join, both because some data points won't be in the COW dataset and because some will be unable to join with confidence, but we expect there to be no systemic bias in the 2nd cause, and we expect any systemic bias from the first cause to have minimal to no impact on the mechanism of fractionalization's impact on conflict onset which we will be trying to test.

We plan on re-running the data creation with this updated dataset and re-running the same models above with an indicator variable for which side began the conflict and separating the dataset into two based on which side instigated. We expect the dataset creation to be the most time-intensive component of this part of the extension, given the code to run the models is already created, but we don't expect this to be unfeasible.

Topographic Features

We hope to use two datasets to test the impact of terrain features in our analysis. The first is the measure of terrain ruggedness published in "Terrain ruggedness and land cover: Improved data for most research designs" by Andrew Shaver, David B. Carter, and Tsering Wangyal Shawa (2019). This data is available at the 1 x 1 km level, so we can join it into our existing data with spatial joins to include in our analysis. While this measure is relatively new, it has already made an extensive impact on the literature, including with use

in other conflict research (eg, Carter, Shaver, and Wright, 2019), so it will allow us to compare and validate our results. We've downloaded the ruggedness data from the Harvard dataverse.

We'll also use the PRIO-GRID data to include the average mountainous terrain in our analysis, which we've downloaded from the PRIO-GRID site. We expect this to be less precise and informative than the ruggedness measure above, but given the extensive historical use of 'mountainous terrain' as a variable in conflict research, we feel it will provide for interesting comparison with historical analyses (eg, Fearon and Laitin, 2003).

Extrastate Conflict Inclusion

We'll be able to use the original UCDP dataset with revised filtering to test this mechanism, so there isn't new data to gather, only slight modifications to the filters used in generating the dataset.

References

Cederman, Lars-Erik, Seraina Rüegger, and Guy Schvitz. 2022. "Redemption through Rebellion: Border Change, Lost Unity, and Nationalist Conflict." American Journal of Political Science 66 (1): 24–42. https://doi.org/10.1111/ajps.12634.

Carter, David B., Andrew C. Shaver, and Austin L. Wright. 2019. "Places to Hide: Terrain, Ethnicity, and Civil Conflict." The Journal of Politics, September. https://doi.org/10.1086/704597.

Fearon, James D., and David D. Laitin. 2003. "Ethnicity, Insurgency, and Civil War." American Political Science Review 97 (1): 75–90. https://doi.org/10.1017/S0003055403000534.

Shaver, A., Carter, D. B., & Shawa, T. W. (2019). Terrain ruggedness and land cover: Improved data for most research designs. Conflict Management and Peace Science, 36(2), 191–218. https://doi-org.ezp-prod1.hul.harvard.edu/10.1177/0738894216659843

Sources Used

https://stackoverflow.com/questions/16498849/logistic-regression-with-robust-clustered-standard-errors-in-r

https://www.r-bloggers.com/2021/05/clustered-standard-errors-with-r/

https://statisticsglobe.com/name-variables-in-for-loop-dynamically-in-r

https://unc-libraries-data.github.io/R-Open-Labs/Extras/Stargazer/Stargazer.html