Final Report

Michael Li

March 1, 2021

Introduction

At the tail end of former President Trump's term, the rise of social unrest within the United States became an undeniable reality of American life. The Black Lives Matter movement and the recent violent insurrection of pro-Trump supporters at the Capitol are just two of the campaigns by U.S. civilians of dissatisfaction with U.S. institutions and state actors. The events associated with these protests garnered international attention—- mostly solidarity for the BLM movement and condemnation for the Capitol Hill riot—- but the ugly truths surrounding social movements in other countries, especially autocratic nations where governmental abuses are most egregious, often fly under the radar. Given this, we are interested in investigating social unrest globally and developing a better understanding of the power that grassroots movements hold in challenging governments. It is imperative to recognize that civil resistance is also far from monolithic; oftentimes, the seeds of the most visible campaigns are sown in the form of smaller movements. For example, while media and academia have recognized the political potency of the 18-day occupation of Tahrir Square in 2011 that led to the removal of Egyptian President Hosni Mubarak, there has been scant attention given to the efficacy of the concurrent wave of labor strikes and road-blocking demonstrations in the country at large. Literature on labor strikes and other acts of omission is particularly thin, as this form of protest is less visible than many of its counterparts [1]. As alluded to previously, the extent to which 'quieter' protests are understood in authoriarian regimes is particularly abysmal. Existing studies showing the success of small, tactical movements at acheiving political aims are limited to single-country analyses and only examine first-world countries [2]. The primary goals for this study are as follows:

- What are the most important determinants of a successful protest?
- What are the cumulative effects of repeated instances of social resistance against a given political target?
- Does the use of violence by a government lead to a 'chilling effect', whereby future protests are deterred?

Data

This paper uses the NAVCO 3.0 dataset assembled as part of the Nonviolent and Violent Campaigns and Outcomes (NAVCO) data project, the first and only project to collect systematic, global data on both violent insurgencies and nonviolent civil resistance campaigns [3]. This was done by manually coding variables of interest for events according to news reports. Coders used news wire reports from Agence France Presse (AFP), the world's oldest and a leading international news agency, to access relevant articles. After binning the articles according to country and time period of interest, coders followed a rigorous set of instructions specified by Chenoweth, Pinckney, and Lewis to identify relevant events and appropriately translate text in articles to values usable for data analysis.

There are two reasonable choices for response variables: (i) the regime response to campaign activity (ii) the number of fatalities or injuries caused due to campaign activity. This paper will primarily focus on the former response variable, as the number of casualties is intrinsically tied to the number of participants involved in the campaign and is only tangentially related to efficacy. The predictor variables fall into seven main categories: location, date, perpetrator characteristics, perpetrator goals, perpetrator methods, target characteristics, and economic impact. We will focus heavily on the subset of predictors related to the methods that perpetrators utilize. Existing literature suggests that tactical innovation and the use of variant methods is critical for the success of civil resistance [4]. However, it is unclear which innovations are effective, warranting an analysis of the comparative utility of different forms of protest. For the cumulative effects research goal, we created a variable to track the number and magnitude of events in the three months prior to a major concession. The designation of a three month time frame was chosen based on the notion that social resistance movements rarely generate continued attention beyond several months following their conclusions.

```r
# reclean data to categorize NA values in continuous variables as zero (assume no or negligble values f
df_clean[is.na(df_clean$damage), c("damage")] <- "minor"
df_clean[is.na(df_clean$econ_impact), c("econ_impact")] <- "none"
df_clean[is.na(df_clean$fatal_casu), c("fatal_casu")] <- 0
df_clean[is.na(df_clean$injuries), c("injuries")] <- 0
#df_clean[is.na(df_clean$num_partic_event), c("num_partic_event")] <- 0

# df_clean %>%
#   filter(!is.na(nv_categ), !is.na(nc_type), !is.na(nv_commission),
#          !is.na(nv_concentration))

## reorder variables actor_id, verb_10?, camp_goals, tactical_choice, st_posture
df_clean$actor_id <- factor(df_clean$actor_id, ordered = FALSE)
df_clean$camp_goals <- factor(df_clean$camp_goals, ordered = FALSE)
df_clean$actor_id <- factor(df_clean$actor_id, ordered = FALSE)
df_clean <- df_clean %>%
  dplyr::mutate(tactical_choice = factor(case_when(as.integer(tactical_choice) == 1 ~ 3.0, # violent (1
                           as.integer(tactical_choice) == 2 ~ 1.0, # non-violent (2) --> 1
                           TRUE ~ 2.0), order = TRUE, levels = c(1, 2, 3))) # mixed(3) --> 2
```

```r
# collapse levels of categorical variables
df_ord <- df_clean
levels(df_ord$actor_id) <- list(state = c("state", "local state"),
                                international = c("international"),
                                `non-state` = c("non-state", "nonaligned"))
levels(df_ord$camp_goals) <- list(`Regime Change` = c("Regime Change"),
                                `Institutional Reform` = c("Institutional Reform"),
                                `Policy Change` = c("Policy Change"),
                                `Territorial Secession` = c("Territorial Secession"),
                                `Autonomy` = c("Autonomy", "Anti-occupation"),
                                `Unknown` = c("Unknown"),
                                `Missing` = c("Missing"))
```

```r
df_pwp <- df_clean %>%
  group_by(country_name, year) %>%
  filter(target_3 %in% c("COP", "GOV", "JUD", "MIL", "LLY", "SPY", "LEG", "TOP")) %>%
  ungroup()

df_pwp <- df_pwp %>% filter(!is.na(tactical_choice) & !is.na(st_posture))

df_pwp$id <- df_pwp %>% group_by(country_name, year) %>%
  group_indices()

df_pwp <- df_pwp %>%
  ungroup() %>%
```

```r
  arrange(id, date) %>%
  group_by(id) %>%
  dplyr::mutate(diff_time = as.double(date - lag(date))) # get time difference b/w protests

df_pwp$diff_time[is.na(df_pwp$diff_time)] <- 0
df_temp <- df_pwp %>% group_by(id) %>%
  dplyr::mutate(Stop = cumsum(diff_time) + 1)
df_pwp$Stop <- df_temp$Stop

# classify governmental responses into binary variable for concessions Y/N
df_pwp <- df_pwp %>%
  dplyr::mutate(concessions = if_else(st_posture == "Full Accomodations" |
                        st_posture == "Material Concessions" |
                        st_posture == "Nonmaterial Concessions", 1, 0 ))

# consolidate same-day protests
df_pwp <- df_pwp %>%
  group_by(id, date) %>%
  dplyr::mutate(num_protests_day = n()) %>%
  ungroup()

df_pwp <- df_pwp %>%
  group_by(id, date) %>%
  slice(which.min(st_posture)) %>%
  ungroup()
```

```r
# filter to get final concession made in response to a campaign
df_wce_conc <- df_pwp %>%
  filter(concessions == 1) %>%
  group_by(id) %>%
  slice(n()) %>%
  ungroup()

df_wce <- rbind(df_pwp %>%
                       filter(concessions == 0), df_wce_conc) %>%
  arrange(id, date) %>%
  group_by(id) %>%
  dplyr::mutate(last_concession = min(which(concessions == 1 | row_number() ==n()))) %>%
  filter(row_number() <= last_concession) %>%
  dplyr::select(-last_concession)

# match start dates to match events with complete information
df_wce <- df_wce %>%
  group_by(id) %>%
  dplyr::mutate(Start = lag(Stop)) %>%
  ungroup()
df_wce$Start[is.na(df_wce$Start)] <- 0

# convert id to a double
df_wce$id <- as.double(df_wce$id)

# get number of protests
df_wce <- df_wce %>%
  group_by(id) %>%
```

```r
  #filter(concessions == 1) %>%
  dplyr::mutate(enum = row_number()) %>%
  dplyr::mutate(enum = if_else(enum >= 8, 8, as.double(enum))) %>%
  ungroup()


# continent variable
df_wce <- df_wce %>%
  dplyr::mutate(region = case_when(country_name %in% c("algeria", "egypt",
                                                        "kenya", "libya",
                                                        "madagascar", "morocco",
                                                     "sierra_leone", "south_sudan",
                                                     "sudan", "tanzania", "tunisia") ~ "Africa",
                               country_name %in% c("iraq", "jordan", "syria",
                                                        "turkey", "yemen") ~ "Middle East",
                               country_name %in% c("bahrain", "china",
                                                        "india", "pakistan",
                                                        "south_korea", "uzbekistan") ~ "Asia",
                               country_name %in% c("estonia", "ukraine") ~ "Europe",
                               TRUE ~ "Americas")) # Mexico, U.S.
```

```r
df_wce$tactical_choice <- as.integer(df_wce$tactical_choice)

# checkWCE(as.data.frame(df_wce), id = "id", event = "concessions", start = "Start",
#           stop = "Stop", expos = "tactical_choice")
```

```r
df_wce <- df_wce %>%
  dplyr::select(id, date, geo_scope, actor_id, camp_goals, tactical_choice, damage, econ_impact,
              fatal_casu, injuries,
              concessions, num_protests_day)
df_wce$damage <- as.double(df_wce$damage)
df_wce$econ_impact <- as.double(df_wce$econ_impact)
df_wce[is.na(df_wce$damage), c("damage")] <- 0
df_wce[is.na(df_wce$econ_impact), c("econ_impact")] <- 0
df_wce[is.na(df_wce$fatal_casu), c("fatal_casu")] <- 0
df_wce[is.na(df_wce$injuries), c("injuries")] <- 0
#df_wce[is.na(df_wce$num_partic_event), c("num_partic_event")] <- 0
df_wce$doy <- as.double(strftime(df_wce$date, format = "%j"))
```

```r
df_wce$geo_scope <- as.character(df_wce$geo_scope)
df_wce$actor_id <- as.character(df_wce$actor_id)
df_wce$camp_goals <- as.character(df_wce$camp_goals)

non_protest_day_maker <- function(df) {
  ly <- seq(from = 1992, to = 2012, by = 4) # years that are leap years
  year <- as.double(format(date, "%Y")) # extract year
  days_ly <- 1:366
  days_nly <- 1:365
  ly_indicator <- year %in% ly
  ly_newdays <- days_ly[!(1:366) %in% df$doy]
  nly_newdays <- days_nly[!(1:365) %in% df$doy]
  if(ly_indicator) {
    days_to_add <-days_ly[!(1:366) %in% df$doy]
  }
```

```r
  else {
    days_to_add <- days_nly[!(1:365) %in% df$doy]
  }
  num_rows <- length(days_to_add)
  id <- df$id[1]
  zeros <- rep(0, num_rows)
  new_rows <- bind_cols(rep(id, num_rows), rep(as.Date("1982-01-13", format = "%m/%d/%y"), num_rows), re
                  rep("", num_rows), zeros, zeros, # camp_goals, tactical_choice, damage
                  zeros, zeros, zeros, # econ_impact, fatal_casu, injuries
                  zeros, zeros, # num_partic_event, concessions, num_protests_day
                  days_to_add) # doy
  colnames(new_rows) <- c("id", "date", "geo_scope", "actor_id", "camp_goals",
                        "tactical_choice", "damage", "econ_impact", "fatal_casu", "injuries",
                        "concessions", "num_protests_day", "doy")
  df_comb <- rbind(df, new_rows)
  return(df_comb)
}
```

```r
# match start dates to match events with complete information
df_pwp_final <- df_pwp

# convert id to a double
df_pwp_final$id <- as.double(df_pwp_final$id)

# get number of protests that led to concessions
df_pwp_final <- df_pwp_final %>%
  group_by(id) %>%
  dplyr::mutate(row_count = row_number()) %>%
  filter(concessions == 1 | row_count == max(row_count)) %>%
  # dplyr::mutate(enum = if_else(concessions == 1,
  #                              row_number(),
  #                              as.integer(0))) %>%
  dplyr::mutate(enum = row_number()) %>%
  dplyr::mutate(enum = if_else(enum >= 8, 8, as.double(enum))) %>%
  ungroup()

# match start dates to events with complete info
df_pwp_final <- df_pwp_final %>%
  group_by(id) %>%
  dplyr::mutate(Start = lag(Stop)) %>%
  ungroup()
df_pwp_final$Start[is.na(df_pwp_final$Start)] <- 0

# continent variable
df_pwp_final <- df_pwp_final %>%
  dplyr::mutate(region = case_when(country_name %in% c("algeria", "egypt",
                                                      "kenya", "libya",
                                                      "madagascar", "morocco",
                                                    "sierra_leone", "south_sudan",
                                                    "sudan", "tanzania", "tunisia") ~ "Africa",
                               country_name %in% c("iraq", "jordan", "syria",
                                                      "turkey", "yemen") ~ "Middle East",
                               country_name %in% c("bahrain", "china",
                                                      "india", "pakistan",
```

```r
                                                  "south_korea", "uzbekistan") ~ "Asia",
                                    country_name %in% c("estonia", "ukraine") ~ "Europe",
                                    TRUE ~ "Americas")) # Mexico, U.S.
```

```r
df_chill <- df_clean %>%
  filter(target_3 %in% c("COP", "GOV", "JUD", "MIL", "LLY", "SPY", "LEG", "TOP")) %>%
  dplyr::mutate(repress = as.double(st_posture) %in% 5:7) %>%
  group_by(repress) %>%
  dplyr::mutate(repress_id = as.double(row_number())) %>%
  dplyr::mutate(repress_id = if_else(repress, repress_id, 0)) %>%
  ungroup()
```

```r
df_chill_condense <- df_chill[, c("country_name", "date")]
# functions to tell if events fell within a year of a repressive response
prior_detector <- function(repress_id, date, country){
  temp <- df_chill_condense %>%
    filter(country_name == country)
  dates <- temp$date
  fun <- sum(date - 365 < dates & date > dates)
  num <- ifelse(repress_id > 0, fun, 0)
  return(num)
}

# near_detector <- function(repress_id, date, country){
#   temp <- df_chill_condense %>%
#     filter(country_name == country)
#   dates <- temp$date
#   fun1 <- sum(date - 365 < dates & date > dates)
#   fun2 <- sum(date < dates & date + 365 > dates)
#   num1 <- ifelse(repress_id > 0, fun1, 0)
#   num2 <- ifelse(repress_id > 0, fun2, 0)
#   return(list(num1, num2))
# }

post_detector <- function(repress_id, date, country){
  temp <- df_chill_condense %>%
    filter(country_name == country)
  dates <- temp$date
  fun <- sum(date < dates & date + 365 > dates)
  num <- ifelse(repress_id > 0, fun, 0)
  return(num)
}

prior_detector <- Vectorize(prior_detector)
post_detector <- Vectorize(post_detector)
```

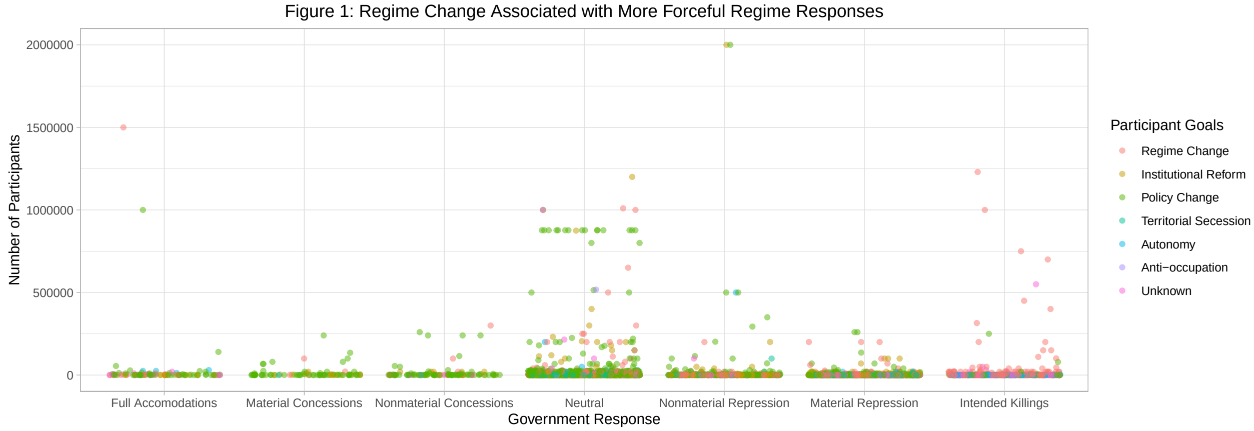
```r
# vectorize? Takes a couple minutes to run
df_chill_final <- df_chill %>%
  dplyr::mutate(prior = prior_detector(repress_id, date, country_name)) %>%
  dplyr::mutate(post = post_detector(repress_id, date, country_name))
```

```r
df_chill_fn <- df_chill_final %>%
  filter(repress_id > 0)
```

```r
# chilling effect eda
df_chill_eda <- df_chill_fn %>%
  pivot_longer(cols = c("prior", "post"),
               names_to = "Timing",
               values_to = "freq_protests")

mu <- ddply(df_chill_eda, "Timing", summarise, grp.mean = mean(freq_protests))
```

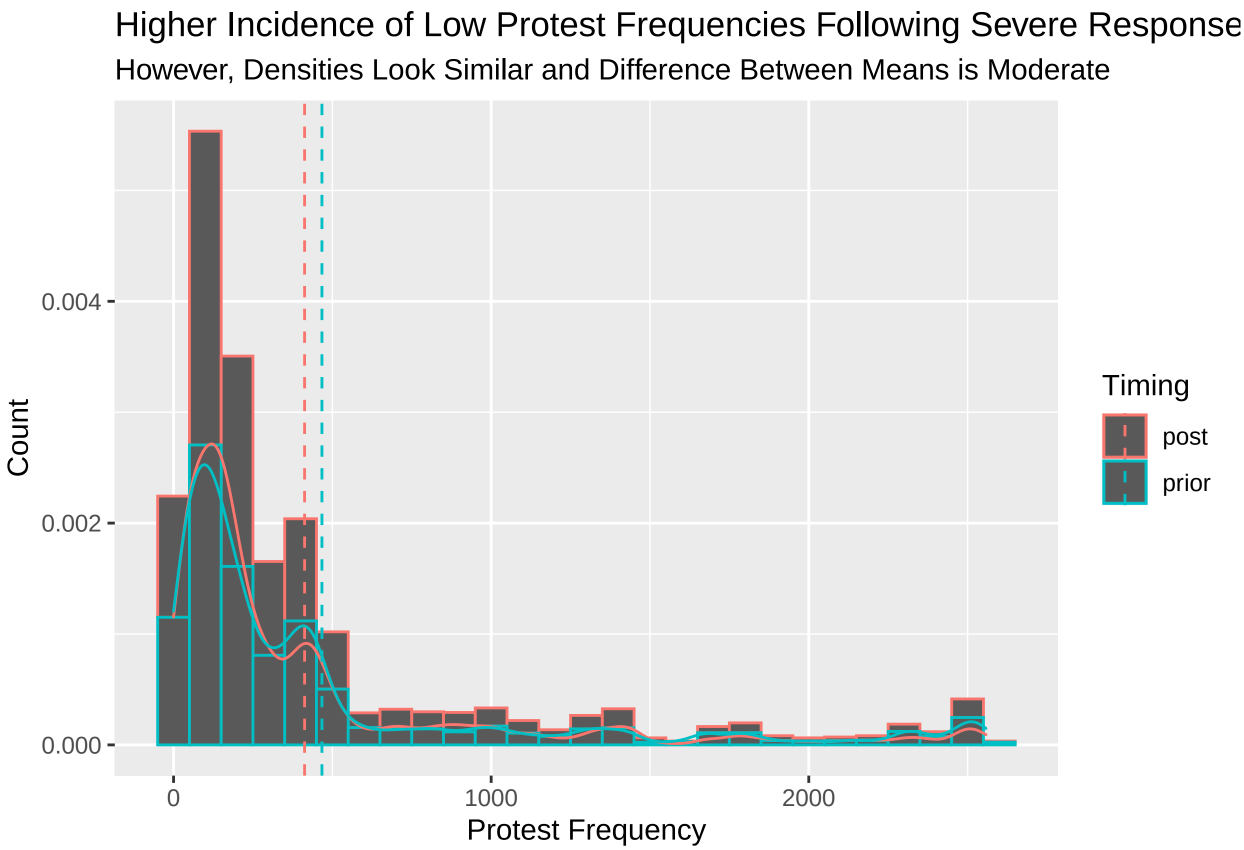
Exploratory Data Analysis

Figure 1: Regime Change Associated with More Forceful Regime Responses



There are several notable insights from this figure. Regime responses of the highest severity, in which the regime resorts to repression with the intent of killing participants, appear to be triggered be triggered by campaigns in which participants are requesting regime change. This trend seems highly plausible: direct challenges against autocratic regimes would be likely to cause governments to lash out expeditiously and forcefully. Figure 1 also seems to offer some evidence for the idea that large campaigns with greater than 50,000 participants tend to elicit neutral regime responses.

```r
ggplot(df_chill_eda, aes(x = freq_protests, color = Timing)) +
  geom_histogram(aes(y =..density..), binwidth = 100) +
  geom_density(alpha = 0.2) +
  geom_vline(data = mu, aes(xintercept = grp.mean, color = Timing),
             linetype = "dashed") +
  labs(title = "Higher Incidence of Low Protest Frequencies Following Severe Responses",
       subtitle = "However, Densities Look Similar and Difference Between Means is Moderate",
       x = "Protest Frequency",
       y = "Count")
```



Methodology

Determinants of Protest Success

To model the factors that influence protest success the most, we opted for an ordinal logistic regression model, otherwise known as a cumulative logit model for ordinal responses. Protest success will be simply defined as the response used by the government, with full accomodation denoting the highest level of success and material and/or physical repression resulting in death denoting the lowest. The response variable has seven levels in total. Since we are examining a classification problem with ordered classes, it would not be appropriate to use a multinomial logistic regression model nor other similar models for nominal variables. Other classes of models, including the ordered probit model, are also commonly implemented for classification problems involving an ordinal response variable. In contrast to the ordered probit model, logistic regression affords greater interpretability; the coefficients β in logistic regression allow for natural interpretations in terms of log odds. Hence, the logit link function is preferrable since the stated goals focus more on the interpretability of the predictors' relationships with the response. A cumulative logit model for an ordinal response with J levels is similar to a binary logistic model, as the levels 1 to j form a single category and the levels $j+1$ to J form a second category. The log-odds of falling at or below the category comprised of levels 1 to j has a linear relationship with the predictors. The logits of the cumulative probabilities are therefore given by the following form:

$$logit(P(Y \leq j)) = \frac{P(Y \leq j)}{P(Y > j)} = \alpha_j + \beta x,\ \ for\ j = 1,\ 2\ ...\ J-1$$

The key assumption underpinning the ordinal logistic regression model is the proportional odds assumption, which implies that only a single set of coefficients is necessary for each covariate in the model. This is because the assumption states that the coefficients that describe the relationship between the lowest ordinal category versus all higher categories are the same as those that describe the relationship between the next lowest category and all higher categories, and so on. To determine how well each predictor upholds the proportional odds assumption, we created a function to calculate the log odds that governmental response was greater or equal to a given level of governmental response from individual logistic regressions with a single predictor. We then tested the equality of coefficients for a given predictor across its corresponding binary logistic regression models. The differences between the logits for each partition of the dependent variable should be similar for a given predictor regardless of that predictor's level if the proportional odds assumption holds. A visualization indicating the validity of the proportional odds assumption for our predictor set has been included in the appendix.

The base model containing relevant protest-level predictors is specified below:

$$\begin{aligned}
logit(P(Government\ Response \leq j)) = \beta_{0j} &+ \beta_1 I(Scope_i = local) \\
&+ \beta_2 I(Scope_i = subnational) + \beta_3 I(Scope_i = national) \\
&+ \beta_4 I(Scope_i = international\ regional) + \beta_5 I(Scope_i = global) \\
&+ \beta_6 I(ActorType_i = state) + \beta_7 I(ActorType_i = non - state) \\
&+ \beta_8 I(ActorType_i = international) + \beta_9 I(ActorType_i = non - aligned) \\
&+ \beta_{10} I(ActorType_i = local\ state\ actor) + \beta_{11} I(Goals_i = regime\ change) \\
&+ \beta_{12} I(Goals_i = institutional\ reform) + \beta_{13} I(Goals_i = policy\ change) \\
&+ \beta_{14} I(Goals_i = territorial\ secession) + \beta_{15} I(Goals_i = greater\ autonomy) \\
&+ \beta_{16} I(Goals_i = anti - occupation) + \beta_{17} I(Tactic_i = violent) \\
&+ \beta_{18} I(Tactic_i = non - violent) + \beta_{19} I(Tactic_i = mixed) \\
&+ \beta 20 NumPartcipants + \beta_{21} Damage + \beta 22 EconImpact \\
&for\ j = 1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7
\end{aligned}$$

Cumulative Protest Effects

Protests do not occur in a vacuum; oftentimes, individual protests are part of a much larger campaign designed to bring about sweeping institutional change. Given this, we posit that there is likely to be a cumulative effect of protests on governmental response, whereby past protests influence the success of future ones within the same campaign. We will follow the inclusion rule used by Pinckney that approximates the definition of a campaign within the civil resistance literature. A campaign must consist of at least three distinct physical events separated by less than one year, with the events meaningfully linked through common actors or goals [5]. To categorize event-level data into campaigns, we will use the NAVCO 1.3 list of maximalist campaigns. Since NAVCO 3.0 is a disaggregated dataset that does not sort individual protests into campaigns, we will analyze the stated protest goals, target, actors, and timing of the events in our dataset and assign them to the campaigns to which they are most closely aligned. Protests that do not clearly seem to be part of any campaign will be discarded.

Once the data has been sorted into campaigns, we will use a weighted cumulative exposure (WCE) model. The WCE model has wide applications within the field of pharmacoepidemiology. In this case, the WCE model is advantageous for three reasons. First, it is able to capture the cumulative effect of past protests on governmental response. Second, it is able to weight the relative importance of past protests depending on their timing. Finally, information concerning the severity of an event can be integrated. We chose this model over recurrent event models such as the Anderson-Gill (AG) model, which suffer from the independence assumption. The AG model assumes that the instantaneous risk of experiencing an outcome of interest at time t remains the same irrespective of whether or not previous events have been experienced. This makes recurrent event independence a poor assumption for this problem. The WCE model is specified as follows:

$$WCE(u) = \sum_{t \leq u} w(u-t)X(t)$$

Here u is the current time when governmental response is evaluated, $t \leq u$ indexes times of protests preceding u, and $X(t)$ represents the past protest intensity at time t. Here intensity will be defined as the number of participants in the protest, as this is a reasonably valid way to measure the strength of a protest.Given that our ultimate reponse variable of interest is quantitative, we will follow the Danieli et al. procedure for implementing the WCE model to assess longitudinal changes in the outcome [6]. This WCE model will be incorporated into a linear effects mixed model, whereby we model governmental response Y, a repeated-over-time outcome, at the k^{th} protest for campaign i. The model will thus take the form below:

$$Y_i(t_{ik}) = \beta_0 + b_{i,0} + \beta_{WCE} WCE_i(t_{ik}) + \sum_{s=1}^{p} (\beta_s Z_{s,ik}) + \sum_{s=1}^{r} (b_{i,s} L_{s,ik}) + \epsilon_{ik}$$

β_0 represents the expected value of the governmental response to a campaign at time t_{ik} assuming that there were no prior protests in the campaign. This is the response that would be expected given that the values of all p fixed effect covariates were set to zero. $b_{i,0}$ $\mathcal{N}(0, \sigma_b^2)$ represents the random intercept for campaign i. The fixed effect, β_{WCE}, represents the expected change in governmental response associated with a unit increase in the WCE metric. β_s represents the fixed effect of the covariates Z_s, while $b_{i,s}$ represents the random effect of a set of covariates L_s that are included in Z_s. These covariates L_s will be the grouping factors that we are seeking to control, namely the geographic scope of protests and the type of actor involved in the protest. We chose to use a mixed effects model because there is reason to believe that observations are not necessarily independent of one another. Protests are likely to 'interact' with one another. For example, an actor conducting multiple protests would be expected to be met with similar governmental responses for each protest. This model allows us to therefore capture variance in the dependent governmental response variable due to relevant clusters.

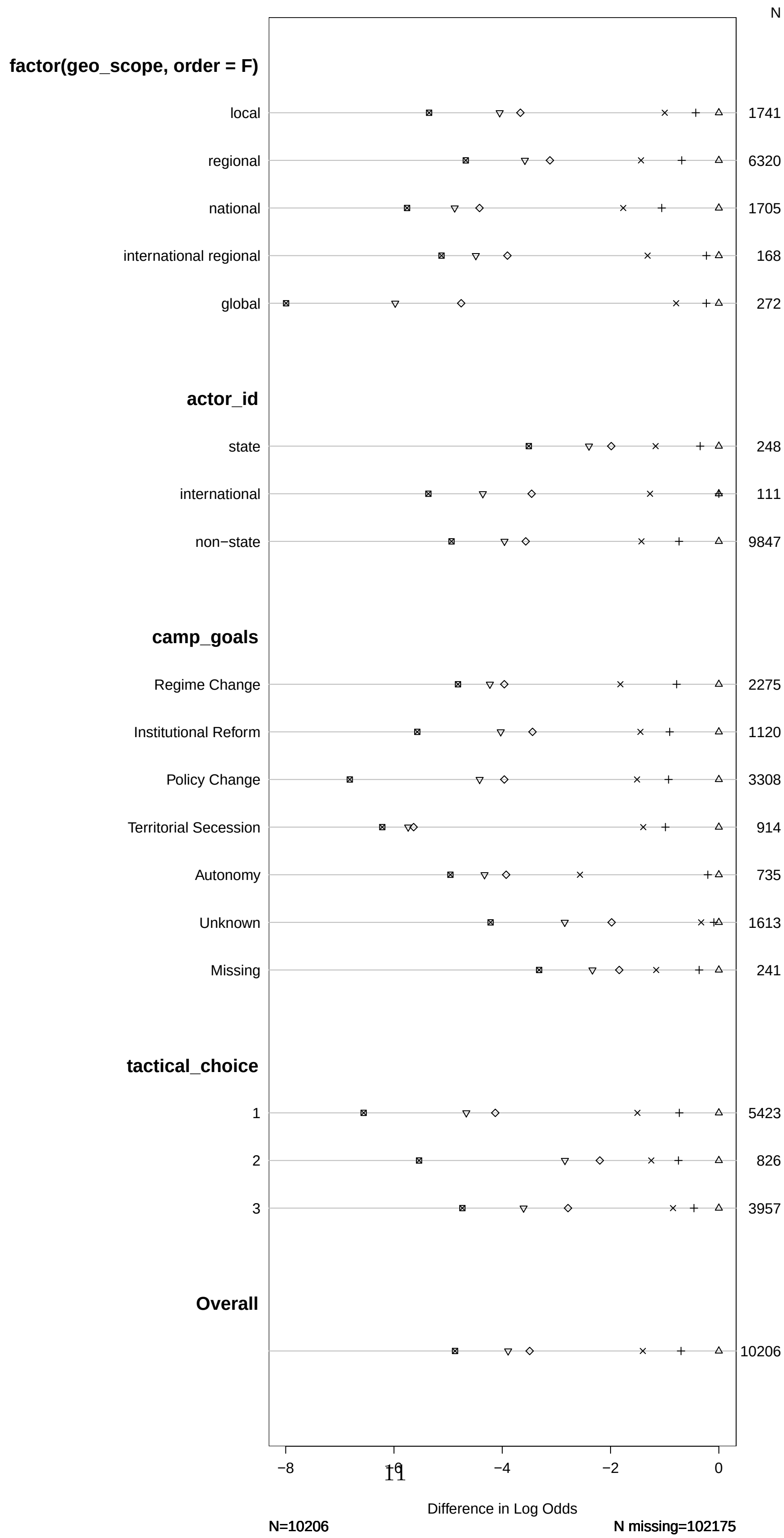
Chilling Effect

To determine whether or not a chilling effect stemming from severe governmental responses exists, we will be examining the frequency of protests a year prior to and following such a response. A severe response will be defined as any governmental response that is of level five, which corresponds to non-material and non-physical repression, or higher. A two proportions Z-test will be used to examine if the difference in protest frequencies is statistically significant. This hypothesis test is effective in comparing population proportion differences between two groups, and it is chosen over the Fisher-exact test because of the sufficiently large sample size.

Results

Protest Success

Figure 3: Proportional Odds Assumption Satisfied by Data



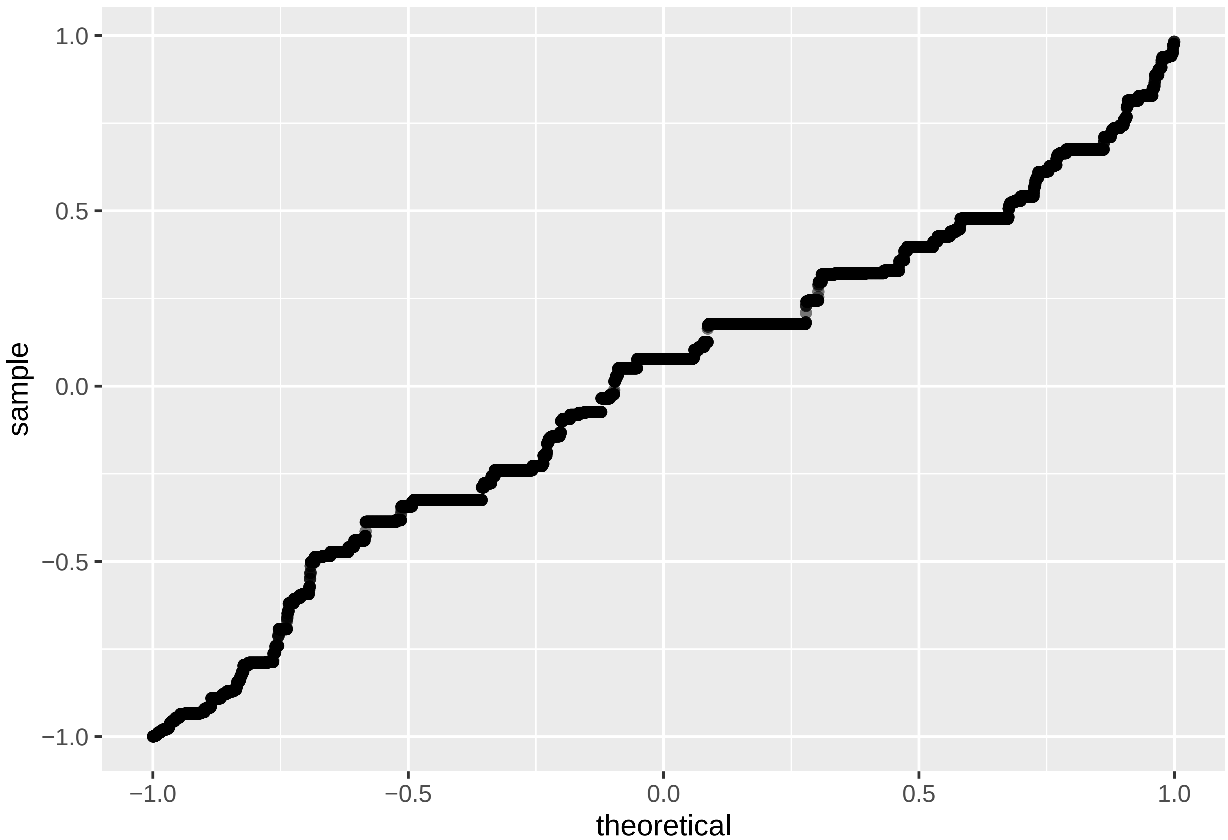
```r
m_ord <- polr(st_posture ~ factor(geo_scope, order = F) + actor_id + camp_goals + tactical_choice,
          data = df_clean,
          #na.action = na.omit,
          Hess = TRUE)
#summary(m_ord)
mod_output_df <- as.data.frame(summary(m_ord)$coefficients)
mod_output_df$p_value <- pnorm(abs(mod_output_df[, "t value"]), lower.tail = FALSE) * 2
mod_output_df[, c("Value", "p_value")]
```

```
##                                                          Value      p_value
## factor(geo_scope, order = F)regional                0.7734783 2.921515e-45
## factor(geo_scope, order = F)national               -0.4655204 2.473723e-11
## factor(geo_scope, order = F)international regional  0.3819241 1.682136e-02
## factor(geo_scope, order = F)global                 -0.1201351 3.583731e-01
## actor_idnon-state                                   0.3202297 2.159738e-01
## actor_idinternational                               0.5202144 1.310984e-01
## actor_idnonaligned                                  0.4595712 5.398501e-01
## actor_idlocal state                                -1.9232323 1.173458e-01
## camp_goalsInstitutional Reform                     -0.8224311 2.995882e-27
## camp_goalsPolicy Change                            -1.2955020 1.638654e-94
## camp_goalsTerritorial Secession                    -0.7848413 2.380953e-21
## camp_goalsAutonomy                                 -1.1153169 2.742708e-33
## camp_goalsAnti-occupation                          -2.0448334 1.965923e-07
## camp_goalsUnknown                                  -1.4438474 2.395803e-95
## tactical_choice.L                                   2.1314738 0.000000e+00
## tactical_choice.Q                                  -0.2152418 3.621041e-04
## Full Accomodations|Material Concessions            -5.6309534 6.531277e-92
## Material Concessions|Nonmaterial Concessions       -4.8910580 1.709835e-72
## Nonmaterial Concessions|Neutral                    -4.1575006 7.519558e-54
## Neutral|Nonmaterial Repression                     -1.4376824 6.652579e-08
## Nonmaterial Repression|Material Repression         -0.7844356 3.187817e-03
## Material Repression|Intended Killings               0.9791423 2.350933e-04
```

```r
1-pchisq(deviance(m_ord), df.residual(m_ord))
```

```
## [1] 0
```

```r
pres <- presid(m_ord)
ggplot(data.frame(y = pres), aes(sample = y)) +
  stat_qq(distribution = qunif, dparams = list(min = -1, max = 1), alpha  = 0.5)
```




```
#resids(m)
#predict(m_ord, type = "probs")
#autoplot(resids(m_ord), what = "covariate", x  = predict(m, type = "probs"))

#residuals(m_ord)
```

Cumulative Protests

```
#checkWCE(as.data.frame(df_wce_final), id = "id", event = "concessions", start = "Start",
#          stop = "Stop", expos = "tactical_choice")
m_wce <- WCE(as.data.frame(df_wce_final), analysis = "Cox", cutoff = 366, nknots = 1:3,
    constrained = "R", aic = FALSE, MatchedSet = NULL,
    id = "id",
    event = "concessions",
    start = "Start", stop = "Stop", expos = "tactical_choice",
    covariates = c("geo_scope", "damage", "econ_impact", "fatal_casu", "injuries",
                   "num_protests_day"))
#m_wce
#summary(m_wce)
#plot(m_wce)
nonusers <- rep(1, 366)
users <- rep(0, 366)
# for all models
#HR.WCE(m_wce, nonusers, users, allres = TRUE)
```

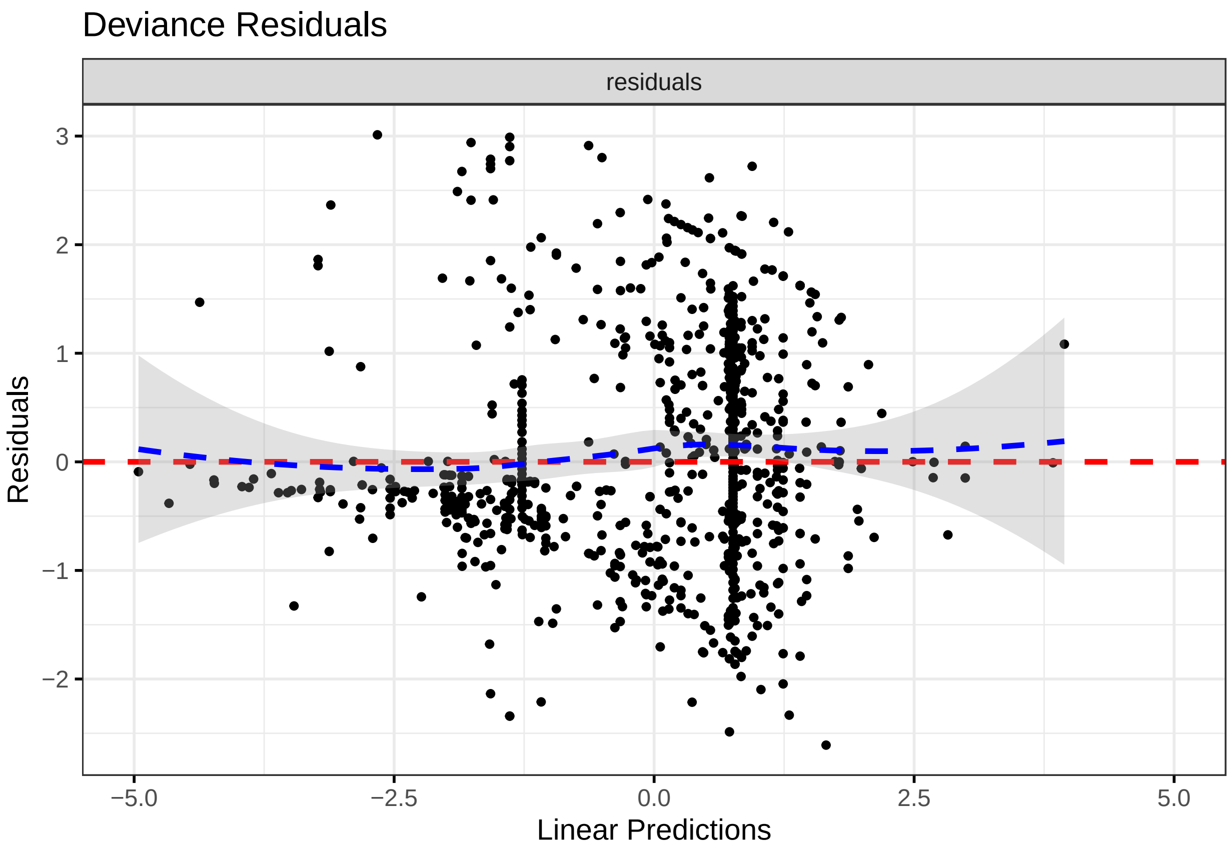
PWP Model

```r
pwp1 <- coxph(Surv(Stop, concessions) ~ cluster(id)+ strata(enum) + factor(tactical_choice, order = F) *
              + region * strata(enum) + factor(geo_scope, order = F) * strata(enum) +
                factor(camp_goals, order = F) * strata(enum),
            data = df_pwp_final)
```

```
## Warning in fitter(X, Y, istrat, offset, init, control, weights = weights, : Ran
## out of iterations and did not converge
```

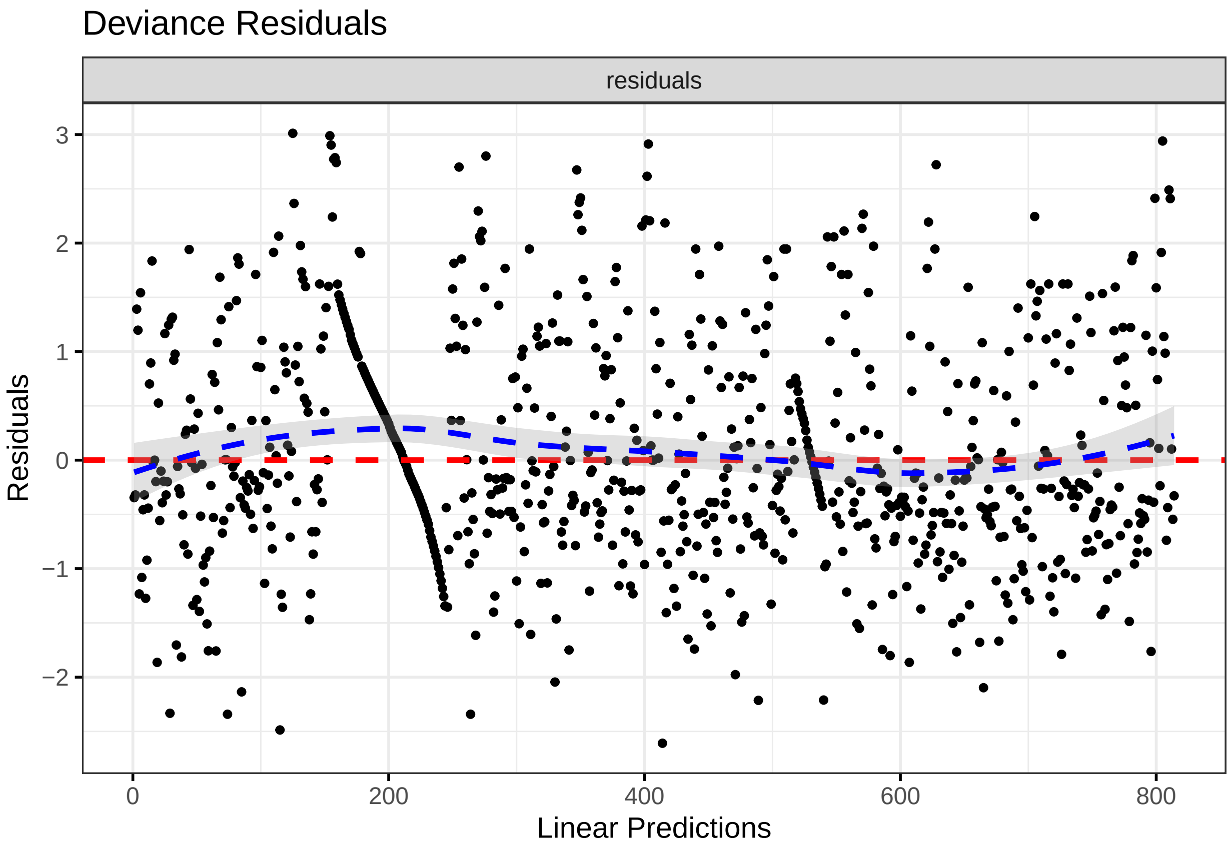
```r
#summary(pwp1)

ggcoxdiagnostics(pwp1, type = "deviance", linear.predictions = T) +
  labs(title = "Deviance Residuals",
       x = "Linear Predictions",
       y = "Residuals") +
  xlim(c(-5,5))
```

```
## `geom_smooth()` using formula 'y ~ x'
## Warning: Removed 10 rows containing non-finite values (stat_smooth).
## Warning: Removed 10 rows containing missing values (geom_point).
```




```r
ggcoxdiagnostics(pwp1, type = "deviance", linear.predictions = F) +
  labs(title = "Deviance Residuals",
       x = "Linear Predictions",
       y = "Residuals")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



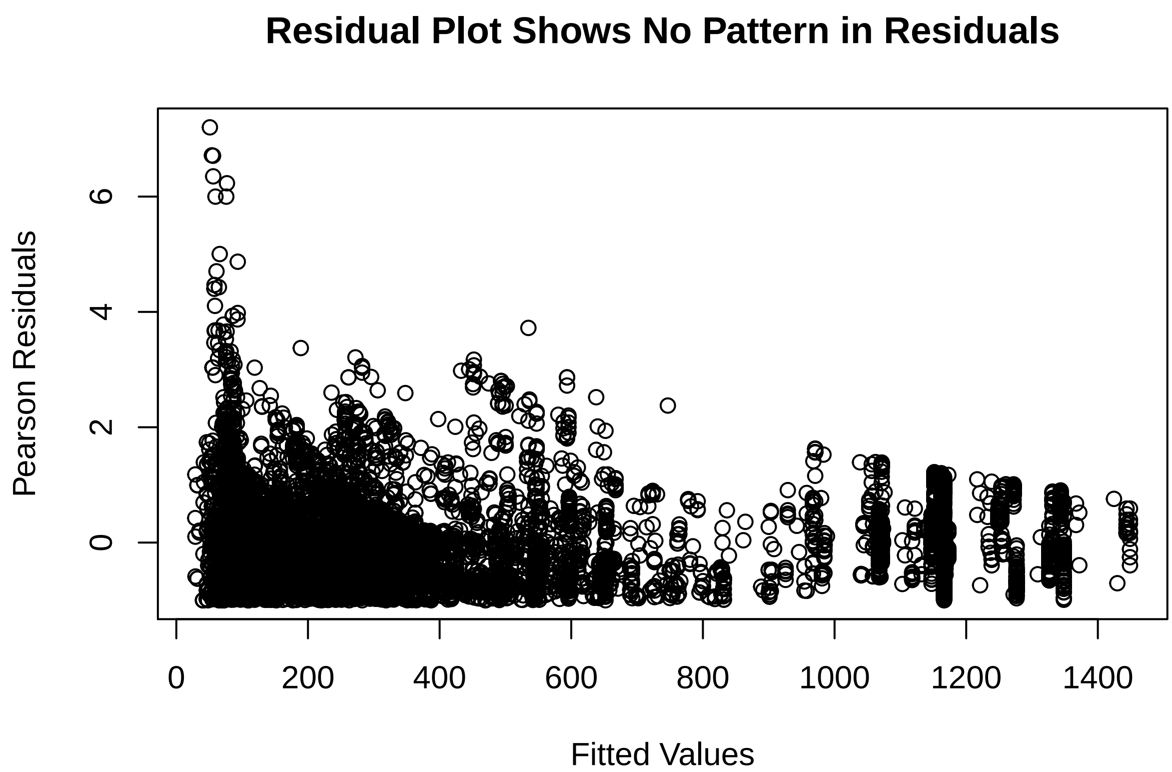
Chilling Effect

```
# create continent variable
df_chill_glm <- df_chill_eda %>%
  dplyr::mutate(region = case_when(country_name %in% c("algeria", "egypt",
                                                         "kenya", "libya",
                                                         "madagascar", "morocco",
                                                      "sierra_leone", "south_sudan",
                                                      "sudan", "tanzania", "tunisia") ~ "Africa",
                                   country_name %in% c("iraq", "jordan", "syria",
                                                         "turkey", "yemen") ~ "Middle East",
                                   country_name %in% c("bahrain", "china",
                                                         "india", "pakistan",
                                                         "south_korea", "uzbekistan") ~ "Asia",
                                   country_name %in% c("estonia", "ukraine") ~ "Europe",
                                   TRUE ~ "Americas")) # Mexico, U.S.

# Poisson Model
m_chill_poiss <- glm(freq_protests ~ Timing + region + geo_scope +
                       actor_id + camp_goals + tactical_choice + st_posture + year,
                     data = df_chill_glm, family = poisson)
#summary(m_chill_poiss)
```

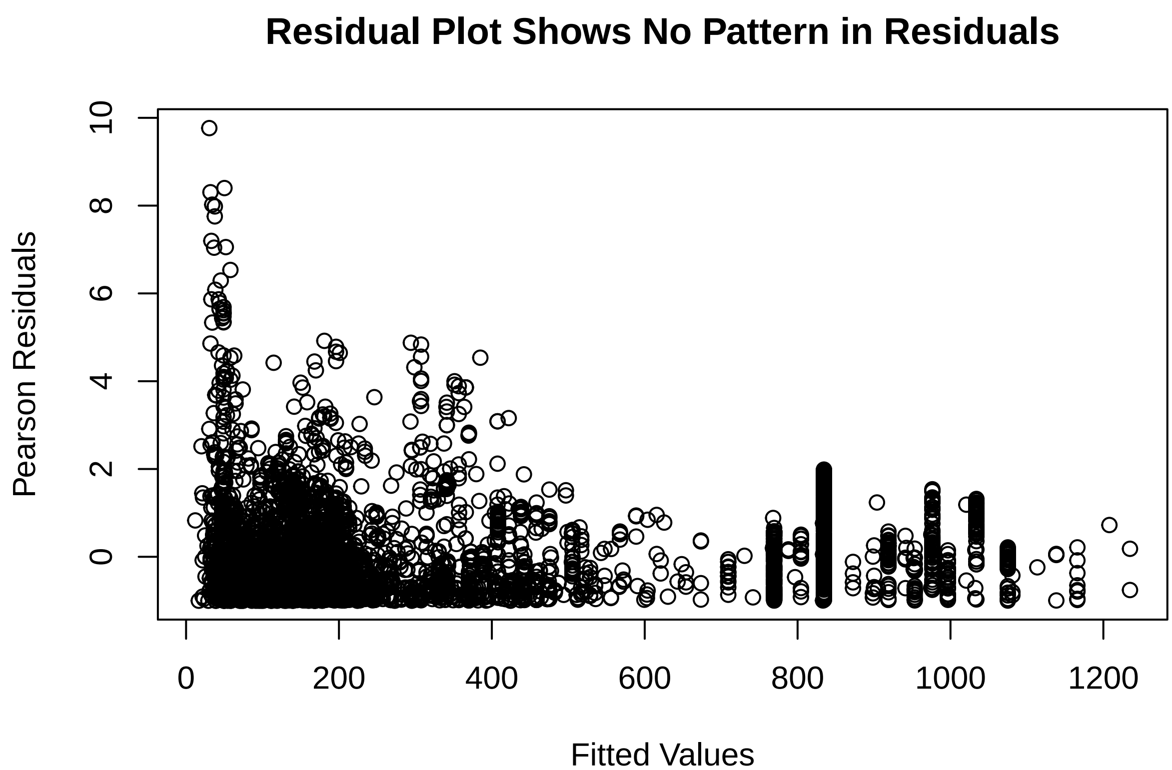
```r
# significant evidence of overdispersion
dispersiontest(m_chill_poiss)
```

```
##
##  Overdispersion test
##
## data:  m_chill_poiss
## z = 53.609, p-value < 2.2e-16
## alternative hypothesis: true dispersion is greater than 1
## sample estimates:
## dispersion
##   199.5395
```

```r
# Negative Binomial Model
m_chill_nb <- glm.nb(freq_protests ~ Timing + region + geo_scope +
                       actor_id + camp_goals + tactical_choice + st_posture + year,
    data = df_chill_glm)

# Negative Binomial Model for mean differences
df_chill_diff <- df_chill_glm %>%
  group_by(repress_id) %>%
  dplyr::mutate(diff = freq_protests[Timing == "prior"] - freq_protests) %>%
  slice_tail() %>%
  ungroup()
m_chill_nb_diff <- glm.nb(abs(diff) ~ region + geo_scope +
                       actor_id + camp_goals + tactical_choice + st_posture + year,
    data = df_chill_diff)

#summary(m_chill_nb_diff)
#summary(m_chill_nb)
plot(m_chill_nb$fitted.values,m_chill_nb$residuals, xlab = "Fitted Values", ylab = "Pearson Residuals",
```



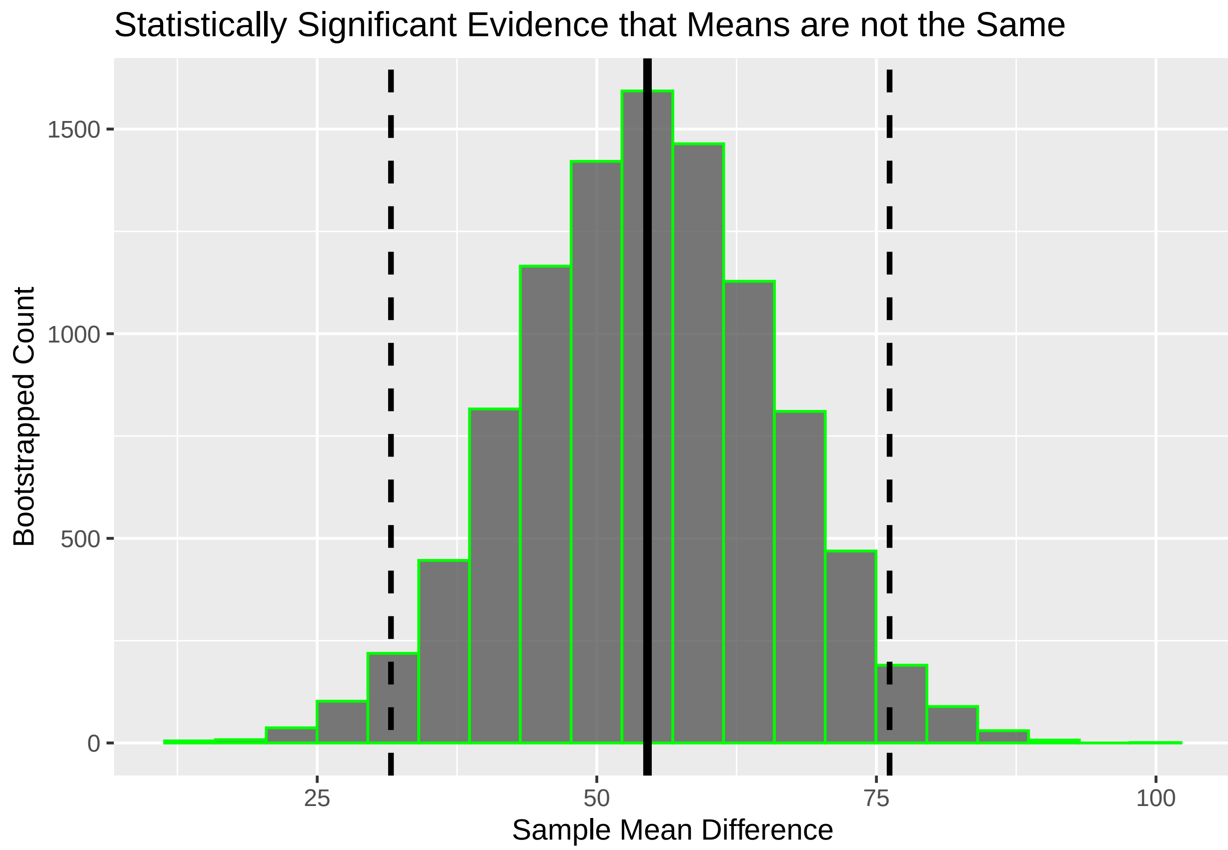

```
plot(m_chill_nb_diff$fitted.values,m_chill_nb_diff$residuals, xlab = "Fitted Values", ylab = "Pearson Re
```




```
# t test for difference between two populations, may be biased
t.test(x = df_chill_fn$prior, y = df_chill_fn$post)
```

```
##
##  Welch Two Sample t-test
##
## data:  df_chill_fn$prior and df_chill_fn$post
## t = 4.7806, df = 10217, p-value = 1.772e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  32.17188 76.89173
## sample estimates:
## mean of x mean of y
##  467.1515  412.6197
```

```
# bootstrap the mean difference of the two samples
bootstrapped <- two.boot(df_chill_fn$prior, df_chill_fn$post, mean, 10000)

bootstrapped_mean_diff <- data.frame(bootstrapped$t)
colnames(bootstrapped_mean_diff) <- "mean_diffs"

ggplot(bootstrapped_mean_diff, aes(x = mean_diffs)) +
  geom_histogram(color = "green", bins = 20, alpha = 0.8) +
  geom_vline(xintercept = bootstrapped$t0, size = 1.5) +
  geom_vline(xintercept = quantile(bootstrapped_mean_diff$mean_diffs, 0.025), size = 1,
             linetype = "dashed") +
  geom_vline(xintercept = quantile(bootstrapped_mean_diff$mean_diffs, 0.975), size = 1,
             linetype = "dashed") +
  labs(title = "Statistically Significant Evidence that Means are not the Same",
       x = "Sample Mean Difference",
       y = "Bootstrapped Count")
```



References

[1] Chenoweth E, Pinckney J, Lewis O. “Days of rage: Introducing the NAVCO 3.0 dataset.” SAGE. Journal of Peace Research. 55 (4): 524-534.

[2] Huet-Vaughn E. “Quiet Riot: Estimating a Causal Effect of Protest Violence.” Working Paper. 1-41.

[3] Chenowth E.

[4] McAdam D. “Tactical Innovation and the Pace of Insurgency.” American Sociological Association. American Sociological Review. 48 (6): 735-754.

[5]

[6] Danieli C, Sheppard T, Costello R, Dixon W, Abrahamowicz M. “Modeling of cumulative effects of time-varying drug exposures on within-subject changes in a continuous outcome.” SSMR. 29(9): 2554-2568.