

PS531 Pre-Analysis Plan Final Paper

Myung Jung Kim

1 Research Question and Theoretical Approach

(1) DESCRIBE A SUBSTANTIVE QUESTION IN SOCIAL SCIENCE. WHAT THEORY ARE YOU ASSESSING? WHY SHOULD ANYONE CARE? (2 PARAGRAPHS)

The international community has been striving to end impunity for grave human rights violations. The effort culminated around 1998 with the rise the International Criminal Court (ICC) and Universal Jurisdiction (UJ) which enabled the overriding of domestic amnesties for international crimes (hereafter, SV amnesties). For this reason, the ICC and UJ's challenge against amnesties has been viewed as a transition from the "era of impunity" to "era of accountability." Surprisingly, however, recent studies report that amnesties that include impunity for international crimes have increased over the past few decades (Mallinder (2012), 95). Why do we witness this trend despite the rise of international regimes challenging amnesties? Under what conditions do states grant amnesty for serious crimes? This paper tackles these questions by providing a theory of amnesty provision for serious crimes in the context of civil wars.

Such international legal development has been the subject of the well-known 'peace vs. justice' debate. While the justice camp argues that the push for anti-impunity would bring peace in the long term (Kim and Sikkink 2010; Olsen, Payne, and Reiter 2010), the peace camp argues that the lack of an amnesty option may leave warlords with no option but to continue fighting and ultimately worsen human rights conditions (Goldsmith and Krasner (2003), Ginsburg (2009), Prorok (2017)). The fundamental assumption of this debate is that the development of international legal regimes would indeed restrict the use of amnesty and hold perpetrators accountable. Yet, the recent finding raises a fundamental question to this debate. This study brings crucial evidence to advance the "justice vs. peace" debate and raise implications about the effectiveness of international law and institutions. Also, by demonstrating the international legal effect on state's granting of amnesty, this study provides a different perspective on conflict bargaining which was traditionally focused on mere state-rebel power-symmetry (Cunningham (2016)).

2 Hypothesis

(2) WHAT ONE OR TWO HYPOTHESES THAT ARISE FROM THE THEORY ARE YOU PLANNING TO ASSESS? WHY OR HOW DOES THE THEORY JUSTIFY YOUR EXPECTATIONS ABOUT THESE HYPOTHESES? (1 OR 2 PARAGRAPHS)

I argue that the international anti-amnesty regimes have conditional effect in discouraging amnesties for serious crimes. I argue that the international regime can deter SV amnesties in electoral democracy. Before the ICC and UJ, states' provision of amnesty had been considered as a matter of

state sovereignty. Despite some human rights treaties that obligate states to prosecute for serious crimes, the international legal standard on amnesty was largely vague before the advent of the two regimes. In this sense, the advent of the anti-amnesty regime demonstrated a clear message to the international community that the state's granting of amnesties for serious crimes goes against international norms and rules. Yet, there is no international enforcement mechanism that punishes states' violation of the anti-amnesty norm. Even the ICC – the most institutionalized anti-amnesty regime – cannot punish its signatory states when they grant an amnesty to a perpetrator of heinous human rights crimes and violate the norm embedded in the ICC Rome Statute. However, existing studies demonstrate that international regimes can still induce compliance from democratic states. This claim is based on the logic that elections in democracies enable domestic constituents to punish leaders who deviate from their public announcements (Fearon (1994)) and that the fear of such punishment, in turn, may deter democratic leaders from violating international agreements (McGillivray and Smith. (2000)). Under this logic, international regimes can induce state compliance by supporting domestic constituents who favor state's compliance with the international regime (Dai (2005)). Based on the domestic constituency model, I argue that the anti-amnesty regimes can deter SV amnesties in states in electoral democracy by empowering pro-justice (or anti-amnesty) constituents who exercise electoral power to punish government that grants amnesties for serious crimes. If my theoretical claim is correct, I should observe the following empirical implication.

***Hypothesis:** After the advent of the anti-impunity regime, states grant less SV amnesties as they face higher levels of constituent costs.*

3 Research Design

3.1 Data

(3) WHAT DATA AND RESEARCH DESIGN WILL HELP YOU ANSWER THIS QUESTION? WHY ARE YOU MAKING THESE CHOICES? (REMEMBER THAT A STATISTICAL MODEL IS NOT A RESEARCH DESIGN.) (2 PARAGRAPHS)

Since it is difficult to make an experiment with state's provision of conflict amnesty, I conduct an observational study. The main subject of comparison is the state's provision of amnesties for serious crimes before and after the emergence of anti-amnesty regimes. Specifically, based on the hypothesis, I should expect the effect of constituent costs on the number of SV amnesties to be conditional by the international justice regime. I use year-1998 as a cutoff point for the emergence of the international justice regime which is widely perceived as the critical juncture for the transition from the era of impunity to the era of accountability (Dancy (2018); Kremaric (2018), Daniels (2020)). Hence, I compare the the size of effects of constituent costs on state's number of amnesties before and after 1998.

The outcome variable is the number of amnesties for serious violations that states granted to rebel groups during civil conflicts, and I use Dancy's Conflict Amnesty Data which provides information on states' issue of amnesty for civil wars from 1945 to 2014 (Dancy (2018)). I make two modification on the original Dancy's Conflict Amnesty dataset. First, as the original dataset identifies amnesties for "serious crimes" based on the *de jure* scope of amnesty law, I add variables for rebel groups' actual involvement in serious crimes. Specifically, I incorporate three different crimes – civilian

killing, child soldier, and sex crimes using the UCDP One-sided violence dataset, the Haer and Böhmelt (2017) dataset and the SVAC dataset respectively.¹ Second, I collapse every yearly observations of dyad-conflict into a dyad-conflict observation in order to prevent overfitting. My unit of analysis is dyad-conflict, and the data have observations of 496 dyad conflicts of 105 countries.

For the before and after 1998 comparison, I categorize civil conflicts based on their years of starts and ends as following:

1. Pre-98 conflicts: conflicts that started and ended before 1998 (e.g., 1980- 1995)
2. Post-98 conflicts: conflicts that started and ended after 1998 (e.g., 2000-2000)
3. Cross-98 conflicts: conflicts that started before 1998 and ended after 1998 (e.g., 1980-2010)
4. Ongoing-98 conflicts: post-98 and cross-98 conflicts jointly

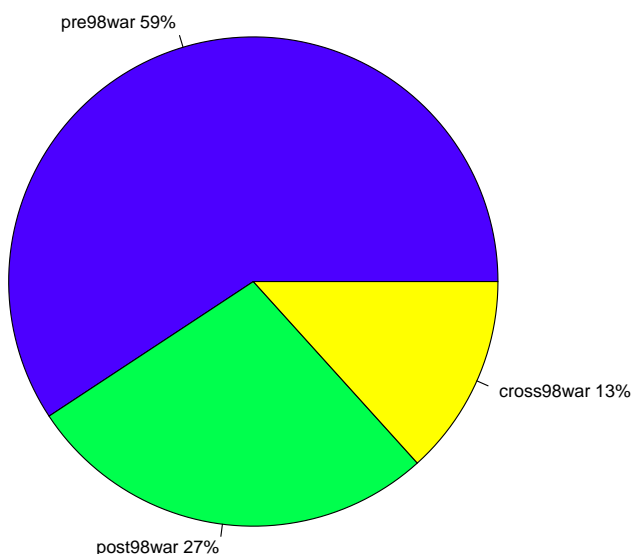


Figure 1: Categories of Conflicts by Time

In the dataset, pre-98 conflicts comprise about 59% of observations (N =295), post-98 conflicts about 27 % (N =136), and cross-98 conflicts about 13% (N = 67). Around 40% of conflicts in the data were ongoing after 1998 (N = 203). Comparing amnesties granted in Pre-98 wars with amnesties granted in Post-98 wars is the sharpest and the most conservative approach (dropping Cross-98 observations). However, including amnesties in cross-98 wars in the analysis is also theoretically appropriate. States usually grant amnesties at the end of a conflict, and therefore, if wars were ongoing in 1998, it is likely that amnesty deals in the war were likely affected by the anti-amnesty regime. Hence, I can make two comparisons: 1) amnesties in pre-98 wars vs. post-98 wars (using `Post98` dummy) and 2) amnesties in pre-98 wars vs. wars that were ongoing in 98 (using `Ongoing98` dummy). My plan is to conduct analyses for both comparison groups (one using as a robust check). However, here, I show pre-98 vs. ongoing-98 comparison which involves more observations. For intuitive understanding, hereafter, I call the ongoing-98 as post-98.

(4) WHAT ARE THE ADVANTAGES AND DISADVANTAGES OF THIS RESEARCH DESIGN TO ADDRESSING YOUR QUESTION (2 PARAGRAPHS)

¹I select these three crimes as they are among the most common human rights crimes committed by rebel groups, and they are the only available data to capture rebel's serious crimes.

Observational study is the only feasible option for this study since it is difficult to conduct a randomized experiment with state's provision of conflict amnesty. Observational studies have multiple advantages as they have low complexity, low cost, and low ethical constraints compared to randomized experiments. Also, as Rosenbaum emphasizes, observational studies can provide useful information when supplemented with tools and strategies such as sensitivity analysis, elaboration of a casual theory and quasi-experimental devices (Rosenbaum (2017)). Especially, they can provide useful information on important variables and their associations—which is often as useful as learning causal relationships. However, observational studies are far limited in drawing causal inference than randomized experimental design. Unlike experimental design, researchers conducting observational studies do not have a control over the data generation process and hence do not know how the independent variables arise. For this reason, researchers face “the statistical uncertainty given a particular set of modeling assumptions” and “the theoretical uncertainty about which modeling assumptions are correct” (Alan S. Gerber and Kaplan (2004)). The principle drawback of observational design comes from its lack of randomization procedure. Since the treated subjects and non-treated subjects are not randomly selected, it suffers from the problems of confounders and bias.

One approach to account for this limitation is to conduct a propensity score matching which adjusts observable covariates and makes an as-if randomized comparison. Propensity score matching pairs subjects based on their propensity score — the conditional probability of treatment given the observed covariates (Observational Studies (2010), 72). By this, it makes it possible to draw and compare the treated and control subjects. There are multiple advantages of using propensity score matching. Propensity scores provide a scalar summary of all the covariate information and there is no limit on the number of covariates for adjustment. Therefore, compare to other matching techniques that need to match multiple covariates, propensity score strategy is simpler, and it also overcomes the problem of dimensionality. Lastly, matching allows researchers to know whether they adjusted covariates enough. This is something impossible by adding “control variables” using a multiple regression model. However, propensity matching strategy is still limited compare to the randomized experiment. First, it only matches observed covariates, and therefore, unobserved covariates can still make bias in treatment effect (Observational Studies (2010), 73). Second, it matches on the variables that researcher chooses to include in the matching process, and therefore there can be bias from this process. Third, it discards unmatched units (Rubin (2001)). Fourth, it is difficult to see the effect of matching variables on the outcome variable (Arane Thavaneswaran (2008)). Therefore, this study also suffers potential bias from omitted covariates. Lastly, since the core purpose of matching is to mimic experiment, it is difficult to have a continuous variable as a treatment. This study has a continuous treatment variable (**democracy**) which makes propensity score matching less appealing. Yet, I conduct a propensity score matching in this pre-analysis for making a comparable comparison.

3.2 Measurement

5) DESCRIBE YOUR MEASURES AND ANY INDICES YOU CONSTRUCT (1 PARAGRAPH)

Post-1998 as the indicator for International Anti-amnesty regime:

Before the rise of the ICC and UJ in 1998, the preexisting human rights treaties served as the main anti-impunity regimes by obligating states to prosecute offenders of serious crimes. However, the pre-98 system lacked an executive institution, and several ad hoc/mixed international tribunals were at best temporal and local. There was no mechanism to monitor or accuse when states let

the offenders go free (Peskin (2005)). On the other hand, in 1998, the ICC and UJ began to override domestic amnesties for international crimes. For this reason, year-1998 is widely perceived as the critical juncture for the transition from the era of impunity to the era of accountability (Dancy (2018), Krmaric (2018), Daniels (2020)), and I use the year-1998 as a tipping point for the emergence of anti-amnesty regimes.

Measure for Constituent Cost: Electoral democracy index

I measure constituent cost using “electoral democracy index” from the V-Dem democracy indices. State leaders are more likely to fear constituent costs under democratic regimes than authoritarian regimes because democratic constituents have more channels to punish leaders. More specifically, elections serve as a central device that holds politicians accountable to their constituents (Barro 1973). For this reason, constituent cost is interpreted in terms of electoral accountability (Dai (2005)). This variable (`v2xpolyarchy`) has a scale from 0 (low) to 1 (high). As the score gets closer to 1, states face higher electoral accountability, and in this study, higher constituent costs.

3.3 Making Comparison Using Propensity Score Matching

(6) USE DATA TO MAKE THE CASE THAT YOUR RESEARCH DESIGN ALLOWS YOU TO INTERPRET OBSERVED QUANTITIES AS THEORETICALLY RELEVANT AND CLEAR:

(6-1) EXPLAIN HOW YOU WILL MAKE THE CASE FOR INTERPRETABLE COMPARISONS (2 PARAGRAPHS PLUS SOME TABLES OR FIGURES)

I make an interpretable comparison using propensity score matching. My main comparison is SV amnesties in pre-98 vs. post-98, but my theory predicts that countries with high electoral democracy score should grant significantly less than authoritarian regimes after 1998. Therefore, there are two potential variables that I can consider as treatment: “Post-1998” and “democracy.” Post-1998 can be considered as treatment since it represents the advent of the anti-amnesty regimes. However, because it is a simple year indicator, it is empirically challenging to find reasonable confounders that would make the post-98 more likely. Instead, I make democracy as a treatment. Since propensity-score matching is to mimic the randomized experiment, if it were a randomized experiment, it is hypothetically randomizing countries and make some democratic and leave others as authoritarian and examining whether the treated and non-treated groups are different in terms of their provision of SV amnesties. Additionally, in order to compare the treatment effects in pre- and post-1998, I conduct the propensity score matching analysis for pre-98 observations and post-98 observations by splitting the data into two.

(*NOTE: I tried to use two-by-two factorial design to capture the interaction effect by making both `post98` and `democracies` as treatments. However, I was unable to fully conduct the `DeclareDesign` process so decided to split the dataset into pre- and post-98 and compare the treatment effects (`democracy`) in the two periods.)

To conduct this strategy, I categorize countries into two groups: democratic (countries with high constituent costs) vs. authoritarian (countries with low constituent cost) by changing the electoral democracy index as a binary variable. Based on the cut-off point used in a V-dem report, I consider countries with democratic index above 0.5 as democratic and 0.5 and below as authoritarian (Lührmann (2021)). Using the binary variable (`DEM`) as a treatment, I match observations based on their propensity to the treatment. Based on the democratization literature, democracy (the treatment in this study) is related to states’ judicial independence (`judicialinde`), number of years at war (`yearsatwar`), number of rebel groups (`numdyads`), and ethnic diversity (`ethnic`).

Table 1 shows that the covariates in the control (DEM=0) and treated group (DEM=1) are quite different for both pre- and post-98 data. This makes it difficult to make a good comparison because it is likely that treatment effects are biased by the covariates. This shows why propensity score matching can be useful in this study.

Table 1: Before Matching Treatment and Control Groups in Pre- and Post-98 Data

<PRE-98>	DEM=0	DEM=1	adj.diff	adj.diff.null.sd	std.diff	z	
judicialinde	0.21	0.65	0.44	0.04	2.51	11.80	
yearsatar	4.12	5.78	1.66	0.86	0.31	1.93	
numdyads	1.71	1.31	-0.40	0.17	-0.38	-2.36	*
ethnic	0.06	0.16	0.09	0.04	0.36	2.24	*
sv	0.27	0.47	0.20	0.07	0.43	2.68	**
<POST-98>	DEM=0	DEM=1	adj.diff	adj.diff.null.sd	std.diff	z	
judicialinde	0.20	0.70	0.50	0.06	3.25	9.07	***
yearsatar	3.72	4.75	1.03	1.03	0.21	1.00	
ethnic	0.09	0.00	-0.09	0.05	-0.35	-1.64	
numdyads	1.76	1.72	-0.04	0.20	-0.04	-0.19	
sv	0.48	0.57	0.09	0.11	0.18	0.85	

(6-2) EXPLAIN HOW YOU WILL JUDGE THE SUCCESS OF YOUR ADJUSTMENT STRATEGY (1 PARAGRAPH)

I will judge the success of the adjustment by looking at how well the balance of covariate distributions in the treatment and control groups are done after matching. As discussed above, propensity matching, unlike multiple linear regression, can show whether adjustment is done enough. Table 2 compares the treatment subjects and control subjects after matching in Pre-98 and Post-98 observations. After the adjustment, the treatment and control groups are more similar on all covariates for both pre- and post-98 data. For this reason, I consider the adjustments as successful. Figure 2 and 3 show the `xBalance` results showing the balance of covariate distributions after matching (Ben B. Hansen (2008)) for Pre-98 and Post-98 data respectively. They show that every covariate became closer to 0 in standardized differences after matching, and the balancing was done more successfully in the Pre-98 dataset. Table 3 and 4 show the structures of matched sets for Pre- and Post-98 dataset respectively. There are 31.2 and 10.3 equivalent number of matched pairs.

Table 2: After Matching Treatment and Control Groups in Pre- and Post-98 Data

<PRE-98>	DEM=0	DEM=1	adj.diff	adj.diff.null.sd	std.diff	z
judicialinde	0.52	0.53	0.01	0.02	0.04	0.33
yearsatar	5.42	4.72	-0.70	1.26	-0.13	-0.56
numdyads	1.26	1.21	-0.05	0.13	-0.04	-0.35
ethnic	0.16	0.13	-0.03	0.09	-0.11	-0.32
sv	0.36	0.36	0.00	0.12	0.01	0.02
<POST-98>	DEM=0	DEM=1	adj.diff	adj.diff.null.sd	std.diff	z
judicialinde	0.44	0.50	0.06	0.04	0.37	1.41
yearsatar	6.44	4.65	-1.80	1.99	-0.37	-0.90
ethnic	0.02	0.00	-0.02	0.06	-0.07	-0.33
numdyads	1.47	1.46	-0.01	0.41	-0.01	-0.03
sv	0.58	0.50	-0.08	0.18	-0.16	-0.44

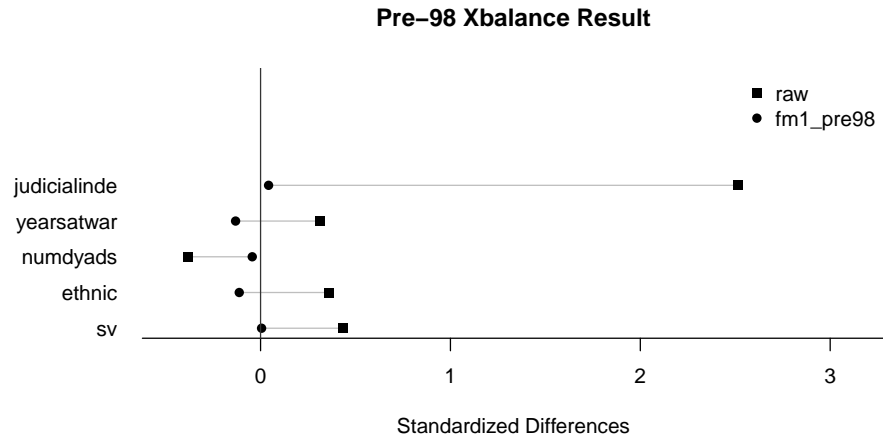


Figure 2: Balance Test for Pre-98

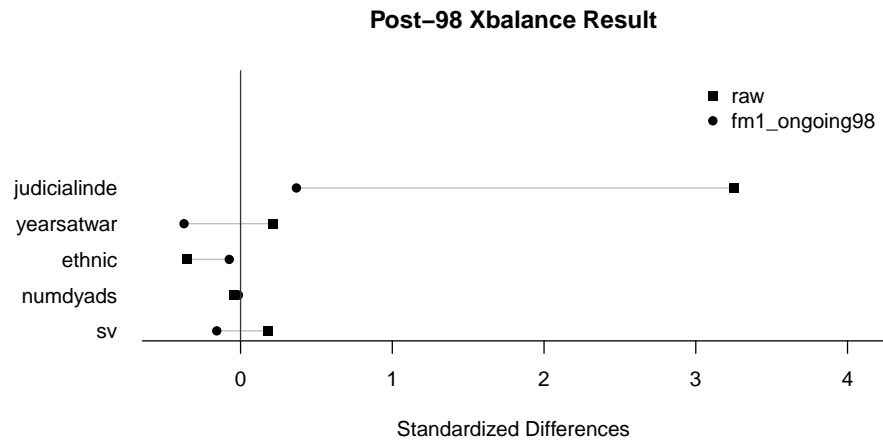


Figure 3: Balance Test for Post-98

Table 3: Structure of Matched Sets for Pre-98

	x
11:1	1
6:1	1
4:1	1
3:1	1
2:1	3
1:1	6
1:2	3
1:3	1
1:4	1
1:5	1
1:10	1
1:37	1
1:200	1

Table 4: Structure of Matched Sets for Post-98

	x
19:1	1
3:1	1
2:1	1
1:1	2
1:4	1
1:91	1

7. EXPLAIN YOUR PLANS FOR ANY MISSING DATA OR EXTREME OUTCOME OR COVARIATE VALUES YOU MAY ENCOUNTER WHEN YOU GET THE REAL DATA (OR PERHAPS YOU HAVE THE BACKGROUND DATA BUT NOT THE REAL OUTCOMES, SO YOU CAN EXPLAIN YOUR PLANS FOR SUCH DATA ISSUES IN THAT CASE HERE TOO). (1 OR 2 PARAGRAPHS)

Theories behind propensity score analysis assume that the covariates are fully observed (PAUL R. ROSENBAUM (1983)). However, in practice, missingness in the covariates is sometimes inevitably. There are two options to deal with the missingness — 1) imputation such as filling the mean values or zero or 2) omitting the observations. In my case, the missing data is mainly caused because the merged datasets cover different time range. Hence, assuming the NAs as 0 or the mean value would be inappropriate. As long as missingness does not depend both on the outcome variable and treatment variable, this bias is generally small. Since there is no theoretical base to believe that the missingness in this study is related to any of these, I am planning to ignore the missing data.

Regarding the extreme outcome, researchers need to be mindful about the bias-variance trade-offs (Lee BK (2010)). Trimming the extreme values changes the population, and it can introduce bias depending on the cut-off (Cole (2008)). For this study, I will not remove the extreme values unless I find some theoretical justification for such decision.

3.4 Performance of the tests

8. EXPLAIN HOW YOU WILL JUDGE THE PERFORMANCE OF THOSE TESTS. WILL YOU ONLY USE SIMPLE FALSE POSITIVE RATE AND POWER? OR DO YOU NEED TO ADD FAMILY-WISE ERROR RATE? FALSE DISCOVERY RATE? OR SOMETHING ELSE? EXPLAIN WHY YOU MADE THIS CHOICE. (1 PARAGRAPH)

Family-wise error rate (FWER) is the probability of making one or more false discoveries, or type 1 errors (i.e., incorrectly rejecting the null hypothesis when the null hypothesis is true). This is inflated usually when performing multiple hypotheses tests. In this case, p-value has to be adjusted using Bonferroni correction or adjusting false discovery rate. However, this study does not involve any multiple testing. Also, I collapse all the observations into country-event, so there is little concern with overfitting issue. Hence, I will judge the performance of tests by looking at the false positive rate and power.

9. SHOW AND EXPLAIN HOW YOUR TEST PERFORMS IN REGARDS THOSE PROPERTIES (AT LEAST YOU WILL SHOW FALSE POSITIVE RATE AND POWER). (2–4 PARAGRAPHS)

The power of a test is denoted by “ $(1-\beta)$,” and it is the probability of a true positive or the “probability of avoiding a false negative.” It ranges from 0 to 1, and as the power increases, the probability of making type II error (false negative) decreases. A false positive rate is the probability of a type I error.

Table 5 shows powers and false positive rates using two different estimators for pre- and post-98 datasets. Here, I show the `lm_robust` and `lm` estimator not propensity score matching OLS, since I can easily compare the model using `DeclareDesign`. Both are based on 0.05 thresholds for p-value (i.e., $\text{power} = \text{mean}(\text{p.value} < 0.05)$; the same for false positive rates) and are based on 500 times of simulations.

The powers of the estimators are both very low—all four power values are less than 0.1. It indicates that these models are likely to cause a high false negative. The false positive rates are also low—all less than 0.1. It indicates that the two models are less likely to cause type I error. The low power could be caused due to frequent zero outcome variables or due to the change of outcome variable as a binary outcome.

Table 5: Power and False Positive Rates

	estimator_label	power	se(power)	false_positive_rate
1	<code>lm_robust</code>	0.08	0.01	0.08
2	<code>lm</code>	0.04	0.01	0.04
2	<code>lm_robust</code>	0.00	0.00	0.00
2	<code>lm</code>	0.06	0.01	0.06

3.5 Estimators

10. WHAT STATISTICAL ESTIMATORS DO YOU PLAN TO USE? EXPLAIN WHY YOU CHOSE THESE ESTIMATORS. ESPECIALLY EXPLAIN WHAT IS YOUR TARGET OF ESTIMATION — WHAT IS THE ESTIMAND? (1 PARAGRAPH)

Based on the information presented in Table 5 and 6, and considering the benefit of keeping the treatment (`democracy index`) as continuous variable, I will use `lm_robust` as the statistical

estimators. As seen in Table 1, this observational study suffers bias from background confounders. In order to tackle this issue, adjustment for covariate is very critical. While propensity score matching would be a great way to overcome the problem, the nature of outcome variable in this study makes it difficult to use the matching strategy. `lm_robust` has efficiency in large samples and low bias in small samples as well as similarities to design-based randomization estimators (Samii and Aronow 2012).

The question of interest in this study is whether being democratic makes the country grant less SV amnesties compared to authoritarian regimes. Also, the study predicts such pattern more or only conspicuous after 1998. In this sense, the target of estimation (estimand) is the difference in differences of the treatment effect or the interaction effect between Post-98 and democracy variables. In this pre-analysis plan, however, my estimand is the average treatment effect (**democracy**).

11. EXPLAIN HOW YOU WILL JUDGE THE PERFORMANCE OF THOSE ESTIMATORS (ESPECIALLY BIAS AND MSE)? (1 PARAGRAPH)

The performance of estimators can be evaluated by looking at characteristics like the bias, consistency, coverage, and MSE. The RMSE (Root Mean Squared Error) is the standard deviation of the residuals that measures how well the data values fit the line of best fit. Unbiased estimator means that the estimator or test statistic is accurate to approximate the parameter. Lastly, coverage rates refer to the coverage probability of the confidence intervals. The coverage probability shows how often we obtain a confidence interval that contains the true population parameter if we were to repeat the entire sampling and analysis process. I will judge the performance of estimators using bootstrapping (i.e., resampling with replacement).

12. SHOW AND EXPLAIN HOW YOUR ESTIMATOR PERFORMS IN REGARDS THOSE PROPERTIES (AT LEAST BIAS AND MSE). (2–4 PARAGRAPHS)

For convenience, when feasible, I use the `diagnose_design` function in `DeclareDesign` to generate a table showing the characteristics discussed in 11. Table 6 shows the diagnose for estimators using two different estimators and each with 500 simulations. Biases are generally low—0.01 or below. Since bias is close to zero, it indicates that the estimators for the treatment coefficients are unbiased. Also, the RMSEs are below 0.2 under both estimators. It means that the data values do not deviate much from the fitted line.

Table 6: Diagnosing Estimators

	estimator_label	bias	rmse	power	coverage
1	Pre-LM	-0.07	0.07	0.00	1.00
2	Pre-GLM	-0.28	0.28	0.00	1.00
3	Post-LM	0.16	0.16	0.00	1.00
4	Post-GLM	0.99	0.99	0.00	1.00

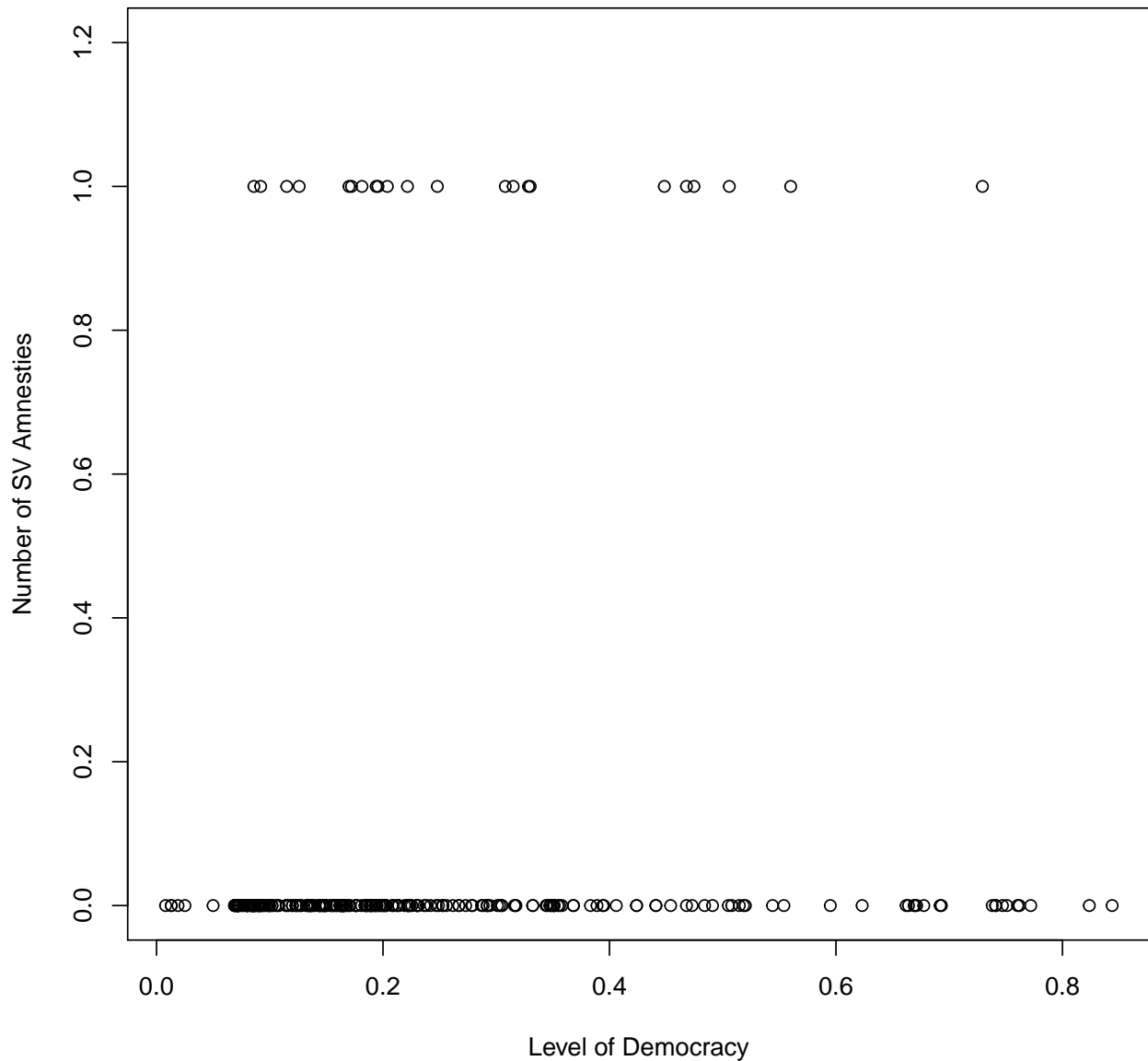
Table 7: Diagnosing Estimators

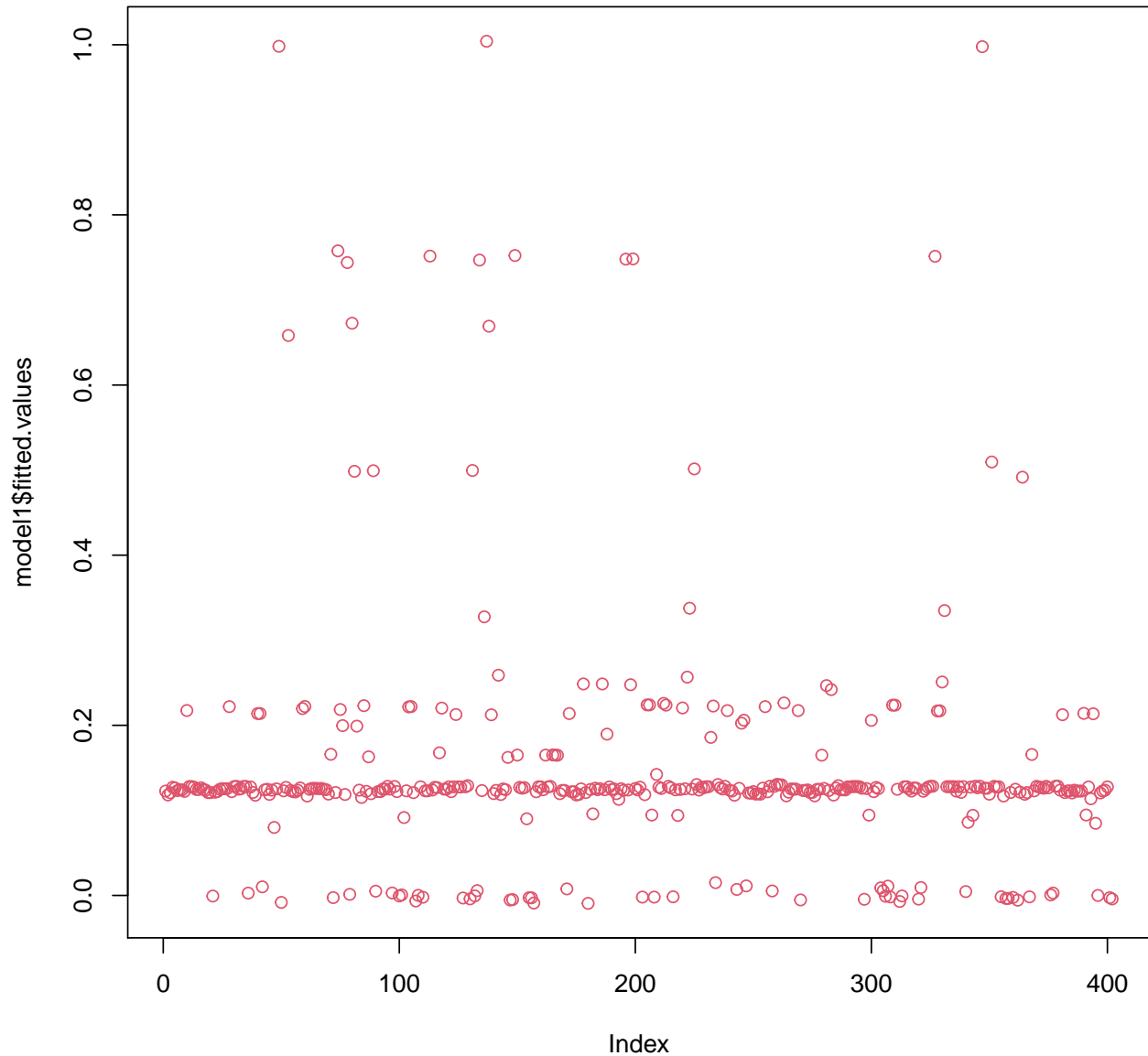
	estimator_label	se(bias)	rmse	power	se(power)
1	<code>lm_robust</code>	0.01	0.11	0.08	0.01
2	<code>lm</code>	0.01	0.18	0.02	0.01
1	<code>lm_robust</code>	0.00	0.06	0.00	0.00
2	<code>lm</code>	0.01	0.19	0.04	0.01

13. MAKE ONE MOCK FIGURE OR TABLE OF THE KIND YOU PLAN TO MAKE WHEN YOU USE THE ACTUAL OUTCOME.

Using fake data, I show a mock figure which shows the real data values in black, and the fitted values of the `lm_robust` model in red. I use propensity score matching and fixed effects to see the number of SV amnesties by democracy score. This figure shows that the actual data points are below the predicted line is above – which means that democratic countries grant more sv amnesties. This is the opposite direction of what the theory is predicting (the figure here does not show the interaction effect). Hence, if the actual study exhibit similar figure, it would mean that the theory has weak statistical evidence.

M4





14. INCLUDE A CODE APPENDIX AND A LINK TO THE GITHUB REPOSITORY FOR THIS PAPER.

All data and RMD file are uploaded in the following github repository: <https://github.com/mjkim12/Preanalysis>

15. INCLUDE REFERENCES WITH APPROPRIATE IN-TEXT CITATIONS.

4 The Appendix

Here is the code appendix.

```
library (readr)
urlfile="https://raw.githubusercontent.com/mjkim12/Preanalysis/main/amnesty_mjk_0402.csv"

mydata<-read_csv(url(urlfile))
```

```

library(tidyverse)
library(car)
library(optmatch)
library(RIttools)
library(pscl)
library(DeclareDesign)
library(mosaic)
library(estimatr)
library(tidyverse)
library(xtable)
library(fabricatr)
library(randomizr)

#Simple pie chart (Composition of conflict)
slices <- c(294, 136, 66)
lbls <- c("pre98war", "post98war", "cross98war")
pct <- round(slices/sum(slices)*100)
lbls <- paste(lbls, pct) # add percents to labels
lbls <- paste(lbls,"%",sep="") # ad % to labels
pie(slices, labels = lbls, col= topo.colors(length(lbls)))
#Make some variables as binary
mydata$dummyoutcome <- ifelse(mydata$sum_hram>0,1,0) #making SVAmnesty into dummy
mydata$sv[which(is.na(mydata$sv))] =0
mydata$DEM <- ifelse(mydata$mean_v2xpolyarchy>0.5, 1, 0) #making dem score into dummy

wrdat <- mydata %>%dplyr:: select(dummyoutcome, sum_hram, DEM, judicialinde, yearsatwar,ethnic)

#Removing Missing Data
sum(is.na(wrdat)) #262
dim(wrdat) #dim: 514,11
wrdat <- na.omit(wrdat)
dim(wrdat) #dimension: 402, 11

#Categorizing conflicts by years of start and end yrs
df_post98 <- wrdat[which (wrdat$post98war==1), ] #dim:79, 11
df_pre98 <- wrdat[which (wrdat$post98war==0), ] #dim: 323, 11
df_ongoing98 <- wrdat[which (wrdat$warend_post98==1), ] #dim:128,11
# Create linear predictors for pre-98 data
df_pre98_2 <- df_pre98

glm_ps_pre98 <-
  glm(DEM ~ judicialinde +yearsatwar + numdyads + ethnic + sv, data = df_pre98_2, family = binomial)

pscore_pre98 <- predict(glm_ps_pre98, type = "response")

# Make distance matrices
psdist_pre98 <-match_on(DEM~ pscore_pre98, data=df_pre98_2)

```

```

# Fullmatching
fm1_pre98 <- fullmatch(psdist_pre98, data = df_pre98_2)
fm1_pre98_summary <- summary(fm1_pre98, data = df_pre98_2, min.controls = 0, max.controls = Inf)

#xtable(fm1_pre98_summary$matched.set.structures, caption = "Structure of Matched Sets for pre-98")

df_pre98_2$fm1_pre98 <- NULL
df_pre98_2[names(fm1_pre98), "fm1_pre98"] <- fm1_pre98
head(df_pre98_2)
#Run xBalance to assess the balance properties of the match
xb1_pre98 <- xBalance(DEM ~ judicialinde + yearsatwar+ numdyads+
                     ethnic + sv,
                     strata = list(raw=NULL, fm1_pre98 = ~fm1_pre98),
                     data = df_pre98_2,
                     report = "all")

plot(xb1_pre98, main="Pre-98 Xbalance Result")

#xtable(xb1_pre98, caption = "Balanced Treatment and Control Groups")

#I also tried the matching based on ranked mahalanobis distance, but propensity score produced

##### Ongoing98 #####
# Create linear predictors for pre-98 data
df_ongoing98_2<-df_ongoing98

glm_ps_ongoing98 <-
  glm(DEM ~ judicialinde +yearsatwar + numdyads + ethnic + sv, data = df_ongoing98_2, family = "binomial")

pscore_ongoing98 <- predict(glm_ps_ongoing98, type = "response")

# Make distance matrices
psdist_ongoing98 <-match_on(DEM~ pscore_ongoing98,data=df_ongoing98_2)

# Fullmatching
fm1_ongoing98 <- fullmatch(psdist_ongoing98, data = df_ongoing98_2)

fm1_ongoing98_summary <- summary(fm1_ongoing98, data = df_ongoing98_2, min.controls = 0, max.controls = Inf)
#There are 10.3 effective sample size

#xtable(fm1_ongoing98_summary$matched.set.structures, caption = "Structure of Matched Sets for ongoing98")

df_ongoing98_2$fm1_ongoing98 <- NULL
df_ongoing98_2[names(fm1_ongoing98), "fm1_ongoing98"] <- fm1_ongoing98
head(df_ongoing98_2)
#Run xBalance to assess the balance properties of the match
xb1_ongoing98 <- xBalance(DEM ~ judicialinde + yearsatwar+ ethnic + numdyads + sv,

```

```

    strata = list(raw=NULL, fm1_ongoing98 = ~fm1_ongoing98),
    data = df_ongoing98_2,
    report = "all")

plot(xb1_ongoing98, main="Post-98 Xbalance Result")

#xtable(xb1_ongoing98, caption = "Balanced Treatment and Control Groups 2")
##### DeclareDesign PRE-98 #####

#Matched dataset

pop_pre98_2 <- declare_population(df_pre98_2) #fm1_pre98

pot.outcome_pre <- declare_potential_outcomes(Y ~ 0.02*Z + sum_hram)

estimand_pre <- declare_inquiry(ATE = mean(Y_Z_1 - Y_Z_0))

assignment_pre <- declare_assignment(m=50, label="Z")

reveal_pre <- declare_reveal(Y,Z)

samp_pre <- declare_sampling(n=100)

estimator_pre1 <- declare_estimator(Y~ DEM,
    model = lm_robust,
    term = 'DEM',
    inquiry = 'ATE',
    label = "lm_robust")

estimator_pre2 <- declare_estimator(Y~ DEM + judicialinde + yearsatwar+ ethnic + numdyads + sv
    model = lm,
    term = 'DEM',
    inquiry = 'ATE',
    label = "lm")

design_pre1 <- pop_pre98_2 +pot.outcome_pre+ estimand_pre +assignment_pre + reveal_pre + samp_
design_pre2 <- pop_pre98_2 +pot.outcome_pre+ estimand_pre +assignment_pre + reveal_pre + samp_

diagnose1 <- diagnose_design(design_pre1 ,design_pre2)

#xtable(head(diagnose1$diagnosands_df[,c(3,5,6, 7,8, 9,10)]))

##### DeclareDesign POST 98 #####

```

```

#Declare Population
pop_ongoing98_2 <- declare_population(df_ongoing98_2) #fm1_ongoing98

pot.outcome_post <- declare_potential_outcomes(Y ~ -0.01*Z + sum_hram)

estimand_post <- declare_inquiry(ATE = mean(Y_Z_1 - Y_Z_0))

assignment_post <- declare_assignment(m=50, label="Z")

reveal_post <- declare_reveal(Y,Z)

samp_post <- declare_sampling(n=100)

estimator_post1 <- declare_estimator(Y~ DEM,
                                     model = lm_robust,
                                     term = 'DEM',
                                     inquiry = 'ATE',
                                     label = "lm_robust")

estimator_post2 <- declare_estimator(Y~ DEM + judicialinde + yearsatwar+ ethnic + numdyads + sv,
                                     model = lm,
                                     term = 'DEM',
                                     inquiry = 'ATE',
                                     label = "lm")

design_post1 <- pop_ongoing98_2 +pot.outcome_post+ estimand_post +assignment_post + reveal_post

design_post2 <-pop_ongoing98_2 +pot.outcome_post+ estimand_post +assignment_post + reveal_post

diagnose1_post <- diagnose_design(design_post1 ,design_post2)
#xtable(head(diagnose1_post$diagnosands_df[,c(3,5,6, 7,8, 9,10)]))
design_pre <-diagnose_design(design_pre1, design_pre2, diagnosands = declare_diagnosands(power
design_post <-diagnose_design(design_post1, design_post2, diagnosands = declare_diagnosands(po

#xtable(head(design_pre$diagnosands_df[,c(3,5,6,7)]))
#xtable(head(design_post$diagnosands_df[,c(3,5,6,7)]))
library(pscl)
#ALL Propensity Score#
wrdat_zero <- wrdat
glm_ps <-
  glm(DEM ~ judicialinde + ethnic + sv, data = wrdat_zero, family = binomial())

```



```

pscore <- predict(glm_ps, type = "response")
psdist <- match_on(DEM~ pscore, data=wrdat_zero)
fm_all <- fullmatch(psdist, data = wrdat_zero)
fm_summary <- summary(fm_all, data = wrdat_zero, min.controls = 0, max.controls = Inf)

#ps into the dataset
wrdat_zero$fm_all <- NULL
wrdat_zero[names(fm_all), "fm_all"] <- fm_all
head(wrdat_zero)
diagnose1 <- diagnose_design(design_pre1 , design_pre2)
diagnose1_post <- diagnose_design(design_post1 , design_post2)

xtable(head(diagnose1$diagnosands_df[, c(3,6,7,9,10)]))
xtable(head(diagnose1_post$diagnosands_df[, c(3,6,7,9,10)]))
fakedata <- draw_data(design_pre1)
fakedata$sum_hram

#ALL Propensity Score#
glm_ps <-
  glm(DEM ~ judicialinde + yearsatwar + numdyads + ethnic + sv, data = wrdat, family = binomial)

pscore <- predict(glm_ps, type = "response")
psdist <- match_on(DEM~ pscore, data=wrdat)
fm_all <- fullmatch(psdist, data = wrdat)
fm_summary <- summary(fm_all, data = wrdat, min.controls = 0, max.controls = Inf)

#lm_robust
model1 <- lm_robust(sum_hram ~ mean_v2xpolyarchy, fixed_effects = ~fm_all, data = wrdat)

plot(wrdat$mean_v2xpolyarchy[wrdat$warend_post98==0],
     wrdat$sum_hram[wrdat$warend_post98==0],
     xlab="Level of Democracy",
     ylab = "Number of SV Amnesties", ylim = c(0,1.2),
     main = "M4")
abline(plot(model1$fitted.values, col = 2))

df_pre98_2 <- df_pre98

initialate <- mean(df_pre98_2$sum_hram[df_pre98_2$DEM == 1]) - mean(df_pre98_2$sum_hram[df_pre98_2$DEM == 0])

library(arm)
balfmla <- lm(DEM ~ judicialinde + yearsatwar + numdyads + ethnic + sv, data = df_pre98_2)

## Bayes Generalized Linear Model
glm2 <- bayesglm(balfmla, data = df_pre98_2, family = binomial)
#boxplot(glm2,
#main = "", names = c("Control", "Treatment"),

```

```

#xlab = ""
#)
## create linear predictors
pscores2 <- predict(glm2)

## make distance matrices
psdist_2 <- match_on(DEM ~ pscores2, data = df_pre98_2)
#as.matrix(psdist)[1:5,1:5]

## Rank-Based Mahalanobis distance
mhdist <- match_on(DEM ~ judicialinde + yearsatwar + numdyads + ethnic + sv, data = df_pre98_2,
fm22 <- fullmatch(mhdist, data = df_pre98_2)
fm22sum <- summary(fm22, data = df_pre98_2, min.controls = 0, max.controls = Inf)

## Add matched set indicators back to data
df_pre98_2$fm22 <- NULL
df_pre98_2[names(fm22), "fm22"] <- fm22

## Balance test to see if I "adjusted enough"
xb22 <- xBalance(DEM ~ judicialinde + yearsatwar + numdyads + ethnic + sv,
strata = list(raw = NULL, fm22 = ~fm22),
data = df_pre98_2,
report = c(
"std.diffs", "z.scores", "adj.means",
"adj.mean.diffs", "chisquare.test", "p.values"
)
)
#plot(xb22)

## What is the biggest difference within set.
diffswithinsets<-df_pre98_2 %>% group_by(fm22) %>% summarize(meandiff = mean(sum_hram[DEM==1])
summary(diffswithinsets$meandiff)

#DIFF PRE MATCHING
with(df_pre98_2, mean(sum_hram[DEM==1]) - mean(sum_hram[DEM==0]))
## What are the distances like?
tmp2 = df_pre98_2$sum_hram
names(tmp2) <- rownames(df_pre98_2)
absdist2 <- match_on(tmp2, z = df_pre98_2$DEM)
qtl2 <- quantile(as.vector(absdist2),seq(0,1,.1))

#declare population
pop_pre98_2 <- declare_population(df_pre98_2)

## declare estimator
estimator22 <- declare_estimator(sum_hram~DEM+judicialinde + yearsatwar + numdyads + ethnic +
                                model = lm,

```

```

        term = 'DEM',
        estimand = 'DEM',
        label = "OLS")

#declare estimand
make_estimands22 <- function(data){
  bs <- coef(lm(sum_hram~DEM, data=df_pre98_2))
  return(data.frame(estimand_label= 'DEM',
                    estimand=bs['DEM'],
                    stringsAsFactors = FALSE))
}
estimand22 <- declare_estimands(handler=make_estimands22,
                               label="Pop_Relationships")

design1_plus_estimands22 <- pop_pre98_2 + estimand22
#kable(estimand1(df_pre98_2), caption = "Estimands 22")

design_full22 <- design1_plus_estimands22 +
  estimator22

#run_design(design_full22)
df2_ongoing98 <-df_ongoing98

initialate <- mean(df2_ongoing98$sum_hram[df2_ongoing98$DEM == 1]) - mean(df2_ongoing98$sum_hram[df2_ongoing98$DEM == 0])

library(arm)
balfmla_33 <- lm(DEM ~ judicialinde +yearsatwar + numdyads + ethnic + sv, data = df2_ongoing98)

## Bayes Generalized Linear Model
glm3 <- bayesglm(balfmla_33, data = df2_ongoing98, family = binomial)
#boaxplot(glm3,main = "", names = c("Control", "Treatment"), xlab = "")
## create linear predictors
pscores3 <- predict(glm3)

## make distance matrices
psdist_3 <- match_on(DEM ~ pscores3, data = df2_ongoing98)
#as.matrix(psdist)[1:5,1:5]

## Rank-Based Mahalanobis distance
mhdist_3 <- match_on(DEM ~ judicialinde +yearsatwar + numdyads + ethnic + sv, data = df2_ongoing98)
fm33 <- fullmatch(mhdist_3, data = df2_ongoing98)
fm33sum <- summary(fm33, data = df2_ongoing98, min.controls = 0, max.controls = Inf)

## Add matched set indicators back to data
df2_ongoing98$fm33 <- NULL
df2_ongoing98[names(fm33), "fm33"] <- fm33

```

```

## Balance test to see if I "adjusted enough"
xb33 <- xBalance(DEM ~ judicialinde + yearsatwar + numdyads + ethnic + sv,
strata = list(raw = NULL, fm33 = ~fm33),
data = df2_ongoing98,
report = c(
"std.diffs", "z.scores", "adj.means",
"adj.mean.diffs", "chisquare.test", "p.values"
)
)
#plot(xb33)

## What is the biggest difference within set.
diffswithinsets<-df_pre98_2 %>% group_by(fm22) %>% summarize(meandiff = mean(sum_hram[DEM==1])
summary(diffswithinsets$meandiff)

#DIFF PRE MATCHING
with(df_pre98_2, mean(sum_hram[DEM==1]) - mean(sum_hram[DEM==0]))
## What are the distances like?
tmp2 = df_pre98_2$sum_hram
names(tmp2) <- rownames(df_pre98_2)
absdist2 <- match_on(tmp2, z = df_pre98_2$DEM)
qtl2 <- quantile(as.vector(absdist2),seq(0,1,.1))

## declare estimator
estimator22 <- declare_estimator(sum_hram~DEM+judicialinde + yearsatwar + numdyads + ethnic + s
                                model = lm,
                                term = 'DEM',
                                estimand = 'DEM',
                                label = "OLS")

#declare estimand
make_estimands22 <- function(data){
  bs <- coef(lm(sum_hram~DEM, data=df_pre98_2))
  return(data.frame(estimand_label= 'DEM',
                    estimand=bs['DEM'],
                    stringsAsFactors = FALSE))
}
estimand22 <- declare_estimands(handler=make_estimands22,
                              label="Pop_Relationships")
pop_pre98_2 <- declare_population(df_pre98_2)

design1_plus_estimands22 <- pop_pre98_2 + estimand22
#kable(estimand1(df_pre98_2), caption = "Estimands 22")

design_full22 <- design1_plus_estimands22 +
  estimator22

```

```

#run_design(design_full22)
# Below is the attempt I made to perform a Two-by-Two Factorial design using DeclareDesign (For

# A: post98, B: democracy
#population <- wrdat
#potential_outcomes <- declare_potential_outcomes(
# Y_A_0_B_0 = sum_hram[warend_post98==0 & DEM==0],
# Y_A_0_B_1 = sum_hram[warend_post98==0 & DEM==1],
# Y_A_1_B_0 = sum_hram[warend_post98==1 & DEM==0],
# Y_A_1_B_1 = sum_hram[warend_post98==1 & DEM==1])

#weight_A <- 0
#weight_B <- 0
#prob_A <- 0.5
#prob_B <- 0.5
#estimand_1 <- declare_inquiry(
# ate_A = weight_B * mean(Y_A_1_B_1 - Y_A_0_B_1)
# + (1-weight_B)*mean(Y_A_1_B_0 - Y_A_0_B_0))

#estimand_2 <- declare_inquiry(
# ate_B = weight_A * mean(Y_A_1_B_1 - Y_A_1_B_0)
# + (1-weight_A) * mean(Y_A_0_B_1 - Y_A_0_B_0))

#estimand_3 <- declare_inquiry(interaction = #mean((Y_A_1_B_1 -
# Y_A_1_B_0) - (Y_A_0_B_1 - Y_A_0_B_0)))

#stuck_here:
#assign_A <- declare_assignment(complete_ra(wrdat, prob = prob_A))

#assign_A <- declare_assignment(A = complete_ra(wrdat, prob = prob_A))

#assign_B <- declare_assignment(B = block_ra(prob = prob_B, blocks = A))

#reveal_Y <- declare_reveal(Y_variables = sum_hram, assignment_variables = c(A, B))

#estimator_1 <- declare_estimator(Y ~ A + B, model = lm_robust,
#term = c("A", "B"), inquiry = c("ate_A", "ate_B"), label = "No_Interaction")
#estimator_2 <- declare_estimator(Y ~ A + B + A:B, model = lm_robust, term = "A:B", inquiry =

#Didn't run here:
#two_by_two <- population + potential_outcomes + estimand_1 + estimand_2 + estimand_3 + assign

#ALL Propensity Score#
glm_ps <-
  glm(DEM ~ judicialinde + yearsatwar + numdyads + ethnic + sv, data = wrdat, family = binomial)

pscore <- predict(glm_ps, type = "response")

```

```

psdist <- match_on(DEM~ pscore, data=wrdat)
fm_all <- fullmatch(psdist, data = wrdat)
fm_summary <- summary(fm_all, data = wrdat, min.controls = 0, max.controls = Inf)

#1. Declare Population
# pop <- declare_design(wrdat)

#2. Declare Truth: For the pre-analysis, I should not look at the outcome variable, so I am fo
#truth = declare_inquiry(truth = 0)

#3. Declare Estimators
#est1 = declare_estimator (y~ mean_v2xpolyarchy*post98war, term = "mean_v2xpolyarchy:post98war"
#est1b = declare_estimator (y~ mean_v2xpolyarchy*warend_post98, term = "mean_v2xpolyarchy:ware
#est2 = declare_estimator(y~mean_v2xpolyarchy*post98war, term = "mean_v2xpolyarchy:post98war",
#est2b = declare_estimator(y~mean_v2xpolyarchy*warend_post98, term = "mean_v2xpolyarchy:warend
#est3 = declare_estimator(y~mean_v2xpolyarchy*post98war, term = "mean_v2xpolyarchy:post98war",

# Declare estimator using a custom handler for zero-inflated count model
#est3 <- declare_estimator(y ~ mean_v2xpolyarchy *post98war |sv, term = "mean_v2xpolyarchy:pos

#How to bring zeroinflated into declare design?
#estimator_function(y ~ mean_v2xpolyarchy *post98war |sv, term = "mean_v2xpolyarchy:post98war"

#I can put control variables above: est1 = declare_estimator (y~ mean_v2xpolyarchy*post98war +

#4. Declare Design
#design = pop + truth + est1 + est1b + est2 + est2b
#forcing estimand to be 0

#5. Run Design Once
#run_design(design)

#6. Run 500 simulations of design
#simulate_design (design)

#7. diagnose_design(design) # check performance
#diagnosands <- declare_diagnosands(bias = mean(est - estimand),
#      power = mean(p <= .05),
#      mean_est = mean(est))
#diagnose_design(design, diagnosands = NULL,
#      # sims = 1000, bootstrap_sims = 1)
##Test the validity of the analysis?

```

```

###conditional_treatment_effect
#lm(sum_hram~DEM*warend_post98,data=wrdat)

#y_z_0 = wrdat$sum_hram
#y_z_1 = wrdat$sum_hram -0.05

#pop_CATE <- declare_population(wrdat)

#potentialoutcome_CATE <- declare_potential_outcomes(
  # y_z_0 = y_z_0,
  # y_z_1 = y_z_0 - 0.05)

wrdat$warend_post98
#estimand_CATE <- declare_inquiry(CATE_diff =
  #   mean(y_z_1[warend_post98 ==1]
  #   - y_z_0[warend_post98==1]) -
  #   mean(y_z_1[warend_post98 ==0]
  #   - y_z_0[warend_post98==0]))

#make_estimands_cate <- function(data) {
#bs_cate <- coef(lm(sum_hram ~ DEM, data = wrdat))
#return(data.frame(
#estimand_label = c("exposure"),
#estimand = bs_cate[c("exposure")],
#stringsAsFactors = FALSE))}

#Declare Estimand
#estimand_CATE <- declare_inquiry (handler = make_estimands_cate, label = "Pop_relationships")

#kable(estimand_CATE(wrdat))

#design1 <- pop_98 + estimand1

#estimator1_yh <- declare_estimator(sum_hram ~DEM*warend_post98, fixed.effects = ~fm1_pre98,
  #coeffecient_name="DEM:warend_post98",

#sampling <- declare_sampling(n=250)
#assignment <- declare_assignment(m=125)

#design_cate <- pop_CATE + potentialoutcome_CATE +estimand_CATE+ sampling+assignment

#diagnose_design(design_cate)

#difwithinsets <- df_ongoing98%>%
#group_by(fm1_ongoing98)%>%

```

```

# summarize(meandiff = mean(sum_hram[DEM==1]) - mean(sum_hram[DEM==0]))
#Zero Inflated polyarchy - PRE98

#model2<-formula(sum_hram~mean_v2xpolyarchy+warend_post98 | warend_post98*mean_v2xpolyarchy)
#modelpre98_2 <- zeroinfl(model2, dist="poisson", data=df_pre98_2)
#summary(modelpre98_2)

#df_pre98_2$predict <- predict(modelpre98_2)

#stat2 <- function(df, inds) {
# model <- formula(sum_hram~mean_v2xpolyarchy+warend_post98 | #warend_post98*mean_v2xpolyarchy)
# predict(
#   zeroinfl(model2, dist = "poisson", data = df[inds, ]),
#   newdata = df)
#}
#set.seed(2018)
#library(boot)
#res2 <- boot(df_pre98_2, stat, R = 100)

#CI2 <- setNames(as.data.frame(t(sapply(1:nrow(df_pre98_2), function(row)
# boot.ci(res2, conf = 0.95, type = "basic", index = row)$basic[, 4:5]))),
# c("lower", "upper"))

#df_all_pre98poly <- cbind.data.frame(df_pre98_2, response = predict(modelpre98_2), CI2)

# Drawing option 1 -- ggplot2 plot with confidence intervals
#ggplot(df_all_pre98poly, aes(mean_v2xpolyarchy, sum_hram)) +
# geom_point() +
#   geom_ribbon(aes(ymin = lower, ymax = upper))
# Drawing option 2 -- Install plotrix package
# install.packages("plotrix")
#library("plotrix")
#plotCI(x = df_all_pre98poly$mean_v2xpolyarchy,
#       y = df_all_pre98poly$sum_hram,
#       li = df_all_pre98poly$lower,
#       ui = df_all_pre98poly$upper)

```

References

- Alan S. Gerber, Donald P. Green, and Edward H. Kaplan. 2004. "The Illusion of Learning from Observational Research, Problems and Methods in the Study of Politics." *Cambridge University Press Cambridge University Press*.
- Arane Thavaneswaran, Lisa Lix. 2008. "Propensity Score Matching in Observational Studies." *Lecture Note*.

- Ben B. Hansen, Jake Bowers. 2008. "Covariate Balance in Simple, Stratified and Clustered Comparative Studies." *Statist.*
- Cole, Hernán, S. R. 2008. "Constructing Inverse Probability Weights for Marginal Structural Models." *American Journal of Epidemiology.*
- Cunningham, David. 2016. "Preventing Civil War: How the Potential for International Intervention Can Deter Conflict Onset." *World Politics.*
- Dai, Xinyuan. 2005. "Why Comply? The Domestic Constituency Mechanism." *International Organization.*
- Dancy, Geoff. 2018. "Deals with the Devil? Conflict Amnesties, Civil War, and Sustainable Peace." *International Organization* 72 (2): 387–421. <https://doi.org/10.1017/S0020818318000012>.
- Daniels, Lesley-Ann. 2020. "How and When Amnesty During Conflict Affects Conflict Termination." *Journal of Conflict Resolution* 64 (9): 1612–37. <https://doi.org/10.1177/0022002720909884>.
- Fearon, James D. 1994. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review.*
- Ginsburg, Tom. 2009. "The Clash of Commitments at the International Criminal Court." *Chicago Journal of International Law.*
- Goldsmith, Jack, and Stephen D. Krasner. 2003. "The Limits of Idealism." *Daedalus.*
- Krcmaric, Daniel. 2018. "Should I Stay or Should I Go? Leaders, Exile, and the Dilemmas of International Justice." *American Journal of Political Science* 62 (2): 486–98. <https://doi.org/https://doi.org/10.1111/ajps.12352>.
- Lee BK, Stuart EA., Lessler J. 2010. "Improving Propensity Score Weighting Using Machine Learning." *Stat Med.*
- Lührmann, Marcus Tannenber, Anna. 2021. "Regimes of the World (RoW): Opening New Avenues for the Comparative Study of Political Regimes (Article)." *Politics and Governance [Online].*
- Mallinder, Louise. 2012. "Amnesties' Challenge to the Global Accountability Norm?" In *Amnesty in the Age of Human Rights Accountability*, edited by Francesca Lessa and Leigh A. Payne, 69–96. Cambridge: Cambridge University Press.
- McGillivray, Fiona, and Alastair Smith. 2000. "Trust and Cooperation Through Agent Specific Punishments." *International Organization.*
- Observational Studies, Design of. 2010. "Design of Observational Studies." *Springer Series in Statistics.*
- PAUL R. ROSENBAUM, DONALD B. RUBIN. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* Volume 70 Issue 1: Pages 41–55.
- Peskin, Victor. 2005. "Beyond Victor's Justice? The Challenge of Prosecuting the Winners at the International Criminal Tribunals for the Former Yugoslavia and Rwanda." *Journal of Human Rights* 4.2.
- Prorok, Alyssa K. 2017. "The (in)compatibility of Peace and Justice? The International Criminal Court and Civil Conflict Termination." *International Organization.*

Rosenbaum, Paul R. 2017. "Observation and Experiment: An Introduction to Causal Inference." *Harvard University Press*.

Rubin, Donald B. 2001. "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation." *Health Services and Outcomes Research Methodology* 2.