

# PS531 Final Project

## Pre-Analysis: Ratification Delays: Exploring Factors in International Environmental Agreements

Ji Soo Yoo

### 1 Introduction

#### 1.1 Research Question

To mitigate the global climate crisis, international organizations and nation-states have set goals such as reducing greenhouse gas emissions and promoting sustainable development through various international agreements. Since the late 1800s, states have created about 700 multilateral agreements and more than 1,000 bilateral agreements related to environmental protection (Mitchell, 2003). Regarding ratifying multilateral and bilateral environment agreements, some nation-states take less time. On the other hand, other states take longer to ratify, or other states never ratify international agreements. This creates a situation where states might agree to reduce emissions but fail to make actual reductions.

Since more international environmental agreements are expected, it is important to understand what influences nation-states' behavior in ratifying agreements (Jacobson & Weiss, 1990). To understand which factors affect nation-states' ratification of international environmental agreements, academia has studied the characteristics of international agreements, the role of domestic institutions, domestic audience costs, economic conditions, and scientific information and technology. Among these factors, this paper will focus on the role of domestic political institutions. George Tsebelis (2002) argues that the factor of political institutions in determining ratification and compliance with international agreements can be understood in the context of veto players theory.

Nation-states' constitution specifies whose consensus is required to change domestic policies from the status quo. In other words, consensus of veto players is required to ratify international agreements. Tsebelis emphasizes the role of political institutions since political power distribution among veto players such as presidents, prime ministers, and legislators set by their political institutions determines the ratification and compliance process of international environment agreements. In addition, Lauren Peritz (2020) states that more veto players imply more actors to block ratifying international agreements. Based on the theoretical framework of the veto-players theory for understanding the role of political institutions, this paper explores *whether types of domestic political institutions affect the duration of ratifying the Paris Agreement, an international environmental agreement.*

## 2 Theory

IR scholars have explored how domestic political institutions affect international cooperation (Neumayer, 2002; Dai, 2006; Kroll & Shogren, 2008; Bättig & Bernauer, 2009; Paquin, 2010). For example, Neumayer (2002) suggests that democracies sign and ratify more multilateral environmental agreements, participate in more environmental intergovernmental organizations, and comply better with the Convention on International Trade in Endangered Species of Fauna and Flora. Similarly, Bättig and Bernauer (2009) examine whether democracy affects levels of political commitment to climate change mitigation and emission reduction. The authors found that the effect of democracy on policy output (levels of political commitment to climate change) is positive. However, it is still ambiguous regarding policy outcomes—emission reduction. Kroll and Shogren (2008) examine how the domestic political system, presidential or parliamentary, affects international environmental agreements in bargaining one’s contribution to reducing emissions. The authors indicate that countries with the presidential system contribute less than the benchmark, while countries with the parliamentary system contribute higher than the benchmark. Domestic political institutions shape countries’ commitment to climate mitigation actions. Based on this theory, the project examines the differences in ratification of international agreements between countries with parliamentary systems and those with non-parliamentary systems (presidential and semi-presidential). The following hypothesis is proposed:

**Hypothesis:** Countries with a parliamentary system are inclined to take less time to ratify international environmental agreements

## 3 Research Design

The main comparison of this study is domestic political institution type such as presidential, semi-presidential, and parliamentary system. To test the hypothesis, I employ a quasi-randomized comparison using propensity score matching with observational data. Research design and identification strategies are discussed below.

### 3.1 Variables and Measures

I analyzed cross-sectional data from various countries, categorizing their political institutions based on the classification in “Principles of Comparative Politics”. Countries with a parliamentary system were coded as ‘1’, while others were coded as ‘0’. The key dependent variable in this study is the duration (in months) taken by countries to ratify the Paris Agreement, calculated by subtracting the signature date from the ratification date. Additionally, I incorporated data on the 2016 gross domestic product (GDP) and carbon dioxide (CO<sub>2</sub>) emissions of member states from the Paris Agreement, sourced from the World Bank Open Dataset. The year 2016 was chosen as it marks the opening of the Paris Agreement for signatures. GDP is used as an indicator of a country’s economic progress and power, reflecting the value of all final goods and services produced within a specific time frame. CO<sub>2</sub> emissions, measured from the burning of fossil fuels like coal, natural gas, and oil, serve as an environmental impact indicator.

## 4 Identification Strategy

Researchers should have a comparison group to make a causal inference. A comparison group allows them to understand an effect of treatments by analyzing differences between treated groups and control groups.

In randomized experiments, researchers have controls and can randomly assign participants to treated and untreated groups. Random assignments balances on observed and unobserved confounding factors (Green & Gerber, 2012) and allows researchers to claim that there is no influence of confounders on treatment effects, an interest of the study.

However, in observational studies, researchers do not have control in random assignments of treatments to participants (Rosenbaum, 2010). Observational studies encounter bias problems since confounding factors between two groups are not balanced. This implies that researchers cannot claim that differences in outcomes between treated and control groups are due to treatment effects. Therefore, observational studies have a disadvantage of making causal inferences due to potential confounding and biased estimates, while they are quick, cheap, and simple to organize (Mariani & Pêgo-Fernandes, 2014).

### 4.1 Propensity Score (PS) Matching

One way to address the limitation of observational studies is to use a propensity score matching. A propensity score matching allows researchers to design as-if randomized study. Researchers estimate the probability of receiving treatment based on observed covariates (Rosenbaum, 2010), also known as propensity score. Based on the similar propensity score, researchers can pair/match subjects in the treatment group to subjects in the controlled group, which replicates randomized experiment (Rubin & Thomas, 2000). In other words, propensity score matching allows researchers to design quasi-randomized experiments.

However, problems of potential confounding and biased estimates can still rise even after propensity score matching due to the main three reasons. First and most critically, propensity score matching does not address bias arising from unobserved covariates. Second, true propensity score is unknown because we do not have information on unobserved covariates (Imai, 2005). Third, the propensity score depends on model specification (i.e., choice of observed covariates) (Rosenbaum, 2010). Despite these limitations, propensity score matching help observational studies minimize bias in estimates of treatment effects.

#### 4.1.1 Treatment

A treatment for this project is binary indicator of domestic political institution. The control group is countries with non-parliamentary types. The treatment group is countries with parliamentary type. The hypothesis poses that countries with a parliamentary system will take less time to ratify international political agreements because the prime minister is typically a member of the political party that compose most of the legislative branch. This project uses classification of countries by system of government in the book called “Principles of Comparative Politics” by Clark et al. (2017). The dummy variable TRT is coded 1 if

countries are based a parliamentary system and 0 if they are not based a parliamentary system.

#### 4.1.2 Propensity Score Matching Model Specification and Matching Method

To estimate propensity scores, this project uses a package called “matchit” in R. As the default setting, “matchit” uses logistic regression. Logistic regression is commonly used to estimate propensity score when treatment is a binary outcome.

Based on propensity scores, there are various ways of matching treated and controlled units such as nearest neighbor matching, full matching, and Mahalanobis distance. First, nearest neighbor matching is to pair each treated subject with each controlled subject within the smallest distances (Stuart, 2010).

Second, Full matching creates matched sets, which each set has one treated subject and one or more control subjects. Subjects that cannot be placed into any matched set are discarded. Lastly, the Mahalanobis distance technique measures the distance between covariates of the treated and controlled unit (Amusa et al., 2019). Units are matched on the shortest distance. The matched ones are balanced across multiple covariates. I explore various matching methods to check which method provides better matched data sets.

The project explores these methods to find the most effective matching approach. Nearest neighbor matching involves 1:1 pairing, where each treated unit is matched with one control unit, focusing on the closest match based on a specified distance measure. In contrast, full matching creates sub-classes (1:k, where  $k \geq 1$ ), and then assigns weights based on subclass. The full matching specification can lead to a loss in precision compared to 1:1 matching, as units in full matching may contribute less to the sample than unweighted units in nearest neighbor matching.

I define the propensity score using covariates that may influence the ratification of international environmental agreements, as suggested in previous studies. Stein (2008) posits that concerns about economic development deter countries from ratifying international environmental agreements like the Kyoto Protocol, due to uncertainties about how these agreements might impact their economic growth. Moreover, countries heavily reliant on natural resources often show reluctance in ratifying international climate agreements. Consequently, I include covariates such as GDP per capita, and the percentages of GDP from Oil and Coal Rents.

#### 4.1.3 Missing Data

Propensity score matching assumes the full observation of covariates within the data (Rosenbaum & Rubin, 1983, 984). Considering the common challenge encountering missing covariate values in research, there are two strategies handling this issue: mean imputation and omission of observations. The mean imputation strategy operates under the assumption that the mechanism behind the missing data is ignorable, replacing missing values with means derived from responding units with the sample (Mattei et al., 2011). Conversely, omitting observation is alternative way when the missing covariates are not contingent upon the outcome and treatment variable. In the absence of sufficient evidence of the association, this

project opts to disregard the missing data

#### 4.1.4 Balance of Covariates

**4.1.4.1 Judge the Success of the Adjustment Strategy** To judge success of each matching method, I conduct balance tests (Rosenbaum, 2010). Balance tests produce a standardized differences in means test between the control and treatment group on covariates. It is essential to conduct a balance test because a similar distribution of covariates between the treated and control groups indicates the successful application of the chosen propensity score (Hansen & Bowers, 2008). Plotting the results of balance tests help visualize differences between raw and adjusted data. Figure 1 and Table 1 show that covariates are not balanced before matching. The chi-square and p-value are 29.42 and 0.00 respectively. Small p-value means that covariates are not balanced. In other words, there are differences between the treatment and control group.

```
##      strata      raw
##      stat      trt=0    trt=1  adj.diff  std.diff      z
## vars
## oil          1.01e+00  1.51e-01  -8.64e-01 -4.06e-01 -1.77e+00 .
## gdpp          4.75e+03  2.27e+04  1.80e+04  1.41e+00  5.26e+00 ***
## Coal          1.76e-01  5.27e-02  -1.23e-01 -2.35e-01 -1.04e+00
## ---Overall Test---
##      chisquare df  p.value
## raw      29.4   3  1.83e-06
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

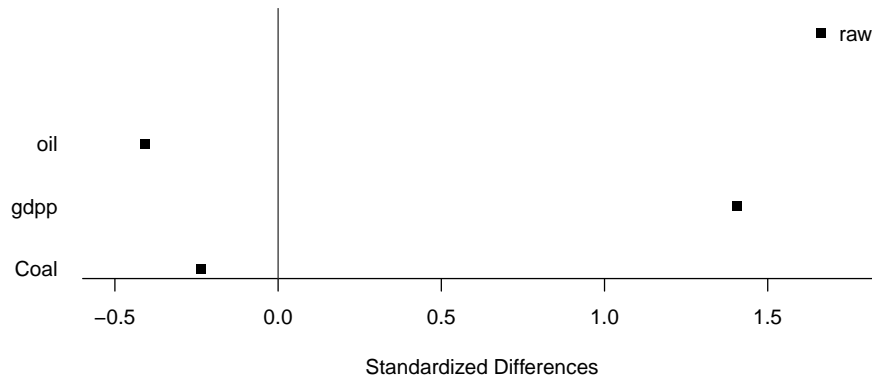


Figure 1: Balance Test Before Matching

## NULL

	chisquare	df	p.value
raw	29.41752	3	1.8e-06

To minimize the differences between the treatment and control groups, this project explores three matching methods: full matching, Mahalanobis distance matching, and nearest-neighbor matching. Figure 2 and Table 2 show the results of full matching. After applying full matching, the chi-square value is 29.42, and the p-value is 0.00. Notably, these values are the same as those observed prior to matching, suggesting that the full matching method may not have been effective in this context.

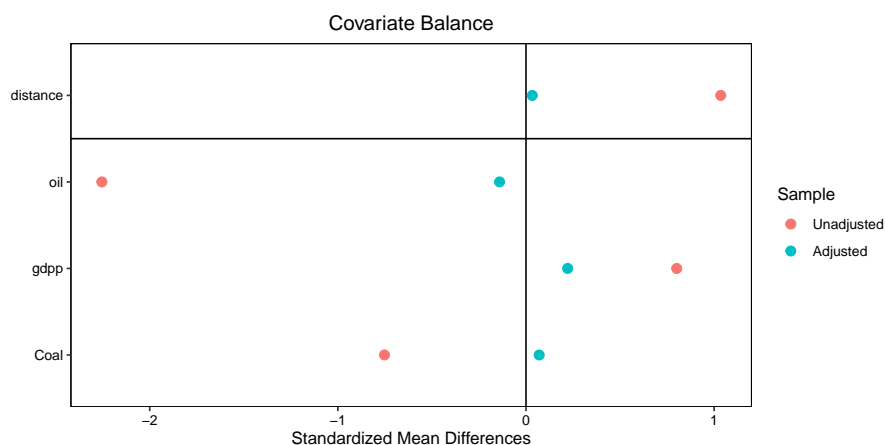


Figure 2: Balance Test for Full Matching

	chisquare	df	p.value
raw	29.41752	3	1.8e-06

Figure 3 and Table 3 show the results of full matching. After applying full matching with the Mahalanobis distance, the chi-square value is 29.42, and the p-value is 0.00. Notably, these values are the same as those observed prior to matching, suggesting that the full matching with the Mahalanobis distance method may not have been effective in this context

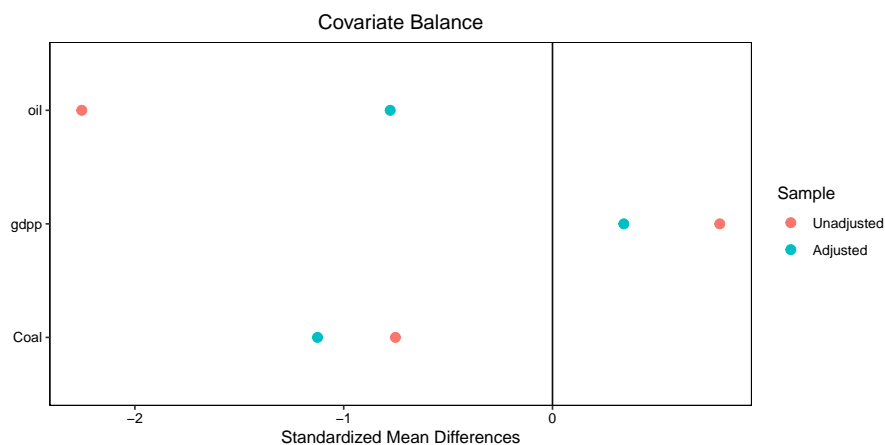


Figure 3: Balance Test Full Matching with Mahalanobis distance

	chisquare	df	p.value
raw	29.41752	3	1.8e-06

The balance test results for nearest neighbor matching are displayed in Figure 4 and Table 4. The chi-square value has notably decreased from 29.42 to 1.2, and the p-value has significantly increased from 0.00 to 0.75. According to Table 4, nearest neighbor matching results in the highest p-value when compared to the other two methods. In the context of balance tests, a higher p-value suggests a better balance among covariates. Consequently, these results indicate that nearest neighbor matching has successfully minimized differences between the treatment and control groups.

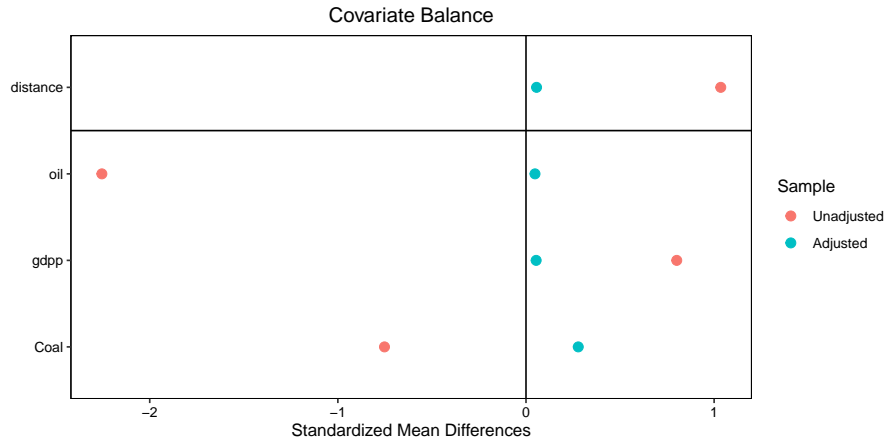


Figure 4: Balance Test for Nearest Neighbor Matching

	chisquare	df	p.value
raw	1.218604	3	0.7485458

## 5 Simulated Data

For the pre-analysis of this study, I create simulated population based on the original dataset and randomly select 500 observations. The random sampling allows researchers to take design-based approaches and to avoid make distributional assumptions.

## 6 Estimand and Estimator

The objective of this research is to test whether some nation-states takes longer or less time to ratify the Paris Agreement, depending on their type of political institution. In other words, this test is to check whether there is an association between political institution type and duration of ratifying the the Paris Agreement. Hence, the target of estimation is  $\beta_1$  (a difference in duration of ratifying the Paris Agreement (**time**), depending on political

institution type of nation-states (**political inst**). I construct the estimand using linear regression model through the DeclareDesign. The estimand by the political institution type is -2.213.

$$Y_{time} = \beta_0 + \beta_1 X_{politicalinst} + \epsilon$$

I will show the values of the estimand for two estimators – linear regression, and the propensity score matching.

Table 5: Estimand for Political Institution Types

estimand_label		estimand
trt	lm	-2.740572

## 6.1 Linear Regression

In this model, the outcome is expressed in the same unit as the dependent variable. In the context of this research, the unit of outcome is duration, measured in months. To use linear regression, certain assumptions must be met. The first assumption is that the error terms (residuals) are normally distributed and independent. In this study, the error terms are generated using “rnorm(N, mean = 0, sd = 1)” in defining potential outcomes through DeclareDesign to ensure their normal distribution and independence. The second assumption is homoscedasticity, which means the residuals should have constant variance across the levels of independent variables. A significant fluctuation in the variance of residuals across these levels can lead to biased estimates. The third assumption is the absence of multicollinearity, indicating that the independent variables should not be highly correlated with each other. To confirm these assumptions, I conduct homoscedasticity and multicollinearity tests. After evaluating the remaining two assumptions, I will use a robust linear regression model if one of the assumptions does not hold.

### 6.1.1 Influential Values

From the linear regression analysis, I initially identify outliers using Cook’s distance. Recognizing that not all outliers necessarily exert a significant influence on the model, I further investigate the presence of potentially influential observations by examining the standardized residual errors. Data points with an absolute value of the standardized residual exceeding a certain threshold (commonly 2 or 3) are considered for their potential impact on the regression results. Based on the threshold (standardized error >3), I filtered thirty potential influential outliers from the data.



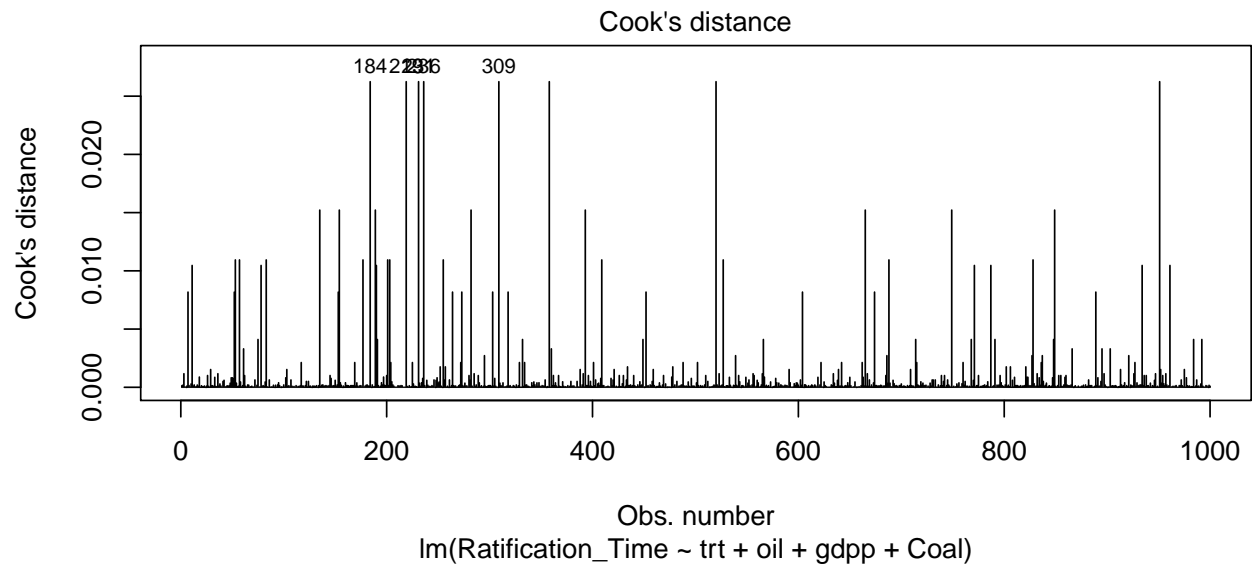
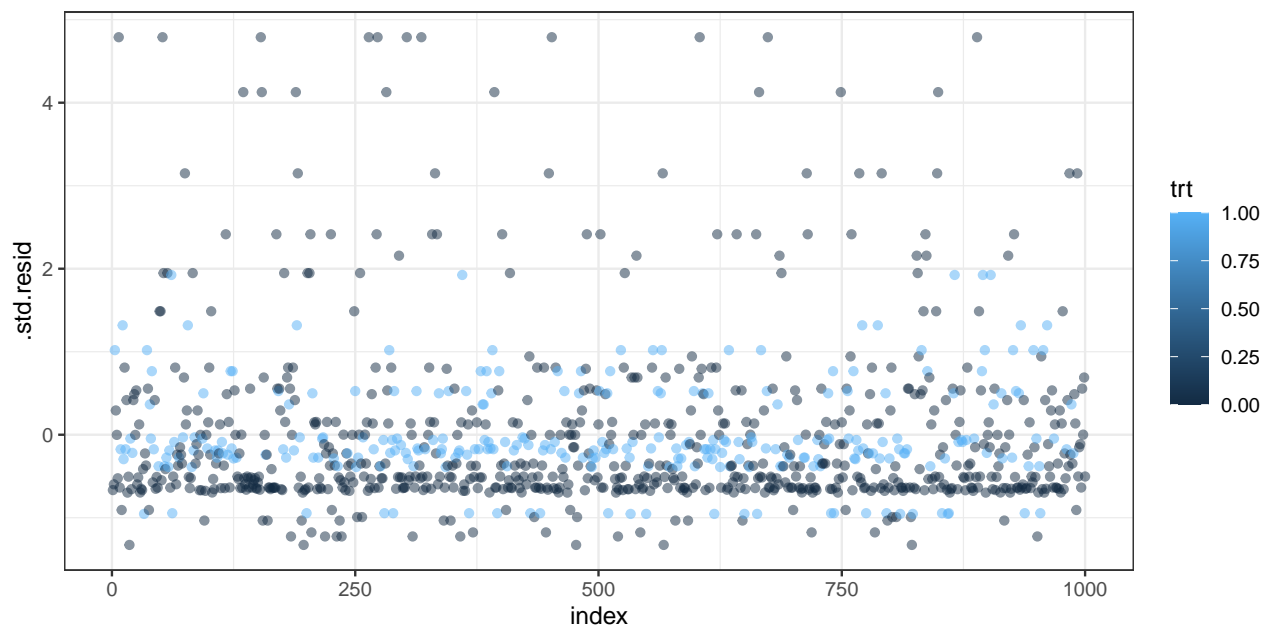


Figure 5: Cook's Distance



#### Homoscedasticity

The results of homoscedasticity test suggest that the variance of the error terms varies across different values of the independent variables. It is also known as the heteroscedasticity problem. Given that the assumption of homoscedasticity is not met, I will use a robust linear regression estimator. This approach adjusts for the heteroscedasticity, providing less biased estimates.

Table 6: Breusch-Pagan Test Results for Homoscedasticity Test

Test	BP_Statistic	Degrees_of_Freedom	P_Value
Breusch-Pagan Test	22.636	4	0.0001497

### 6.1.2 Multicollinearity

This project assesses multicollinearity using the `vif()` function from the `car` package in R. Variance Inflation Factors (VIF) allow researchers to check whether variables are highly correlated, in order to discern their individual effects on the dependent variable. VIF values are indicative of the level of correlation among variables. Generally, a VIF value of 1 suggests no correlation. The table indicates an absence of multicollinearity among these variables, as their VIF values are close to 1.

Table 7: Assessing Multicollinearity

	x
trt	1.458726
oil	1.036222
gdpp	1.422413
Coal	1.013147

### 6.1.3 Diagnosis of the linear regression model

To assess the performance of the estimator, I check values of Bias, RMSE, and Power. The value of bias indicates the difference between the mean estimate and mean estimand. A value of bias close to zero indicates that the estimate is unbiased. The value of bias (=0.35) can be considered as low but not ideal. It suggests that the prediction is off by 0.35 units on average. Values of RMSE (Root Mean Square Error) implies accuracy of a statistical model; a lower value of RMSE is better. The value of RMSE (0.61) indicates that the statistical model has the prediction error by 0.61. A value of power implies the probability of correctly rejecting the null hypothesis when it is false—this checks a Type II error. In the context of my research, a Type II error would occur if I conclude that there is no difference in ratification duration across different government systems, when in fact, there is a difference. A higher value of power is better, as it decreases the likelihood of making a Type II error. The power value of 1 implies a certainty in correctly rejecting a false null hypothesis.

```
##      Design Inquiry Estimator Term N Sims Mean Estimand Mean Estimate   Bias
## 1 design_lm      trt lm_robust  trt   500      -2.74      -2.39    0.35
## 2                                     (0.00)      (0.02) (0.02)
##      SD Estimate   RMSE   Power Coverage
## 1          0.50    0.61    1.00    0.89
## 2      (0.02) (0.02) (0.00)  (0.02)
```

Design	Inquiry	Estimator	N Term Sims	Mean Estimand	Mean Estimate	Bias	RMSE	Power	Coverage
design_lm	trt	lm_robust	500	-2.74 (0.00)	-2.39 (0.02)	0.35 (0.02)	0.61 (0.02)	1.00 (0.00)	0.89 (0.02)

Type I Error involves incorrectly rejecting the null hypothesis when it is actually true. In the context of my research, a Type I error would occur if I conclude that there is a difference in the ratification duration across different government systems, when, in fact, no such difference exists. The value of type 1 error (0.056) is above the significant level (0.05). The value implies that it is to be careful in interpreting the estimates.

```
##          design sim_ID term estimator      estimate std.error statistic
## 1 design_lm_null      1  trt lm_robust -0.028423629 0.08314301 -0.3418643
## 2 design_lm_null      2  trt lm_robust -0.002956422 0.09500376 -0.0311190
## 3 design_lm_null      3  trt lm_robust  0.019222155 0.08434895  0.2278885
## 4 design_lm_null      4  trt lm_robust  0.020245634 0.08491486  0.2384227
## 5 design_lm_null      5  trt lm_robust  0.050666961 0.09067219  0.5587927
## 6 design_lm_null      6  trt lm_robust -0.191363030 0.08712405 -2.1964434
##      p.value   conf.low   conf.high  df      outcome inquiry
## 1 0.73252518 -0.1915794  0.13473214 995 Ratification_Time    trt
## 2 0.97518088 -0.1893871  0.18347429 995 Ratification_Time    trt
## 3 0.81977976 -0.1463001  0.18474441 995 Ratification_Time    trt
## 4 0.81160232 -0.1463871  0.18687840 995 Ratification_Time    trt
## 5 0.57642891 -0.1272637  0.22859763 995 Ratification_Time    trt
## 6 0.02829031 -0.3623310 -0.02039507 995 Ratification_Time    trt
```

Table 9: Type I Error Rate

Type_I_Error_Rate
0.056

## 6.2 Linear Regression Estimator (after matching)

Similar to the approach taken with the actual dataset, I explored various matching methods for the simulated data. These methods included nearest neighbor matching and nearest neighbor matching using propensity score matching. For each set of matched simulated data, I constructed an estimand and a robust linear regression estimator. Subsequently, the results from these estimators were compared to evaluate their performance and consistency across different matching techniques

### 6.2.1 Nearest Neighbor Matching

The table shows that covariates across the control and treatment groups remain unbalanced even after performing nearest neighbor matching. This study proceeded to construct an

estimand and a robust linear regression estimator using the nearest neighbor matched data. Comparing these results to those in Table 1, this study observed an increase in bias by 0.09 and a decrease in coverage by 0.05. The value of power (=1) is high; this suggests that the model is unlikely to make a Type II error. The value of type 1 error (0.056) is above the significant level (0.05). The value implies that it is to be careful in interpreting the estimates.

Table 10: Estimand 2

	estimand_label	estimand
trt	nearest_estimand	-2.8375

##	Design	Inquiry	Estimator	Term	N	Sims	Mean	Estimand
## 1	design_lm_ps	trt	lm_nearest_matching	trt	500			-2.84
## 2								(0.00)
##	Mean	Estimate	Bias	SD	Estimate	RMSE	Power	Coverage
## 1		-2.39	0.44		0.50	0.67	1.00	0.84
## 2		(0.02)	(0.02)		(0.02)	(0.02)	(0.00)	(0.02)

Table 11: Type I Error Rate

Design	Inquiry	Estimator	N	Term	Sims	Mean Esti- mand	Mean Esti- mate	Bias	RMSE	Power	Coverage
design_lm_trps	lm_nearest_matching	trt	500			-2.84 (0.00)	-2.39 (0.02)	0.44 (0.02)	0.67 (0.02)	1.00 (0.00)	0.84 (0.02)

Type_I_Error_Rate
0.056

### 6.2.2 Nearest Neighbor Matching Using Propensity Score

For another matching method, I calculated propensity score and performed nearest-neighbor matching based on propensity score. The percentage balance improvements for each variable shows that differences between the treated and control group are reduced. The diagnosis of this statistical model has the same value of the statistical model used nearest neighbor matching without using propensity scores. The model is unlikely to make type II error (Power = 1), whereas there is likely to make type I error (=0.056). The values imply that it is careful to interpret effects of treatment on the outcome variable.

Covariate	Std..Mean.Diff.	Var..Ratio	eCDF.Mean	eCDF.Max
distance	96.1	90.4	96.0	81.6
oil	95.8	74.1	50.7	4.1

Covariate	Std..Mean.Diff.	Var..Ratio	eCDF.Mean	eCDF.Max
gdpp	86.9	60.0	89.3	78.3
Coal	67.3	81.4	-80.1	-150.0

##	Design	Inquiry	Estimator	Term	N	Sims	Mean	Estimand
## 1	design_lm_ps_nr	trt	lm_ps_matching	trt	500			-2.84
## 2								(0.00)
##	Mean	Estimate	Bias	SD	Estimate	RMSE	Power	Coverage
## 1	-2.39	0.44	0.50	0.67	1.00	0.84		
## 2	(0.02)	(0.02)	(0.02)	(0.02)	(0.00)	(0.02)		

Table 14: Type I Error Rate

Design	Inquiry	Estimator	Term	N	Mean Esti- mand	Mean Esti- mate	Bias	RMSE	Power	Coverage
design_lm_ps_nr	trt	lm_ps_matching	trt	500	-2.84 (0.00)	-2.39 (0.02)	0.44 (0.02)	0.67 (0.02)	1.00 (0.00)	0.84 (0.02)

Type_I_Error_Rate
0.056

## 7 Mock Result

I will show the results of statistical tests using the first estimator since the estimator has the least bias estimate among three estimators. The table shows that the effect of treatment is -2.7498. In the context of my research, it is understood that countries with parliamentary system is likely to ratify 2.7 months faster from their signature date compare to countries with other political systems (presidential or semi-presidential system). The diagnostics of estimators and the performance of statistical tests suggest the need for caution in interpreting the results. To address these concerns, I conducted a sensitivity analysis.

Table 16: Treatment from Estimator 1

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	9.7857	0.3258	30.0326	0.0000
trt	-2.7498	0.6468	-4.2515	0.0000
Coal	0.8840	0.5067	1.7447	0.0814
oil	0.1365	0.1004	1.3592	0.1744
gdpp	0.0000	0.0000	0.7080	0.4791

## 8 Sensitivity Analysis

Sensitivity analysis tests inform researchers whether variations in outcomes are due to the treatment or to unobservable covariates although propensity score matching addresses bias arising from observable covariates. A low R2 value suggests that it is likely that unobservable covariates significantly influence the variation in outcomes. In the context of this research, the quicker ratification of the Paris Agreement by countries with parliamentary systems, as compared to their signature date, may be attributed more to unobservable covariate-related reasons than to the type of political institution (where the treatment is having a parliamentary system).

Statistic	Value
Model Formula	Ratification_Time ~ trt + Coal + oil + gdpp
<i>Null Hypothesis</i>	
q	1
reduce	TRUE
<i>Unadjusted Estimates of 'trt'</i>	
Coefficient Estimate	-2.74984
Standard Error	0.64679
t-value	-4.25149
<i>Sensitivity Statistics</i>	
Partial R2 of treatment with outcome	0.01784
Robustness Value, q = 1	0.126
Robustness Value, q = 1 alpha = 0.05	0.06996

Table 17: Sensitivity Analysis to Unobserved Confounding

## 9 Replication Data

All data and codes (in. Rmd) can be found in the following github repository: <https://github.com/jisoosy2/ps531preanalysis>

## 10 Appendix

```
library(formatR)
library(knitr)
library(readxl)
library(readr)
library(dplyr)
library(tidyverse)
library(car)
library(optmatch)
library(Matching)
```

```

library(RItools)
library(pscl)
library(DeclareDesign)
library(mosaic)
library(estimatr)
library(tidyverse)
library(xtable)
library(fabricatr)
library(randomizr)
library(WeightIt)
library(cobalt)
library(arm)
library(stats)
library(MatchIt)
library(MASS)
# Load data This data contains countries, their signature
# and ratification date, the time that they took to ratify
# (in months), CO2 and GDP in 2016, types of domestic
# political institutions (presidential, semi-presidential,
# and parliamentary system)

df <- read_excel("Desktop/2020 Fall /Quant2_Final/FinalProject_Yoo.xlsx")

# Check NAs in this data
na.count2 <- sum(is.na(df$Signature))
na.count <- sum(is.na(df$Ratification))

# sum(is.na(df)) #120

# Omit NAs
df <- na.omit(df)

# These dataset are for oil rent (% of GDP), coal rent (%
# of GDP), and GDP per capita
df1 <- read_excel("Downloads/ps531_oil.xls")
df2 <- read_excel("Downloads/ps531_gdpp.xls")
df3 <- read_excel("Downloads/ps531_coal.xls")

# Merge the first two dataset
merge1 <- merge(df, df1, by = "Participant", all = FALSE)

# Merge the result with the third data
merge2 <- merge(merge1, df2, by = "Participant", all = FALSE)

```

```

# Merge the result with the fourth data
final_merge <- merge(merge2, df3, by = "Participant", all = FALSE)

# Assign treatment based on system. Countries with a
# parliamentary system is assigned as 1 and otherwise as 0.
final_merge <- final_merge %>%
  mutate(trt = if_else(system == "pm", 1, 0))

# Count treatment and control
final_merge %>%
  group_by(trt) %>%
  summarise(count = n())

# Omit NAs
final_merge <- na.omit(final_merge)

treated <- (final_merge$trt == 1)
cov <- final_merge[, c(11, 12, 13)]
std.diff <- apply(cov, 2, function(x) 100 * (mean(x[treated]) -
  mean(x[!treated]))/(sqrt(0.5 * (var(x[treated]) + var(x[!treated])))))

xb <- xBalance(trt ~ oil + gdpp + Coal, strata = list(raw = NULL),
  data = final_merge, report = c("std.diffs", "z.scores", "adj.means",
  "adj.mean.diffs", "chisquare.test", "p.values"))
xb #small p-value means that covariates are not balanced
plot(xb)
kable(xb$overall)

# Full matching using the ps
m.out1 <- matchit(trt ~ oil + gdpp + Coal, method = "full", data = final_merge)

# Create a matched dataset
matched_data1 <- match.data(m.out1)
love.plot(m.out1)

# Check balance for the matched data set
balance_check1 <- xBalance(trt ~ oil + gdpp + Coal, strata = list(raw = NULL),

```



```

    data = matched_data1, report = "all")
kable(balance_check1$overall)
# Full matching using the Mahalanobis distance
m.out2 <- matchit(trt ~ oil + gdpp + Coal, method = "full", distance = "mahalanobis",
    data = final_merge)
# Create the second matched dataset
matched_data2 <- match.data(m.out2)
love.plot(m.out2)

# Check balance for the second matched dataset
balance_check2 <- xBalance(trt ~ oil + gdpp + Coal, strata = list(raw = NULL),
    data = matched_data2, report = "all")

kable(balance_check2$overall)
# Nearest neighbor matching using propensity score
m.out3 <- matchit(trt ~ oil + gdpp + Coal, method = "nearest",
    caliper = 0.2, data = final_merge)

# Create the third matched dataset
matched_data3 <- match.data(m.out3)
love.plot(m.out3)

# Check balance for the third matched dataset
balance_check3 <- xBalance(trt ~ oil + gdpp + Coal, strata = list(raw = NULL),
    data = matched_data3, report = "all")
kable(balance_check3$overall)

# writing regression model dependent variable is only
# positive

# hist(df$Ratification_Time)
# summary(glm(Ratification_Time~system+CO2+GDP+pop+CRIScore,
# family = 'poisson', data = matched_data3))

# model =
# glm(Ratification_Time~system+CO2+GDP+pop+CRIScore, family
# = 'poisson', data = matched_data3) Anova(model)

# system is significant variable/siginificant factor in
# predicting ratification time or system influences
# influence time

# unshown category is the reference group. whether
# parliamentary system is different from presidential

```

```

# system whether semi-presidential system is different from
# presidential system

# matched_data3$system = as.factor(matched_data3$system)
# contrasts(matched_data3$system) = contr.treatment(3, base
# = 2) contrasts(matched_data3$system) model2 =
# glm(Ratification_Time~system+CO2+GDP+pop+CRIScore, family
# = 'poisson', data = matched_data3) summary(model2)
# mean_count exp(7.808e-01) #if you change from the
# reference group to system 1, the ratification time
# increase twice. 2.183218

# Create a simulated data
fake_population <- declare_model(N = 1000, data = final_merge,
  handler = resample_data)

N <- final_merge$Participant

# glm(Ratification_Time~trt+gdpp+oil+Coal, data =
# final_merge) Use coefficients from logistic regression to
# construct potential outcomes Add normal distribution
# error terms
outcomes <- declare_potential_outcomes(Ratification_Time ~ trt *
  -2.41 + 2.921e-05 * gdpp + 0.3186 * oil + 0.7143 * Coal +
  rnorm(N, mean = 0, sd = 1), assignment_variables = "trt")

assignment <- declare_assignment(assignment_variable = "trt")
treatment_outcomes <- declare_reveal(outcome_variables = "Ratification_Time",
  assignment_variables = "trt")

design <- fake_population + outcomes

set.seed(12345)
df_sim <- draw_data(design)

## Sampling 500 observations from the population
set.seed(12345)
N <- as.numeric(nrow(final_merge))
sampling_1 <- declare_sampling(S = draw_rs(N = N, n = 1000))

```

```

design2 <- fake_population + sampling_1
set.seed(12345)
df1_fake <- draw_data(design2)

df1_fake

# Declare an estimand
make_estimand1 <- function(data) {
  model <- lm_robust(Ratification_Time ~ trt, data = df1_fake)
  bs <- coef(model)
  return(data.frame(estimand_label = c("lm"), estimand = bs["trt"],
    stringsAsFactors = FALSE))
}

estimand1 <- declare_inquiry(handler = make_estimand1, label = "pop_relationship")
design1_and_estimand <- fake_population + sampling_1 + estimand1

# Generate the estimand
estimand1_result <- estimand1(df1_fake)
kable(estimand1_result, caption = "Estimand for Political Institution Types\\label{tab:}
lm_fake <- lm(Ratification_Time ~ trt + oil + gdpp + Coal, data = df1_fake)
plot(lm_fake, which = 4, id.n = 5)

# Calculate standardized residuals
standardized_resid <- rstandard(lm_fake)

# Identify outliers
potential_outliers <- which(abs(standardized_resid) > 3)

# Check the number of rows that are potential influential
# points
outlier_rows <- rownames(df1_fake)[potential_outliers]

# Extract model results: computes the standardized residuals
# and the Cook's distance

library(broom)

model.data <- augment(lm_fake) %>%

```

```

mutate(index = 1:n())
model.data %>%
  top_n(3, .cooksd)

# Plot the standardized residuals:
ggplot(model.data, aes(index, .std.resid)) + geom_point(aes(color = trt),
  alpha = 0.5) + theme_bw()

# Filter potential influential data points
filtered <- model.data %>%
  filter(abs(.std.resid) < 3)

# There are 30 influential data points '7' '52' '75' '135'
# '153' '154' '189' '191' '264' '273' '282' '303' '318'
# '332' '393' '449' '452' '566' '604' '665' '674' '714'
# '749' '768' '791' '848' '849' '889' '984' '992'

# xtable(filtered)
library(lmtest)
library(sandwich)

# Perform homoscedasticity test
hsd_test <- bptest(lm_fake)

# Create a data frame for kable
hsd_results <- data.frame(Test = "Breusch-Pagan Test", BP_Statistic = 22.636,
  Degrees_of_Freedom = 4, P_Value = 0.0001497)

kable(hsd_results, caption = "Breusch-Pagan Test Results for Homoscedasticity Test",
  align = "c")

collinearity <- car::vif(lm_fake)
kable(collinearity, caption = "Assessing Multicollenearity \\label{tab:VIF}")

# Declare estimator1
lm_robust_estimator1 <- declare_estimator(Ratification_Time ~
  trt + oil + gdpp + Coal, model = lm_robust, term = c("trt"),
  inquiry = c("trt"), label = "lm_robust")

design_lm <- design1_and_estimand + lm_robust_estimator1

```

```

set.seed(123345)

# Simulate
sim_full <- simulate_design(design_lm, sims = 500)

diag1_lm <- diagnose_design(sim_full)

estimator1perform <- reshape_diagnosis(diag1_lm, digits = 2,
  select = NULL, exclude = NULL)
estimator1perform
kable(estimator1perform[, c(1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12)])

# Redesign the potential outcomes under the null hypothesis
# since the original potential outcome has the coefficient
# of -2.41. This would lead to the higher value of type 1
# error rate.
null_outcomes <- declare_potential_outcomes(Ratification_Time ~
  0 * trt + 2.921e-05 * gdpp + 0.3186 * oil + 0.7143 * Coal +
  rnorm(N, mean = 0, sd = 1), assignment_variables = "trt")

design_null <- fake_population + null_outcomes + assignment +
  treatment_outcomes

# Add the estimator to new design for the null hypothesis
# test.
design_lm_null <- design_null + lm_robust_estimator1

# Set the seed for reproducibility
set.seed(12345)

# Simulate
sim_null <- simulate_design(design_lm_null, sims = 500)

# Extract p-values from the simulations
term_filtered <- sim_null[sim_null$term == "trt", ]
estimator_filtered <- term_filtered[term_filtered$estimator ==
  "lm_robust", ]
head(estimator_filtered)
p_values <- estimator_filtered$p.value

# Define the statistical significance level
alpha <- 0.05

```

```

# Calculate the Type I error rate
type_1_errors <- p_values < alpha
type_1_error_rate <- mean(type_1_errors)

type_1_error_df <- data.frame(Type_I_Error_Rate = type_1_error_rate)

# Use kable to create a table
kable(type_1_error_df, caption = "Type I Error Rate", align = "c")

# Nearest neighbor matching
fnearest <- matchit(trt ~ oil + gdpp + Coal, method = "nearest",
  caliper = 0.2, data = df1_fake)
matched_data_nr <- match.data(fnearest)
balance_sim_check1 <- xBalance(trt ~ oil + gdpp + Coal, strata = list(raw = NULL),
  data = matched_data_nr, report = "all")
kable(balance_sim_check1$overall)

# Construct estimand using (nearest-neighbor) matched data
ps_estimand <- function(data) {
  bs <- coef(lm_robust(Ratification_Time ~ trt, data = matched_data_nr))
  return(data.frame(estimand_label = "nearest_estimand", estimand = bs["trt"],
    stringsAsFactors = FALSE))
}
estimand2 <- declare_inquiry(handler = ps_estimand, label = "pop_relationship2")
design_estimand2 <- fake_population + sampling_1 + estimand2

kable(estimand2(df1_fake), caption = "Estimand 2 \\label{tab:estmnd2}")

library(estimatr)
# Create lm nearest matching estimator
estimator2_lm_ps <- declare_estimator(Ratification_Time ~ trt +
  gdpp + oil + Coal, model = lm_robust, term = c("trt"), inquiry = c("trt"),
  label = "lm_nearest_matching")

design_lm_ps <- design_estimand2 + estimator2_lm_ps

set.seed(123345)
sim_full_ps <- simulate_design(design_lm_ps, sims = 500)

diag1_lm_ps <- diagnose_design(sim_full_ps)

estimator2perform <- reshape_diagnosis(diag1_lm_ps, digits = 2,

```

```

    select = NULL, exclude = NULL)
estimator2perform
kable(estimator2perform[, c(1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12)])

# Declare potential outcomes under the null hypothesis
null_outcomes_nr <- declare_potential_outcomes(Ratification_Time ~
  0 * trt + 2.921e-05 * gdpp + 0.3186 * oil + 0.7143 * Coal +
  rnorm(N, mean = 0, sd = 1), assignment_variables = "trt")

# Define the null hypothesis design
design_null_nr <- fake_population + null_outcomes_nr + assignment +
  treatment_outcomes

# Combine the null hypothesis design with the estimator
design_lm_null_nr <- design_null_nr + estimator2_lm_ps

# Set the seed for reproducibility
set.seed(12345)

# Simulate the design
sim_null_nr <- simulate_design(design_lm_null_nr, sims = 500)
term_filtered_nr <- sim_null_nr[sim_null_nr$term == "trt", ]
estimator_filtered_nr <- term_filtered_nr[term_filtered_nr$estimator ==
  "lm_nearest_matching", ]
p_values_nr <- estimator_filtered_nr$p.value

# Define the statistical significance level
alpha <- 0.05

# Calculate the Type I error rate
type_1_errors_nr <- p_values_nr < alpha
type_1_error_rate_nr <- mean(type_1_errors_nr)
type_1_error_df_nr <- data.frame(Type_I_Error_Rate = type_1_error_rate_nr)

# Use kable to create a table
kable(type_1_error_df_nr, caption = "Type I Error Rate", align = "c")
# Create a propensity score model
ps_model <- glm(trt ~ oil + gdpp + Coal, data = df1_fake, family = "binomial")

# Calculate propensity scores and nearest-neighbor matching
propensity_scores <- predict(ps_model, type = "response")
psmatching <- matchit(trt ~ oil + gdpp + Coal, data = df1_fake,
  method = "nearest", caliper = 0.2)

```

```

# Perform matching
matched_data_ps <- match.data(psmatching)

# Check balance
balance_check_ps <- summary(psmatching)

# Extract balance of improvement from the summary
percent_balance_improvement <- data.frame(Covariate = c("distance",
  "oil", "gdpp", "Coal"), `Std. Mean Diff.` = c(96.1, 95.8,
  86.9, 67.3), `Var. Ratio` = c(90.4, 74.1, 60, 81.4), `eCDF Mean` = c(96,
  50.7, 89.3, -80.1), `eCDF Max` = c(81.6, 4.1, 78.3, -150))

kable(percent_balance_improvement, digits = 1)

# Construct an estimand using matched data
ps_estimand_ps <- function(data) {
  bs <- coef(lm_robust(Ratification_Time ~ trt, data = matched_data_ps))
  return(data.frame(estimand_label = "ps_estimand", estimand = bs["trt"],
    stringsAsFactors = FALSE))
}
estimand3 <- declare_inquiry(handler = ps_estimand_ps, label = "pop_relationship")
design_estimand3 <- fake_population + sampling_1 + estimand3
estimator3_lm_ps <- declare_estimator(Ratification_Time ~ trt +
  gdpp + oil + Coal, model = lm_robust, term = c("trt"), inquiry = c("trt"),
  label = "lm_ps_matching")

design_lm_ps_nr <- design_estimand3 + estimator3_lm_ps

set.seed(123345)
sim_full_ps_nr <- simulate_design(design_lm_ps_nr, sims = 500)

diag1_lm_ps_nr <- diagnose_design(sim_full_ps_nr)

estimator3perform <- reshape_diagnosis(diag1_lm_ps_nr, digits = 2,
  select = NULL, exclude = NULL)
estimator3perform
kable(estimator3perform[, c(1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12)])

# Declare potential outcomes under the null hypothesis
estimator3_lm_ps <- declare_estimator(Ratification_Time ~ trt +
  gdpp + oil + Coal, model = lm_robust, term = c("trt"), inquiry = c("trt"),
  label = "lm_ps_matching")

```



```

null_outcomes_nr_ps <- declare_potential_outcomes(Ratification_Time ~
  0 * trt + 2.921e-05 * gdpp + 0.3186 * oil + 0.7143 * Coal +
  rnorm(N, mean = 0, sd = 1), assignment_variables = "trt")

# Define the null hypothesis design
design_null_nr_ps <- fake_population + null_outcomes_nr + assignment +
  treatment_outcomes

# Combine the null hypothesis design with the estimator
design_lm_null_nr_ps <- design_null_nr_ps + estimator3_lm_ps

# Set the seed for reproducibility
set.seed(12345)

# Simulate the design
sim_null_nr_ps <- simulate_design(design_lm_null_nr_ps, sims = 500)
term_filtered_nr_ps <- sim_null_nr_ps[sim_null_nr_ps$term ==
  "trt", ]
estimator_filtered_nr_ps <- term_filtered_nr_ps[term_filtered_nr_ps$estimator ==
  "lm_ps_matching", ]
p_values_nr_ps <- estimator_filtered_nr_ps$p.value

# Define the statistical significance level
alpha <- 0.05

# Calculate the Type I error rate
type_1_errors_nr_ps <- p_values_nr_ps < alpha
type_1_error_rate_nr_ps <- mean(type_1_errors_nr_ps)
type_1_error_df_nr_ps <- data.frame(Type_I_Error_Rate = type_1_error_rate_nr_ps)

kable(type_1_error_df_nr_ps, caption = "Type I Error Rate", align = "c")

est1 <- lm(Ratification_Time ~ trt + Coal + oil + gdpp, data = df1_fake)
est1_summary <- summary(est1)
# Extract coefficients
coefficients_df <- as.data.frame(est1_summary$coefficients)
# Use kable to create a table
kable(coefficients_df, caption = "Treatment from Estimator 1",
  digits = 4)

library(sensemakr)
model.sensitivity <- sensemakr(model = est1, treatment = "trt",
  benchmark_covariates = c("oil", "Coal", "gdpp"), kd = c(0.01,

```

```
0.05, 0.1), ky = 1:3, q = 1, alpha = 0.05, reduce = TRUE)
```

## 11 Reference

- Amusa, L.B., Zewotir, T., & North, D. (2019). A Weighted Covariate Balancing Method of Estimating Causal Effects in Case-Control Studies. *Modern Applied Science*.
- Bättig, M. B., & Bernauer, T. (2009). National institutions and global public goods: are democracies more cooperative in climate change policy?. *International organization*, 63(2), 281-308.
- Clark, W. R., Golder, M., & Golder, S. N. (2017). *Principles of comparative politics*. CQ Press.
- Dai, X. (2006). The conditional nature of democratic compliance. *Journal of Conflict Resolution*, 50(5), 690-713.
- Kroll, S., & Shogren, J. F. (2013). Domestic politics and climate change: international public goods in two-level games. In *The Politics of Climate Change* (pp. 108-128). Routledge.
- Gerber, A. S., & Green, D. P. (2012). Field experiments: Design analysis and interpretation (Illustrated ed.). W. W. Norton & Company.
- Hansen, B. B., & Bowers, J. (2008). Covariate balance in simple, stratified and clustered comparative studies. *Statistical Science*, 219-236.
- Imai, K. (2005). Do get-out-the-vote calls reduce turnout? The importance of statistical methods for field experiments. *American Political Science Review*, 99(2), 283-300.
- Jacobson, D., and Edith, W. (1990). National Implementation and Compliance with International Environmental Accords. *Items*. Social Science Research Council.
- Mariani, A. W., & Pego-Fernandes, P. M. (2014). Observational studies: why are they so important?. *Sao Paulo Medical Journal*, 132, 01-02.
- Mattei, A., Mealli, F., & Rubin, D. B. (2011). Missing data and imputation methods. *Modern Analysis of Customer Surveys: With Applications Using R*, 129-154.
- Mitchell, R. B. (2003). International environmental agreements: a survey of their features, formation, and effects. *Annual review of environment and resources*, 28(1), 429-461.
- Neumayer, E. (2002). Do democracies exhibit stronger international environmental commitment? A cross-country analysis. *Journal of peace research*, 39(2), 139-164.
- Paquin, S. (2010). Federalism and compliance with international agreements: Belgium and Canada compared. *The Hague Journal of Diplomacy*, 5(1-2), 173-197.

- Peritz, L. (2020). When are international institutions effective? The impact of domestic veto players on compliance with WTO rulings. *International Studies Quarterly*, 64(1), 220-234.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55.
- Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American statistical Association*, 79(387), 516-524.
- Rosenbaum, P. R. (2010). Design sensitivity and efficiency in observational studies. *Journal of the American Statistical Association*, 105(490), 692-702.
- Rubin, D. B., & Thomas, N. (2000). Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association*, 95(450), 573-585.
- Stein, J. (2008). The International Law and Politics of Climate Change: Ratification of the United Nations Framework Convention and the Kyoto Protocol. *The Journal of Conflict Resolution*, 52(2), 243-268. <http://www.jstor.org/stable/27638605>
- Stuart E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science : a review journal of the Institute of Mathematical Statistics*, 25(1), 1-21. <https://doi.org/10.1214/09-STS313>
- Tsebelis, G. (2002). *Veto players: How political institutions work*. Princeton University Press.