Problem Set - Non-Compliance

Yuxuan Mei

1. First, let's start with the data that compares the control and treatment (calls encouraging people to vote). Read the data called 'noncompliance_treat_small.csv' using the function fread. Suggestion: To keep this dataset straight with the other one we'll load in later, I would give it a name like data_treatvscontrol. Remember that the variables are defined on the assignment Google Drive.

```
library(data.table)
library(fixest)
library(broom)
library(lfe)
```

```
## Loading required package: Matrix
```

```
### use the fread function to read the data.
data_treatvscontrol <- fread('noncompliance_treat_small.csv')
```

2. What is the sample size of this dataset?

```
nrow(data_treatvscontrol)
```

```
## [1] 293412
```

The sample size of the dataset is 293,412

3. Calculate the intent-to-treat effect (ITT) of using regression like we have in previous assignments. That is, what is the average treatment effect (ATE) of being assigned to the treatment group on voter turnout? You should use a regression function to do this. In this problem set, we'll be using the function 'feols' from the fixest package. This is a more powerful version of the 'lm' function with a very similar syntax. Note the code below, which provides a template for the regression.

```
# Your regression
# Note, 'hetero' here changes the standard errors to be heteroskedasticity robust.
# You don't have to worry about knowing what that means.
#this_reg <- feols(outcome ~ treatment_assignment, data = this_data, se = 'hetero')
this_reg <- feols(voted_aug2008 ~ treatment_attempt_turnout_call + voted_nov2002, data = data_treatvsco

# Use the function 'etable' or 'modelsummary' to output it nicely.
etable(this_reg)
```

```
##                                      this_reg
## Dependent Var.:                   voted_aug2008
##
## (Intercept)                    0.1080*** (0.0010)
## treatment_attempt_turnout_call  0.0121** (0.0045)
```

```
## voted_nov2002                  0.1174*** (0.0014)
## ______________________________ __________________
## S.E. type                      Heteroskedas.-rob.
## Observations                              293,412
## R2                                        0.01992
## Adj. R2                                   0.01991
```

The ITT of voter turnout is 0.012, which is statistically significant at a 5% significance level.

4a. First, calculate the compliance rate. In other words, what proportion of people in the treatment condition were successfully contacted? Save this value as a variable called alpha.

```
alpha <- sum(data_treatvscontrol[treatment_attempt_turnout_call == 1, contacted])/nrow(data_treatvsconti
```

The compliance rate is 57.4% ### 4b. Divide your ITT estimate by alpha and the standard error by alpha. How do you interpret this result? (4 points)

```
ITT <- 0.0121
se <- 0.0045
CACE <- ITT/alpha
CACE_se <- se/alpha
```

ITT/alpha is the complier average casual effect (CACE). The CACE (ATE for compliers) of phone call on voter turnout is 2.1%, with a standard error of 0.0078.

4c. Instead of calculating the CACE by hand, we can do it using the 'feols' function in R. Notice the standard errors are slightly different.

```
CACE_reg <- feols(voted_aug2008 ~ 1 | contacted ~ treatment_attempt_turnout_call, data = data_treatvscon

etable(CACE_reg)
```

```
##                            CACE_reg
## Dependent Var.:      voted_aug2008
##
## (Intercept)     0.1870*** (0.0007)
## contacted        0.0209** (0.0079)
## _______________ __________________
## S.E. type       Heteroskedas.-rob.
## Observations                293,412
## R2                          0.00018
## Adj. R2                     0.00017
```

The CACE is still 2.1%, with a standard error of 0.0079.

5. Let's turn to the placebo data. Read the datafile called 'noncompliance_placebo.csv'.

Suggestion: to keep things straight, I would give this dataset a name like data_with_placebo.

Note: The treatment_attempt_turnout_call variable is still there, but has a slightly different meaning now. When it is set to 1, that still means that someone was in the treatment group attempted with a turnout call. However, when it is set to 0 in this dataset, it now means that they're in the placebo group attempted with a placebo call.

```
data_with_placebo <- fread('noncompliance_placebo.csv')
```

6. What is the sample size of this data set? How much smaller or larger is it than the other dataset?

```
nrow(data_with_placebo)
```

```
## [1] 47540
```

```
nrow(data_with_placebo) - nrow(data_treatvscontrol)
```

```
## [1] -245872
```

The sample size of the placebo dataset is 47540, which is about 1/6 of the size of the other dataset.

7. Using regression, examine whether, just within the placebo group, those who answered the phone turn out to vote at the same rates as those who don't answer the phone. What is your interpretation of this result? Does this indicate that the placebo caused people to vote at a higher rate? Or do you interpret this pattern in another way? No more than two sentences.

```
###  Hint: to run a regression in the subset of the data that is in the placebo group:
placebo_reg <- feols(voted_aug2008 ~ contacted, data = data_with_placebo[treatment_attempt_turnout_call

etable(placebo_reg)
```

```
##                         placebo_reg
## Dependent Var.:       voted_aug2008
##
## (Intercept)      0.1618*** (0.0036)
## contacted        0.0438*** (0.0050)
## _______________ __________________
## S.E. type        Heteroskedas.-rob.
## Observations                23,761
## R2                         0.00312
## Adj. R2                    0.00307
```

In the placebo group, compliers and non-compliers voter turnout rate is statistically different (compliers' voter turnout rate is 4.38% higher). This doesn't mean the placebo caused people to vote at a higher rate.It only shows the turnout rate of compliers in the placebo group (who would have answered the call if in treatment group).

8a. To estimate the CACE, run a regression that calculates the effect of treatment on turnout only among those who were successfully contacted.

```
CACE_reg2 <- feols(voted_aug2008 ~ treatment_attempt_turnout_call, data = data_with_placebo[contacted ==

etable(CACE_reg2)
```

```
##                                     CACE_reg2
## Dependent Var.:                   voted_aug2008
##
## (Intercept)                    0.2056*** (0.0035)
## treatment_attempt_turnout_call 0.0306*** (0.0051)
## ______________________________ __________________
## S.E. type                      Heteroskedas.-rob.
## Observations                                26,793
## R2                                         0.00136
## Adj. R2                                    0.00132
```

8b. How do we interpret these estimates?

The CACE (the treatment effect on compliers) is about 3.06% with a standard error of 0.0051.

8c. This dataset has a much smaller sample size than the first dataset you looked at. Why is the standard error in this dataset not many times bigger?

The problem with ITT/alpha is that the standard error is amplified, thus have to be make up by a much larger sample size. The standard error using the placebo group method does not have this limitation, thus is smaller.

9: Run a regression to test whether the effects of the turnout call vary by whether the person voted in 2002. Interpret the results.

```
# Reminder, we can add an interaction effect as follows:
# feols(outcome ~ treatment*covariate, data = name_of_data, se = 'hetero')

final_reg <- feols(voted_aug2008 ~ treatment_attempt_turnout_call*voted_nov2002, data = data_with_placel

etable(final_reg)
```

```
##                                                 final_reg
## Dependent Var.:                             voted_aug2008
##
## (Intercept)                                0.1335*** (0.0054)
## treatment_attempt_turnout_call               -0.0063 (0.0075)
## voted_nov2002                              0.1031*** (0.0069)
## treatment_attempt_turnout_call x voted_nov2002 0.0517*** (0.0099)
## ____________________________________       __________________
## S.E. type                                  Heteroskedas.-rob.
## Observations                                           26,793
## R2                                                    0.02241
## Adj. R2                                               0.02230
```

The CATE for voter turnout rate is 4.54% when taking the voting information in 2002 as a covariate.

BONUS: Propose and conduct a test of the assumptions required for a placebo analysis.

Conduct a difference in difference calculation on the regression with an interaction term between treatment and placebo to see if the placebo will have an effect on the turnout rate.

```
data_with_placebo[treatment_attempt_turnout_call == 0 & contacted == 1, placebo:=1]
data_with_placebo[is.na(data_with_placebo)] <- 0

assumption_reg <- feols(voted_aug2008 ~ treatment_attempt_turnout_call*placebo, data = data_with_placebo
```

```
## Variables 'placebo' and 'treatment_attempt_turnout_call:placebo' have been removed because of collin
```

```
etable(assumption_reg)
```

```
##                                  assumption_reg
## Dependent Var.:                    voted_aug2008
##
## (Intercept)                    0.2056*** (0.0035)
## treatment_attempt_turnout_call 0.0306*** (0.0051)
## ______________________________ __________________
## S.E. type                      Heteroskedas.-rob.
```

```
## Observations                          26,793
## R2                                  0.00136
## Adj. R2                             0.00132
```

The regression results shows that the placebo does not have an effect on the outcome.

How long did this problem set take you? What is the difficulty level?

3.5 hours, it's reasonable.