Problem Set 3

Anita Ye - yy3557 - 001

Due Dec 1, 2023

This homework must be turned in on Brightspace by Dec. 1, 2023. It must be your own work, and your own work only – you must not copy anyone's work, or allow anyone to copy yours. This extends to writing code. You may consult with others, but when you write up, you must do so alone.

Your homework submission must be written and submitted using Rmarkdown. No handwritten solutions will be accepted. **No zip files will be accepted. Make sure we can read each line of code in the pdf document.** You should submit the following:

1. A compiled PDF file named yourNetID_solutions.pdf containing your solutions to the problems.
2. A .Rmd file containing the code and text used to produce your compiled pdf named yourNetID_solutions.Rmd.

Note that math can be typeset in Rmarkdown in the same way as Latex. Please make sure your answers are clearly structured in the Rmarkdown file:

1. Label each question part
2. Do not include written answers as code comments.
3. The code used to obtain the answer for each question part should accompany the written answer. Comment your code!

Question 1 (Total: 100)

Does US military assistance strengthen or further weaken fragile and conflict-affected foreign governments? Aid may bolster state capacity and suppress violence from nonstate actors such as paramilitary groups. On the other hand, aid may be diverted to those same violent groups. To answer the question, Dube and Naidu (2015)(https://www.journals.uchicago.edu/doi/10.1086/679021?mobileUi=0) leverage changes in the allocation of US military aid to Colombian military bases. They test whether Colombian municipailites in which military bases are located have more or less paramilitary violence when the level of U.S. miliary aid increases, relative to Colombian municipalities in which miliary bases are not located.

For this problem, you will need the 'bases_replication_file.dta' file. The variables you will need are:

- parattq - DV here is paramilitary attacks
- bases6 - indicator variable whether or not there is a base in the municipality
- lrmilnar col - (logged) U.S. military and narcotics aid to Colombia
- bases6xlrmilnar col - the treatment i.e., the interaction between the level of U.S. military and narcotics aid and whether or not there is a base in the municipality
- lnnewpop - is log of population

Part a (60 points)

The treatment in this case is a continuous 'intensity' variable that changes over time. The authors use the interaction between the level of U.S. military and narcotics aid and whether a base exists in a municipality. How many units are in the 'control' group (no bases)? Does the bases variable change over time or is it a unit-constant factor? How about the logged military aid variable, does it change across units for a given year? What do the authors seem to be assuming about how military aid is allocated?

```r
library(tidyverse)
library(haven)
library(estimatr) # for lm with robust se : ?lm_robust()

# Load bases data
bases <- haven::read_dta("bases_replication_final.dta")

# How many observations are in the ``no bases group"
no_bases_count <- sum(bases$bases6 == 0, na.rm = TRUE)
print(paste("Number of observations in the no bases group:", no_bases_count))
```

```
## [1] "Number of observations in the no bases group: 16272"
```

```r
control_group_municipalities <- bases %>%
group_by(municipality) %>%
summarize(all_no_bases = all(bases6 == 0)) %>%
filter(all_no_bases) %>%
nrow()
#print(head(bases))
print(paste("Number of unique municipalities in the control group:", control_group_municipalities))
```

```
## [1] "Number of unique municipalities in the control group: 904"
```

```r
## How about each of them?
bases6_dist <- bases %>% count(year, bases6)
print(bases6_dist)
```

```
## # A tibble: 36 x 3
##      year bases6     n
##     <dbl>  <dbl> <int>
##   1  1988      0   904
```

```
## 2  1988      1    32
## 3  1989      0   904
## 4  1989      1    32
## 5  1990      0   904
## 6  1990      1    32
## 7  1991      0   904
## 8  1991      1    32
## 9  1992      0   904
## 10 1992      1    32
## # i 26 more rows
```

```r
## How many municipalities do we have
lrmilnar_col_dist <- bases %>% count(year, lrmilnar_col)
print(lrmilnar_col_dist)
```

```
## # A tibble: 18 x 3
##     year lrmilnar_col     n
##    <dbl>        <dbl> <int>
##  1  1988       -3.91    936
##  2  1989       -3.68    936
##  3  1990       -1.90    936
##  4  1991       -2.43    936
##  5  1992       -2.32    936
##  6  1993       -2.73    936
##  7  1994       -3.40    936
##  8  1995       -3.49    936
##  9  1996       -2.78    936
## 10  1997       -2.97    936
## 11  1998       -2.53    936
## 12  1999       -1.29    936
## 13  2000       -0.108   936
## 14  2001       -2.98    936
## 15  2002       -1.02    936
## 16  2003       -0.782   936
## 17  2004       -0.669   936
## 18  2005       -0.559   936
```

```r
# Check does the bases variable change over time.
bases_over_time <- bases %>%
group_by(year) %>%
summarize(mean_bases = mean(bases6))
print(bases_over_time)
```

```
## # A tibble: 18 x 2
##     year mean_bases
##    <dbl>      <dbl>
##  1  1988     0.0342
##  2  1989     0.0342
##  3  1990     0.0342
##  4  1991     0.0342
##  5  1992     0.0342
##  6  1993     0.0342
##  7  1994     0.0342
##  8  1995     0.0342
##  9  1996     0.0342
## 10  1997     0.0342
```

```
## 11  1998     0.0342
## 12  1999     0.0342
## 13  2000     0.0342
## 14  2001     0.0342
## 15  2002     0.0342
## 16  2003     0.0342
## 17  2004     0.0342
## 18  2005     0.0342
```

```r
unique_aid_by_year <- bases %>%
group_by(year) %>%
summarize(unique_aid = n_distinct(`lrmilnar_col`))
print(unique_aid_by_year)
```

```
## # A tibble: 18 x 2
##     year unique_aid
##    <dbl>      <int>
##  1  1988          1
##  2  1989          1
##  3  1990          1
##  4  1991          1
##  5  1992          1
##  6  1993          1
##  7  1994          1
##  8  1995          1
##  9  1996          1
## 10  1997          1
## 11  1998          1
## 12  1999          1
## 13  2000          1
## 14  2001          1
## 15  2002          1
## 16  2003          1
## 17  2004          1
## 18  2005          1
```

```r
values_by_year <- bases %>%
group_by(year) %>%
summarise(lrmilnar_col_value = list(unique(lrmilnar_col)))
values_by_year$numeric_value <- sapply(values_by_year$lrmilnar_col_value, function(x) as.numeric(x))
print(values_by_year)
```

```
## # A tibble: 18 x 3
##     year lrmilnar_col_value numeric_value
##    <dbl> <list>                     <dbl>
##  1  1988 <dbl [1]>                  -3.91
##  2  1989 <dbl [1]>                  -3.68
##  3  1990 <dbl [1]>                  -1.90
##  4  1991 <dbl [1]>                  -2.43
##  5  1992 <dbl [1]>                  -2.32
##  6  1993 <dbl [1]>                  -2.73
##  7  1994 <dbl [1]>                  -3.40
##  8  1995 <dbl [1]>                  -3.49
##  9  1996 <dbl [1]>                  -2.78
## 10  1997 <dbl [1]>                  -2.97
```

```
## 11  1998 <dbl [1]>                    -2.53
## 12  1999 <dbl [1]>                    -1.29
## 13  2000 <dbl [1]>                    -0.108
## 14  2001 <dbl [1]>                    -2.98
## 15  2002 <dbl [1]>                    -1.02
## 16  2003 <dbl [1]>                    -0.782
## 17  2004 <dbl [1]>                    -0.669
## 18  2005 <dbl [1]>                    -0.559
```

The dataset includes 16,272 observations in the 'control' group, representing municipalities without military bases, and there are 936 unique municipalities, 904 of them with no bases.

The bases6 variable, indicating the presence of a military base, is unit-constant, meaning the status of a municipality having a base does not change over time. This constancy is further evidenced by the consistent count of 904 municipalities without a base (bases6 = 0) and 32 with a base (bases6 = 1) each year.

Additionally, the lrmilnar_col variable, representing logged U.S. military and narcotics aid to Colombia, varies annually but remains constant across all municipalities within a single year. This pattern suggests that aid distribution is influenced by specific characteristics of the municipalities, like the intensity of paramilitary activities or alignment with U.S. interests.

The study hypothesizes that fluctuations in U.S. military and narcotics aid predominantly affect municipalities with military bases, potentially leading to different levels of paramilitary violence. By comparing paramilitary attacks in municipalities with and without bases against changes in military aid levels, the study aims to isolate the direct impact of this aid on paramilitary violence. The assumption is that changes in military aid are more pronounced in areas with bases, allowing for an analysis of the localized effects of military aid while controlling for other factors.

Part b (20 points)

The authors use a common empirical strategy called two-way fixed effects to estimate the average treatment effect of military aid. The model they estimate includes fixed effects for both time periods and units (and includes logged population as an additional covariate):

$$Y_{it} = \gamma_t + \alpha_i + \tau D_{it} + \beta X_{it} + \epsilon_{it}$$

What assumptions are the authors making in order to identify the treatment effect of military aid?

The study assumes constant treatment effects, proposing that the effect of the treatment (military aid) on the outcome (paramilitary violence) is consistent across different units and time periods. This is expressed mathematically as E[Yi(1) - Yi(0)|Xi = x] = E[Yi(1) - Yi(0)|Xi = w], indicating that the average treatment effect remains constant over time. Furthermore, the study makes a parallel trends assumption, which suggests that any selection bias present at different times is identical. Specifically, it's presumed that the trends in potential outcomes under control conditions from one time period to the next in the treated group mirror those observed in the control group. The model incorporates time-trend shocks (gamma t) and unit-level intercept shifts (alpha i) to adjust for potential confounding factors.

The authors hypothesize that without U.S. military aid, the trends in paramilitary violence in Colombian municipalities with and without military bases would have been similar over time. This premise is crucial for attributing changes in violence directly to variations in military aid. The model includes time fixed effects (gamma t), addressing factors that change over time but are uniform across all municipalities, like national policies or global events. Additionally, unit fixed effects (alpha i) account for unchanging, municipality-specific characteristics such as geographic location or historical violence levels. The study assumes that all relevant variables influencing both military aid distribution and paramilitary violence levels are either incorporated in the model or captured by the fixed effects, including variables like logged population. The allocation of military aid is considered exogenous, meaning it's not influenced by unobserved factors that also affect paramilitary violence levels. The assumption extends to the impact of military aid, which is presumed consistent across different municipalities and time periods, although the magnitude of this effect might vary. Lastly, the study assumes no interference between units, implying that the level of paramilitary violence in one municipality is not influenced by the military aid received by another municipality.

Part c (20 points)

Using the two-way fixed effects estimator, estimate the effect of U.S. military and narcotics aid on the number of paramilitary attacks, including log of population as a covariate. The two sets of fixed effects are for municipality (municipality) and year (year). Cluster your standard errors at the unit level (see the cluster argument in lm_robust. Report a 95% confidence interval for your estimate and interpret your results.

```r
#?lm_robust (set se_type to "CR0")
# Fit Regression using lm_robust

# Summarize
# Load necessary library
library(estimatr)

# Fit Regression using lm_robust
model <- lm_robust(
  paratt ~ bases6xlrmilnar_col + lnnewpop + factor(municipality) + factor(year),
  data = bases,
  clusters = bases$municipality,
  se_type = "CR0"
)



# Summarize the results
coefficients_summary <- summary(model)$coefficients[c("bases6xlrmilnar_col", "lnnewpop"), ]

# 95% Confidence Interval
confint_model <- confint(model)[c("bases6xlrmilnar_col", "lnnewpop"), ]

print("Coefficients Summary:")
```

```
## [1] "Coefficients Summary:"
```

```r
print(coefficients_summary)
```

```
##                      Estimate Std. Error  t value    Pr(>|t|)   CI Lower
## bases6xlrmilnar_col 0.1503116 0.06008643 2.501590 0.012533674 0.03239170
## lnnewpop            0.1178481 0.04523456 2.605266 0.009326138 0.02907504
##                      CI Upper  DF
## bases6xlrmilnar_col 0.2682315 935
## lnnewpop            0.2066211 935
```

```r
print("Confidence Intervals:")
```

```
## [1] "Confidence Intervals:"
```

```r
print(confint_model)
```

```
##                          2.5 %    97.5 %
## bases6xlrmilnar_col 0.03239170 0.2682315
## lnnewpop            0.02907504 0.2066211
```

```r
# 95% Confidence Interval
```

```r
#confint(model, level = 0.95)
```

In the conducted analysis utilizing a two-way fixed effects model, empirical evidence is presented on the relationship between U.S. military and narcotics aid and paramilitary violence in Colombian municipalities. The statistical analysis estimates that an incremental increase in the logged quantity of military aid in municipalities with military bases correlates with an elevation of 0.1503 in the average number of paramilitary attacks. This effect is statistically significant, as evidenced by a 95% confidence interval ranging from 0.0324 to 0.2682, excluding zero, thereby leading to the rejection of the null hypothesis at the alhpa = 0.05 significance level.

The interaction term coefficient of 0.1503 for bases6xlrmilnar_col is particularly revealing. It demonstrates a positive and significant association between the joint impact of U.S. military and narcotics aid and the presence of military bases on the frequency of paramilitary attacks. This association is substantiated by a robust standard error of 0.0601, a t-value of 2.502, and a p-value of 0.0125. Furthermore, the analysis also identifies a statistically significant positive correlation between the population size (indicated by the coefficient of 0.1178 for lnnewpop) and the incidence of paramilitary attacks.

These empirical findings imply that while the primary objective of U.S. military aid is to decrease violence, its interaction with specific local variables, such as the presence of military bases, can yield paradoxical outcomes. This highlights the complexities inherent in the effects of foreign aid in conflict zones, underscoring the need for nuanced understanding and evaluation of such interventions.

```r
tidy(lm_robust(paratt~bases6+lrmilnar_col+bases6xlrmilnar_col+lnnewpop, data = bases, fixed_effects= ~
               se_type = "CR0",
               clusters =municipality))
```

```
## 1 coefficient  not defined because the design matrix is rank deficient

##                  term      estimate    std.error    statistic     p.value
## 1              bases6 -2.256998e+10 1.756246e+13 -0.001285127 0.998974892
## 2        lrmilnar_col            NA           NA           NA          NA
## 3 bases6xlrmilnar_col  1.503119e-01 6.001576e-02  2.504539638 0.012430338
## 4            lnnewpop  1.178481e-01 4.523070e-02  2.605488898 0.009320115
##        conf.low    conf.high  df  outcome
## 1 -3.448897e+13 3.444383e+13 935  paratt
## 2            NA           NA  NA  paratt
## 3  3.253066e-02 2.680931e-01 935  paratt
## 4  2.908264e-02 2.066135e-01 935  paratt
```