Problem Set 3

Yiyun (Leo) Yao - yy3959 - (Recitation) 002

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_base <- bases %>%
  filter(bases6 == 0)
no_base_num <- no_base %>% nrow()
no_base_num
```

```
## [1] 16272
```

There are 16272 observations in the control group (no bases).

```r
## How about each of them?

bases6_count_total <- bases %>%
  group_by(year) %>%
  summarise(count_total = n(), unique_bases6_count_total = n_distinct(bases6),
            .groups = 'drop')
bases6_count_total
```

```
## # A tibble: 18 x 3
##     year count_total unique_bases6_count_total
##    <dbl>       <int>                     <int>
##  1  1988         936                         2
##  2  1989         936                         2
##  3  1990         936                         2
##  4  1991         936                         2
##  5  1992         936                         2
##  6  1993         936                         2
##  7  1994         936                         2
##  8  1995         936                         2
##  9  1996         936                         2
## 10  1997         936                         2
## 11  1998         936                         2
## 12  1999         936                         2
## 13  2000         936                         2
## 14  2001         936                         2
## 15  2002         936                         2
## 16  2003         936                         2
## 17  2004         936                         2
## 18  2005         936                         2
```

```r
bases6_count_control <- no_base %>%
  group_by(year) %>%
  summarise(count_control = n(), unique_bases6_count_control = n_distinct(bases6),
            .groups = 'drop')
bases6_count_control
```

```
## # A tibble: 18 x 3
##     year count_control unique_bases6_count_control
##    <dbl>         <int>                       <int>
##  1  1988           904                           1
##  2  1989           904                           1
##  3  1990           904                           1
##  4  1991           904                           1
##  5  1992           904                           1
##  6  1993           904                           1
##  7  1994           904                           1
##  8  1995           904                           1
##  9  1996           904                           1
## 10  1997           904                           1
## 11  1998           904                           1
## 12  1999           904                           1
## 13  2000           904                           1
## 14  2001           904                           1
## 15  2002           904                           1
## 16  2003           904                           1
## 17  2004           904                           1
## 18  2005           904                           1
```

The bases variable is constant in time. It does not change over time.

```r
logaid_count_total <- bases %>%
  group_by(municipality) %>%
  summarise(count_total = n(), .groups = 'drop')
logaid_count_total
```

```
## # A tibble: 936 x 2
##    municipality count_total
##           <dbl>       <int>
##  1         5001          18
##  2         5002          18
##  3         5004          18
##  4         5021          18
##  5         5030          18
##  6         5031          18
##  7         5034          18
##  8         5036          18
##  9         5038          18
## 10         5040          18
## # i 926 more rows
```

```r
logaid_count_control <- no_base %>%
  group_by(municipality) %>%
  summarise(count_control = n(), .groups = 'drop')
logaid_count_control
```

```
## # A tibble: 904 x 2
##    municipality count_control
##           <dbl>         <int>
##  1         5002            18
##  2         5004            18
##  3         5021            18
##  4         5030            18
##  5         5031            18
##  6         5034            18
##  7         5036            18
##  8         5038            18
##  9         5040            18
## 10         5042            18
## # i 894 more rows
```

The logged military aid variable is constant in unit. It does not change across units for a given year.

```r
## How many municipalities do we have
municipalities_count_total <- bases %>%
  group_by(year) %>%
  summarise(count_total = n_distinct(municipality), .groups = 'drop')
municipalities_count_total
```

```
## # A tibble: 18 x 2
##     year count_total
##    <dbl>       <int>
```

```
## 1 1988 936
## 2 1989 936
## 3 1990 936
## 4 1991 936
## 5 1992 936
## 6 1993 936
## 7 1994 936
## 8 1995 936
## 9 1996 936
## 10 1997 936
## 11 1998 936
## 12 1999 936
## 13 2000 936
## 14 2001 936
## 15 2002 936
## 16 2003 936
## 17 2004 936
## 18 2005 936
```

```
## How many municipalities do we have
municipalities_count_control <- no_base %>%
  group_by(year) %>%
  summarise(count_control = n_distinct(municipality), .groups = 'drop')
municipalities_count_control
```

```
## # A tibble: 18 x 2
##     year count_control
##    <dbl>         <int>
##  1  1988           904
##  2  1989           904
##  3  1990           904
##  4  1991           904
##  5  1992           904
##  6  1993           904
##  7  1994           904
##  8  1995           904
##  9  1996           904
## 10  1997           904
## 11  1998           904
## 12  1999           904
## 13  2000           904
## 14  2001           904
## 15  2002           904
## 16  2003           904
## 17  2004           904
## 18  2005           904
```

In the control group (no base), we have 904 municipalities. Overall, we have 936 municipalities.

The authors seem to be assuming that military aid is allocated based on the presence of military bases in the municipalities. From the results above, I'm inferring that the author is assuming the military aid is consistent across the years for each municipality and that the aid is uniformly distributed for each municipality within the same year.

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?

1. SUTVA
2. Ignorability
3. Unconfoundedness
4. Time-invariant unobserved heterogeneity / Unobserved confounding is constant within each group. For example, any unobserved factors that could potentially influence both the treatment assignment and the outcome are assumed to be constant over time. Another assumption that also lies in this category would be that, for each municipality/year, there are unobserved factors that remain constant over time and are unique to each municipality/year, and these factors would affect the outcome. This helps control for unobserved factors that are constant within each municipality and do not change over time.
5. Parallel trends / Common trends
6. Outcome model is linear / Independence of errors

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
model <- tidy(lm_robust(paratt ~ bases6xlrmilnar_col + lnnewpop
                        + factor(municipality) + factor(year),
                        data = bases, clusters = municipality, se_type = 'CR0'))

# Summarize
summary(model)
```

```
##      term              estimate         std.error           statistic
##  Length:955         Min.   :-3.129   Min.   :0.0009686   Min.   :-2922.508
##  Class :character   1st Qu.:-2.801   1st Qu.:0.2207440   1st Qu.:  -12.394
##  Mode  :character   Median :-2.727   Median :0.2439929   Median :  -11.052
##                     Mean   :-2.644   Mean   :0.2373292   Mean   :  -14.453
##                     3rd Qu.:-2.641   3rd Qu.:0.2657525   3rd Qu.:   -9.907
##                     Max.   : 3.255   Max.   :0.3431439   Max.   :   24.494
##     p.value            conf.low         conf.high            df
##  Min.   :0.000000   Min.   :-3.434   Min.   :-2.838   Min.   :935
##  1st Qu.:0.000000   1st Qu.:-3.278   1st Qu.:-2.342   1st Qu.:935
##  Median :0.000000   Median :-3.239   Median :-2.238   Median :935
##  Mean   :0.002458   Mean   :-3.109   Mean   :-2.178   Mean   :935
##  3rd Qu.:0.000000   3rd Qu.:-3.155   3rd Qu.:-2.127   3rd Qu.:935
##  Max.   :0.970358   Max.   : 2.994   Max.   : 3.515   Max.   :935
##    outcome
```

```
##  Length:955
##  Class :character
##  Mode  :character
##
##
##
```

```r
# Extracting the treatment effect
model[2, ]
```

```
##                  term  estimate  std.error statistic    p.value  conf.low
## 2 bases6xlrmilnar_col 0.1503116 0.06008643   2.50159 0.01253367 0.0323917
##   conf.high  df outcome
## 2 0.2682315 935  paratt
```

From the summary of the results, we can see that municipalities, time, and population all have different impacts on the paramilitary attacks. The estimate for the treatment effect is 0.15 at a 95% confidence interval [0.032, 0.268]. The standard error is 0.06, which is very small. This estimate of 0.15 means that the U.S. military and narcotics aid increases the average number of paramilitary attacks by 0.15, including log of population as a covariate. This shows the positive relationship between military aids and number paramilitary attacks. More aids causes higher level of paramilitary violence.