

# Problem Set 3

Andrew Shao - as13381 - 005

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 your-NetID\_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 municipalities in which military bases are located have more or less paramilitary violence when the level of U.S. military aid increases, relative to Colombian municipalities in which military 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
- bases6xlrmlnar\_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?

```
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"
n_control = sum(bases$bases6 == 0)
print(n_control)
```

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

```
# Print number of municipalities whose bases6 value changes between years
bases %>%
  group_by(municipality) %>%
  summarise(bases6_change = n_distinct(bases6) > 1) %>%
  filter(bases6_change) %>%
  nrow()
```

```
## [1] 0
```

```
# Print number of years whose lrmilnar_col value changes between municipalities
bases %>%
  group_by(year) %>%
  summarise(lrmilnar_col_change = n_distinct(lrmilnar_col) > 1) %>%
  filter(lrmilnar_col_change) %>%
  nrow()
```

```
## [1] 0
```

There are 16,272 units in the control group of no bases. Since there are no municipalities whose bases6 column has more than 1 unique value between different years, we can conclude that the bases variable doesn't change over time. Because there are no years whose lrmilnar\_col column has more than 1 unique value between different municipalities, we can also conclude that the logged military aid variable doesn't change across units for a given year. This implies that the authors assume military aid is allocated evenly between municipalities with no spillover effects.

### 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 authors are assuming the relationship between the outcome and independent variables is linear. They also assume that the unobserved confounders are constant within groups. Additionally, the authors assumed the parallel trends assumption to be met, that the selection bias is the same at different times which causes the trend in potential outcomes between times in the treated group is the same as in the control group. In other words, in the absence of treatment, the difference between treatment and control group outcomes is constant between times.

### 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.

```
##?lm_robust (set se_type to "CRO")
# Fit Regression using lm_robust
lin_reg = lm_robust(paratt ~ bases6 + lrmilnar_col + bases6xlrmlnar_col + lnnewpop,
                    data = bases,
                    clusters = municipality,
                    fixed_effects = municipality + year,
                    se_type = 'CRO')
# Summarize
tidy(lin_reg)
```

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

	term	estimate	std.error	statistic	p.value
## 1	bases6	NA	NA	NA	NA
## 2	lrmilnar_col	2.167531e+09	3.389352e+11	0.00639512	0.994898831
## 3	bases6xlrmlnar_col	1.503116e-01	6.008642e-02	2.50158975	0.012533666
## 4	lnnewpop	1.178481e-01	4.523458e-02	2.60526492	0.009326166

```
##
```

	conf.low	conf.high	df	outcome
## 1	NA	NA	NA	paratt
## 2	-6.629943e+11	6.673294e+11	935	paratt
## 3	3.239171e-02	2.682314e-01	935	paratt
## 4	2.907500e-02	2.066211e-01	935	paratt

The 95% confidence interval for the treatment is [3.239171e-02, 2.682314e-01]. This interval notably doesn't include 0 so we conclude that the effect is significant and reject the null hypothesis that there is no treatment effect; US military and narcotics aid to Columbia increases the intensity of paramilitary attacks.