

# Problem Set 3

Petra Ivanovic - pi2018 - Section 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")

# Let's filter our df for only those variables relevant to us
# Selecting specific columns
bases <- bases %>%
  select(municipality, year, paratt, bases6,
         lrmilnar_col, bases6xlrmlnar_col, lnnewpop)
```

bases

```
## # A tibble: 16,848 x 7
##   municipality year paratt bases6 lrmilnar_col bases6xlrmlnar_col lnnewpop
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1      5113  1988     0     0    -3.91     0    -4.93
## 2      5310  1988     0     0    -3.91     0    -4.69
## 3      5861  1988     0     0    -3.91     0    -4.35
## 4      5809  1988     0     0    -3.91     0    -4.30
## 5     54405  1988     0     0    -3.91     0    -3.27
## 6      5134  1988     0     0    -3.91     0    -4.59
## 7     41668  1988     0     0    -3.91     0    -3.73
## 8      5055  1988     0     0    -3.91     0    -4.43
```

```
## 9      18460 1988      0      0      -3.91      0      -4.28
## 10     47692 1988      0      0      -3.91      0      -4.12
## # i 16,838 more rows
```

Here we can see our table filtered for only features relevant for our analysis as indicated in the question.

```
# How many observations are in the "no bases group"
```

```
no_bases_count <- bases %>%
  filter(bases6 == 0) %>%
  nrow()
```

```
no_bases_count
```

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

```
# Count how many municipalities do not have bases
```

```
count_municipalities_with_only_zero <- bases %>%
  group_by(municipality) %>% # Group by municipality
  summarise(all_zero = all(bases6 == 0)) %>% # Check if all bases6 values are 0
  filter(all_zero) %>% # Keep only those with all 0s
  nrow() # Count the number of such municipalities
```

```
print(count_municipalities_with_only_zero)
```

```
## [1] 904
```

We can see that in our table there are 16272 values in the “no bases group” which indicates that 16272 of our units do not have bases. When we filter this for distinct municipalities we can see that there are 904 municipalities have no bases.

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

```
# Checking if the bases variable changes over time
```

```
bases_change_over_time <- bases %>%
  group_by(municipality) %>%
  summarise(unique_bases_values = n_distinct(bases6))
```

```
bases_change_over_time
```

```
## # A tibble: 936 x 2
```

```
##   municipality unique_bases_values
```

```
##           <dbl>           <int>
```

```
## 1           5001             1
```

```
## 2           5002             1
```

```
## 3           5004             1
```

```
## 4           5021             1
```

```
## 5           5030             1
```

```
## 6           5031             1
```

```
## 7           5034             1
```

```
## 8           5036             1
```

```
## 9           5038             1
```

```
## 10          5040             1
```

```
## # i 926 more rows
```

```
count_non_1_values <- bases_change_over_time %>%
  filter(unique_bases_values != 1) %>%
  nrow()

print(paste("Number of municipalities with more than 1 bases values:",
            count_non_1_values))
```

```
## [1] "Number of municipalities with more than 1 bases values: 0"
```

From above output we can see that for each of our municipalities there is only 1 unique value for bases which indicates it does not change through time as if it did there would be multiple unique values and not just 1. Further from the “count\_non\_1\_values” being 0 it also shows us that there are no municipalities with any number of unique bases values but 1. Due to this we can conclude that bases is unit-constant for each municipality - status of military bases does not change over time for any of the municipalities we have information on. In other words, bases variable is constant for all of our municipalities for the time period covered by the data.

```
# Examining changes in logged military aid across units
aid_variation_across_units <- bases %>%
  group_by(year, municipality) %>%
  summarise(unique_aid_values = n_distinct(lrmilnar_col))
```

```
## ‘summarise()’ has grouped output by ‘year’. You can override using the
## ‘.groups’ argument.
```

```
aid_variation_across_units
```

```
## # A tibble: 16,848 x 3
## # Groups:   year [18]
##   year municipality unique_aid_values
##   <dbl>         <dbl>         <int>
## 1  1988           5001             1
## 2  1988           5002             1
## 3  1988           5004             1
## 4  1988           5021             1
## 5  1988           5030             1
## 6  1988           5031             1
## 7  1988           5034             1
## 8  1988           5036             1
## 9  1988           5038             1
## 10 1988           5040             1
## # i 16,838 more rows
```

```
# Number of municipalities with non unique aid values
non_unique_aid_count <- aid_variation_across_units %>%
  filter(unique_aid_values != 1) %>%
  nrow()

print(paste("Number of municipalities that have multiple aid values:",
            non_unique_aid_count))
```

```
## [1] "Number of municipalities that have multiple aid values: 0"
```

In the above table we can see that when we group our data by year and municipality, logged military aid value does not change. We can see this by the fact that “non\_unique\_aid\_count” is 0 which indicates that in our table above there are no values that have more than 1 unique value from this groupation. This shows that within each year all municipalities were getting same amount of aid. However, this does not mean that they were getting same amount of aid through all years. Let’s investigate this further.

```
# Military aid by municipality
aid_variation_across_units_all_time <- bases %>%
  group_by(municipality) %>%
  summarise(unique_aid_values = n_distinct(lrmilnar_col))

aid_variation_across_units_all_time
```

```
## # A tibble: 936 x 2
##   municipality unique_aid_values
##   <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
```

```
# Number of municipalities with non unique aid values
non_18_aid_count <- aid_variation_across_units_all_time %>%
  filter(unique_aid_values != 18) %>%
  nrow()

print(paste("Number of municipalities that have more/less than 18 aid values:",
  non_18_aid_count))
```

```
## [1] "Number of municipalities that have more/less than 18 aid values: 0"
```

```
# Military aid through years for municipality 5113
bases_municipality_5113 <- bases %>%
  filter(municipality == 5113) %>%
  arrange(year)

bases_municipality_5113
```

```
## # A tibble: 18 x 7
##   municipality year paratt bases6 lrmilnar_col bases6xlrmlnar_col lnnewpop
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1     5113 1988      0      0    -3.91      0    -4.93
## 2     5113 1989      0      0    -3.68      0    -4.93
## 3     5113 1990      0      0    -1.90      0    -4.93
## 4     5113 1991      0      0    -2.43      0    -4.93
```

## 5	5113	1992	0	0	-2.32	0	-4.93
## 6	5113	1993	0	0	-2.73	0	-4.93
## 7	5113	1994	0	0	-3.40	0	-4.93
## 8	5113	1995	0	0	-3.49	0	-4.94
## 9	5113	1996	0	0	-2.78	0	-4.94
## 10	5113	1997	0	0	-2.97	0	-4.94
## 11	5113	1998	0	0	-2.53	0	-4.95
## 12	5113	1999	0	0	-1.29	0	-4.95
## 13	5113	2000	0	0	-0.108	0	-4.95
## 14	5113	2001	2	0	-2.98	0	-4.96
## 15	5113	2002	0	0	-1.02	0	-4.96
## 16	5113	2003	0	0	-0.782	0	-4.96
## 17	5113	2004	0	0	-0.669	0	-4.96
## 18	5113	2005	0	0	-0.559	0	-4.97

From the output above we can see that when not grouped by year, each municipality has 18 unique values for aid. This indicates that military aid changed for each municipality 18 times most likely at the start / end of each year as we have seen that when we group by year there is only 1 unique value so we are safe to assume that it does not change within a same year. Further, second table shows all the values for municipality 5113 (randomly chosen as an example) in which we can see exactly what was said before, we have 18 rows (years), each with a different year value and for each year logged military aid value changes and “bases” value stays constant. To conclude, we have shown that the amount of military aid does not change within a same year for each municipality but it does change every year for each municipality.

```
## How many municipalities do we have
number_of_municipalities <- bases %>%
  distinct(municipality) %>%
  nrow()

print(paste("Number of unique municipalities:", number_of_municipalities))
```

```
## [1] "Number of unique municipalities: 936"
```

We can see that in total we have data on 936 unique municipalities. Above we saw that 904 of these have no bases so we can conclude 32 municipalities have bases.

### What do the authors seem to be assuming about how military aid is allocated?

The authors seem to be assuming that the allocation of military aid is consistent (does not change) within each year for each municipality, but it does vary annually. We saw this from the pattern we found above in which each municipality receives a unique military aid value every year, but the value remains constant within that year for all municipalities. This indicates there is an assumption of uniform distribution (unit-constant) of military aid within a year and then changes in following years. This kind of assumption of uniformity in aid within a year could indicate that the allocation process is not accounting for real-time changes in each municipality as they could have different needs 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?

As always there are multiple assumptions that need to be considered when identifying treatment effect. Let's look at some assumptions:

1. Fixed effects assumption - this assumption states that there are both unit-specific and time-specific unobserved heterogeneities that the fixed effects,  $\alpha_i$  for units and  $\gamma_t$  for time period, are controlling for. This means that we account for all differences across units and across time.
2. Random errors - this assumption states that error term  $\epsilon_{it}$  is randomly distributed with mean of 0 and no correlation with our treatment or outcome
3. Constant treatment effect - this assumption states that treatment effect is constant across all units and over time, which is why a single treatment effect  $\tau$  is estimated
4. **Parallel trends** - this assumption is associated with two-way fixed effect models and it states that if there was no treatment the treated and control groups would have followed parallel paths over time, in other words, the selection bias in time 1 is the same as the selection bias in time 0 which means that any pre-existing trends in our outcome variable are similar for both groups and that we can identify treatment effect by deviations from this common trend after the treatment is introduced. Due to this, the authors are assuming that any confounding that happens in the data would be completely constant over time in order to identify the treatment effect of bases6 on military aid.

As always we also need to consider assumptions of

5. SUTVA - as usual, this assumption states there can be no spillover within treatment and that there is only one unit of treatment
6. Random Sampling - this assumption states that data is randomly sampled from the population so that our estimated parameters are unbiased representations of the population parameters

### 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 "CR0")

# Fit Regression using lm_robust
model <- lm_robust(
  paratt ~ bases6 + lrmilnar_col + bases6xlrmlnar_col + lnnewpop +
    factor(municipality) + factor(year),
  data = bases,
  clusters = municipality,
  se_type = "CR0"
)

# Extracting summary
model_summary <- summary(model)

## 2 coefficients not defined because the design matrix is rank deficient

# Extracting coefficients
coefficients <- model_summary$coefficients
```

```
# Removing factor levels from the coefficients to not have 80 pages
non_factor_coefficients <- coefficients[!grepl("factor",
                                             rownames(coefficients)), ]

# Display the coefficients for non-factor variables
print(non_factor_coefficients)
```

```
##              Estimate Std. Error  t value    Pr(>|t|)    CI Lower
## (Intercept)           NA          NA         NA         NA         NA
## bases6                2.8012786  0.27019263  10.367709  6.482785e-24  2.27102442
## lrmilnar_col          -0.1159936  0.05710417  -2.031262  4.251082e-02 -0.22806074
## bases6xlrmlnar_col    0.1503116  0.06008643   2.501590  1.253367e-02  0.03239170
## lnnewpop              0.1178481  0.04523456   2.605266  9.326138e-03  0.02907504
##              CI Upper  DF
## (Intercept)           NA  NA
## bases6                3.331532861  935
## lrmilnar_col          -0.003926366  935
## bases6xlrmlnar_col    0.268231458  935
## lnnewpop              0.206621101  935
```

```
# Extracting R-squared value
cat("Multiple R-squared:", model_summary$r.squared, "\n")
```

```
## Multiple R-squared: 0.2219161
```

```
cat("Adjusted R-squared:", model_summary$adj.r.squared, "\n")
```

```
## Adjusted R-squared: 0.1744884
```

The model we built estimates the effect of U.S. military and narcotics aid on the number of paramilitary attacks in Colombian municipalities, controlling for the presence of a military base, the interaction between the aid and the presence of a military base, and the log of the population, with municipality and year fixed effects.

Looking at the 95% confidence intervals for each variable we can notice the following:

1. For bases6 variable coefficient is 2.80 which indicates that municipalities with a military base have, as we would expect, an increase in paramilitary attacks compared to municipalities without a base, while holding all other variables constant. The 95% confidence interval for bases6 estimate ranges from approximately 2.27 to 3.33. As 0 is not in the interval we can conclude that the effect is statistically significant.
2. For lrmilnar\_col variable coefficient is -0.12 which indicates that an increase in logged U.S. military and narcotics aid leads to a decrease in the number of paramilitary attacks, while holding all other variables constant. The 95% confidence interval for lrmilnar\_col estimate ranges from approximately -0.23 to -0.004. As 0 is not in the interval we can conclude that the effect is statistically significant, however since the upper CI is very close to 0 we can say that the statistical significance is weaker than for other significant variables.
3. For lnnewpop variable coefficient is 0.12 which indicates that as the log of the population increases, the number of paramilitary attacks is expected to increase as well, while holding all other variables constant. The 95% confidence interval for lnnewpop estimate ranges from approximately 0.03 to 0.21. As 0 is not in the interval we can conclude that the effect is statistically significant.



4. Finally, the `bases6xlrmlnar_col` variable is the interaction term and has the coefficient of 0.15 which implies (as expected) that the effect of U.S. aid on paramilitary attacks is different for municipalities with a military base compared to those without. The positive coefficient suggests that the presence of a military base might diminish the negative effect of the aid on paramilitary violence or even make it positive. The 95% confidence interval for `bases6xlrmlnar_col` estimate ranges from approximately 0.03 to 0.27. As 0 is not in the interval we can conclude that the effect is statistically significant.

Further, looking at our R-squared results we can see that only about 22.19% of the variance in paramilitary attacks is explained by the model. This results gets even lower, down to 17.44%, when we calculate the adjusted R-squared that accounts for the number of predictors in the model.

All of these findings suggest that the presence of a military base is associated with an increase in paramilitary attacks in Colombia but the presence of U.S. military aid tends to reduce those attacks. However, the interaction effect (ATE) indicates that the impact of U.S. aid is moderated by the presence of a military base. In other words, we estimate the effect of each additional (logged) quantity of military aid raises the average number of paramilitary attacks in municipalities with military bases by .150. The positive and statistically significant coefficient for the logged population variable indicates that larger municipalities tend to experience more paramilitary attacks. The R-squared values suggest that the model has low to moderate explanatory power.