Problem Set 3

Ghina Al Shdaifat - gha2009 - 003 - Collaborated: Petra Ivanovic(pi2018)&Kendall Brogle (klb9313)

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("/Users/ghinaalshdaifat/Desktop/bases_replication_final.dta")
# Filtering out variables # Selecting specific columns
bases <- bases %>%
select(municipality, year,paratt, bases6, lrmilnar_col, bases6xlrmilnar_col,
       lnnewpop)
bases
```

```
## # A tibble: 16,848 x 7
##    municipality  year paratt bases6 lrmilnar_col bases6xlrmilnar_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
```

```r
# How many observations are in the "no bases group"
noBases_count <- bases %>% filter(bases6 == 0) %>% nrow()
noBases_count
```

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

```r
# Count municipalities with zero
municipalitiesWzero <- bases %>%
  group_by(municipality) %>% # Group by municipality
  summarise(all_zero = all(bases6 == 0)) %>% # Filter with base equal to 0
  filter(all_zero) %>%
  nrow()
# Print the number of municipalities with zero
print(municipalitiesWzero)
```

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

- 16272 values are in the "no bases group," indicating that 16272 of our units do not have bases.
- Filtering out the distinct municipalities indicates a total of 904 municipalities that have no bases.

```r
## How about each of them?
change <- bases %>%
  group_by(municipality) %>%
  summarise(unique_bases_values = n_distinct(bases6))
change
```

```
## # 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
```

```r
countNon1Values <- change %>% filter(unique_bases_values
                                               != 1) %>% nrow()
# Number of municipalities with a base of value of > 1
countNon1Values
```

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

There is a single unique base value for each of our municipalities, indicating no change as time passes. With a count of 0, we can additionally illustrate that there are no municipalities with any number of unique bases values but 1. Therefore, we can conclude that the bases are unit-constant for each individual municipality. In other words, status of military bases do not change as time passes by for any of the municipalities. The bases variable is constant for all municipalities for the time period presented in the data.

```r
# Examining changes in logged military aid across units
aidVar <- 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.
```

```r
aidVar
```

```
## # 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
```

```r
# Number of municipalities with multiple aid values
nonUniqueAidCount <- aidVar %>% filter(unique_aid_values != 1) %>% nrow()
print(paste(nonUniqueAidCount))
```

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

The logged value of military aid remains constant across municipalities. This is evidenced by the "non_unique_aid_count" being zero, suggesting that for each grouping by year and municipality, there is no variation in military aid – all municipalities receive an identical amount of aid within a given year. Nonetheless, this uniformity in aid distribution doesn't imply that the amount of aid received was consistent across different years. Further examination is required to understand how aid distribution varied year over year.

```r
aidVarAll <- bases %>% group_by(municipality) %>%
  summarise(unique_aid_values = n_distinct(lrmilnar_col))

aidVarAll
```

```
## # 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
```

```r
# Number of municipalities that have more or less than 18 aid value
non18 <- aidVarAll %>% filter(unique_aid_values != 18) %>% nrow()
non18
```

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

```r
# Aid for municipality for the year 5113
municipality5113 <- bases %>% filter(municipality == 5113) %>% arrange(year)
municipality5113
```

```
## # A tibble: 18 x 7
##    municipality  year paratt bases6 lrmilnar_col bases6xlrmilnar_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
```

Without considering the year, each municipality experiences 18 distinct instances of changes in military aid. This pattern suggests that the fluctuations in aid likely occur at the beginning or end of each year, given the observation that grouping the data by year results in a single unique value of aid per municipality, implying stability within the same year. Additionally, the second table provides a detailed view for a specific municipality, number 5113, chosen for illustrative purposes. The data for this municipality aligns with the previous observation, displaying 18 separate entries, each corresponding to a different year, and showing variations in the logged military aid value, while the "bases" value remains unchanged. Overall, the analysis confirms that the quantity of military aid allocated to each municipality remains consistent throughout any given year but varies from year to year.

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

```
## [1] 936
```

The dataset includes information on 936 distinct municipalities. It was determined that 904 of these municipalities do not have any military bases, hence, there are bases present in 32 of the municipalities.

The researchers appear to be working under the presumption that the distribution of military aid to each municipality remains the same throughout a given year, but shifts from one year to the next. This inference is drawn from the identified trend where each municipality is allocated a distinct value of military aid that does not vary within that particular year, although it does so across different years. This indicates there is a notion that the military aid is evenly distributed (constant for each unit) during a single year, followed by adjustments in subsequent years. Such an assumption of annual uniformity suggests that the process of allocating aid might not take into account the varying needs that could emerge within municipalities over the course of 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?

The authors' methodology includes the use of fixed effects in their statistical model, suggesting they believe that by incorporating these effects, they are adjusting for all immutable characteristics of the municipalities that might also impact paramilitary violence. These characteristics are constant within each municipality over time, which helps separate the impact of U.S. aid from other variables.

Regarding parallel trends, the assumption is that in the absence of changes in U.S. aid, the municipalities with military bases would exhibit trends in paramilitary violence similar to those without bases. If this assumption holds, it implies that observed differences in trends post-aid allocation are attributable to the aid.

The authors also presume that the error term (ϵ_{it}) captures all the unobserved factors that could influence paramilitary attacks, such as economic conditions. This presumes these unobserved factors are not correlated with the military aid, suggesting that the results should be free from bias related to these unmeasured variables.

Furthermore, the study assumes the principle of SUTVA (Stable Unit Treatment Value Assumption) and independence of treatment assignment. This means they suppose that an increase in military aid in one municipality doesn't affect paramilitary violence in another, indicating no spillover effects. It also implies that the allocation of military aid is not based on any characteristics of the municipalities that simultaneously affect their levels of paramilitary violence, such as a heightened need for aid. Essentially, they are considering the allocation of military bases and the consequent aid to be independent of the outcomes they are measuring.

They also assumed controlled treatment effect - (aid) is constant between units and time.

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

# Fit the linear model with robust standard errors
fitted_model <- lm_robust(
  paratt ~ bases6 + lrmilnar_col + bases6xlrmilnar_col + lnnewpop +
    factor(municipality) + factor(year),
  data = bases,
  clusters = municipality,
  se_type = "CR0"
)

# Summarize the fitted model
summary_fitted <- summary(fitted_model)
```

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

```r
# Retrieve the model's coefficients
coefficients_fitted <- summary_fitted$coefficients

# Filter out factor levels from the coefficients to shorten the output
main_coefficients <- coefficients_fitted[!grepl("factor",
                                               rownames(coefficients_fitted)), ]

# Output the main coefficients
print(main_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
## bases6xlrmilnar_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
## bases6xlrmilnar_col  0.268231458 935
## lnnewpop             0.206621101 935
```

```r
# Report the R-squared values
cat("Computed R-squared:", summary_fitted$r.squared, "\n")
```

```
## Computed R-squared: 0.2219161
```

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

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

The coefficient for the variable indicating the presence of military bases (bases6) stands at roughly 2.801, with a standard error near 0.270. This reflects a positive association between military bases and the dependent variable, which in this context is the frequency of paramilitary attacks. The statistical significance of this variable is confirmed by a t-value of 10.367 and a p-value well below 0.001.

Regarding the variable representing logged military aid (lrmilnar_col), the coefficient is near -0.115 with a standard error around 0.057, pointing to a negative correlation with the dependent variable. The reduction in paramilitary attacks with increased military aid is statistically significant, as the p-value falls below 0.05.

For the interaction term between the presence of bases and logged military aid (bases6xlrmilnar_col), the coefficient is close to 0.150 with a standard error of 0.060. This positive coefficient infers that the impact of military aid on paramilitary attacks is contingent upon whether a municipality has a military base, with the effect being statistically significant (p-value under 0.05).

The logged population size (lnnewpop) yields a coefficient of around 0.118 and a standard error of 0.045, suggesting that an increase in population size correlates with a rise in paramilitary attacks, with statistical significance indicated by a p-value less than 0.01.

As for the model's explanatory power, the multiple R-squared value is approximately 0.221, meaning the model accounts for about 22.1% of the variation in the dependent variable. After adjusting for the number of predictors, the adjusted R-squared value is about 0.174, signifying that roughly 17.4% of the variation is explained.

Each coefficient is accompanied by a 95% confidence interval, delineating the range within which we can be 95% certain the actual parameter value for the population is found. Additionally, degrees of freedom, listed as 935 for each coefficient, are crucial for the computation of both the t-value and the subsequent p-value.

Overall, the analysis infers that military bases correlate with a rise in paramilitary attacks, that increased military aid is linked to a decrease in such attacks, but this link is altered when a base is present. Furthermore, the model illustrates a correlation between higher population sizes and an uptick in paramilitary attacks. The R-squared values indicate that while the model sheds light on some of the variation in the dependent variable, there remains a notable portion that is unaccounted for.