

Untitled

2024-08-01

To show that the parallel trends assumption is satisfied, we need to prove that:

$$\mathbb{E}[Y_{i1}(0) - Y_{i0}(0) \mid D_i = 1] = \mathbb{E}[Y_{i1}(0) - Y_{i0}(0) \mid D_i = 0].$$

The potential outcomes under control are given by:

$$Y_{i0}(0) = \delta_0 + u_i + \epsilon_{i0},$$

$$Y_{i1}(0) = \delta_1 + u_i + \epsilon_{i1}.$$

For a given unit i , the difference in potential outcomes under control between $t = 1$ and $t = 0$ is:

$$Y_{i1}(0) - Y_{i0}(0) = (\delta_1 + u_i + \epsilon_{i1}) - (\delta_0 + u_i + \epsilon_{i0}),$$

$$Y_{i1}(0) - Y_{i0}(0) = \delta_1 - \delta_0 + (\epsilon_{i1} - \epsilon_{i0}).$$

Next, we take the expectation of this difference conditional on D_i .

For $D_i = 1$:

$$\mathbb{E}[Y_{i1}(0) - Y_{i0}(0) \mid D_i = 1] = \mathbb{E}[\delta_1 - \delta_0 + (\epsilon_{i1} - \epsilon_{i0}) \mid D_i = 1].$$

Given $\mathbb{E}[\epsilon_{i0} \mid D_i = 1] = \eta_i(1)$ and $\mathbb{E}[\epsilon_{i1} \mid D_i = 1] = \eta_i(1)$, we have:

$$\mathbb{E}[\epsilon_{i1} - \epsilon_{i0} \mid D_i = 1] = \mathbb{E}[\epsilon_{i1} \mid D_i = 1] - \mathbb{E}[\epsilon_{i0} \mid D_i = 1] = \eta_i(1) - \eta_i(1) = 0.$$

Thus,

$$\mathbb{E}[Y_{i1}(0) - Y_{i0}(0) \mid D_i = 1] = \delta_1 - \delta_0.$$

For $D_i = 0$:

$$\mathbb{E}[Y_{i1}(0) - Y_{i0}(0) \mid D_i = 0] = \mathbb{E}[\delta_1 - \delta_0 + (\epsilon_{i1} - \epsilon_{i0}) \mid D_i = 0].$$

Given $\mathbb{E}[\epsilon_{i0} \mid D_i = 0] = \eta_i(0)$ and $\mathbb{E}[\epsilon_{i1} \mid D_i = 0] = \eta_i(0)$, we have:

$$\mathbb{E}[\epsilon_{i1} - \epsilon_{i0} \mid D_i = 0] = \mathbb{E}[\epsilon_{i1} \mid D_i = 0] - \mathbb{E}[\epsilon_{i0} \mid D_i = 0] = \eta_i(0) - \eta_i(0) = 0.$$

Thus,

$$\mathbb{E}[Y_{i1}(0) - Y_{i0}(0) \mid D_i = 0] = \delta_1 - \delta_0.$$

##1.2 First, let's break down the expectation of the estimator:

$$\begin{aligned} \mathbb{E}[\hat{\tau}] &= \frac{1}{n_t} \mathbb{E} \left[\sum_{i=1}^n (Y_{i1} - Y_{i0}) D_i \right] - \frac{1}{n_c} \mathbb{E} \left[\sum_{i=1}^n (Y_{i1} - Y_{i0}) (1 - D_i) \right] \\ &= \frac{1}{n_t} \sum \mathbb{E}[\mathbb{E}[(\tau_i D_i + \delta_1 - \delta_0 + \epsilon_{i1} - \epsilon_{i0}) D_i] \mid D_i] - \frac{1}{n_c} \sum \mathbb{E}[\mathbb{E}[(\tau_i D_i + \delta_1 - \delta_0 + \epsilon_{i1} - \epsilon_{i0}) (1 - D_i)] \mid D_i] \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n_t} \sum \mathbb{E}[\tau_i + \delta_1 - \delta_0 + \epsilon_{i1} - \epsilon_{i0}] \Pr(D_i = 1) - \frac{1}{n_c} \sum \mathbb{E}[\delta_1 - \delta_0 + \epsilon_{i1} - \epsilon_{i0}] \Pr(D_i = 0) \\
&= \frac{1}{n_t} \sum \mathbb{E}[\tau_i + \delta_1 - \delta_0 + \epsilon_{i1} - \epsilon_{i0}] \frac{n_t}{n} - \frac{1}{n_c} \sum \mathbb{E}[\delta_1 - \delta_0 + \epsilon_{i1} - \epsilon_{i0}] \frac{n_c}{n} \\
&= \frac{1}{n} \sum \mathbb{E}[\tau_i + \delta_1 - \delta_0 + \epsilon_{i1} - \epsilon_{i0}] - \frac{1}{n} \sum \mathbb{E}[\delta_1 - \delta_0 + \epsilon_{i1} - \epsilon_{i0}] \\
&= \frac{1}{n} \sum \tau_i + \delta_1 - \delta_0 - \frac{1}{n} \sum \delta_1 - \delta_0 = \frac{1}{n} \sum \tau_i
\end{aligned}$$

So we've shown that the expectation of the estimator equals the ATE:

$$\mathbb{E}[\hat{\tau}] = \tau$$

and thus the estimator is unbiased for the ATE in this setting.

##1.3

$$\begin{aligned}
\mathbb{E}[\hat{\tau}_t = 1] &= \frac{1}{n_t} \sum_{i=1}^n \mathbb{E}[Y_{i1} D_i] - \frac{1}{n_c} \sum_{i=1}^n \mathbb{E}[Y_{i1} (1 - D_i)] \\
&= \frac{1}{n_t} \sum_{i=1}^n \mathbb{E}[\mathbb{E}[Y_{i1} D_i | D_i]] - \frac{1}{n_c} \sum_{i=1}^n \mathbb{E}[\mathbb{E}[Y_{i1} (1 - D_i) | D_i]] \\
&= \frac{1}{n_t} \sum_{i=1}^n \mathbb{E}[Y_{i1} D_i | D_i = 1] \cdot P(D_i = 1) - \frac{1}{n_c} \sum_{i=1}^n \mathbb{E}[Y_{i1} (1 - D_i) | D_i = 0] \cdot P(D_i = 0) \\
&= \frac{1}{n_t} \sum_{i=1}^n \mathbb{E}[Y_{i1} D_i | D_i = 1] \cdot \frac{n_t}{n} - \frac{1}{n_c} \sum_{i=1}^n \mathbb{E}[Y_{i1} (1 - D_i) | D_i = 0] \cdot \frac{n_c}{n} \\
&= \frac{1}{n} \left(\sum_{i=1}^n \mathbb{E}[Y_{i1} D_i | D_i = 1] - \mathbb{E}[Y_{i1} (1 - D_i) | D_i = 0] \right) \\
&= \frac{1}{n} \left(\sum_{i=1}^n \mathbb{E}[\tau_i] + \mu_i(1) - \mu_i(0) \right) \\
Bias &= \frac{1}{n} \left(\sum_{i=1}^n \mu_i(1) - \mu_i(0) \right)
\end{aligned}$$

##2.1

$$\begin{aligned}
\mathbb{E}[\hat{\tau}_{Wald}] &= \mathbb{E} \left[\frac{1}{p} \left(\frac{1}{n_t} \sum_{i=1}^n Y_i Z_i - \frac{1}{n_c} \sum_{i=1}^n Y_i (1 - Z_i) \right) \right] \\
&= \frac{1}{p} \left(\frac{1}{n_t} \sum_{i=1}^n \mathbb{E}[Y_i Z_i] - \frac{n_c}{n} \sum_{i=1}^n \mathbb{E}[Y_i (1 - Z_i)] \right)
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{p} \left(\frac{1}{n_t} \sum_{i=1}^n \mathbb{E}[Y_i | Z_i = 1] P(Z_i = 1) - \frac{1}{n_c} \sum_{i=1}^n \mathbb{E}[Y_i | Z_i = 0] P(Z_i = 0) \right) \\
&= \frac{1}{p} \left(\frac{1}{n_t} \sum_{i=1}^n \mathbb{E}[Y_i | Z_i = 1] \frac{n_t}{n} - \frac{1}{n_c} \sum_{i=1}^n \mathbb{E}[Y_i | Z_i = 0] \frac{n_c}{n} \right) \\
&= \frac{1}{np} \left(\sum_{i=1}^n \mathbb{E}[Y_i | Z_i = 1] - \sum_{i=1}^n \mathbb{E}[Y_i | Z_i = 0] \right) \\
&= \frac{1}{p} (\mathbb{E}[\tau_{ITT}]) \\
&= \frac{\tau_{ITT}}{p} = \text{LATE}
\end{aligned}$$

##3

```
library(haven)
library(estimatr)
ohie <- read_dta("OHIE.dta")
model_bp <- lm_robust(tab2bp_hyper ~ treatment, data = ohie)
summary(model_bp)
```

```
##
## Call:
## lm_robust(formula = tab2bp_hyper ~ treatment, data = ohie)
##
## Standard error type: HC2
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper DF
## (Intercept)   0.1591    0.004795 33.1816 3.215e-231 0.14971 0.16851 12186
## treatment    -0.0016    0.006621 -0.2417 8.090e-01 -0.01458 0.01138 12186
##
## Multiple R-squared:  4.796e-06 , Adjusted R-squared:  -7.727e-05
## F-statistic: 0.05842 on 1 and 12186 DF, p-value: 0.809
```

```
model_depression <- lm_robust(tab2phqtot_high ~ treatment, data = ohie)
summary(model_depression)
```

```
##
## Call:
## lm_robust(formula = tab2phqtot_high ~ treatment, data = ohie)
##
## Standard error type: HC2
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper DF
## (Intercept)   0.30367    0.006034 50.327 0.000e+00 0.29184 0.31549 12159
## treatment    -0.03493    0.008207 -4.257 2.091e-05 -0.05102 -0.01885 12159
##
## Multiple R-squared:  0.001493 , Adjusted R-squared:  0.001411
## F-statistic: 18.12 on 1 and 12159 DF, p-value: 2.091e-05
```

```
model_needmet <- lm_robust(tab5_needmet_med_inp ~ treatment, data = ohie)
summary(model_needmet)
```

```
##
## Call:
## lm_robust(formula = tab5_needmet_med_inp ~ treatment, data = ohie)
##
## Standard error type: HC2
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper    DF
## (Intercept)  0.61241   0.006378   96.02 0.00e+00  0.59990   0.6249 12214
## treatment    0.03446   0.008746    3.94 8.19e-05  0.01732   0.0516 12214
##
## Multiple R-squared:  0.001272 , Adjusted R-squared:  0.00119
## F-statistic: 15.52 on 1 and 12214 DF,  p-value: 8.19e-05
```

```
model_catastrophic <- lm_robust(tab4_catastrophic_exp_inp ~ treatment, data = ohie)
summary(model_catastrophic)
```

```
##
## Call:
## lm_robust(formula = tab4_catastrophic_exp_inp ~ treatment, data = ohie)
##
## Standard error type: HC2
##
## Coefficients:
##           Estimate Std. Error t value  Pr(>|t|) CI Lower  CI Upper    DF
## (Intercept)  0.05382   0.003003  17.923 6.769e-71  0.04794  0.059711 11793
## treatment   -0.01527   0.003879   -3.936 8.336e-05 -0.02287 -0.007665 11793
##
## Multiple R-squared:  0.001329 , Adjusted R-squared:  0.001245
## F-statistic: 15.49 on 1 and 11793 DF,  p-value: 8.336e-05
```

Treatment can decrease blood pressure level (ITT of -0.0016). However this is not significant since CI includes 0. Treatment can decrease depression, with an ITT of -0.035, and a p-value less than 0.05, CI shows significance. Treatment can decrease catastrophic medical expenditure, with an ITT of -0.01527, and CI shows significance. Treatment can also increase need met, indicating that more people think that they get treated everytime when they need it, illustrated by ITT of 0.034 and a CI showing significance.

##3.2

```
model_bp_2 <- lm_robust(tab2bp_hyper ~ ohp_all_ever_admin, data = ohie)
model_depression_2 <- lm_robust(tab2phqtot_high ~ ohp_all_ever_admin, data = ohie)
model_needmet_2 <- lm_robust(tab5_needmet_med_inp ~ ohp_all_ever_admin, data = ohie)
model_catastrophic_2 <- lm_robust(tab4_catastrophic_exp_inp ~ ohp_all_ever_admin, data = ohie)
summary(model_bp_2)
```

```
##
## Call:
## lm_robust(formula = tab2bp_hyper ~ ohp_all_ever_admin, data = ohie)
##
## Standard error type: HC2
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|) CI Lower  CI Upper
```

```
## (Intercept)          0.16345    0.003966  41.212  0.00000  0.15567  0.171220
## ohp_all_ever_admin -0.01805    0.007162  -2.521  0.01173 -0.03209 -0.004014
##                      DF
## (Intercept)          12186
## ohp_all_ever_admin  12186
##
## Multiple R-squared:  0.0005003 , Adjusted R-squared:  0.0004183
## F-statistic: 6.354 on 1 and 12186 DF,  p-value: 0.01173
```

```
summary(model_depression_2)
```

```
##
## Call:
## lm_robust(formula = tab2phqtot_high ~ ohp_all_ever_admin, data = ohie)
##
## Standard error type:  HC2
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper
## (Intercept)    0.27129   0.004773  56.833 0.000e+00  0.26194  0.28065
## ohp_all_ever_admin 0.04932   0.009237   5.339 9.516e-08  0.03121  0.06742
##                      DF
## (Intercept)    12159
## ohp_all_ever_admin 12159
##
## Multiple R-squared:  0.002438 , Adjusted R-squared:  0.002356
## F-statistic: 28.5 on 1 and 12159 DF,  p-value: 9.516e-08
```

```
summary(model_needmet_2)
```

```
##
## Call:
## lm_robust(formula = tab5_needmet_med_inp ~ ohp_all_ever_admin,
##           data = ohie)
##
## Standard error type:  HC2
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper
## (Intercept)    0.61281   0.005220 117.399 0.000e+00  0.60258  0.62305
## ohp_all_ever_admin 0.06127   0.009482   6.461 1.078e-10  0.04268  0.07985
##                      DF
## (Intercept)    12214
## ohp_all_ever_admin 12214
##
## Multiple R-squared:  0.003297 , Adjusted R-squared:  0.003216
## F-statistic: 41.75 on 1 and 12214 DF,  p-value: 1.078e-10
```

```
summary(model_catastrophic_2)
```

```
##
## Call:
## lm_robust(formula = tab4_catastrophic_exp_inp ~ ohp_all_ever_admin,
##           data = ohie)
##
## Standard error type:  HC2
```

```
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper
## (Intercept)    0.04894   0.002351  20.812 1.658e-94  0.04433  0.053546
## ohp_all_ever_admin -0.01073   0.004052  -2.647 8.128e-03 -0.01867 -0.002784
##              DF
## (Intercept)    11793
## ohp_all_ever_admin 11793
##
## Multiple R-squared:  0.0005371 , Adjusted R-squared:  0.0004523
## F-statistic: 7.007 on 1 and 11793 DF,  p-value: 0.008128
```

By looking at CIs, everything now are significant. With a decrease in blood pressure (-0.018), decrease in catastrophic medical expenditure (-0.011), and a positive impact in need met (0.06) as expected. However, now the depression level increased in this naive regression (0.049) There might be unobserved confounders between ohp_all_ever_admin and different outcomes. ##3.3

```
compliance_selected <- sum(ohie$treatment == 1 & ohie$ohp_all_ever_admin == 1) / sum(ohie$treatment == 1)
compliance_non_selected <- sum(ohie$treatment == 0 & ohie$ohp_all_ever_admin == 0) / sum(ohie$treatment == 0)
print(c(compliance_selected, compliance_non_selected))
```

```
## [1] 0.3818357 0.8545455
```

```
IV_effect <- lm_robust(ohp_all_ever_admin ~ treatment, data = ohie)
summary(IV_effect)
```

```
##
## Call:
## lm_robust(formula = ohp_all_ever_admin ~ treatment, data = ohie)
##
## Standard error type: HC2
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper    DF
## (Intercept)    0.1455   0.003467  41.95      0  0.1387   0.1523 20743
## treatment      0.2364   0.005891  40.12      0  0.2248   0.2479 20743
##
## Multiple R-squared:  0.07189 , Adjusted R-squared:  0.07184
## F-statistic: 1610 on 1 and 20743 DF,  p-value: < 2.2e-16
```

The regression shows high F statistic, extremely low p-value, with a CI showing significance, indicating a strong explainability on prediction of compliance rate. ##3.4

```
bp_iv <- iv_robust(tab2bp_hyper ~ ohp_all_ever_admin | treatment, data = ohie)
summary(bp_iv)
```

```
##
## Call:
## iv_robust(formula = tab2bp_hyper ~ ohp_all_ever_admin | treatment,
##           data = ohie)
##
## Standard error type: HC2
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper
## (Intercept)    0.1601   0.008181 19.5670 5.714e-84  0.14404  0.17611
## ohp_all_ever_admin -0.0063   0.026059 -0.2417 8.090e-01 -0.05738  0.04478
```

```

##                                DF
## (Intercept)                   12186
## ohp_all_ever_admin            12186
##
## Multiple R-squared:  0.0002882 , Adjusted R-squared:  0.0002062
## F-statistic: 0.05844 on 1 and 12186 DF,  p-value: 0.809

depression_iv <- iv_robust(tab2phqtot_high ~ ohp_all_ever_admin | treatment, data = ohie)
summary(depression_iv)

##
## Call:
## iv_robust(formula = tab2phqtot_high ~ ohp_all_ever_admin | treatment,
##           data = ohie)
##
## Standard error type:  HC2
##
## Coefficients:
##              Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper
## (Intercept)      0.3248   0.01039  31.256 2.383e-206  0.3045  0.3452
## ohp_all_ever_admin -0.1376   0.03286  -4.188 2.838e-05 -0.2020 -0.0732
##              DF
## (Intercept)      12159
## ohp_all_ever_admin 12159
##
## Multiple R-squared:  -0.03258 , Adjusted R-squared:  -0.03267
## F-statistic: 17.54 on 1 and 12159 DF,  p-value: 2.838e-05

cata_iv <- iv_robust(tab4_catastrophic_exp_inp ~ ohp_all_ever_admin | treatment, data = ohie)
summary(cata_iv)

##
## Call:
## iv_robust(formula = tab4_catastrophic_exp_inp ~ ohp_all_ever_admin |
##           treatment, data = ohie)
##
## Standard error type:  HC2
##
## Coefficients:
##              Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper
## (Intercept)      0.06314   0.00508  12.429 3.030e-35  0.05319  0.07310
## ohp_all_ever_admin -0.06036   0.01543  -3.913 9.163e-05 -0.09060 -0.03012
##              DF
## (Intercept)      11793
## ohp_all_ever_admin 11793
##
## Multiple R-squared:  -0.01096 , Adjusted R-squared:  -0.01105
## F-statistic: 15.31 on 1 and 11793 DF,  p-value: 9.163e-05

needmet_iv <- iv_robust(tab5_needmet_med_inp ~ ohp_all_ever_admin | treatment, data = ohie)
summary(needmet_iv)

##
## Call:
## iv_robust(formula = tab5_needmet_med_inp ~ ohp_all_ever_admin |
##           treatment, data = ohie)

```

```
##
## Standard error type: HC2
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper
## (Intercept)      0.5915   0.01086  54.448 0.000e+00  0.57022  0.6128
## ohp_all_ever_admin 0.1355   0.03442   3.935 8.354e-05  0.06798  0.2029
##              DF
## (Intercept)      12214
## ohp_all_ever_admin 12214
##
## Multiple R-squared:  -0.001537 , Adjusted R-squared:  -0.001619
## F-statistic: 15.49 on 1 and 12214 DF,  p-value: 8.354e-05
```

Blood pressure is still not significant. while the result in 3.2 shows significance, IV might mitigated some biases. Decrease in depression is larger in IV estimation than previous linear regression in 3.1, while in 3.2, naive regression even shows positive impact on depression. For IV estimation, it shows significance according to CI. This suggests that IV dealt with some biases and it is more accurate. Decrease in catastrophic expenditure is larger in magnitude than previous linear regressions, also showing significance according to CI. Again, there is a larger effect on increasing need met metric than previous estimations and this effect is still statistically significant. This again illustrates the capability of IV estimation in figuring out the true relationship.

##3.5 Exclusion Restriction: The instrument used must affect the outcomes only through Medicaid enrollment and not through any other pathway. This means that the instrument should not have a direct effect on the outcomes, other than through its effect on the treatment variable (Medicaid enrollment). Monotonicity (No Defiers): This assumption implies that the instrument affects all individuals in the same direction Homogeneous Treatment Effect: This assumption implies that the treatment effect of Medicaid enrollment is the same for all individuals in the sample Independence (Random Assignment of the Instrument): The instrument must be as good as randomly assigned, meaning that it is not correlated with any unobserved confounders that could also affect the outcome.

##4.1

```
library(haven)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
data <- read_dta("bases_replication_final.dta")
```

```
sum(data$bases6==0)
```

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

```
time_variation <- data %>%
  group_by(municipality) %>%
  summarize(changes_over_time = var(bases6,na.rm = TRUE))
print(time_variation)
```



```
## # A tibble: 936 x 2
##   municipality changes_over_time
##   <dbl> <dbl>
## 1      5001      0
## 2      5002      0
## 3      5004      0
## 4      5021      0
## 5      5030      0
## 6      5031      0
## 7      5034      0
## 8      5036      0
## 9      5038      0
## 10     5040      0
## # i 926 more rows

time_variation <- data %>%
  group_by(municipality) %>%
  summarize(aid_change = var(lrmilnar_col, na.rm = TRUE))
print(time_variation)
```

```
## # A tibble: 936 x 2
##   municipality aid_change
##   <dbl> <dbl>
## 1      5001    1.41
## 2      5002    1.41
## 3      5004    1.41
## 4      5021    1.41
## 5      5030    1.41
## 6      5031    1.41
## 7      5034    1.41
## 8      5036    1.41
## 9      5038    1.41
## 10     5040    1.41
## # i 926 more rows
```

It seems like the number of bases is constant while military aid is not. Author may assume that in each muni, military aid is not uniform, ie different military aid each year in different muni. Also, author might assume that the military aid in each muni are changing in different years. And it seems like author assumes same change of aid in different munis, ie, same variance of aid in all munis.

##4.2 No Time-Varying Confounders (Conditional Independence): The key assumption is that after controlling for both unit and time fixed effects, any remaining variation in the treatment variable D_{it} is exogenous. This means that there are no unobserved confounders that vary over time and across units. And the assignment of military aid is as good as random. Parallel Trends Assumption: No need to explain Consistency and Linearity: the effect is assumed to be linear. Error term is assumed to be uncorrelated with treatment.

##4.3

```
# Load necessary libraries
library(foreign)
library(plm)
```

```
##
## Attaching package: 'plm'

## The following objects are masked from 'package:dplyr':
##
```

```

##      between, lag, lead
library(lmtest)

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric

# Convert data to panel data format
pdata <- pdata.frame(data, index = c("municipality", "year"))
colnames(pdata)

## [1] "department"          "municipality"
## [3] "year"                 "gueratt"
## [5] "paratt"               "govatt"
## [7] "oil"                  "coca"
## [9] "evercoca"             "rmil_colombia"
## [11] "lrmil_colombia"        "rmil_world_nolac"
## [13] "lrmilwnl"              "rnar_colombia"
## [15] "lrnar_colombia"        "rnar_world_nolac"
## [17] "lrnarwnl"              "rmilnar_col"
## [19] "lrmilnar_col"          "captives"
## [21] "operation_antinarcotics" "PAR_political_homicide"
## [23] "bases6"                "borbases6"
## [25] "meanheight"            "stdheight"
## [27] "bases6xlrmlwnl"        "bases6xlrnarwnl"
## [29] "bases6xlrmlnar_col"    "bases6xlrml_colombia"
## [31] "bases6xlrnar_colombia" "coca00"
## [33] "bases6xlrnworxely"     "bases6xlrnmcxely"
## [35] "bases6xely"            "GUER_political_homicide"
## [37] "urban88"               "lmilexp"
## [39] "sample"                "yearsq"
## [41] "bases6xyear"           "bases6xyearxpost2000"
## [43] "bases6xyearxpost2001i" "bases6xpost2001i"
## [45] "pGUERpres"             "DELpres8890"
## [47] "meanslope"             "borbases6v2"
## [49] "frkid"                 "weaponsops"
## [51] "PAR_tot_hom"           "GUER_tot_hom"
## [53] "meansloplexlrmlnar_col" "meansloplexlwnl"
## [55] "sumtotcas7587v2_pop88" "sumgueratt7587v2_pop88"
## [57] "gaviriauxbases6"       "samperxbases6"
## [59] "pastranaxbases6"       "uribexbases6"
## [61] "gp0003mediboth"        "newpop"
## [63] "lnnewpop"              "paratt_ind"
## [65] "colmilexp_2000_bnusd"   "bases6xlcmlilexp_2000_bnusd"
## [67] "endogbaseonly"         "endogbaseexpxpost2001"

# Run the two-way fixed effects regression
model <- plm(paratt ~ lrmilnar_col + bases6xlrmlnar_col + lnnewpop,
             data = pdata,
             model = "within",
             effect = "twoways")

```

```

# Summary of the model
summary(model)

## Twoways effects Within Model
##
## Call:
## plm(formula = paratt ~ lrmilnar_col + bases6xlrmlnar_col + lnnewpop,
##      data = pdata, effect = "twoways", model = "within")
##
## Unbalanced Panel: n = 936, T = 8-18, N = 16606
##
## Residuals:
##      Min.      1st Qu.      Median      3rd Qu.      Max.
## -3.276524 -0.129748 -0.029909  0.053485  13.828254
##
## Coefficients:
##              Estimate Std. Error t-value Pr(>|t|)
## bases6xlrmlnar_col 0.150312   0.017293  8.6918 < 2.2e-16 ***
## lnnewpop           0.117848   0.034770  3.3894 0.0007022 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    3412.5
## Residual Sum of Squares: 3391.9
## R-Squared:      0.0060294
## Adj. R-Squared: -0.054558
## F-statistic: 47.4696 on 2 and 15651 DF, p-value: < 2.22e-16

# Clustered standard errors
coeftest(model, vcovHC(model, type = "HC1", cluster = "group"))

##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## bases6xlrmlnar_col 0.150312   0.060090  2.5014 0.012379 *
## lnnewpop           0.117848   0.045237  2.6051 0.009193 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Coefficients and standard errors from your model output
coef_bases6xlrmlnar_col <- 0.150312
se_bases6xlrmlnar_col <- 0.060090

coef_lnnewpop <- 0.117848
se_lnnewpop <- 0.045237

# Calculate 95% confidence intervals
ci_bases6xlrmlnar_col <- c(
  coef_bases6xlrmlnar_col - 1.96 * se_bases6xlrmlnar_col,
  coef_bases6xlrmlnar_col + 1.96 * se_bases6xlrmlnar_col
)

ci_lnnewpop <- c(
  coef_lnnewpop - 1.96 * se_lnnewpop,

```

```

    coef_lnnewpop + 1.96 * se_lnnewpop
)

# Print the confidence intervals
cat("95% CI for bases6xlrmlnar_col: [", ci_bases6xlrmlnar_col[1], ", ", ci_bases6xlrmlnar_col[2], "]")

## 95% CI for bases6xlrmlnar_col: [ 0.0325356 , 0.2680884 ]

cat("95% CI for lnnewpop: [", ci_lnnewpop[1], ", ", ci_lnnewpop[2], "]\n")

## 95% CI for lnnewpop: [ 0.02918348 , 0.2065125 ]

```

The interaction term between U.S. military aid and the presence of a base (bases6xlrmlnar_col) has a positive and statistically significant effect on the number of paramilitary attacks. This suggests that in municipalities with a U.S. military base, increased aid is associated with more paramilitary violence. The population size (lnnewpop) also has a positive and statistically significant effect, indicating that larger populations are associated with more paramilitary attacks.