Econometrics II - Assignment 4

Uncensored sloths

30 Jan 2022

Question 1

a) Despite having exactly the same pre-treatment outcomes, it happens to be the case that parallel trends assumption is violated. How is this possible? Explain what it means for parallel trends assumption to be violated, and give an example of how it could be violated.

parallel trend assumption: The selection bias should be constant over time (eta__t - eta__c) it is an identifying assumption so we cannot test for it

Example: employment in brger restaurants - treatment is the minimum income - but we include taxes into the group specific effects (beforehand the same and after increase in Penselvenia)

b) biased estimation (which direction depends on the change in eta) as part of the change in the eta is part of the estimation

Question 2

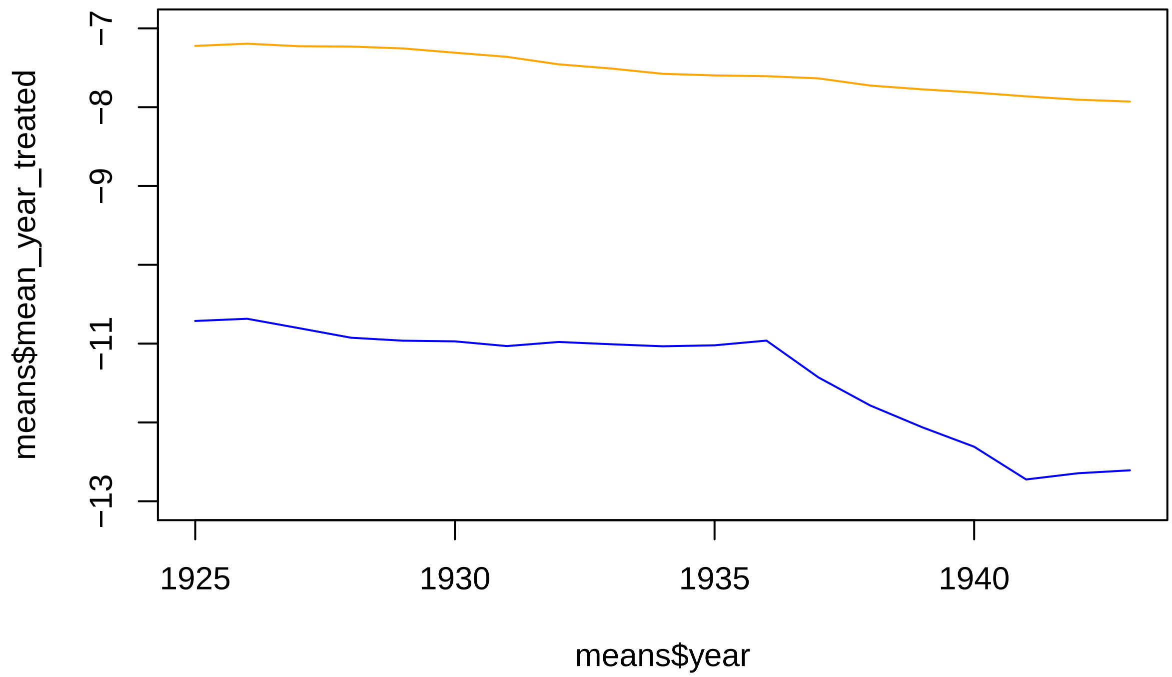
```r
# Load data
data <- read.csv("assignment5.csv")
```

a)

```r
ex1 <- data[data$treated == 1,] %>%
    group_by(year) %>%
          summarise(mean_year_treated = mean(lnm_rate))

ex2 <- data[data$treated == 0,] %>%
    group_by(year) %>%
          summarise(mean_year_untreated = mean(lnm_rate))

means = merge(x=ex1,y=ex2,by="year")

head(means)
```

```
##   year mean_year_treated mean_year_untreated
## 1 1925         -10.71272           -7.222623
## 2 1926         -10.68480           -7.194308
## 3 1927         -10.80431           -7.226814
## 4 1928         -10.92541           -7.231104
## 5 1929         -10.96293           -7.254658
## 6 1930         -10.97147           -7.309226
```

```r
plot(means$year, means$mean_year_treated,type="l",col="blue", ylim=c(-13,-7))
points(means$year, means$mean_year_untreated,type="l",col="orange")
```



Generally, there seems to be a common downward trend which can be explained by advancement of treatment methods, diet and hygiene and is independent from the Sulfa drug. However, in 1937 you can see a substantial drop in mortality due to scarlet fever which stabilizes around 1942, increasing even a little bit (probably due to war and shortage in medicine).

b) Using only data for the years 1936 and 1937, make a table with the mean log mortality rate for treated and control diseases before and after the introduction of sulfa drugs. Use the numbers from the table to calculate the difference-in-differences estimator.

```r
means[means$year == 1936 | means$year == 1937, ]
```

```
##    year mean_year_treated mean_year_untreated
## 12 1936         -10.96190           -7.606973
## 13 1937         -11.42854           -7.634611
```

```r
treated_dif <- means$mean_year_treated[means$year == 1937] -
               means$mean_year_treated[means$year == 1936]
control_dif <- means$mean_year_untreated[means$year == 1937] -
               means$mean_year_untreated[means$year == 1936]
did <- treated_dif - control_dif
```

c) Using only data for the years 1936 and 1937, estimate a difference-in-difference regression. Comment on your results.

```r
data$year_control <- ifelse(data$year == 1937, 1, 0)
data$post <- ifelse(data$year >= 1937, 1, 0)
data$D <- data$post*data$treated
```

```r
m1 <- feols(lnm_rate ~ D | treated + year, data = subset(data, data$year == 1936| data$year == 1937),
            se = 'standard')
summary(m1)
```

```
## OLS estimation, Dep. Var.: lnm_rate
## Observations: 192
## Fixed-effects: treated: 2,  year: 2
## Standard-errors: IID
##    Estimate Std. Error  t value Pr(>|t|)
## D -0.439008   0.218887 -2.00564 0.046329 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.750306     Adj. R2: 0.848869
##                  Within R2: 0.020948
```

d) Using all years, estimate a difference-in-difference regression. To do that, you need to create an indicator variable equal to 1 for the years 1937-1943 and equal to 0 for the years 1925-

1936. What is the interpretation of the difference-in-differences coefficients? What do you conclude about the effect of sulfa drugs on mortality rates?

```r
m2 <- feols(lnm_rate ~ D | treated + year, data,
            se = 'standard')
summary(m2)
```

```
## OLS estimation, Dep. Var.: lnm_rate
## Observations: 1,721
## Fixed-effects: treated: 2,  year: 19
## Standard-errors: IID
##    Estimate Std. Error  t value  Pr(>|t|)
## D -0.866724   0.060468 -14.3336 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.605666     Adj. R2: 0.914713
##                  Within R2: 0.107824
```

As it will be discussed later, we first assume that there is no clustering in sample and treatment.

e) Estimate an event-study specification. Comment on your results.

```r
m3 <- feols(lnm_rate ~ treated*i(year, ref = 1936) | treated + year, data,
 se = 'standard')
```

```
## Variables 'treated', 'year::1925' and 17 others have been removed because of collinearity (see $colli
```

```r
summary(m3)
```

```
## OLS estimation, Dep. Var.: lnm_rate
## Observations: 1,721
## Fixed-effects: treated: 2,  year: 19
## Standard-errors: IID
##                      Estimate Std. Error   t value   Pr(>|t|)
## treated:year::1925 -0.135176   0.190306 -0.710310 4.7761e-01
## treated:year::1926 -0.135567   0.187186 -0.724239 4.6902e-01
## treated:year::1927 -0.222575   0.180546 -1.232787 2.1783e-01
## treated:year::1928 -0.339383   0.177195 -1.915308 5.5623e-02 .
## treated:year::1929 -0.353351   0.175174 -2.017141 4.3839e-02 *
## treated:year::1930 -0.307318   0.175174 -1.754360 7.9551e-02 .
```

```
## treated:year::1931 -0.314916   0.175174 -1.797736 7.2398e-02 .
## treated:year::1932 -0.167430   0.174220 -0.961026 3.3668e-01
## treated:year::1933 -0.144043   0.173300 -0.831174 4.0599e-01
## treated:year::1934 -0.102580   0.173761 -0.590350 5.5504e-01
## treated:year::1935 -0.069522   0.173300 -0.401162 6.8835e-01
## treated:year::1937 -0.439008   0.173300 -2.533218 1.1392e-02 *
## treated:year::1938 -0.704925   0.174240 -4.045723 5.4524e-05 ***
## treated:year::1939 -0.932996   0.173761 -5.369431 8.9992e-08 ***
## treated:year::1940 -1.139170   0.174240 -6.537948 8.2465e-11 ***
## treated:year::1941 -1.505930   0.175259 -8.592617  < 2.2e-16 ***
## treated:year::1942 -1.384778   0.174240 -7.947544 3.4544e-15 ***
## treated:year::1943 -1.323763   0.174739 -7.575677 5.8611e-14 ***
## ... 19 variables were removed because of collinearity (treated, year::1925 and 17 others [full set i
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.593665     Adj. R2: 0.917231
##                   Within R2: 0.142828
```

f) Argue at which level you need to cluster standard errors. Implement your suggested clusterrobust standard errors. Comment on your results.

state level - as there is potential heterogeneity on state level - states with higher scarlet fever rate receive more treatment

```r
m1_clustered <- feols(lnm_rate ~ D | treated + year, data = subset(data, data$year == 1936| data$year ==
            cluster = 'state')
summary(m1_clustered)
```

```
## OLS estimation, Dep. Var.: lnm_rate
## Observations: 192
## Fixed-effects: treated: 2,  year: 2
## Standard-errors: Clustered (state)
##    Estimate Std. Error  t value Pr(>|t|)
## D -0.439008   0.250152 -1.75497 0.082489 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.750306     Adj. R2: 0.848869
##                   Within R2: 0.020948
```

```r
m2_clustered <- feols(lnm_rate ~ D | treated + year, data,
            cluster = 'state')
summary(m2_clustered)
```

```
## OLS estimation, Dep. Var.: lnm_rate
## Observations: 1,721
## Fixed-effects: treated: 2,  year: 19
## Standard-errors: Clustered (state)
##    Estimate Std. Error  t value   Pr(>|t|)
## D -0.866724   0.170408 -5.08616 1.8343e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.605666     Adj. R2: 0.914713
##                   Within R2: 0.107824
```

```r
m3_clustered <- feols(lnm_rate ~ treated*i(year, ref = 1936) | treated + year, data,
 cluster = 'state')
```

```
## Variables 'treated', 'year::1925' and 17 others have been removed because of colli
```

```r
summary(m3_clustered)
```

```
## OLS estimation, Dep. Var.: lnm_rate
## Observations: 1,721
## Fixed-effects: treated: 2,  year: 19
## Standard-errors: Clustered (state)
##                     Estimate Std. Error   t value   Pr(>|t|)
## treated:year::1925 -0.135176   0.148563 -0.909894 3.6518e-01
## treated:year::1926 -0.135567   0.149337 -0.907792 3.6629e-01
## treated:year::1927 -0.222575   0.117451 -1.895042 6.1128e-02 .
## treated:year::1928 -0.339383   0.100636 -3.372398 1.0793e-03 **
## treated:year::1929 -0.353351   0.129361 -2.731515 7.5165e-03 **
## treated:year::1930 -0.307318   0.122704 -2.504539 1.3965e-02 *
## treated:year::1931 -0.314916   0.148109 -2.126252 3.6075e-02 *
## treated:year::1932 -0.167430   0.122456 -1.367261 1.7477e-01
## treated:year::1933 -0.144043   0.139702 -1.031075 3.0512e-01
## treated:year::1934 -0.102580   0.115933 -0.884816 3.7849e-01
## treated:year::1935 -0.069522   0.098989 -0.702314 4.8420e-01
## treated:year::1937 -0.439008   0.250893 -1.749783 8.3385e-02 .
## treated:year::1938 -0.704925   0.239964 -2.937626 4.1502e-03 **
## treated:year::1939 -0.932996   0.230479 -4.048081 1.0529e-04 ***
## treated:year::1940 -1.139170   0.241185 -4.723215 8.0109e-06 ***
## treated:year::1941 -1.505930   0.229506 -6.561604 2.7862e-09 ***
## treated:year::1942 -1.384778   0.237230 -5.837276 7.3536e-08 ***
## treated:year::1943 -1.323763   0.231502 -5.718148 1.2412e-07 ***
## ... 19 variables were removed because of collinearity (treated, year::1925 and 17 others [full set i
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.593665     Adj. R2: 0.917231
##                  Within R2: 0.142828
```

year as we choose turbocolosis as a control based on the argumentation that at that point of time there was no medication against it - treatment level

g) Do a test of whether the prior trends differ between the treated and control groups. What do you conclude?

```r
linearHypothesis(m3_clustered, c("treated:year::1925=0", "treated:year::1926=0", "treated:year::1927=0"
                                  "treated:year::1929=0", "treated:year::1930=0", "treated:year::1931=0
```

```
## Warning: In vcov.fixest(model, complete = FALSE):
##  'complete' is not a valid argument of function vcov.fixest (fyi, some of
## its main arguments are 'vcov' and 'ssc').
```

```
## Linear hypothesis test
##
## Hypothesis:
## treated:year::1925 = 0
## treated:year::1926 = 0
## treated:year::1927 = 0
## treated:year::1928 = 0
## treated:year::1929 = 0
## treated:year::1930 = 0
## treated:year::1931 = 0
## treated:year::1932 = 0
```

```
## treated:year::1933 = 0
## treated:year::1934 = 0
## treated:year::1935 = 0
##
## Model 1: restricted model
## Model 2: lnm_rate ~ treated * i(year, ref = 1936) | treated + year
##
##   Df  Chisq Pr(>Chisq)
## 1
## 2 11 38.547  6.322e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```