Assignment 5

Group name: Foodies with hoodies

Contents

1 Question 1: 2
1.1 (i) . . . 2
1.2 (ii) . . . 2

2 Question 2: 3
2.1 (i) . . . 3
2.2 (ii) . . . 4
2.3 (iii) . . . 5
2.4 (iv) . . . 5
2.5 (v) . . . 5
2.6 (vi) . . . 7
2.7 (vii) . . . 7

```
# load packages
if(!require(pacman)){install.packages("pacman")}

p_load(devtools,tidyverse,dplyr,ggplot2,latex2exp,stargazer, fixest,
 ↪  modelsummary)
```

1 Question 1:

1.1 (i)

I'm not 100% sure about this one, but here is my answer:

Suppose the differences in outcomes between the treatment and the control group is:

$$Y_{g1} - Y_{g0} = (\alpha_1 - \alpha_0) + \delta D_g + (U_{g1} - U_{g0}), \tag{1}$$

in which: δ is the treatment effect.

The parallel trends assumption state that without the intervention of the treatment $(\delta = 0)$, the difference of between the control and treatment group $(\alpha_1 - \alpha_0)$ remain constant over time. Since in this example, they only look at the pre-treatment period, the parallel trends assumption could be violated due to the fact that after the treatment period, the differences in outcomes of control and treatment groups are not constant over time anymore.

When parallel trend is violated, it means that $\alpha_1 - \alpha_0$ changes over time and this means it is no longer to estimate Equation 1 using OLS. If we continue estimating it using OLS, we will have a biased and inconsistent estimator (?).

One example for violation of this assumption could be that the there exists autocorrelation in the treatment group after getting treated. Specifically, the outcome of the next time lag is influenced by the outcome of the previous lag. Before the treatment previous, both the control and the treatment group have the same time trend because both experience no treatment. However, after receiving the treatment,the treated group has a steeper slope in their outcomes and is no longer parallel to the control group.

1.2 (ii)

The difference-in-difference estimator is an OLS estimator Equation 1, which can be written in the form below:

$$Y_{g1} - Y_{g0} = \beta_0 + \delta D_g + U_g. \tag{2}$$

The main problem with applying the OLS estimator in this case is that the estimator will be inconsistent. This is because the estimator treats the time trend β_0 as a constant, however, in this case, the parallel trends assumption is not satisfied and thus β_0 is not constant. This makes the error term changes over time and is correlated with the treatment variable D_g.

MAYBE WE CAN STILL ELABORATE MORE HERE?

2 Question 2:

```
dfData = read.csv("assignment5.csv")
attach(dfData)
```

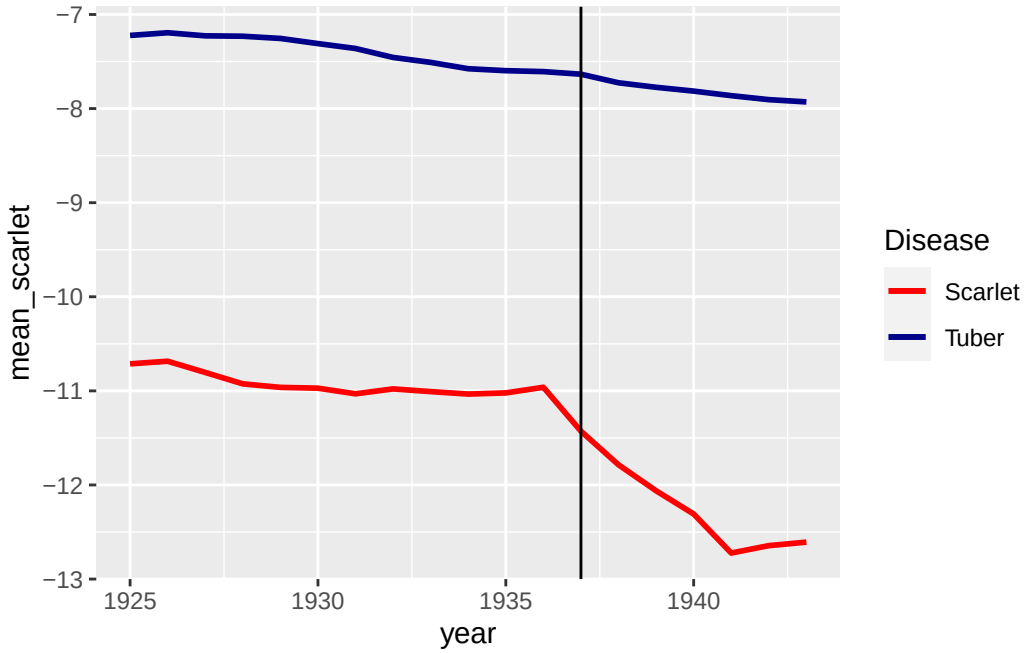
2.1 (i)

```
# Get subgroups and mean per year
df_scarlet = dfData[dfData$treated==1,]
mean_scarlet = df_scarlet %>% group_by(year) %>%
  summarise(mean_rate=mean(lnm_rate),
            .groups = 'drop')
df_tuber = dfData[dfData$treated==0,]
mean_tuber = df_tuber %>% group_by(year) %>%
  summarise(mean_rate=mean(lnm_rate),
            .groups = 'drop')

# reformat into dataframe
df_grouped =
 ↪  data.frame(mean_tuber$year,mean_scarlet$mean_rate,mean_tuber$mean_rate)
names(df_grouped) = c("year","mean_scarlet","mean_tuber")

# Plot
ggplot() +
  geom_line(data=df_grouped,aes(y=mean_scarlet,x=
   ↪  year,colour="Scarlet"),size=1 )+
  geom_line(data=df_grouped,aes(y=mean_tuber,x=
   ↪  year,colour="Tuber"),size=1) +
  scale_color_manual(name = "Disease", values = c("Tuber" = "darkblue",
   ↪  "Scarlet" = "red")) +
```

```
  geom_vline(xintercept = 1937) #add a vertical line indicating
   ↪  treatment year
```



2.2 (ii)

```
# get mean effects
mean_treated_1936 = df_scarlet[df_scarlet$year==1936,] %>%
  summarise(mean_rate=mean(lnm_rate),
            .groups = 'drop')
mean_treated_1937 = df_scarlet[df_scarlet$year==1937,] %>%
  summarise(mean_rate=mean(lnm_rate),
            .groups = 'drop')

mean_control_1936 = df_tuber[df_tuber$year==1936,] %>%
  summarise(mean_rate=mean(lnm_rate),
            .groups = 'drop')
mean_control_1937 = df_tuber[df_tuber$year==1937,] %>%
  summarise(mean_rate=mean(lnm_rate),
            .groups = 'drop')
```

	Before treatment (1936)	After treatment (1937)	Difference
Treatment	-10.96	-11.43	0.47
Control	-7.61	-7.63	0.02
Difference	-3.35	-3.8	0.45

The difference-in-differences estimator is 0.45

HERE MAYBE I ROUNDED IT TOO MUCH, you should recalculate before rounding to the 3rd or 4th decimal probably. Because I hear others got 0.439, which is 0.44.

2.3 (iii)

```
# Create indicator variable
dfData$indicator <- ifelse(dfData$year >=1937, 1, 0)

# Get subdata for the year 1936 and 1937
dfSub = dfData[dfData$year==1936 | dfData$year==1937,]

# DiD model
DiD1 = feols(lnm_rate ~ indicator*treated|year + treated, data = dfSub,
 ↪  se="standard")
```

2.4 (iv)

```
DiD2= feols(lnm_rate ~ indicator*treated| year + treated, data = dfData,
 ↪  se="standard")
msummary(list(DiD1,DiD2), stars = c('*' = .1, '**' = .05, '***' = .01))
```

(For (iii) and (iv), I took out the group-specific and time-specific effect. However, if you want to interpret those you can take out the "|year + treated" part in the model specification)

2.5 (v)

For this question, we can take out the two-way fixed effect of group and year and create an interaction variable of year and treated group, making the year 1936 the reference year and thus normalize its correficients to 0.

	(1)	(2)
indicator × treated	−0.439**	−0.867***
	(0.219)	(0.060)
Num.Obs.	192	1721
R2	0.851	0.916
R2 Adj.	0.849	0.915
R2 Within	0.021	0.108
R2 Within Adj.	0.016	0.107
AIC	442.6	3200.1
BIC	455.6	3314.5
RMSE	0.75	0.61
Std.Errors	IID	IID
FE: year	X	X
FE: treated	X	X

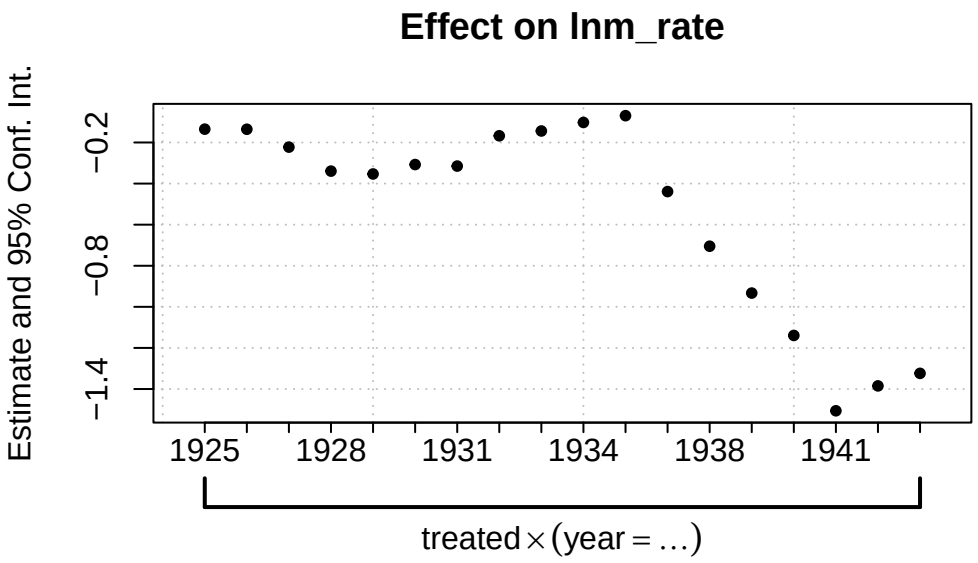
* p < 0.1, ** p < 0.05, *** p < 0.01

```
es <- feols(lnm_rate ~ i(year, treated, ref = 1936)|year+treated, data =
 ↪  dfData)
summary(es)
```

```
OLS estimation, Dep. Var.: lnm_rate
Observations: 1,721
Fixed-effects: year: 19,  treated: 2
Standard-errors: Clustered (year)
                    Estimate Std. Error       t value  Pr(>|t|)
year::1925:treated -0.135176   5.59e-13 -2.420236e+11 < 2.2e-16 ***
year::1926:treated -0.135567   5.58e-13 -2.428443e+11 < 2.2e-16 ***
year::1927:treated -0.222575   5.60e-13 -3.975258e+11 < 2.2e-16 ***
year::1928:treated -0.339383   5.62e-13 -6.038792e+11 < 2.2e-16 ***
year::1929:treated -0.353351   5.59e-13 -6.319386e+11 < 2.2e-16 ***
year::1930:treated -0.307318   5.60e-13 -5.490184e+11 < 2.2e-16 ***
year::1931:treated -0.314916   5.60e-13 -5.626043e+11 < 2.2e-16 ***
year::1932:treated -0.167430   5.60e-13 -2.987636e+11 < 2.2e-16 ***
year::1933:treated -0.144043   5.62e-13 -2.565318e+11 < 2.2e-16 ***
year::1934:treated -0.102580   5.60e-13 -1.831545e+11 < 2.2e-16 ***
year::1935:treated -0.069522   5.62e-13 -1.237185e+11 < 2.2e-16 ***
year::1937:treated -0.439008   5.58e-13 -7.861498e+11 < 2.2e-16 ***
year::1938:treated -0.704925   5.54e-13 -1.272684e+12 < 2.2e-16 ***
year::1939:treated -0.932996   5.51e-13 -1.692804e+12 < 2.2e-16 ***
```

```
year::1940:treated -1.139170   5.48e-13 -2.077248e+12 < 2.2e-16 ***
year::1941:treated -1.505930   5.78e-13 -2.604525e+12 < 2.2e-16 ***
year::1942:treated -1.384778   5.44e-13 -2.545701e+12 < 2.2e-16 ***
year::1943:treated -1.323763   5.86e-13 -2.258105e+12 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.593665     Adj. R2: 0.917231
                 Within R2: 0.142828
```

```r
coefplot(es)
```



2.6 (vi)

We need to cluster standard errors at he fixed effect levels: namely year and treated

2.7 (vii)

For this, we can use the Wald test to check to check on the pretrends from the event study model.

```r
# https://lrberge.github.io/fixest/reference/wald.html
wald(es, keep="year::[1925192619271928192919301931193219331934193519341935]$")
```

```
[1] NA
```

we can use the placebo test, in which we pick fake treatment periods before the actual treatment period and see if there is a significant effect.

```r
# Create fake indicator variables
dfData$D_fake1 <- ifelse(dfData$year >=1928, 1, 0)
dfData$D_fake2 <- ifelse(dfData$year >=1930, 1, 0)
dfData$D_fake3 <- ifelse(dfData$year >=1932, 1, 0)
dfData$D_fake4 <- ifelse(dfData$year >=1934, 1, 0)

# Test fake models
DiD1_fake = feols(lnm_rate ~ D_fake1*treated|year + treated, data =
 ↪  dfData, cluster = "treated^year")
DiD2_fake = feols(lnm_rate ~ D_fake2*treated|year + treated, data =
 ↪  dfData, cluster = "treated^year")
DiD3_fake = feols(lnm_rate ~ D_fake3*treated|year + treated, data =
 ↪  dfData, cluster = "treated^year")
DiD4_fake = feols(lnm_rate ~ D_fake4*treated|year + treated, data =
 ↪  dfData, cluster = "treated^year")
msummary(list(DiD1_fake,DiD2_fake,DiD3_fake,DiD4_fake), stars = c('*' =
 ↪  .1, '**' = .05, '***' = .01))
```

	(1)	(2)	(3)	(4)
D_fake1 × treated	−0.407***			
	(0.092)			
D_fake2 × treated		−0.358***		
		(0.105)		
D_fake3 × treated			−0.386***	
			(0.116)	
D_fake4 × treated				−0.513***
				(0.127)
Num.Obs.	1721	1721	1721	1721
R2	0.907	0.907	0.907	0.909
R2 Adj.	0.905	0.906	0.906	0.908
R2 Within	0.011	0.014	0.020	0.040
R2 Within Adj.	0.011	0.013	0.020	0.039
AIC	3376.8	3372.3	3361.1	3326.8
BIC	3491.3	3486.7	3475.6	3441.2
RMSE	0.64	0.64	0.63	0.63
Std.Errors	by: treated^year	by: treated^year	by: treated^year	by: treated^year
FE: year	X	X	X	X
FE: treated	X	X	X	X

* p < 0.1, ** p < 0.05, *** p < 0.01