

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)	4
2.4	(iv)	4
2.5	(v)	5
2.6	(vi)	5
2.7	(vii)	5

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

p_load(devtools,tidyverse,dplyr,ggplot2,latex2exp,stargazer)
```

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:

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

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:

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

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)
```

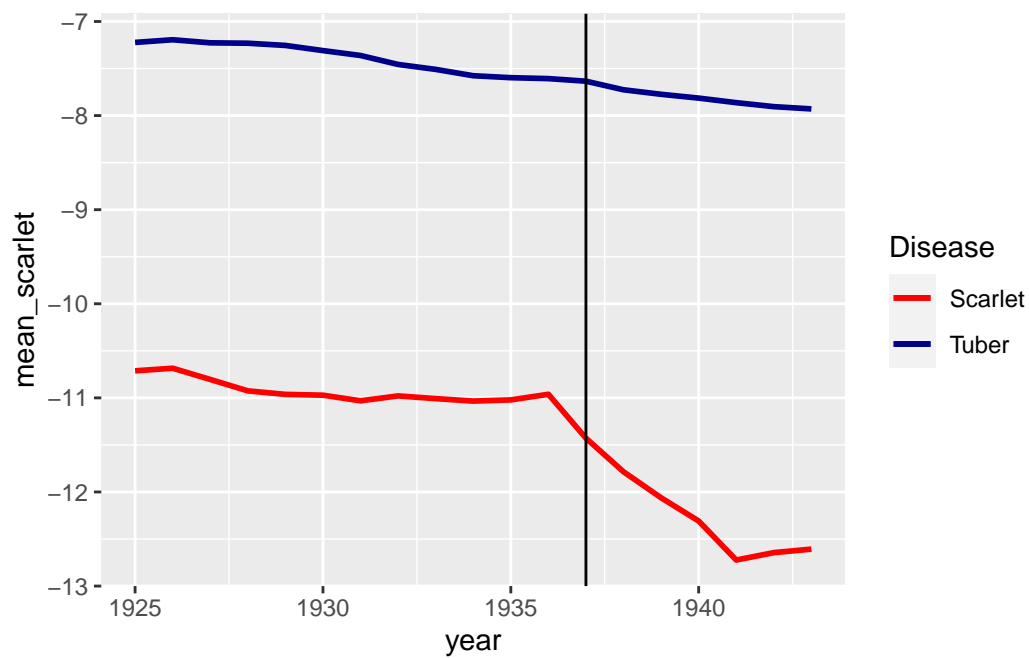
2.1 (i)

```
# 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)

2.3 (iii)

```
#DiD1 = lm()...
```

2.4 (iv)

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

# Get difference of each group
df_grouped$difference = df_grouped$mean_scarlet -df_grouped$mean_tuber
```

```
# Difference-in-difference estimator
DiD2 = lm(difference ~ indicator, data = df_grouped)
stargazer(DiD2, type="text")
```

```
=====
                        Dependent variable:
                        -----
                                difference
-----
indicator                        -0.871***
                                (0.118)

Constant                        -3.546***
                                (0.072)

-----
Observations                      19
R2                               0.761
Adjusted R2                      0.747
Residual Std. Error      0.249 (df = 17)
F Statistic              54.024*** (df = 1; 17)
=====
Note:                *p<0.1; **p<0.05; ***p<0.01
```

2.5 (v)

2.6 (vi)

2.7 (vii)