Impact of Closure Policies on COVID-19 Cases

Team 7: Zhaoyan Zhi, Arkan Chatterjee, Beepa Bose, Sricharan Sridhar

5/7/2022

Contents

Background

The COVID-19 war has not yet been won, with countries still heavily infected, losing lives and creating lasting economic effects. The government has processes and procedures in place to implement policies during an epidemic or a pandemic. In order to successfully tackle the pandemic, it is important to understand different policies, and the impact it has on the population. Decision-makers and curious individuals can leverage this information to stay informed and refocus their efforts to the right channels to deal with / fight the pandemic.

Being an air-bourne disease, known to be notoriously infectious. In such cases, closure and containment policies play a vital role in curbing the spread and decreasing deaths. Our team got together to analyze and identify the best closure policy, and recommend the same to decision-makers in power to implement the same.

Threats to Causal Inference

Selection Bias

This happens when we are taking a particular section of the population and inferring the results for the whole population. In our case, we choose specific time frames to make analysis and the result may only apply to those period. We are assuming that this is not the case for the following analysis.

Omitted Variable Bias

This happens when we are not including a policy that is correlated with the policy that we want to observe and that also impacts the number of cases. In our case, we are particularly selecting the timelines such that there is no other policy change during that timeline.

Simultaneity Error

This happens when a factor impacts the outcome and the outcome also impacts the factor, due to this we will not be able to estimate the correct value of the impact because of the factor on the outcome. In our case, this is not happening.

Measurement Error

This happens when we are not able to measure our factor and we use another to measure our main factor. The error in that measurement process can lead to an incorrect estimate of the impact of the factor on the outcome. In our case this is not an issue since we exactly know the date when the policy was implemented.

Data Exploration

The dataset used here is collected by The Oxford Covid-19 Government Response Tracker (OxCGRT) and give systematic information on several different common policies governments have taken during the 2-year COVID-19 pandemic (2020-01 to 2022-01). This dataset records these policies on a scale to reflect the extent of government action, and aggregates these scores into a suite of policy indices. They have been collecting information like number of cases, policies implemented countrywide.

Specifically speaking, this daily panel dataset shows 21 indicators of government policy response also includes statistics on the number of reported Covid-19 cases and deaths in each country on a daily basis.

The specific data we used in this project are:

Unit of Analysis: Regions like Austria

Outcome variable: number of cases in Austria

Treatment variable: Whether the Closure policies are implemented. We focus on three policies here: Stay Home Policy, School Closing Policy, Cancel Public Event Policy

Time period: (around 2 months for each policy)

2020/04/01-2020/06/01(Stay Home Policy with policy ENDED at 2020/05/01), 2020/05/06-2020/06/30(School Closing Policy with policy STARTED at 2020/06/03), 2021/06/01-2021/08/01(Cancel Public Event Policy with policy ENDED at 2021/07/01)

Important Feature: number of cases daily, policy type, region

Data Source:https://github.com/OxCGRT/covid-policy-tracker/tree/master/data

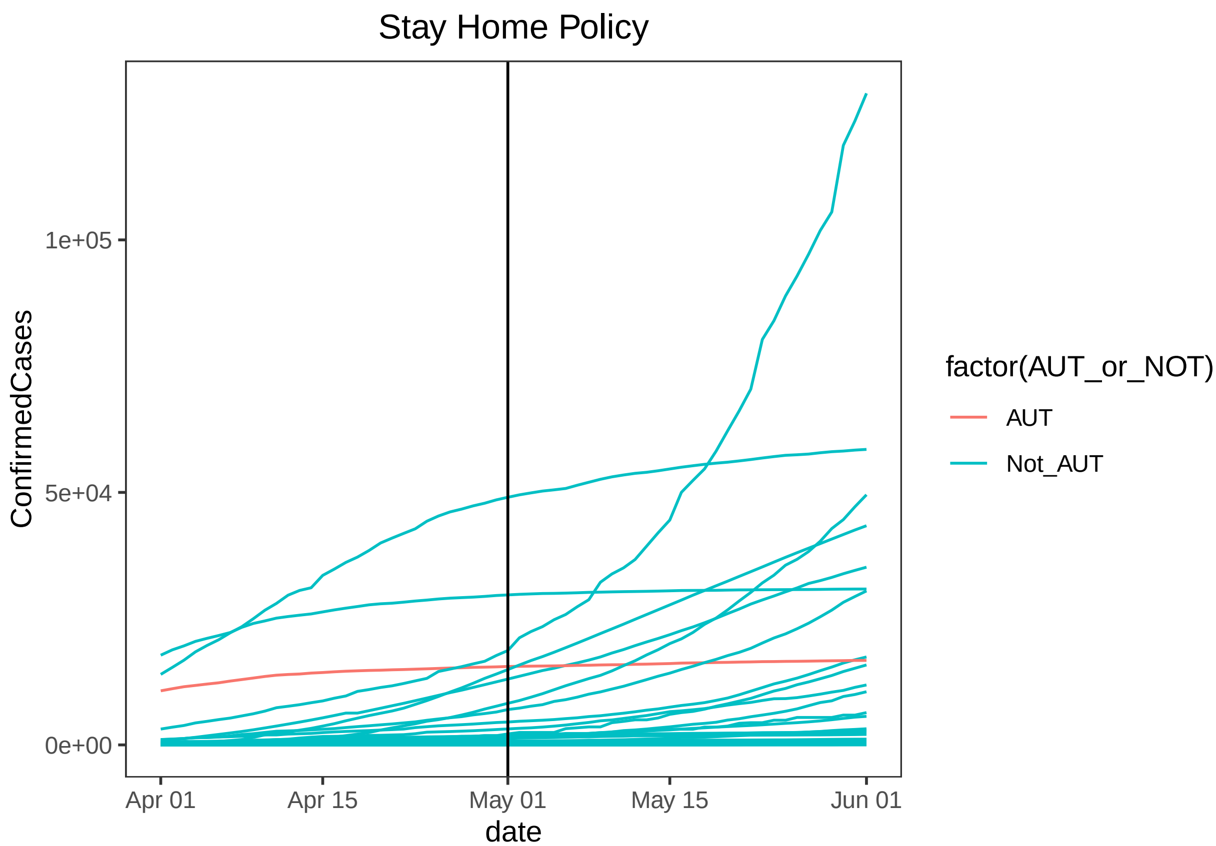
```r
ori_df=read.csv("covid-data.csv")
ori_df["date"]=ymd(ori_df$Date)

e2=filter(ori_df,date>="2020-04-01")
e2=filter(e2,date<="2020-06-01")
# remove regions with high proportion of null value
e2=filter(e2,CountryCode%in%c("BRA","CAN","AUS","CHN")==FALSE)
e2["AUT_or_NOT"]=ifelse(e2$CountryCode%in%c("AUT"),"AUT","Not_AUT")
```

Case Number for Different Regions from 2020 Apr 01 to 2020 Jun 01

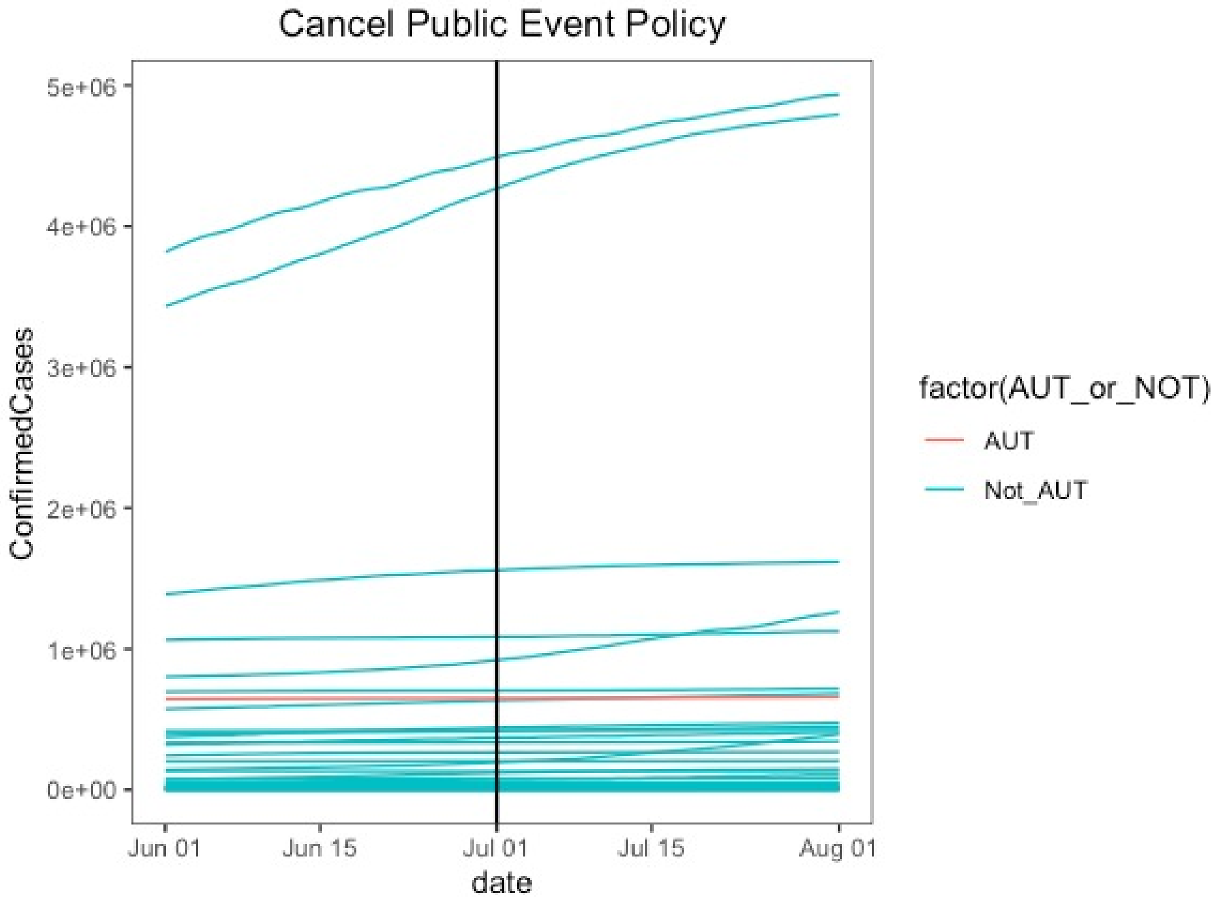
(black vertical line represent the Stay Home police implementation day for AUT)

```
# visualize the trend
ggplot(data=e2,aes(x=date,y=ConfirmedCases,color=factor(AUT_or_NOT),
                   group=CountryCode)) + #alpha=treat
  geom_line() +
  geom_vline(xintercept=as.numeric(as.Date("2020-05-01")),color="black") +
  ggtitle("Stay Home Policy")   +
  theme_bw() + theme(panel.grid=element_blank()) +
  theme(plot.title = element_text(hjust = 0.5))
```



Case Number for Different Regions from 2021 Jun 01 to 2021 Aug 01

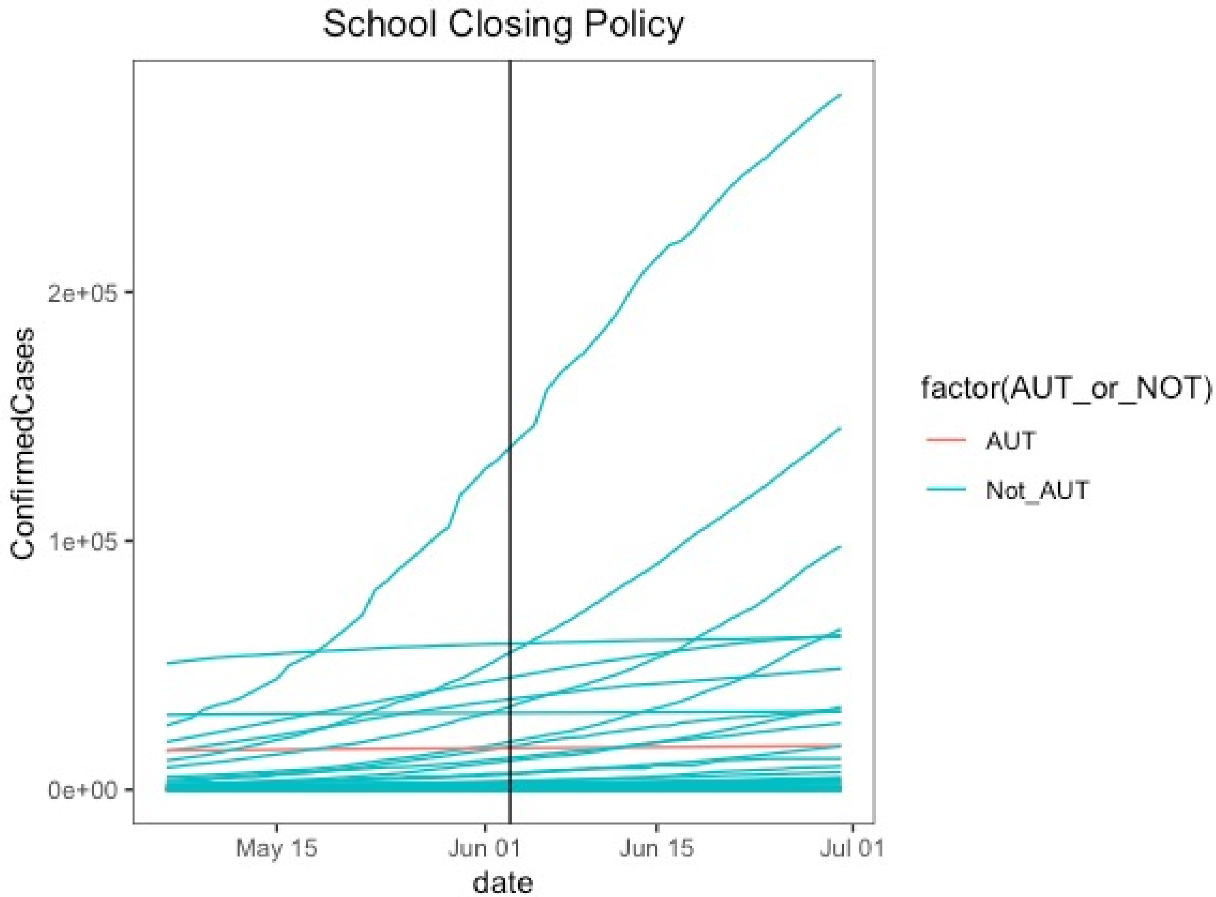
(black vertical line represent the Cancel Public policy implementation day for



AUT)

Case Number for Different Regions from 2020 May 06 to 2020 Jun 30

(black vertical line represent the School policy implementation day for AUT)



Noted: Since the codes for the graph of "School Closing Policy" and "Cancel Public Event Policy" are similar to "Stay Home Policy" except for the time frame, we do not show the codes for them due to the space limiation.

Method

Since it is hard to find a parallel trend for AUT, we decide to generate a synthesized trend as control counterfactual with synthetic control method first and then utilize difference-in-difference method to test the impact of policy.

Synthetic Control Method

First of all, for each policy, we leverage Lasso to determine the regions and their corresponding weights to generate the synthetic trend for both pre-period and post period.

```
### select time frame
e2=filter(ori_df,date>="2020-05-06")
e2=filter(e2,date<="2020-06-30")
e2=filter(e2,CountryCode%in%c("BRA","CAN","AUS","CHN")==FALSE)

# long table to wide table
e2.wide <- e2 %>% pivot_wider(id_cols=c("date"),names_from=c("CountryCode"),
                              values_from="ConfirmedCases")
#View(e2.wide)
e2.wide.train <- subset(e2.wide,date<as.numeric(as.Date("2020-06-03")))

# transform to lasso
e2.wide.train_mm <- model.matrix(`AUT`~., e2.wide.train)
#View(e2.wide.train_mm)
#View(e2.wide.train)
lasso <- cv.glmnet(e2.wide.train_mm, e2.wide.train$`AUT`,
                   standardize=TRUE,alpha=1,nfolds=5)
ests <- as.matrix(coef(lasso,lasso$lambda.1se))

# result of lasso
names(ests[ests!=0,])
```

```
## [1] "(Intercept)" "ARE"         "BDI"         "BGR"         "BIH"
## [6] "BLR"         "BRB"
```

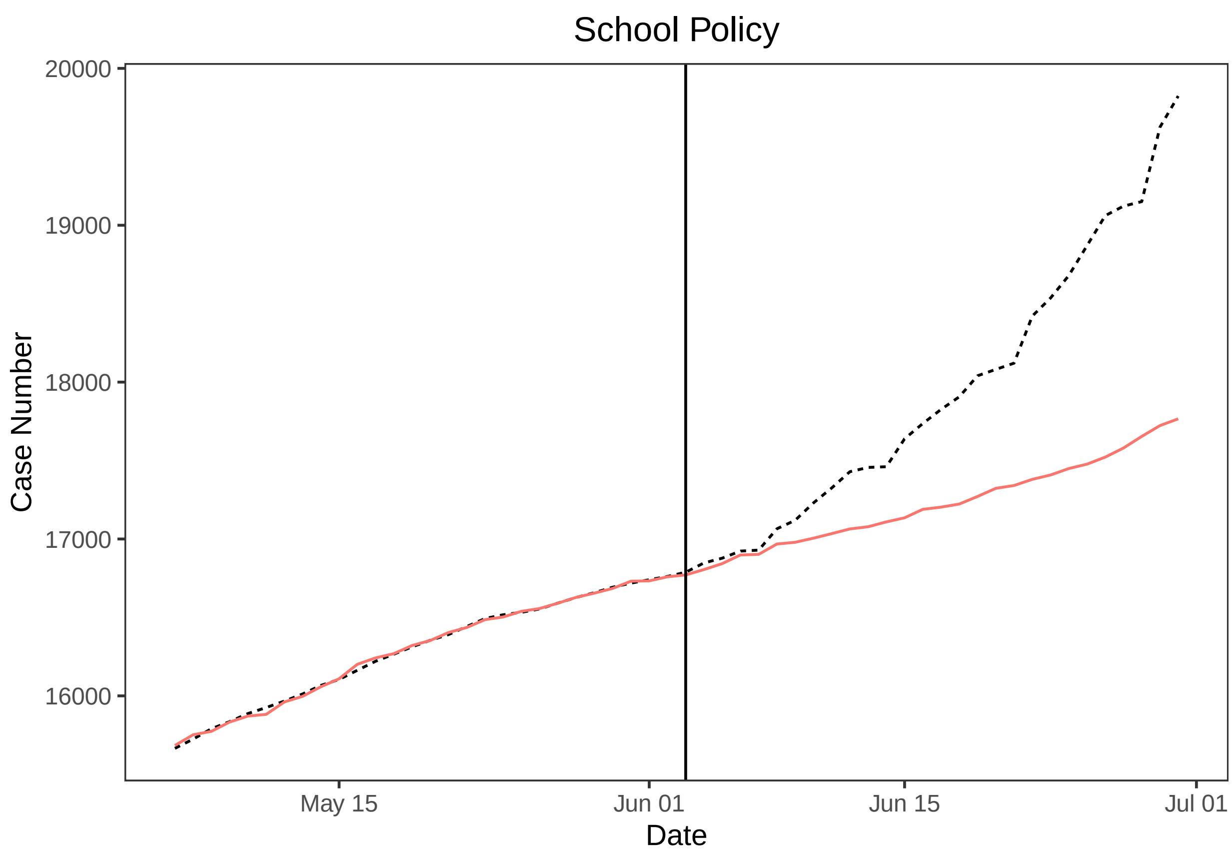
```
ests[ests!=0,]
```

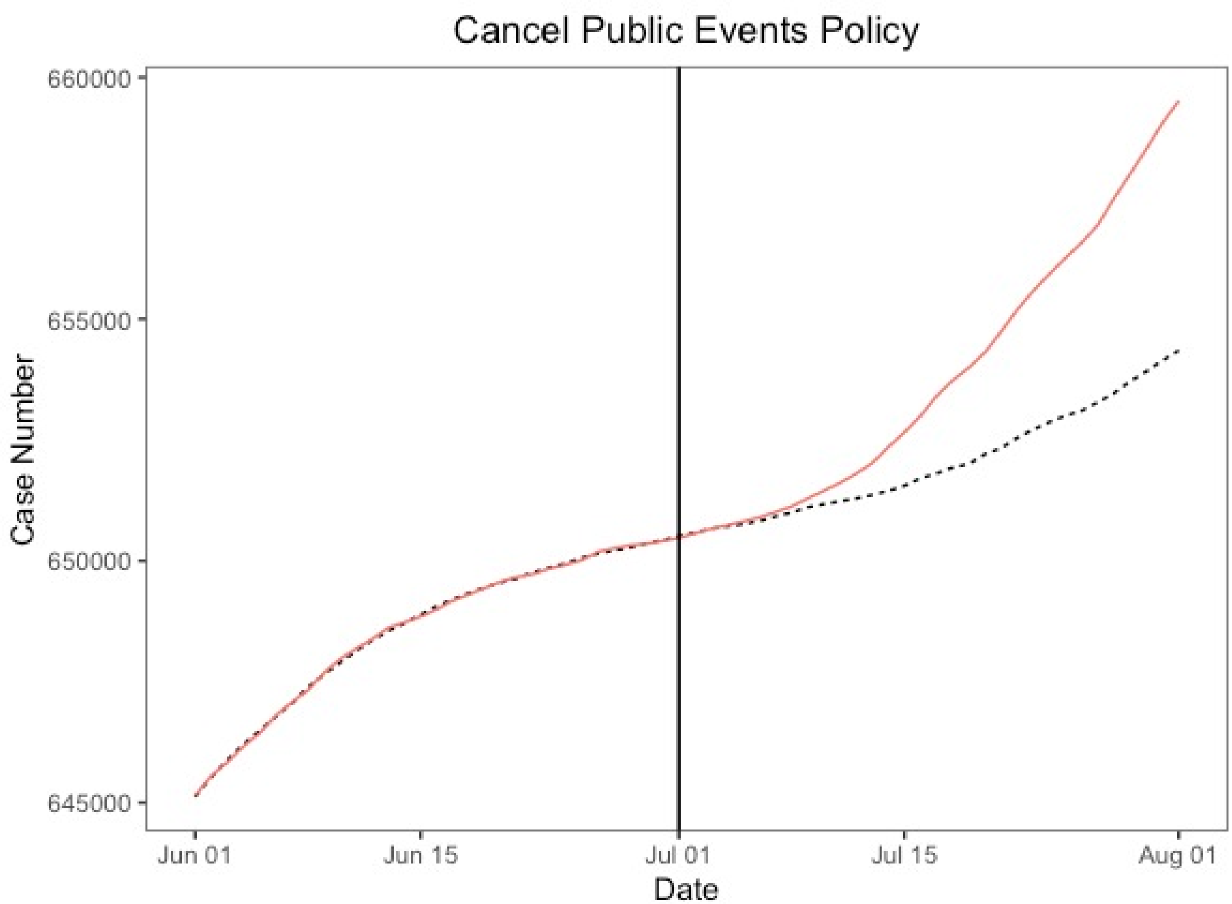
```
##  (Intercept)          ARE          BDI          BGR          BIH          BLR
## 1.317242e+04 3.620403e-03 8.600408e-02 4.756453e-01 4.992106e-01 1.129724e-02
##          BRB
## 4.992421e+00
```

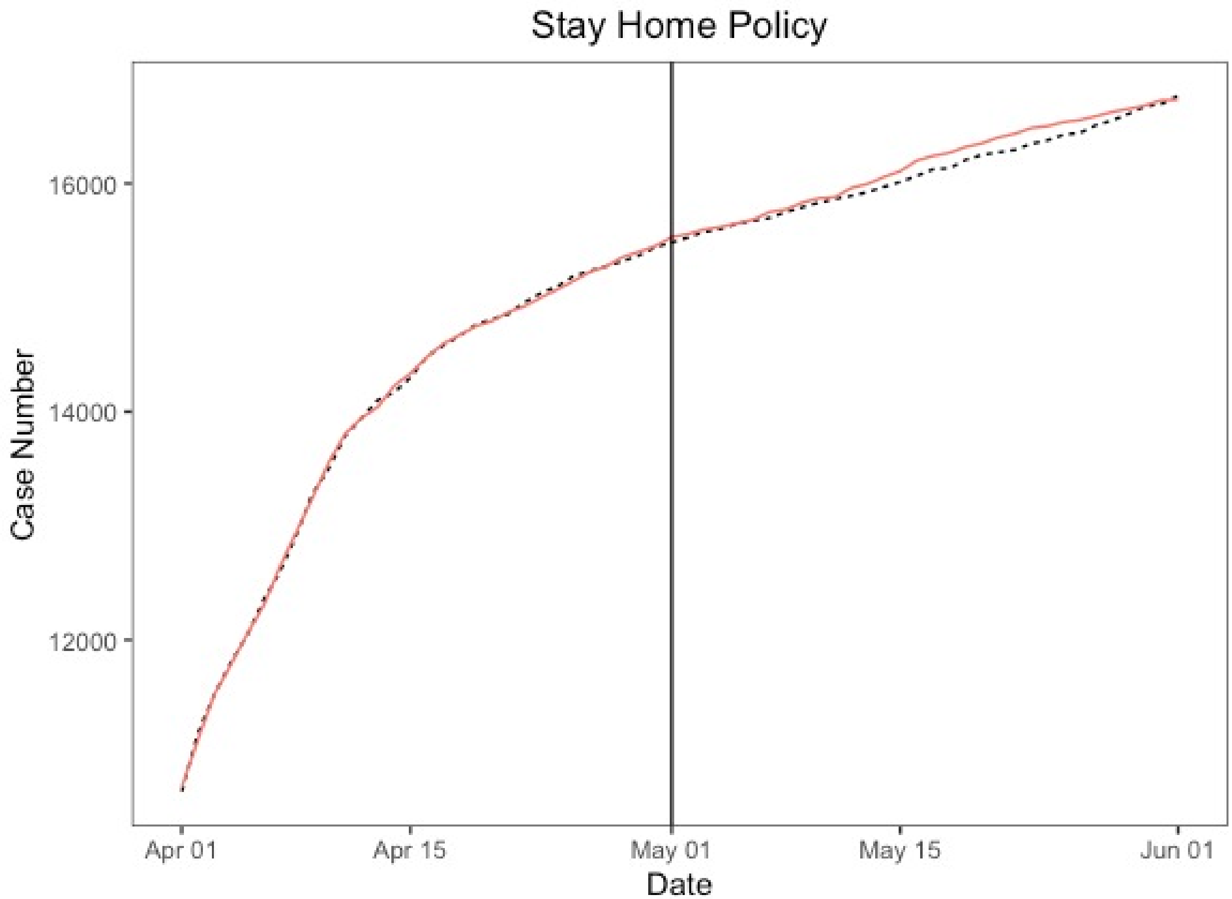
```
# synthesized function generation
fml.rhs <- paste(c(names(ests[ests!=0,]))[2:length(names(ests[ests!=0,]))],
                 collapse="+")
fml <- as.formula(paste("`AUT`~",fml.rhs))
synth <- lm(data=e2.wide.train,formula=fml)

# predict synthesized trend after experiment
e2.wide$synth <- predict(synth,newdata = e2.wide)
#View(e2.wide)

OLS_plot <- ggplot(data=e2.wide) +
  geom_line(aes(y=synth,x=date,linetype="solid"),show.legend = FALSE) +
  geom_line(aes(y=`AUT`,x=date,color="red",linetype="dashed"),
            show.legend = FALSE) +
  geom_vline(xintercept=as.numeric(as.Date("2020-06-03")),color="black") +
  ggtitle("School Policy")   +
  theme_bw() + theme(panel.grid=element_blank()) +
  theme(plot.title = element_text(hjust = 0.5))
OLS_plot=OLS_plot + xlab("Date") + ylab("Case Number")
OLS_plot
```




Cancel Public Events Policy
660000
655000
650000
645000
Case Number
Jun 01
Jun 15
Jul 01
Jul 15
Aug 01
Date


Stay Home Policy
16000
14000
12000
Case Number
Apr 01
Apr 15
May 01
May 15
Jun 01
Date

Noted: Since the codes for the graph of "Stay Home Policy" and "Cancel Public Event Policy" are similar to "School Closing Policy" except for the time frame and treatment date, we do not show the codes for them due to the space limiation.

From the graphs above we can tell that:

1) For the cancel public event policy, the actual case number go higher than the synthetize trend with the policy ending as the intervention.

2) For the school closing policy, the actual case num go lower with the policy start as intervention.

3) For the stay home policy, we can tell that the difference between actual trend and resembled trend is not that different.

Difference in Difference

After generating the synthetic trend, we then leverage the difference-in-difference method to see whether there significant difference in case number between pre-period and post-period. One important assumption of DiD method is the "parallel trend assumption", and we can tell from the previous graphs that the synthetic trend is overlapping with treatment trend in pre-period for each policy(which is a special form of parallel), therefore the "parallel trend assumption" is fulfiled.

With Did we can find out:

1. the change in post period compared to pre-period in control group

2. the change in post period compared to pre-period in treatment group

3. And most importantly, to find out the difference of this two changes. This can tell us whether there is an impact of the policy and if there is, how large the impact would be.

```
### DID Dinference-in-difference
# generate table needed
#View(e2.wide)
df1=select(e2.wide,"date","synth")
df1=rename(df1, case=synth)
df1["treat"]=0
df2=select(e2.wide,"date","AUT")
df2=rename(df2, case=AUT)
```

```r
df2["treat"]=1
df=rbind(df1,df2)
df$post=ifelse(df$date>="2020-06-03",1,0)
#View(df)
# conduct regression
school=lm(case~treat+post+treat*post,data=df)
#summary(school)
```

School Policy Result

```r
summary(school)
```

```
##
## Call:
## lm(formula = case ~ treat + post + treat * post, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1143.69  -321.20   -16.04   279.13  1891.87
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.629e+04  9.938e+01 163.872  < 2e-16 ***
## treat        4.221e-12  1.406e+02   0.000 1.000000
## post         1.645e+03  1.406e+02  11.708  < 2e-16 ***
## treat:post  -7.139e+02  1.988e+02  -3.591 0.000497 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 525.9 on 108 degrees of freedom
## Multiple R-squared:  0.6423, Adjusted R-squared:  0.6323
## F-statistic: 64.63 on 3 and 108 DF,  p-value: < 2.2e-16
```

Stay Home Policy Result

```
Call:
lm(formula = case ~ treat + post + treat * post, data = df)

Residuals:
    Min      1Q  Median      3Q     Max 
-3237.8  -407.5   144.8   575.6  1543.9 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.391e+04  1.848e+02  75.248  < 2e-16 ***
treat       3.179e-12  2.614e+02   0.000    1.000    
post        2.186e+03  2.573e+02   8.499 6.17e-14 ***
treat:post  6.836e+01  3.638e+02   0.188    0.851    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1012 on 120 degrees of freedom
Multiple R-squared:  0.554,	Adjusted R-squared:  0.5429 
F-statistic: 49.69 on 3 and 120 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = case ~ treat + post + treat * post, data = df)

Residuals:
    Min      1Q  Median      3Q     Max 
-3426.3 -1221.7    72.8  1265.1  5701.5 

Coefficients:
              Estimate Std. Error  t value Pr(>|t|)    
(Intercept)  6.485e+05  3.440e+02 1885.204  < 2e-16 ***
treat       -1.334e-10  4.865e+02    0.000  1.00000    
post         3.412e+03  4.788e+02    7.125 8.39e-11 ***
treat:post   1.861e+03  6.772e+02    2.748  0.00692 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1884 on 120 degrees of freedom
Multiple R-squared:  0.6001,	Adjusted R-squared:  0.5901 
F-statistic: 60.02 on 3 and 120 DF,  p-value: < 2.2e-16
```

Noted: Since the codes for the graph of “Stay Home Policy” and “Cancel Public Event Policy” are similar to “School Closing Policy” except for the time frame and treatment date, we do not show the codes for them due to the space limiation.

Result

For ‘Public Events Cancel policy’, the coefficient of treatment is 1861, indicating there are 1861 more cases daily after the policy ENDED. The p-value is 0.0004, indicating that the probability of observing this or more extreme number if there is no impact of the treatment in AUT is very low. In other word, there is a statistically significant effect of this policy on the COVID- 19 cases.

For ‘School Closure policy’, the coefficient of treatment is 784, indicating there are 784 less cases daily after the policy STARTED. The p-value is 0.0069, indicating that the probability of observing this or more extreme number if there is no impact of the treatment in AUT is very low. In other word, there is a statistically significant effect of this policy on the COVID- 19 cases.

For ‘Stay home policy’, the coefficient of treatment is 6.8, indicating there are 6.8 more cases daily after the policy ENDED. The p-value is 0.851, indicating that the probability of observing this or more extreme number if there is no impact of the treatment in AUT is high. In other word, we do not reject the hypothesis that there is no the effect of this policy on the COVID- 19 cases.

Therefore, we can safely say that the ‘Public Events Cancel’ policy and ‘School Closure’ policy have impact on the overall COVID-19 cases number while “Stay Home policy” does not.

Recommendation on the basis of the result: Other similar countries can reference the policy implementation in AUT and consider prioritizing the implementation of the “School Closure policy” and ‘Public Events Cancel policy’ before the “Stay Home Policy” to control the number of cases.

Limitations and Future Scope

1. Apart from the specific closure policies, we have looked into for this project namely Public events cancel policy, School closure policy, Stay home policy. There can be other policies in the dataset like healthcare, vaccination and containment policies which took place during the time period. Those can be correlated with the closure policies. For this project, we have assumed those to be constant during the policy impact analysis. So, taking the other policies in consideration can be a part of our future scope.

2. Another limitation is the cases may be under reported. Without knowing the actual effect it is hard to infer causality. This can arise when the countries or hospitals are under-reporting COVID cases. Considering the level of strictness of the policies can also be part of future scope.

3. Interference Effect: We do not know how stringent the policies were. We have assumed no interference that is, people in the areas where a certain policy was implemented adhered to the policy. No person from areas without the policy came into the place where a policy was being implemented or vice versa.

4. Becuase the Synthetic Control Method is Average Treatment on the Treated (not ATE), which means the result might not generalize to other units. Hence, the result above might onot apply to other countries which have difference from AUT.