

1. Briefly explain the policy variations induced by the NBP program.

What does variable “nbp” = 1 indicate?

The `nbp` variable is an indicator variable for if an observation is regulated by the Nitrogen Oxides (NOx) Budget Program (NBP). The NBP cap and trade program created to reduce the regional transport of NOx emissions from power plants and other large combustion sources in areas where emission concentration are at its highest.

b. What does variable “summer” = 1 indicate?

The `summer` variable is an indicator variable for if it is a summer month

c. What does variable “post = 1” indicate?

The `post` variable indicates if the year of a given observation is after 2003

The NBP operated a cap-and-trade system for over 2,500 electricity generating units and industrial boilers in the Eastern and Midwestern U.S. between 2003 and 2008. The market only operated between May 1 and September 30 since ozone pollution is highest during the summer.

(source (http://joseph-s-shapiro.com/research/NOx_Defenses_DGS.pdf)).

2. Replicate Panel A (States Participating in NBP) of the Appendix Fig. 2.

```

pacman::p_load(ggplot2,
               dplyr,
               magrittr)
# Reading data
nbp = haven::read_dta("nbp.dta")
skim_df = skimr::skim(nbp)

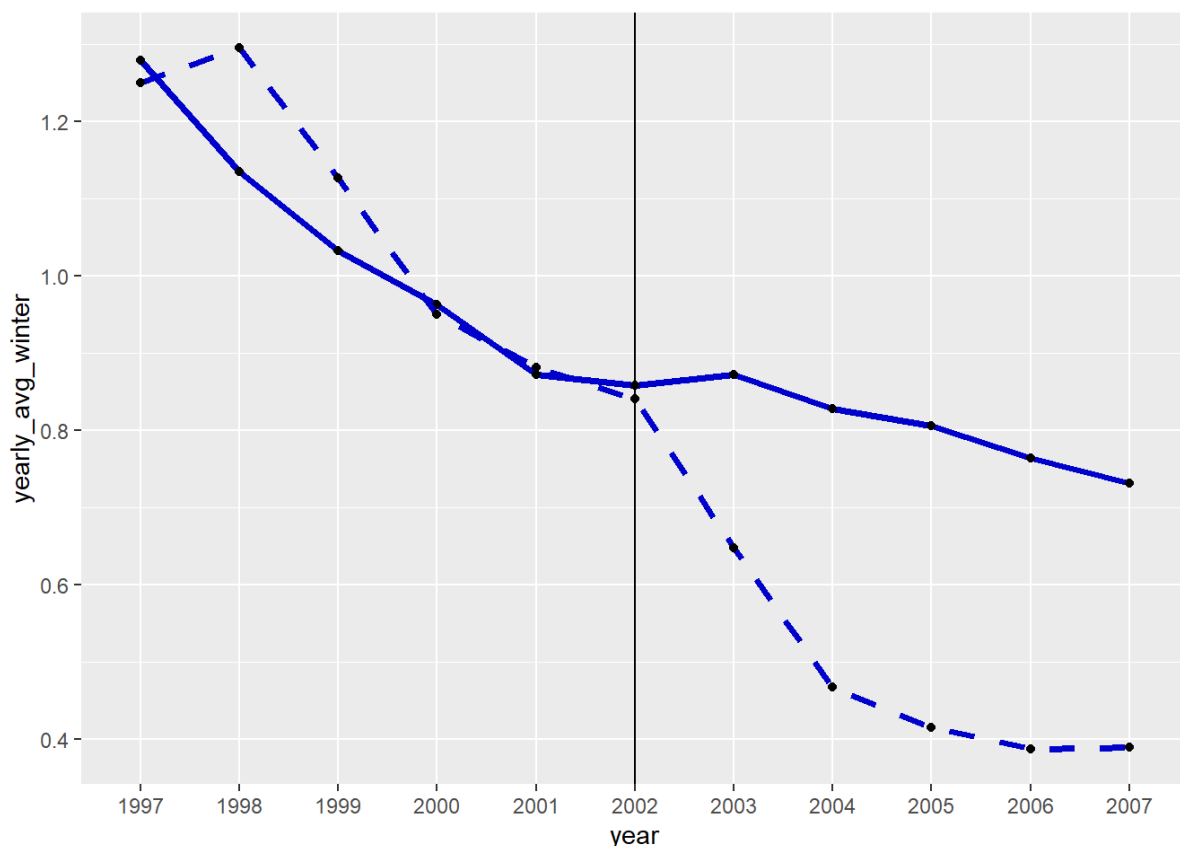
# Cleaning

nbp = nbp %>%
  mutate(
    fips_county = as.character(fips_county),
    year = as.factor(year),
    summer = as.factor(summer),
    state = as.factor(fips_state),
    nbp = as.factor(nbp),
    post_2003 = as.factor(post)) %>%
  rename(emissions = nox_emit,
         county_code = fips_county)

# Creating means by year
merged_df = nbp %>% group_by(year) %>% filter(summer==1 & nbp==1) %>%
  summarise(yearly_avg_summer = mean(emissions))
nbp %>% group_by(year) %>% filter(summer==0 & nbp==1) %>%
  summarise(yearly_avg_winter = mean(emissions)) %>%
  merge(.,merged_df, by="year") -> merged_df

ggplot(merged_df, aes(x = year))+
  geom_line(aes(y = yearly_avg_winter,
               group = 1),
           color = "mediumblue",
           size = 1.25)+
  geom_point(aes(x = year,y=yearly_avg_winter))+
  geom_line(aes(y = yearly_avg_summer,
               group = 2),
           color = "mediumblue",
           size = 1.25,
           linetype = "dashed")+
  geom_point(aes(x = year,y=yearly_avg_summer))+
  geom_vline(xintercept = "2002") -> panel_A
panel_A

```



3. State the parallel trends assumption underlying Panel A of Appendix Fig. 2. Does the graphical pattern suggest the parallel trends assumption is likely to hold?

The parallel trends assumption states that, although treatment and comparison groups may have different levels of the outcome prior to the start of treatment, their trends in pre-treatment outcomes should be the same (Ryan et. al. ([https://journals.sagepub.com/doi/abs/10.1177/0962280218814570?journalCode=smaa#:~:text=The%20parallel%20trends%20assumption%20states,outcomes%20should%20be%20the%20same.\)\)](https://journals.sagepub.com/doi/abs/10.1177/0962280218814570?journalCode=smaa#:~:text=The%20parallel%20trends%20assumption%20states,outcomes%20should%20be%20the%20same.)))).

In this case, the solid line representing summer months would be the treatment because the policy is only enforced during summer months and is not in effect during the winter.

Higher temperatures are generally associated with higher ozone levels in general (EPA (<https://www.epa.gov/air-trends/trends-ozone-adjusted-weather-conditions>)), so we would expect to see a difference between the two lines on the same graph. Since the policy is only in effect during summer months, then the trends that exist during winter months should continue because they are not affected by the policy. The parallel trends assumption in this context would be holding the assumption that the same trend should can be found in winter months for both the treatment and the control group, and as a result we can assume that changes in our outcome variable (emissions) can be attributed to exposure to the treatment (i.e. being an NBP-participating state)

4. Clearly state, estimate, and report a difference-in-differences regression that examines the effect of the NBP program on NOx emissions corresponding to Panel A of Appendix Fig. 2. Provide a one-sentence interpretation of the coefficient for the interaction term “summer*post”.

To compute the difference-in-difference estimate of the casual effect of the treatment, we can create simple linear models from the data included in the panel representation

```
nbp %<>% mutate(year = as.integer(as.character(year)))
sum_treat_pre = nbp %>% group_by(year) %>%
  filter(summer==1 & nbp==1 & year<2003)
sum_treat_post = nbp %>% group_by(year) %>%
  filter(summer==1 & nbp==1 & year>=2003)
win_treat_pre = nbp %>% group_by(year) %>%
  filter(summer==0 & nbp==1 & year<2003)
win_treat_post = nbp %>% group_by(year) %>%
  filter(summer==0 & nbp==1 & year>=2003)

mod_tb = lm(emissions~year, data = sum_treat_pre)
mod_ta = lm(emissions~year, data = sum_treat_post)
mod_ub = lm(emissions~year, data = win_treat_pre)
mod_ua = lm(emissions~year, data = win_treat_post)
stargazer::stargazer(mod_tb, mod_ta, mod_ub, mod_ua, type = "text",
  column.labels = c("TB","TA","UB","UA"))
```

```
##
## =====
##
##                               Dependent variable:
## -----
##                               emissions
##                               TB           TA           UB           U
A
##                               (1)           (2)           (3)
## (4) -----
## -----
## year                -0.099***           -0.060***           -0.085***           -0.
035
##                               (0.027)           (0.014)           (0.026)           (0.0
26)
##
## Constant            199.331***           119.963***           170.357***           70.
420
##                               (53.703)           (28.427)           (52.727)           (53.
068)
##
## -----
## Observations                7,110           5,925           7,110           5,9
25
## R2                        0.002           0.003           0.001           0.0
003
## Adjusted R2              0.002           0.003           0.001           0.0
001
## Residual Std. Error    3.868 (df = 7108)      1.543 (df = 5923)      3.797 (df = 7108)      2.881 (df
= 5923)
## F Statistic            13.631*** (df = 1; 7108) 17.672*** (df = 1; 5923) 10.314*** (df = 1; 7108) 1.721 (df
= 1; 5923)
## =====
## Note:
***p<0.01                                *p<0.1; **p<0.05;
```

```
sum_treat_pre %>% ungroup() %>% summarise(mean(emissions)) -> TB
sum_treat_post %>% ungroup() %>% summarise(mean(emissions)) -> TA
win_treat_pre %>% ungroup() %>% summarise(mean(emissions)) -> UB
win_treat_post %>% ungroup() %>% summarise(mean(emissions)) -> UA
```

And now to estimate the causal effect of the treatment

```
calc1 = ((TA-TB)-(UA-UB))
calc1
```

```
## mean(emissions)
## 1 -0.3731687
```

Or we could define a `lm` model formally defining interaction variables to understand interactions that may be present in our data

```
mod_lm = nbp %>% mutate(year = as.integer(year)) %>%
  lm(emissions~year+
      nbp+nbp*post_2003+
      post_2003+post_2003*summer+
      summer+summer*nbp,
      data = .)
summary(mod_lm)
```

```
##
## Call:
## lm(formula = emissions ~ year + nbp + nbp * post_2003 + post_2003 +
##     post_2003 * summer + summer + summer * nbp, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.177 -0.752 -0.532 -0.397  67.413
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   90.365905   14.327695   6.307 2.86e-10 ***
## year          -0.044961    0.007166  -6.274 3.53e-10 ***
## nbp1           0.596836    0.038227  15.613 < 2e-16 ***
## post_20031     0.222400    0.055153   4.032 5.53e-05 ***
## summer1        0.154348    0.037292   4.139 3.50e-05 ***
## nbp1:post_20031 -0.286792    0.045609  -6.288 3.24e-10 ***
## post_20031:summer1 -0.196647    0.045508  -4.321 1.55e-05 ***
## nbp1:summer1   -0.200792    0.045420  -4.421 9.85e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.678 on 55850 degrees of freedom
## Multiple R-squared:  0.008982, Adjusted R-squared:  0.008858
## F-statistic: 72.32 on 7 and 55850 DF, p-value: < 2.2e-16
```

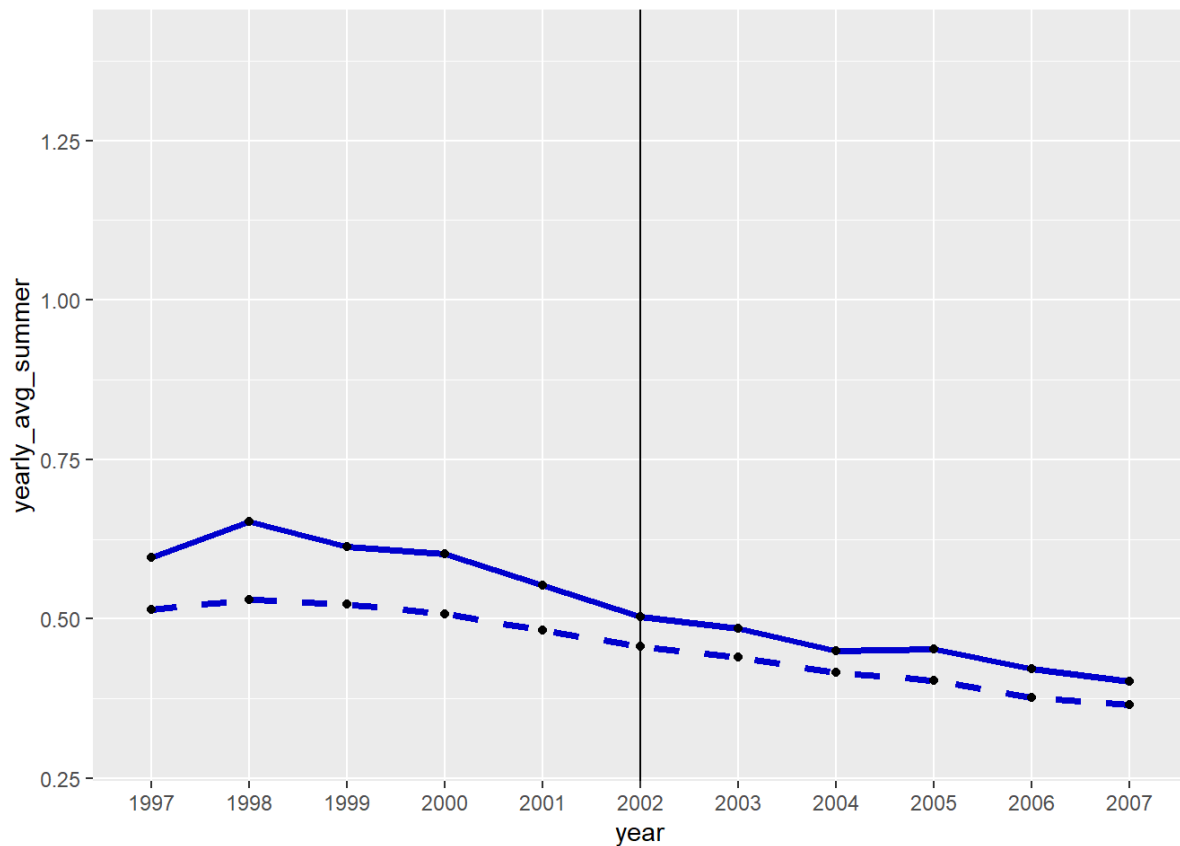
The `summer*post` interaction tells us how significant that the characterization of an observation being from a summer month post-2003 has in explaining the relationship that exists in the data.

5. Replicate Panel B (States Not Participating in NBP) of the Appendix Fig. 2.

```
merged_df2 = nbp %>% group_by(year) %>% filter(summer==1 & nbp==0) %>%
  summarise(yearly_avg_summer = mean(emissions))%>%
  mutate(year = as.numeric(year))
nbp %>% group_by(year) %>% filter(summer==0 & nbp==0) %>%
  summarise(yearly_avg_winter = mean(emissions)) %>%
  merge(.,merged_df2, by="year") -> merged_df2
merged_df2 %<>% ungroup() %>% mutate(year = as.factor(year))
```

```
p1 = ggplot(merged_df2, aes(x = year))+
  geom_line(aes(y = yearly_avg_summer,
               group = 1),
            color = "mediumblue",
            size = 1.25)+
  geom_point(aes(y=yearly_avg_summer))
```

```
panel_B = p1 +
  geom_line(aes(y = yearly_avg_winter,
               group = 2),
            color = "mediumblue",
            size = 1.25,
            linetype = "dashed")+
  geom_point(aes(y=yearly_avg_winter)) +
  geom_vline(xintercept = "2002")+
  ylim(0.3,1.4) # Adjusting to scale as previous image
panel_B
```



6. Explain what is the point of Panel B of the Appendix Fig. 2.

The purpose of showing Panel B is because it allows us to compare trends that are occurring with the treatment group (NBP states) and the control group (non-NBP states) at the same time. Even though the values may be higher in absolute terms for NBP-states (hence the idea for the policy), we would still expect trends to “continue as normal” in the absence of the treatment. We can see that the dashed line representing winter months in both panels is nearly identical from simple visual observation. The trend for summer months, however, is noticeably different. Furthermore, we can also see there seems to be an inflection point in the trend for summer-emissions for NBP states, where the trends are no longer identical as they were prior to the implementation of the treatment/policy.

7. Clearly state, estimate, and report a difference-in-differences regression that examines the effect of the NBP program on NOx emissions corresponding to Panel B of Appendix Fig. 2. Provide a one-sentence interpretation of the coefficient for the interaction term “summer*post”.

```
nbp %<>% mutate(year = as.integer(as.character(year)))
sum_ctrl_pre = nbp %>% group_by(year) %>%
  filter(summer==1 & nbp==0 & year<2003)
sum_ctrl_post = nbp %>% group_by(year) %>%
  filter(summer==1 & nbp==0 & year>=2003)
win_ctrl_pre = nbp %>% group_by(year) %>%
  filter(summer==0 & nbp==0 & year<2003)
win_ctrl_post = nbp %>% group_by(year) %>%
  filter(summer==0 & nbp==0 & year>=2003)

mod_cb = lm(emissions~year, data = sum_ctrl_pre)
mod_ca = lm(emissions~year, data = sum_ctrl_post)
mod_ub = lm(emissions~year, data = win_ctrl_pre)
mod_ua = lm(emissions~year, data = win_ctrl_post)
stargazer::stargazer(mod_cb, mod_ca, mod_ub, mod_ua, type = "text",
  column.labels = c("CB","CA","UB","UA"))
```



```
##
## =====
##
##                               Dependent variable:
## -----
##                               emissions
##                               CB          CA          UB          UA
##                               (1)        (2)        (3)        (4)
## -----
## year                        -0.099***    -0.060***    -0.013      -0.019
##                               (0.027)      (0.014)      (0.014)      (0.015)
##
## Constant                    199.331***    119.963***    26.047      38.556
##                               (53.703)      (28.427)      (27.825)      (30.878)
## -----
## Observations                 7,110         5,925         8,124         6,770
## R2                          0.002          0.003          0.0001        0.0002
## Adjusted R2                 0.002          0.003          -0.00002       0.0001
## Residual Std. Error    3.868 (df = 7108)    1.543 (df = 5923)    2.142 (df = 8122)    1.792 (df = 6768)
## F Statistic             13.631*** (df = 1; 7108) 17.672*** (df = 1; 5923) 0.843 (df = 1; 8122) 1.527 (df = 1; 6768)
## =====
## Note:                               *p<0.1; **p<0.05; ***
p<0.01
```

```
sum_ctrl_pre %>% ungroup() %>% summarise(mean(emissions)) -> CB
sum_ctrl_post %>% ungroup() %>% summarise(mean(emissions)) -> CA
win_ctrl_pre %>% ungroup() %>% summarise(mean(emissions)) -> UB
win_ctrl_post %>% ungroup() %>% summarise(mean(emissions)) -> UA
```

And now to estimate the causal effect of the treatment

```
calc2 = ((CA-CB)-(UA-UB))
calc2
```

```
## mean(emissions)
## 1 -0.04215831
```

8. Clearly state, estimate, and report a triple-difference regression that examines the effect of NBP program on NOx emissions. Discuss how the coefficient for the interaction term “nbpsummerpost” relates to your answers to question 4 and question

To calculate the triple-difference regression estimator, we simply solve for the difference between the difference-in-difference estimate of the effect of treatment in participating states and the estimate from non-participating states.

```
calc3 = calc2 - calc1
calc3
```

```
## mean(emissions)
## 1 0.3310104
```

To look at this relationship from the perspective of a linear model, we can create the following to tell us how significant the triple interaction is at defining the relationship in the data and understanding the true causality of trends, specifically on NOx emissions. Modeling with DDD allows us to gain a more insightful model in terms of describing causal inference, and since there are parallel trends, we can have confidence in the results from the model.

```
mod_lm_3D = nbp %>% mutate(year = as.integer(year)) %>%
  lm(emissions~year+
      nbp+nbp*post_2003+
      post_2003+post_2003*summer+
      summer+summer*nbp+
      nbp*post_2003*summer,
      data = .)
summary(mod_lm_3D)
```

```
##
## Call:
## lm(formula = emissions ~ year + nbp + nbp * post_2003 + post_2003 +
##     post_2003 * summer + summer + summer * nbp + nbp * post_2003 *
##     summer, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.170 -0.801 -0.519 -0.390  67.373
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   90.401016   14.326137   6.310 2.81e-10 ***
## year          -0.044961    0.007165  -6.275 3.52e-10 ***
## nbp1           0.521607    0.043482  11.996 < 2e-16 ***
## post_20031     0.145155    0.059112   2.456 0.014067 *
## summer1        0.084126    0.042010   2.003 0.045233 *
## nbp1:post_20031 -0.121287    0.064494  -1.881 0.060032 .
## post_20031:summer1 -0.042158    0.062310  -0.677 0.498672
## nbp1:summer1   -0.050333    0.061492  -0.819 0.413063
## nbp1:post_20031:summer1 -0.331010    0.091208  -3.629 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.677 on 55849 degrees of freedom
## Multiple R-squared:  0.009216, Adjusted R-squared:  0.009074
## F-statistic: 64.94 on 8 and 55849 DF, p-value: < 2.2e-16
```

9. Suppose that your job is to provide a retrospective analysis on the impact of the NBP program for the Environmental Protection Agency. Write a short summary of 200 or fewer words describing what you have found in your analysis of the NBP data. In particular, based on your triple-difference estimate, provide a calculation on how much NO_x emissions in total has the NBP program reduced from 2003 to 2007?

From looking at the data and validating the methods used by the authors in this publication, the results from this quasi-experiment suggest that NO_x NBP cap-and-trade markets have positive effects when measured in terms of several measurements such as lower levels of NO_x emissions, ozone concentration, and mortality rates. The health benefit is also passed on to the pharmaceutical companies. The decreased need for additional expenditures that would otherwise be demanded without the treatment/regulatory standards, the valued amount of the health benefit savings would outweigh the cost of implementation and adherence to this program, while at the same time also providing many other benefits associated with lower levels of NO_x not present in this data.

In our analysis of the same data used by the authors, we used several models and different metrics to assess the accuracy/confidence of the claim made in their research. The data was modeled with triple difference regression since (since the parallel trends assumption was present and verified) and the estimator of the treatment effect was approximately 0.33, thus suggesting a statistically significant and reproducible analysis.