Empirical Project 3

The Effect of the U.S. EPA NOx Budget Trading Program

Devin Bunch

1.Briefly explain the policy variations induced by the NBP program. a. What does variable “nbp”= 1 indicate? b. What does variable “summer”= 1 indicate? c. What does variable “post = 1”indicate?

a. The variable nbp is equal to 1 if the state is regulated by the NBP, and thus must take action against their NOx emission problem. Indicator—dummy variable.
b. The variable summer is in reference to the season of the year. When “summer” is equal to 1, the data was collected during a time period wherein the season was currently Summer. Indicator—dummy variable.
c. The variable post describes the time period after treatment started. Here, the treatment was administered in the year 2003, so our post variable serves as an indicator equal to 1 if the data observation was recorded after the cut-off time treatment started.

##done

Replicate Panel A (States Participating in NBP) of the Appendix Fig. 2. Hint: To obtain the solid line, compute year-over-year averages of “nox_emit” for all counties with “nbp”=1 and “summer”=1.

```
#Load packages and libraries
library(pacman)
p_load(readr,dplyr, tidyverse, ggplot2, skimr, haven, stargazer, tidymodels, skimr
, janitor, magrittr, datasets, rpart.plot, baguette, glmnet, tune, haven, ranger,
data.table, parallel, sandwich,  huxtable, rdrobust, fixest, scales)

#more specifically,
#library(tidyverse)  # for ggplot(), %>%, mutate(), and friends
library(broom)  # Convert models to data frames
library(modelsummary)  #to create side-by-side regression tables
library(hrbrthemes)  # To exactly replicate the look of the figures
```

```
## NOTE: Either Arial Narrow or Roboto Condensed fonts are required to use these t
hemes.
```

```
##       Please use hrbrthemes::import_roboto_condensed() to install Roboto Conden
sed and
```

```
##       if Arial Narrow is not on your system, please see https://bit.ly/arialnar
row
```

```
#load data from downloads, using the haven package, R cannot download ".dta" files
nbp_df <- read_dta("~/Downloads/nbp.dta")
```

```
# This tells us we have seven columns, we are given variable names, and all observ
ations are numeric
#names(nbp_df)
#skim(nbp_df)

#try to just view it
#nbp_df
```

```
#Panel A (States Participating in NBP)

#let's create a grouped data set containing mean observations for each year and su
mmer month

fig2 <- nbp_df %>%
  group_by(summer, year, nbp) %>%
  summarize(nox_mean = mean(nox_emit))
```

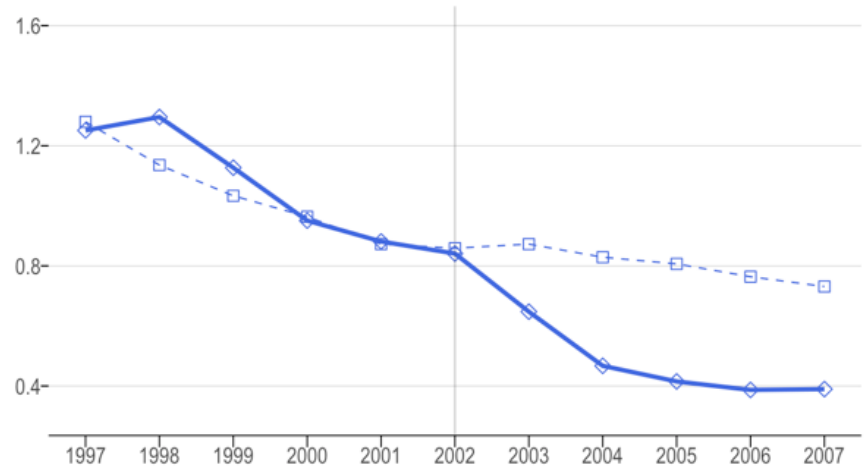
```
## `summarise()` has grouped output by 'summer', 'year'. You can override using th
e `.groups` argument.
```

```
#now let's  plot with ggplot, beimg sure to use filters to display the right obser
vations
#only include participating states in this part
pan_a  <- ggplot(data = fig2 %>% filter(nbp == 1), aes(x = year, y = nox_mean, sha
pe = factor(summer))) +
  # match the shapes in the article
  geom_point(data = fig2 %>% filter(nbp == 1 & summer == 1),
             color = "royalblue", shape = 5, size = 2.3) + ## open diamond shape
  # create a different line for each season, summer
  geom_line(data = fig2 %>% filter(nbp == 1 & summer == 1),
            aes(x = year, y = nox_mean), size = 1.1, color = "royalblue") +
#create the verticle line like in the article
  geom_vline(xintercept = 2002, alpha = 0.2) + ## the year treatment started
  #the figure in the paper does not have axis labels
  labs(x = " ", y = " ", title = "Figure 2. Summer-Equivalent Seasonal NOx Emissio
ns (Mil. Tons)", subtitle = "(A) States Participating in NBP",
       color = "summer") +
  scale_x_continuous(breaks = seq(1997, 2007, 1),
                     labels = c("1997", "1998", "1999", "2000", "2001", "2002", "2
003", "2004", "2005", "2006", "2007")) +
  scale_y_continuous(breaks = seq(0.4, 1.6, 0.4), labels =  seq(0.4, 1.6, 0.4), li
mits = c(0.3, 1.6)) +
  #remove the grid lines using themes
  theme_ipsum() +
  theme(panel.grid.major.x = element_blank(),
        panel.grid.minor = element_blank()) +
  # match the shapes in the article
  geom_point(data = fig2 %>% filter(nbp == 1 & summer == 0),
             color = "royalblue", shape = 0, size = 2.3) +  ## open box shape
  # create a different line for each season, winter
  geom_line(data = fig2 %>% filter(nbp == 1 & summer == 0),
            aes(x = year, y = nox_mean),
            linetype = "dashed", size = 0.4, color = "royalblue") +
  ## making the format of the x axis manually
  #geom_hline(yintercept = 0.3, colour = "black", size = 0.45) +
  theme(axis.ticks.x = element_line(size = 0.34),
        axis.ticks.length = unit(6, "pt"),
        axis.line.x = element_line(size = 0.34),
        axis.ticks.length.y = unit(4, "pt"),
        axis.ticks.y = element_line(size = 0.4))

pan_a
```

Figure 2. Summer-Equivalent Seasonal NOx Emissions (Mil. Tor

(A) States Participating in NBP



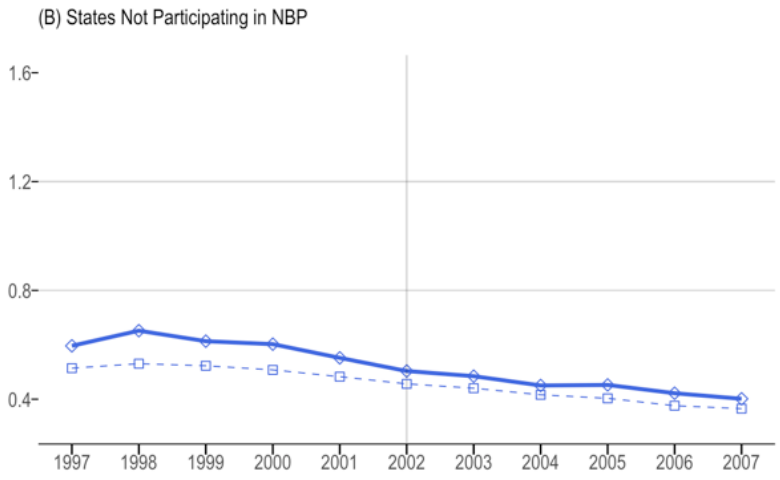

```
#now to add the finishing touches to the legend

#Wow, it looks impeccable


##notes, if you will

#make the line colors match the paper exactly
#scale_color_manual(values = c("royalblue", "royalblue"))
#pan_a + scale_fill_continuous(guide_legend())
```

```r
#replicate the same code for panel b, making the necessary changes to just include
states not participating in the program
#only include participating states in this part
pan_b  <- ggplot(data = fig2 %>% filter(nbp == 0), aes(x = year, y = nox_mean)) +
  # match the shapes in the article
  geom_point(data = fig2 %>% filter(nbp == 0 & summer == 1),
             color = "royalblue", shape = 5, size = 2) +
  # create a different line for each season, summer
  geom_line(data = fig2 %>% filter(nbp == 0 & summer == 1),
            aes(x = year, y = nox_mean), size = 1.1, color = "royalblue") +
#create the verticle line like in the article
  geom_vline(xintercept = 2002, alpha = 0.2) + ## the year treatment started ## th
ere's no treatment for panel B, but Eric has the line written in panel B in the in
structions, so we will include it here
  #the figure in the paper does not have axis labels
  labs(x = " ", y = " ", title = " ", subtitle = "(B) States Not Participating in
NBP",
       color = "summer") +
  scale_x_continuous(breaks = seq(1997, 2007, 1)) +
  scale_y_continuous(breaks = seq(0.4, 1.6, 0.4), labels = seq(0.4, 1.6, 0.4), lim
its = c(0.3, 1.6)) +
  #remove the grid lines using themes
  theme_ipsum() +
  theme(panel.grid.major.x = element_blank(),
        panel.grid.minor = element_blank(),
        panel.grid.major.y = element_blank()) +
  ## make the only two horizontal lines
  geom_hline(yintercept = 0.8, alpha = 0.2, size = 0.4) +
  geom_hline(yintercept = 1.2, alpha = 0.2) +
  # match the shapes in the article
  geom_point(data = fig2 %>% filter(nbp == 0 & summer == 0),
             color = "royalblue", shape = 0, size = 2) +
  # create a different line for each season, winter
  geom_line(data = fig2 %>% filter(nbp == 0 & summer == 0),
            aes(x = year, y = nox_mean),
            linetype = "dashed", size = 0.34, color = "royalblue") +
    theme(axis.ticks.x = element_line(),
          axis.ticks.length = unit(6, "pt"),
          axis.line.x = element_line(size = 0.34),
          axis.ticks.length.y = unit(4, "pt"),
          axis.ticks.y = element_line(size = 0.4))
# Note in the appendix, it's zoomed in a little because panel b is closer to the x
axis and harder to see from further away
pan_b
```



Notes: The data underlying Appendix Figure 2 is expressed as summer-equivalent since the summer period has 5 months while the winter period has 7 months. Specifically, the summer equivalent of winter emissions is actual winter emissions multiplied by 5/7. These graphs show summary statistics describing total emissions, not regression results. Summer defined as May- September, winter as January-April and October-December. NBP participating states include: Alabama, Connecticut, Delaware, District of Columbia, Illinois, Indiana, Kentucky, Maryland, Massachusetts, Michigan, Missouri, New Jersey, New York, North Carolina, Ohio, Pennsylvania, Rhode Island, South Carolina, Tennessee, Virginia, and West Virginia. States not participating in NBP include: Arkansas, Arizona, California, Colorado, Florida, Idaho, Kansas, Lousiana, Minnesota, Montana, Nebraska, Nevada, New Mexico, North Dakota, Oklahoma, Oregon, South Dakota, Texas, Utah, Washington, Wyoming. Alaska, Georgia, Hawaii, Iowa, Maine, Mississippi, Missouri, New Hampshire, Vermont, and Wisconsin are excluded from the main analysis sample.

Looks great Devin, keep going! :-)

we define Post=0.5 in 2003 and Post=1.0 in 2004 through 2007. A

```
#regression question 4, DD estimator
#make a new filtered data frame for just the participating states
nbp1 <- nbp_df %>% filter(nbp == 1)

#now let's run the regression
reg1 = lm(nox_emit ~ summer + post + summer*post, data = nbp1)

#display our results nicely
#summary(reg1)
tidy(reg1)
```

term	estimate	std.error	statistic	p.value
(Intercept)	1.02	0.0383	26.7	5.23e-155
summer	0.0338	0.0542	0.623	0.533
post	-0.223	0.0569	-3.93	8.59e-05
summer:post	-0.373	0.0804	-4.64	3.51e-06

```
stargazer(reg1, keep.stat = c("n", "rsq"), model.numbers = FALSE,style = 'qje', ty
pe = 'text')
```

```
##
## ================================================
##                             nox_emit
## ------------------------------------------------
## summer                       0.034
##                             (0.054)
##
## post                        -0.223***
##                             (0.057)
##
## summer:post                 -0.373***
##                             (0.080)
##
## Constant                     1.024***
##                             (0.038)
##
## N                            26,070
## R2                           0.005
## ================================================
## Notes:      ***Significant at the 1 percent level.
##              **Significant at the 5 percent level.
##              *Significant at the 10 percent level.
```

```
# regression for question 7
#make a new data frame for just nbp == 0:
df_2 = nbp_df %>% filter(nbp ==0)

#now put it into our regression
reg7 = lm(nox_emit ~ post + summer + post*summer, data = df_2)

#and display the results
tidy(reg7)
```

term	estimate	std.error	statistic	p.value
(Intercept)	0.502	0.023	21.8	7.72e-105
post	-0.102	0.0341	-2.99	0.00277
summer	0.0841	0.0325	2.59	0.00974
post:summer	-0.0422	0.0483	-0.873	0.382

```
stargazer(reg7,  keep.stat = c("n", "rsq"), model.numbers = FALSE,style = 'qje', t
ype = 'text')
```

```
##
## ===================================================
##                                 nox_emit
## ---------------------------------------------------
## nbp                             0.522***
##                                 (0.043)
##
## summer                          0.084**
##                                 (0.042)
##
## post                            -0.102**
##                                 (0.044)
##
## nbp:post                        -0.121*
##                                 (0.065)
##
## summer:post                      -0.042
##                                 (0.062)
##
## nbp:summer                       -0.050
##                                 (0.062)
##
## nbp:summer:post                -0.331***
##                                 (0.091)
##
## Constant                        0.502***
##                                 (0.030)
##
## N                                55,858
## R2                               0.009
## ===================================================
## Notes:          ***Significant at the 1 percent level.
##                  **Significant at the 5 percent level.
##                  *Significant at the 10 percent level.
```

```
##
## ================================================
##                             nox_emit
## ------------------------------------------------
## post                       -0.102***
##                             (0.034)
##
## summer                      0.084***
##                             (0.033)
##
## post:summer                  -0.042
##                             (0.048)
##
## Constant                    0.502***
##                             (0.023)
##
## N                            29,788
## R2                           0.001
## ================================================
## Notes:      ***Significant at the 1 percent level.
##              **Significant at the 5 percent level.
##              *Significant at the 10 percent level.
```

```
# regression for question 8, using the whole unfiltered data frame,
reg8 = lm(nox_emit ~ nbp + summer + post + nbp*post + post*summer + nbp*summer + n
bp*post*summer, data = nbp_df)

#and displaying results
tidy(reg8)
```

term	estimate	std.error	statistic	p.value
(Intercept)	0.502	0.0297	16.9	5.84e-64
nbp	0.522	0.0435	12	4.3e-33
summer	0.0841	0.042	2	0.0453
post	-0.102	0.0441	-2.32	0.0205
nbp:post	-0.121	0.0645	-1.88	0.0601
summer:post	-0.0422	0.0623	-0.676	0.499
nbp:summer	-0.0503	0.0615	-0.818	0.413
nbp:summer:post	-0.331	0.0912	-3.63	0.000286

```
stargazer(reg8,  keep.stat = c("n", "rsq"), model.numbers = FALSE,style = 'qje', t
ype = 'text')
```