Assignment 2 - EDS 241: Environmental Policy Evaluation

Juliet Cohen

01/30/2022

The goal of this assignment is to provide a simple test of whether the effects of air quality regulations are the same across locations with different racial mix. To this end you will test if the NOx Budget Program, a cap-and-trade market for nitrogen oxides (NOx) emissions from power plants lead to similar effects in counties that are predominantly white versus counties that are predominantly African American. The data are a subset of data sample I used in the following paper: https://olivierdeschenes.weebly.com/uploads/1/3/5/0/135068654/defensive-investmentsand-the-demans-for-air-quality.pdf. You can also get more information on the NOx Budget Program, here: https://www.epa.gov/airmarkets/nox-budget-trading-program

The data included in the file NBP.xls, which is available on Gauchospace, are: fips (fips code identifying each county), NBP (indicator =1 if the county was regulated under the NOx Budget Program), PctBlack (fraction of the county population that is African American), and Dnox_masstons (change in annual NOx emissions from all power plants in a county between 2000 and 2008 (in tons)). Note that the NBP market was in effect in 212 of the 485 counties in the sample from 2003 to 2008, so the 2008-2000 change give us a sense of the program's effect on emissions. If emissions of NOx from power plants declined in a county, then Dnox_masstons should be negative.

Finally, it may be useful to review the RegressionSpecification.pdf notes on Gauchospace before doing the assignment.

0.0.1 Load Packages

```
library(here)
library(dplyr)
library(tidyverse)
library(estimatr)
library(modelsummary)
library(knitr)
```

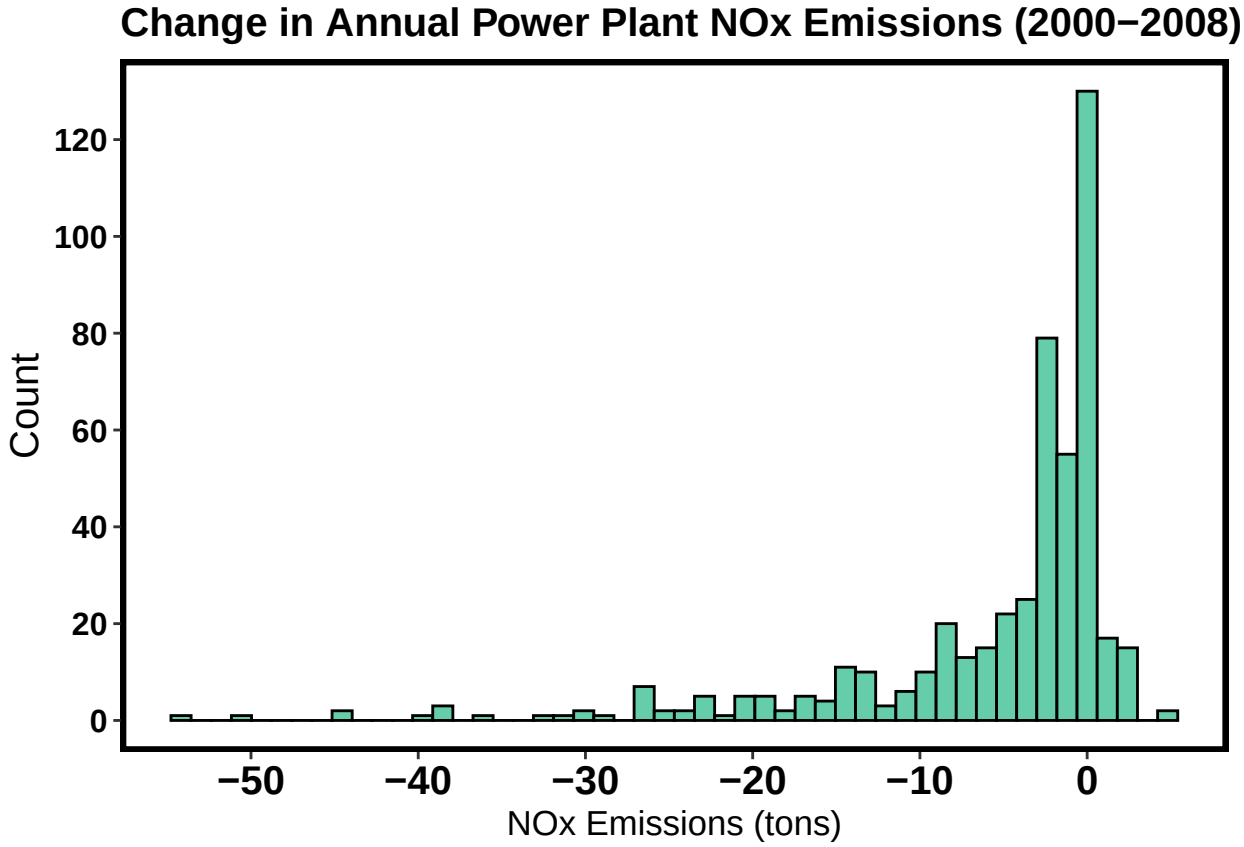
0.0.2 Read in Data

```
data <- read.csv(here("NBP.csv"))

# check for NA values, there are none
#map(data, ~sum(is.na(.)))
```

0.1 (a) Make a histogram depicting the distribution of Dnox_masstons.

```
emissions_histogram <- ggplot(data = data, aes(x = Dnox_masstons)) +
  geom_histogram(bins = 50, color = "black", fill = "aquamarine3") +
  ggtitle("Change in Annual Power Plant NOx Emissions (2000-2008)") +
   xlab("NOx Emissions (tons)") +
   ylab("Count") +
```

```
   theme(panel.background = element_blank(),
         axis.title.x = element_text(color = "black", size = 13),
         axis.text.x = element_text(face = "bold", color = "black", size = 15),
         axis.title.y = element_text(color = "black", size = 15),
         axis.text.y = element_text(face = "bold", color = "black", size = 12),
         plot.title = element_text(color="black", size = 15, face = "bold"),
         panel.border = element_rect(colour = "black", fill = NA, size = 2)) +
  scale_y_continuous(breaks = seq(0, 160, by = 20)) +
  scale_x_continuous(breaks = seq(-60, 6, by = 10))

emissions_histogram
```



0.2 (b) Create an indicator =1 if the county has PctBlack above the sample median, and =0 otherwise (in the rest of the assignment, I refer to this variable as 'D'). What is the average of PctBlack for counties above the median, i.e. counties for which D=1?

```
# find median of the PctBlack values
median_pct_blk <- median(data$PctBlack)
median_pct_blk
```

```
## [1] 5
```

```
data_blk_med <- data %>%
  mutate(abv_bel_med_blk = case_when(
    PctBlack > median_pct_blk ~ 1,
```

```
    PctBlack <= median_pct_blk ~ 0))

data_blk_med_summary <- data_blk_med %>%
  group_by(abv_bel_med_blk) %>%
  summarise(mean = mean(PctBlack))

data_blk_med_summary$mean[1]
```

```
## [1] 1.724409
```

```
data_blk_med_summary$mean[2]
```

```
## [1] 19.90909
```

The average of `PctBlack` for counties above the median (D=1) is .

0.3 (c) Estimate a regression of Dnox_masstons on NBP. Interpret the estimated intercept and the coefficient on NBP.

```
nox_nbp_model <- lm_robust(formula = Dnox_masstons ~ NBP, data = data)
nox_nbp_model
```

```
##              Estimate Std. Error   t value     Pr(>|t|)  CI Lower  CI Upper  DF
## (Intercept) -3.615385  0.4207345 -8.593030 1.180567e-16 -4.442081 -2.788689 483
## NBP         -3.908200  0.7957798 -4.911158 1.241163e-06 -5.471818 -2.344582 483
```

```
nox_nbp_model$coefficients[1]
```

```
## (Intercept)
##   -3.615385
```

```
nox_nbp_model$coefficients[2]
```

```
##     NBP
## -3.9082
```

```
nox_nbp_model_table <- tidy(nox_nbp_model)
nox_nbp_model_table %>%
  select(term, estimate, std.error, p.value, conf.low, conf.high) %>%
  kable()
```

term	estimate	std.error	p.value	conf.low	conf.high
(Intercept)	-3.615385	0.4207345	0.0e+00	-4.442081	-2.788689
NBP	-3.908200	0.7957798	1.2e-06	-5.471818	-2.344582

- For the regression of `Dnox_masstons` on `NBP`, the estimated intercept is -3.6153846. This means that when NBP is held constant at 0, meaning no counties were regulated under the NOx Budget Program, the amount of change in annual NOx emissions from all power plants in a county between 2000 and 2008 was -3.6153846 tons. Since this number is negative, we can conclude that annual NOx emissions decreased by that amount when all else is held constant.

- The estimated coefficient on NBP is -3.9082003. This means that when counties were regulated under the NOx Budget Program, the amount of change in annual NOx emissions from all power plants in a county between 2000 and 2008 was -3.9082003 tons. Since this number is negative, we can conclude that annual NOx emissions decreased by that amount when counties were regulated but all else is held constant.

0.4 (d) Create an interaction between the variables NBP and D. Estimate a regression of Dnox_masstons on NBP, D, and this interaction. Interpret each estimated regression coefficient, including the intercept.

```
nox_nbp_d_model <- lm_robust(formula = Dnox_masstons ~ NBP + abv_bel_med_blk + NBP:abv_bel_med_blk, data
nox_nbp_d_model
```

```
##                       Estimate Std. Error   t value     Pr(>|t|)  CI Lower
## (Intercept)         -2.601351  0.4690004 -5.546587 4.814718e-08 -3.522894
## NBP                 -6.332611  1.2158251 -5.208488 2.828147e-07 -8.721596
## abv_bel_med_blk     -2.214649  0.8589630 -2.578282 1.022558e-02 -3.902432
## NBP:abv_bel_med_blk  5.035403  1.5917320  3.163474 1.657634e-03  1.907796
##                       CI Upper  DF
## (Intercept)         -1.6798086 481
## NBP                 -3.9436262 481
## abv_bel_med_blk     -0.5268653 481
## NBP:abv_bel_med_blk  8.1630105 481
```

```
nox_nbp_d_model$coefficients[1]
```

```
## (Intercept)
##   -2.601351
```

```
nox_nbp_d_model$coefficients[2]
```

```
##       NBP
## -6.332611
```

```
nox_nbp_d_model$coefficients[3]
```

```
## abv_bel_med_blk
##       -2.214649
```

```
nox_nbp_d_model$coefficients[4]
```

```
## NBP:abv_bel_med_blk
##            5.035403
```

```
nox_nbp_d_model_table <- tidy(nox_nbp_model)
nox_nbp_d_model_table %>%
  select(term, estimate, std.error, p.value, conf.low, conf.high) %>%
  kable()
```

term	estimate	std.error	p.value	conf.low	conf.high
(Intercept)	-3.615385	0.4207345	0.0e+00	-4.442081	-2.788689
NBP	-3.908200	0.7957798	1.2e-06	-5.471818	-2.344582

Question: does the variables being turned off mean that NO counties were regulated, or are we talking about 1 county hypothetically?

- The intercept is -2.6013514, which represents the change in annual NOx emissions from all power plants in a county between 2000 and 2008 (in tons) when the county was not regulated under the NOx Budget Program and the county's PctBlack is below the sample median, meaning the fraction of the county population that is African American is below the sample median.

- The estimated regression coefficient for NBP is -6.3326109, which represents the change in annual NOx emissions from all power plants in a county between 2000 and 2008 (in tons) when the county is regulated

under the NOx Budget Program and the county's PctBlack is below the sample median, meaning the fraction of the county population that is African American is below the sample median.

- The estimated regression coefficient for D (which is represented by the variable `abv_bel_med_blk` in my code) is -2.2146486, which represents the change in annual NOx emissions from all power plants in a county between 2000 and 2008 (in tons) when the county is not regulated under the NOx Budget Program and the county's PctBlack is above the sample median, meaning the fraction of the county population that is African American is above the sample median.

- The estimated regression coefficient for the interaction between NBP and D (which is represented by the variable `abv_bel_med_blk` in my code) is 5.0354034, which represents the difference in the change in annual NOx emissions from all power plants in a county between 2000 and 2008 (in tons) when the county is regulated under the NOx Budget Program and the county's PctBlack is above the sample median, meaning the fraction of the county population that is African American is above the sample median, compared to the change in annual NOx emissions from all power plants in a county between 2000 and 2008 (in tons) when the county is not regulated under the NOx Budget Program or the county's PctBlack is below the sample median, meaning the fraction of the county population that is African American is below the sample median.

0.5 (e) What is the predicted Dnox_masstons in a county that was not regulated under NBP and where PctBlack is above the sample median (i.e., where D=1)? Report the 95% confidence interval for this prediction. Make sure to use "heteroskedasticity-robust" standard errors.

```
not_reg_abv_blk <- data.frame(NBP = c(0), abv_bel_med_blk = c(1))

CI <- predict(object = nox_nbp_d_model, newdata = not_reg_abv_blk, se.fit = TRUE, interval = "confidenc
CI
```

```
## $fit
##         fit       lwr       upr
## [1,] -4.816 -6.229991 -3.402009
##
## $se.fit
##         1
## 0.7196221
```

The predicted Dnox_masstons in a county that is not regulated under NBP and where PctBlack is above the sample median is -4.816 tons, which represents the change in annual NOx emissions from all power plants in a county between 2000 and 2008 (in tons). The 95% confidence interval for this prediction is [-6.229991, -3.402009].