EDS241: Assignment 2

Desik Somasundaram

02/06/2022

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 gives 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.

1 Clean and plot data

The following code loads and cleans the data.

```
# Load data
NBPdata <- read_excel("NBP.xls", sheet = 1, na = "NA")

# Clean data
NBPdata <-janitor::clean_names(NBPdata)
```

Make a histogram depicting the distribution of dnox_masstons.

```
nbphist<-ggplot(NBPdata, aes(x= NBPdata$dnox_masstons))+
  geom_histogram()+
  labs(x = "Change in annual NOx emissions from all power plants in a county between
2000 and 2008 (in tons)", y = "Count")
```

Figure 1: Change in annual NOx emissions from all power plants in a county between 2000 and 2008 (in tons) in CA

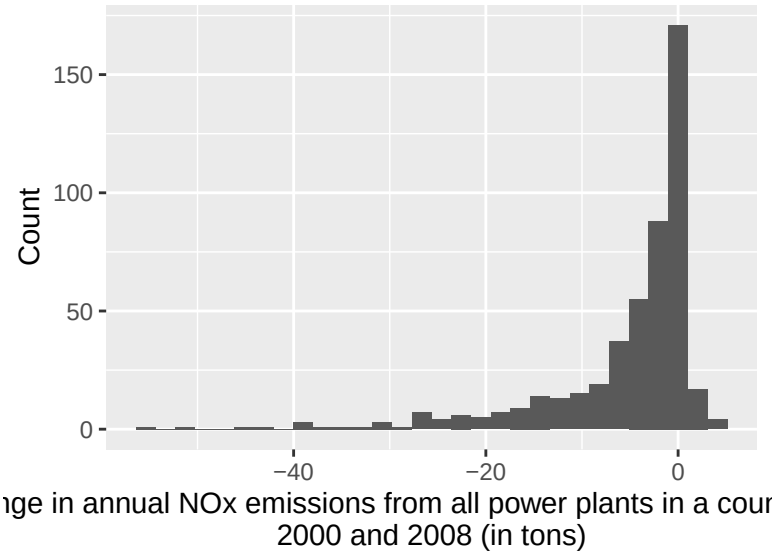

Figure 1 shows a skewed left distribution for the change in annual NOx emissions from all power plants in a county from 2000 to 2008.

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?

```
#compute median percent black
pct_black_median <- median(NBPdata$pct_black)
#create indicator based on median percent black
NBPdata<- NBPdata %>%
  mutate(D = case_when(
    pct_black <= pct_black_median ~ 0,
    pct_black > pct_black_median ~ 1))

#find average percent black for counties above the median
NBPdata_PctBlack_avg <- NBPdata %>%
          filter(D == 1) %>%
          group_by(D) %>%
          summarize(mean(pct_black))
```

The average of PctBlack for counties above the median is ***

D	mean(pct_black)
1	19.3

***.

Estimate a regression of Dnox_masstons on NBP. Interpret the estimated intercept and the coefficient on NBP.

$$Y_i = \beta_0 + \beta_1 X_{1i} + u_i \tag{1}$$

where Y_i is Dnox_masstons i, X_{1i} is the NBP, and u_i the regression error term.

```
model_1 <- lm_robust(formula =  dnox_masstons ~ nbp, data=NBPdata)
```

Table 1 shows the estimated coefficients from estimating equation (1).

	(1)
(Intercept)	-3.622 ***
	(0.420)
nbp	-3.920 ***
	(0.796)
N	485
R2	0.052

*** p < 0.001; ** p < 0.01; * p < 0.05.

In model (1), the estimated β_0 coefficient implies that if the county was **NOT regulated** under the NOx Budget Program and if there is a **decline of 3.622 tons** in the change in annual NOx emissions from all power plants in a county between the 2000 and 2008. The estimated β_1 coefficient implies that if the county was **regulated** under the NOx Budget Program there is **an additional decline of 3.92 tons** in the change in annual NOx emissions from all power plants in a county between 2000 and 2008. Both β_0 and β_1 are ***statistically significant at the 5%***.

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.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} * X_{2i} + u_i \tag{2}$$

where Y_i is Dnox_masstons i, X_{1i} is the NBP, X_{2i} is the D and u_i the regression error term.

In R, we run the following code:

```
model_2 <- lm_robust(formula =  dnox_masstons ~ nbp + D + nbp*D, data=NBPdata)
```

Table 2 shows the estimated coefficients from estimating equation (2).

In model (2), the estimated β_0 coefficient implies that if the county was **NOT regulated** under the NOx Budget Program **AND** the county has PctBlack **below** the sample median, there is a **decline of 2.418 tons** in the change in annual NOx emissions from all power plants in a county between the 2000 and 2008. The estimated β_1 coefficient implies that if the county was **regulated** under the NOx Budget Program **AND** the county has PctBlack **below** the sample median, there is **an additional decline of 7.141 tons** in the change in annual NOx emissions from all power plants in a county between 2000 and 2008. The estimated β_2 coefficient implies that if the county was **NOT regulated** under the NOx Budget Program **AND** the county has PctBlack **above** the sample median, there is **an additional decline of 2.588 tons** in the change in annual NOx emissions from all power plants in a county between 2000 and 2008. The estimated β_3 interaction coefficient implies that if the county was **regulated** under the NOx Budget Program **AND** the county has PctBlack **above** the sample median, the **decline** in the change in annual NOx emissions from all power plants in a county between 2000 and 2008 **gets negated by 6.372 tons**. In other words,

	(1)
(Intercept)	-2.418 ***
	(0.442)
nbp	-7.141 ***
	(1.257)
D	-2.588 **
	(0.853)
nbp:D	6.372 ***
	(1.614)
N	485
R2	0.086

*** p < 0.001; ** p < 0.01; * p < 0.05.

the change in annual NOx emissions increases by 6.372 tons when both NBP and D are equal to 1. β_0, β_1, β_2 and β_3 are ***statistically significant at the 5%***.

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.

```
PredCounty=data.frame(nbp=c(0), D=c(1))
predict(model_2, newdata=PredCounty, se.fit=TRUE, interval='confidence')
```

```
## $fit
##            fit       lwr       upr
## [1,] -5.006106 -6.440065 -3.572147
##
## $se.fit
##         1
## 0.7297841
```

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) is **-5 tons** with with **lower limit of -6.44 tons** and **upper limit of -3.57 tons** for the **95% confidence level**.