EDS 241 Take Home Final

Mia Forsline

3/18/2022

Introduction

We plan to examine the effect of opening a garbage incinerator on housing values in North Andover, Massachusetts during the early 1980s. Construction began in 1981, and the incinerator was operational by 1985. We hypothesize that the placement of an incinerator will decrease the price of houses nearby the incinerator.

Data description

We will use housing price data of houses sold in 1978 (before people knew about the incinerator's construction) and in 1981 (after construction had started). The data used are from the paper: K.A. Kiel and K.T. McClain (1995): "House Prices During Siting Decision Stages: The Case of an Incinerator from Rumor Through Operation," *Journal of Environmental Economics and Management* 28, 241-255.

Data variables:

- `rprice` = inflation-adjusted sales price of the house
- `nearinc` = 1 if the house is located near the incinerator, = 0 otherwise
- `age` = age of the house
- `land` = square footage of the lot
- `area` = square footage of the house
- `rooms` = number of rooms in the house
- `year` indicator = 1978 or 1981

Research question

What is the impact of the opening of a garbage incinerator on housing values in North Andover, MA ?

Set up

Read in the data

```
data <- read_csv(here("data", "KM_EDS241.csv"))

data_clean <- data %>%
  clean_names() %>%
```

```
  mutate(year = as.factor(year),
         age = as.numeric(age),
         rooms = as.numeric(rooms),
         area = as.numeric(area),
         land = as.numeric(land),
         nearinc = as.factor(nearinc),
         rprice = as.numeric(rprice)
  )
```

(a) Using the data for 1981, estimate a simple OLS regression of real house values on the indicator for being located near the incinerator in 1981. What is the house value "penalty" for houses located near the incinerator?

```
data_1981 <- subset(data_clean, year == 1981)

mdl <- lm_robust(rprice ~ nearinc, data = data_1981)
inc_coef <- round(mdl$coefficients[[2]], digits = 2) * -1
```

The house value "penalty" for houses located near the incinerator is approximately $30688.27.

Table 1 shows that, in 1981, on average, houses located near an incinerator had inflation-adjusted sales prices that are approximately $30688.27 lower than houses not located near incinerator.

Table 1: Being near an incinerator lowers housing price

	Dependent variable:
	Housing Price
Incinerator Presence	−30,688.270***
	(6,243.167)
Observations	142
R^2	0.165
Note:	*p<0.1; **p<0.05; ***p<0.01

(a) Does this estimated coefficient correspond to the 'causal' effect of the incinerator (and the negative amenities that come with it) on housing values? Explain why or why not.

This estimated coefficient most likely corresponds to what we would expect the 'causal' effect of the incinerator to be on housing prices. For example, garbage incinerators emit many pollutants and increase ambient pollution nearby. Thus, according to land market theory, housing values are expected to fall. However, due to possible omitted variables bias, it may be difficult to definitively use the estimated coefficient to determine a causal effect of the incinerator on housing prices.

(b) Using the data for 1978, provide some evidence the location choice of the incinerator was not "random", but rather selected on the basis of house values and characteristics. [Hint: in the 1978 sample, are house values and characteristics balanced by nearinc status?]

```
data_1978 <- subset(data_clean, year == 1978)

#Calculate the unadjusted mean difference by hand
data_inc <- subset(data_1978, nearinc == 1)
inc_mean_price <- round(mean(data_inc$rprice), digits = 2)

data_no_inc <- subset(data_1978, nearinc == 0)
no_inc_mean_price <- round(mean(data_no_inc$rprice), digits = 2)

diff <- no_inc_mean_price - inc_mean_price

#Calculate the unadjusted mean difference using a linear regression

mdl3 <- lm_robust(age ~ nearinc, data = data)
age_coef <- round(mdl3$coefficients[[2]], digits = 2)

mdl4 <- lm_robust(rooms ~ nearinc, data = data)
rooms_coef <- round(mdl4$coefficients[[2]], digits = 2) * -1

mdl5 <- lm_robust(area ~ nearinc, data = data)
area_coef <- round(mdl5$coefficients[[2]], digits = 2) * -1
```

In 1978, on average, houses located near the incinerator cost approximately \$63692.86 while houses not located near the incinerator cost approximately \$82517.23. This difference of approximately \$18824.37 demonstrates that the location choice of the incinerator was unlikely to be random.

Furthermore, on average, houses located near the incinerator are approximately 24.03 years older, have approximately 0.72 fewer rooms, and have approximately 312.29 less square footage than houses located far from the incinerator (p-value < 0.05). This also illustrates how the incinerator was likely not placed randomly.

(c) Based on the observed differences in (b), explain why the estimate in (a) is likely to be biased downward (i.e., overstate the negative effect of the incinerator on housing values).

Since houses located near the incinerator are more likely to be older, have fewer rooms, and have less square footage, they are already more likely to have lower housing prices than houses located far from the incinerator regardless of the incinerator's presence. In other words, the groups being compared (houses near the incincerator vs houses far from the incinerator) are not well balanced and thus are likely to overestimate the negative effect of the incinerator on housing values.

(d) Use a difference-in-differences (DD) estimator to estimate the causal effect of the incinerator on housing values without controlling for house and lot characteristics. Interpret the magnitude and sign of the estimated DD coefficient.

```
# DD REGRESSION, Y = Wind+Solar installed capacity (MW), using lm package
DD_cap1 <- lm(formula = rprice ~ nearinc, data=data_clean)
se_DD_cap1 <- starprep(DD_cap1,
                       stat = c("std.error"),
                       se_type = "HC2", alpha = 0.05)
#30 years = 30 time dummies
#48 states = 48 state dummies

DD_cap2 <- lm(formula = rprice ~ nearinc, data=data_clean)
se_DD_cap2 <- starprep(DD_cap2,
                       stat = c("std.error"),
                       se_type = "CR2",
                       clusters=data_clean$year,
                       alpha = 0.05)

se_models <- list(se_DD_cap1[[1]], se_DD_cap2[[1]])
stargazer(DD_cap1, DD_cap2,
          se = se_models,
          keep = c("nearinc"),
          type="text")
```

```
##
## ===============================================================
##                                     Dependent variable:
##                                 ----------------------------
##                                           rprice
##                                     (1)            (2)
## ---------------------------------------------------------------
## nearinc1                         -24,456.640*** -24,456.640***
##                                   (4,418.959)    (5,943.057)
##
## ---------------------------------------------------------------
## Observations                          321            321
## R2                                   0.115          0.115
## Adjusted R2                          0.112          0.112
## Residual Std. Error (df = 319)   31,210.950     31,210.950
## F Statistic (df = 1; 319)        41.317***      41.317***
## ===============================================================
## Note:                              *p<0.1; **p<0.05; ***p<0.01
```

```
# DD REGRESSION, Y = Wind+Solar generation (GWh), using plm package
DD_gen1 <- plm(rprice ~ nearinc,
               index = c("year"),
               model = "within",
               effect = "twoways",
```

```
               data = data_clean)

near_inc_coef <- round(DD_gen1$coefficients[[1]], digits = 2) * -1

# Calculate standard errors (note slightly different procedure with plm package)
se_DD_gen1 <- coeftest(DD_gen1, vcov = vcovHC(DD_gen1, type = "HC2"))[, "Std. Error"]
```

Table 2 shows...

Table 2: Title

	Dependent variable:
	Housing Price
Incinerator Presence	−25,052.490***
	(2.043)
Observations	321
R^2	0.114
Note:	*p<0.1; **p<0.05; ***p<0.01

Without controlling for housing or lot characteristics, the estimated DD coefficient of `nearinc` tells us that, on average, houses located near an incinerator are $25052.49 cheaper than houses located far from an incinerator.

(e) Report the 95% confidence interval for the estimate of the causal effect on the incinerator in (d).

```
#use the plm::tidy() function to summarize information about the components of a plm model
conf <- tidy(DD_gen1, conf.int = TRUE, conf.level = 0.95)
conf_low <- round(conf$conf.low, digits = 2)
conf_high <- round(conf$conf.high, digits = 2)

confint(DD_gen1)
```

```
##              2.5 %    97.5 %
## nearinc1 -36618.02 -13486.96
```

The lower bound of the confidence interval is -36618.02, and the upper bound of the confidence interval is -13486.96.

(f) How does your answer in (d) changes when you control for house and lot characteristics?

```
# DD REGRESSION, Y = Wind+Solar generation (GWh), using plm package
DD_gen2 <- plm(rprice ~ nearinc + age + rooms + area + land + year,
               index = c("year"),
               model = "within",
               effect = "twoways",
               data = data_clean)

near_inc_coef2 <- round(DD_gen2$coefficients[[1]], digits = 2) * -1

# Calculate standard errors (note slightly different procedure with plm package)
se_DD_gen2 <- coeftest(DD_gen2, vcov = vcovHC(DD_gen2, type = "HC2"))[, "Std. Error"]
```

Table 3 shows...

Table 3: Title

	Dependent variable:
	Housing Price
Incinerator Presence	$-8{,}878.357^{***}$
	(37.077)
age	-202.309^{***}
	(1.004)
rooms	$8{,}500.440^{***}$
	(50.379)
area	23.377^{***}
	(0.008)
land	0.021^{***}
	(0.003)
Observations	321
R^2	0.586
Note:	*p<0.1; **p<0.05; ***p<0.01

(f) Test the hypothesis that the coefficients on the house and lot characteristics are all jointly equal to 0.

```
hyp <- linearHypothesis(DD_gen2, c("age=0", "rooms=0", "area=0", "land=0"), white.adjust = "hc2")
p <- hyp$`Pr(>Chisq)`[[2]]
```

0

(g) Using the results from the DD regression in (f), calculate by how much did real housing values change on average between 1978 and 1981.

(h) Explain (in words) what is the key assumption underlying the causal interpretation of the DD estimator in the context of the incinerator construction in North Andover.