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` indicator = 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

- double check that each variable is the appropriate class

```
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
```

In 1981, the house value "penalty" for houses located near the incinerator was 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 of our simple model 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?]

```
#subset the data to only 1978 data
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 of various house/lot characteristics using a linear regressio
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.

Moreover, on average, houses located near the incinerator were approximately 24.03 years older, had approximately 0.72 fewer rooms, and had approximately 312.29 less square footage than houses located far from the incinerator (p-value < 0.05). The statistically significant differences in houses near or not near the incinerator further illustrate 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 (p-value < 0.05), 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 control group (houses far from the incinerator) and the treatment group (houses near the incinerator) being compared are not well balanced. Thus, we 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 using the `lm` package
DD_cap1 <- lm(formula = rprice ~ nearinc + year , data=data_clean)
se_DD_cap1 <- starprep(DD_cap1,
                       stat = c("std.error"),
                       se_type = "HC2", alpha = 0.05)

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

near_inc_coef <- round(DD_cap1$coefficients[[2]], digits = 2) * -1
```

Table 2 compares two methods of calculating standard errors of an `lm()` model estimating the effect of being near an incinerator on housing prices in North Andover, MA. The HC2 method does not acknowledge that housing price observations are grouped by year (either 1978 or 1981). As a result, the standard error values for the HC2 method are lower than the CR2 method that clusters the standard errors based on year. No matter the standard error calculation method, the model also demonstrates that being near an incinerator, on average, decreases housing prices and that housing prices appreciate over time.

Table 2: HC2 vs CR2 Standard Errors

	Dependent variable:	
	Housing Price	
	(1)	(2)
Incinerator Presence	−23,896.000***	−23,896.000***
	(4,363.427)	(5,869.232)
Year 1981	15,290.320***	15,290.320***
	(3,403.823)	(182.880)
Observations	321	321
R^2	0.167	0.167
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 and holding the year fixed, houses located near an incinerator are $23896 cheaper than houses

located far from an incinerator. In short, the sign of the DD estimator is negative, and the magnitude is 23896.

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

```
#use lm_robust() model to calculate confidence intervals
DD_cap1 <- lm_robust(formula = rprice ~ nearinc + year , data=data_clean)

conf <- confint(DD_cap1)
nearinc_conf <- conf[2,]

conf_low <- round(nearinc_conf[[1]], digits = 2)
conf_high <- round(nearinc_conf[[2]], digits = 2)
```

The lower bound of the confidence interval is -32480.83, and the upper bound of the confidence interval is -15311.16.

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

```
DD_cap3 <- lm(formula = rprice ~ nearinc + year + age + rooms + area + land, data=data_clean)
se_DD_cap3 <- starprep(DD_cap3,
                       stat = c("std.error"),
                       se_type = "HC2", alpha = 0.05)

DD_cap4 <- lm(formula = rprice ~ nearinc + year + age + rooms + area + land, data=data_clean)
se_DD_cap4 <- starprep(DD_cap4,
                       stat = c("std.error"),
                       se_type = "CR2",
                       clusters=data_clean$year,
                       alpha = 0.05)

near_inc_coef2 <- round(DD_cap3$coefficients[[2]], digits = 2) * -1
year_coef <- round(DD_cap3$coefficients[[3]], digits = 2)
```

Table 3 shows that when controlling for house and lot characteristics, the presence of the garbage incinerator no longer has a statistically significant causal effect on housing prices in North Andover, MA. Instead, housing prices only change significantly over time.

Table 3: Incinerator presence does not predict housing prices

	Dependent variable:	
	Housing Price	
	(1)	(2)
Incinerator Presence	−2,604.816	−2,604.816
	(4,363.427)	(5,869.232)
Year - 1981	9,019.277***	9,019.277***
	(3,403.823)	(182.880)
House Age	−260.659	−260.659
Number of Rooms	6,593.785	6,593.785
House Square Footage	24.293	24.293
Lot Size	0.120	0.120
Observations	321	321
R^2	0.604	0.604
Note:	*p<0.1; **p<0.05; ***p<0.01	

When controlling for housing and lot characteristics (`age`, `rooms`, `area`, and `land`), the incinerator indicator coefficient 2604.82 decreases in magnitude and is no longer statistically significant. Thus, the incinerator indicator no longer demonstrates a causal effect on housing prices.

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

```
DD_cap3 <- lm_robust(formula = rprice ~ nearinc + year + age + rooms + area + land, data=data_clean)
hyp <- linearHypothesis(DD_cap3, c("age=0", "rooms=0", "area=0", "land=0"), white.adjust = "hc2")
p <- hyp$`Pr(>Chisq)`[[2]]
```

The p-value $= 0 < 0.05$, so we can reject the null hypothesis and conclude that the house and lot characteristics are all jointly equal to zero.

(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.

```
DD_cap3 <- lm_robust(formula = rprice ~ nearinc + year + age + rooms + area + land, data=data_clean)
```

Holding all other variables fixed, from 1978 to 1981, housing values increased by $9019.28.

(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.

The key assumption is the parallel trends assumption, which says the treatment group D (houses near the incinerator) and the control group (houses not near the incinerator) would have the same trends in their outcomes (change housing prices at the same rate over time) regardless of the treatment (the placement of the incinerator). In other words, the control group provides a valid counterfactual for the temporal evolution of the mean outcomes in the treatment group in the absence of a change in treatment.