

EDS241: Take-Home Final Exam

Halina Do-Linh

03/18/2022

The question for this take-home final exam asks you to examine the impact of the opening of a garbage incinerator on housing values in North Andover, MA. The data for the exercise are a subset of the data in 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.

The key variables for the analysis are:

- `rprice` (inflation-adjusted sales price of house)
- `nearinc` (=1 if house 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).

These variables are contained in the CSV file `KM_EDS241.csv`.

Load Data

```
incinerator_data <- read_csv(here::here("assignments/data/KM_EDS241.csv")) %>%  
  # made nearinc and year factors  
  mutate(nearinc = as.factor(nearinc),  
         year = as.factor(year))
```

Question A

```
# filter for 1981  
incinerator_1981 <- incinerator_data %>%  
  filter(year == 1981)  
  
# simple OLS regression  
model_1 <- lm_robust(data = incinerator_1981,  
                    formula = rprice ~ nearinc)  
  
huxreg(model_1)
```

	(1)
(Intercept)	101307.515 *** (2944.810)
nearinc1	-30688.274 *** (6243.167)
N	142
R2	0.165

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

Answer: The house value “penalty” for houses located near the incinerator is -30688.27. This estimated coefficient of `nearinc` does correspond to the “casual” effect of the incinerator because we expect houses near the incinerator to negatively affect housing values, but this effect may be exaggerated due to omitted variables bias.

Question B

```
# filter for 1978
incinerator_1978 <- incinerator_data %>%
  filter(year == 1978)

# df of means of house values and characteristics nearinc(0) vs nearinc(1)
incinerator_diff <- incinerator_1978 %>%
  group_by(nearinc) %>%
  summarize(mean_rprice = mean(rprice),
             mean_age = mean(age),
             mean_land = mean(land),
             mean_area = mean(area),
             mean_room = mean(rooms))

# linear regression of other covariates regressed by nearinc to show OVB
model_2 <- lm_robust(age ~ nearinc, data = incinerator_1978)
model_3 <- lm_robust(land ~ nearinc, data = incinerator_1978)
```

Answer: `model_2` and `model_3` provide some evidence that the location choice of the incinerator was not “random” and was instead selected based on the basis of house values and characteristics. This is because both models show coefficients that are non-zero and are statistically significant, and so there is a correlation between `nearinc` and some house values and characteristics. Omitted variable bias could be occurring and other variables may be acting on `rprice` in addition to `nearinc`.

Question C

Answer: Because houses that are near incinerators are of lower value to begin with. Houses near incinerators have lower value, are older, and the square footage of the house lot and area is smaller compared to houses not near incinerators. This means there is an overestimated negative effect of the incinerator on housing values.

Question D

```
# diff in diff estimator using regression
DD_inc <- lm_robust(data = incinerator_data,
                    formula = rprice ~ nearinc + year)

huxreg(DD_inc)
```

	(1)
(Intercept)	84103.881 *** (1986.559)
nearinc1	-23895.996 *** (4363.427)
year1981	15290.315 *** (3403.823)
N	321
R2	0.167

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

Answer: The estimated DD coefficient has a negative sign, which means the the effect is a decrease. The magnitude is 23896, which means on average housing prices for houses near incinerators will change by that magnitude, while holding time constant.

Question E

```
# 95% confidence interval
DD_inc_ci <- confint(DD_inc)
```

Answer: -32480.83 and -15311.16.

Question F

```
model_4 <- lm_robust(data = incinerator_data,
                    formula = rprice ~ nearinc + year + age + rooms + area + land)

linearHypothesis(model_4,
                  c("age = 0",
                    "rooms = 0",
                    "area = 0",
                    "land = 0"),
                  white.adjust = "hc2")
```

Res.Df	Df	Chisq	Pr(>Chisq)
318			
314	4	135	3.85e-28

Answer: `nearinc` is no longer statistically significant which tells us that the model in question D was overestimating the effect of a house being near an incinerator now that we are considering other variables.

With a p-value of 0.00000000000000022 we can reject the null hypothesis that the coefficients on the house and lot characteristics are all jointly equal to 0.

Question G

```
mod4_ht <- huxreg(model_4)
restack_across(mod4_ht, 7)
```

	(1)		(1)		(1)
(Intercept)	-14144.356	age	-260.659 ***	land	0.120
	(10765.286)		(50.524)		(0.135)
nearinc1	-2604.816	rooms	6593.785 ***	N	321
	(5819.305)		(1547.520)	R2	0.604
year1981	9019.277 ***	area	24.293 ***	*** p < 0.001; ** p < 0.01; * p < 0.05.	
	(2291.266)		(3.993)		

Answer: The real housing values on average between 1978 and 1981 increased by 9019.28, holding everything else constant.

Question H

Answer: Homes not near an incinerator are a valid counterfactual for the temporal evolution of the mean outcomes in homes that are near an incinerator in absence of a change in homes that are near an incinerator. This is the parallel trend assumption.