EDS241: Take Home Finalr

Clarissa Boyajian

03/17/2022

1 Final Assignment

In this final assignment I will examine the impact of the opening of a garbage incinerator on housing values in North Andover, MA. The data for this assignment are a subset of the data in the paper: K.A. Kiel and K.T. McClain (1995); "Housing Prices During Siting Decision Stages: The Case of an Incinerator from Rumor Through Operation," *Journal of Environmental Economics and Management* 28, 241-255.

1.1 Load and clean data

The following code loads and cleans the data.

```
data_house_values <- read.csv(here::here("KM_EDS241.csv"))
```

1.2 Question 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? 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.

The code chunk estimates a simple OLS regression of real house values on the indicator for being located near the incinerator in 1981:

```
data_house_values_1981 <- data_house_values %>%
  filter(year == 1981)

model_a <- lm_robust(data = data_house_values_1981,
                     formula = rprice ~ nearinc)

model_a_table <- huxreg(model_a)
restack_across(model_a_table, 5)
```

	(1)		(1)
(Intercept)	101307.515 ***	N	142
	(2944.810)	R2	0.165
nearinc	-30688.274 ***	*** p < 0.001; ** p < 0.01; * p < 0.05.	
	(6243.167)		

Answer The "penalty" for houses located near the incinerator is a decrease of $30688.27 in the inflation-adjusted sales price. The estimated coefficient corresponds partially to the "causal" effect of the incinerator on housing values. However, housing prices are likely lower in the area where the incinerator was more likely to be sited, meaning that this OLS likely suffers from omitted variable bias.

1.3 Question B

Using the data for 1978, provide some evidence the location 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?]

The code chunk below calculates the average inflation-adjusted sales price, age, square footage of lot, square footage of house, and number of rooms for all houses located near the incinerator and all houses not located near the incinerator:

```
data_house_values_1978 <- data_house_values %>%
  filter(year == 1978) %>%
  group_by(nearinc) %>%
  summarise(mean_rprice = mean(rprice),
            mean_rooms = mean(rooms),
            mean_area = mean(area),
            mean_land = mean(land),
            mean_age = mean(age))

huxtable(data_house_values_1978)
```

nearinc	mean_rprice	mean_rooms	mean_area	mean_land	mean_age
0	8.25e+04	6.83	2.07e+03	5.26e+04	12.7
1	6.37e+04	6.04	1.83e+03	2.18e+04	39.8

Answer Houses that were not located near the incinerator on average had higher sale prices (by $18824.37), number of rooms (by 0.79), square footage of land (by 30729.13 ft^2), and square footage of house (by 240.11 ft^2), and on average were younger (by -27.04 years). This shows that the incinerator location was chosen to be near the "less desirable" homes.

1.4 Question 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).

Answer The difference in question (b) show that the various housing characteristics are likely omitted variables from the model in question (a) because they are correlated to whether a home is near the incinerator or not.

1.5 Question D

Use a difference-in-difference (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.

The code chunk below calculates the causal effect of the incinerator on housing values:

```
model_d <- lm(data = data_house_values,
              formula = rprice ~ nearinc + as.factor(year))

model_d_table <- huxreg(model_d)
restack_across(model_d_table, 7)
```

	(1)		(1)
(Intercept)	84103.881 ***	N	321
	(2544.127)	R2	0.167
nearinc	-23895.996 ***	logLik	-3766.506
	(3697.775)	AIC	7541.011
as.factor(year)1981	15290.315 ***	*** p < 0.001; ** p < 0.01; * p < 0.05.	
	(3408.762)		

Answer The estimated DD coefficient is $-23896. This means that the incinerator had a negative impact on housing values, with the value dropping on average by $23896 between 1978 and 1981.

1.6 Question E

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

The code chunk below calculates the 95% confidence interval for the estimate of the causal effect on the incinerator in (d):

```
model_e <- lm_robust(data = data_house_values,
              formula = rprice ~ nearinc + as.factor(year))
```

Answer There is 95% confidence that the interval (-32480.83, 80195.42) contains the true average causal effect of the incinerator on housing values.

1.7 Question F

How does your answer in (d) change when you control for house and lot characteristics? Test the hypothesis that the coefficients on the house and lot characteristics are all jointly equal to 0.

The code chunk below calculates the causal effect of the incinerator on housing values with controls for housing and lot characteristic:

```
model_f <- lm(data = data_house_values,
              formula = rprice ~ nearinc + as.factor(year) + rooms + area + land + age)

model_f_table <- huxreg(model_f)
restack_across(model_f_table, 9)
```

	(1)		(1)		(1)
(Intercept)	-14144.356	area	24.293 ***	logLik	-3647.262
	(9857.979)		(2.058)	AIC	7310.524
nearinc	-2604.816	land	0.120 ***	*** p < 0.001; ** p < 0.01; * p < 0.05.	
	(3008.760)		(0.032)		
as.factor(year)1981	9019.277 ***	age	-260.659 ***		
	(2450.197)		(38.756)		
rooms	6593.785 ***	N	321		
	(1636.364)	R2	0.604		

The code chunk below tests the hypothesis that all coefficients on the house a lot characteristics are jointly equal to 0:

```
model_f_hyp_test <- linearHypothesis(model = model_f,
                                     c("rooms = 0", "area = 0", "land = 0", "age = 0"),
                                     white.adjust = "hc2")
huxtable(model_f_hyp_test)
```

Res.Df	Df	F	Pr(>F)
318			
314	4	33.7	2.21e-23

Answer When controlling for housing and lot characteristics, the causal effect of the incinerator on housing values is reduced to $-2604.82 and is no longer statistically significant. This indicates that the original coefficient overestimated the negative effect of the incinerator on housing values due to omitted variable bias. We can reject the null hypothesis that the coefficients on housing and lot characteristics are jointly equal to 0 because the p-value of the linear hypothesis test is 0.00000000000000000000002206466, which is statistically significant.

1.8 Question 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.

Answer On average the real housing values increased by $9019.28 between 1978 and 1981.

1.9 Question 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.

Answer The key assumption underlying the causal interpretation of the DD estimator is the parallel trend assumption. Houses that are not near the incinerator are a valid counterfactual (or control) for houses near the incinerator (or treatment) for the temporal change in the mean outcomes in absence of a change in treatment. A change in treatment is not possible as homes do not move around from year to year.