EDS 241: Final Exam

Paloma Cartwright

2022-03-16

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.

Background: The construction of a new garbage incinerator in North Andover in the early 1980s was controversial due to the increases in ambient pollution that it would create. Rumors of the incinerator began after 1978. The construction started in 1981, and the incinerator began operating in 1985. In Economics, land market theory suggests that local amenities are capitalized in housing values, and predicts that the prices of houses located near the incinerator would fall compared to the price of houses located further away from the incinerator. By 1981, you can assume that all market participants had full information on the upcoming garbage incinerator, so that housing values had capitalized the upcoming arrival of the incinerator.

Data: The authors of the paper collected data on prices of houses that sold in 1978 (before the upcoming construction of the incinerator was public knowledge) and in 1981 (after the construction had started). 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), and
- a year indicator (1978 or 1981).

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

(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.

```
data_1981 <- data %>%
 filter(year == 1981)
model1 <- lm_robust(rprice ~ nearinc, data = data_1981)</pre>
summary(model1)
##
## Call:
## lm_robust(formula = rprice ~ nearinc, data = data_1981)
##
## Standard error type: HC2
##
##
  Coefficients:
##
             Estimate Std. Error t value
## (Intercept)
               101308
                           2945 34.402
## nearinc1
               -30688
                           6243 -4.915
##
                                                                          Pr(>|t|)
## nearinc1
             0.0000024423503623929663697744892048024922814875026233494281768798828125000\\
##
             CI Lower CI Upper DF
## (Intercept)
                95485
                       107130 140
                       -18345 140
## nearinc1
               -43031
##
## Multiple R-squared: 0.1653,
                                Adjusted R-squared: 0.1594
## F-statistic: 24.16 on 1 and 140 DF, p-value: 0.000002442
```

The penalty for houses being located near the incinerator is 30688.2737598. This coefficient does correspond to the causal effect of the incinerator because it is demonstrating that house values will significantly decrease if a house a near an incinerator compared to a house that is not. There is a possibility that omitted variables are exaggerating this causal relationship.

(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 nearing status?]

Near Incinerator	Average House Price
0	8.25e + 04
1	6.37e + 04

```
mean_diff <- avg_house_price_78$`Average House Price`[1] -
    avg_house_price_78$`Average House Price`[2]

mean_diff

## [1] 18824.37

model2 <- lm_robust(age ~ nearinc, data = data)

model3 <- lm_robust(area ~ nearinc, data = data)

model4 <- lm_robust(rooms ~ nearinc, data = data)</pre>
```

The mean difference between average house prices for those near the incinerator versus not is 18824.37. This shows that it is likely that the incinerator location choice was not random. Additionally, houses that are near an incinerator are on average 24.0319444 years older than those that are not near the incinerator. The square footage of houses near incinerators are also on average 312.2911111 less than those further away. Finally, there are on average 0.7166667 less rooms in houses near incinerators. All of these things show that the placement of the incinerators was likely not random.

(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).

Because of the fact that the houses near the incinerator are older, smaller and have less rooms, they will already have a lower value, regardless of the placement of the incinerator so this creates an overestimation of the negative impact of the incinerators 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.

```
DD1 <- lm_robust(formula = rprice ~ nearinc + year, data = data)
summary(DD1)</pre>
```

```
##
## Call:
## lm_robust(formula = rprice ~ nearinc + year, data = data)
##
## Standard error type: HC2
##
## Coefficients:
##
         Estimate Std. Error t value
           84104
                   1987
                       42.336
## (Intercept)
           -23896
                   4363
                       -5.476
## nearinc1
           15290
                   3404
                        4.492
## year1981
##
## nearinc1
         ## year1981
         ##
         CI Lower CI Upper DF
                 88012 318
## (Intercept)
           80195
## nearinc1
           -32481
                 -15311 318
            8593
                 21987 318
## year1981
##
## Multiple R-squared: 0.1674,
                       Adjusted R-squared: 0.1621
## F-statistic: 21.15 on 2 and 318 DF, p-value: 0.000000002379
```

```
# DD1 <- plm(rprice ~ nearinc,
# index = c("year"),
# model = "within",
# effect = "twoways", data = data)
# summary(DD1)</pre>
```

The estimated DD coefficient is -23896. This means that on average for houses that are near the incinerator, their value decreases (negative sign) by 23896 (magnitude).

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

```
confint <- confint(DD1)
conf_nearinc <- confint[2,]
low <- round(conf_nearinc[[1]], 2)
high <- round(conf_nearinc[[2]], 2)</pre>
```

The 95% confidence interval for the estimate of the causal effect on the incinerator is [-32480.83, -15311.16].

(f) How does your answer in (d) changes 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.

```
## lm_robust(formula = rprice ~ nearinc + year + age + rooms + area +
##
##
## Standard error type:
##
## Coefficients:
                  Estimate Std. Error t value
                                                     Pr(>|t|)
                                                                  CI Lower
##
## (Intercept) -14144.3562 10765.2862 -1.3139 0.189843745252 -35325.5703
                            5819.3055 -0.4476 0.654738768621 -14054.5772
## nearinc1
                -2604.8161
## year1981
                 9019.2767
                            2291.2664 3.9364 0.000101916484
                                                                 4511.1007
                 -260.6588
                               50.5237 -5.1591 0.000000440517
                                                                 -360.0667
## age
## rooms
                 6593.7854
                            1547.5197
                                       4.2609 0.000026950500
                                                                 3548.9666
## area
                   24.2933
                                3.9928 6.0843 0.000000003402
                                                                   16.4372
## land
                    0.1197
                                0.1349   0.8878   0.375327708821
                                                                   -0.1456
##
                 CI Upper DF
## (Intercept)
                7036.8580 314
                8844.9450 314
## nearinc1
## year1981
               13527.4528 314
## age
                -161.2509 314
                9638.6042 314
## rooms
                  32.1493 314
## area
## land
                   0.3851 314
```

```
##
## Multiple R-squared: 0.6039 , Adjusted R-squared: 0.5963
## F-statistic: 89.07 on 6 and 314 DF, p-value: < 0.00000000000000022
nearinc_coeff <- DD2$coefficients[[2]]</pre>
```

The coefficient for nearinc is now -2604.8160767 which has decreased significantly from the model which only included nearinc and year as variables. This new coefficient is also no longer statistically significant which means that it is no longer a good indicator for the price of homes. This means that being near an indicator does not have as much of an effect on housing prices as the other now included variables do.

```
linHyp1 <- linearHypothesis(DD2, c("age=0", "rooms=0", "area=0", "land=0"), white.adjust = "hc2")
linHyp1$^Pr(>Chisq)^[2]
```

[1] 0.0000000000000000000000000003851232

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

```
year_coeff <- DD2$coefficients[[3]]</pre>
```

Holding everything else constant, from 1978 to 1981 housing values increase 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.