EDS 241: Final Exam

Paloma Cartwright

2022-03-15

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`.

```
data <- read_csv(here("KM_EDS241.csv")) %>%
  mutate(nearinc = as.factor(nearinc),
         year = as.factor(year))
```

```
## Rows: 321 Columns: 7
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (7): year, age, rooms, area, land, nearinc, rprice
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

(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)
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|)
## (Intercept) 0.000000000000000000000000000000000000000000000000000000000000000000003633
## nearinc1    0.000002442350362392966369774489204802492281487502623349428176879882812500
##             CI Lower CI Upper  DF
## (Intercept)    95485   107130 140
## nearinc1      -43031   -18345 140
##
## 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 30688.274. 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 `nearinc` status?]

```
data_1978 <- data %>%
  filter(year == 1978)
```

```
avg_house_price_78 <- data_1978 %>%
  group_by(nearinc) %>%
  summarize(avg_price = round(mean(rprice), 2)) %>%
  rename("Near Incinerator" = "nearinc",
         "Average House Price" = "avg_price")
avg_house_price_78
```

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

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

mean_diff
```

```
## [1] 18824.37
```

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

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

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

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.

```r
# DD_cap1 <- lm_robust(formula = rprice ~ nearinc, data=data)
# summary(DD_cap1)

DD1 <- plm(rprice ~ nearinc,
               index = c("year"),
               model = "within",
               effect = "twoways", data = data)

summary(DD1)
```

```
## Twoways effects Within Model
##
## Call:
## plm(formula = rprice ~ nearinc, data = data, effect = "twoways",
##     model = "within", index = c("year"))
##
## Unbalanced Panel: n = 2, T = 142-179, N = 321
##
## Residuals:
##    Min. 1st Qu.  Median 3rd Qu.    Max.
##  -74834  -12152       0   12152   74834
##
## Coefficients:
##          Estimate Std. Error t-value  Pr(>|t|)
## nearinc1 -25052.5     5900.9 -4.2455 0.0000395 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    119120000000
## Residual Sum of Squares: 105530000000
## R-Squared:      0.11406
## Adj. R-Squared: -1.025
## F-statistic: 18.0247 on 1 and 140 DF, p-value: 0.000039498
```

```
# these two should produce the same thing if we added year in the lm_robust one above. So we should dis
```

The estimated DD coefficient is -25052.4883116. This means that for houses that are near the incinerator, their value decreases (negative sign) by 25052.4883116 (magnitude).

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

```
low <- confint(DD1)[1]
high <- confint(DD1)[2]
```

The 95% confidence interval for the estimate of the causal effect on the incinerator is [-36618.0177002, -13486.9589231].

(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.

```
DD2 <- plm(rprice ~ nearinc + age + rooms + area + land,
               index = c("year"),
               model = "within",
               effect = "twoways",
               data = data)
summary(DD2)
```

```
## Twoways effects Within Model
##
## Call:
## plm(formula = rprice ~ nearinc + age + rooms + area + land, data = data,
##     effect = "twoways", model = "within", index = c("year"))
##
## Unbalanced Panel: n = 2, T = 142-179, N = 321
##
## Residuals:
##     Min.  1st Qu.   Median  3rd Qu.     Max.
## -49958.9  -6537.7      0.0   6537.7  49958.9
##
## Coefficients:
##              Estimate   Std. Error t-value           Pr(>|t|)
## nearinc1 -8878.357469  4413.997215 -2.0114          0.0462583 *
## age       -202.309147    51.541679 -3.9252          0.0001371 ***
## rooms     8500.439559  2088.463008  4.0702 0.0000792649908328 ***
## area        23.376524     2.918406  8.0100 0.0000000000004591 ***
## land         0.020809     0.046259  0.4498          0.6535404
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    119120000000
## Residual Sum of Squares: 49347000000
## R-Squared:      0.58573
## Adj. R-Squared: 0.025251
```

```
## F-statistic: 38.4579 on 5 and 136 DF, p-value: < 0.000000000000000222
```

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

```
## [1] 0.00000000000000000000000000000001864039
```

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.