

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`.

```
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

```
data_1981 <- data %>%
  filter(year == 1981)

model1 <- lm_robust(rprice ~ nearinc, data = data_1981)
summary(model1)
```

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 is near an incinerator compared to a house that is not. There is a possibility that omitted variables are exaggerating this causal relationship.

- ```
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
```

2

```
mean_diff
```

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

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

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

summary(DD1)
```

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)
```

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.

```
DD2 <- lm_robust(rprice ~ nearinc + year + age + rooms + area + land,
 data = data)
```

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

summary(DD2)
```

```
##
Call:
lm_robust(formula = rprice ~ nearinc + year + age + rooms + area +
land, data = data)
##
Standard error type: HC2
##
Coefficients:
Estimate Std. Error t value Pr(>|t|) CI Lower
(Intercept) -14144.3562 10765.2862 -1.3139 0.189843745252 -35325.5703
nearinc1 -2604.8161 5819.3055 -0.4476 0.654738768621 -14054.5772
year1981 9019.2767 2291.2664 3.9364 0.000101916484 4511.1007
age -260.6588 50.5237 -5.1591 0.000000440517 -360.0667
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
nearinc1 8844.9450 314
year1981 13527.4528 314
age -161.2509 314
rooms 9638.6042 314
area 32.1493 314
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]]
```

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]
```

[illegible]

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_coef <- DD2$coefficients[[3]]
```

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.