EDS241: Take Home Final

Alexandra Yousefivand

03/17/2022

```r
# Load data
data_raw <- read_csv(here("KM_EDS241.csv"))

data <- data_raw %>%
  mutate(year = as.factor(year),
         nearinc = as.factor(nearinc))
```

1 (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.

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

```r
model <- lm_robust(formula = rprice ~ nearinc, data = data_1981)
summary(model)
```

```
##
## 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) 0.0000000000000000000000000000000000000000000000000000000000000000000003633
## nearinc1    0.0000024423503623929663697202790961782170597871299833059310913085937500000
##             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
```

University of California, Santa Barbara **Olivier Deschenes**
Bren School of Env. Science & Management **Winter 2022**

EDS 241
Take Home Final

Due by 5:00 pm on 3/18/22
Return your R-markdown document through Gauchospace
(including signed cover sheet)

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.

Use a R-markdown document like in the assignments to produce your answer sheet.

Pledge of honor:
By taking this exam you are pledging to work alone on the exercises. Slack or email me at olivier@econ.ucsb.edu for any clarifying questions. I do reserve the right to decline answering some questions.

Name: Alexandra Yousefivand

Signature: 

Figure 1: Take Home signed cover sheet

```
penalty <- abs(round(model$coefficients[2]))
```

The house value "penalty" for houses located near the incinerator is 30688; in other words, on average, houses near the incinerator cost $30688 less than houses not near the incinerator. The very low p-value indicates that this is a statistically significant result and this estimated coefficient correlates with price. This might correspond to the 'causal' effect of the incinerator; however, there are other variables that may contribute to the difference in housing prices as well, which implies the possibility of omitted variables bias.

2 (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?]

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

```
data_nearinc <- data_1978 %>% filter(nearinc == 1)
data_not_nearinc <- data_1978 %>% filter(nearinc == 0)
```

```
# unadjusted mean difference
nearinc_mean_price <- mean(data_nearinc$rprice)

not_nearinc_mean_price <- mean(data_not_nearinc$rprice)

difference_price <- not_nearinc_mean_price - nearinc_mean_price
difference_price
```

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

Houses near the incinerator cost, on average, $18824 less than houses not near the incinerator.

```
# unadjusted mean difference
nearinc_mean_age <- mean(data_nearinc$age)

not_nearinc_mean_age <- mean(data_not_nearinc$age)

difference_age <- not_nearinc_mean_age - nearinc_mean_age
difference_age
```

```
## [1] -27.03775
```

Houses near the incinerator are, on average, 27 years older than houses not near the incinerator.

```r
# unadjusted mean difference
nearinc_mean_rooms <- mean(data_nearinc$rooms)

not_nearinc_mean_rooms <- mean(data_not_nearinc$rooms)

difference_rooms <- not_nearinc_mean_rooms - nearinc_mean_rooms
difference_rooms
```

```
## [1] 0.793554
```

Houses near the incinerator have, on average, 0.79 fewer rooms than houses not near the incinerator.

```r
# unadjusted mean difference
nearinc_mean_area <- mean(data_nearinc$area)

not_nearinc_mean_area <- mean(data_not_nearinc$area)

difference_area <- not_nearinc_mean_area - nearinc_mean_area
difference_area
```

```
## [1] 240.1132
```

Houses near the incinerator have, on average, 240 less square footage (of the house) than houses not near the incinerator.

```r
# unadjusted mean difference
nearinc_mean_land <- mean(data_nearinc$land)

not_nearinc_mean_land <- mean(data_not_nearinc$land)

difference_land <- not_nearinc_mean_land - nearinc_mean_land
difference_land
```

```
## [1] 30729.13
```

Houses near the incinerator have, on average, 30729 less square footage (of the lot) than houses not near the incinerator.

```r
# unadjusted mean difference using linear regression
model_age   <- lm_robust(formula = age ~ nearinc, data = data_1978)
summary(model_age)
```

```
##
## Call:
## lm_robust(formula = age ~ nearinc, data = data_1978)
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error t value    Pr(>|t|) CI Lower CI Upper  DF
```

```
## (Intercept)    12.75      3.227   3.951 0.000112327     6.38    19.12 177
## nearinc1       27.04      5.759   4.695 0.000005329    15.67    38.40 177
##
## Multiple R-squared:  0.1106 ,    Adjusted R-squared:  0.1055
## F-statistic: 22.04 on 1 and 177 DF,  p-value: 0.000005329
```

```r
model_rooms <- lm_robust(rooms ~ nearinc, data = data_1978)
summary(model_rooms)
```

```
##
## Call:
## lm_robust(formula = rooms ~ nearinc, data = data_1978)
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error t value
## (Intercept)   6.8293    0.07183  95.081
## nearinc1     -0.7936    0.15895  -4.992
##
## (Intercept) 0.0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
## nearinc1    0.000001419911976227534418978221131268924182222690433263778686523437500000000000000000000000000000000000000
##             CI Lower CI Upper  DF
## (Intercept)    6.688   6.9710 177
## nearinc1      -1.107  -0.4799 177
##
## Multiple R-squared:  0.1481 ,    Adjusted R-squared:  0.1433
## F-statistic: 24.92 on 1 and 177 DF,  p-value: 0.00000142
```

```r
model_area  <- lm_robust(area ~ nearinc, data = data_1978)
summary(model_area)
```

```
##
## Call:
## lm_robust(formula = area ~ nearinc, data = data_1978)
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error t value
## (Intercept)   2074.8      45.83  45.273
## nearinc1      -240.1     120.21  -1.997
##
## (Intercept) 0.0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
## nearinc1    0.04731528349435318303584097066050162538886070251464843750000000000000000000000000000000000000000000000000
##             CI Lower CI Upper  DF
## (Intercept)   1984.3 2165.196 177
## nearinc1      -477.4   -2.876 177
##
## Multiple R-squared:  0.03091 ,   Adjusted R-squared:  0.02543
## F-statistic:  3.99 on 1 and 177 DF,  p-value: 0.04732
```

```
model_land  <- lm_robust(land ~ nearinc, data = data_1978)
summary(model_land)
```

```
##
## Call:
## lm_robust(formula = land ~ nearinc, data = data_1978)
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error t value                      Pr(>|t|) CI Lower
## (Intercept)    52569       4635  11.341 0.000000000000000000009291    43422
## nearinc1      -30729       7141  -4.303 0.000027779594032855775510 50   -44821
##             CI Upper  DF
## (Intercept)    61716 177
## nearinc1      -16637 177
##
## Multiple R-squared:  0.08082 ,   Adjusted R-squared:  0.07563
## F-statistic: 18.52 on 1 and 177 DF,  p-value: 0.00002778
```

Additionally, each of these coefficients (or mean difference values) are statistically significant ($p = 0.05$). The above evidence implies the location choice of the incinerator was not "random", but rather selected on the basis of housing prices and characteristics.

3 (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).

The estimate in (a), which is based on the observed differences in (b), is likely to be biased downward because this value captures the impact of other characteristics related to housing price (such as the age and size of the home) other than location relative to the incinerator. Before construction of the incinerator in 1978, homes near the incinerator site were older, smaller, and cost less, on average. Because the previous estimate absorbs the affect of these housing characteristics, it is likely to overstate the negative effect of the incinerator on housing values.

4 (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.

```
diff_diff <- lm_robust(formula = rprice ~ nearinc, data = data)
summary(diff_diff)
```

```
##
## Call:
## lm_robust(formula = rprice ~ nearinc, data = data)
```

```
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error t value
## (Intercept)    91035       1793  50.783
## nearinc1      -24457       4419  -5.534
##
## (Intercept) 0.000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
## nearinc1    0.000000065202409033234613334323315214646754611749202013015747070312500000000000000000000
##             CI Lower CI Upper  DF
## (Intercept)    87509    94562 319
## nearinc1      -33151   -15763 319
##
## Multiple R-squared:  0.1147 ,    Adjusted R-squared:  0.1119
## F-statistic: 30.63 on 1 and 319 DF,  p-value: 0.0000000652
```

The DD estimator is -24457, which implies houses near the incinerator are worth, on average, $24457 less than houses not near the incinerator.

5 (e)

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

```
conf_low  <- diff_diff$conf.low[[2]]
conf_high <- diff_diff$conf.high[[2]]
```

There is a 95% probability that the estimate of the causal effect on the incinerator is between $-33151 and $-15763.

6 (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.

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

```
##
## 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
```

The `nearinc1` variable coeffcicent is no longer statistically significant; whereas the coefficents of `year1981`, `age`, `rooms`, and `area` are statistically significant and non-zero. This implies that these other variables impact housing prices more than being located near the incinerator.

```
linear_hypothesis <- linearHypothesis(model = model_control,
                                      c("age=0",
                                        "rooms=0",
                                        "area=0",
                                        "land=0"),
                                      white.adjust="hc2")
summary(linear_hypothesis)
```

```
##      Res.Df          Df        Chisq         Pr(>Chisq)
##  Min.   :314   Min.   :4   Min.   :134.7   Min.   :0
##  1st Qu.:315   1st Qu.:4   1st Qu.:134.7   1st Qu.:0
##  Median :316   Median :4   Median :134.7   Median :0
##  Mean   :316   Mean   :4   Mean   :134.7   Mean   :0
##  3rd Qu.:317   3rd Qu.:4   3rd Qu.:134.7   3rd Qu.:0
##  Max.   :318   Max.   :4   Max.   :134.7   Max.   :0
##                NA's   :1   NA's   :1       NA's   :1
```

```
p_value_lin_hyp <- linear_hypothesis$`Pr(>Chisq)`[2]
p_value_lin_hyp
```

```
## [1] 0.00000000000000000000000000003851232
```

Because the p-value « 0, we reject the null hypothesis that the coefficients on the house and lot characteristics are all jointly equal to zero. Therefore, we must control for these previously omitted variables.

7 (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.

```
price_increase <- model_control$coefficients[[3]]
price_increase
```

```
## [1] 9019.277
```

Holding all other variables constant, housing prices increased by $9019, on average, between 1978 and 1981.

8 (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.

The key assumption underlying the causal interpretation of the DD estimator is that the control group provides a valid counterfactual for the temporal evolution of the mean outcomes in the treatment group in absence of a change in treatment. In this example, the key assumption is that the trend in housing price is the same, whether or not a house is located near the incinerator. In other words, the parallel trends assumption indicates that the trend is the same for both the treatment (near incinerator) and control (not near incinerator) groups.