EDS241: Final

Joe DeCesaro

03/18/2022

1 Take Home Final

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.

1.1 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.

1.2 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)
- year indicator (1978 or 1981).

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

1.2.1 Read in the data

```
house_data <-read.csv(here("KM.csv"))
```

1.3 (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.

```
house_data_1981 <- house_data %>%
  filter(year == 1981) # filter to just get 1981

# effect of being near incinerator on prices in 1981
model1 <- lm_robust(rprice ~ nearinc,
                    data = house_data_1981)
summary(model1)
```

```
##
## Call:
## lm_robust(formula = rprice ~ nearinc, data = house_data_1981)
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper  DF
## (Intercept)   101308       2945  34.402 3.633e-70    95485   107130 140
## nearinc       -30688       6243  -4.915 2.442e-06   -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 house "penalty" for houses near the incinerator is an average decrease in value of 30688.27 dollars compared to houses not near the incinerator. The estimated coefficient corresponds with the 'causal' effect of the incinerator as the p-value deems it a statistically significant difference at the 0.001 level. However, the incinerator could have been placed in an area that already had lower housing prices generally (ie. a worse neighborhood, less land, etc) so this could be omitted variables bias (OVB). We explore if this is OVB in the next question.

1.4 (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?]

```r
# lets check if the prices were lower in these areas already
house_data_1978 <- house_data %>%
  filter(year == 1978)

# house prices that would be near/away incinerator in 1978
model2.1 <-  lm_robust(formula = rprice ~ nearinc,
                       data = house_data_1978)

model2.1_sum <- summary(model2.1)
model2.1_sum
```

```
##
## Call:
## lm_robust(formula = rprice ~ nearinc, data = house_data_1978)
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper  DF
## (Intercept)    82517       1878  43.932 3.949e-97    78811    86224 177
## nearinc       -18824       6010  -3.132 2.031e-03   -30685    -6964 177
##
## Multiple R-squared:  0.08167 ,   Adjusted R-squared:  0.07648
## F-statistic:  9.81 on 1 and 177 DF,  p-value: 0.002031
```

Based on the housing information from 1978, the incinerator was placed near houses that, on average, were sold for 18824.37 adjusted dollars less than houses not near the incinerator placement. The p-value for the difference in houses near the incinerator is 0.00203 and therefore statistically significant at the 0.01 level. This indicates that placement of the incinerator was not random but near homes that were of less value already, lets look at another variable to be sure.

```r
# did houses near the incinerator have more, less, or the same land?
model2.2 <-  lm_robust(formula = land ~ nearinc,
                       data = house_data_1978)

model2.2_sum <- summary(model2.2)
model2.2_sum
```

```
##
## Call:
## lm_robust(formula = land ~ nearinc, data = house_data_1978)
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper  DF
## (Intercept)    52569       4635  11.341 9.291e-23    43422    61716 177
## nearinc       -30729       7141  -4.303 2.778e-05   -44821   -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
```

Based on the housing information from 1978, the incinerator was placed near houses that, on average, had 30729.13 less of lot square footage than houses not near the incinerator placement. The p-value for the difference in houses near the incinerator is 0.00003 and therefore statistically significant at the 0.001 level. This confirms that placement of the incinerator was not random.

1.5 (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 difference in housing prices near the incinerator in 1982 is likely affected by the fact that they were on average worth less in 1978 and had smaller lot sizes than those homes further away from the incinerator. There could be additional omitted housing characteristic variables. The affect on price we are seeing in (a) is likely overstated from omitting variables from the regression and they should be controlled for.

1.6 (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.

```
 house_data <-  house_data %>%
  # make binary variable for years
  mutate(post1981 = case_when(year == 1978 ~ 0,
                              year == 1981 ~ 1),
         # make treatment dummy binary variable for houses in 1981 that are near the inc
         treatment_dummy = nearinc*post1981)

# make the DD model for different binary possibilities
model3 <- lm_robust(formula = rprice ~ post1981 + nearinc + treatment_dummy,
                    data = house_data)

# to be able to get the confidence intervals by calling them
model3_table <- tidy(model3) %>%
  select(term, estimate, std.error, p.value, conf.low, conf.high)

model3_sum <- summary(model3)
model3_sum
```

```
##
## Call:
## lm_robust(formula = rprice ~ post1981 + nearinc + treatment_dummy,
##     data = house_data)
##
## Standard error type:  HC2
##
## Coefficients:
##                 Estimate Std. Error t value   Pr(>|t|) CI Lower CI Upper  DF
## (Intercept)        82517       1878  43.932 7.429e-137    78822    86213 317
## post1981           18790       3493   5.380  1.452e-07    11918    25662 317
## nearinc           -18824       6010  -3.132  1.897e-03   -30649    -7000 317
## treatment_dummy   -11864       8666  -1.369  1.720e-01   -28914     5186 317
```

```
##
## Multiple R-squared:  0.1739 ,    Adjusted R-squared:  0.1661
## F-statistic: 17.72 on 3 and 317 DF,  p-value: 1.169e-10
```

The estimated DD causal effect of the incinerator on housing values without controlling for other characteristics is an adjusted sales price difference of -11863.9. This value represents the difference between adjusted sales price of houses compared to those that are not near the incinerator and sold in 1981 (control) AND the difference compared to houses sold in 1978 and would be near the incinerator (treated pre-policy intervention). The negative sign means that the houses in the area that were near the incinerator (treated) had a lower price sales than those not near the incinerator (not treated) but, it should be noted that the p-value is 0.172 and therefore not significant. This is confirmed with the confidence interaval below.

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

The 95% confidence interval for the estimate of the causal effect on the incinerator is [-28913.8, 5186]. The 95% confidence interval crosses 0 so the DD estimate is not statistically different from 0.

1.8 (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.

```
# make the DD model for different binary possibilities and control for housing characteristics
model4 <- lm_robust(formula = rprice ~ post1981 + nearinc + treatment_dummy +
                      # control variables
                      age + rooms + area + land,
                    data = house_data)

# to be able to get the confidence intervals by calling them
model4_table <- tidy(model4) %>%
  select(term, estimate, std.error, p.value, conf.low, conf.high)

summary(model4)
```

```
##
## Call:
## lm_robust(formula = rprice ~ post1981 + nearinc + treatment_dummy +
##     age + rooms + area + land, data = house_data)
##
## Standard error type:  HC2
##
## Coefficients:
##                    Estimate Std. Error t value       Pr(>|t|)    CI Lower
## (Intercept)     -17688.8531  11070.584 -1.5978 0.111090982713 -39471.0244
## post1981         13093.9319   2795.311  4.6842 0.000004195095   7593.9555
## nearinc           3514.1412   7149.521  0.4915 0.623402359190 -10553.0565
## treatment_dummy -13320.1540   6785.662 -1.9630 0.050533201725 -26671.4332
## age               -266.3383     50.716 -5.2516 0.000000279088   -366.1251
## rooms             6969.0020   1542.265  4.5187 0.000008832216   3934.4851
## area                23.7821      3.901  6.0962 0.000000003194     16.1063
```

```
## land                 0.1268      0.137  0.9254 0.355473122621     -0.1428
##                     CI Upper  DF
## (Intercept)      4093.3181 313
## post1981        18593.9082 313
## nearinc         17581.3389 313
## treatment_dummy    31.1252 313
## age              -166.5515 313
## rooms           10003.5188 313
## area               31.4579 313
## land                0.3964 313
##
## Multiple R-squared:  0.612 , Adjusted R-squared:  0.6034
## F-statistic: 79.94 on 7 and 313 DF,  p-value: < 2.2e-16
```

The estimated DD causal effect of the incinerator on housing values while controlling for other characteristics is an adjusted sales price difference of -13320.15. The negative sign means that the houses in the area that were near the incinerator (treated) had a lower price sales than those not treated and sold in 1978. It should be noted that the confidence interval is [-26671.43, 31.13] which does cross 0 so the DD estimate is not statistically different from zero but it is close.

```
# linear hypothesis test
hyp_test <- linearHypothesis(model4,
                             c("age = 0", "rooms = 0", "area = 0", "land = 0"),
                             white.adjust = "hc2")
hyp_test
```

Res.Df	Df	Chisq	Pr(>Chisq)
317			
313	4	138	7.39e-29

The proper F-statistic is 138 which is greater than 10, therefore the housing characteristics are not a weak instrument. The p-value of the linear hypothesis test is statistically significant so we can reject the null hypothesis that the coefficients on the housing characteristics are all jointly equal to 0.

1.9 (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. (for the control group per Slack)

The real housing values change on average between 1978 and 1981 for the houses not near the incinerator (the control group), when controlling for other housing characteristics, is 13093.93.

1.10 (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 the the the parallel trend assumption which states 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 the context of the incinerator construction in North Andover, this means that if the treatment did not happen (incinerator was never built), the house sale price difference from 1978-1981 for the houses that would have been treated (near the incinerator) is the same for those houses that were never treated (not near the incinerator). In numbers this means that, if the incinerator was never built the houses that would have been near the incinerator would have seen an average increase in the adjusted sale price of 13093.93 from 1978-1981 when controlling for the other housing characteristics. This is according to the model from question (f).