Assignment #2

ECON 5783 — University of Arkansas

Prof. Kyle Butts

These assignments should be completed in groups of 1 or 2 but submitted individually. My preference is for you to use Rmarkdown files to have your code, results, and your answers to the questions intermixed. Since I am not requiring you to code in R for these assignments, you can use latex or microsoft word to write up answers alternatively. Not that in these cases, I would like you to upload your code seperately.

Theoretical Questions

1. Define the following conditional expectation function:

$$p(n) \equiv \mathbb{E}[\text{Home Value}_i \mid \text{Num Rooms}_i = n],$$

   where i denotes a given home in Massachusetts.

   i. In words, describe how to think about $p(2)$.

   ii. Say you have a sample of parcels, how would you go about estimating $p(n)$? How would you estimate this in a regression?

   iii. How does the "fully-flexible" conditional expecation function differ from a linear regression model where $\text{Home Value}_i = \text{Num Rooms}_i\beta + u_i$?

   iv. Why might we not believe that $p(3) - p(2)$ be the causal effect of increasing from 4 to 5 rooms on homeprice?

2. Now let's think about a more complicated conditional expectation of worker's wages as a function of sex (M/F) and having a college-degree (0/1):

$$w(g, d) \equiv \mathbb{E}[\text{Wages}_i \mid \text{Gender}_i = g, \text{College Degree}_i = 1]$$

   i. Say I include a set of indicator variables for gender and for whether or not the worker has a college degree in a regression. Describe a scenario where this regression model is

not the conditional expecation function (hint: think about interactions).

Coding Exercise

This exercise involves a sample dataset of homes in Massachussets. A randomly selected sample of the dataset appears in this folder `data/MA_parcels_sample.parquet`. You can open this dataset using `arrow::read_parquet("data/MA_parcels_sample.parquet")` after installing the arrow package in R.

1. Estimate the conditional expecation function from the theoretical question using regression (the `i` function from the `fixest` package will help you):

$$p(n) \equiv \mathbb{E}[\text{Home Value}_i \mid \text{Num Rooms}_i = n].$$

   What is the predicted change in home price from going from 4 to 5 rooms?
2. Now estimate a linear regression of Home Value on the number of rooms. What is the predicted change in home price from going from 4 to 5 rooms? How does it differ from part (i)
3. Now I want to use binscatter regression to estimate the relationship between the lot size of the home to the total value controlling (linearly) for the number of rooms. Use the `binsreg` package to do this. Use the setting `line = c(0, 0)` and `line = c(2, 2)`.
4. Do you think this is the causal effect of lot size on home value? Why or why not?