Midterm Project: Due November 10, 2023

1 Description

Your midterm project is to answer a causal question using data and a causal inference method of your choice (from two-stage least squares, difference-in-difference, or regression discontinuity). **You will be graded based on a group write-up and 8 minute in class presentation.** Your write-up should take the form of a 3 to 5 page single-spaced paper. This paper should have 4 main components:

1. An explanation of the causal question that details the setting and why society should care about the answer
2. An explanation of the data
3. An initial attempt to answer the causal question using multiple linear regression
4. A better answer to the causal question using two-stage least squares, difference-in-difference, or regression discontinuity

The rubric in the next section provides a detailed breakdown of how points will be assigned. You may complete the project in your homework group or an alternative group of up to 5 people. Please submit one copy of the paper per group. Please put the names of all group members at the top.

2 Rubric

Your write-up and in class presentation should address the following points:

1. (10 points) What causal question are you answering?
    - What are the key details of the setting?
    - Why should society care about the answer?
2. (20 points) What data will you use to answer this question?
    - Where did you find the data?
    - How was this data originally collected?
    - What are the important variables in the data?
    - What are some key summary statistics of the data? (i.e. What is "in" the data?)
3. (20 points) *Estimate a multiple linear regression that attempts to answer the question*
    - What is your outcome variable and why?
    - What are your explanatory variables and why?
    - What are your estimates for the coefficients?
    - How do you interpret these results?
    - What does your regression say about the causal question?

4. (10 points) *Explain why the multiple linear regression may provide an incorrect answer to your question*
5. (35 points) *Use a two-stage least squares, difference-in-difference, or regression discontinuity analysis to provide a better answer to your question.* BE CAREFUL TO CHOOSE AN APPROPRIATE METHOD!
    - *Explain the two-stage least squares, difference-in-difference, or regression discontinuity analysis you run*
    - What are your parameter estimates?
    - How do you interpret these results? (i.e. What does this analysis say about the causal question?)
6. (5 points) Provide a key takeaway, policy recommendation, or simple summary of your findings

3 Replicating an Existing Economics Paper

A valid approach to this assignment is to "replicate" an existing economics paper. That is, find an economics paper that uses two-stage least squares, difference-in-difference, or regression discontinuity analysis to answer a causal question with publicly available data. You can then base your write-up on the causal question, data, and analysis method in the paper. However, write-ups that are directly based on an existing paper in this way (i.e. that use the paper's question, data, and analysis method) will receive a maximum grade of B+.

4 ChatGPT

You may use ChatGPT or similar models to assist with the write-up. That said, remember that these models hallucinate. They will generally provide incorrect information on causal inference methods.

5 Data

You are encouraged to use any and all data that you can find for this project. Potential sources for useful data include:

- https://dataverse.harvard.edu/
- https://www.openicpsr.org/openicpsr/
- https://www.kaggle.com/datasets
- https://www.data.gov/
- https://www.bls.gov/data/
- https://data.cdc.gov/
- https://data.seattle.gov/
- https://opendata.cityofnewyork.us/
- https://www.nber.org/research/data