Evaluation of Labour Policy

Replication: Marginal returns to medical care

Topic: Regression discontinuity design (RDD)

Reading: Douglas Almond, Joseph Doyle, Amanda Kowalski, and Heidi Williams (2010). Estimating marginal returns to medical care: Evidence from at-risk newborns. *Quarterly Journal of Economics*, 125 (2): 591-634.

Data description:

The National Center for Health Statistics (NCHS) birth cohort–linked birth/infant death files include data for a complete census of births occurring each year in the United States for the years 1983–1991 and 1995–2002 (approximately 66 million births). The data include information reported on birth certificates linked to information reported on death certificates for infants who die within one year of birth. The sample for the lab session includes about 200,000 newborns who fall within the pilot bandwidth of three ounces (85 grams) around the threshold of 1,500 grams. The sample contains information birthweight (*dbirwt*) and mortality after one year (*death1year*). Given that you do not have information of year of birth, state of birth, and newborn characteristics, you won't be able to replicate Almond et al.'s results.

Deadline:

The solutions must be submitted by May 3, 2023 (23.59h).

You have to submit two files:

- One PDF file with your answers
- One R source file with your coding

or one Output (html or PDF) from an RMarkdown including your code, output, and interpretations.

Submit your solutions via StudyNet.

QUESTIONS:

Q1: Preliminaries.

Based on the paper's definitions, create the following variables necessary for the analysis:

a) The running variable.

b) The dummy variable.

c) The interaction term between them.

[Hint: center the running variable around the cutoff]

Q2: Graphical analysis.

a) Produce a graph of the density of the running variable. Do you observe any violations of the RDD assumptions? Explain.

b) Draw a graph of one-year mortality against birthweight. Work on the details of the graph to best represent your estimation strategy. What do you conclude?

Q3: RDD estimation.

a) Run the RDD regression using OLS. Choose how to compute the standard errors and motivate your answer.

b) Interpret the results found in part a) and comment on statistical significance.

Q4: Run the following specification checks and comment on their results:

a) Use a 25-grams bandwidth around the cutoff.

b) Allow for a quadratic functional form of the running variable.

c) Use the package 'rdrobust' to compute the symmetric and asymmetric optimal bandwidths.

Q5: Discussion of the identification strategy.

Discuss, in 2-3 sentences, the validity of the identification strategy of Almond et al. What do you think are the threats to internal validity? Use one example to support your discussion.