Graded Homework Exercise 2

Felix Weinhardt

May/June 2024

This problem set will be graded. Please work alone or together in groups of two. Send your solutions as Stata do-file and PDF to weinhardt@europa-uni.de before June 7th 0:00am.

Problem 1

Think about the following RD situation: X is your running variable and you know the assignment rule that everyone with $X <= c$ should get treatment T, so $T = 1$, and everyone with $X > c$ not ($T = 0$). What you are interested in is getting the estimate for the effect of T on the outcome Y. In the data that you collected with a colleague, however, for $X > c$ you observe individuals with $T = 0$ and also with $T = 1$. For $X <= c$ you only observe individuals with $T = 1$.

1. Your frustrated colleague who put most of effort into the data-collection states: "Since we observe $T = 0$ and $T = 1$ for $X > c$ this is not a setting where we can learn anything about the effect of T on Y." (Assume that we are in a world of homogeneous potential outcomes). Do you share this view? Explain.

2. Your colleague is now very excited and tells you that she believes the homogeneous treatment effects assumption can be defended and as a result we can just compare average outcomes of individuals with $X <= c$ and $T = 1$ to average outcomes of individuals with $X > c$ and $T = 0$. Your colleague has even produced some estimates of this effect for various functional forms in X and bandwidth choices. How much would you trust these estimates? Explain using the potential outcomes framework.

3. You recall that in some lecture you were told that "RDD is IV" and to use the "ivregress" command in Stata to estimate effects in RDD, so you tell your colleague to try it out. The resulting treatment effect differs significantly from the effect found in Q2. Provide at least two reasons why this could be the case.

4. Your colleague really digs herself into the data and now noticed that there is a very large share of observations with $X > c$ and $T = 1$. Is this something you think you should be worried about? How would you assess this?

5. A senior colleague whom you show your data points out that the individuals who have $X > c$ and $T = 1$ could be a non-random selection of all individuals close to c and that these look like individuals with lower levels of education. If this was the case, what would this mean for the interpretation of your result from the ivregress command? And how could you assess the types of individuals that your estimates speak to?

6. Simulate the data generating process. For this you should set a seed such that your data can be replicated. Colour observations with $T = 0$ and $T = 1$ differently and clearly indicate the threshold

value c in the graphs. Create one graph where there is a negative effect of the treatment and one graph where there is no effect. Please provide your code (Stata or R) and the graphs as solutions.

7. Now provide a simulated example where the continuity of the conditional regression function is not met (to simplify things, you may move to a sharp RD design). What stronger assumption would you require to be able to estimate the discontinuity in this case? Your discussion on this can be less formal. Provide again a graph, code and also a small discussion.

Problem 2

Read the teaching note by Weinhardt 2024 (Github and here). From the field of your own interest, find two recently published empirical papers papers (from the AEA journals) that when interpreting results discuss coefficient movements or stability from including additional controls. Provide the paper/PDF, reference with page number and highlight the text section in the PDF that discusses coefficient movements. Are heterogeneous effects explicitly ruled our or modeled? Provide a brief comment on the paper/section.