Graded Homework Exercise 2

Felix Weinhardt

May/June 2025

This exercise contains two problems and will be graded. Please work alone or together in groups of two. Upload your solutions to the relevant Github repository 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 for $X <= c$ you observe individuals with $T = 0$ and for $X > c$, you observe individuals with $T = 1$ as well as individuals with $T = 0$.

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 $T = 0$ to average outcomes of individuals with $X > c$ and $T = 1$. 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. Write down the Stata command/specification for this, allowing for a second-degree polynomial on both sides of the cutoff. Your colleagues runs this regression. 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 $T = 0$. 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 $T = 0$ 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 above? And how could you assess the types of individuals that your estimates speak to?

Problem 2

Read the study "Social Housing, Neighborhood Quality and Student Performance", Journal of Urban Economics 2014, vol. 82: pp.12-31. https://doi.org/10.1016/j.jue.2014.06.001

1. What estimator is being used to identify the reduced form effect of living in a social housing neighborhood on test scores?
2. You vaguely recall that in the lectures we discussed different settings of this estimator, and that in some circumstances the implicit weighting might produce bad results. From the institutional setting of this study, do you think that these recent advances in the literature are relevant (a priori)? Explain why.
3. Given the empirical evidence presented, would you suggest to use a different estimator that weighs groups differently? Explain why/why not.