5304 Econometrics Autumn 2023[1]

Problem set 4
Regression Discontinuity Design

Part 1: RDD in practice

For the first part of this assignment, we use the data set from Angrist and Lavy (1999). In this paper, the authors investigate the impact of class size on student test performance for fourth- and fifth graders in Israel. In particular, the twelfth century rabbinic scholar Maimonides proposed a maximum class size of 40 (known as Maimonides' rule) which still lives on today. This rule creates a nonlinear relationship between cohort size, i.e. the number of incoming students in a school year, and actual class size. For example, 40 incoming students should in principle result in one big class, while the addition of one additional students splits the cohort into two classes of (on average) 20.5 students each. We will focus on the sample of fifth graders, which is available in the data set *PS4_AngristLavy99.dta*.

1. Restrict the sample to include only schools with cohort sizes below 160. Plot a linear relationship between actual class size and cohort size, allowing for discontinuities at 41, 81 and 121. Do schools appear to follow Maimonides' rule in class division? *Hint: use "binscatter" and its RD option. In R you can use ggplot's stat_summary_bin for bins, and geom_smooth on different subsamples of the data to get linear fits with discontinuities.*

2. Plot the same kind of relationship as in (2), but now with average reading and math scores on the y-axis, respectively, and discuss briefly what you find.

We are now going to depart from the exact methodology used in Angrist and Lavy (1999), but the general takeaways are the same. In (1) and (2), you have investigated the first stage and reduced form relationships in a fuzzy regression discontinuity design.[2] Now we are going to estimate the effect of class size on test scores with 2SLS, using the discontinuous nature of Maimonides' rule as an instrument.

3. Let us focus only on the cutoff around 41 – restrict the sample to only include cohort sizes between 0 and 80.

4. Generate a new variable of the following form:

$$Cohort\_Recentered = Cohort\_Size - 41$$

   This is simply the cohort size of a given school centered at 41. Further, generate a dummy variable (*Above*) that equals 1 if a school has a cohort size equal to or greater than 41.

   Now, write down the following two equations: (i) the first stage equation of class size on the *Above* dummy as well as the running variable ($Cohort\_Recentered$), allowing the running variable to have different slopes on each side of the cutoff, and (ii) the reduced form equation of math scores on the same variables. Denote the coefficients on *Above* as γ_1 and π_1 in the first stage and reduced form equations, respectively.[3]

---

[1] TA: Petter Berg, petter.berg@phdstudent.hhs.se

[2] I suggest reading the chapter "Fuzzy RD is IV" in Mostly Harmless Econometrics which also provides an exposition based on Angrist and Lavy (1999).

[3] Note that the functional form used here is exactly the same as the one used in the figures from (1). Cohort size enters linearly in the expression, and its slope is allowed to vary on each side of each cutoff. Also, "running variable" is common RD lingo for the variable in which cutoffs determine something of interest: treatment, or in this case class division.

Show the regression estimates of the two equations in a table. What is the interpretation of γ_1 and π_1? Why do you think we center the cohort size variable at 41, so that it is equal to zero when the discontinuity occurs? *Hint: it only simplifies interpretation of the coefficients. Try running the same regressions without re-centering and see what happens!*

5. Using your estimates from the previous question, compute $\hat{\beta}_1 = \hat{\pi}_1/\hat{\gamma}_1$. Further, run a 2SLS regression with $Cohort\_Recentered_i$ and $\left(Cohort\_Recentered \times Above_i\right)$ as exogenous regressors, and class size instrumented by $Above_i$. Report the estimates in a table. Is the estimated effect of class size on math scores the same as $\hat{\beta}_1$? Even if estimates are the same, why should you always use 2SLS in practice?

Part 2: Interpreting published results

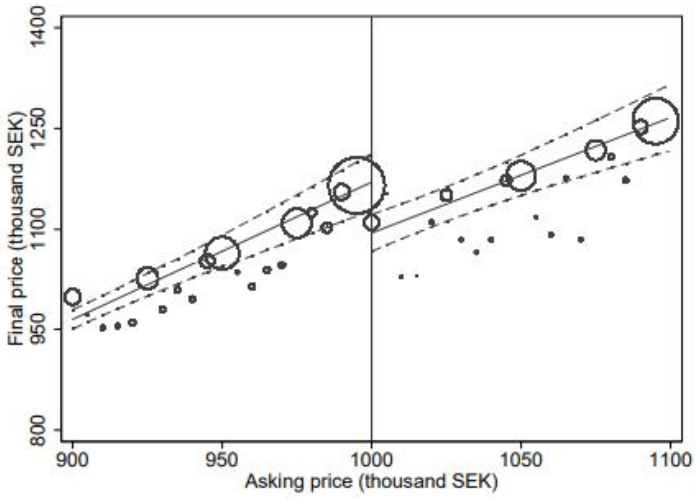
In Repetto and Solís (2019), the authors investigate the so called "left-digit bias" in the Swedish housing market. In particular, they estimate the effect on final sales prices in housing auctions of the starting price being at or just below a round number (e.g. 1,990,000 SEK vs. 2,000,000 SEK). The authors find that setting starting prices just below leads to $3-5\%$ higher final prices, and that this is driven by more people participating in the auction.

1. Figure 1 in Repetto and Solís (2019) shows the discontinuity in final sales prices at the 1 million SEK cutoff. Write down an econometric specification that would allow you to estimate the size of the discontinuity at the 1 million SEK cutoff. Define all variables that you introduce, and motivate why you set up the specification in the chosen way. Ignore control variables, just focus on the simple RD. Which coefficient in your set-up measures the size of the "jump" at the cutoff? *Hint: don't worry, there is not just one right way to do this.*

2. In most RDD settings, being just above or just below the cutoff can be argued to be as good as random. Why is this not the case here?

3. When an apartment in Sweden turns out to be hard to sell (for whatever reason) it is common for the housing ad to be revised with an "accepted price", such that the first interested buyer willing to pay the price gets the apartment. Assume that such accepted prices are more commonly set to multiples of a million. Would this be problematic?

Figure 1: The discontinuity in final prices around the 1-million asking price threshold



References

Angrist, J. and V. Lavy (1999): "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *The Quarterly Journal of Economics*, 114, 533–575.

Repetto, L. and A. Solís (2019): "The Price of Inattention: Evidence from the Swedish Housing Market," *Journal of the European Economic Association*, 18, 3261–3304.