5304 Econometrics Autumn 2023

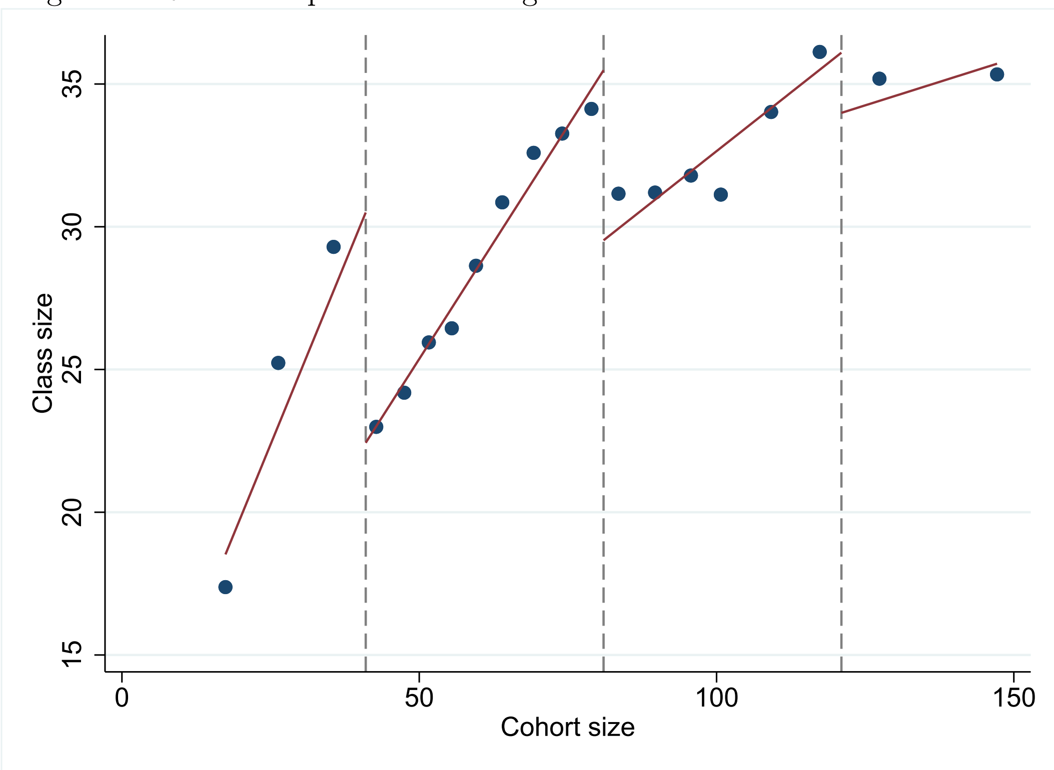
Problem set 4: Group 15[1]

Part 1: RDD in practice (Angrist and Lavy (1999))

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.

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?

Figure 1: Relationship between fifth grade actual class size and cohort size



[1] 42624@student.hhs.se, 42613@student.hhs.se, 42632@student.hhs.se, 25164@student.hhs.se

Figure 1 plots the average of actual and predicted (i.e. line of best fit within the intervals) class sizes against enrollment in fifth grade. From the raw data, we observe discontinuous drops in class sizes at 41, 81, 121 such that there are no violations of Maimonides' rule (maximum observed class size at 36 to 37).

We note through simple mathematical reasoning that the drop at 41 is larger than the drop at 81, which in turn is larger than the drop at 121. Since it makes no sense to have an additional class with only one child, the school would proceed to redistribute students evenly across n + 1 classes (taking n to be the original number of classes). The fall is more pronounced when the cohort size is small.

However, class sizes are not uniquely determined by the Maimonides' rule which calls for a fuzzy RDD approach over a sharp RDD (which assumes a one-dimensional threshold rule).

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.

Figure 2: Relationships between fifth grade average scores and cohort size

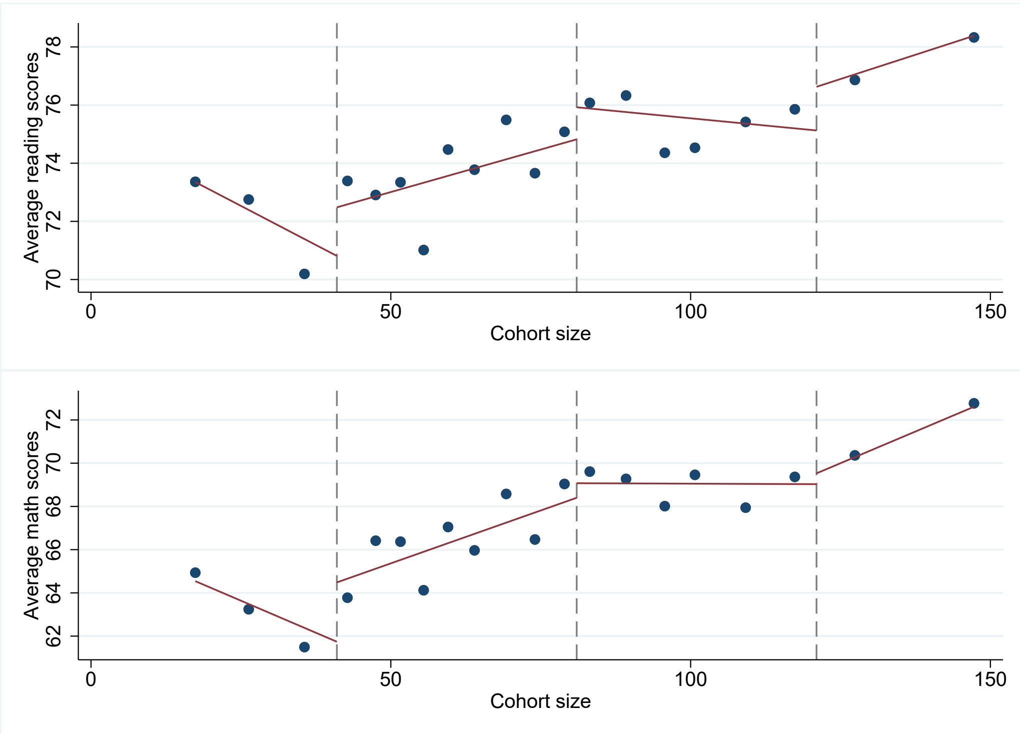


Figure 2 plots the average reading and math scores against enrollment in fifth grade. Close to the threshold of enrollment levels of 40, there seems to be a jump in average scores. At higher enrollment levels of 80 and 120, the discontinuity seems somewhat weaker, especially for math scores.

Intuitively, assuming that smaller class sizes do improve average test scores, we believe that this concurs with the findings in question 1 which show how at the interval ends, class size falls more when overall cohort size is small. Thus, we can speculate that the larger increase in academic performance is positively correlated to the larger fall in class size.

The good news is that the data seems to be linear within the intervals, simplifying functional form specification. However, we remain sceptical of the lack of observations, especially near the thresholds.

Overall in the data, we find that average test scores are positively correlated with cohort size. Selection bias comes from pupils selecting schools (and sometimes, from schools selecting pupils). Schools that have better funding, higher quality of teachers and located in densely populated urban areas are more highly sought after, thereby increasing cohort size. There is also selection bias from the quality of students that enroll into schools of different sizes.

Although the RDD approach involves us applying the "randomization inference" principle that schools arbitrarily close to the intervals (e.g. schools of cohort size 40 and 41) are fundamentally not significantly different from each other, we would need to conclusively prove that there is no manipulation at the boundaries to believe that there is co-variate smoothness. It is possible for parents who believe that the class size will shrink once exceeding the threshold, and thus move their children to these schools. Since these parents care more about their child's education, we can reasonably expect the child to have higher academic performance due to higher levels of educational investment from parents. We would need to test this later using a McCrary density test.

Thus, the inability to rule out omitted variable bias at the cutoff will violate the assumptions underlying randomization inference.

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: $Cohort\_Recentered = Cohort\_Size - 41$ 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. 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?

$$
\begin{aligned}
classize_i &= \gamma_0 + \gamma_1 Above_i + \gamma_2 Cohort\_Recentered_i \\
&\quad + \gamma_3 Above_i \times Cohort\_Recentered_i + \epsilon_{1i} \\
avgmath_i &= \pi_0 + \pi_1 Above_i + \pi_2 Cohort\_Recentered_i \\
&\quad + \pi_3 Above_i \times Cohort\_Recentered_i + \epsilon_{2i}
\end{aligned}
$$

We first see that $Above_i$ is the "instrument". The first stage equation from $classsize_i$ involves a constant γ_0, the impact of the instrument γ_1 near the threshold (i.e., the sudden drop at discontinuity), γ_2 parameter corresponding to the normalized cohort size, and γ_3 which shows the effect of crossing the threshold as we get further above the threshold (i.e., cohort size increasing beyond threshold), and some error term ϵ_{1i}.

γ_1 is the first stage which shows the effect (on class size) of the instrument of following the Maimonides' rule at the threshold. The coefficient of -11.38, which we take to be statistically significant (at 0.001% significance level), is a strong first stage giving us confidence that there is a sizeable jump in treatment probability (i.e., class size falls a lot). Thus, the "instrument" is 'relevant' and a discontinuity exists.

π_1 is the reduced-form effect of $Above_i$ (enrollment level changes from 40 to 41) that reduces class size and in turn affect average math score. Assuming exclusion restriction and other RD assumptions hold, this is the **"intent-to-treat"** effect (Lee and Lemieuxa (2010)).

The value of 3.533 (statistically significant at 5% significance level) shows that there is a positive correlation between smaller class size and math scores. Similar to the IV interpretation, π_1 is the local average treatment effect (LATE)(focusing only on compliers) at the cutoff cohort size of 41.

Centering cohort size at 41 is a normalization that ensures the treatment effect (class size has on average scores) at $Cohort\_Size = 41$ is still the coefficient on $Above_i$.

Additionally, it allows us to accurately control for the interaction effects in γ_3 and π_3. Using the normalised variable in the control polynomial gives it flexibility to take different forms to the left and the right of the threshold.

Thus, the standard local linear specification involves normalisation of the running variable.

	(i) classize	(ii) avgmath
Above	-11.38***	3.533*
	(1.232)	(1.666)
Cohort_Recentered	0.673***	-0.167
	(0.0569)	(0.0855)
Above × *Cohort_Recentered*	-0.338***	0.274**
	(0.0602)	(0.0933)
Constant	33.72***	60.81***
	(1.190)	(1.448)
Number of classes	1185	1184

Standard errors in parentheses corrected for within-school correlation
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

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 $(Cohort\_Recentered \times Above_i)$ 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?

$$\hat{\beta}_1 = \hat{\pi}_1/\hat{\gamma}_1 = 3.5327379/-11.383923 \approx -0.31$$

	2SLS
Class size	-0.310
	(0.161)
Cohort_Recentered	0.0421
	(0.0658)
Above × *Cohort_Recentered*	0.169*
	(0.0812)
Mean math score	71.27***
	(4.037)
Number of classes	1184

Standard errors in parentheses corrected for within-school correlation
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The estimated effect of class size on math scores is the same as $\hat{\beta}_1$. The result is an imprecise

$$-0.31(s.e. = 0.161),$$

which illustrates the loss of precision by shrinking sample about the discontinuity (Angrist and Pischke (2009)). Shrinking the bandwidth decreases the extent of omitted variable bias by making schools more comparable (with closer cohort size) but the elimination of observations creates inefficiency from data loss. Since we lack sufficient observations near the threshold, the class size parameter estimate loses its statistical significance from imprecision.

We use 2SLS in practice since it produces more appropriate standard errors.

Furthermore, 2SLS yields the better theoretical interpretation. In this study, we are carrying out a fuzzy RD (analogous to instrumental variables which utilize 2SLS, and not a sharp RD (analogous to matching). Fuzzy RD is used here since the Maimonides' rule is only part of the decision-making calculus that determines class sizes. After all, the Maimonides' rule is not an inviolable legal obligation, but a tradition/ heuristic. Other factors such as availability of teachers, level of school funding concurrently influence class sizes. Since probability is no longer a discrete jump from 0 to 1, then the OLS estimate cannot be understood as the average treatment effect. Thus, we rely upon the IV approach to obtain the LATE, given the exclusion restriction and monotonicity assumptions hold.

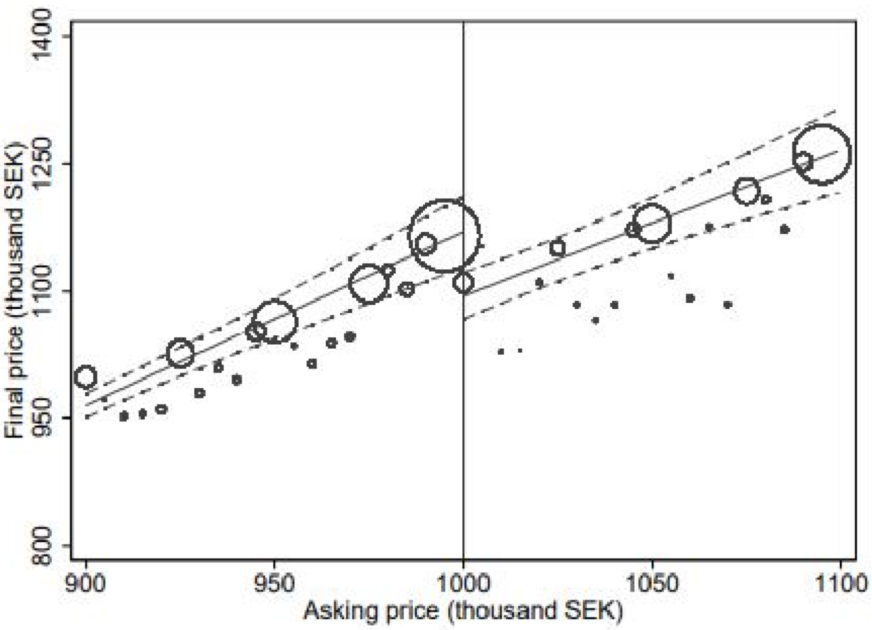
Therefore, even if the OLS and 2SLS estimates coincide, it is always preferred to use the 2SLS approach.

Part 2: Interpreting published results (Repetto and Solís (2019))

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.

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



A specification that would allow for estimating the size of the discontinuity:

$$p_i = \alpha + \beta \mathbb{1}(x_i \geq c) + \gamma(x_i - c) + \delta \mathbb{1}(x_i \geq c) \times (x_i - c) + \epsilon_i,$$

where p_i is the the final sale price of apartment i, and $x_i - c$ is the running variable, defined as the distance between the asking price, x_i, and the threshold, c (1 million SEK). The running variable is assumed to have a linear effect on the final price, but the slope can be different at each side of each threshold. The coefficient of interest is β, which captures the discontinuity in the final price (in absolute terms) as apartments cross the 1-million threshold, assuming a homogeneous effect.

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?

Interpreting the discontinuity as a causal effect in most regression-discontinuity designs relies on the assumption that agents are unable to perfectly manipulate the running variable since they can set the initial asking price. In this setting, the running variable is perfectly manipulable by the seller. This systematic bunching of apartments around thresholds can potentially be a consequence of selection, which can take at least two forms: selection based on apartment characteristics and selection due to real estate agents (Repetto and Solís (2019)).

Arguably, price determination based on apartment characteristics around the threshold is more likely to be as-good-as-random since the seller does not decide on the characteristics. Furthermore, since the characteristics that distinguish a house just above or below the threshold are likely to be small, they are not as likely to be systematically attributable in one way or another to a specific price range of apartments. However, if sellers know that setting the price just below a round number will attract more people to the bidding process and yield a higher final price, there will be more asking prices right below the threshold. This intentional grouping of asking prices beneath the threshold violates random assignment.

The credibility of the assumption of no manipulation of the running variable around the cutoff could be assessed with a (McCrary) density test. The density of the running variable should evolve smoothly around the cutoff. If there are jumps that could be indication of a violation.

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?

If accepted prices are more commonly set to multiples of a million, it could potentially introduce a bias in the estimation of the left-digit bias effect. This is because the choice of accepted prices aligns with the threshold of a million, and the observed prices could be influenced not only by the left-digit bias but also by the pricing convention related to accepted prices. This may confound the estimates of the jump at the 1 million SEK cutoff, making it challenging to disentangle the effect of left-digit bias from the effect of accepted price conventions.

If there is a general trend of final prices being higher than the asked prices, there will be a drop in this trend at the thresholds if accepted prices are usually set to the thresholds, since final prices will then be equal to asked prices, and not higher, for a lot of apartment sales at the thresholds.

To address this, carrying out a "donut RD" by dropping sample observations at the heap points would provide a robustness check.

References

Angrist, J. D. 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.

Angrist, J. D. and J.-S. Pischke (2009): *Mostly harmless econometrics : an empiricist's companion*, Princeton: Princeton University Press.

Lee, D. S. and T. Lemieuxa (2010): "Regression Discontinuity Designs in Economics," *Journal of economic literature*, 48, 281–355.

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.