Problem Set 4 ECON8825

Federico Rodari

December 9, 2022

1 Question 1 - Differences-in-differences

(a) Consider the following twfe model

$$Y_{it} = \alpha_i + \delta_t + \beta^{DD} D_{it} + u_{it}$$

and the observed outcome as function of potential outcome and individual treatment effect

$$Y_{it} = Y_{it}(0) + \tau_{it} D_{it}$$

Under the parallel trend assumption, we know that $\beta^{DD} = \text{ATT} \equiv \mathbb{E}\left[\mathbb{E}\left[\tau_{it} \mid g_i\right]\right]$, and therefore

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}\left[Y_{it} \mid g_i\right]\right] &= \mathbb{E}\left[\mathbb{E}\left[Y_{it}(0) \mid g_i\right]\right] + \mathbb{E}\left[\tau_{it} \mid g_i\right] D_{it} \\ &= \mathbb{E}\left[\mathbb{E}\left[Y_{it}(0) \mid g_i\right]\right] + \beta^{DD} D_{it}\end{aligned}$$

therefore it must be that $Y_{it}(0) = \alpha_i + \delta_t + u_{it}$

(b) Using FWL, we can write

$$\hat{\beta}^{\text{twfe}} = \frac{\frac{1}{NT}\sum_i^N \sum_t^T \widetilde{D}_{it}\widetilde{Y}_{it}}{\frac{1}{NT}\sum_i^N \sum_t^T \widetilde{D}_{it}^2} = \frac{\sum_i^N \sum_t^T \widetilde{D}_{it} Y_{it}}{\sum_i^N \sum_t^T \widetilde{D}_{it}^2}$$

where the second equality comes from the fact that double demeaning is idempotent. From $Y_{it} = Y_{it}(0) + \tau_{it} D_{it}$, we can substitute and obtain $\widetilde{D}_{it} Y_{it} = \widetilde{D}_{it} Y_{it}(0) + \widetilde{D}_{it} D_{it} \tau_{it}$ so

$$\begin{aligned}\frac{\sum_i^N \sum_t^T \left[\widetilde{D}_{it} Y_{it}(0) + \widetilde{D}_{it} D_{it} \tau_{it}\right]}{\sum_i^N \sum_t^T \widetilde{D}_{it}^2} &= \sum_i^N \sum_t^T \frac{\widetilde{D}_{it}}{\sum_i^N \sum_t^T \widetilde{D}_{it}^2} Y_{it}(0) + \sum_i^N \sum_t^T \frac{\widetilde{D}_{it} D_{it}}{\sum_i^N \sum_t^T \widetilde{D}_{it}^2} \tau_{it} \\ &= \sum_i^N \sum_t^T \frac{\widetilde{D}_{it}}{\sum_i^N \sum_t^T \widetilde{D}_{it}^2} \left[\alpha_i + \delta_t + u_{it}\right] + \sum_i^N \sum_t^T \frac{\widetilde{D}_{it} D_{it}}{\sum_i^N \sum_t^T \widetilde{D}_{it}^2} \tau_{it} \\ &= \sum_i^N \sum_t^T \frac{\widetilde{D}_{it}}{\sum_i^N \sum_t^T \widetilde{D}_{it}^2} u_{it} + \sum_i^N \sum_t^T \frac{\widetilde{D}_{it} D_{it}}{\sum_i^N \sum_t^T \widetilde{D}_{it} D_{it}} \tau_{it}\end{aligned}$$

given $\sum_i^N \sum_t^T \alpha_i \widetilde{D}_{it} = \sum_i^N \sum_t^T \delta_t \widetilde{D}_{it} = 0$, which implies $\omega_{it} = \frac{\widetilde{D}_{it} D_{it}}{\sum_i^N \sum_t^T \widetilde{D}_{it} D_{it}}$.

(c) For the weights to be appropriate we would need them to sum to 1. This condition could be violated as pointed out in Goodman-Bacon (2021), since β^{twfe} can imply negative weights arising from 2×2 DiDs where the early treatment cohorts are used as control group for later treated cohorts. In this context, the weights appear to be well-behaving since $\widetilde{D}_{it} D_{it} = (D_{it} - \bar{D}_i)(D_{it} - \bar{D}_t)$, so with a

binary treatment the numerator is always positive.
If the weights sum to one

$$\mathbb{E}[\mathbb{E}[\hat{\beta}^{\text{twfe}} \mid \widetilde{D}_{it}]] = \sum_i^N \sum_t^T \frac{\widetilde{D}_{it}}{\sum_i^N \sum_t^T \widetilde{D}_{it}^2} \mathbb{E}[u_{it} \mid \widetilde{D}_{it}] + \sum_i^N \sum_t^T \frac{\widetilde{D}_{it}^2}{\sum_i^N \sum_t^T \widetilde{D}_{it}^2} \mathbb{E}[\mathbb{E}[\tau_{it} \mid \widetilde{D}_{it}]]$$

$$\mathbb{E}[\hat{\beta}^{\text{twfe}}] = \sum_i^N \sum_t^T \frac{\widetilde{D}_{it}^2}{\sum_i^N \sum_t^T \widetilde{D}_{it}^2} \tau^{\text{ATT}} = \tau^{\text{ATT}}$$

2 Question 2 - Regression discontinuity

Let's first have a look at the distribution of the two raw (not binned) outcomes of interest.

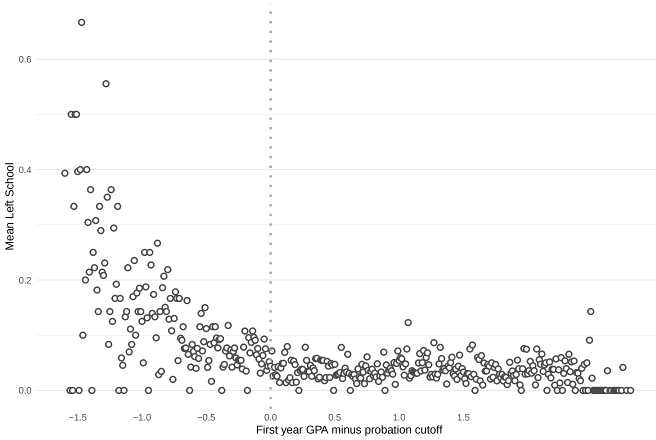

Figure 1: Left School

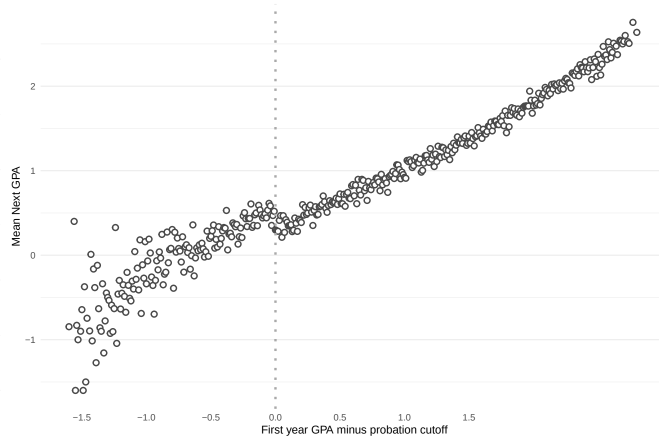

Figure 2: Next Year GPA

Figure 3: Distribution of Outcome Variables v. Relative GPA

While the outcome left school might look to exhibit some nonlinearities, it is likely linear that in a neighborhood of the cutoff. For this reason, I chose to adopt a linear model in this design.
To estimate the causal effect of being on probation, I adopt two approaches. The first one is to replicate the results in Lindo et al. (2010) by adopting a uniform kernel with symmetric bandwidths of size 0.6, which entails the estimation of a simple linear regression model by restricting the sample to the observations that fall within the bandwidth. Similarly, I try to replicate the same results using the package rdrobust provided by Cattaneo, though it requires an adjustment: since the default model assumes a treatment for $c \geq 0$, but our is implied by $c < 0$, we need to estimate the model using $-X$ as a running variable. Surprisingly (for me), the results are consistent for left school but not for nextGPA[1]. Finally, I use the package rdrobust to estimate a linear model with a triangular kernel and optimal bandwidth chosen to minimize the MSE.
The results using the following regression

$$Y_i = \beta_0 + \beta_1 1(\text{GPAcutoff}_i < 0) + \beta_2 \text{GPAcutoff}_i + \beta_3 \text{GPAcutoff}_i \cdot 1(\text{GPAcutoff}_i < 0) + \varepsilon_i$$

[1] I have read carefully the documentation but I could not figure out what was driving different results.

Table 1: Estimated Discontinuity in Left School

	First year GPA < cutoff	Constant (control mean)
Estimate	0.018	0.001
SE	0.008	0.014

Table 2: Estimated Discontinuity in Next GPA

	First year GPA < cutoff	Constant (control mean)
Estimate	0.24	0.722
SE	0.03	0.051

While using the package we get the following results

Variable	MSE-Optimal Bandwitdh	RD Estimator	Robust Inference p-value	Robust Inference Conf.Int	Eff. Number Observations
Left Voluntarily School	0.6	0.018	0.225	[-0.008, 0.035]	(8131, 4698)
Left Voluntarily School	0.492	0.017	0.131	[-0.005, 0.038]	(6216, 4026)
Next Year GPA	0.6	0.199	0.003	[0.048, 0.223]	(7473, 4047)
Next Year GPA	0.233	0.065	0.473	[-0.075, 0.161]	(2363, 2014)

Comparison between Lindo et al. (2010) specification and the one implied by optimal bandwitdh and triangular kernel.

The resulting effect can be visualized in

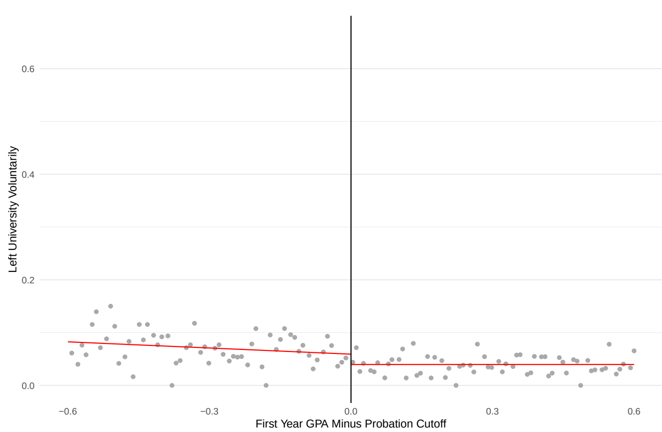

Figure 4: Left School

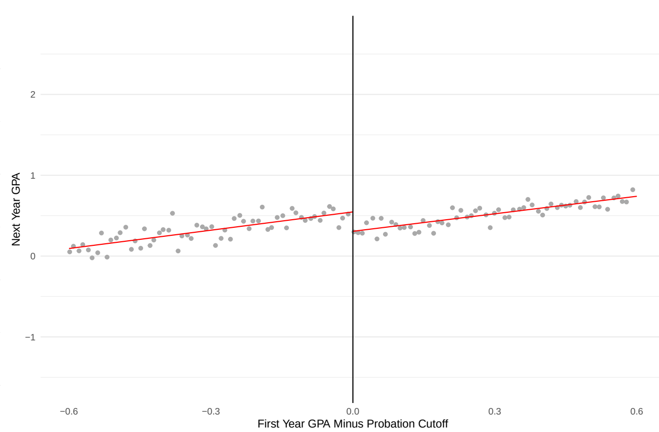

Figure 5: Next Year GPA

Figure 6: Visualization of RD Estimates

From the above results, we can conclude that being put on probation increases the probability of leaving school permanently by 1.8 percentage points, and leades students to improve their GPA by approximately

0.24 points.
The estimates of the effect on subsequent year GPA are likely to be biased (in either direction) due to a composition effect. If academic probation led the worst students to drop out, on the extensive margin we would observe upward bias, given that now the sample of students has higher GPA by construction, even without a true intensive margin effect. Alternatively, we could have the best students to drop out and find a downsized effect (downward bias) through the same composition effect mechanism.
Given this mechanism, a natural way to address this bias would be to provide best and worst case scenarios to bound the estimated effect. If we assume that the students at the bottom (top) drop out in the treatment group, we should drop also the bottom (top) ones in the control group to make the groups comparable.
I perform a series of robustness checks on a set of predetermined variables but no placebo variables, as I could not think about any other variable that could be be logically follow after being put on probation but that would not likely be affected by it. The reason for performing such checks is to test the existence of significant discontinuity in observables that could be symptomatic of treatment manipulation. I include students' high school relative grades, credits attempted in the first year at college, age at entry, gender, whether the student was born in North America and whether English was the student's first language. As an example, think about the language dummy: potentially, we could observe that English speaker students know the schooling system rules better and could tweak the rules in order to be just above the probation cutoff.

Variable	MSE-Optimal Bandwitdh	RD Estimator	Robust Inference p-value	Robust Inference Conf.Int	Eff. Number Observations
HS grade percentile ranking	0.6	-0.114	0.907	[-2.453, 2.178]	(8131, 4698)
Credits attempted in 1st Year	0.6	-0.025	0.218	[-0.087, 0.02]	(8131, 4698)
Age at Entry	0.6	0.033	0.209	[-0.028, 0.126]	(8131, 4698)
Male	0.6	0.001	0.443	[-0.07, 0.031]	(8131, 4698)
Born in North America	0.6	0.021	0.167	[-0.01, 0.059]	(8131, 4698)
English is 1st language	0.6	-0.017	0.91	[-0.044, 0.049]	(8131, 4698)

Using uniform (rectangular) kernel, linear model, symmetric custom bandwidth

Variable	MSE-Optimal Bandwitdh	RD Estimator	Robust Inference p-value	Robust Inference Conf.Int	Eff. Number Observations
HS grade percentile ranking	0.508	-0.034	0.812	[-2.461, 1.928]	(6597, 4160)
Credits attempted in 1st Year	0.36	-0.046	0.071	[-0.114, 0.005]	(4438, 3280)
Age at Entry	0.512	0.042	0.165	[-0.021, 0.123]	(6706, 4200)
Male	0.51	-0.013	0.43	[-0.066, 0.028]	(6597, 4160)
Born in North America	0.52	0.022	0.088	[-0.004, 0.059]	(6808, 4268)
English is 1st language	0.363	0.004	0.656	[-0.039, 0.062]	(4438, 3280)

Using triangular kernel, linear model, symmetric MSE optimal bandwidth

In the two tables below I conduct a similar analysis as before, comparing the methodology in Lindo et al. (2010) with one that uses optimal bandwidths and triangular weights. We can observe that in the standard setting we do not observe any significant discontinuity, though in the second setting we do observe that some of the coefficients are significant at the 10% level. Again, using the package I was not able to exactly replicate the paper's results, and I do not think the inconsistency comes from the different number of observations as some of the coefficients are too difficult to be driven by a small difference in sample size.