Homework 2

Jordi Torres

October 24, 2025

1 Question 1

Following Cattaneo et al., I select windows around each cutoff where covariates are balanced at the 15 % level (i ALSO use the test of diff in means to be as less restrictive as possible). For the population running variable, no such window exists (recommended [0; 0]); the local-randomization assumption therefore fails.For the time running variable, a balanced window of ± 1.6 units is obtained with 147 observations (24 below, 123 above).
Within this window, as shown in table **??**, randomization-based inference yields a significant wage discontinuity of 326 units ($p < 0.01$), while the teacher competency score shows no significant change ($p \approx 0.98$).
This holds only if the following assumptions hold:

1. As if random treatment assignment-¿covariates test
2. SUTVA-¿argue this with the same way they do it in the paper, cite them.SUTVA may be violated if the policy triggered spillovers through teacher sorting around the population cutoff – e.g. if teachers who chose a position in a high-bonus school just below the threshold would have otherwise chosen a position just above the threshold in the absence of the wage bonus policy.
3. No manipulation at the cutoff-¿rdsensitivity

Table 1: Local Randomization RD Estimates

Running Variable	**Outcome**	**Recommended Window**	**Obs. Below**	**Obs. Above**	**Effect**	**p-value**
Population	Wage	[0 ; 0]	0	20	–	–
Time	Wage	[-1.61 ; 1.61]	24	123	326.43	0.000
Population	Score	[0 ; 0]	0	20	–	–
Time	Score	[-1.61 ; 1.61]	24	123	-0.08	0.986

In table XX I show some preliminary evidence of graphical RD around the cutoff. I have restricted the dataset to be around the cutoff and then I have grouped the dataset in similar bins -not using the Cattaneo et al approach- but very similarly. Just to provide some evidence. I have also fitted a kernel. We can observe that just looking at the graph there seems not to be a dis
Figure XX presents the randomization sensitivity analysis for the four combinations of running and outcome variables. I will consider only the time variable, as the balance test already did not support local effects of population. For this variable the wage effect is positive and significant in narrow windows, remaining robust as the window widens—consistent with a credible local randomization design. The score outcome shows small and insignificant effects, indicating no impact on academic performance. **Revise this, unclear**
Finally, in tables XX and XX I show the results of the min-pvalue. Here we observe that clearly population does not capture a local effect, as the p-values of balancedness only breach 0.15 with larger bandwidths, while the opposite is true for the time variable, where the effect seems more locally robust.

2 Question 2

2.1 Continuity parametric

$$Y_{is} = \alpha + \beta_1 X_{is} + \tau D_{is} + \beta_2 (D_{is} \times X_{is}) + \varepsilon_{is}, \tag{1}$$

Table 2: Continuity-based (Global Parametric) RD Estimates

Outcome	Running Variable	RD Estimate ($\hat{\tau}$)	Std. Error	t-value	p-value	Significance
Wage	Population	284.32	9.78	29.08	< 0.001	***
Score	Population	4.49	1.08	4.17	< 0.001	***
Wage	Time	324.21	8.48	38.25	< 0.001	***
Score	Time	-1.52	8.48	-0.18	0.858	

Notes: Clustered standard errors at the school level.

Specification: $Y = \alpha + \beta_1 X + \tau D + \beta_2 D \times X$. The coefficient τ measures the RD discontinuity at the cutoff.

2.2 Continuity non-parametric

Table 3: Continuity-based (Local Non-Parametric) RD Estimates using `rdrobust`

Outcome	Running Var	RD Estimate ($\hat{\tau}$)	Std. Error	p-value	Bandwidth (h)	N_L / N_R	Significance
Wage	Population	-205.18	23.45	< 0.001	161.5	10,130 / 4,658	***
Wage	Time	347.15	12.85	< 0.001	32.8	6,151 / 8,637	***
Score	Population	-8.19	2.07	< 0.001	211.0	10,130 / 4,658	***
Score	Time	4.88	1.82	0.007	27.8	6,151 / 8,637	**

Notes: Estimates computed using `rdrobust` with triangular kernel and MSE-optimal bandwidths.

$Y_{is} = \alpha + \tau D_{is} + f_-(X_{is}) + f_+(X_{is}) + \varepsilon_{is}$, where τ measures the discontinuity at the cutoff.

Cluster-robust standard errors at the school level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

add also graphs

2.3 Assumptions and test

I think I may need to add stuff on the sensitivity? rd sensitivity.
No manipulation (rddensity) –¿this is key, but do it more systematic, review this once I am done with this bullshit.
Covariates smooth at the cutoff
Placebo cutoffs (optional): run RD at fake cutoffs (e.g., ±20 units) to show no spurious jumps.
Robustness:

Table 4: Robustness of Local RD Estimates to Polynomial Order and Kernel Choice

Specification	Outcome	Running Var	$\hat{\tau}$	SE	p-value	Bandwidth (h)	Obs (L/R)
p=2, Triangular	Wage	Population	-227.06	37.67	< 0.001	159.9	10,130 / 4,658
	Wage	Time	335.92	18.92	< 0.001	29.1	6,151 / 8,637
	Score	Population	-7.38	2.84	0.009	215.2	10,130 / 4,658
	Score	Time	1.27	2.87	0.658	24.7	6,151 / 8,637
p=3, Triangular	Wage	Population	-228.46	39.30	< 0.001	252.6	10,130 / 4,658
	Wage	Time	333.28	21.08	< 0.001	40.5	6,151 / 8,637
	Score	Population	-6.14	3.39	0.070	231.9	10,130 / 4,658
	Score	Time	0.82	3.14	0.794	36.3	6,151 / 8,637
p=1, Epanechnikov	Wage	Population	-201.69	22.14	< 0.001	161.9	10,130 / 4,658
	Wage	Time	352.15	13.43	< 0.001	28.2	6,151 / 8,637
	Score	Population	-8.04	2.20	0.0002	176.4	10,130 / 4,658
	Score	Time	5.19	1.85	0.005	25.2	6,151 / 8,637
p=2, Epanechnikov	Wage	Population	-229.13	37.17	< 0.001	150.9	10,130 / 4,658
	Wage	Time	336.95	18.84	< 0.001	28.6	6,151 / 8,637
	Score	Population	-6.91	2.98	0.020	191.2	10,130 / 4,658
	Score	Time	1.08	2.88	0.709	23.6	6,151 / 8,637

Notes: Local linear regressions estimated with `rdrobust`. Bandwidths are MSE-optimal.
Triangular and Epanechnikov kernels used with varying polynomial orders $p \in \{1, 2, 3\}$.
Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

3 3

Table 5: Joint Regression Discontinuity using Distance to Rurality Frontier

Outcome	RD Estimate ($\hat{\tau}$)	Std. Error	p-value	Bandwidth (h)	N_L / N_R	Significance
Wage	-132.72	11.45	< 0.001	294.85	12,192 / 2,596	***
Score	-7.75	1.31	< 0.001	710.48	12,192 / 2,596	***

Notes: Running variable is the signed Euclidean distance to the rurality frontier,
defined as $d_i = \sqrt{(\text{population}_i - 500)^2 + (\text{time}_i - 120)^2}$, with negative values for treated units.
Local linear estimates computed using `rdrobust` with triangular kernel and MSE-optimal bandwidths.
Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

4 4

Correct but pending to add the correct graph!!!