

Homework 2

Jordi Torres

October 26, 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 differences in means to be less restrictive)¹. For the population running variable, no such window exists; 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 ($p \approx 0.98$).

The validity of these results rests on three identification assumptions:

1. **Local as-if random assignment.** Within the selected window around the cutoff, treatment status is assumed to be as good as randomly assigned. This requires that the distribution of pre-treatment covariates be balanced on both sides of the threshold. I check this by testing differences in means and using the Cattaneo et al. balance criterion (15% level).
2. **Stable Unit Treatment Value Assumption (SUTVA).** Each unit's potential outcomes depend only on its own treatment status. SUTVA may be violated if teachers strategically sort across nearby vacancies—e.g., if teachers prefer high-bonus schools just below the threshold—introducing interference across units.
3. **No precise manipulation of the running variable.** Agents should not be able to precisely control their position relative to the cutoff (population or time). This is verified graphically and through density tests (`rddensity`), ensuring no discontinuity in the distribution of the running variable at the cutoff.

Under these conditions, treatment assignment can be interpreted as locally random, and the estimated discontinuity corresponds to a causal effect of the high wage bonus on the outcome.

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 graph to be around the cutoff and then I have grouped the observations in similar bins in an ad hoc way (not using fully Cattaneo et al). Just to provide some evidence. We can observe that just looking at the graph there seems be a jump around both cutoffs and both variables.

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,

¹I have included all the X variables in the dataset. The prob of balancedness is lower the more variables we include. But we want balance on all the potential variables that can affect potential outcomes, so including them all seems reasonable here

remaining robust as the window widens, which is consistent with a credible local randomization design. The score outcome shows small and insignificant effects, indicating no impact on academic performance. Finally, in tables XX I show the results of the min-pvalue across different bandwidths. 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 (so the sample is not balanced in characteristics at the neighborhood of the cutoff), while the opposite is true for the time variable, where the effect seems more locally robust.

2 Question 2

2.1 Continuity parametric

For this model, I have simply specified:

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

I could have included different specifications for the running variable, fit polynomials etc. The assumptions of this method are the same as in the method below and will be mentioned at the end of this section, along with evidence in their favor.

The results of this method can be found in table 1. Results from the global parametric specification show a large and statistically significant wage discontinuity at both the population and time cutoffs (≈ 284 and 324 units, respectively), consistent with the structure of wage bonuses across the rurality frontier. Teacher scores, however, do not exhibit any significant discontinuity, suggesting that higher wages in rural areas did not translate into differences in measured teacher quality near the cutoff.

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

Next, I used the state-of art method of Cattaneo et al. This one corrects for the optimal bandwidth of the kernel to optimize in the trade-off variance-bias of the estimate. In my baseline specification I have chosen a triangular kernel, their baseline method to compute bandwidth h (mserd) and have fitted a line ($p=1$). I have also run different specifications combining these three elements, which I show in table 4. In table 3 I report the results of my main specification. The local estimates confirm a clear and statistically significant discontinuity in teacher wages at both rurality thresholds, consistent with the policy-driven wage bonuses. In contrast, the teacher competency score shows only minor or no significant jumps, suggesting limited sorting or selection effects around the cutoff. Compared to the global parametric specification, the local estimates rely on weaker functional-form assumptions and therefore provide a more credible estimate of the causal discontinuity.

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: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

To visualize the discontinuities estimated with the local polynomial RD, I use the `rdplot` command from the `rdrobust` package, which plots the fitted local polynomial on each side of the cutoff using the optimal bandwidths selected by the CCT procedure. The figures show the sharp discontinuity in wages at both the population and time thresholds, while the jump is less strong for teacher scores, which is both consistent with the numerical estimates reported in Table 3. **add comment**

2.3 Identification Assumptions and Validity Checks

The identification of the local average treatment effect in the continuity-based RD design relies on the assumption that, in the absence of treatment, potential outcomes evolve smoothly around the cutoff. In practice, this assumption may fail if there is strategic sorting or bunching around the cutoff, as agents could manipulate the running variable to obtain the treatment (e.g., by misreporting population or distance).

To assess this, I propose the following standard validity checks:

1. **No manipulation at the cutoff:** tested using the `rdensity` procedure, which examines whether the distribution of the running variable is continuous at the threshold. The estimated density test statistics (reported in Appendix ??) show no evidence of bunching or sorting, supporting the assumption of local randomization. **REVISE**
2. **Covariate smoothness:** observable covariates such as population, time to provincial capital, and school-level characteristics remain balanced around the cutoff. This supports the idea that unobservables are also likely to be smooth.
3. **Placebo cutoffs (optional):** as an additional falsification test, I could estimate RD effects at fake cutoffs (e.g., ± 20 units from the true threshold). The absence of discontinuities at these placebo thresholds would be more evidence in support of the validity of this method. I have not done this due to time constraints.

Overall, the empirical evidence is consistent with the identifying assumptions of the RD design. To further support that my estimates are not due to kernel selection, p selection or method of bandwidth selection, I estimate the same model across different specifications. Effects on wages are robust across specifications, while the effects on score are more affected by the assumptions or functional forms we introduce.

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

Question 3. I compute:

$$d_i = \sqrt{(\text{population}_i)^2 + (\text{time}_i)^2},$$

and assign a negative sign to schools located on the “urban” side of the threshold (i.e., those with $D_{\text{pop}} = 1$ or $D_{\text{time}} = 1$). This way, the discontinuity at zero captures the jump when moving from urban (negative) to rural (positive) locations²

Using this signed distance, I run the same local linear RD as before, using Cattaneo et al approach. Results are reported in Table 5. Both outcomes show large and statistically significant negative jumps at the frontier. Given that the distance measure combines both time and population, the estimated effects are smaller than when using each running variable separately. This may happens because this metric pulls together towns that differ along either dimension but end up at a similar overall distance from the cutoff, making the comparison less sharp.

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

Here now I will be considering effects along the L axis formed by time and population around the cutoff (0,0). I will define different boundary points b_1, b_2, \dots, b_n and estimate local rd in each of this points, using the `rd2` command in R. I will consider points in the (population, 0) dimension and in the (0, -time) dimension. Given that I consider a lot of boundary points, I will show only the output of the effects (and CI) of the boundary effects on the two outcomes we are considering.

I report the results in graphs X and X. Points from b1 to b20 report increasing departures along the population axis and keeping time at 0. b_1 , for example, is the effect of increasing population by 10 as we move along the time 0 frontier. The opposite (i.e decrease time by 5) is the case for points 20-40. As we can see, there is evidence of a local effect of population that decreases as we move up, and a local effect of time that also increases as time gets negative and population is 0. For score there does not seem to be a local effect at the boundary.

²This measure is not perfect, since it implicitly puts more weight on whichever dimension has higher variance; wich means eaning I might exaggerate cases that are far in one direction but close in the other. Still, itis a simple and intuitive way to summarize both running variables into a single metric.

Table 6: Conditional Logit (Fixed-Effects) Estimates

Variable	Coefficient	Std. Error	z	P> z
wage	-1.6489	0.0456	-36.13	0.000
distance	-0.0209	0.00018	-115.57	0.000
index_ccpp	0.0273	0.0136	2.00	0.046
time	0.1185	0.00546	21.69	0.000
time2	-0.00136	0.00014	-9.43	0.000
logpop	0.0673	0.0193	3.49	0.000
logpop2	-0.0094	0.0009	-10.36	0.000
multigrado	-0.1213	0.0367	-3.28	0.001
bilingue	-0.1913	0.0417	-4.59	0.000
unidocente	0.0621	0.1260	0.49	0.623
frontera	0.8065	0.0483	16.70	0.000
vraem	0.3702	0.0676	5.48	0.000
Model Fit				
Log-likelihood	-73,839.99			
Observations	65,125,137			
LR χ^2 (12)	22,987.83 ($p < 0.001$)			
Pseudo R^2	0.6088			

Notes: Conditional (fixed-effects) logit model estimated with teacher-level grouping (`group(id.teach)`).
The dependent variable is `match`. Standard errors are clustered at the teacher level.

Table 7: Conditional Logit and Mixed Logit Estimates

Variable	Conditional Logit			Mixed Logit		
	Coef.	Std. Err.	z	Coef.	Std. Err.	z
wage	-1.6489	0.0456	-36.13	-0.8053	0.1665	-4.84
distance	-0.0209	0.0002	-115.57	-0.0550	0.0026	-21.29
index_ccpp	0.0272	0.0136	2.00	0.1203	0.0454	2.65
time	0.1185	0.0055	21.69	0.1821	0.0283	6.45
time2	-0.00136	0.00014	-9.43	-0.00401	0.00110	-3.65
logpop	0.0673	0.0193	3.49	-0.0676	0.0707	-0.96
logpop2	-0.0094	0.0009	-10.36	-0.0040	0.0036	-1.11
multigrado	-0.1213	0.0367	-3.28	-0.0864	0.1279	-0.68
bilingue	-0.1913	0.0417	-4.59	-0.4751	0.1514	-3.08
unidocente	0.0621	0.1260	0.49	-0.5799	0.4534	-1.28
frontera	0.8065	0.0483	16.70	0.9359	0.1831	5.11
vraem	0.3702	0.0676	5.48	0.6058	0.2328	2.60
SD (Random parameters)						
wage (SD)				-0.0536	0.1961	-0.27
distance (SD)				-0.0243	0.0014	-17.50
Model fit						
Log-likelihood	-73,839.99			-3,291.52		
Observations	65,125,137			660,678		
Pseudo R^2 / LR χ^2	0.6088 / 22,987.83			LR $\chi^2(2)=1053.99, p < 0.001$		

Notes: The mixed logit model allows for random coefficients on wage and distance. The sign of the estimated standard deviations is irrelevant; they should be interpreted as positive. Clustered standard errors at the teacher level for both models.

Table 8: RD Replication Estimates using Conditional and Mixed Logit Models

Outcome	Running Var.	Conditional Logit			Mixed Logit		
		$\hat{\tau}$	SE	p-value	$\hat{\tau}$	SE	p-value
Wage	Population	-0.2055	0.0309	< 0.001	-0.1807	0.0363	< 0.001
Wage	Time	0.2502	0.0244	< 0.001	0.2514	0.0277	< 0.001
Score	Population	-6.4892	3.6061	0.072	-6.9488	4.1439	0.094
Score	Time	-6.7586	4.5362	0.136	-7.8370	4.7963	0.102
Observations		Same sample (RD window)			Same sample (RD window)		
Kernel / Method		Triangular, MSE-optimal h			Triangular, MSE-optimal h		

Notes: RD estimates computed using local linear regression around the cutoff, with triangular kernel and MSE-optimal bandwidths.

$\hat{\tau}$ measures the local treatment effect at the cutoff estimated through the clogit and mixlogit assignment models. Significance at 10%, 5%, and 1% levels are indicated by p-values < 0.1, < 0.05, and < 0.01 respectively.

Table 9: Average Outcomes by Model and Rurality Status (Observed vs Counterfactual)

Model	Rurality	Mean (Observed)	Mean (Counterfactual)	Difference (Δ)
Clogit	Urban (0)	94.13	92.89	-1.23
Clogit	Rural (1)	119.24	121.69	2.45
Mixlogit	Urban (0)	94.83	94.31	-0.52
Mixlogit	Rural (1)	119.41	120.75	1.34

Notes: The table reports mean predicted outcomes (teacher scores) under the observed and counterfactual wage regimes. Rurality = 1 indicates rural schools, while 0 denotes urban ones. The difference Δ represents the change in mean outcome after the counterfactual adjustment to wages.