Replication Study Ariana Alves Antunes

Applied Statistics II – 4 April, 2023

What is the paper about?

The research is regarding educational performance information and electoral accountability, relying on two different research designs.

- First, a regression discontinuity design is employed to compare municipalities across Brazil that barely met a school quality target and those that barely missed it
- Second, a field experiment in the state of Pernambuco that examines the effect of providing information about a municipality's ANA performance on votes for the incumbent mayor's re-election is analysed.

The main goal is the measure education quality to study whether and how voters respond to signals of public education quality.

1. IDEB* Basic Education Development index
2. ANA National Literacy Assessment

Model

The project has two different parts, this paper is focusing only on the regression model (RDD*) used to identify the effect of meeting the IDEB target on electoral outcomes in municipalities across Brazil. Which I was able to replicate as followed

○ Model:

$$T_{mj} = \begin{cases} 1 & \text{if } D_{mj} \geq -0.05 \quad \text{(rounding, IDEB score} \geq \text{IDEB target)} \\ 0 & \text{if } D_{mj} < -0.05 \quad \text{(rounding, IDEB score} < \text{IDEB target)} \end{cases}$$

$$\tau = \mathbb{E}[Y_{1mj} - Y_{0mj}|D_{mj} = c] = \lim_{D_{mj}\downarrow c} \mathbb{E}[Y_{1mj}|D_{mj} = c]$$
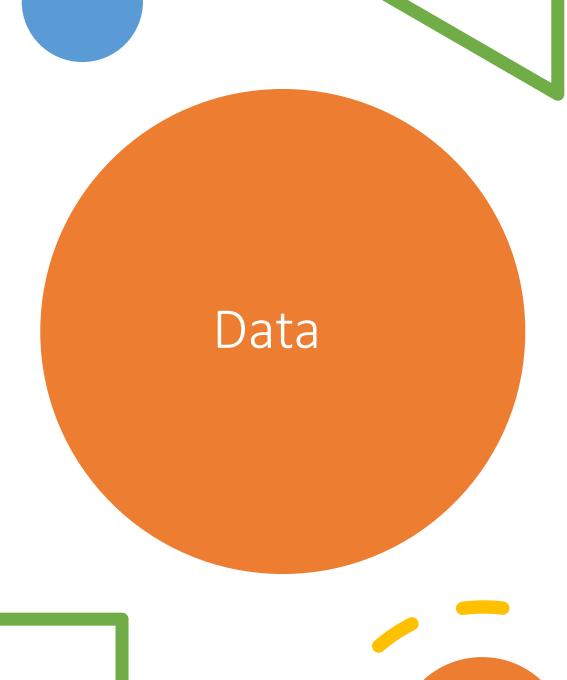$$- \lim_{D_{mj}\uparrow c} \mathbb{E}[Y_{0mj}|D_{mj} = c].$$

"Where Y1i,j and Y0i,j represent the potential outcome of interest (vote share or re-election of the mayor) under treatment (having met the IDEB target) and under control (having missed it).
If average potential outcomes are continuous, we can estimate the local average treatment effect (LATE) around the cut-off:
c = -.05 by taking the difference in means above and below the threshold"

Data

For election outcomes, data from Brazil's Superior Electoral Court was used.

For IDEB scores and targets, the Ministry of Education's IDEB results for primary education at the level of the municipality.

For balance checks and further specifications, we use data from the 2010 census and from the Basic Municipal Information data set for 2009, both administered by Brazil's official statistics agency (IBGE, Instituto Brasileiro de Geografia e Estatística), as well from the Ministry of Education's yearly school census.

Three IDEB waves (2007, 2011, and 2015) was used, the results of which were published before the municipal elections of 2008, 2012, and 2016.

The effective sample excludes municipality-period observations (1) where the mayor is not eligible to run for re-election because of term limits.

When using vote share as the dependent variable, we also exclude observations where eligible mayors choose not to run.

Data Replication

Data 2020 replicated 2023

- RDD analysis – replicate the data on the Table 1. Effect of Meeting the IDEB Target on Re-election of the Mayor

	Linear	Linear	Robust	Robust
IDEB target met	-0.079*	-0.085**	-0.098*	-0.11**
	(0.045)	(0.043)	(0.053)	(0.056)
Election cycle fixed effects	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Bandwidth	0.396	0.388	0.396	0.388
Observations	1805	1770	1805	1770

Main Findings / Output comparison

What could have been added?

Observational research designs such as RDDs are subject to concerns about statistical modelling assumptions. RDD analysis of the effect of meeting IDEB targets relies on aggregate data, limiting the ability to test mechanisms about how voters process information generated by standardized tests.

A better approach could have been to run an OLS with municipalities elections from two years and look within municipalities at the effect of increases in the log of municipality spending on public education on whether the incumbent party was re-elected for mayor's office.

For municipality i in election year t:

$$Reelection_{i,t} = \beta_0 + \beta_1 * \ln(Educ_{i,t}) + \beta_1 * \ln(Budget_{i,t})$$
$$+ \delta' \mathbf{X}_{i,t} + \eta_i + \gamma_t + \varepsilon_{i,t}.$$

Main Findings / Output comparison

What could have been added?

Ordinary Least Squares regression (OLS) is a common technique for estimating coefficients of linear regression equations which describe the relationship between one or more independent quantitative variables and a dependent variable (simple or multiple linear regression).

```
> linear.m <- lm(incumbent_mayor_reelected ~ ideb_gap_centered, data = m)
> summary(linear.m)

Call:
lm(formula = incumbent_mayor_reelected ~ ideb_gap_centered, data = m)

Residuals:
    Min      1Q  Median      3Q     Max
-0.2813 -0.2787 -0.2783  0.7208  0.7237

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        0.2782528  0.0045500   61.16   <2e-16 ***
ideb_gap_centered 0.0008133  0.0062382    0.13    0.896
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4483 on 14166 degrees of freedom
  (2542 observations deleted due to missingness)
Multiple R-squared:  1.2e-06,   Adjusted R-squared:  -6.939e-05
F-statistic: 0.017 on 1 and 14166 DF,  p-value: 0.8963
```

Conclusion

Conceptual replications

Conceptual replication is when research is conducted by using different methods to repeat the original study.

Despite difference in methods, the new data is similar to the original study findings, by trying a different approach it can be observed the same results as per original paper but with high statistical evidence.

Questions

Thank you for your time!