

ECON2390 Fall 2024

Applied Econometrics

Problem Set 2

Adrien Foutelet

October 2024

1 Regression Discontinuity Design

1.1

We run OLS to estimate the following model:

$$Y_m = \beta_0 + \beta_1 \text{PANwin}_m + \epsilon_m$$

We get the following estimates (see table 1).

Table 1: Simple OLS regression model

	Y
(mean) PANwin	0.904 (0.06)
Constant	43.63*** (4.58)
Observations	152

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The positive correlation between variables Y and PANwin that the coefficient suggests does not appear to be significant.

We may be facing endogeneity:

- It is where the homicide rate is high and the drug trafficking organisations have strong presence that the party promoting radical anti drug policies is elected (reverse causality);
- PAN may have prioritized campaigning in dangerous areas where their agenda gave them the best chances to win (selection);
- In some places, the local culture may be more martial, making PAN most popular, a cultural trait that may result from weaker state capacity, which is suitable for criminal organisations to thrive (omitted variable bias).

We need to isolate exogenous variations in PANwin to infer causal results.

For these reasons, β_1 should be positively biased and β_0 should be negatively biased.

1.2

The key assumption necessary for the fuzzy regression discontinuity design to be valid in Dell's analysis is:

- The continuity assumption: there is continuity of the potential outcomes at the cutoff.

In practice, it is necessary that:

- All relevant factors besides treatment vary smoothly at the threshold between a PAN victory and loss.
- There is absence of selective sorting around the PAN win-loss threshold.

1.3

Investigating whether the regression discontinuity design shows evidence of manipulation, we do a McCrary test for the following null hypothesis:

$$H_0 : \text{"The density of the running variable is continuous at the cutoff"}$$

We get the following results (see table 2):

Our high p-value implies that we fail to reject the null hypothesis, meaning that we find no evidence of significant difference in the density of the running variable on either side of the cutoff.

We get the following plot of the estimated density of the running variable on both sides of the cutoff (see figure 1). No visual discontinuity can be found in the density at the cutoff: the density is smooth.

Table 2: McCrary density test

	margin_victory
LogDensityRatio	-0.0763
StdError	24.19
PValue	0.939
Bandwidth_left	0.0135
Bandwidth_right	0.0157

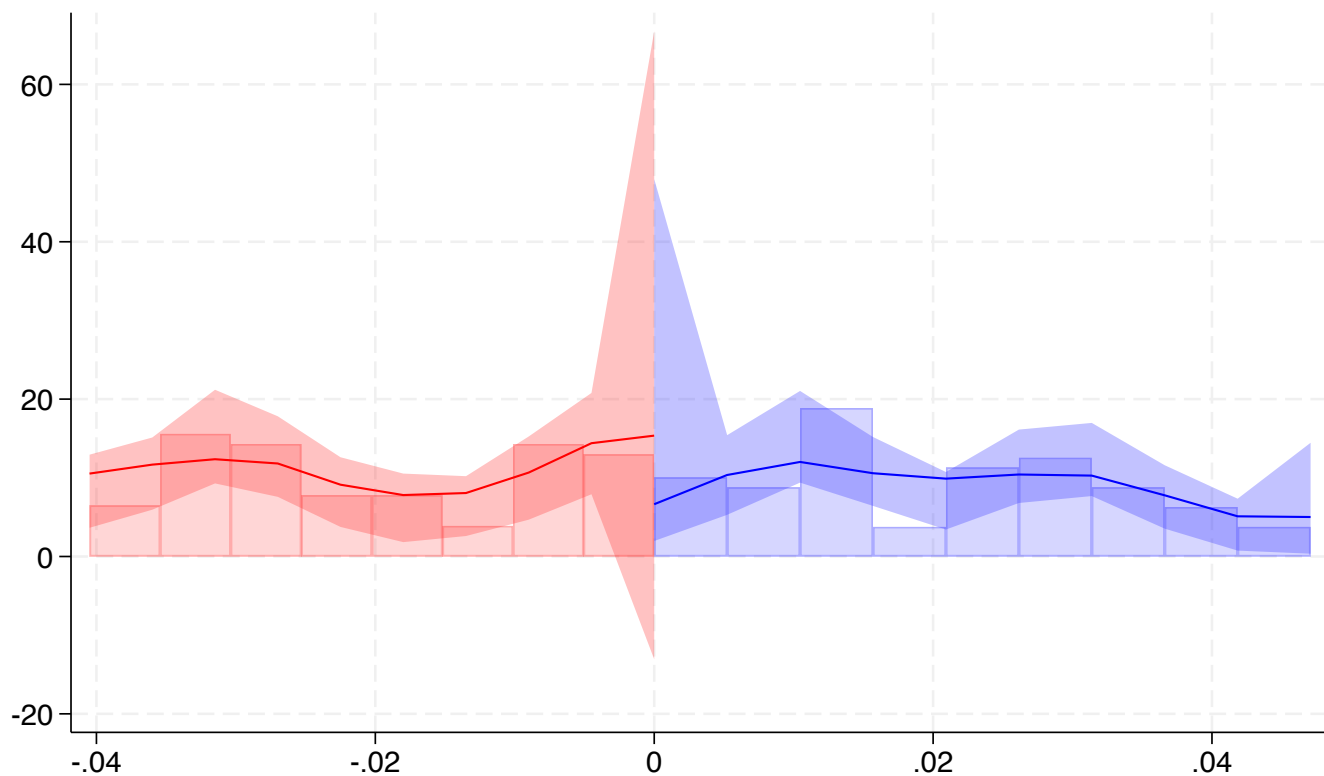


Figure 1: McCrary test result plot

It makes sense because it is very unlikely that the voters, in a general election, know the results in advance and change their behavior accordingly to get a precise margin of victory close to the cutoff.

1.4

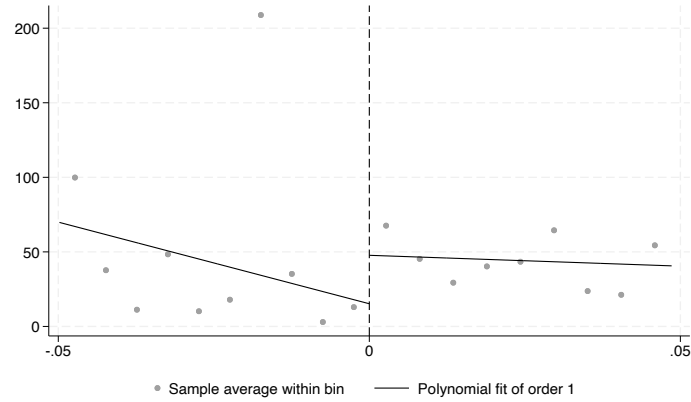


Figure 2: Binned scatter plot of PAN victory margin and homicide rate (number of bins selected via mimicking variance evenly-spaced method using spacings estimators, linear specification)

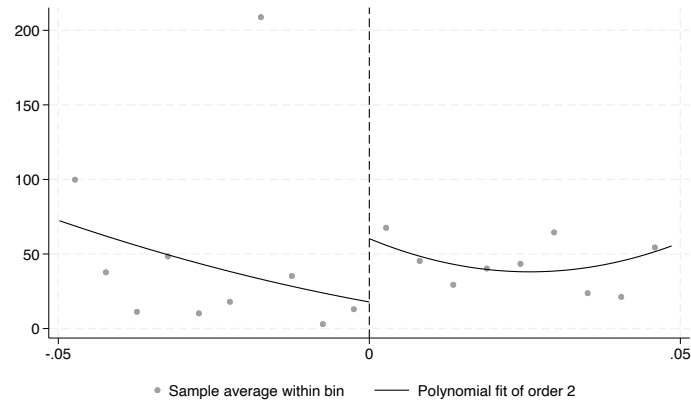


Figure 3: Binned scatter plot of PAN victory margin and homicide rate (number of bins selected via mimicking variance evenly-spaced method using spacings estimators, quadratic specification)

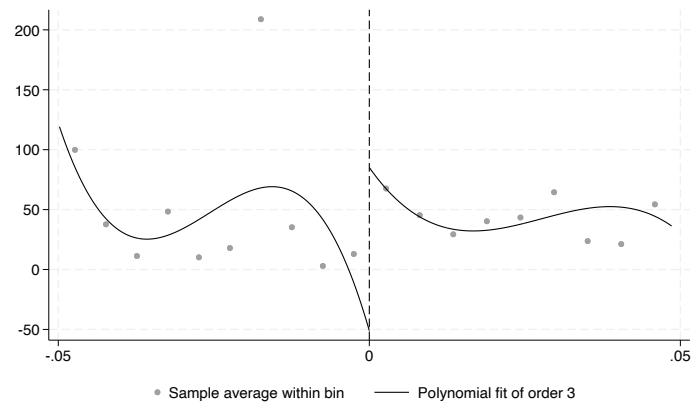


Figure 4: Binned scatter plot of PAN victory margin and homicide rate (number of bins selected via mimicking variance evenly-spaced method using spacings estimators, cubic specification)

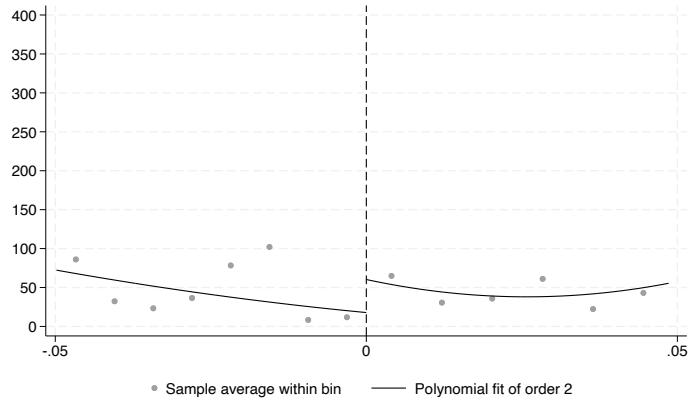


Figure 5: Binned scatter plot of PAN victory margin and homicide rate (number of bins selected via IMSE-optimal evenly-spaced method using spacings estimators, cubic specification)

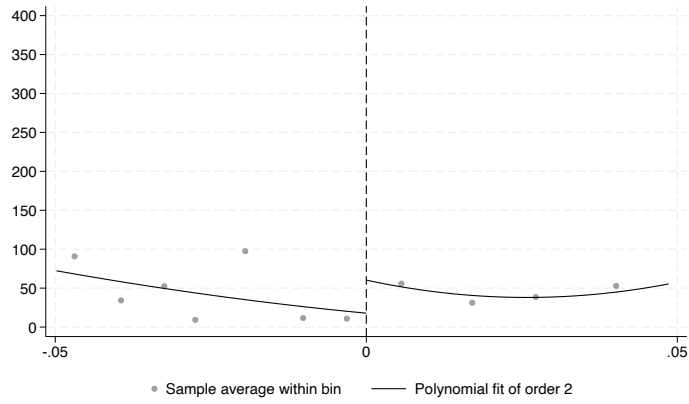


Figure 6: Binned scatter plot of PAN victory margin and homicide rate (number of bins selected via IMSE-optimal quantile-spaced method using spacings estimators, cubic specification)

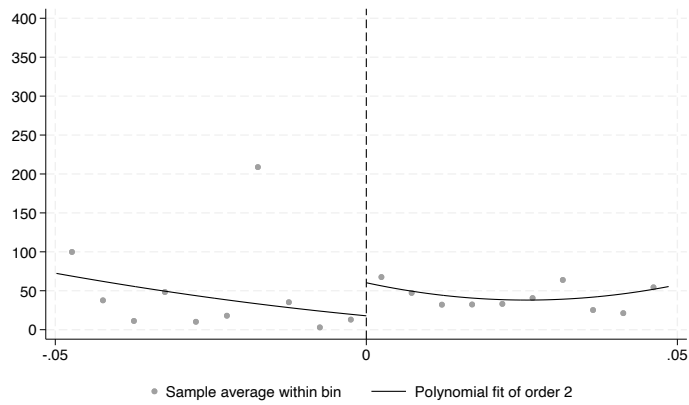


Figure 7: Binned scatter plot of PAN victory margin and homicide rate (number of bins normatively set at 10, cubic specification)

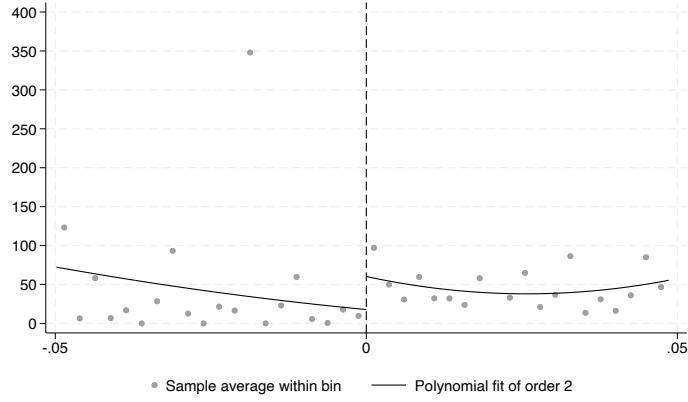


Figure 8: Binned scatter plot of PAN victory margin and homicide rate (number of bins bormatively set at 20, cubic specification)

The discontinuity at the cutoff is robust to the three polynomial specifications (see figures 2, 3, 4, 5, 6, 7, 8). However, the degree three specification seems to suffer from overfitting: there is no justification explaining the variations clost to the cutoff and the large discontinuity seems fabricated.

The discontinuity at the cutoff is robust to the five procedures for bin number selection, with no major difference in the size of the cutoff.

1.5

We get the following estimates (see table 3).

We get significant RDD estimates for the non-parametric-triangular-kernel-degree 1 case, for the parametric-degree 1 case, and for the parametric-degree 2 case. Hence the significance seems to be better overall for the parametric models in our case.

The magnitudes of the estimates of the parametric and non-parametric estimates are close to one another.

The magnitudes are greater than that of our rdplot discontinuities. That is due to the fact that the bandwidth used by default with rdplot differs from the optimal bandwidth selected by rdrobust. For example, if we choose the same bandwidth with rdplot for the parametric-linear case, we get a consistent discontinuity (see figure 9).

Table 3: Non-parametric and parametric RDD models

	Non-para tri lin	Non-para tri quad	Para lin	Para quad
RD Estimate	80.25*	93.93		
	(2.25)	(1.92)		
(mean) spread			979.5	2697.6
			(0.36)	(0.39)
(mean) PANwin			69.99**	81.79*
			(2.91)	(2.57)
(mean) PANwin=1 \times (mean) spread			-5770.9	-12951.7
			(-1.50)	(-1.31)
(mean) spread squared				260133.0
				(0.54)
(mean) PANwin=1 \times (mean) spread squared				138247.6
				(0.21)
Constant			12.95	14.37
			(0.78)	(0.66)
Observations	152	152	38	54

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

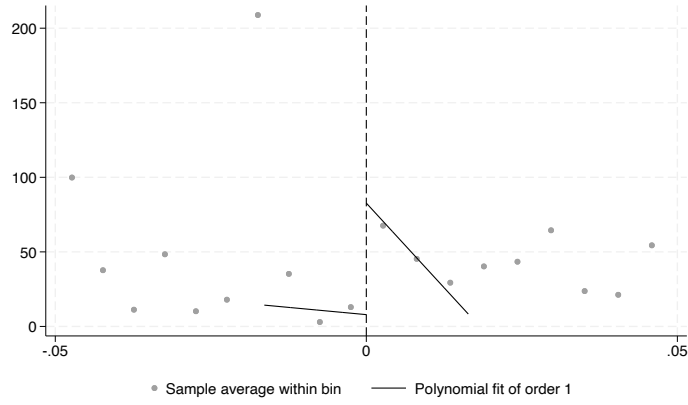


Figure 9: Binned scatter plot of PAN victory margin and homicide rate matching with a bandwidth matching the standard rdrobust one (number of bins selected via mimicking variance evenly-spaced method using spacings estimators, linear specification)

The parametric-specific RDD assumptions:

- A functional form is assumed for $E(Y | X)$, allowing for different parameters on either side of the cutoff.
- All observations are used regardless of their distance from the cutoff (the functional form holds globally).

The non-parametric-specific RDD assumptions:

- Local continuity of the potential outcomes i.e., potential outcomes are assumed to be smooth within a small neighborhood of the cutoff and the estimation focuses only on this local region,
- A bandwidth must be chosen around the cutoff to balance:
 - Bias from including too many observations far from the cutoff,
 - Variance from using too few observations,
- Observations are weighted based on their distance from the cutoff, typically giving more weight to those closer to it,
- The small sample nature of the local estimates calls for bias correction and robust variance estimation.

Parametric RDD is typically more precise (lower standard errors) but risks bias if the model is misspecified. Non-parametric RDD sacrifices precision (higher standard errors) for robustness, making it better suited for estimating local effects near the cutoff with fewer assumptions. The choice depends on the situation:

- If the functional form is likely well-specified: parametric methods give more precise estimates,
- If there is uncertainty about the functional form: non-parametric methods give more reliable but less precise results.

1.6

We get the following estimates (see table 4) from the donut hole approach. This tests whether the observed treatment effect is driven by potential manipulation or discontinuities unrelated to the treatment itself. Indeed it reveals whether the units near the cutoff influence the results.

More precisely, if individuals manipulate their running variable, it is likely that those near the cutoff be different across treatment and control groups. Therefore, a similar treatment effect with and without the donut hole suggests no manipulation near the cutoff. Note what is interpreted as manipulation could also just be local anomalies around the cutoff, like outliers. The McCrary test invalidated these possibilities.

Here we do find a plummeting treatment effect after excluding 20% of the optimal bandwidth on both sides. This apparent inconsistency with the results from the McCrary test is most proably explained by the fact that removing that many observations from the optimal bandwidth, especially the most weighted ones, is likely to reduce precision. We would need a criteria to choose an appropriate donut hole size.

Table 4: RDD model robustness

	Para tri lin	Para tri lin 5%	Para tri lin 10%	Para tri lin 20%
RD_Estimate	80.25*	80.25*	98.36*	48.95
	(2.25)	(2.25)	(2.14)	(1.00)
Observations	152	152	151	144

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

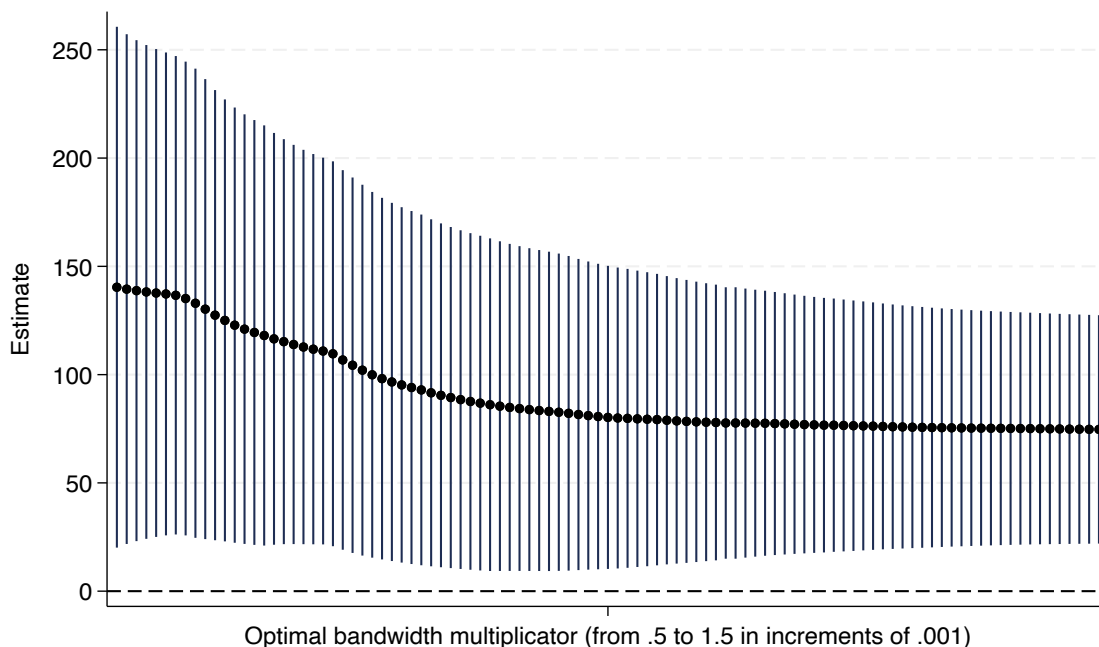


Figure 10: RDD model robustness

Using a loop to re-estimate the model with bandwidths ranging from 0.5 to 1.5 times the optimal bandwidth, we get the following graph (see figure 10). Recall that a bandwidth must be chosen around the cutoff to balance:

- Bias from including too many observations far from the cutoff,
- Variance from using too few observations,

Here, as we increase the bandwidth, the confidence interval decreases, betraying shrinking variance. The magnitude of the estimates plateaus at the level of the one computed with the optimal bandwidth. This is likely to be specific to our model and dataset as it is not necessarily the case in the context of increasing bias.

1.7

John Marshall's concern about compensating differentials in Politician Characteristic Regression Discontinuity (PCRD) designs hints at the idea that these designs fail to isolate the pure effect of a politician's characteristic (e.g., gender, party affiliation) because other characteristics, known as compensating differentials, influence both the election outcome and downstream the effects.

For example, if voters are biased against women, female candidates in close elections may need to compensate by having higher levels of competence or qualifications than their male opponents. The characteristic of interest (e.g., gender) correlates with compensating differentials that offset its impact (e.g., competence).

This means that PCRD designs compare candidates who are not equivalent on other important traits, leading to biased estimates of the effect of the characteristic of interest, since the effect measured is a compound treatment effect of the characteristic of interest and the compensating differentials.

Marshall recommends to attempt to measure compensating differentials to bound or reinterpret the results.

Dell uses close elections to argue that PAN mayoral victories causally increase drug-related violence. The compensating differentials from winning PAN candidates may be:

- Their strong-willed disruptive and uncompromising characters (prone to get drug dealers' backs up),
- Their expected lower level of corruption (prone to increase violence from the mobsters as the mayor would never compromise with them and cast a blind eye),
- Their stronger connections with federal authorities (prone to make anti crime policies even stronger locally and increase violence in retaliation),
- Their populist views rejecting sophisticated understanding of deviance (prone to make them both attractively radical and acting without tact).

There would then be selection of these traits for the mayors in the treatment group.

1.8

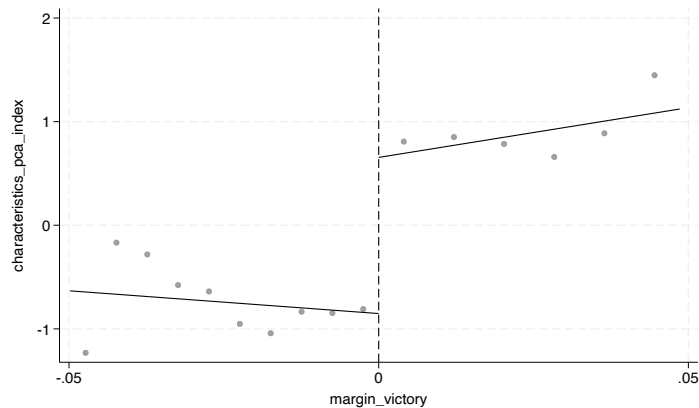


Figure 12: RDD model robustness

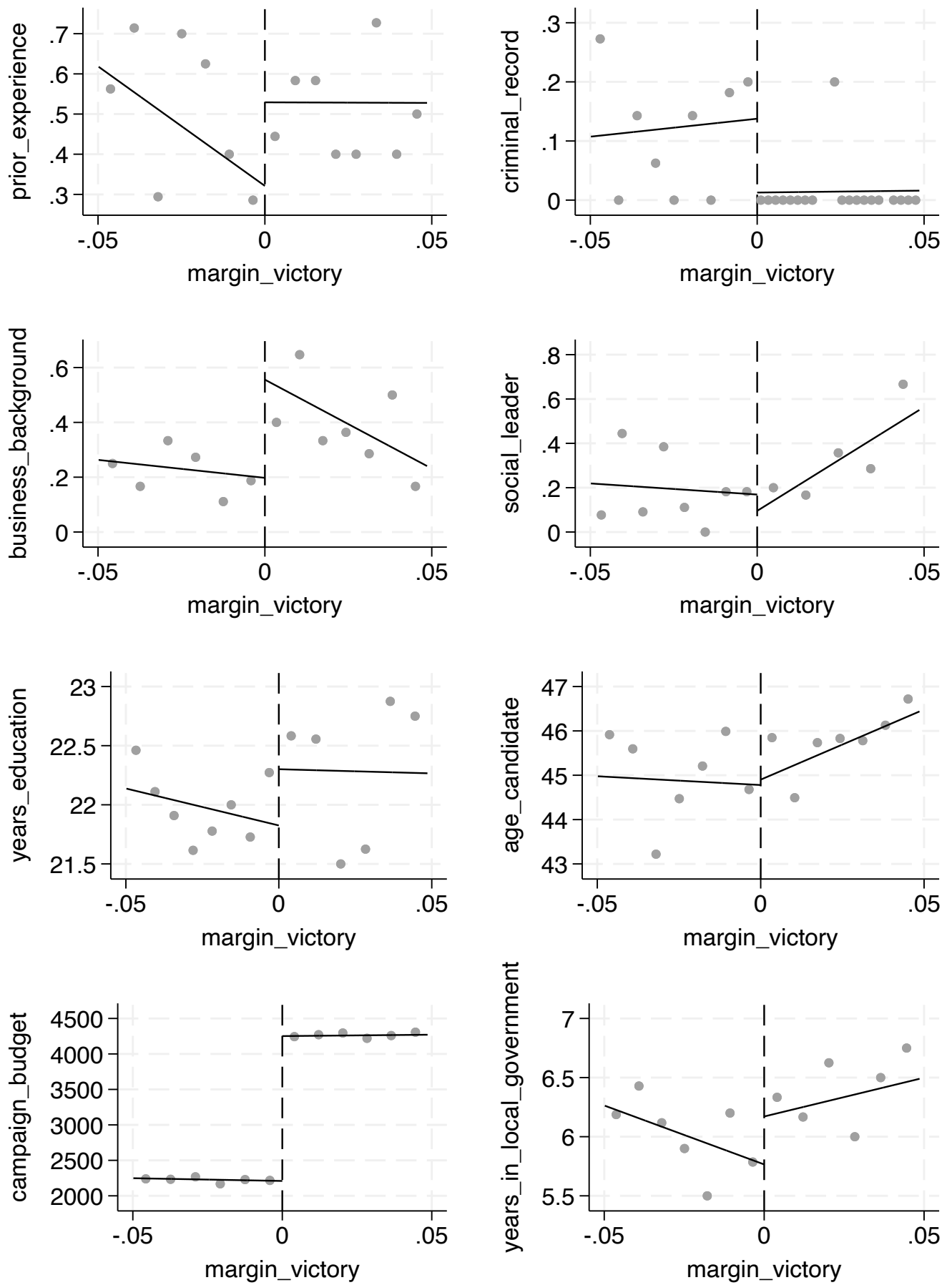


Figure 11: RDD model robustness

Table 5: Candidate characteristic RDD analysis

	pca_index	prior_experience	criminal_record	business_background	social_leader	years_education	age_candidate	campaign_budget	years_in_local_government
RD_Estimate	1.934** (2.84)	0.179 (0.60)	-0.303 (-1.26)	0.0150 (0.05)	0.0274 (0.11)	0.525 (0.60)	1.629 (0.98)	2038.0*** (21.42)	0.673 (0.93)
Observations	152	152	152	152	152	152	152	152	152

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Testing for discontinuities in candidate characteristics, we get the following estimates (see table 5 and figure 11). I include a principal component analysis index of all the characteristics together (see table 5 and figure 12). The RDD estimates are strongly significant for campaign budget and principal component analysis index. This implies that compensating differentials affect Dell's results: what we thought was the treatment effect of a PAN victory seems in fact to be the compound treatment effect of a PAN victory and the compensating differential in terms of campaign budget.

Covariate adjustment in a RDD is not sufficient to address the problem of compensating differentials, because these differentials introduce biases that are inherently linked to both observable and unobservable characteristics. Even if observable covariates balanced or controlled for, unobservable characteristics (e.g., competence) could vary discontinuously at the threshold, biasing the treatment effect. Also, covariates may include variables affected by the treatment itself (e.g., campaign funding postulated by voter - donors - and party organization expectations of winning). Including these covariates would introduce post-treatment bias (conditioning on factors influenced by the treatment), contaminating causal estimates.

All in all, covariate adjustment can reduce noise and account for observable imbalances, but it cannot fully address biases introduced by compensating differentials, especially when unobservable factors are at play.

1.9

To bound the effect of a PAN mayor on the homicide rate, we implement:

$$\hat{\tau}_{PCRD}^{corr} = \hat{\tau}_{PCRD} - \sum_k \hat{\gamma}_k \hat{\delta}_k$$

With:

- index i for candidates,
- index d for district,
- index k for covariates,
- $\hat{\gamma}_k$ an estimate of the LATE of each compensating differential Z_{idk} ,
- $\delta_k = E[Z_{idk} \mid \Delta_d = 0, X_{id} = 1, X_{jd} = 0] - E[Z_{idk} \mid \Delta_d = 0, X_{id} = 0, X_{jd} = 1]$,
- $\Delta_d = V_{id} - V_{0d}$,

- V_{id} and V_{0d} the vote shares of the most popular candidates of types $X_{id} = 1$ and $X_{id} = 0$ respectively,
- X_{id} a binary characteristic that helps candidate i win votes.

We implement it using the procedure described and get normally distributed correction terms centered on 5.26 and with a sample standard deviation of 2.66 (see figure 13).

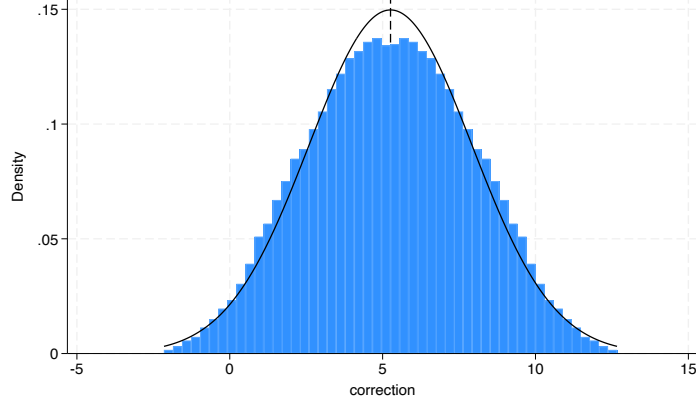


Figure 13: Distribution of the computed correction term

Recalling that $\hat{\tau}_{PCRD} = 80.25^* (2.25)$, the sign and the significance level of our estimate are robust to the (observable) effect of compensating differentials.

2

2.1

We compute the 2SLS estimator and the Wald estimator, as described (see table 6). The 2SLS and residualized Wald estimators are equal.

2.2

A researcher observing (Y_i, D_i, Z_i, X_i) would identify the conditional-on- X_i LATESs $E[Y_i(1) - Y_i(0) \mid X_i, G = CP]$ by running 2SLS on the dataset restricted to each level of X_i respectively. This is an intuition that we can suspect is not relevant considering how unbalanced our data is by construction.

Table 6: IV estimation

	2SLS	Wald
D	7.836*** (52.56)	7.836*** (52.27)
x1	0.137* (2.56)	
x2	0.0724 (1.54)	
x1x2	0.247*** (3.48)	
Constant	0.929*** (10.22)	1.096*** (13.18)
Observations	10000	10000

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: Unconditional and conditional LATEs

	2SLS X=(0,0)	LATE X=(0,0)	2SLS X=(1,0)	LATE X=(1,0)	2SLS X=(0,1)	LATE X=(0,1)	2SLS X=(1,1)	LATE X=(1,1)
D	6.680*** (16.95)	5.911*** (60.13)	8.434*** (37.83)	8.126*** (93.68)	6.502*** (12.48)	7.086*** (73.80)	9.139*** (47.82)	9.062*** (103.33)
Constant	1.587*** (6.97)	0.0624 (0.99)	0.828*** (8.78)	-0.0380 (-1.00)	1.808*** (5.70)	-0.0179 (-0.24)	0.600*** (5.08)	-0.0129 (-0.16)
Observations	2500	417	2500	833	2500	417	2500	833

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We compute estimated and real coefficients (see table 7). As we can see, the 2SLS coefficients are close but upper-biased.

2.3

We compute the convex-weighted average of the conditional LATEs by appropriately using sample means and variances through a series of collapses (see table 8). The groups that get the most weight are $X_i = (1, 0)$ and $X_i = (1, 1)$. For the 2SLS estimator to identify the unconditional LATE, we need to have:

$$\forall x \in \{(0, 0), (1, 0), (0, 1), (1, 1)\}, E[Y_i(1) - Y_i(0) \mid X_i, G = CP] = E[Y_i(1) - Y_i(0) \mid G = CP] \text{ (constant)}$$

In our data generating process, contrary to what we did, would, $Y_i(1)$ to be built orthogonal to X_i .

Table 8: Weights and convex weighted average of the conditional LATEs

Beta 2SLS	7.836
Weight00	0.542
Weight10	1.633
Weight01	0.300
Weight11	1.525
LATEsCvxWghtedAvg	8.104

2.4

If we exclude the $X_1.X_2$ interaction term from the set of controls, then the estimated effect through x_1 of $X_i = (1, 0)$ is the same as that of $X_i = (1, 1)$. Also, the estimated effect through x_2 of $X_i = (0, 1)$ is the same as the estimated effect of $X_i = (1, 1)$. It is still a convex-weighted average of treatment effects, this time of two treatment effects only.

2.5

Table 9: Estimated and true unconditional LATEs

	2SLS	LATE
D	8.227*** (69.70)	8.116*** (158.84)
Constant	0.883*** (13.39)	-0.0122 (-0.34)
Observations	10000	2500

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A researcher observing (Y_i, D_i, Z_i, X_i) would identify the unconditional LATEs $E[Y_i(1) - Y_i(0) \mid G = CP]$ by running 2SLS on the dataset without controlling for X_i . This is an intuition that we can suspect is not relevant considering how unbalanced our data is by construction.

We compute estimated and real coefficients (see table 9). As we can see, the 2SLS coefficient is close but upper-biased.

References

- [1] Dell, Melissa. "Trafficking Networks and the Mexican Drug War." *American Economic Review* 105, no. 6 (2015): 1738-1779.
- [2] Marshall, J. (2024), Can Close Election Regression Discontinuity Designs Identify Effects of Winning Politician Characteristics?. *American Journal of Political Science*, 68: 494-510.
- [3] Joshua D. Angrist, Guido W. Imbens, 1995. "Average Causal Response with Variable Treatment Intensity," NBER Technical Working Papers 0127, National Bureau of Economic Research, Inc.