On the perils of Normalizing-and-Pooling in RD designs

Margherita Fort (Univ. of Bologna, CESifo, IZA)

Andrea Ichino (EUI, CEPR, CESifo, IZA)

Enrico Rettore (Univ. of Padova, FBK-IRVAPP, IZA)

Giulio Zanella (Univ. of Bologna, IZA)

SIDE webinar

Dec. 9, 2020

1. The (sharp) RDD in pills

Individuals in a target population are either exposed – $\boldsymbol{D}$=1 – or unexposed – $\boldsymbol{D}$=0 – to an intervention. The target is measuring the *average causal effect* of exposure on an outcome $\boldsymbol{Y}$ using a sample of individuals from the target population.

$\boldsymbol{Y_1}$ and $\boldsymbol{Y_0}$ are the *potential* outcomes under exposure/non exposure to the intervention.

$\boldsymbol{Y} = \boldsymbol{Y_0} + (\boldsymbol{Y_1} - \boldsymbol{Y_0})\boldsymbol{D}$ is the outcome we observe in the data.

The RDD selection rule. To select individuals into exposure they are ranked according to a *continuous* observable variable $\boldsymbol{X}$ - the *running variable*, predetermined to the intervention.

$f(x)$ is the pdf of $\boldsymbol{X}$.

Then, the exposure status of individuals *deterministically* follows from the rule:

$$\boldsymbol{D} = \mathrm{I}(\boldsymbol{X} \geq \boldsymbol{c}), \qquad (1)$$

c – the *cutoff* - a point in the support of ***X*** known in advance.

Examples. Incentives to students awarded on the basis of an observed test score; mandatory training programmes for individuals with more than a certain number of months in unemployment.

We name it *Allocation Rule 1*.

To the left of the cutoff ***c*** we observe:

$$E\{\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{c} - \varepsilon\} = E\{\boldsymbol{Y_0}|\boldsymbol{X} = \boldsymbol{c} - \varepsilon\} \qquad (2)$$

ε a ‘small’ *positive* number. To the right of the cutoff ***c*** we observe:

$$\begin{aligned} E\{\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{c} + \varepsilon\} &= E\{\boldsymbol{Y_1}|\boldsymbol{X} = \boldsymbol{c} + \varepsilon\} \\ &= E\{\boldsymbol{Y_0}|\boldsymbol{X} = \boldsymbol{c} + \varepsilon\} + E\{\boldsymbol{Y_1} - \boldsymbol{Y_0}|\boldsymbol{X} = \boldsymbol{c} + \varepsilon\}. \qquad (3) \end{aligned}$$

Comparing individuals immediately to the right of the cutoff to those immediately to the left of it:

$$\begin{aligned} E\{\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{c} + \varepsilon\} - E\{\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{c} - \varepsilon\} &= E\{\boldsymbol{Y_1} - \boldsymbol{Y_0}|\boldsymbol{X} = \boldsymbol{c} + \varepsilon\} \\ &\quad + [E\{\boldsymbol{Y_0}|\boldsymbol{X} = \boldsymbol{c} + \varepsilon\} - E\{\boldsymbol{Y_0}|\boldsymbol{X} = \boldsymbol{c} - \varepsilon\}] \qquad (4) \end{aligned}$$

The RDD *identifying restriction*. E{$\boldsymbol{Y_0}$|$\boldsymbol{X}$} is a continuous function of $\boldsymbol{X}$ at $\boldsymbol{X}$=$\boldsymbol{c}$:

$$E\{\boldsymbol{Y_0}|\boldsymbol{X} = \boldsymbol{c} + \varepsilon\} \approx E\{\boldsymbol{Y_0}|\boldsymbol{X} = \boldsymbol{c} - \varepsilon\} \qquad (5)$$

If (5) holds, the difference in (4) identifies the average causal effect on *marginal* individuals exposed to the intervention, E{$\boldsymbol{Y_1}$ – $\boldsymbol{Y_0}$|$\boldsymbol{X}$ = $\boldsymbol{c}$ + ε}.

The intuition behind the requirement of *continuity* E{$\boldsymbol{Y_0}$|$\boldsymbol{X}$} at $\boldsymbol{X}$=$\boldsymbol{c}$. For the difference in (4) to deserve a causal interpretation it must be that in the absence of the intervention no discontinuity would be observed at $\boldsymbol{c}$.

To implement it:

run the regression of the outcome ***Y*** on the running variable ***X*** separately to the right and to the left of the cutoff ***c***.

Measure the jump at ***c***: $E\{\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{c} + \varepsilon\} - E\{\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{c} - \varepsilon\}$

This is done using nonpar methods, e.g. Local Linear Regression. Rules for selecting the optimal bandwidth (see Calonico, Cattaneo and Titiunik, 2014; Imbens and Kalyanaraman, 2012).

1.1. The multi-cutoff case.

What if the intervention is *implemented in J different sites*. Individuals are ranked on the running variable separately in each site, the cutoff ***c**_j* possibly varying across sites.

Example: Fort, Ichino and Zanella (2019), non-/cognitive costs of daycare for children 0-2.

Nothing new if the sample size is enough large for an analysis site by site. If not, the ony way out is to content yourself with an across-site average causal effect.

The *Normalizing-and-Pooling* (NP) estimator (Cattaneo et al., 2016):

Normalize the running variable in each site: $\boldsymbol{X} - \boldsymbol{c_j}$

Pool together the J sites and proceed as before using the *normalized* running variable and the *cutoff at zero*. As if it were a standard RDD.

The case we analyze - we name it *Allocation Rule 2*:

In each site it is the *number of available slots* to be predetermined, K_j, j = 1, J.

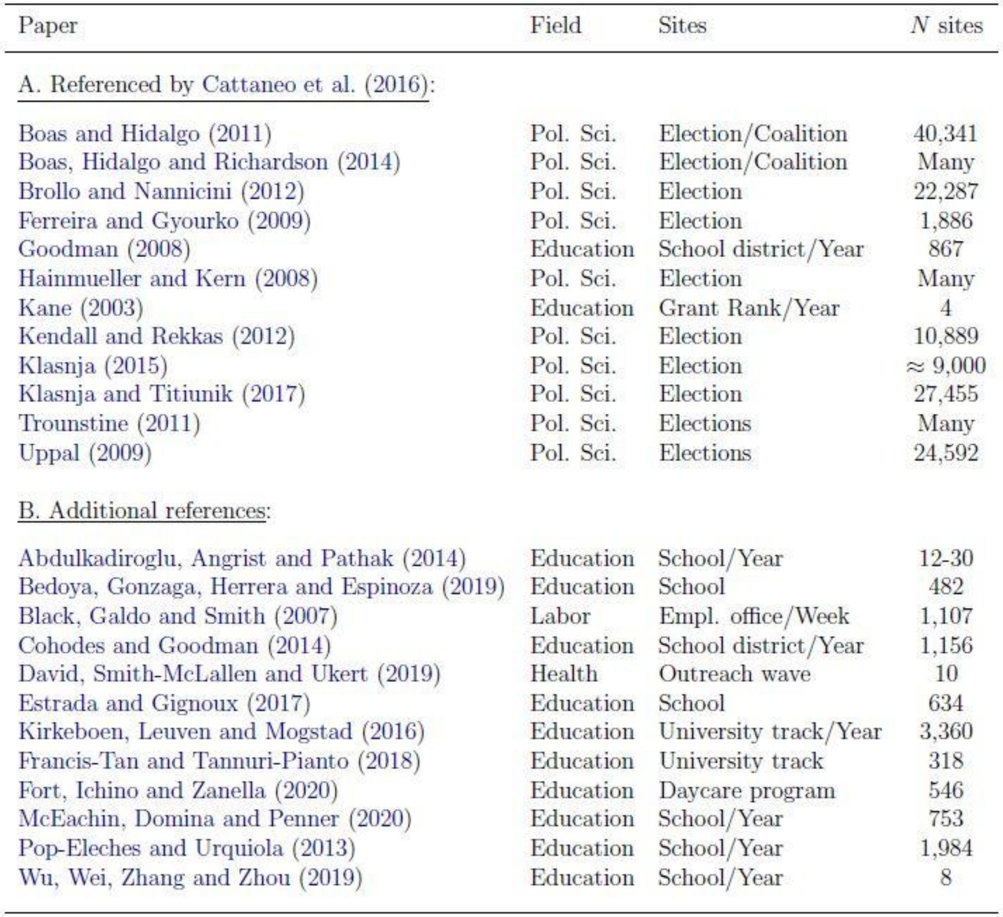
There are N_j applicants in site j. Slots are filled starting from the highest-score applicant, until exhaustion. The cutoff $\boldsymbol{c_j}$ *ex-post* coincides with the score of the marginal unit exposed to the intervention.

We show the NP estimator might fail in this set-up. Even if the continuity condition holds in each site!

To avoid adding a layer of difficulty, we focus on the case in which units do *not* choose their site.

2. Multi-site RDD with *Allocation Rule 2*: examples

In our cursory survey we found 24 papers in economics/political science published since 2007 falling into this set-up.



Paper	Field	Sites	N sites
A. Referenced by Cattaneo et al. (2016):			
Boas and Hidalgo (2011)	Pol. Sci.	Election/Coalition	40,341
Boas, Hidalgo and Richardson (2014)	Pol. Sci.	Election/Coalition	Many
Brollo and Nannicini (2012)	Pol. Sci.	Election	22,287
Ferreira and Gyourko (2009)	Pol. Sci.	Election	1,886
Goodman (2008)	Education	School district/Year	867
Hainmueller and Kern (2008)	Pol. Sci.	Election	Many
Kane (2003)	Education	Grant Rank/Year	4
Kendall and Rekkas (2012)	Pol. Sci.	Election	10,889
Klasnja (2015)	Pol. Sci.	Election	≈ 9,000
Klasnja and Titiunik (2017)	Pol. Sci.	Election	27,455
Trounstine (2011)	Pol. Sci.	Elections	Many
Uppal (2009)	Pol. Sci.	Elections	24,592
B. Additional references:			
Abdulkadiroglu, Angrist and Pathak (2014)	Education	School/Year	12-30
Bedoya, Gonzaga, Herrera and Espinoza (2019)	Education	School	482
Black, Galdo and Smith (2007)	Labor	Empl. office/Week	1,107
Cohodes and Goodman (2014)	Education	School district/Year	1,156
David, Smith-McLallen and Ukert (2019)	Health	Outreach wave	10
Estrada and Gignoux (2017)	Education	School	634
Kirkeboen, Leuven and Mogstad (2016)	Education	University track/Year	3,360
Francis-Tan and Tannuri-Pianto (2018)	Education	University track	318
Fort, Ichino and Zanella (2020)	Education	Daycare program	546
McEachin, Domina and Penner (2020)	Education	School/Year	753
Pop-Eleches and Urquiola (2013)	Education	School/Year	1,984
Wu, Wei, Zhang and Zhou (2019)	Education	School/Year	8

3. The analogy to the Randomized Control Trial (RCT)

If the RDD identifying restriction (5) holds – i.e. if the exposure status of units is truly determined by the sign of ($\boldsymbol{X} - \boldsymbol{c}$) - it is *as if* in the vicinity of the cutoff a RCT took place.

Thistlethwhite and Campbell forcefully made this point sixty years ago.

An intervention implemented in J different sites following the RDD protocol is analogue to a *stratified* RCT, each site being a stratum within which the exposure status is randomly determined locally at the cutoff $\boldsymbol{c_j}$.

If $\boldsymbol{c_j}$ is *determined in advance* (*Allocation Rule 1*) the expected number of units in a ε right/left neighbourhood of the cutoff is $N_j\, f_j\, \varepsilon$, where f_j is the density of $\boldsymbol{X}$ at the cutoff in site j.

That is, locally at the cutoff:

$$\Pr(\boldsymbol{D}=1 | \boldsymbol{X} \approx \boldsymbol{c_j}, \text{ site } j) = 0.5 \tag{6}$$

Due to the ‘local’ randomness of ***D***, the regression:

$$Y_{ij} = \alpha + \beta D_{ij} + v_{ij} \qquad\qquad j = 1, J \qquad\qquad i = 1, N_j \tag{7}$$

estimated restricting the sample to units ‘close’ to the cutoff identifies the average causal effect.

Note that as far as identification is concerned there is no need to include *site fixed-effects* in the regression.

The site fixed-effects are:

$$E\{\boldsymbol{Y_0}|\boldsymbol{X}\approx \boldsymbol{c_j}, \text{site } j\} = \alpha + E\{v|\boldsymbol{X}\approx \boldsymbol{c_j}, \text{site } j\}, \qquad j = 1, J \qquad (8)$$

Condition (6) implies zero correlation between the fixed-effect and ***D***.

Still, including site fixed-effects results in a *more precise* estimate of the causal parameter as far as sites are heterogeneous wrt the average outcome $\boldsymbol{Y_0}$ (see for instance Athey and Imbens, 2016).

Also note that in this set-up you don’t need to adjust the *standard* standard error by the *Moulton factor*: it is *one* because – again *thanks to (6)* - the intraclass correlation of ***D*** is zero.

When it is the number of slots per site – not the cutoff $\boldsymbol{c_j}$! – to be determined in advance (*Allocation Rule 2*) the cutoff $\boldsymbol{c_j}$ is the score of the marginal unit exposed to the intervention.

We show:

- Pr($\boldsymbol{D}$=1|$\boldsymbol{X}$≈$\boldsymbol{c_j}$, site j) ≠ 0.5 and varying across sites. As a consequence…
- …the site fixed-effect (8) might be correlated to ***D***.

Bottom line: the regression of ***Y*** on ***D*** (locally at the cutoff) might fail to identify the causal parameter.

4. The anatomy of the problem

Under *Allocation Rule 2*, in site j:

- not exposed, $N_j - K_j$ units to the left of the cutoff
- exposed, 1 unit *exactly* at the cutoff $\boldsymbol{c_j}$
$K_j - 1$ units to the right of the cutoff

The *expected* number of unexposed/exposed units in a neighborhood of $\boldsymbol{c_j}$:

- $N_j f_j \varepsilon$ in $(\boldsymbol{c_j} - \varepsilon, \boldsymbol{c_j})$, the same as under *Allocation Rule 1*
- $N_j f_j \varepsilon + 1 - N_j f_j \varepsilon / K_j$ in $[\boldsymbol{c_j}, \boldsymbol{c_j} + \varepsilon)$

As a result:

$$\Pr(\boldsymbol{D}=1|\boldsymbol{X}\approx\boldsymbol{c_j}, \text{ site j}) = [1- 1/K_j + 1/N_j\ f_j\ \varepsilon] \ / \ [2- 1/K_j + 1/N_j\ f_j\ \varepsilon] \neq 0.5 \qquad (9)$$

In addition, this probability *varies across sites* as a function of N_j, f_j and K_j.

The main implication is that if the site fixed-effect E{$\boldsymbol{Y_0}$| $\boldsymbol{X}\approx\boldsymbol{c_j}$, site j} is correlated to Pr($\boldsymbol{D}$=1|$\boldsymbol{X}\approx\boldsymbol{c_j}$, site j_j) the NP is *biased*.

But wait… biased wrt what?

4.1. The parameter of interest

Under *Allocation Rule 1* there is no ambiguity in the definition of the causal parameter:

$$\sum_j w_j \, E\{\boldsymbol{Y_1} - \boldsymbol{Y_0} | \boldsymbol{X}=\boldsymbol{c_j}, \text{ site } j\} \qquad (10)$$

where w_j is the weight of site j in the weighted average providing the average causal effect across sites at the cutoff:

$$w_j \propto N_j \, f_j \, \varepsilon \qquad (11)$$

This is no longer the case under *Allocation Rule 2*.

As a result of (9), the weight of site j among marginally *unexposed* individuals – i.e. those in ($\boldsymbol{c_j}$ -ε, $\boldsymbol{c_j}$) – is the same as in under *Allocation Rule 1*, $w_j^- = w_j$.

On the other hand, the weight of site j among marginally *exposed* units – i.e. those in [$\boldsymbol{c_j}$, $\boldsymbol{c_j}$ + ε) is:

$$w_j^+ \propto 1 + (1\text{-}1/K_j)\ N_j\ f_j\ \varepsilon \qquad (12)$$

Bottom line: under *Allocation Rule 2* there are *two* different causal parameters, one for those marginally unexposed to the intervention – ATNT - the other one for those marginally exposed to it – the ATT.

They coincide if and only if:

$$\textstyle\sum_j (w_j^- - w_j^+)\ \mathrm{E}\{\boldsymbol{Y_1} - \boldsymbol{Y_0} | \boldsymbol{X}{=}\boldsymbol{c_j}, \text{ site j}\} = 0 \qquad (13)$$

That is, iff the difference between the two weights is uncorrelated to the average causal effect.

In the following, we focus on the average treatment effect on marginally unexposed units (ATNT).

This is often - albeit not always – a parameter of interest in a RDD because it aswers the question ‘what would be the impact of a marginal expansion of the intervention?’

The analysis for the ATT develops by analogy.

4.2. The bias of the NP estimator (for the ATNT)

The bias is:

$$\sum_j (w_j^+ - w_j^-) \, E\{\boldsymbol{Y_1} | \boldsymbol{X}{=}\boldsymbol{c_j}, \text{site } j\} \qquad (14)$$

Ingredients of the bias:

- The average value (across sites) of $N_j f_j$ and of $N_j f_j/K_j$
- The degree of across site heterogenity wrt $E\{\boldsymbol{Y_1}|\boldsymbol{X}\approx\boldsymbol{c_j}$, site j$\}$, $N_j f_j$ and $N_j f_j/K_j$
- The correlation between $E\{\boldsymbol{Y_1}|\boldsymbol{X}\approx\boldsymbol{c_j}$, site j$\}$ and $N_j f_j$
- The correlation between $E\{\boldsymbol{Y_1}|\boldsymbol{X}\approx\boldsymbol{c_j}$, site j$\}$ and $N_j f_j/K_j$

In particular, the bias is zero if *at least one* of the following conditions hold:

- N_j f_j grows large in each site
- in the absence of across site heterogeneity either wrt E{$\boldsymbol{Y_1}$| $\boldsymbol{X}$≈$\boldsymbol{c_j}$, site j} or wrt (N_j f_j, K_j)
- corr{E{$\boldsymbol{Y_1}$|$\boldsymbol{X}$≈$\boldsymbol{c_j}$, site j}, N_j f_j} = corr{E{$\boldsymbol{Y_1}$|$\boldsymbol{X}$≈$\boldsymbol{c_j}$, site j}, N_j f_j/K_j} = 0

The first condition implies there is no need to pool data across sites. But here we focus on the case in which N_j is 'small', i.e. one cannot help pooling…

The second condition means that either the fixed-effect or the (average of the) explanatory variable ***D*** does not vary across sites.

The third condition implies that the site fixed-effect is uncorrelated to the explanatory variable ***D***.

4.3. What if excluding marginal participants

The origin of the problem is that as a result of *Allocation Rule 2* there is one exposed unit exactly at the cutoff in each site.

What if we throw away those units?

The probability of exposure to the intervention around the cutoff:

$$\Pr(\boldsymbol{D}=1|\boldsymbol{X}\approx\boldsymbol{c_j}, \text{site } j) = (1-1/K_j) / (2-1/K_j) < 0.5 \qquad (15)$$

still varying across sites, unless K_j is constant (or large) across sites.

That is, even here there is room for across site correlation between the site fixed-effect and the (average of the) explanatory variable ***D***.

4.4. Digging into the bias of the NP estimator

Discussion so far points to the existence of a bias of the NP estimator (whether or not units at the cutoff are kept in the sample).

The question is how much this bias is relevant in practice.

Keeping units at the cutoff in the sample, the leading term of the bias *when N_j is 'small'* is:

$$\text{bias}_1 = -\ \text{Cov}\{\ E\{\boldsymbol{Y_1} \mid \boldsymbol{X} \approx \boldsymbol{c_j}, \text{ site j}\},\ N_j f_j / \overline{Nf}\ \} \tag{16}$$

where $\overline{Nf}$ is the average of $N_j f_j$ across sites.

For example, in an educational context the (absolute) value of this covariance is large when schools that attract students with the best outcome under exposure are also more popular, i.e., they attract a larger fraction of applicants (in general and in a neighborhood of the cutoff).

Discarding units at the cutoff from the sample the bias is:

$$\text{bias}_2 = [-\ \text{bias}_1 - \text{Cov}\{\ \text{E}\{\boldsymbol{Y_1} |\ \boldsymbol{X} \approx \boldsymbol{c_j}, \text{site j}\}, \text{h}_j/\bar{h}\ \}] * \bar{h}\ /\ (\overline{Nf} - \bar{h}) \qquad (17)$$

where:

$$\text{h}_j = \text{f}_j\ /\ (\text{K}_j/\text{N}_j) = \text{f}(\boldsymbol{c_j}\ |\ \boldsymbol{X} \geq \boldsymbol{c_j})$$

is the *hazard function* of the distribution of the running variable at the cutoff $\boldsymbol{c_j}$.

Then:

$$\text{bias}_2 / \text{bias}_1 = [- 1 + \text{Cov}\{ \text{E}\{\boldsymbol{Y_1}|\ \boldsymbol{X}\approx\boldsymbol{c_j}, \text{site j}\}, \text{h}_\text{j}/\bar{h}\ \} / \text{Cov}\{ \text{E}\{\boldsymbol{Y_1}|\ \boldsymbol{X}\approx\boldsymbol{c_j}, \text{site j}\}, \text{N}_\text{j}\text{f}_\text{j}/\overline{Nf}\ \}] * \bar{h} / (\overline{Nf}\text{-}\bar{h}) \qquad (18)$$

$\bar{h}$ / ($\overline{Nf}$-$\bar{h}$) < 1. Then, whether bias$_2$ < bias$_1$ it depends on the sign and the size of the two covariances.

5. Adding site fixed-effect to the picture.

The intuition here is straightforward. Regressing ***Y*** on ***D*** pooling all sites together provides a biased estimate when there is correlation between the site fixed-effect and the explanatory variable.

Then… add site fixed-effect!

The estimand of the fixed-effect estimator (FE) is:

$$\sum_j p_j \, [E\{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{c_j}+\varepsilon, \text{site } j\} - E\{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{c_j}-\varepsilon, \text{site } j\}] = \sum_j p_j \, E\{\boldsymbol{Y_1} - \boldsymbol{Y_0}|\boldsymbol{X}=\boldsymbol{c_j}, \text{site } j\} \qquad (16)$$

the equality following from the continuity of E{$\boldsymbol{Y_0}|\boldsymbol{X}$, site j} at $\boldsymbol{X}=\boldsymbol{c_j}$ in each site.

Contrast it to the estimand of NP:

$$\sum_j w_j^+ E\{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{c_j} + \varepsilon, \text{site } j\} - \sum_j w_j^- E\{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{c_j} - \varepsilon, \text{site } j\} = \sum_j w_j^+ E\{\boldsymbol{Y_1}|\boldsymbol{X}=\boldsymbol{c_j}, \text{site } j\} - \sum_j w_j^- E\{\boldsymbol{Y_0}|\boldsymbol{X}=\boldsymbol{c_j}, \text{site } j\} \tag{17}$$

The trick of FE is straightforward:

- the NP estimator evaluates the average outcome *across sites* separately right and left to the cutoff, using different weights. Then, take the difference.

- the FE estimator evaluate the average causal effect within in each site. Then, take the average across sites.

Note however that the FE weight of site j in (16) is:

$$p_j \propto [1 + (2\text{-}1/K_j)\, N_j\, f_j\, \varepsilon]\ \Pr(\boldsymbol{D}{=}1|\boldsymbol{X}{\approx}\boldsymbol{c_j},\ \text{site j})\ \Pr(\boldsymbol{D}{=}0|\boldsymbol{X}{\approx}\boldsymbol{c_j},\ \text{site j}) \neq N_j\, f_j\, \varepsilon \qquad (18)$$

That is, the FE estimand (16) is a *meaningful causal parameter* but the weights are *not* those required to get the ATNT (nor the ATT).

Whether this set of weights makes a relevant difference wrt the set of weight one has in mind it is an empirical issue.

Reweighting. It requires estimating the quantity in (17). But be careful: with ‘small’ N_j estimating it implies more noise into the estimate.

But… removing units at the cutoff:

$$p_j \propto (K_j-1)/[2*(2*K_j-1)] * N_j f_j \varepsilon \neq N_j f_j \varepsilon \qquad (19)$$

that is no need to estimate the reweighting factor, it depends only on K_j.

5.1. Implementing the nonpar fixed-effect RDD

To estimate the two averages E{$\boldsymbol{Y}|\boldsymbol{X}$ = $\boldsymbol{c}$ + ε} and E{$\boldsymbol{Y}|\boldsymbol{X}$ = $\boldsymbol{c}$ - ε} one typically uses local linear regressions which require selecting an optimal bandwidth around the cutoff.

To our knowledge, there is no theory yet to select the optimal bandwidth for an RDD in the presence of group fixed-effect.

In our experiments (not reported here) we implemented the following heuristic procedure:

1) Start with a tentative bandwidth, bw_0
2) At iteration j, apply the within-group transformation using bw_{j-1} and pool the wg-transformed data across groups. Evaluate the optimal bandwidth on the wg-transformed data, bw_j and estimate the causal effect.
3) Replace bw_{j-1} by bw_j and repete the steps. Up to convergence.

6. Alternative estimators

6.1. Double Normalizing

It exploits the analogy to first-differencing in panel data models. The NP normalizes the running variable by taking the difference $\boldsymbol{X}$-$\boldsymbol{c}_j$ separately in each site.

Do the same with the outcome of *unexposed* individuals by taking the difference between their outcome and the outcome of the exposed individual at the cutoff, separately in each site.

Then, pool the samples of 'doubly-normalized' unexposed individuals across sites and run a standard RDD.

Exposed individuals other than the one at the cutoff do *not* contribute to this estimator.

In our experiments the DN estimator performs well, just slightly less precise than the FE.

6.2. Rank distance

Abdulkadiroglu, Angrist and Pathak (2014) use *ranks* of individuals in their own site/group as the running variable, then proceed using the NP estimator.

This amounts to convert the running variable in the original metric using its empirical cdf, a *non-decreasing monotonic* mapping.

This mapping preserves the order of individuals *within* each site. Note however that if the number of individuals varies across sites *this mapping is no longer order preserving*, i.e. an individual 'close' to the cutoff in the original metric might end up 'far away' of it in the rank metric.

In Abdulkadiroglu, Angrist and Pathak (2014) this is presumably not a problem because the size of their groups is approximately the same.

We show that 'small' sites/groups are penalized by this strategy, i.e. they get a weight *smaller* than the one they would get in the original metric.

6.3. Symmetric distance

Boas and Hidalgo (2011) posit that in each site the cutoff relevant for unexposed individuals is the score of the last exposed individual, while the cutoff relevant for exposed individuals is the score of the first unexposed individual.

This way there is no individual exactly at the cutoff.

Note however that this strategy introduces in each site a *positive difference* between the cutoff relevant for unexposed individuals - $\boldsymbol{c^u_j}$ - and the cutoff relevant for the exposed ones – $\boldsymbol{c^e_j}$, since by definition $\boldsymbol{c^u_j} > \boldsymbol{c^e_j}$.

As an implication, if E{$\boldsymbol{Y_0}$|$\boldsymbol{X}$} varies with $\boldsymbol{X}$ around the cutoff point(s):

$E\{\boldsymbol{Y_0}|\ \boldsymbol{c^e_j} + \varepsilon\} - E\{\boldsymbol{Y_0}|\ \boldsymbol{c^u_j} - \varepsilon\} \neq 0$

That is, a violation of the RDD identifying restriction.

The resulting bias converges to zero as N_j grows large because ($\boldsymbol{c^u_j} - \boldsymbol{c^e_j}$) -> 0, but it might be non negligible when the number of individuals per site is small.

7. Summing up

The strenght of the RDD comes from its close analogy to a RCT, even if only locally at the cutoff relevant for selection into exposure.

This analogy breaks down when the RDD is implemented pooling data from J different groups/sites, if the following conditions *jointly* hold:

- the probability of exposure at the cutoff varies across groups/sites
- this probability is correlated to the site average potential outcomes.

If selection into exposure is defined in each group/site by setting in advance a *cutoff point* in the support of the running variable - i.e. *Allocation Rule 1*, in our terminology - the first condition does <u>not</u> hold. Hence, the problem does <u>not</u> arise.

Instead, if selection into exposure is defined by setting in advance the *number of exposed individuals* in each group/site – i.e. *Allocation Rule 2*, in our terminology - this problem might arise.

According to our survey of published papers this set-up is quite common in empirical economics/political science.

Under *Allocation Rule 2*, the heuristic procedure to normalize at zero the running variable in each group/site and to pool the J groups/sites together – as if it were a single ranking on the running variable – might produce a biased estimate of the causal effect.

There is no theoretical reason to think this bias is negligible (even if in our empirical exercises we found that the bias of the NP estimator *after removing individuals at the cutoff* is small).

The straightforward solution is adding group/site fixed effect to the regression.

But take care… one has to pay a bit of attention to the weight attached to each site to obtain the intended causal parameter.