Quantitative Economics for the Evaluation of the European Policy

Dipartimento di Economia e Management

Co-funded by the Erasmus+ Programme of the European Union

Project funded by

European Commission Erasmus + Programme –Jean Monnet Action

Project number 553280-EPP-1-2015-1-IT-EPPJMO-MODULE

Irene Brunetti Davide Fiaschi Angela Parenti[1]

11/12/2015

[1] ireneb@ec.unipi.it, davide.fiaschi@unipi.it, and aparenti@ec.unipi.it.

Regression Discontinuity Designs (RDD)

- The idea: assignment to the treatment is determined, either completely or partly, by the value of a predictor (the covariate X_i) being on either side of a fixed threshold.
- RDD exploits the fact that some rules are quite arbitrary and therefore provide good quasi-experiments when you compare people who are just affected by the rule with people who are just not affected by the rule.
- The predictor may itself be associated with the potential outcomes, but this association is assumed to be smooth, and so any discontinuity of the conditional distribution of the outcome is interpreted as evidence of a causal effect of the treatment.

Recent Examples

- Effect of class size on student achievement (class size is determined by a cutt-off in class size)
- Effect of access to credit on development outcomes (loan offer is determined by credit score threshold)
- Effect of party democratic versus republican mayor
- Effect of wages increase for mayors on policy performance (wage jumps at population cutt-offs)
- Effect of an additional night in the hospital, a newborn delivered at 12:05 a.m. will have an extra night of reimbursable care
- Effect of school district boundaries on home values
- Effect of colonial borders on development outcomes

Regression Discontinuity Design (RDD)

There are 2 types of RDD:

- **Sharp RD**: treatment is a **deterministic** function of a covariate X.
- **Fuzzy RD**: exploits discontinuities in the probability of treatment conditional on a covariate X (the discontinuity is then used as an IV).

Sharp RDD

- In the Sharp RD design the assignment D_i is a deterministic and discontinuous function of a covariate x_i:

$$D_i = 1\{X_i \geq x_0\}$$

- All units with a covariate value of at least x_0 are assigned to the treatment group (and participation is mandatory for these individuals).
- all units with a covariate value less than x_0 are assigned to the control group (members of this group are not eligible for the treatment).

Thus:

$$D = \begin{cases} 1, & \text{if } x_i > x_0 \\ 0, & \text{if } x_i < x_0 \end{cases}$$

where x_0 is a known threshold or cutt-off.

Sharp RDD

- Once we know x_i we know D_i.
- As highlighted by Imbens and Lemieux (2008) there is no value of x_i at which you observe both treatment and control observations $\Longrightarrow$ the method relies on extrapolation across covariate values.

Identification strategy in SRD

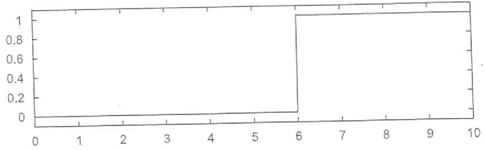

Fig. 1. Assignment probabilities (SRD).

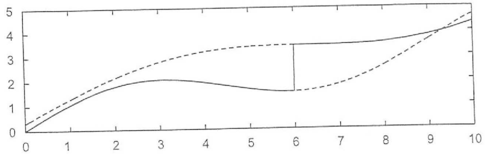

Fig. 2. Potential and observed outcome regression functions.

Key Identifying Assumptions

- Key identifying assumptions:

$$E[Y_{0i}|X_i] \text{ and } E[Y_{1i}|X_i] \text{ are continuos in } X_i \text{ at } X_0$$

- This means that all other unobserved determinants of Y are continuously related to the running variable X.
- This allows us to use average outcomes of units just below the cutt-off as a valid counterfactual for units right above the cutt-off.
- This assumption can not be directly tested. But there are some tests which give suggestive evidence whether the assumption is satisfied.
- $Y_1; Y_0 \perp D|X$
- $0 < Pr(D = 1|X = x) < 1$ (always violated in Sharp RDD)

Identification Result

- The treatment effect is identified at the threshold as:

$$\begin{aligned}\alpha_{SRDD} &= E[Y_1 - Y_0|X = x_0] \\ &= E[Y_1|X = x_0] - E[Y_0|X = x_0] \\ &= \lim_{x \to x_{0+}} E[Y_1|X = x_0] - \lim_{x \to x_{0-}} E[Y_0|X = x_0]\end{aligned} \tag{1}$$

- Without further assumptions on α_{SRDD} is only valid at the threshold.

Estimate $\alpha_{SRDD} = E[Y_1|X = x_0] - E[Y_0|X = x_0]$

- Trim the sample to a reasonable window around the threshold x_0 (discontinuity sample):
  - $x_0 - h \leq X_i \leq x_0 + h$, were h is some positive value that determines the size of the window;
  - h may be determined by cross-validation.
- Re-code running variable to deviations from threshold: $\tilde{X} = X - x_0$
  - $\tilde{X} = 0$ if $X = x_0$
  - $\tilde{X} > 0$ if $X > x_0$ and thus $D = 1$
  - $\tilde{X} < 0$ if $X < x_0$ and thus $D = 0$
- Decide a model for $E[Y|X]$:
  - linear, same slope for $E[Y_0|X]$ and $E[Y_1|X]$;
  - linear, different slopes;
  - non-linear;
  - always start with visual inspection (scatter plot with kernel/lowess) to check which model is appropriate.

SRDD Estimation: Linear with Same Slope

- $E[Y_0|X]$ is linear and treatment effect, α, does not depend on X:
$$E[Y_0|X] = \mu + \beta X; \qquad E[Y_1 - Y_0|X] = \alpha$$
- Therefore $E[Y_1|X] = \alpha + E[Y_0|X] = \alpha + \mu + \beta X$
- Since D is determined given X, we have that:

$$\begin{aligned} E[Y|X;D] &= D \cdot E[Y_1|X] + (1-D) \cdot E[Y_0|X] \\ &= \mu + \alpha D + \beta X \\ &= (\mu - \beta x_0) + \alpha D + \beta(X - x_0) \\ &= \gamma + \alpha D + \beta \tilde{X} \end{aligned} \tag{2}$$

- So we just run a regression of Y on D and $\tilde{X}$
- The key difference between this regression and regressions we have investigated in previous lectures is that D_i is not only correlated with X_i but it is a deterministic function of X_i.

SRDD Estimation: Differential Slopes

- $E[Y_0|X]$ and $E[Y_1|X]$ are distinct linear functions of X, so the average effect of the treatment $E[Y_1 - Y_0|X]$ varies with X:
$$E[Y_0|X] = \mu_0 + \beta_0 X; \qquad E[Y_1|X] = \mu_1 + \beta_1 X$$
- So $E[Y_1 - Y_0|X] = \alpha(X) = (\mu_1 - \mu_0) + (\beta_1 - \beta_0)X$ we have:

$$\begin{aligned} E[Y|X;D] &= D \cdot E[Y_1|X] + (1-D) \cdot E[Y_0|X] \\ &= \gamma + \beta_0(X - x_0) + \alpha D + \beta_1((X - x_0) \cdot D) \end{aligned} \tag{3}$$

- Regress Y on $(X - x_0)$, D and the interaction $((X - x_0) \cdot D)$, the coeffcient of D reflects the average effect of the treatment at $X = x_0$

SRDD Estimation: Non-linear Case

- Suppose the non-linear relationship is $E[Y_{0i}|x_i] = f(X_i)$ for some reasonably smooth function $f(X_i) \Longrightarrow E[Y_0|X]$ and $E[Y_1|X]$ are distinct non-linear functions of X and the average effect of the treatment $E[Y_1 - Y_0|X]$ varies with X.
- Include quadratic and cubic terms in $(X - x_0)$ and their interactions with D in the equation:

$$\begin{aligned} E[Y|X;D] = \gamma_0 + \gamma_1(X - x_0) + \gamma_p(X - x_0)^2 + \alpha_0 D \\ + \alpha_1((X - x_0) \cdot D) + \alpha_p((X - x_0)^2 \cdot D) \end{aligned} \tag{4}$$

or

$$\begin{aligned} E[Y|X;D] = \gamma_0 + \gamma_1(X - x_0) + \gamma_2(X - x_0)^2 + \gamma_3(X - x_0)^3 + \alpha_0 D \\ + \alpha_1((X - x_0) \cdot D) + \alpha_2((X - x_0)^2 \cdot D) + \alpha_3((X - x_0)^3 \cdot D) \end{aligned} \tag{5}$$

SRDD Estimation: Non-linear Case

- We can also use a nonparametric kernel method.

Sharp RDD: Falsification Checks

- *Sensitivity:* Are results sensitive to alternative specifications?
- *Balance Checks*: Do covariates Z jump at the threshold?
- Check if jumps occur at *placebo thresholds* c^*?
- *Sorting*: Do units sort around the threshold?

Sensitivity to Specification

- $Y = f(X) + \alpha D + \epsilon$: A miss-specified control function $f(X)$ can lead to a spurious jump: Take care not to confuse a nonlinear relation with a discontinuity.
- More flexibility (e.g.adding polynomials) reduces bias, decreases efficiency.
- Check sensitivity to size of bandwidth (i.e. estimation window).

Balance Checks Test

- Test for comparability of agents around the cutt-off:
  - Visual tests: Plot$E[Z|X;D]$ and look for jumps, ideally the relation between covariates and treatment should be smooth around threshold,
  - Run the RDD regression using Z as the outcome:
    $$E[Z|X;D] = \beta_0 + \beta_1(X - x_0) + \alpha_z D + \beta_3((X - x_0) \cdot D)$$
    ideally should yield $\alpha_z = 0$ if Z is balanced at the threshold.
- Finding a discontinuity in Z does not necessarily invalidate the RDD:
  - Can incorporate Z as additional controls into our main RDD regression. Ideally, this should only impact statistical significance, not magnitude of treatment effect.
  - Alternatively, regress the outcome variable on a vector of controls and use the residuals in the RDD, instead of the outcome itself.
- Balance checks address only observables, not unobservables

Placebo Threshold

- Let c^* be a placebo thershold value. Run the regression of:
$$E[Y|X; D] = \beta_0 + \beta_1(X - c^*) + \alpha D + \beta_3((X - c^*) \cdot D)$$
and check if α is large and significant?
- Usually we split the sample to the left and the right of the actual threshold x_0 in order to avoid miss-specification by imposing a zero jump at x_0.
- The existence of large placebo jumps does not invalidate the RDD, but does require an explanation.
- Concern is that the relation is fundamentally discontinuous and jump at cutt-off is contaminated by other factors.
- Maybe data exists in a period where there was no program.

Sorting Around the Threshold

- Can subjects behavior invalidate the local continuity assumption?
  - Can administrators strategically choose what assignment variable to use or which cutt-off point to pick?
  - Either can invalidate the comparability of subjects near the threshold because of sorting of agents around the cutt-off, where those below may differ on average form those just above.

- What else changes at x_0? Continuity violated in the presence of other programs that use a discontinuous assignment rule with the exact same assignment variable and cut-off.

Sorting Around the Threshold

- Example: Beneficial job training program offered to agents with income $< x_0$. Concern, people will withhold labor to lower their income below the cut-off to gain access to the program.
- Test for discontinuity in density of forcing variable:
  - Visual Histogram Inspection:
    - Construct equal-sized non-overlapping bins of the forcing variable such that no bin includes points to both the left and right of the cut-off.
    - For each bin, compute the number of observations and plot the bins to see if there is a discontinuity at the cut-off
  - Formal tests (e.g. McCrary, 2008)

Fuzzy Regression Discontinuity Design

- Fuzzy RD exploits discontinuities in the probability of treatment conditional on a covariate.
- Threshold may not perfectly determine treatment exposure, but it creates a discontinuity in the probability of treatment exposure
- Incentives to participate in a program may change discontinuously at a threshold, but the incentives are not powerful enough to move all units from nonparticipation to participation
- We can use such discontinuities to produce instrumental variable estimators of the effect of the treatment (close to the discontinuity)

Fuzzy Regression Discontinuity Design (FRD)

- Probability of being offered a scholarship may jump at a certain score threshold (when applicants are given "special consideration")
- We should not compare recipients with non-recipients (even close to threshold) since they are likely differ along unobservables related to outcome.
- But for applicants with scores close to the threshold we can exploit the discontinuity as an instrument to estimate the LATE for the subgroup of applicants for whom scholarship receipt depends on the difference between their score and the threshold.
  - A complier in the framework is a student who switches from non-recipient to recipient if her scores crosses the threshold

Fuzzy RDD: Discontinuity in $E[D|X]$

Figure 2

Fuzzy RDD: Discontinuity in $E[Y|X]$

Figure 3

Fuzzy RDD: Identification

Identification Assumptions

- Binary instrument Z with $Z = 1\{X > x_0\}$
- Restrict sample to observations close to discontinuity where $E[Y|D;X]$ jumps so that $X \approx x_0$ and thus $E[X|Z=1] - E[X|Z=0] \approx 0$.
- Usual IV assumptions hold (ignorability, first stage, montonicity)

Identification Result

$$\begin{aligned}
\alpha_{FRDD} &= E[Y_1 - Y_0|X = x_0] \\
&= \frac{outcome \;\; discontinuity}{treatment \;\; discontinuity} \\
&\approx \frac{E[Y|Z=1] - E[Y|Z=0]}{E[D|Z=1] - E[D|Z=0]}
\end{aligned} \tag{6}$$

Fuzzy RDD: Identification

- Suppose $E[Y_0|X]$ is linear and treatment effect is constant:

$$E[Y_0|X] = \mu + \beta X; \quad E[Y_1 - Y_0|X] = \alpha; \quad E[Y_1|X] = \alpha + \mu + \beta X.$$

- Suppose also that $E[D|X]$ has a discontinuity at c. For those who are close to c, the average outcomes are:

$$\begin{aligned} E[Y|Z=0] &= \mu + \alpha E[D|Z=0] + \beta E[X|Z=0] \\ E[Y|Z=1] &= \mu + \alpha E[D|Z=1] + \beta E[X|Z=1] \end{aligned}$$

- For those with $X \approx c$ $E[X|Z=1] - E[X|Z=0] \approx 0$. However, $E[D|Z=1] - E[D|Z=0] \neq 0$ because of discontinuity in the assignment probability.

Fuzzy RDD: Estimation

- Cut the sample to a small window above and below the threshold (discontinuity sample).
- Code instrument $Z = 1\{X > x_0\}$
- Fit 2SLS: $Y = \beta_0 + \beta_1(X - c) + \beta_2(Z \cdot (X - c)) + \alpha D$, where D is instrumented with Z.
- Specification can be more flexible by adding polynomials.
- Using a larger window we may also fit 2SLS:

  $$Y = \beta_0 + \beta_1(X - c) + \alpha D + \beta_2(D \cdot (X - c))$$

  where D and $D \cdot (X - c)$ are instrumented with Z and $Z \cdot (X - c)$.
- Also helpful to separately plot (and estimate) the outcome discontinuity and treatment discontinuity.

Graphical Analysis in RD Designs

A graphical analysis should be an integral part of any RD study. You should show the following graphs:

1. **Outcome by forcing variable (X_i)**
   - The standard graph showing the discontinuity in the outcome variable.
   - Construct bins and average the outcome within bins on both sides of the cut-off.
   - You should look at different bin sizes when constructing these graphs (see Lee and Lemieux (2010) for details).
   - Plot the forcing variable X_i on the horizontal axis and the average of Y_i for each bin on the vertical axis.
   - You may also want to plot a relatively exible regression line on top of the bin means.
   - Inspect whether there is a discontinuity at x_0.
   - Inspect whether there are other unexpected discontinuities.

Outcome by forcing variable

Figure 4

Graphical Analysis in RD Designs

2. **Probability of treatment by forcing variable if fuzzy RD**
   - In a fuzzy RD design you also want to see that the treatment variable jumps at x_0.
   - This tells you whether you have a first stage.
3. **Covariates by forcing variable.**
   - Construct a similar graph to the one before but using a covariate as the outcome.
   - There should be no jump in other covariates.
   - If the covariates would jump at the discontinuity one would doubt the identifying assumption.

Covariates by forcing variable

Figure 5

4 **The density of the forcing variable.**

- One should plot the number of observations in each bin.
- This plot allows to investigate whether there is a discontinuity in the distribution of the forcing variable at the threshold.
- This would suggest that people can manipulate the forcing variable around the threshold.
- This is an indirect test of the identifying assumption that each individual has imprecise control over the assignment variable.

Internal and External Validity

- At best, Sharp and Fuzzy RDD estimate the average effect of the sub-population with X_i close to c;
- Fuzzy RDD restricts this subpopulation even further to that of the compliers with X_i close to c;
- Only with strong assumptions (e.g., homogenous treatment effects) can we estimate the overall average treatment effect.
- So, RDD have strong internal validity but may have weak external validity (although it depends...)

An Application of Fuzzy RD on Class Sizes

- Angrist and Lavy (1999) use a fuzzy RD design to analyze the effect of class size on test scores.
- They extend RD in two ways:
  - The causal variable of interest (class size) takes on many values. $\Longrightarrow$ the first stage exploits discontinuities in average class size instead of probabilities of a single treatment.
  - They use multiple discontinuities.
- Angrist and Lavy exploit an old Talmudic rule that classes should be split if they have more than 40 students in Israel.
  - school with 40 students has only one class. $\Longrightarrow$ class size 40.
  - A school with 41 students has two classes. $\Longrightarrow$ class sizes 21 and 20.
- The use the Maimonides' rule of 40 to construct instrumental variables.

An Application of Fuzzy RD on Class Sizes

- The rule is not followed completely strictly $\Longrightarrow$ they have a fuzzy discontinuity design.

Econometric Specification

- They want to estimate the relationship between average achievement and class size:

$$Y_{isc} = \alpha_0 + \rho n_{sc} + \eta_{isc}$$

- Estimating this relationship with OLS may lead to biased results because class size is likely to be correlated with the error term. The 2 main reasons for this are:
  - Parents from higher socioeconomic backgrounds may put their children in schools with smaller classes.
  - Because principals may put weaker students in smaller classes.

An Application of Fuzzy RD on Class Sizes

- Angrist & Lavy therefore use the Maimonides rule in a fuzzy RD design.

$$Y_{isc} = \alpha_0 + \alpha_1 d_s + \rho n_{sc} + \beta_1 e_s + \beta_2 e_s^2 + \eta_{isc} \tag{7}$$

  where Y_{isc} is the test score of student i in school s and class c.
  e_s is enrollment in school s.
  d_s is the percentage of disadvantage students in class
  n_{sc} is class size.

- The variables relate to the previous description as follows:
  - n_{sc} plays the role of D_i.
  - e_s plays the role of X_i.
  - m_{sc} plays the role of Z_i
- The first stage regression is:

$$n_{sc} = \gamma_0 + \gamma_1 d_s + \pi m_{sc} + \delta_1 e_s + \delta_2 e_s^2 + \zeta_{isc}$$

  where m_{sc} is the function describing Maimonides rule.

An Application of Fuzzy RD on Class Sizes: regression results

- There is a positive OLS relationship between class size and test scores. If you control for percentage disadvantaged and total enrollment, however, the relationship turns slightly negative but not significantly;
- Students in schools with more overall enrollment (often in bigger cities) do better on average.
- Average test scores are partly a mirror image of predicted class sizes.

An Application of Fuzzy RD on Class Sizes: regression results

- Because larger schools are often in better-off areas they control for enrolment when they redraw the relationship between class-size and achievement.
- Now test-scores are more of a mirror image to predicted class sizes.
- The effect of class size now is signicantly negative.