The Economics of European Regions: Theory, Empirics, and 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

Davide Fiaschi Angela Parenti[1]

November 28, 2019

[1]davide.fiaschi@unipi.it, and angela.parenti@unipi.it.

Regression Discontinuity Design -Thistlethwaite and Campbell (1960)

- RDD was introduced by Thistlethwaite and Campbell (1960) as a way of estimating treatment effects in non-experimental setting where the treatment is determined by whether an observed "assignment" variable ("forcing" variable) exceeds a *known* cut-off.
- They use RDD to analyse the impact of merit awards on future academic outcomes.
- They use the fact tat the **allocation** of awards was based on an **observed** test score.
- Main idea: individuals with scores **just below** the cut-off (who did no receive the award) were good comparisons to those **just above** the cut-off (who did receive the award).

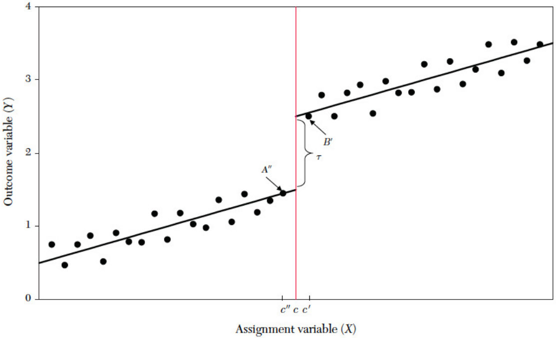
RDD -Thistlethwaite and Campbell (1960) (cont.)

- This assignment generates a sharp *discontinuity* in the treatment (receiving the award) as a *function* of the test score.
- At the same time, there are no reasons, other than the merit award, for future academic outcomes to be a discontinuous function of the test score.

  $\Rightarrow$ the discontinuity jump in the outcome at the cut-off is the *causal effect* of the merit award.

Example Linear RD setup

Figure 1. Lee and Lemieux (2010).



• B' reasonable guess for Y of an individual scoring c (receiving the treatment).
• A'' reasonable guess for Y for the same individual in the counterfactual (not receiving the treatment).
$\Rightarrow B' - A''$ causal estimate.

Example Linear RD setup (cont.)

- In order of the RDD approach to work "all other factors" determining Y must be evolving "smoothly" with respect to X.
- In order to produce a reasonable guess for the treated and untreated states $X = c$ with finite data, one has to use data *away* from the discontinuity
  $\Rightarrow$ the estimate will be dependent on the chosen *functional form*.

Sharp RDD

- In **Sharp RDD** designs the treatment status is a **deterministic** and **discontinuous** function of a covariate X_i.

$$\begin{cases} D_i = 1 & \text{if } X_i \geq c \\ D_i = 0 & \text{if } X_i < c \end{cases}$$

  where c is a **known** threshold or cut-off.

- Once we know X_i we know D_i .
- Imbens and Lemieux (2008): there is no value of X_i at which you observe both treatment and control observations.

RDD in potential outcome framework

Two potential outcomes $Y_i(1), Y_i(0) \Rightarrow$ causal effect: $Y_i(1) - Y_i(0)$
Fundamental problem of causal inference $\Rightarrow$ ATE: $E[Y_i(1) - Y_i(0)]$.

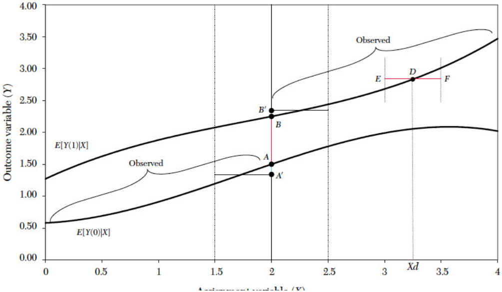


Figure 2. Lee and Lemieux (2010)

- In RDD two underlying relationship between average outcome and X: $E[Y_i(1)|X]$ and $E[Y_i(0)|X]$.
- All individuals to the right of the cut-off are exposed to treatment and all those to the left are denied to treatment.
- We only observe $E[Y_i(1)|X]$ to the right of the cut-off and $E[Y_i(0)|X]$ to the left

ATE at c under **continuity** in the underlying functions

$$\begin{aligned} B - A &= \lim_{\epsilon \downarrow 0} E[Y_i|X_i = c + \epsilon] - \lim_{\epsilon \uparrow 0} E[Y_i|X_i = c + \epsilon] \\ &= E[Y_i(1) - Y_i(0)|X = c] = \tau \Rightarrow \text{average treatment effect at } c\ ! \end{aligned}$$

RDD in potential outcome framework (cont.)

- Suppose that in addition potential outcomes can be described by a linear, constant effects model:

$$E\left[Y_i(0)|X_i\right] = \alpha + \beta X_i$$
$$Y_i(1) = Y_i(0) + \tau$$

- This leads to the regression:

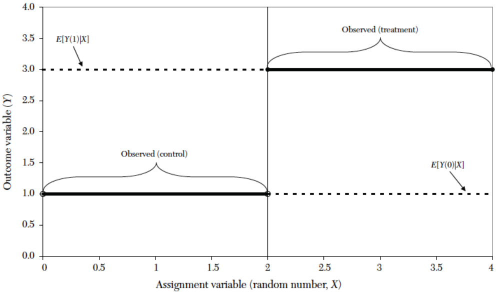
$$Y_i = \alpha + \beta X_i + \tau D_i + \epsilon_i$$

- The key difference of this regression is that D_i is not only correlated with X_i but it is a **deterministic function** of X_i.

RDD as a local randomized experiment

- The randomized experiment can be thought as an RDD where the assignment variable is $X = \nu$, where ν is a randomly generated number, and the cut-off is c.

Figure 3. Lee and Lemieux (2010) where ν follows a uniform distribution.



- The assignment now is random and therefore independent of potential outcomes.
- Moreover, the curves $E[Y_i(1)|X]$ and $E[Y_i(0)|X]$ are flat ($\Rightarrow$ continuous at c).
- The average causal effect is the difference in the mean value of Y just above and just below c.

RDD as a local randomized experiment (Cont.)

Example. *Treatment*: Job search assistance for the unemployed; *Outcome*: founding a job within one month of the treatment

- Assume that for ethical reasons people receiving "bad draw" are compensated; i.e., the lower the random number X the higher the compensation.
- Assume now that people with higher monetary compensation can afford to take more time looking for a job.
- Then, the probability to find a job within one month is higher for those who receive lower compensation (see Figure 2).
- Becomes a "smoothly contaminated" randomize experiment and the simple comparison of means no longer yields a consistent estimate of the treatment effect.
- By focusing around the threshold the RDD yields consistent estimates since people just above and just below the threshold receive essentially the same monetary compensation $\Rightarrow$ **locally** randomize experiment

Key Identifying Assumption

- Key identifying assumption:
  $E\left[Y_i(1)|X\right]$ and $E\left[Y_i(0)|X\right]$ are continuous in X_i at c.
- This means that all other unobserved determinants of Y are continuously related to the forcing X.
- This allows us to average outcomes of units just below the cut-off as a **valid counterfactual** for units right above the cut-off variable.
- This assumption cannot be directly tested. But there are some tests which give suggestive evidence whether the assumption is satisfied.

Identification and interpretation - Lee and Lemieux (2010)

1. How do I know whether an RDD is appropriate for my context? When are the identification assumptions plausible or implausible?

   "When there is a continuously distributed stochastic error component to the assignment variable - which can occur when optimizing agents **do not have precise control over the assignment** variable - then the variation in the treatment will be as good as randomized in a neighbourhood around the discontinuity threshold."

   • If individuals have a great control over the assignment variable we can expect that individuals on one side of the threshold to be *systematically* different from those on the other side.
   • But individual will not always be able to have *precise* control.
   • Precise sorting around the threshold is self-selection!

Identification and interpretation - Lee and Lemieux (2010)

2. Is there any way I can test those assumptions?

   "Yes. As in a randomized experiment, the distribution of observed baseline covariates should not change discontinuously at the threshold."

   • Although is impossible to test this directly, a discontinuity would indicate a **failure** of the identifying assumption.
   • As when we want to asses whether the randomized experiment was carried out properly
   $\Rightarrow$ the treatment and control groups must be similar in their characteristics.
   • If a lagged dependent variable is added as regressor which is pre-determined the local randomization result will imply that the lagged dependent variable will have a continuous relationship with X.

Identification and interpretation - Lee and Lemieux (2010)

3. To what extent are results from RDD generalizable?

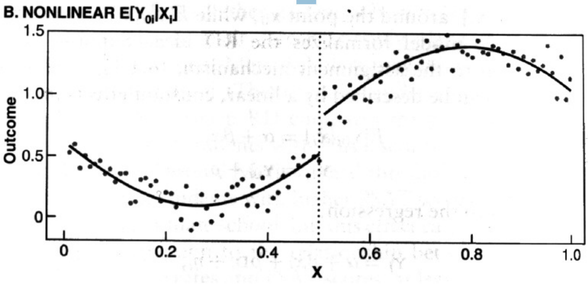
   "The RD estimand can be interpreted as a *weighted average treatment effect*, where the weights are the relative ex ante probability that the value of an individual's assignment variable will be in the neighbourhood of the threshold."

   • If the weights are relatively similar across individuals RDD estimate is closer to the overall average treatment effect.

Sharp Regression Discontinuity - Nonlinear Case

As seen in the example of job search assistance and the probability to find a job, sometimes the trend relation $E\left[Y_i(0)|X\right]$ is nonlinear.

Figure 6.1.1. Angrist and Pischke (2010).



Sharp Regression Discontinuity - Nonlinear Case (cont.)

- Suppose the nonlinear relationship is $E\left[Y_i(0)|X\right] = f(X_i)$ for some reasonably smooth function $f(X_i)$.
- In that case we can construct RDD estimates by fitting:

$$Y_i = f(X_i) + \tau D_i + \eta_i \tag{1}$$

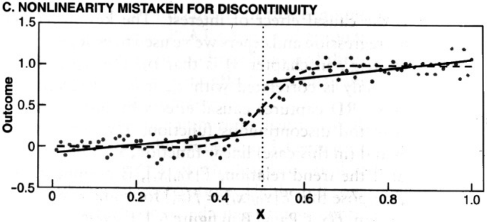
- There are 2 ways of approximating $f(X_i)$:
  - Use a nonparametric kernel method
  - Use a p-th order polynomial, i.e. estimate:

$$Y_i = \alpha + \beta_1 X_i + \beta_2 x_i^2 + \beta_p x_i^p + \tau D_i + \eta_i \tag{2}$$

Internal Validity of RDD Estimates

- The validity of RDD estimates depends crucially on the assumption that the polynomials provide an adequate representation of $E\left[Y_i(0)|X\right]$.
- If not what looks like a jump may simply be a non-linearity in $f(X_i)$ that the polynomials have not accounted for.

Figure 6.1.1. Angrist and Pischke (2010).



Fuzzy RDD

- The treatment is determined **partly** by whether the assignment variable crosses a cut-off point (imperfect compliance).
- In the case of Sharp RDD the probability of treatment *jumps from 0 to 1* when X crosses the cut-off c.
- **Fuzzy RDD** instead allows for a **smaller jump** in the probability of assignment to the treatment at c and only requires:

$$\lim_{\epsilon \downarrow 0} (D = 1 | X = c + \epsilon) \neq \lim_{\epsilon \uparrow 0} (D = 1 | X = c + \epsilon)$$

- Since the probability is less than one, the jump in the relationship between Y and X is no longer the ATE.
- The ATE can be found by dividing the jump in the relationship between Y and X by the jump in the relation between D and X:

$$\tau_F = \frac{\lim_{\epsilon \downarrow 0} E\left[Y | X = c + \epsilon\right] - \lim_{\epsilon \uparrow 0} E\left[Y | X = c + \epsilon\right]}{\lim_{\epsilon \downarrow 0} E\left[D | X = c + \epsilon\right] - \lim_{\epsilon \uparrow 0} E\left[D | X = c + \epsilon\right]}$$

Fuzzy RDD as IV

- The discontinuity in the relationship between D and X therefore becomes an **instrumental variable** for treatment status.
- D_i is no longer deterministically related to crossing a threshold but there is a jump in the *probability* of treatment at c.

$$P\left[D_i = 1 | X_i\right] = \begin{cases} g_1(X_i) & \text{if } X_i \geq c \\ g_0(X_i) & \text{if } X_i < c \end{cases}$$

where $g_1(X_i) \neq g_0(X_i)$.

- $g_1(X_i)$ and $g_0(X_i)$ can be anything as long as they differ at c.

Fuzzy RDD as IV (Cont.)

Parallel to the "Wald" formulation of the treatment effect in an IV setting the two assumptions of Imbens and Angrist (1994) must be satisfied, i.e:

- **monotonicity**: X crossing the cut-off cannot *simultaneously cause* some units to take up and others to reject the treatment
- **excludability**: X crossing the cut-off cannot impact Y except through impacting receipt of treatment.

In this case:

$$\tau_F = E\left[Y(1) - Y(0) | \text{unit is complier}, X = c\right]$$

where "compliers" are those who are treated if they satisfy the cut-off rule $X_i \geq c$ but would not otherwise.

IV vs Fuzzy RDD

- The difference between *sharp* RDD and *fuzzy* RDD is parallel to the difference between randomize experiment with perfect compliance and the case of imperfect compliance (when only the ITT can be estimated).
- In the IV setting is important to verify a strong first-stage relationship, in the fuzzy RDD is important that a discontinuity exists in the relationship between D and X.
- The IV estimate is the LATE (average treatment effect for the subpopulation affected by the instruments), in the fuzzy RDD can be interpreted as a *weighted* LATE, where the weights reflect the ex-ante likelihood the individual's X is close to the threshold.

Fuzzy RDD as TSLS

- The relationship between the probability of treatment and X_i can be written as:

$$P\left[D_i = 1 | X_i\right] = g_0(X_i) + \left[g_1(X_i) - g_0(X_i)\right] T_i$$

where $T_i = 1(X_i \geq c)$.

- T_i is used as an instrument for D_i.
- The estimated first stage would be:

$$D_i = \gamma_0 + \gamma_1 X_i + \gamma_2 X_i^2 + ... + \gamma_p X_i^p + \pi T_i + \nu_{1i}$$

- The fuzzy RDD reduced form is:

$$Y_i = \mu + \phi_1 X_i + \phi_2 x_i^2 + ... + \phi_p X_i^p + \tau\pi T_i + \nu_{2i}$$

Practical Tips for Estimation

I. Graphical Analysis in RD Designs

II. Estimating the f-Function

III. Testing the Validity of the RD Design

I. Graphical Analysis in RD Designs

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.
   - Plot the forcing variable X_i on the horizontal axis and the average of Y_i for each bin on the vertical axis.
   - Optionally also plot a relatively flexible regression line on top of the bin means.
   - Inspect whether there is a discontinuity at c.
   - Inspect whether there are other unexpected discontinuities.

   - As robustness for the choice of the bandwidth look at different bin sizes when constructing these graphs (Lee and Lemieux (2010) for details).

I. Graphical Analysis in RD Designs: Outcome by forcing variable

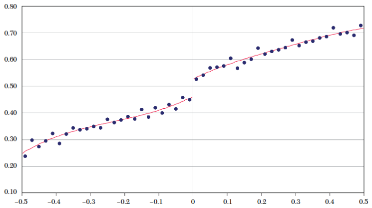


Figure 6. Lee and Lemieux (2010):
Bandwidth of 0.02 (50 bins)

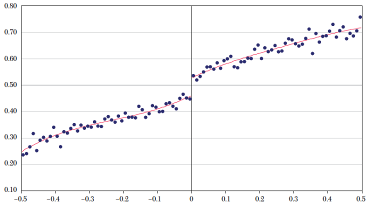


Figure 7. Lee and Lemieux (2010):
Bandwidth of 0.01 (100 bins)

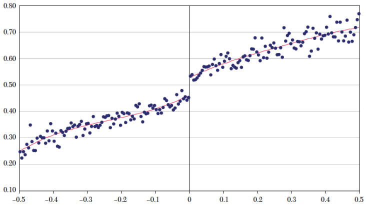


Figure 8. Lee and Lemieux (2010):
Bandwidth of 0.005 (200 bins)

I. Graphical Analysis in RD Designs

2. **Probability of treatment by forcing variable if fuzzy RD**.
   - In a fuzzy RD design we also check if the treatment variable jumps at c.
   - If so, there is a first stage!
3. **Covariates by forcing variable**.
   - Construct similar graphs to the one of the outcome but using a covariate as the "outcome".
   - There should be **no jump** in other covariates (e.g., lagged outcome variable).
   - If the covariates would jump at the discontinuity one would doubt the identifying assumption.

I. Graphical Analysis in RD Designs: Covariates by forcing variable

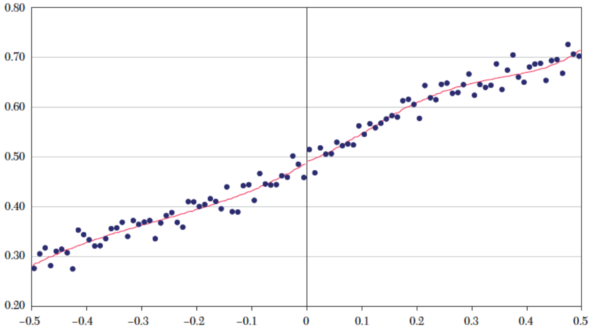


Figure 17. Lee and Lemieux (2010):
Discontinuity in Baseline Covariate (on lagged outcome variable)

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

I. Graphical Analysis in RD Designs: The density of the forcing variable

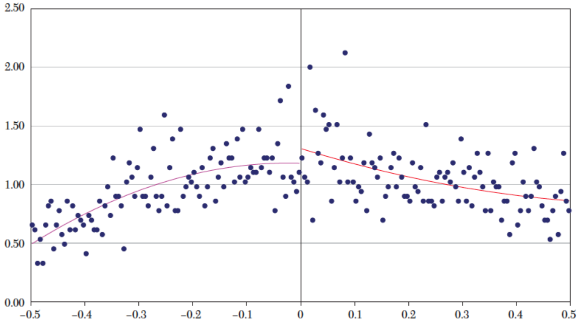


Figure 16. Lee and Lemieux (2010):
Density of the Forcing Variable

II. Estimating the f-Function

- As pointed out before there are essentially two ways of approximating the $f(X_i)$:
  - Kernel regression.
  - Polynomial function.
- There is no right or wrong method. Both have advantages and disadvantages.

II. Estimating the f-Function: the kernel method

- The nonparametric kernel method has its problems in this case because you are trying to estimate regressions at the cut-off point. $\Rightarrow$ "boundary problem":

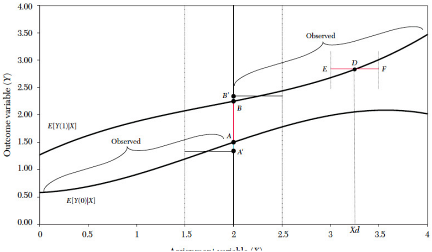


Figure 2. Lee and Lemieux (2010)

- While the "true" effect is AB, with a certain bandwidth a rectangular kernel would estimate the effect as $A'B'$.
- There is therefore systematic bias with the kernel method if the $f(X_i)$ is upwards or downwards sloping.

II. Estimating the f-Function: the kernel method

- The standard solution to this problem is to run local linear regression to reduce the bias.
- The simpler case is the rectangular kernel, which amounts to estimating a standard regression over a window of width h on both sides of the cut-off.
- Other kernel might be chosen but this has little impact in practice.
- While estimating this in a given window of width h around the cut-off is straightforward it is more difficult to choose the bandwidth h.
- See Lee and Lemieux (2010) for two methods to choose the bandwidth (usual trade-off between bias and efficiency).

II. Estimating the f-Function: the polynomial method

- The polynomial method suffers from the problem that uses data far away from the cut-off to estimate the $f(X_i)$ function.
- The equivalent of choosing the right bandwidth for the polynomial method is to use the right order of polynomial.
- See Lee and Lemieux (2010) for a test on the right polynomial.
- Practically:
  - report results for both estimation types;
  - show that including higher order polynomials does not substantially affect the findings;
  - show that the results are not affected by variation in the window around the cut-off.

III. Testing the Validity of the RD Design

1. **Testing the continuity of the density of X**
   - A discontinuity in the density suggests that there is some *manipulation* of X around the threshold.
2. **Explore the sensitivity of the results to the inclusion of baseline covariates**
   - The inclusion of baseline covariates (no matter how they are correlated with outcome) should not affect the estimated discontinuity, if no-manipulation assumption holds.
   - Lee and Lemieux (2010) suggest to simply including the covariates directly, after choosing a suitable order of polynomial
     $\Rightarrow$ significant changes in the estimated effect or increases in the standard errors may be an indication of a misspecified functional form.

References

- Angrist, J. D., and Pischke, J. S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.
- Becker, S. O., Egger, P. H., and Von Ehrlich, M. (2010). Going NUTS: The effect of EU Structural Funds on regional performance. *Journal of Public Economics*, 94(9), 578-590.
- Lee, D. S., and Lemieux, T. (2010). Regression Discontinuity Designs in Economics, *Journal of Economic Literature*, 48(2), 281-355.
- Pellegrini, G., Terribile, F., Tarola, O., Muccigrosso, T., and Busillo, F. (2013). Measuring the effects of European Regional Policy on economic growth: A regression discontinuity approach. *Papers in Regional Science*, 92(1), 217-233.
- Thistlethwaite, D. L., and Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, 51(6), 309.