The basic RDD principle
RDD estimation in a regression framework
The validity of RDD

Econometrics of Evaluation
Regression Discontinuity Design (RDD)

Jean-Noël Senne

Université Paris-Saclay

Problématiques Contemporaines du Développement
M2EIED
Université de Pau et des Pays de l'Adour
2023 - 2024

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD

The basic idea of RDD

- A sometimes useful way to 'mimic' a counterfactual we don't observe is to consider the very structure of the program being evaluated

- **Regression Discontinuity Designs (RDD)** are based on a **precise rule** that determines treatment

- RDD exploit settings in which selection into treatment becomes discretely more likely when a **continuous variable (called the 'forcing' or 'running' variable)** crosses a given threshold

→ Many treatments have **"threshold" assignment rules**
   E.g. scholarship or admission based on academic test scores, means-tested anti-poverty programs, ages cutoffs for pensions or benefits, school map, vote margins...

- Relatively old method : Thistlewaite and Campbell (1960) analyze the impact of scholarships based on school results on future careers

- Only used in economics since the late 1990s, following some influential studies :
  - Angrist and Lavy (1999) study the impact of class size on school results
  - Black (1999) studies the impact of school mapping on housing prices
  - Hahn, Todd, and Van der Klaauw (2001) rigorously specify the required identification assumptions
  - Imbens & Lemieux (2008) provide a detailed practical guide

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD

The intuition behind RDD

- Very simple idea : there exists a **continuous** selection variable Z which has a **discontinuous impact** on treatment probability (due to a set rule)

$\rightarrow$ Around a threshold value $\underline{Z}$, there is a **discontinuity in treatment assignment** (but <u>not</u> in the outcome)

- **Identification assumption :** Treatment discontinuity

  (1) $\quad T^+ = \lim\limits_{Z \to \underline{Z}^+} E(T|Z) \quad \neq \quad T^- = \lim\limits_{Z \to \underline{Z}^-} E(T|Z)$

  (2) $\quad E(Y_0)$ and $E(Y_1)$ are continuous at $\underline{Z}$

$\rightarrow$ Individuals close to the discontinuity are very similar, but have a different treatment exposure (treated and controls respectively)

$\rightarrow$ Around the threshold value $\underline{Z}$, **treatment is as-good-as-random** (no systematic *observable and unobservable* differences between treated and controls)

- **Estimation of average treatment effect :** compare the outcome of individuals just below and just above the discontinuity !

The basic RDD principle
RDD estimation in a regression framework
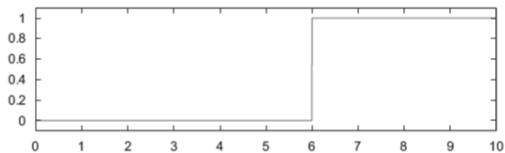The validity of RDD

RDD on a graph

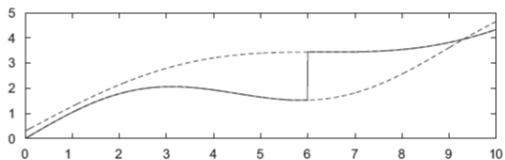Fig. 1. Assignment probabilities (SRD).

Fig. 2. Potential and observed outcome regression functions.

Source : Imbens & Lemieux (2008)

⇒ **How do we use RDD in practice ?**

⇒ **What can we learn from RDD estimations ?**

⇒ **To what extent are RDD estimations valid and robust ?**

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD

Plan

The basic RDD principle

RDD estimation in a regression framework

The validity of RDD

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD

The two regression discontinuity designs

- Let's note Y_i the observed outcome and (Y_{1i}, Y_{0i}) the potential outcomes
- We want to estimate the causal effect of a treatment T_i on Y_i :

$$Y_i = \alpha + \beta_i T_i + \epsilon_i$$

where : $\beta_i = Y_{1i} - Y_{0i}$ is the individual treatment effect
$\alpha = E(Y_{0i})$ is the mean outcome without treatment
$\epsilon_i = Y_{0i} - E(Y_{0i})$ is an error term (with zero mean)

- Let's note Z a **continuous** selection variable with a **discontinuous impact** on the probability to be treated $P(T = 1)$

- There are **2 types** of discontinuity designs :
    - **Sharp design** ('clear')
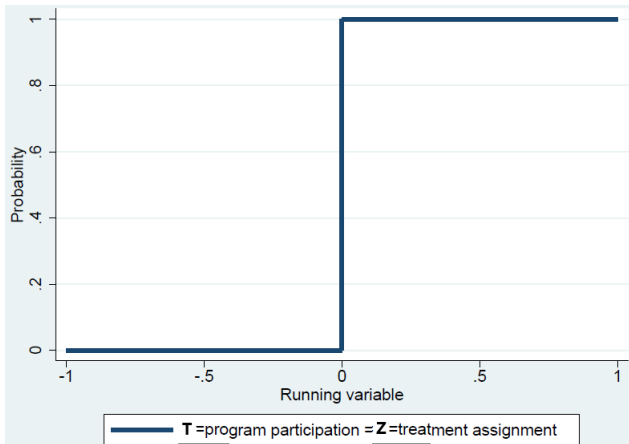   - **Fuzzy design** ('blurred')

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD

Sharp design

- Treatment T is a **deterministic function** of the selection variable Z :

$$\begin{cases} T_i = T(Z_i) = 1_{(Z_i \geq \underline{Z})} & \text{or} \quad P(T_i = 1 | Z_i \geq \underline{Z}) = 1 \\ T_i = T(Z_i) = 0_{(Z_i < \underline{Z})} & \text{or} \quad P(T_i = 1 | Z_i < \underline{Z}) = 0 \end{cases}$$

$\Rightarrow$ Individual i 'automatically' switches from controls $P(T_i = 0)$ to treated $P(T_i = 1)$ when Z_i crosses the threshold $\underline{Z}$

$\Rightarrow$ **(Perfect) discontinuity** in T around $\underline{Z}$

The basic RDD principle
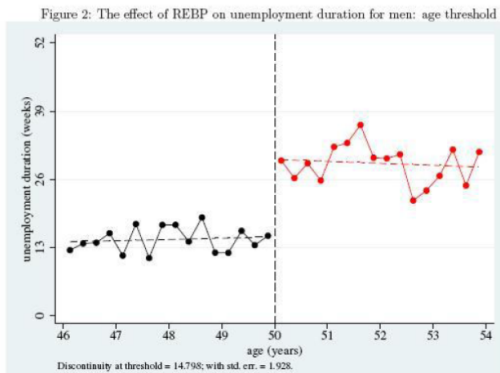RDD estimation in a regression framework
The validity of RDD

Sharp design on a graph

- All individuals with $Z_i \geq \underline{Z}$ are treated
- All individuals with $Z_i < \underline{Z}$ are untreated

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD

Sharp design example

- Lalive (2008) investigates the effect of unemployment benefits on unemployment duration
- In Austria, unemployment benefits are higher when job loss occurs after 50
- → **Sharp discontinuity in the amount of unemployment benefits** around 50
- → **Discontinuity in unemployment duration** that can be attributed to higher benefits
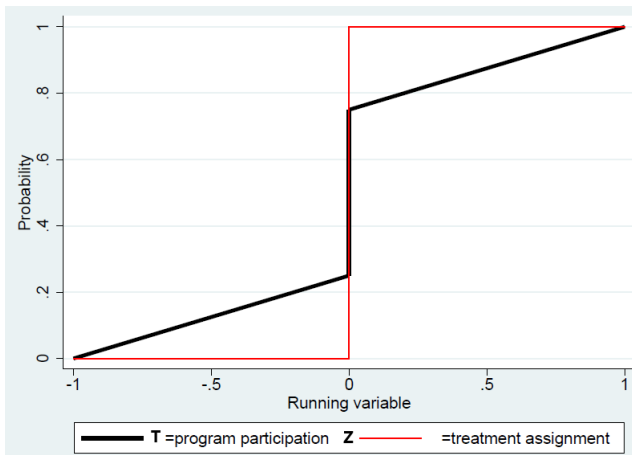

Figure 2: The effect of REBP on unemployment duration for men: age threshold

Discontinuity at threshold = 14.798; with std. err. = 1.928.

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD

Fuzzy design

- The selection variable Z affects the **probability of being treated**, but with imperfect assignment **($\simeq$ imperfect compliance)**

$$0 < P(T_i = 1 | Z_i < \underline{Z}) \quad << \quad P(T_i = 1 | Z_i \geq \underline{Z}) < 1$$

$\Rightarrow$ Some individuals with $Z_i < \underline{Z}$ are treated, while some individuals with $Z_i \geq \underline{Z}$ are untreated

$\Rightarrow$ **(Neat) discontinuity** in the probability of being treated $P(T_i = 1 | Z)$ around $\underline{Z}$

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD

Fuzzy design on a graph

- Some individuals with $Z_i \geq \underline{Z}$ are untreated...

- Some individuals with $Z_i < \underline{Z}$ are treated...

- ... but there is a discrete 'jump' in treatment probability around $\underline{Z}$

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD

Fuzzy design example

- Van Der Klauw (2002) investigates the effect of scholarship on college enrollment
- In the US, scholarship amounts are determined by test scores, but not perfectly (parental income, affirmative action, cover letters, etc.)
- → **Fuzzy discontinuities in the amount of the scholarship** around various threshold for test scores

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD
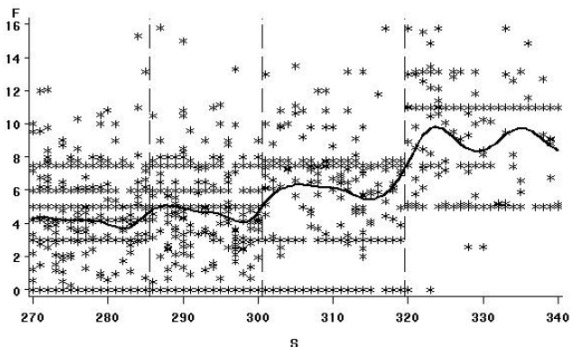
Fuzzy design example

- Angrist & Lavy (1999) investigate the effect of class size on primary school achievement
- In Israel, there is an old 'Maïmonides' rule about 'filling' classes : classes are 'filled' up to a certain threshold (40 pupils), after which a new class is created
- → **Fuzzy discontinuities in class size** around various thresholds for school size (multiples of 40)

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD
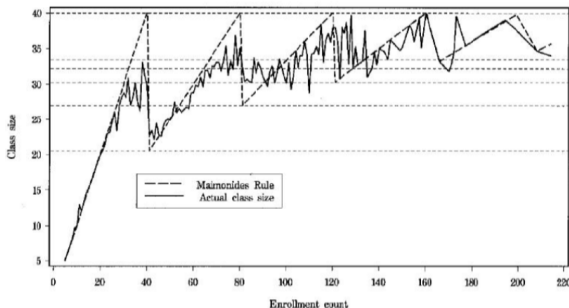
The RDD identification assumptions

- **Identification assumptions :**

  1. There exists a continuous selection variable Z such that (the probability) of treatment T shows a **discontinuity around a threshold $\underline{Z}$** (or a cap $\bar{Z}$)

  $\rightarrow$ **Discontinuity in treatment T**

  2. The potential outcomes Y_1 et Y_0 are **continuous** around the discontinuity threshold $\underline{Z}$
  $\simeq$ The **unobserved component** of the outcome $E(\epsilon|Z)$ is **continuous** at $\underline{Z}$ (as well as the treatment effect β)

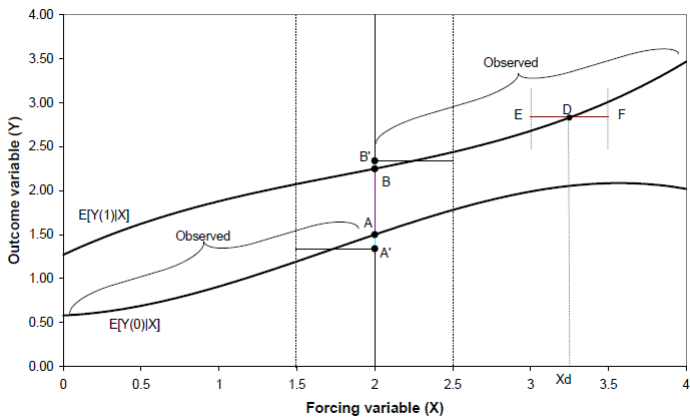 $\rightarrow$ **Continuity in the conditional regression function $E(Y|Z)$**

- Formally, these assumptions write :

$$(1) \quad T^+ = \lim_{Z \to \underline{Z}^+} E(T|Z) \quad \neq \quad T^- = \lim_{Z \to \underline{Z}^-} E(T|Z)$$

$$(2) \quad \epsilon^+ = \lim_{Z \to \underline{Z}^+} E(\epsilon|Z) \quad = \quad \epsilon^- = \lim_{Z \to \underline{Z}^-} E(\epsilon|Z)$$

Note – *In practice, we only need continuity of the (unobserved component of the) outcome in a neighborhood of the threshold $\underline{Z}$.*

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD

Identification assumptions on a graph

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD

Average treatment effect in a sharp design

- Let's compare individuals **in the neighborhood of the threshold :**

$$\lim_{Z \to \underline{Z}^+} E(Y_i|Z_i) - \lim_{Z \to \underline{Z}^-} E(Y_i|Z_i) = [\lim_{Z \to \underline{Z}^+} E(\beta T_i|Z_i) + \lim_{Z \to \underline{Z}^+} E(\epsilon_i|Z_i)]$$

$$- [\lim_{Z \to \underline{Z}^-} E(\beta T_i|Z_i) + \lim_{Z \to \underline{Z}^-} E(\epsilon_i|Z_i)]$$

$$= \lim_{Z \to \underline{Z}^+} E(\beta T_i|Z_i) - \lim_{Z \to \underline{Z}^-} E(\beta T_i|Z_i)$$

$$+ \underbrace{\lim_{Z \to \underline{Z}^+} E(\epsilon_i|Z_i) - \lim_{Z \to \underline{Z}^-} E(\epsilon_i|Z_i)}_{=0}$$

$$= \beta \times [\lim_{Z \to \underline{Z}^+} E(T_i|Z_i) - \lim_{Z \to \underline{Z}^-} E(T_i|Z_i)]$$

$$= \beta \times (T^+ - T^-)$$

- In a **sharp design**, since $T^+ = 1$ and $T^- = 0$, the RDD estimator is :

$$\beta_{RDD} = \lim_{Z \to \underline{Z}^+} E(Y_i|Z_i) - \lim_{Z \to \underline{Z}^-} E(Y_i|Z_i) = Y^+ - Y^-$$

- Warning ! This is only a **local treatment effect** estimated at $Z = \underline{Z}$
$\rightarrow$ If treatment effects are **heterogenous**, then in general :
$$\beta_{RDD} \neq ATT \neq ATE$$

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD

Average treatment effect in a fuzzy design

- To identify the treatment effect in a fuzzy design, we need an additional **monotonicity assumption** :

$$P(T_i = 1 | Z \geq \underline{Z}) > P(T_i = 1 | Z < \underline{Z}) \quad \forall i$$

$\rightarrow$ $T_i(Z_i)$ is **non-decreasing in** Z_i around the discontinuity

$\rightarrow$ There are **no defiers !**

- Similar to an IV setting, we have :
  - **Compliers :** $\lim_{Z \to \underline{Z}^+} E(T_i | Z_i) = 1$ and $\lim_{Z \to \underline{Z}^-} E(T_i | Z_i) = 0$
  - **Never-takers :** $\lim_{Z \to \underline{Z}^+} E(T_i | Z_i) = 0$ and $\lim_{Z \to \underline{Z}^-} E(T_i | Z_i) = 0$
  - **Always-takers :** $\lim_{Z \to \underline{Z}^+} E(T_i | Z_i) = 1$ and $\lim_{Z \to \underline{Z}^-} E(T_i | Z_i) = 1$

- In a **fuzzy design**, the RDD estimator is a **Wald estimator** (with instrument $Z = 1$ when Z is close to $\underline{Z}$) :

$$\beta_{RDD} = \frac{\lim_{Z \to \underline{Z}^+} E(Y_i | Z_i) - \lim_{Z \to \underline{Z}^-} E(Y_i | Z_i)}{\lim_{Z \to \underline{Z}^+} E(T_i | Z_i) - \lim_{Z \to \underline{Z}^-} E(T_i | Z_i)} = \frac{Y^+ - Y^-}{T^+ - T^-}$$

- Warning ! This is only a **very LATE** estimated on **compliers** at $Z = \underline{Z}$

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD

Plan

The basic RDD principle

RDD estimation in a regression framework

The validity of RDD

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD

The two estimation techniques

- There is **no empirical counterpart** to the limit
→ We have to use data away from the threshold $\underline{Z}$

- Since the treatment effect is only **identified locally** around the threshold $\underline{Z}$, we need to **restrict the estimation** to observations for which the selection variable Z tends toward $\underline{Z}$ (from below and above)
→ We have to use observations whose Z_i values are (as) close (as possible) to the discontinuity

- There are **2 approaches** to estimation :
  1. **Non-parametric estimator :** we compute a **weighted average** of the outcome on both sides of $\underline{Z}$ using **Kernel functions** as weights and choosing an estimation 'window' (bandwidth)

  2. **Parametric estimator :** we use **local linear regression** choosing a functional form and an estimation 'window' (bandwidth)

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD

The dilemma with the estimation 'window'

- The RDD estimator is **sensitive the bandwidth** h we use around the threshold $\underline{Z}$ to restrict the sample of estimation :

$$\underline{Z} - h \leq Z_i \leq \underline{Z} + h$$

- We are more confident if the observations are **close to the discontinuity**, since treatment effect is identified around this cutoff point...

- However, we need a sufficient number of observation around the discontinuity to get enough **statistical power** (i.e to detect an effect)
  - ▶ A too narrow window will yield imprecise estimates (too little data)
  - ▶ A too wide window will yield biased estimates (too far from threshold)

- Standard econometric **trade-off between bias and precision :**
  - ▶ **The more observations** (far from the threshold), the **more precise** the estimate but with a risk of bias
  - ▶ **The less observations** (close to the threshold), the **less biased** is the estimate but with a lack of precision

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD

How do we choose the estimation 'window'?

- We can use sophisticated techniques to choose the optimal bandwidth h like **cross (trimmed) validation**

- The idea is to **compare the observed Y and estimated $\hat{Y}$ outcomes** on a the sample located in this window, and keep the bandwidth that provides the closest estimates
  1. First choose a initial h_1, looking at the data (natural 'window?')
  2. Estimate the treatment effect and compute the outcome fitted values $\hat{Y}$
  3. Repeat the estimation using different values of h
  4. Compute a cross-validation criterion $\simeq R^2$ (Akaike's Information Criterion (AIC) or Bayesian Information Criterion (BIC))
  $\rightarrow$ The optimal bandwidth is the value h which maximizes this cross-validation criterion

- In pratice, we choose an **ad-hoc bandwidth** (that seems 'reasonable')
  - ▶ Imbens and Lemieux (2008) suggest dropping half of the observations on either side of the threshold $\underline{Z}$
  - ▶ Always **check the robustness** of the results by re-estimating the effect for slightly larger or smaller 'windows'

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD

Non-parametric estimation

- We want to estimate **2 limits :** $lim_{Z \to \underline{Z}^+} E(Y_i|Z_i)$ and $lim_{Z \to \underline{Z}^-} E(Y_i|Z_i)$

- We can use **Kernel estimators** (that weight the observations according to their distance to the threshold $\underline{Z}$)

- The above limits can be estimated by :

$$\frac{\sum_i 1_{(Z_i \geq \underline{Z})} k_h(Z_i - \underline{Z}) Y_i}{\sum_i 1_{(Z_i \geq \underline{Z})} k_h(Z_i - \underline{Z})} \quad \text{and} \quad \frac{\sum_i 1_{(Z_i < \underline{Z})} k_h(Z_i - \underline{Z}) Y_i}{\sum_i 1_{(Z_i < \underline{Z})} k_h(Z_i - \underline{Z})}$$

where k_h is a kernel function with bandwidth h

- In practice, very simple estimation using a **Kernel uniform function**

$$k_h(Z_i) = 1_{(|Z_i| < h)}$$

$\Rightarrow$ Averaging the outcome on observations with values of Z_i within a distance h of the threshold $\underline{Z}$

- We can use more complex Kernel functions (probability density functions)

- However, **convergence issues** in small samples... so that we **mostly use parametric estimations**

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD

Parametric estimation in a sharp design

- We can alway assume that $E(\epsilon|Z)$ is a **continuous function** $g(Z)$ in a neighbourhood of the threshold $\underline{Z}$

- We can then write a **local linear regression** (to be estimated around the threshold $\underline{Z}$) :
$$Y_i = \alpha + \beta T_i + g(Z_i) + \epsilon_i$$

$\rightarrow$ We need to choose a **functional form for** $g(.)$

- A simple **linear approximation** writes :
$$Y_i = \alpha + \beta T_i + \gamma Z_i^* + \epsilon_i$$

  where $Z_{i*} = (Z_i - \underline{Z})$

- We can allow the **slopes to differ** on either side of the threshold $\underline{Z}$:
$$Y_i = \alpha + \beta T_i + 1_{[Z_i < \underline{Z}]}\gamma_b Z_i^* + 1_{[Z_i \geq \underline{Z}]}\gamma_a Z_i^* + \epsilon_i$$

- We can use more **flexible functional forms (polynomial)** to fit the data :
$$Y_i = \alpha + \beta T_i + 1_{[Z_i < \underline{Z}]}g(Z_i^*) + 1_{[Z_i \geq \underline{Z}]}g(Z_i^*) + \epsilon_i$$

$\Rightarrow$ $\hat{\beta}_{RDD}$ is estimated **through OLS** (observations within a given bandwidth h)

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD

Parametric estimation in a fuzzy design

- In a first step, we need to estimate the **probability of being treated** $P(T_i = 1|Z_i)$ around the threshold $\underline{Z}$ (choosing a bandwidth h)
- The variable $S_i = 1_{[Z_i \geq \underline{Z}]}$ is a **binary instrumental variable** for T_i
  - ▶ It is a strong predictor of treatment T_i (relevance)
  - ▶ It is uncorrelated to ϵ_i around the threshold $\underline{Z}$ (exogeneity)

- We can compute the **Wald estimator :**

$$\beta_{RDD} = \frac{E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0)}{E(T_i|Z_i = 1) - E(T_i|Z_i = 0)}$$

$$= \frac{E(Y_i|Z_i \geq \underline{Z}) - E(E(Y_i|Z_i < \underline{Z})}{E(T_i|Z_i \geq \underline{Z}) - E(E(T_i|Z_i < \underline{Z})}$$

- $\hat{\beta}_{RDD}$ can also be estimated **through 2SLS :**

$$\begin{cases} T_i = \delta + \pi S_i + \gamma Z_i^* + \nu_i & \textit{(First-stage)} \\ Y_i = \alpha + \beta \hat{T}_i + \phi Z_i^* + \epsilon_i & \textit{(Second-stage)} \end{cases}$$

where $\hat{T}_i$ is estimated using S_i and Z_i^* as resp. excluded/included instruments

- Warning ! **(Very) LATE** applies

Note – *In sharp and fuzzy designs, we do not 'need' any covariates other than Z, but they improve precision, may reduce the bias far from the threshold and are useful to check the validity of RDD.*

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD

Plan

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD

Graphical analysis

- The identification of a causal treatment effect under RDD requires **2 assumptions** :
    1. There is a **continuous** selection variable Z such that there is a **discontinuity in treatment probability** around a threshold $\underline{Z}$
    2. Other components of the outcome Y are **continuous at the threshold** $\underline{Z}$

- Even if it's not a formal proof, **graphical analyses** lend credence to the validity of the RDD estimation strategy

- Graphical analyses ensure that :
    1. The treatment variable T_i is **indeed discontinuous** around the known threshold $\underline{Z}$
    2. The outcome variable Y_i shows a **similar discontinuity** around the same threshold $\underline{Z}$
    3. There is no discontinuities in **other determinants** X_i of the outcome Y_i around the same threshold $\underline{Z}$
    4. There is no discontinuities in Y_i **outside of the threshold** $\underline{Z}$
    5. The selection variable Z_i has not been **manipulated**

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD

Graphical analysis of the treatment variable

- **Plot the treatment T_i by values of the selection variable Z_i**
    - Show how treatment probability varies as Z_i crosses the threshold $\underline{Z}$
    - In sharp designs, it should go from an average of 0 to an average of 1
    - In fuzzy designs, there should be a **discrete jump**

- **Example :** Ozier (2011) investigates the impact of secondary schooling in Kenya on various social and labor market participation indicators. He exploits the fact that pupils must pass an exam to enter high-school.

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD
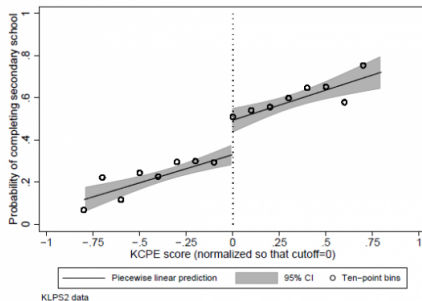
Graphical analysis of the outcome variable

- **Plot the outcome Y_i by values of the selection variable Z_i**

  - Show that there is a jump in the outcome at threshold $\underline{Z}$

  - Insights on the sign and possible significance of the treatment effect

  - If there is not visible jump, forget about it...

- **Example :** Lalive (2008) investigates the impact of unemployment benefits on unemployment duration

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD
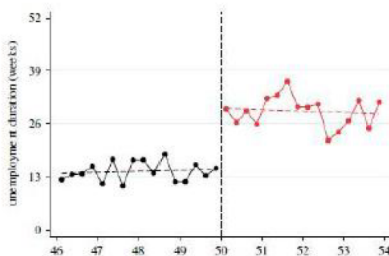
Graphical analysis of other covariates

- **Plot other (relevant) characteristics X_i by values of the selection variable Z_i**

  - It helps detect specification issues (you need individuals below and above the threshold $\underline{Z}$ to be really comparable)

  - If there is an (unjustified) visible jump in another variable X_i at the threshold $\underline{Z}$, the discontinuity in outcome may not be only related to treatment effect...

- **Example :** Lalive (2008) plots the probability of being married

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD
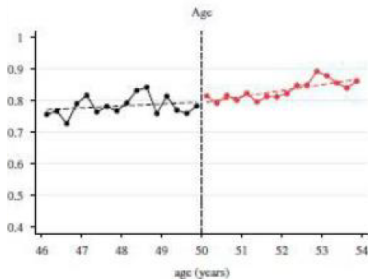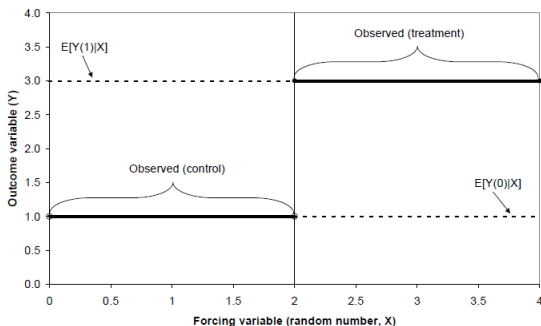
Graphical analysis of the selection variable

- **Plot the density of the selection variable Z_i**
  - Show that there are enough observations around the threshold

  - If there is a jump in density at the threshold $\underline{Z}$, the selection variable may have been manipulated...

  - If individuals can influence the 'forcing' variable (to get treated), this invalidates the analysis (re-introduces selection bias)

- **Example :** Camacho and Conover (2011) investigates the impact of an anti-poverty program whose selection is based on a poverty index score. The rule became known in 1997.

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD
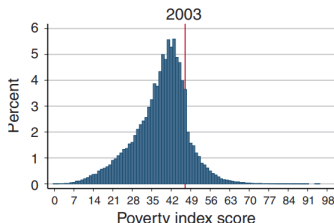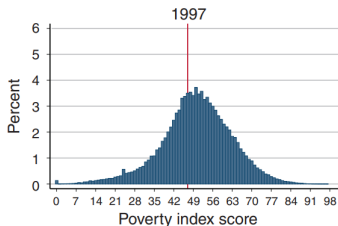
RDD as randomization

- Locally, RDD is close to a **randomization**

- Whether individuals end up below or above the threshold $\underline{Z}$ is in practice a **random draw**

- Requires that the selection variable is **non-manipulable**

- Identification is **only local** ($\simeq$ randomization around the threshold $\underline{Z}$)

Figure 3: Randomized Experiment as a RD Design

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD

RDD as matching

- **The conditional independence assumption (CIA)** from matching models is trivially satisfied in a sharp design

- Conditional on Z, there is **no additional variation in T** :

$$(Y_0, Y_1) \perp T | Z$$

- However, there is **no common support** in a sharp design, since for any specific value of Z, either $P(T = 1|Z) = 1$ or $= 0$ depending on whether $Z \geq \underline{Z}$ (or not)

- But RDD allows deviation from the (sharp) selection rule

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD

RDD as IV

- In fuzzy designs, we use the selection variable Z_i as an **instrument** :

$$S_i = 1_{(Z_i > \underline{Z})}$$

- In fact, the borderline between the two methods is quite **blurred** : identification techniques that would have been called instrument a few years ago are now called RDD

- The main difference is the **assumptions** under which the causal treatment effect is identified :
    - In IV estimation strategies, potential outcomes are **independent** from the instrument (selection variable)

    - In RDD estimation strategies, potential outcomes are **continous** around the instrument (i.e threshold for the selection variable)

  $\Rightarrow$ Identification assumptions are **less strong in RDD than in IV**
  $\Rightarrow$ **Cost :** the RDD estimator is a **(very) LATE**

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD

The disadvantages of RDD

- **External validity :** estimates are local around the threshold $\underline{Z}$

- **Very demanding in terms of data** around the threshold $\underline{Z}$ to get precise estimates

- **Need to assume homogenous effects** to extrapolate the results to observations far from the threshold

- RDD only applies to treatments that have a **precise threshold rule for eligibility**

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD

To sump up

Steps for implementing RDD estimations :

1 – Find a setting in which the treatment has a **precise threshold rule for eligibility**

2 – Have detailed data (and large enough samples) to be able to **restrict to the observations that are close to the discontinuity**

3 – **Run a graphical analysis** to check the relevance of the RDD
The treatment probability and the outcome must show a visible discontinuity around the threshold. There is no other discontinuities (in outcome and other covariates). The selection variable should not be manipulable.

4 – **Use non-parametric or parametric local estimations** by restricting the estimation to observations close to the discontinuity
The choice of the bandwidth might be ad-hoc but check the robustness to various 'windows'. In linear regressions, use (flexible) OLS in a sharp design and (flexible) 2SLS in a fuzzy design.

5 – **Interpret the results :** it may be a **(very) LATE**, i.e the treatment effect for the compliers around the discontinuity (always assess the external validity of your estimation)

The basic RDD principle
RDD estimation in a regression framework
The validity of RDD

References

Angrist & Lavy (1999). "Using Maimonides' rule to estimate the effect of class size on scholastic achievement," Quarterly Journal of Economics, 114(2), 533–575.

Black (1999). "Do better schools matter ? Parental valuation of elementary education," Quarterly Journal of Economics, 114(2), 587–599.

Camacho & Conover (2011). Manipulation of social program eligibility. American Economic Journal : Economic Policy, 3(2), 41-65.

Hahn, Todd & Van der Klaauw (2001). "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," Econometrica, 69(1), 201–09.

Imbens & Lemieux (2008). "Regression discontinuity designs : A guide tp practice," Journal of Econometrics, 142(2), 615–635.

Lalive (2008). "How do extended benefits affect unemployment duration ? A regression discontinuity approach" Journal of Econometrics, 142(2), 785–806.

Ozier (2018). The impact of secondary schooling in Kenya a regression discontinuity analysis. Journal of Human Resources, 53(1), 157-188.

Van der Klaauw (2002). "Estimating the Effect of Financial Aid offers on College Enrollment : A regression discontinuity design" International Economic Review, 43(4), 1249–1487.