Regression Discontinuity Design (RDD)

Traduction mot à mot du S3_txt.pdf

politique publique

M1 ECAP – TD4 – Année 2024/2025

chapitre 2s

Responsable d'enseignement : Benoît SÉVI
benoit.sevi@univ-nantes.fr

QUINTIN DE KERCADIO Pierre

March 28, 2025

Table of contents

Introduction

Très souvent, les individus traités ne sont pas un échantillon aléatoire de la population, donc on ne peut pas facilement trouver un groupe de contrôle avec lequel les comparer.

Un **Regression Discontinuity Design (RDD)** permet de créer un groupe de contrôle très proche du groupe traité : ceux qui **n'ont pas bénéficié du traitement par hasard**.

Pour cela, il faut que le traitement soit attribué selon une **règle claire basée sur un seuil** :

- Les individus **au-dessus** du seuil sont traités
- Ceux **en-dessous** ne le sont pas

Exemple

On veut connaître l'effet d'obtenir une bonne note au bac sur la réussite dans le supérieur.

Comparer les individus ayant au moins **14/20** avec ceux en-dessous n'est **pas une bonne idée** :

- Trop de différences entre les deux groupes : niveau académique antérieur, motivation, environnement familial…
- Ces caractéristiques sont souvent **non observées** mais **corrélées avec le traitement** → biais d'endogénéité.

Meilleure idée : comparer ceux entre **10 et 10.1** (traités) avec ceux entre **9.9 et 9.99** (non traités).

- Ces derniers ont probablement échoué pour des raisons aléatoires (stress, note, etc.)
- Ils sont donc très **comparables** au groupe traité.

Idée principale du RDD

Tirer parti des règles d'allocation du traitement pour comparer des individus aux caractéristiques proches.

Hypothèse :

Bien que les groupes traités et non traités soient **très différents**, les individus **juste au-dessus** et **juste en-dessous** du seuil sont très similaires.

Autre exemple : alcool et accidents de la route

On veut étudier l'effet de la consommation d'alcool sur les accidents.

- Traitement : avoir le droit de boire de l'alcool
- Dans un pays où la vente d'alcool est autorisée à partir de 21 ans :

Mauvaise idée : comparer les <21 ans avec les >21 ans
Meilleure idée : comparer ceux de **20 ans et 10 mois** à ceux de **21 ans et 2 mois**

> Source : Carpenter & Dobkin (2009)

Condition nécessaire

Il faut pouvoir attribuer ou refuser le traitement à partir d'un **critère objectif**.

- Les individus **ne doivent pas pouvoir choisir** s'ils sont éligibles ou non.
- Cette règle claire permet de définir une **discontinuité nette**.

Exemple visuel – Page 7

Page 7 – Graphique : Conditions d'éligibilité au programme

Analyse : - Ce graphique illustre un seuil bien défini pour l'accès à un programme. - Il montre que **le traitement est déterminé uniquement par la variable de score/ressource**. - On a ici une situation idéale pour un Sharp RDD.

Exemple visuel – Page 8

Page 8 – Graphique : Effet du programme

Analyse : - La discontinuité verticale indique une **variation nette du résultat au seuil**, donc un **effet causal du programme**. - Cet effet est mesuré par la différence de l'ordonnée de part et d'autre du seuil.

---

1. Formalisation du RDD

1.1 Sharp RDD

On peut écrire un Sharp RDD ainsi :

$$D_i = \begin{cases} 1 & \text{si } x_i \geq x_0 \\ 0 & \text{si } x_i < x_0 \end{cases}$$

f(x) non-linéaire : deux approches

Supposons que la relation soit non linéaire :

$$\mathbb{E}[Y_i^0 | x_i] = f(x_i)$$

On peut alors estimer l'équation suivante :

$$Y_i = f(x_i) + \rho D_i + \eta_i$$

Deux méthodes principales pour approximer $f(x_i)$:

1. **Polynôme d'ordre p :**

$$Y_i = \alpha + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_p x_i^p + \rho D_i + \eta_i \tag{1}$$

2. **Méthode non paramétrique par noyau** (voir section 2.1.1)

Polynômes différents des deux côtés de la discontinuité

On peut généraliser la fonction $f(x_i)$ en autorisant des expressions différentes de part et d'autre du seuil, en ajoutant des interactions avec D_i.

> Comme le soulignent **Lee et Lemieux (2010)**, il est important d'estimer séparément les deux côtés de la coupure, sinon on utiliserait des valeurs de chaque côté pour estimer l'autre.

On a alors :

$$\mathbb{E}[Y_i^0|x_i] = \alpha + \beta_{01}\tilde{x}_i + \beta_{02}\tilde{x}_i^2 + \cdots + \beta_{0p}\tilde{x}_i^p$$

$$\mathbb{E}[Y_i^1|x_i] = \alpha + \rho + \beta_{11}\tilde{x}_i + \beta_{12}\tilde{x}_i^2 + \cdots + \beta_{1p}\tilde{x}_i^p$$

avec $\tilde{x}_i = x_i - x_0$ (centrage autour du seuil)

Intérêt du centrage :

Cela permet d'interpréter directement le coefficient de D_i comme l'effet du traitement au seuil ($x_i = x_0$), car tous les termes croisés avec $\tilde{x}_i$ s'annulent à ce point.

Modèle de régression complet (interaction polynomiale)

On peut combiner les expressions précédentes en une équation unique :

$$Y_i = \alpha + \sum_{j=1}^{p} \beta_{0j}\tilde{x}_i^j + \rho D_i + \sum_{j=1}^{p} \beta_j^* D_i \tilde{x}_i^j + \eta_i \tag{2}$$

avec :

- $\beta_j^* = \beta_{1j} - \beta_{0j}$

Cas particulier :

L'équation (1) est un cas particulier de (2) où tous les $\beta_j^* = 0$ (pas de différences de pente de chaque côté)

Interprétation de l'effet du traitement

- **Effet au seuil** x_0 : ρ
- **Effet à une distance** $c > 0$ **du seuil** :

$$\rho + \beta_1^* c + \beta_2^* c^2 + \cdots + \beta_p^* c^p$$

Exemple de Sharp RDD : effet de l'incumbency — Lee (2008)

Étude : Lee (2008), *Journal of Econometrics*

- Objectif : estimer l'effet d'avoir remporté une élection précédente sur la probabilité de remporter la suivante
- Hypothèse : les **candidats en poste** ont un avantage (expérience, visibilité, réseau…)

Mais :

- Une régression MCO de la réélection sur le fait d'être sortant est **biaisée** : les sortants sont différents (plus forts ?)

Lee utilise une approche Sharp RDD :

- Compare les candidats **ayant gagné de justesse** avec ceux **ayant perdu de justesse**
- Ex : élection à la Chambre des représentants aux États-Unis

---

Page 17 – Graphique : effet d’avoir gagné l’élection précédente

Analyse graphique :

- Le graphique montre une **discontinuité nette** dans la probabilité de gagner l’élection suivante au point de victoire juste au-dessus de 50 %.
- Cela illustre parfaitement le cadre du Sharp RDD : **gagner l’élection précédente agit comme un “traitement”**.

  On interprète la **différence de probabilité de réélection** au seuil comme un **effet causal de l’incumbency**.

---

2. Fuzzy RDD

2.1 Définition

Un **Fuzzy RDD** est un cas où le **traitement n’est pas parfaitement déterminé par la règle du seuil**.

Exemple : certaines personnes au-dessus du seuil **ne reçoivent pas** le traitement, et certaines en-dessous le **reçoivent quand même**.

Le **RDD devient alors un cas particulier de variable instrumentale (IV)** :

- La variable (D_i) (traitement) est **instrumentée** par la variable (Z_i = 1{x_i x_0})
- On estime alors un effet local : **le LATE**

2.2 Illustration graphique

Page 32 – Graphique : discontinuité dans la probabilité de traitement (Fuzzy)

Analyse :

- Contrairement au Sharp RDD, le **saut de la variable de traitement au seuil n’est pas de 0 à 1**, mais partiel.
- Le traitement devient **plus probable** au-dessus du seuil, mais pas systématique.

Cela reflète un **non-respect imparfait** de la règle, ce qui est courant en pratique.

2.3 Identification dans le Fuzzy RDD

On suppose que :

- La **probabilité de traitement** présente une **discontinuité au seuil** :

$$\lim_{x \downarrow x_0} \mathbb{P}(D_i = 1|x_i) \neq \lim_{x \uparrow x_0} \mathbb{P}(D_i = 1|x_i)$$

- Et que (x_i) **n'affecte pas directement (Y_i)** sauf via (D_i)

On peut alors identifier :

$$\text{Effet causal local} = \frac{\text{Saut de } Y_i}{\text{Saut de } D_i}$$

C'est exactement l'**estimateur de Wald** avec un instrument binaire.

2.4 Estimation

Deux étapes (2SLS) :

Étape 1 :

Estimer la probabilité de traitement conditionnelle au score :

$$D_i = f(x_i) + \gamma Z_i + \nu_i$$

où (Z_i = 1{x_i x_0})

Étape 2 :

Utiliser la variable instrumentée :

$$Y_i = f(x_i) + \rho \widehat{D_i} + \eta_i$$

Page 34 – Exemple empirique

> Exemple classique de Fuzzy RDD : analyse des politiques sociales où certaines personnes éligibles ne réclament pas le droit.

Page 34 – Graphique : discontinuité dans le traitement floue

Analyse :

- La discontinuité est présente dans **l'issue (Y)** et dans la **probabilité de traitement (D)**.
- Il faut utiliser un modèle Fuzzy pour estimer correctement l'effet du programme.

---

3. Tests de validité dans un RDD

3.1 Pourquoi tester ?

Le RDD repose sur une **hypothèse forte** : les individus **ne peuvent pas manipuler précisément leur valeur de (x)** pour être juste au-dessus ou au-dessous du seuil.

Exemple : dans un concours ou une règle d'admissibilité, si des individus peuvent **tricher ou se positionner stratégiquement**, la discontinuité n'est plus exogène.

3.2 Test de McCrary

> McCrary (2008) propose un test de discontinuité dans la densité du score (x_i)

Idée :

- Si les individus manipulent leur score, on observera une **saut dans la densité** de (x_i) au niveau du seuil.
- Si la densité est continue, on peut considérer que le **RDD est crédible**.

Page 38 – Graphique : estimation de la densité de (x_i) (test de McCrary)

Analyse :

- Si la courbe de densité est **continue au seuil**, alors **pas de manipulation détectée**.

- Si une rupture nette est visible → **RDD invalide**.

3.3 Test de discontinuité dans les covariables

> Autre test utile : vérifier que les **caractéristiques observables** (âge, sexe, etc.) ne changent pas brusquement au seuil.

Idée :

- Ces covariables ne devraient pas avoir de **saut au seuil** si le traitement est bien exogène.

Page 39 – Graphiques : test sur des covariables

Analyse :

- Les résultats montrent que les covariables ne présentent **pas de discontinuité nette**.
- Cela renforce la validité du RDD : les groupes autour du seuil sont bien comparables.

3.4 Placebo RDD (faux seuils)

Méthode :

- Répéter l'analyse RDD en plaçant **des seuils fictifs** où il n'y a **pas de règle d'allocation réelle**.
- Si l'on trouve un effet significatif au faux seuil → doute sur la validité du vrai RDD.

Souhaites-tu que je poursuive avec les **pages 46 à 60**, qui vont vers les applications empiriques (ex. Boston schools) et les extensions ?

4. Exemple empirique : Boston Exam Schools

4.1 Contexte

Dans les villes comme **Boston** et **New York**, certaines écoles très sélectives (appelées **exam schools**) sélectionnent leurs élèves sur la base d'un score d'entrée.

Ces écoles offrent :

- Des camarades d'un niveau élevé
- Davantage de ressources
- Potentiellement de meilleurs enseignants

Question :

Est-ce que **fréquenter une école d'élite améliore les résultats scolaires** ?

4.2 Pourquoi pas une simple comparaison ?

Comparer les élèves des exam schools avec les autres **ne donne pas un effet causal**, car :

- Les élèves des exam schools sont **différents ex ante**
- Ils sont probablement **plus motivés**, **mieux préparés**, issus d'un milieu plus favorisé

Peut-on utiliser une expérience aléatoire ?

Non, car si l'entrée était aléatoire, **les exam schools n'auraient plus d'élèves sélectionnés**, donc cela changerait la nature même du traitement.

4.3 Pourquoi c'est un bon cas de Fuzzy RDD

L'entrée dans une exam school dépend d'un **seuil sur le score d'examen**.

Les élèves juste **au-dessus du seuil sont très similaires** à ceux juste **en-dessous**, à l'exception de leur admission.

Mais :

- Certains élèves au-dessus du seuil ne vont **pas** dans l'école
- D'autres en-dessous peuvent aller ailleurs

→ C'est donc un **Fuzzy RDD**, pas un Sharp.

Page 38 – Premier stade : Boston Latin School

Graphique : taux d'inscription en fonction du score à l'examen

Analyse :

- On observe une **forte discontinuité** dans la probabilité d'inscription à BLS au seuil.
- L'effet est partiel → Fuzzy RD.

Page 39 – Qualité des pairs (peer quality)

On constate que **ne pas être admis** à BLS modifie fortement la qualité des pairs.

> Mesurée par les **résultats en maths des camarades 2 ans avant** l'examen d'entrée

Résultat : +0.8 écart-type dans la qualité des pairs à BLS

Analyse :

- L'admission change significativement l'environnement scolaire, même si cela ne garantit pas un effet sur les notes futures.

Page 40 – Effet sur les résultats en maths

Question : Est-ce que la **hausse de la qualité des pairs** améliore les résultats en maths des élèves ?

Résultat : **aucun effet significatif** après 1 ou 2 ans dans l'école.

Graphique : scores de math 7 et 8 année

Analyse :

- Malgré l'amélioration de l'environnement, les scores en mathématiques **ne bougent pas**.

4.4 Estimation de l'effet de la qualité des pairs

Modèle estimé :

$$Y_i = \beta_1 + \beta_2 X_{-i} + \beta_3 X_i + \epsilon_i$$

où :

- (Y_i) : score en maths de l'élève (i)
- (X_{-i}) : score moyen des camarades (4 année)
- (X_i) : score individuel de l'élève (i) (4 année)

Pourquoi pas OLS ?

Le score moyen des camarades (X_{-i}) est probablement **corrélé avec des facteurs non observés** affectant (Y_i).

→ Cela entraînerait un **biais de variable omise**

Solution : utiliser le **Fuzzy RDD comme instrument** de la qualité des pairs

---

Page 42 – Forme réduite (reduced form)

Équation estimée :

$$Y_i = \lambda_1 + \lambda_2 D_i + \lambda_3 R_i + \nu_i$$

- (D_i) : indicatrice = 1 si l'élève est au-dessus du seuil
- (R_i) : score relatif à la coupure (running variable)

Résultat : Pas de saut significatif → **pas d'effet visible du traitement sur (Y_i)**

Page 43 – Premier stade (First Stage)

On estime :

$$X_{-i} = \gamma_1 + \gamma_2 D_i + \gamma_3 R_i + \mu_i$$

Résultat : être admis augmente nettement la qualité des pairs

→ Cela confirme la **validité de l'instrument**

Page 44 – 2SLS

Deuxième étape :

$$Y_i = \beta_1 + \beta_2 \widehat{X}_{-i} + \beta_3 R_i + u_i$$

Résultat : l'effet estimé de la qualité des pairs **reste nul**

Conclusion de l’exemple Boston

- Être admis dans une école d’élite **améliore la qualité des pairs**
- Mais cela **n’améliore pas les performances scolaires mesurées en maths**

Cela va **à l’encontre de l’intuition** selon laquelle de “meilleurs camarades” amélioreraient automatiquement les résultats

---

5. Conclusion et remarques finales

Résumé des étapes clés d’un RDD crédible :

1. **Vérifier qu’il existe une règle de seuil bien définie** :
    - Le traitement est alloué selon une variable continue x et une coupure x_0
2. **Choisir entre un Sharp ou un Fuzzy RDD** :
    - Sharp : traitement parfaitement déterminé par la coupure
    - Fuzzy : saut dans la probabilité du traitement au seuil
3. **Spécifier la forme fonctionnelle de $f(x)$** :
    - Polynomiale ou non paramétrique (noyau / local linear)
4. **Choisir une fenêtre (intervalle) autour du seuil** :
    - Compromis biais/variance
    - Utiliser des outils statistiques pour déterminer la largeur optimale
5. **Estimer l’effet causal** :
    - Discontinuité dans Y au seuil x_0 (Sharp)
    - Ratio des discontinuités (Y/D) pour le Fuzzy $\rightarrow$ LATE

---

6. Tests de robustesse et de validité

Pour renforcer la crédibilité de l'analyse RDD, on recommande :

- **Tester la densité de la variable de score (test de McCrary)** :
  - Une discontinuité indiquerait une manipulation stratégique
- **Tester les covariables** :
  - Elles ne doivent pas sauter au seuil
  - Permet de valider la comparabilité des groupes
- **Placebo thresholds** :
  - Tester l'effet de fausses coupures
  - S'assurer que les résultats ne sont pas dus au hasard

---

7. Portée des résultats – LATE vs ATE

L'effet estimé dans un RDD est **local** :

- Il reflète l'effet causal pour les individus **proches du seuil**
- Ce n'est **pas** un effet moyen pour l'ensemble de la population

  Cela soulève des questions de **validité externe**

---

8. Avantages du RDD

Forte crédibilité causale (quasi-expérimental)

Permet de capturer des effets locaux très précis

Approche intuitive et visualisable (via les graphes de discontinuité)

---

9. Limites du RDD

Requiert une règle de traitement stricte et connue

Applicable uniquement **autour du seuil**

Nécessite beaucoup de données autour du point de coupure

Sensible à la spécification de la fonction $f(x)$ et au choix de la fenêtre

---

10. Références bibliographiques clés

- Lee & Lemieux (2010) – *Regression Discontinuity Designs in Economics*, Journal of Economic Literature
- McCrary (2008) – *Manipulation of the running variable*, Journal of Econometrics
- Hahn, Todd, Van der Klaauw (2001) – *Identification and estimation with RD*, Econometrica
- Angrist & Lavy (1999), Angrist & Krueger (1991)
- Abdulkadiroğlu, Angrist, Pathak (2014) – *Boston Exam Schools*, Econometrica

---

Fin du chapitre

Ce chapitre sur le Regression Discontinuity Design vous aura permis : - De comprendre les fondements théoriques du RDD - D'identifier les différences entre Sharp et Fuzzy - D'appliquer des méthodes économétriques rigoureuses à des cas concrets - De tester la validité des hypothèses et de conduire des analyses robustes

---