

Regression Discontinuity Design (RDD)

Pratiques de la Recherche en Économie

Florentine Oliveira-Roux

2026-03-30

Cette séance

1. Intuition et définitions
2. Sharp RDD
 - 2.1. Estimateur
 - 2.2. Hypothèse d'identification
 - 2.3. Exemples
 - 2.4. Estimation sur `R`
 - 2.5. Importance de la forme fonctionnelle
3. Fuzzy RDD

1. Intuition et définitions

Intuition

De nombreux traitements sont définis selon une **règle/un seuil**

Par exemple:

- le revenu à partir duquel un individu peut bénéficier d'une prestation sociale
- l'âge d'entrée à l'école, âge légal pour avoir droit de vote, consommer de l'alcool (majorité)
- moyenne au bac requise pour pouvoir candidater à certaines écoles

La régression sur discontinuité, ou **Regression Discontinuity Design**, exploite ce(s) seuil(s) pour estimer l'effet causal du traitement

Intuition: **exogénéité locale**

- les individus proches du seuil sont raisonnablement comparables
- cependant ceux au-dessus du seuil sont traités alors que ceux en dessous ne le sont pas
- la discontinuité crée une **quasi-expérience** au voisinage du seuil de discontinuité
 - autour du seuil, l'allocation au traitement est *as good as random*

Définition: *running variable*

Formellement, si l'on revient au framework des outcomes potentiels:

$$Y_i = Y_{1i}D_i + Y_{0i}(1 - D_i)$$

Maintenant,

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

où X_i est appelée *running/forcing variable*, c'est à dire la variable (*continue*) sur laquelle s'applique le critère de sélection dans le traitement

Deux types de régressions sur discontinuité:

- **sharp**: la probabilité de traitement devient certaine au seuil c , i.e. la probabilité pour l'individu i d'être traité passe de 0 à 1 (ex: la consommation légale d'alcool (D_i) à partir d'un certain âge (X_i))
- **fuzzy**: la probabilité de traitement augmente au seuil c mais ne passe pas nécessairement à 1 (ex: bourses sur critères de notes, redoublement)

Définitions: *bandwith* et forme fonctionnelle

Bandwith: intervalle autour du seuil de discontinuité dans lequel on conserve les observations pour estimer l'effet du traitement

⇒ Arbitrage:

- un bandwith trop étroit peut limiter le nombre d'observations
- un bandwith trop large peut inclure des observations moins comparables

Forme fonctionnelle: désigne la spécification de la relation entre l'outcome et la forcing variable dans le modèle de régression. Elle peut être linéaire, polynomiale, etc.

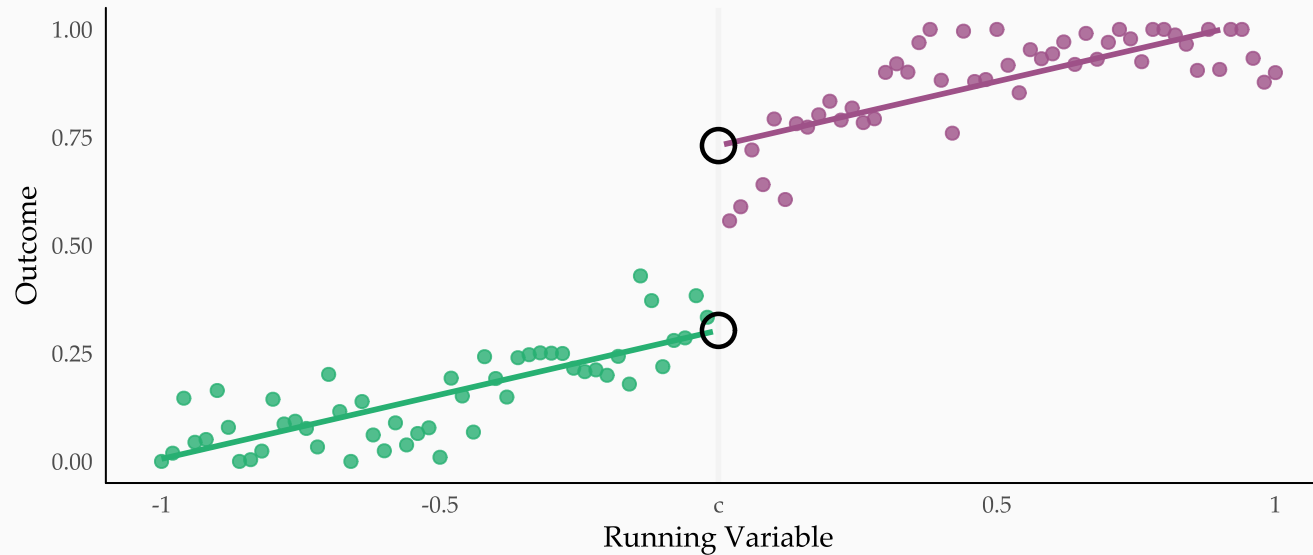
2. Sharp RDD

2.1. Estimateur

Définition

L'estimateur de l'effet causal du traitement D_i sur Y_i est simplement la différence de la moyenne de l'outcome Y_i de part et d'autre du seuil c :

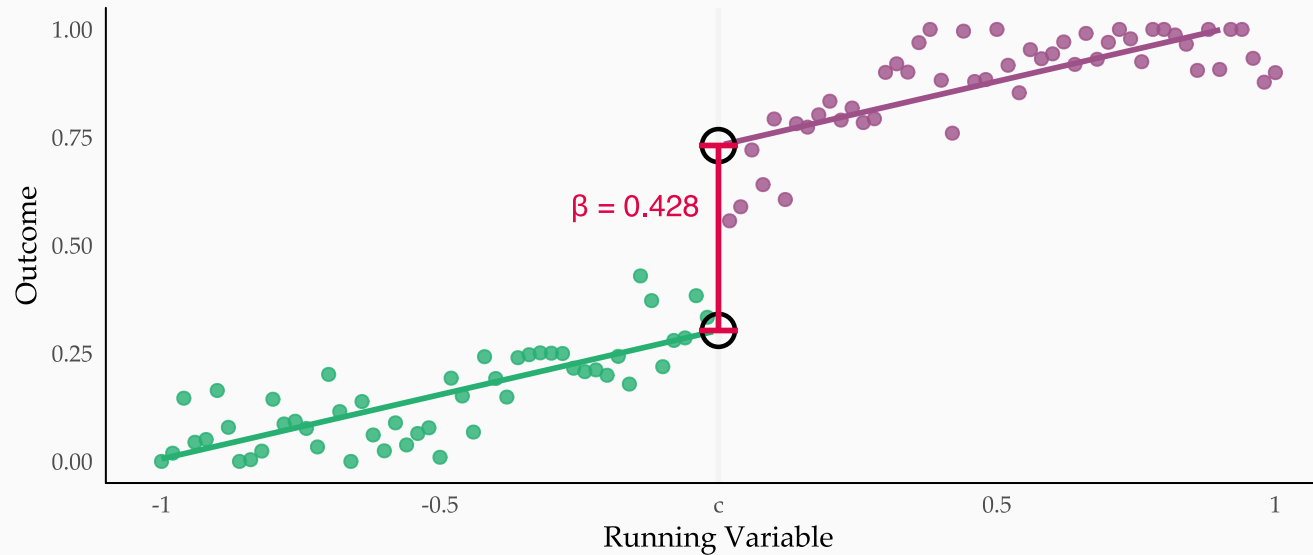
$$\begin{aligned}\beta_{\text{RDD}}^{\text{sharp}} &= \lim_{x \rightarrow c^+} \mathbb{E}(Y_i | X_i = x) - \lim_{x \rightarrow c^-} \mathbb{E}(Y_i | X_i = x) \\ &= \lim_{x \rightarrow c^+} \mathbb{E}(Y_{1i} | X_i = x) - \lim_{x \rightarrow c^-} \mathbb{E}(Y_{0i} | X_i = x)\end{aligned}$$



Définition

L'estimateur de l'effet causal du traitement D_i sur Y_i est simplement la différence de la moyenne de l'outcome Y_i de part et d'autre du seuil c :

$$\begin{aligned}\beta_{\text{RDD}}^{\text{sharp}} &= \lim_{x \rightarrow c^+} \mathbb{E}(Y_i | X_i = x) - \lim_{x \rightarrow c^-} \mathbb{E}(Y_i | X_i = x) \\ &= \lim_{x \rightarrow c^+} \mathbb{E}(\textcolor{violet}{Y}_{1i} | X_i = x) - \lim_{x \rightarrow c^-} \mathbb{E}(\textcolor{teal}{Y}_{0i} | X_i = x)\end{aligned}$$



2.2. Hypothèse d'identification

Continuité

Hypothèse d'identification: $\mathbb{E}(Y_{1i}|X_i = x)$ et $\mathbb{E}(Y_{0i}|X_i = x)$ sont continues en x

$$\text{Donc } \beta_{\text{RDD}}^{\text{sharp}} = \lim_{x \rightarrow c^+} \mathbb{E}(Y_i|X_i = x) - \lim_{x \rightarrow c^-} \mathbb{E}(Y_i|X_i = x) = \mathbb{E}(Y_{1i} - Y_{0i}|X_i = c)$$

Continuité

Hypothèse d'identification: $\mathbb{E}(Y_{1i}|X_i = x)$ et $\mathbb{E}(Y_{0i}|X_i = x)$ sont continues en x

$$\text{Donc } \beta_{\text{RDD}}^{\text{sharp}} = \lim_{x \rightarrow c^+} \mathbb{E}(Y_i|X_i = x) - \lim_{x \rightarrow c^-} \mathbb{E}(Y_i|X_i = x) = \mathbb{E}(Y_{1i} - Y_{0i}|X_i = c)$$

Cela revient à supposer que les individus ne peuvent pas manipuler la valeur de X_i afin de se situer stratégiquement de l'un ou l'autre côté du seuil!

L'estimateur $\beta_{\text{RDD}}^{\text{sharp}}$ est un estimateur **local** de l'effet moyen du traitement (**LATE**).

NB: on n'a pas fait l'hypothèse ici d'assignation aléatoire du traitement D_i (et donc X_i)

NB2: il faut également que les variables de contrôle soient continues en c pour que l'estimateur de la RDD soit un estimateur non biaisé de l'effet causal du traitement

2.3. Examples

Congé maternité et santé mentale

Bütikofer, A., Riise, J., & M. Skira, M. (2021). *The impact of paid maternity leave on maternal health.* AEJ: Economic Policy

Motivation:

- Peu d'évidence sur les effets du congé maternité rémunéré sur la santé mentale des mères... alors même que c'est l'une des justifications de l'existence de ce congé

Congé maternité et santé mentale

Bütikofer, A., Riise, J., & M. Skira, M. (2021). *The impact of paid maternity leave on maternal health.* AEJ: Economic Policy

Motivation:

- Peu d'évidence sur les effets du congé maternité rémunéré sur la santé mentale des mères... alors même que c'est l'une des justifications de l'existence de ce congé
- Effet ambigu:
 - Effet positif si l'emploi augmente le stress ou réduit le temps que la femme consacre à s'occuper d'elle-même et à se remettre des effets physiques de l'accouchement
 - Effet négatif si cela permet à la mère d'avoir davantage d'interactions sociales et cela augmente le revenu

Congé maternité et santé mentale

Bütikofer, A., Riise, J., & M. Skira, M. (2021). *The impact of paid maternity leave on maternal health.* AEJ: Economic Policy

Motivation:

- Peu d'évidence sur les effets du congé maternité rémunéré sur la santé mentale des mères... alors même que c'est l'une des justifications de l'existence de ce congé
- Effet ambigu:
 - Effet positif si l'emploi augmente le stress ou réduit le temps que la femme consacre à s'occuper d'elle-même et à se remettre des effets physiques de l'accouchement
 - Effet négatif si cela permet à la mère d'avoir davantage d'interactions sociales et cela augmente le revenu

Contexte:

- Introduction d'un congé payé maternité en Norvège, au 1er Juillet 1977
 - avant la réforme, aucun congé rémunéré; seulement 12 semaines de congé non rémunérés
 - après la réforme: 4 mois de congés payés et jusqu'à 12 mois de congés non rémunérés

Question de recherche: quel effet l'introduction du congé payé a-t-elle eu sur la santé mentale des mères ?

Congé maternité et santé mentale

Bütikofer, A., Riise, J., & M. Skira, M. (2021). *The impact of paid maternity leave on maternal health.* AEJ: Economic Policy

Question: pourquoi ne peut-on pas simplement comparer la santé mentale moyenne des mères ayant recours au congé maternité et celle des mères n'y ayant pas recours?

Congé maternité et santé mentale

Bütikofer, A., Riise, J., & M. Skira, M. (2021). *The impact of paid maternity leave on maternal health.* AEJ: Economic Policy

Question: pourquoi ne peut-on pas simplement comparer la santé mentale moyenne des mères ayant recours au congé maternité et celle des mères n'y ayant pas recours?

Endogeneité

- certaines caractéristiques inobservables peuvent affecter à la fois le recours au congé maternité et la santé mentale des mères
- causalité inversée: si la santé mentale des mères impacte le recours au congé maternité

Congé maternité et santé mentale

Bütikofer, A., Riise, J., & M. Skira, M. (2021). *The impact of paid maternity leave on maternal health.* AEJ: Economic Policy

Stratégie d'identification: Regression Discontinuity Design!!

Question: quelle est la running variable? quel est le seuil?

Congé maternité et santé mentale

Bütikofer, A., Riise, J., & M. Skira, M. (2021). *The impact of paid maternity leave on maternal health.* AEJ: Economic Policy

Stratégie d'identification: Regression Discontinuity Design!!

Question: quelle est la running variable? quel est le seuil?

- Date de naissance de l'enfant
- 1er juillet 1977

Les mères ayant accouché de leur enfant **avant le 1er juillet 1977** ne bénéficient pas de l'introduction du congé maternité rémunéré. Celles ayant accouché **le 1er juillet 1977 ou après** peuvent en bénéficier.

Congé maternité et santé mentale

Bütikofer, A., Riise, J., & M. Skira, M. (2021). *The impact of paid maternity leave on maternal health.* AEJ: Economic Policy

Stratégie d'identification: Regression Discontinuity Design!!

Question: quelle est l'hypothèse d'identification?

Congé maternité et santé mentale

Bütikofer, A., Riise, J., & M. Skira, M. (2021). *The impact of paid maternity leave on maternal health.* AEJ: Economic Policy

Stratégie d'identification: Regression Discontinuity Design!!

Question: quelle est l'hypothèse d'identification?

Continuité de la running variable au point de discontinuité: ici, cela signifie que le nombre de naissances est continu au point de discontinuité

- Cela signifie que les mères ne peuvent pas stratégiquement choisir d'accoucher avant ou après le 1er juillet 1977

Congé maternité et santé mentale

Bütikofer, A., Riise, J., & M. Skira, M. (2021). *The impact of paid maternity leave on maternal health.* AEJ: Economic Policy

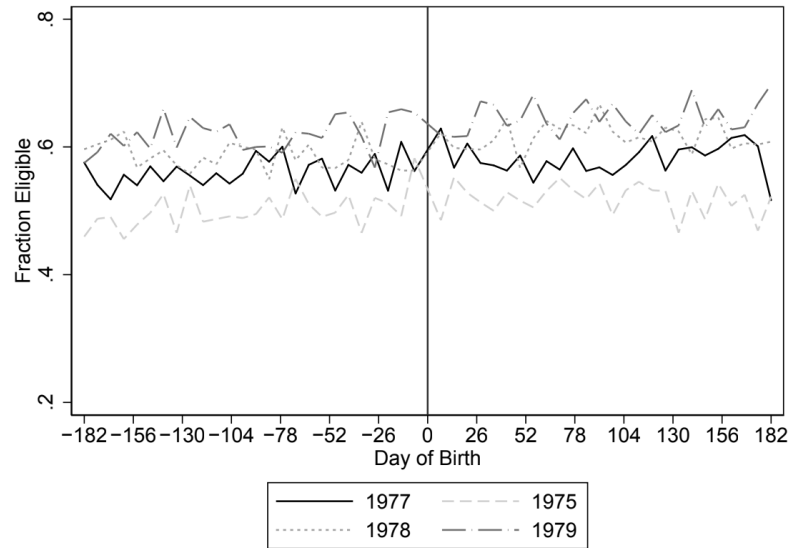


FIGURE 1. PROPORTION OF MOTHERS ELIGIBLE FOR PAID MATERNITY LEAVE

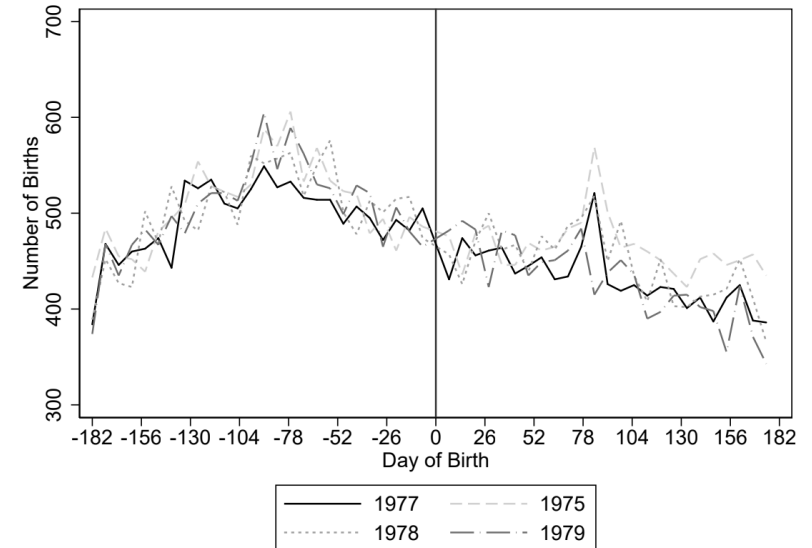
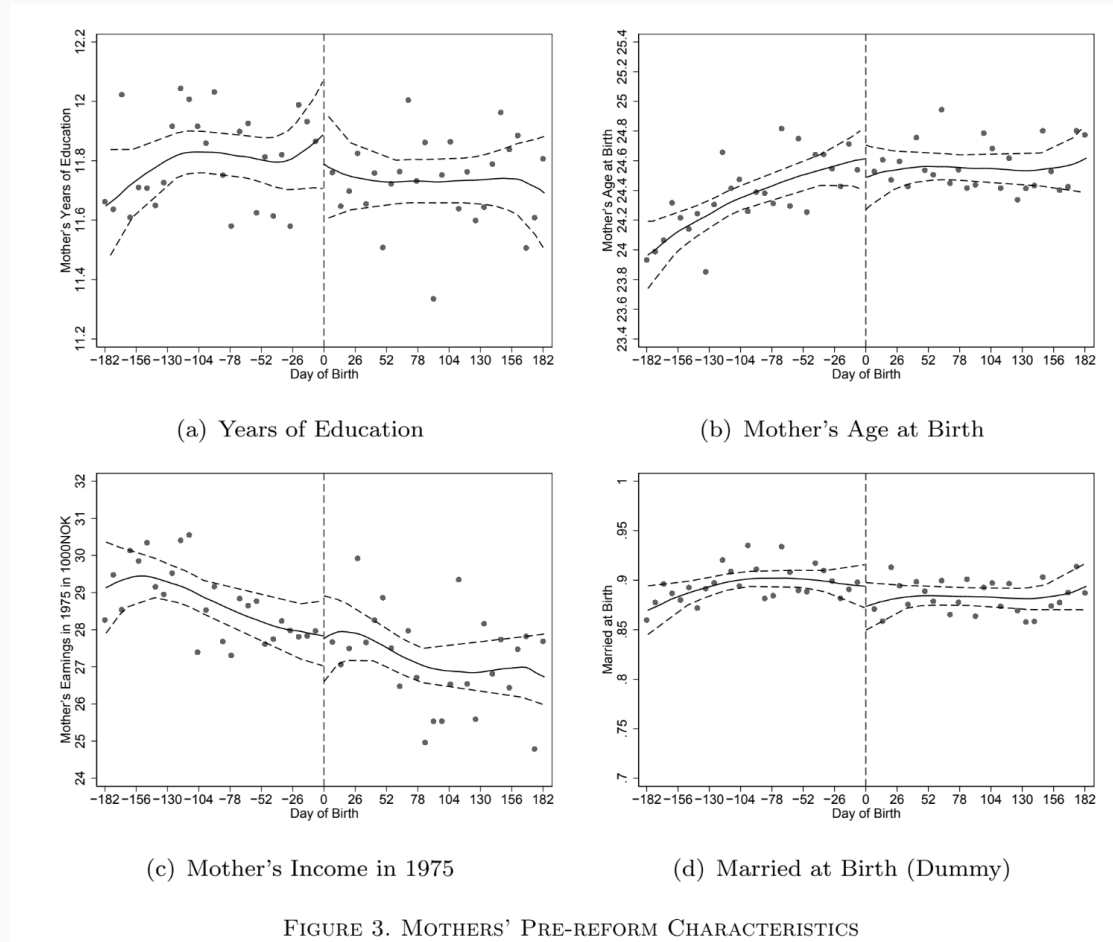


FIGURE 2. NUMBER OF CHILDREN BORN TO ELIGIBLE MOTHERS

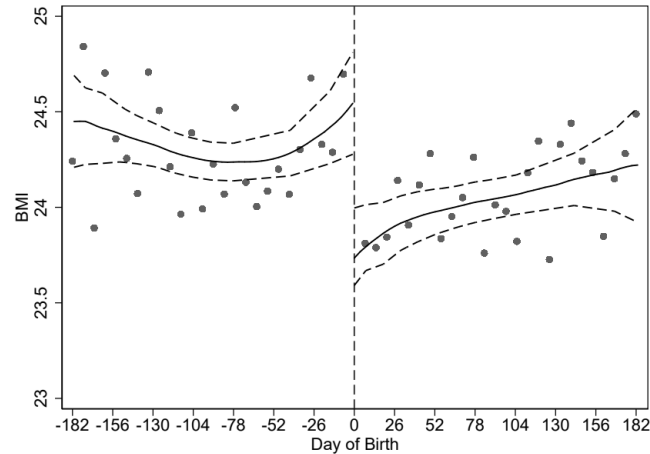
Congé maternité et santé mentale

Bütikofer, A., Riise, J., & M. Skira, M. (2021). *The impact of paid maternity leave on maternal health.* AEJ: Economic Policy

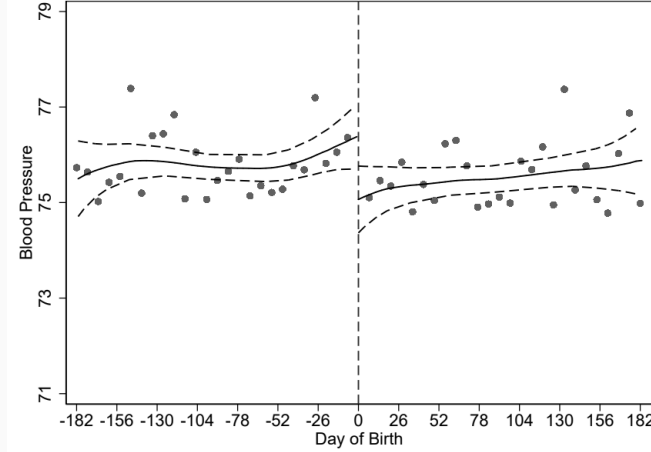


Congé maternité et santé mentale

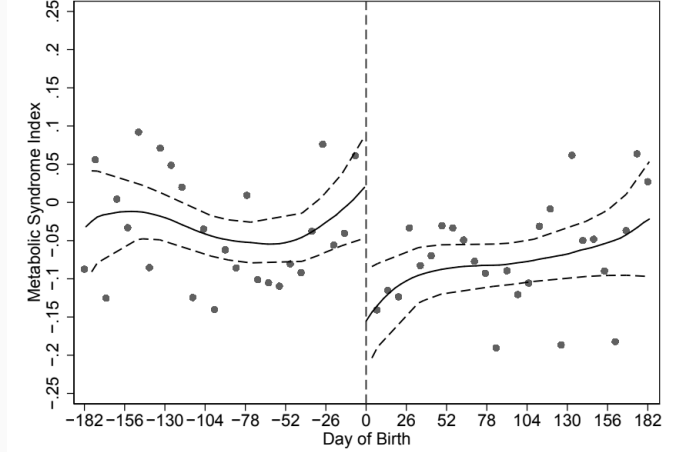
Bütikofer, A., Riise, J., & M. Skira, M. (2021). *The impact of paid maternity leave on maternal health.* AEJ: Economic Policy



(a) Body Mass Index



(d) Blood Pressure (Diastolic)



(g) Metabolic Syndrome Index

Congé maternité, spécialisation et développement des enfants

Canaan, S. (2022) . *Parental leave, household specialization and children's well-being*. Labour Economics.

- Impact de l'allongement de la durée du congé parental en France en 1994: allocation mensuelle pouvant aller jusqu'aux trois ans de l'enfant
 - avant réforme, parents éligibles à partir du 3ème enfant
 - après la réforme, parents éligibles à partir du 2ème enfant
 - $\simeq$ augmentation de la durée du congé parental de 3 ans
- Effets :
 - Négatifs sur l'emploi des mères: éloignées plus longtemps du marché du travail et baisse du salaire
 - Négatifs sur la spécialisation des tâches au sein du ménage: les hommes ne prennent pas ce congé parental et travaillent davantage
 - Négatifs sur le développement verbal des enfants

2.4. Estimation sur R

Estimation sur R

Étape 1: centrer la *running variable*

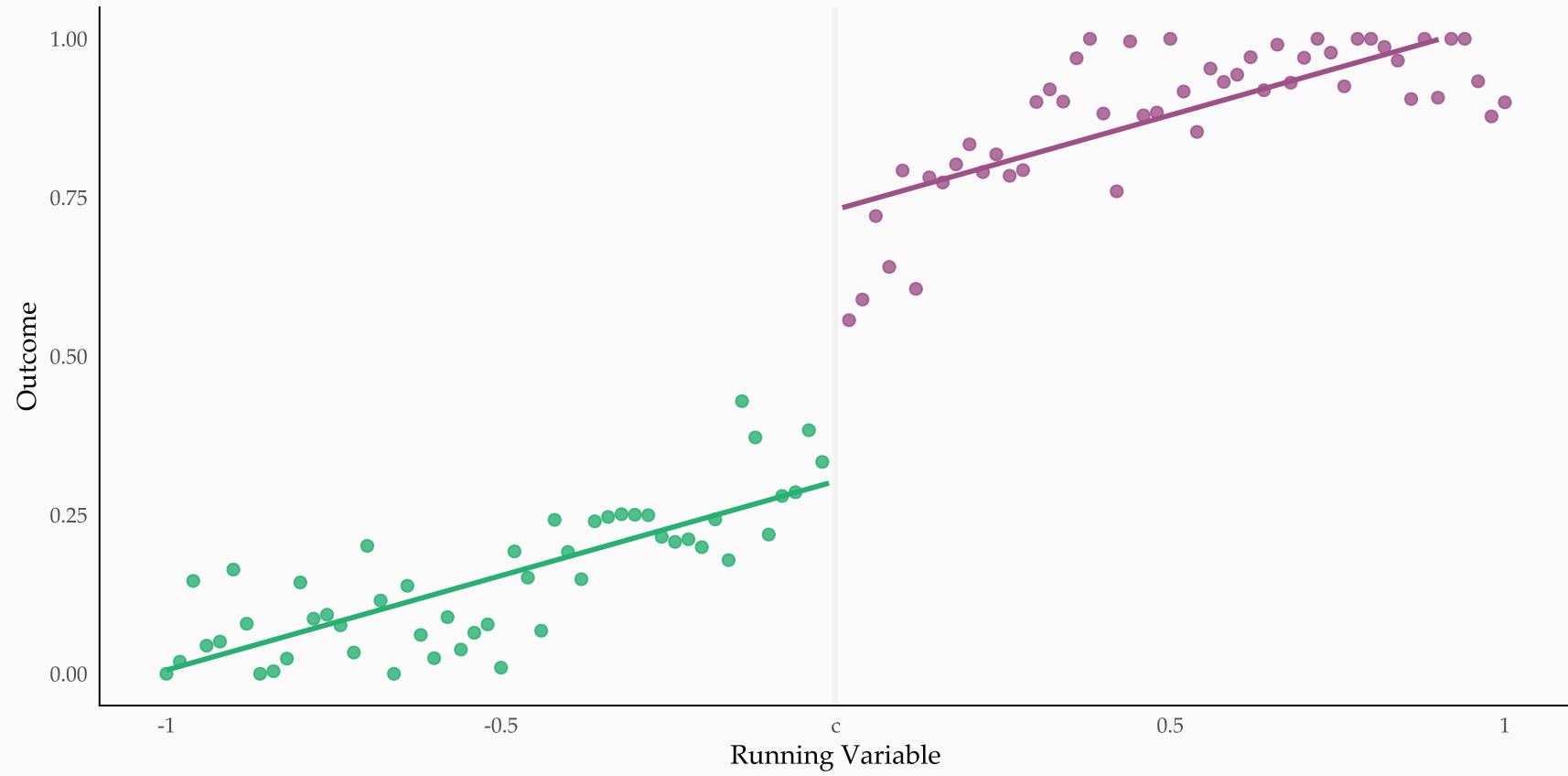
- $\tilde{X}_i = X_i - c$

Étape 2: choisir le modèle à estimer

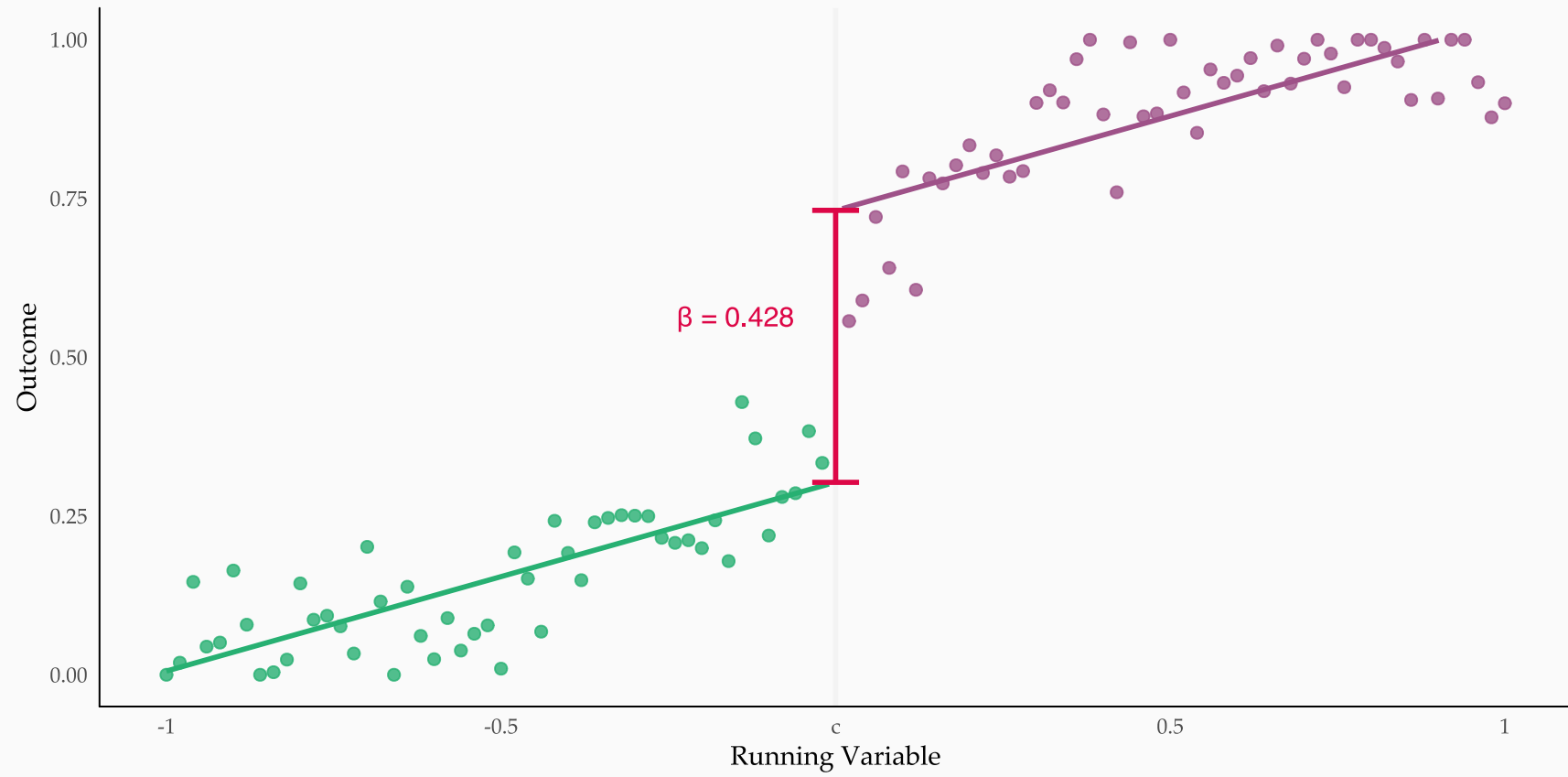
- Linéaire avec pentes communes: $Y_i = \alpha + \beta D_i + \delta \tilde{X}_i + \varepsilon_i$
- Linéaire avec pentes différentes: $Y_i = \alpha + \beta(D_i \times \tilde{X}_i) + \delta \tilde{X}_i + \eta D_i + \varepsilon_i$
- Quadratique: $Y_i = \alpha + \beta D_i + \delta \tilde{X}_i + \lambda \tilde{X}_i^2 + \eta(D_i \times \tilde{X}_i) + \nu(D_i \times \tilde{X}_i)^2 + \varepsilon_i$

2.5. Importance de la forme fonctionnelle

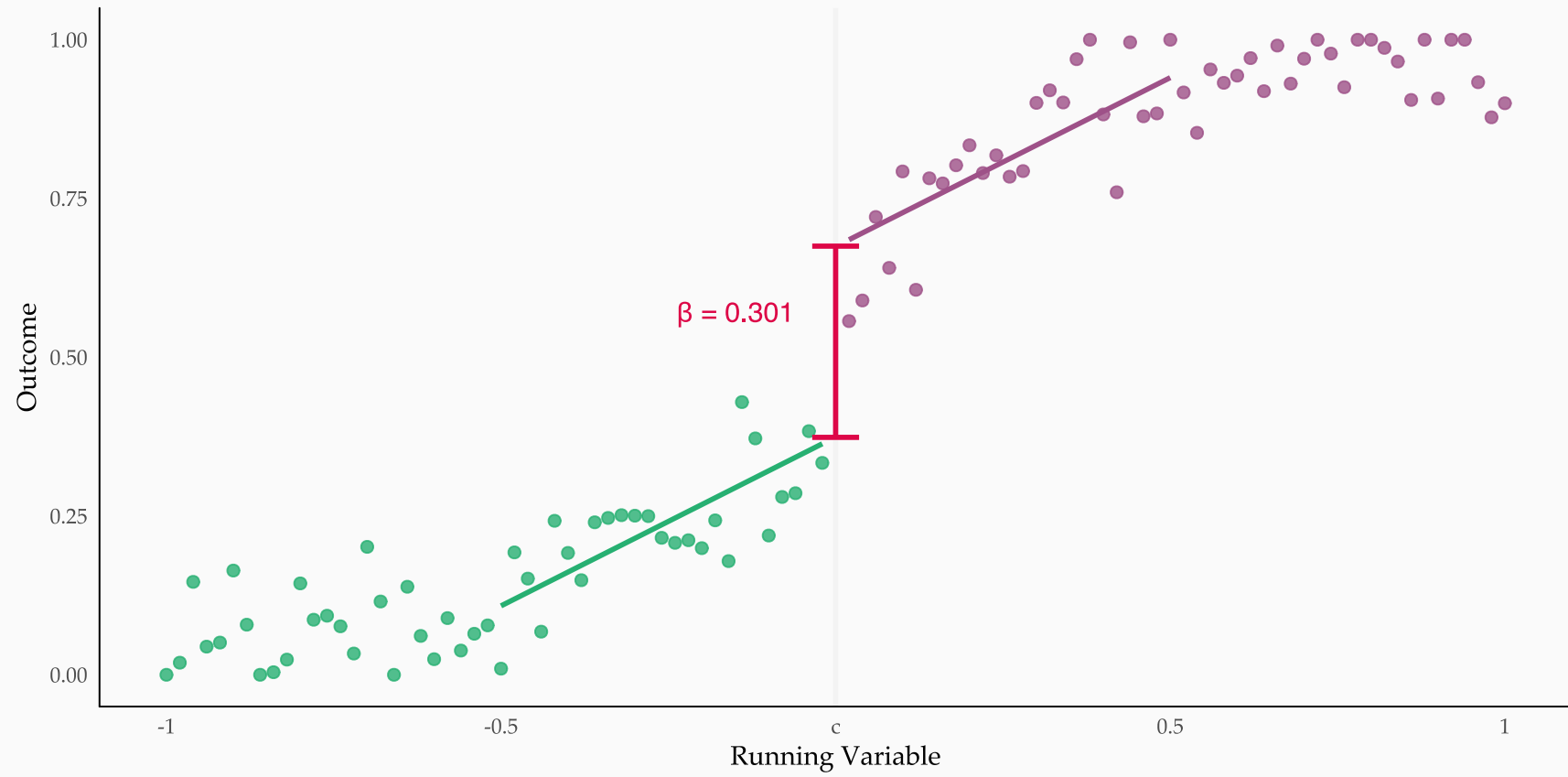
$$Y_i = \alpha + \beta D_i + \delta \tilde{X}_i + \varepsilon_i$$



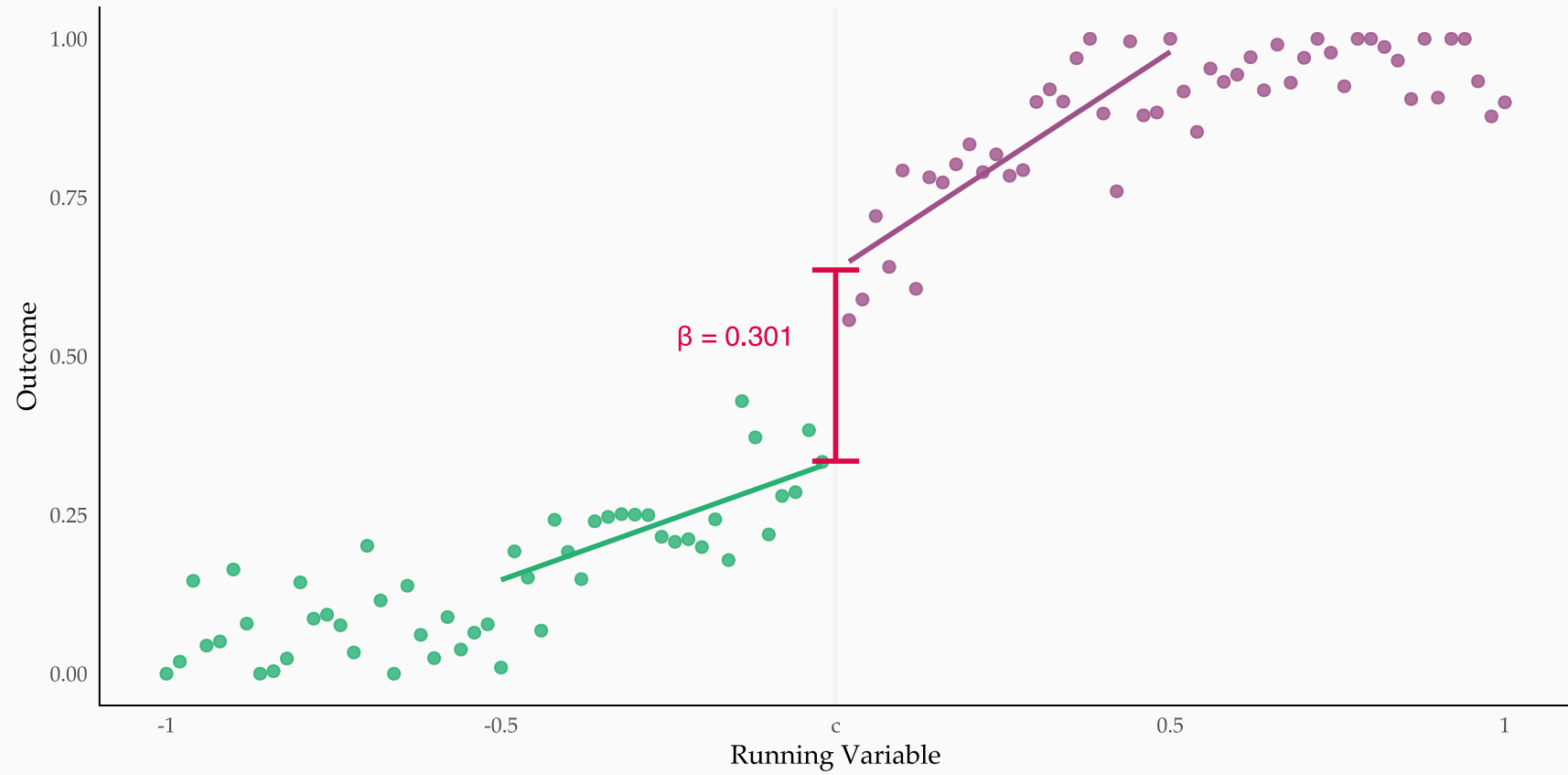
$$Y_i = \alpha + \beta D_i + \delta \tilde{X}_i + \varepsilon_i$$



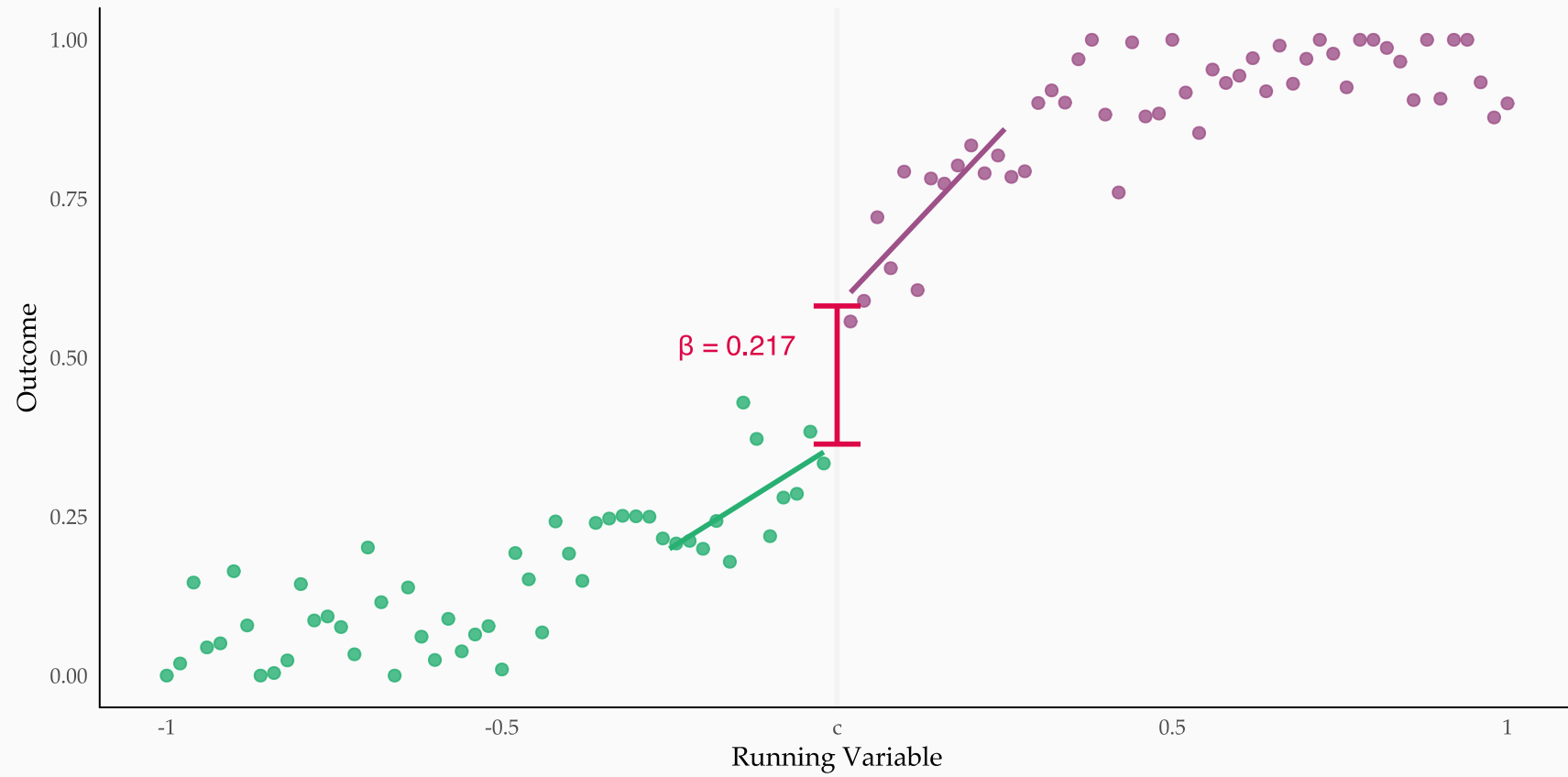
$$Y_i = \alpha + \beta D_i + \delta \tilde{X}_i + \varepsilon_i$$



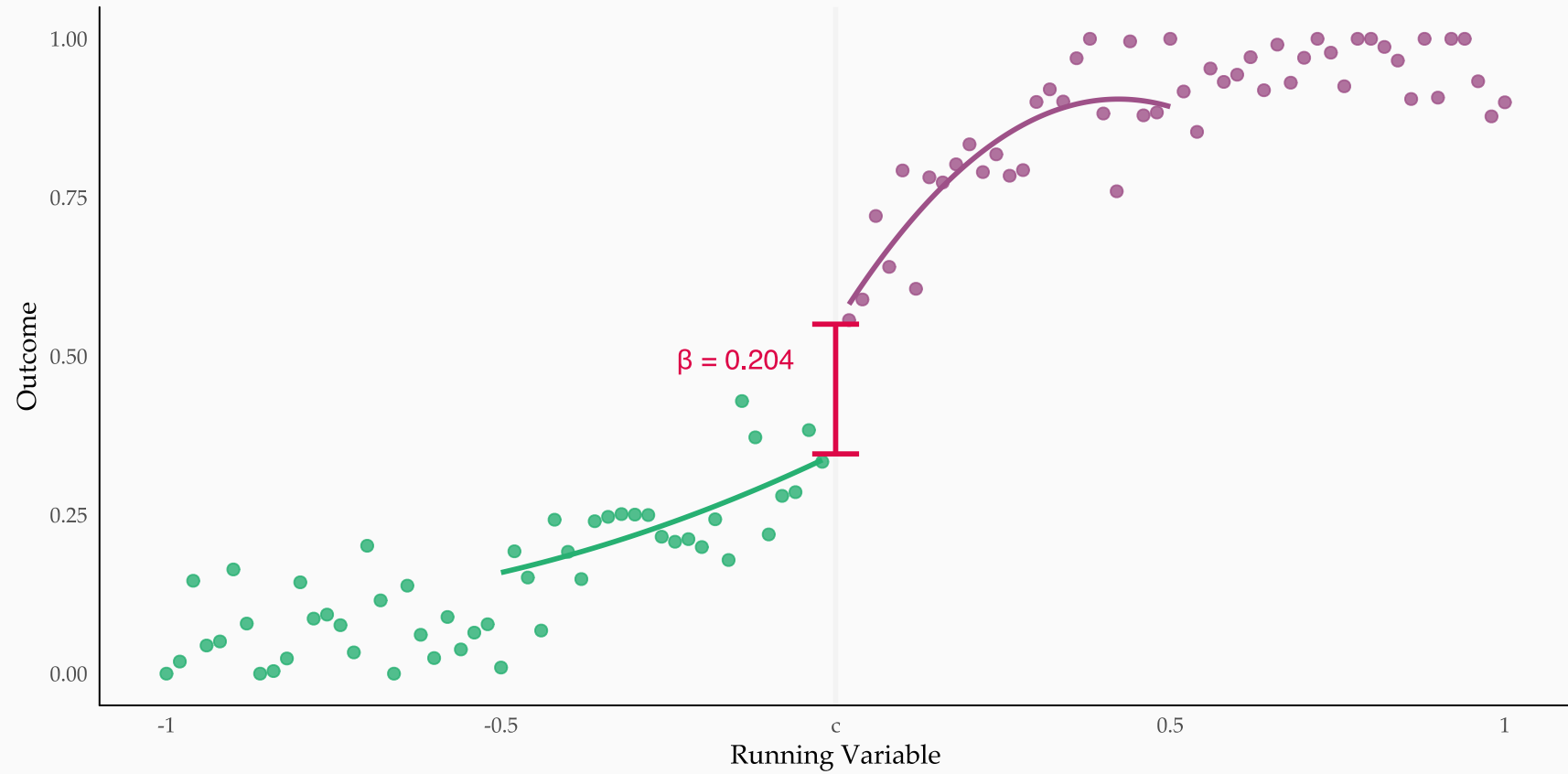
$$Y_i = \alpha + \beta(D_i \times \tilde{X}_i) + \delta\tilde{X}_i + \eta D_i + \varepsilon_i$$



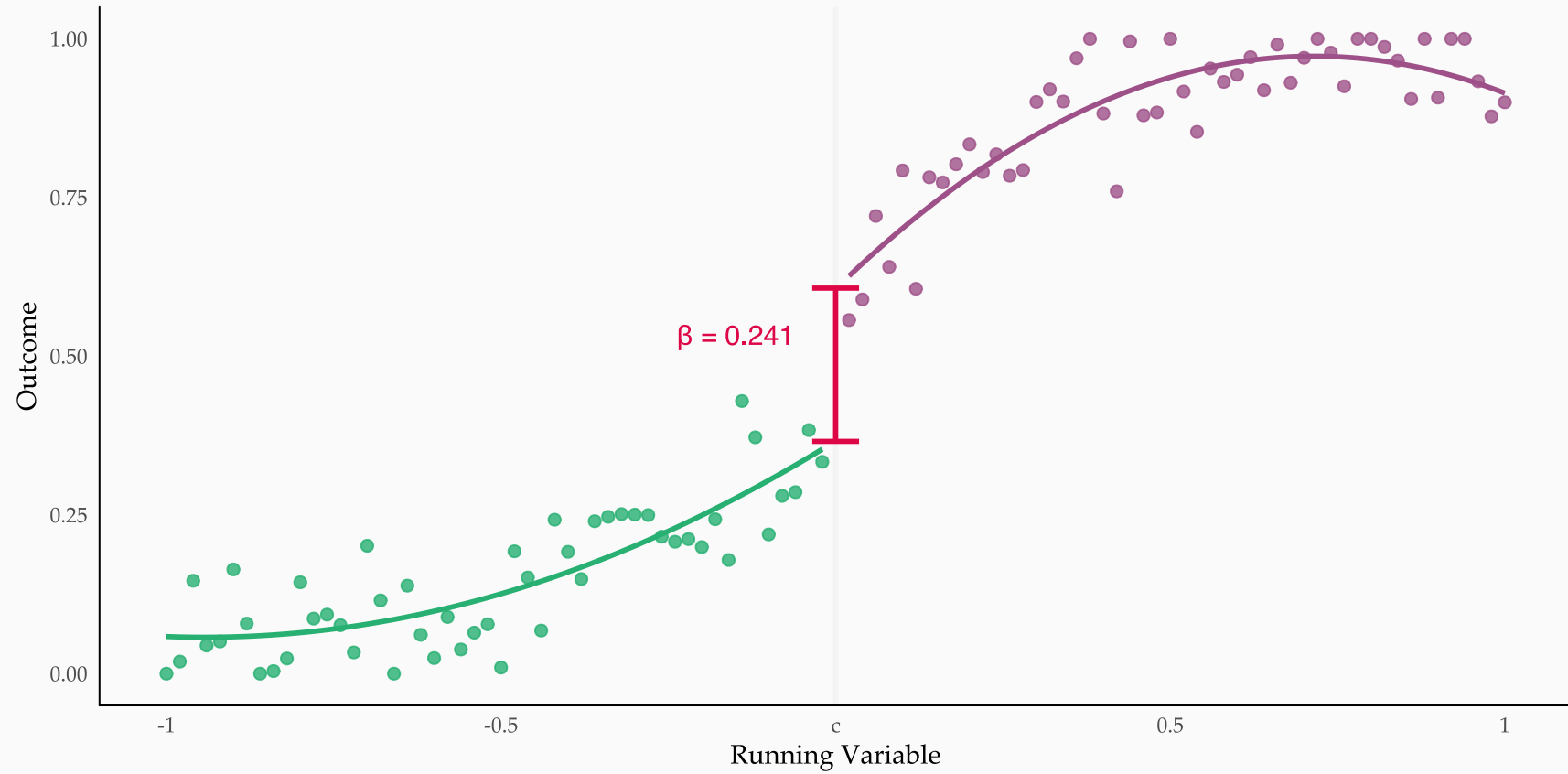
$$Y_i = \alpha + \beta(D_i \times \tilde{X}_i) + \delta\tilde{X}_i + \eta D_i + \varepsilon_i$$



$$Y_i = \alpha + \beta D_i + \delta \tilde{X}_i + \lambda \tilde{X}_i^2 + \eta(D_i \times \tilde{X}_i) + \nu(D_i \times \tilde{X}_i)^2 + \varepsilon_i$$



$$Y_i = \alpha + \beta D_i + \delta \tilde{X}_i + \lambda \tilde{X}_i^2 + \eta(D_i \times \tilde{X}_i) + \nu(D_i \times \tilde{X}_i)^2 + \varepsilon_i$$



Application: Carpenter and Dobkin (2011)

Question de Recherche: quel est l'effet causal de la consommation d'alcool sur la mortalité des jeunes?

Application: Carpenter and Dobkin (2011)

Question de Recherche: quel est l'effet causal de la consommation d'alcool sur la mortalité des jeunes?

Question: pourquoi ne peut-on pas simplement comparer la mortalité des jeunes qui consomment de l'alcool à celle des jeunes qui n'en consomment pas ?

Application: Carpenter and Dobkin (2011)

Question de Recherche: quel est l'effet causal de la consommation d'alcool sur la mortalité des jeunes?

Question: pourquoi ne peut-on pas simplement comparer la mortalité des jeunes qui consomment de l'alcool à celle des jeunes qui n'en consomment pas ?

Biais de sélection/OVB: tout ce qui n'est pas observable et qui impacte à la fois la consommation d'alcool et la mortalité (mode de vie, aversion au risque, environnement social, etc)

Application: Carpenter and Dobkin (2011)

Question de Recherche: quel est l'effet causal de la consommation d'alcool sur la mortalité des jeunes?

Question: pourquoi ne peut-on pas simplement comparer la mortalité des jeunes qui consomment de l'alcool à celle des jeunes qui n'en consomment pas ?

Biais de sélection/OVB: tout ce qui n'est pas observable et qui impacte à la fois la consommation d'alcool et la mortalité (mode de vie, aversion au risque, environnement social, etc)

Carpenter and Dobkin (2011): utilisent l'âge minimum légal à partir duquel un individu est autorisé à consommer de l'alcool (MLDA: Minimum Legal Drinking Age)

Contexte: US

- MLDA: 21 ans

Importer les données:

```
library(masteringmetrics)
data("mla", package = "masteringmetrics")
```

Application: Carpenter and Dobkin (2011)

```
## # A tibble: 6 × 19
##   agecell    all allfitted internal internalfitted external externalfitted
##   <dbl> <dbl>      <dbl>      <dbl>          <dbl>      <dbl>          <dbl>
## 1    19.1  92.8        91.7        16.6          16.7        76.2          75.0
## 2    19.2  95.1        91.9        18.3          16.9        76.8          75.0
## 3    19.2  92.1        92.0        18.9          17.1        73.2          75.0
## 4    19.3  88.4        92.2        16.1          17.3        72.3          74.9
## 5    19.4  88.7        92.3        17.4          17.4        71.3          74.9
## 6    19.5  90.2        92.5        17.9          17.6        72.3          74.9
## # i 12 more variables: alcohol <dbl>, alcoholfitted <dbl>, homicide <dbl>,
## #   homicidefitted <dbl>, suicide <dbl>, suicidefitted <dbl>, mva <dbl>,
## #   mvafitted <dbl>, drugs <dbl>, drugsfitted <dbl>, externalother <dbl>,
## #   externalotherfitted <dbl>
```

Application: Carpenter and Dobkin (2011)

- 1) Quelle est la running variable? Quel est le seuil? Quelle est l'hypothèse d'identification?
- 2) Construisez la variable `above21` qui vaut 1 pour toutes les classes d'âge ≥ 21
- 3) Représenter graphiquement l'évolution **linéaire et quadratique** de la mortalité liée aux accidents de la route (variable `mva`) et de la mortalité globale (`all`) autour du seuil de discontinuité. *Hint: utilisez les commandes `geom_smooth()` de `ggplot` pour tracer des droites de régressions sur un graphique et `poly(x, 2)` pour un polynôme de degré 2 de la variable x .*
- 4) **Pour le prochain cours:** Estimez l'effet d'atteindre le MLDA sur la mortalité globale (linéaire avec pente différente, et quadratique), en utilisant le package `rdrobust`

Solution : Carpenter and Dobkin (2011)

1) Quelle est la **running variable**? Quel est le seuil? Quelle est l'hypothèse d'identification?

Solution : Carpenter and Dobkin (2011)

1) Quelle est la **running variable**? Quel est le seuil? Quelle est l'hypothèse d'identification?

Running Variable: Age

Solution : Carpenter and Dobkin (2011)

1) Quelle est la running variable? Quel est le **seuil**? Quelle est l'hypothèse d'identification?

Running Variable: Age

Solution : Carpenter and Dobkin (2011)

1) Quelle est la running variable? Quel est le **seuil**? Quelle est l'hypothèse d'identification?

Running Variable: Age

Seuil: 21 ans

Solution : Carpenter and Dobkin (2011)

1) Quelle est la running variable? Quel est le seuil? Quelle est l'**hypothèse d'identification**?

Running Variable: Age

Seuil: 21 ans

Solution : Carpenter and Dobkin (2011)

1) Quelle est la running variable? Quel est le seuil? Quelle est l'**hypothèse d'identification**?

Running Variable: Age

Seuil: 21 ans

Hypothèse d'identification: Autour du seuil, l'allocation du traitement est *as good as random*, i.e. les individus autour du seuil de leur 21ème anniversaire, ne diffèrent en moyenne que de par leur accès ou non à l'alcool

Solution : Carpenter and Dobkin (2011)

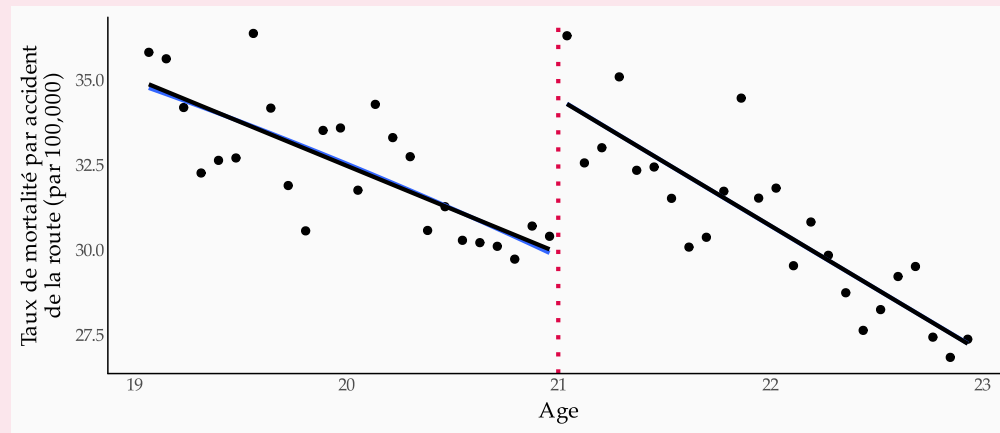
2) Construisez la variable `above21` qui vaut 1 pour toutes les classes d'âge ≥ 21

```
mlda = mlda %>%  
  mutate(above21 = ifelse(agecell  $\geq$  21, 1, 0))
```

Solution : Carpenter and Dobkin (2011)

3) Représenter graphiquement l'évolution **linéaire et quadratique** de la mortalité liée aux **accidents de la route** (variable `mva`) et de la mortalité globale (`all`) autour du seuil de discontinuité

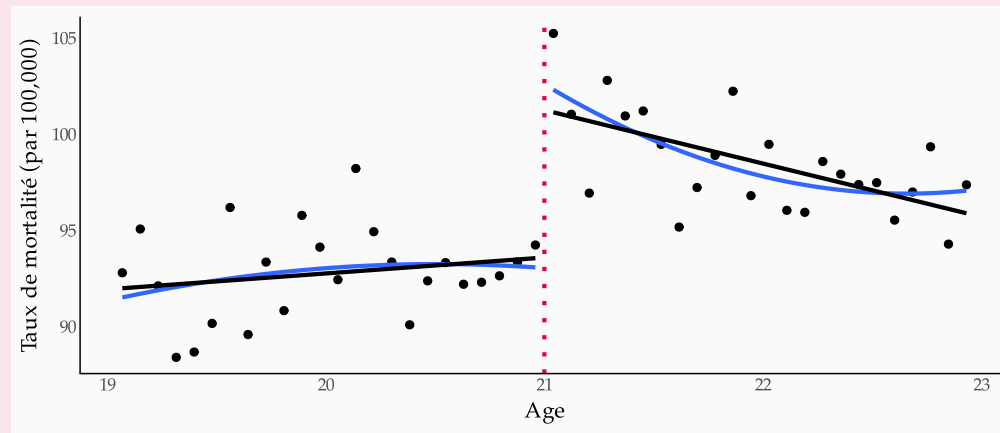
```
mlda %>%  
  ggplot(aes(x = agecell, y = mva)) +  
  geom_point() +  
  geom_smooth(aes(group = above21), se = FALSE, method = "lm", formula = y ~ poly(x, 2)) +  
  geom_smooth(aes(group = above21), se = FALSE, method = "lm", formula = y ~ x, color = "black") +  
  geom_vline(xintercept = 21, color = "#dd0747", linetype = "dotted", linewidth = 1) +  
  labs(y = "Taux de mortalité par accident\nde la route (par 100,000)", x = "Age") +  
  theme_minimum
```



Solution : Carpenter and Dobkin (2011)

3) Représenter graphiquement l'évolution **linéaire et quadratique** de la mortalité liée aux accidents de la route (variable `mva`) et de la **mortalité globale** (`all`) autour du seuil de discontinuité

```
mlda %>%
  ggplot(aes(x = agecell, y = all)) +
  geom_point() +
  geom_smooth(aes(group = above21), se = FALSE, method = "lm", formula = y ~ poly(x, 2)) +
  geom_smooth(aes(group = above21), se = FALSE, method = "lm", formula = y ~ x, color = "black") +
  geom_vline(xintercept = 21, color = "#dd0747", linetype = "dotted", linewidth = 1) +
  labs(y = "Taux de mortalité (par 100,000)", x = "Age") +
  theme_minimum
```



Solution : Carpenter and Dobkin (2011)

4) Estimez l'effet d'atteindre le MLDA sur la mortalité globale (linéaire avec pente différente et quadratique), en utilisant le package `rdrobust`

Solution : Carpenter and Dobkin (2011)

4) Estimez l'effet d'atteindre le MLDA sur la mortalité globale (**linéaire avec pente différente** et quadratique), en utilisant le package `rdrobust`

```
library(rdrobust)

# Linéaire (pentes différentes de chaque côté)
rdd1 = rdrobust(y = mllda$mva, # outcome
               x = mllda$agecell, # running variable
               c = 21, # seuil
               p = 1, # degré du polynôme
               bwselect = "mserd") # bandwidth, ici optimal

# Quadratique
rdd2 = rdrobust(y = mllda$mva,
               x = mllda$agecell,
               c = 21,
               p = 2,
               bwselect = "mserd")
```

Solution : Carpenter and Dobkin (2011)

4) Estimez l'effet d'atteindre le MLDA sur la mortalité globale (linéaire avec pente différente et **quadratique**), en utilisant le package `rdrobust`

```
library(rdrobust)

# Linéaire (pentes différentes de chaque côté)
rdd1 = rdrobust(y = mllda$all, # outcome
               x = mllda$agecell, # running variable
               c = 21, # seuil
               p = 1, # degré du polynôme
               bwselect = "mserd") # bandwidth, ici optimal

# Quadratique
rdd2 = rdrobust(y = mllda$all,
               x = mllda$agecell,
               c = 21,
               p = 2,
               bwselect = "mserd")
```

Solution : Carpenter and Dobkin (2011)

4) Estimez l'effet d'atteindre le MLDA sur la mortalité globale (linéaire avec pente différente et **quadratique**), en utilisant le package `rdrobust`

Modèle	Estimateur	Std. Error	p-value	Bandwidth
Linéaire	9.595	3.591	0.008	0.493
Quadratique	10.280	4.777	0.031	0.698

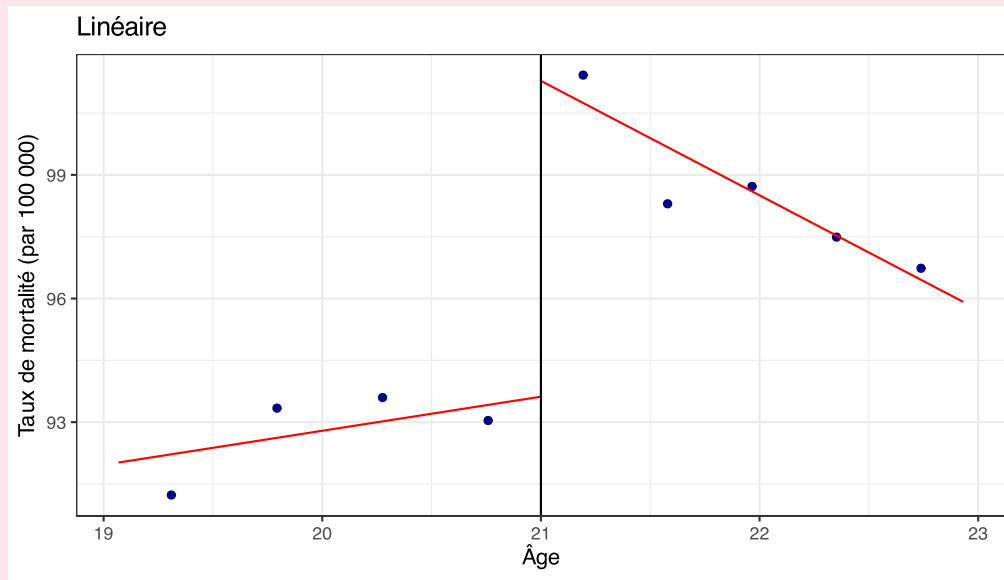
Interprétation : $\hat{\beta}$ mesure le saut du taux de mortalité au seuil de 21 ans — i.e. l'effet causal de l'accès légal à l'alcool

- Linéaire ($\hat{\beta} = \mathbf{9.60}$): atteindre l'âge légal de consommation d'alcool entraîne une hausse d'environ 9.6 décès supplémentaires pour 100 000
- Quadratique ($\hat{\beta} = \mathbf{10.28}$): résultat similaire, une hausse d'environ 10.3 décès pour 100 000

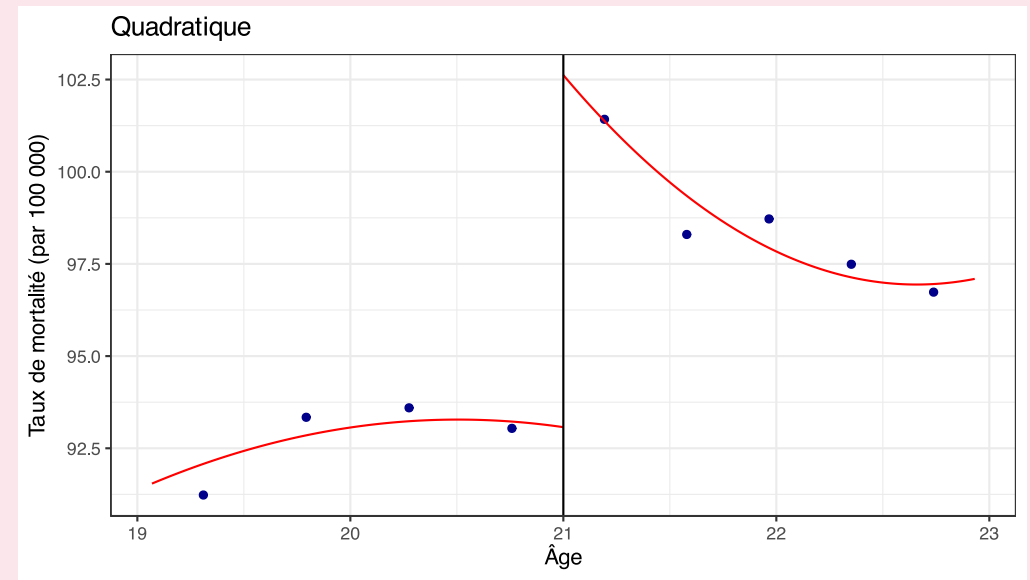
Solution : Carpenter and Dobkin (2011)

4) Estimez l'effet d'atteindre le MLDA sur la mortalité globale (linéaire avec pente différente et **quadratique**), en utilisant le package `rdrobust`

Linéaire ($p = 1$)



Quadratique ($p = 2$)



NB : `rdplot` agrège les observations en bins (moyennes locales) dont le nombre est choisi de manière optimale, donc les points ne représentent pas des observations individuelles mais des moyennes par intervalle d'âge

3. Fuzzy RDD

Définition

Les **fuzzy RDD** exploitent un changement discontinu dans l'assignation au traitement D_i au seuil de c .

Contrairement au *sharp* RDD où la probabilité d'être traité passe de 0 à 1 lorsque X_i passe le seuil de discontinuité c , dans un cas de *fuzzy* RDD, la probabilité de traitement est continue au voisinage de c :

$$0 < \lim_{x \rightarrow c^+} \mathbb{P}(D_i = 1 | X_i = x) - \lim_{x \rightarrow c^-} \mathbb{P}(D_i = 1 | X_i = x) < 1$$

On a maintenant deux effets lorsque X_i franchit le seuil c :

1. L'effet sur l'**outcome**: $\lim_{x \rightarrow c^+} \mathbb{E}(Y_i | X_i = x) - \lim_{x \rightarrow c^-} \mathbb{E}(Y_i | X_i = x)$
2. L'effet sur la **probabilité de traitement**: $\lim_{x \rightarrow c^+} \mathbb{E}(D_i | X_i = x) - \lim_{x \rightarrow c^-} \mathbb{E}(D_i | X_i = x)$

L'estimateur de l'effet du traitement dans le cas d'une *fuzzy* RDD est donc:

$$\beta_{\text{RDD}}^{\text{fuzzy}} = \frac{\lim_{x \rightarrow c^+} \mathbb{E}(Y_i | X_i = x) - \lim_{x \rightarrow c^-} \mathbb{E}(Y_i | X_i = x)}{\lim_{x \rightarrow c^+} \mathbb{E}(D_i | X_i = x) - \lim_{x \rightarrow c^-} \mathbb{E}(D_i | X_i = x)}$$

Lien avec...

La formule de l'estimateur nous rappelle...

Lien avec...

La formule de l'estimateur nous rappelle...

Celle de l'estimateur IV! Ici, l'instrument $Z_i = 1\{X_i \geq c\}$!!

Lien avec...

La formule de l'estimateur nous rappelle...

Celle de l'estimateur IV! Ici, l'instrument $Z_i = 1\{X_i \geq c\}$!!

Angrist, J. and Lavy, V. (1999). *Using Maimonides' rule to estimate the effect of class size on scholastic achievement.* QJE

- Règle de Maïmonide: en Israël, règle stipulant qu'au-delà de 40 élèves par classe, l'école doit scinder la classe en deux
⇒ crée une discontinuité dans la taille des classes dès que la taille des classes dépasse 40 élèves, 80 élèves, 120 élèves.
- En pratique : la règle n'est pas toujours appliquée de manière stricte (certaines écoles de petite taille n'ont pas les moyens matériels de scinder la classe, d'autres peuvent décider de scinder avant d'atteindre 40 élèves) ⇒ le fait d'être au-dessus du seuil de 40 élèves augmente fortement la probabilité que la classe soit effectivement scindée **mais ne passe pas de 0 à 1**

Recap: Regression Discontinuity Design (RDD)

Data: Données observationnelles

Hypothèse d'identification: Continuité des outcomes potentiels au seuil

- Formellement: $\mathbb{E}(Y_{1i}|X_i = x)$ et $\mathbb{E}(Y_{0i}|X_i = x)$ sont continues en x
- Pas de manipulation de la running variable, et les individus comparables autour du seuil

Modèle linéaire à pentes communes: pour tout individu i ,

$$Y_i = \alpha + \beta D_i + \delta \tilde{X}_i + \varepsilon_i \quad \text{où} \quad D_i = 1\{X_i \geq c\}$$

Estimateur de l'effet du traitement: $\hat{\beta}_{\text{RDD}}^{\text{sharp}} = \lim_{x \rightarrow c^+} \mathbb{E}(Y_i|X_i = x) - \lim_{x \rightarrow c^-} \mathbb{E}(Y_i|X_i = x)$

Implémentation sur R:

- Visualisation d' l'outcome autour du seuil avec `ggplot`
- Tests de manipulation: `DCdensity()` du package `rdd` (test de McCrary)
- Estimation: `lm(y ~ D + X + I(D*X), data = data)` ou packages `rdrobust`, `rddensity`

Sources

[Causal inference: The Mixtape, Scott Cunningham](#)

[Mixtape Sessions: RDD, Scott Cunningham](#)

[Mastering Metrics: RDD](#)

[RDD, Edward Rubin](#)