

Evaluation de projets de développement

La régression par discontinuité

Björn Nilsson

`bjorn.nilsson@universite-paris-saclay.fr`

Faculté Jean Monnet
Université Paris-Saclay

M2 GPDS, 1er semestre
Année universitaire 2024-2025

université
PARIS-SACLAY

Présentation à faire pour janvier

- Vous aurez à faire une présentation en binôme pour le mois de janvier.
 - Il s'agira de présenter de manière critique un article de recherche.
 - L'article utilisera une méthode d'évaluation d'impact pour répondre à une question liée à la lutte contre la pauvreté.
- 20 minutes (+ 10 minutes de discussion)
- Ce qui est attendu:
 - Présentation de la motivation, de la question de recherche qui est posée, et de la stratégie mise en oeuvre pour y répondre.
 - Veillez particulièrement à identifier les hypothèses derrière le résultat obtenu et de discuter leur crédibilité.
 - Discutez les résultats (portée et limites)

Plan du cours

- Séance 1 : Introduction
- Séance 2 : Mesurer la pauvreté
- Séance 3 : Lire des travaux d'évaluation d'impact
- Séance 4 : La différence de différence
- **Séance 5 : La régression par discontinuité**
- Séance 6 : Les variables instrumentales
- Séance 7 : Les essais aléatoires randomisés
- Séance 8 : Présentations d'articles
- Séance 9 : Présentations d'articles
- Séance 10 : Examen

Plan du chapitre

1. Avant-propos
2. Introduction RDD
3. Hypothèses et estimation
4. Exemples

C'est quoi la régression par discontinuité ?

La régression par discontinuité

- La régression par discontinuité (ou **RDD**) est une méthode d'évaluation quasi-expérimentale.
- Elle exploite un seuil d'éligibilité pour accéder à un traitement.
- En faisant l'hypothèse que les observations proches du seuil peuvent être considérées comme similaires.
- En comparant les traités et les non-traités proches du seuil, on peut alors estimer l'effet moyen du traitement sur la population proche du seuil.
 - C'est ce qu'on appelle un effet **LATE**—Local average treatment effect.

C'est quoi la régression par discontinuité ?

La régression par discontinuité

- La régression par discontinuité est donc définie en termes d'un *score*, d'un seuil et d'un traitement !
- Le **score** et le **seuil** reflète le traitement qu'un individu se voit accorder.
 - C'est le mécanisme d'assignation (comparer avec tirage au sort pour les RCT).
 - Le seuil agit comme un changement abrupt d'assignation qui nous permet—sous des hypothèses que nous clarifierons par la suite—d'apprendre des choses sur l'effet du programme.
- Parfois le seuil détermine parfaitement l'accès au traitement—on parle de **sharp design**.
- Parfois il influence la probabilité d'être traité, mais on observe quand même des traités et des non-traités des deux côtés du seuil (**fuzzy design**).

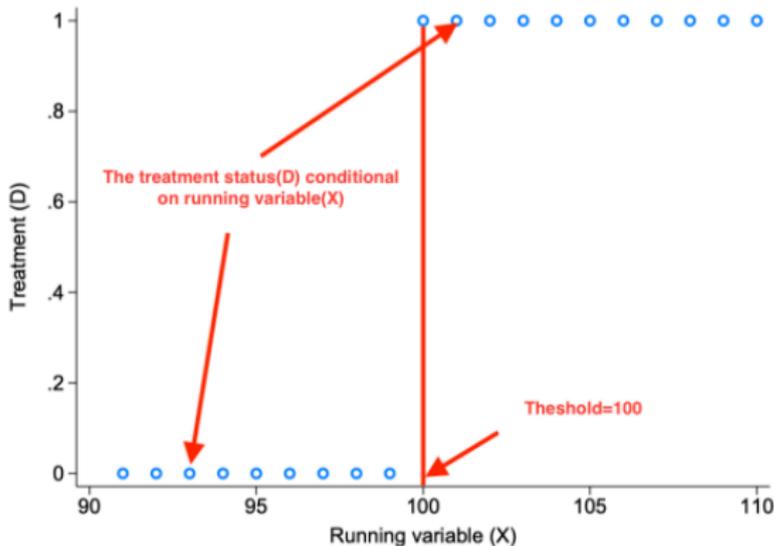
C'est quoi la régression par discontinuité ?

La régression par discontinuité

- L'identification d'un effet provient du fait qu'on compare deux populations proches.
 - Autrement dit, plus on compare des personnes à la limite du seuil, plus on est proche du vrai effet de traverser le seuil.
- L'idéal est $\text{Effet} = \lim_{x \rightarrow c} E[Y_i | X_i = x] - \lim_{x \leftarrow c} E[Y_i | X_i = x]$
- Qui est approximativement égal à: $E[Y_{1,i} | X_i = c] - E[Y_{0,i} | X_i = c]$
- Approximativement, car pour une valeur donnée, on est soit traitée, soit non-traitée, et jamais les deux à la fois (la méthode repose sur l'extrapolation).

C'est quoi la régression par discontinuité ?

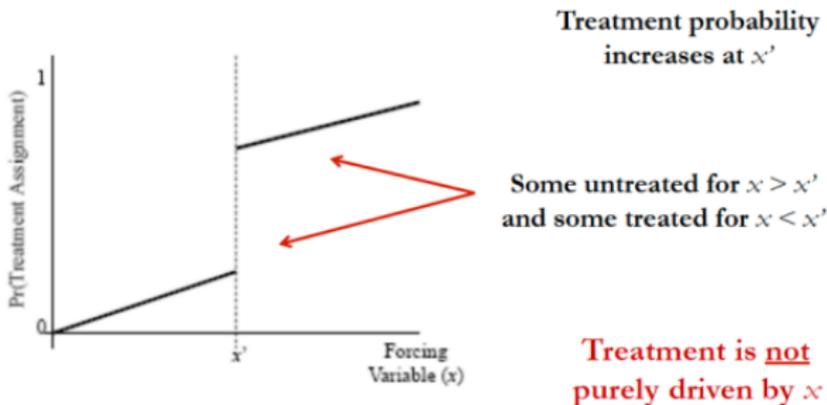
Exemple de sharp design



Source: Qu (2021)

C'est quoi la régression par discontinuité ?

Exemple de fuzzy design



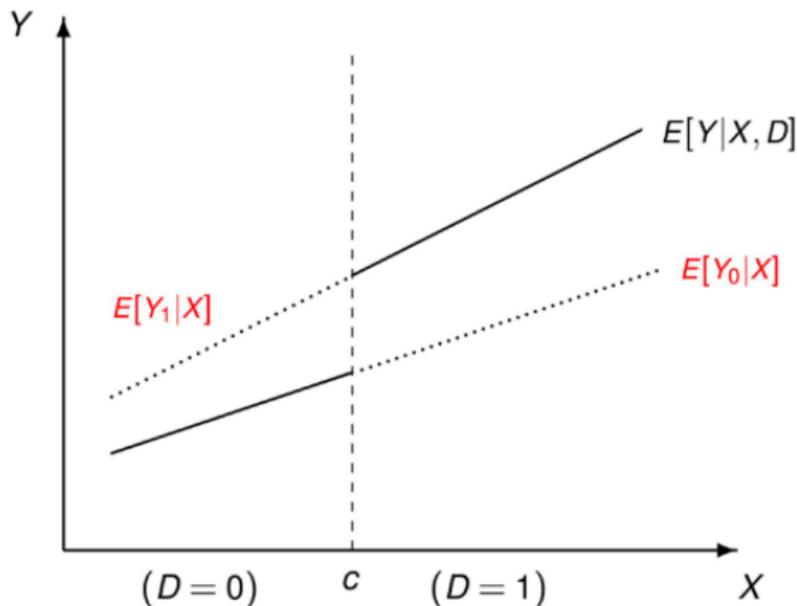
Source: Qu (2021)

Hypothèses de la méthode: sharp design

- **Continuité dans les autres variables pertinentes autour du seuil.**
 - Nécessaire pour pouvoir attribuer l'effet au traitement.
 - Ex: si il y a un saut dans une autre variable non-liée au traitement en même temps que la variable d'assignation, il est difficile d'identifier un effet.
- **Pas de manipulation avec précision**
 - Les individus ne doivent pas pouvoir manipuler le score pour être juste au-dessus ou juste en dessous du traitement.
 - Ex: cas des examens; il y a de l'incertitude sur la notation et de l'incertitude sur les capacités qui font qu'un individu ne peut pas déterminer à l'avance son score de manière précise.
 - Parfois visible dans les données : plein d'observations juste au dessus du seuil.
 - Test de McCrary (2008): on y reviendra.

Hypothèses de la méthode: sharp design

L'intuition



Source: Qu (2021)

Hypothèses de la méthode: sharp design

Comment estimer un effet par RDD ?

- $Y_i = \alpha + \beta D_i + \gamma(X_i - c) + \epsilon_i$
- β donne alors l'effet du traitement.

Rappel : L'effet estimé par RDD est égal à $E[Y_{1,i}|X_i = c] - E[Y_{0,i}|X_i = c]$.

- $E[Y_{1,i}|X_i = c] = \alpha + \beta + \epsilon_i$
- $E[Y_{0,i}|X_i = c] = \alpha + \epsilon_i$
- $E[Y_{1,i}|X_i = c] - E[Y_{0,i}|X_i = c] = \beta$.

Hypothèses de la méthode: sharp design

Choix d'une fenêtre d'estimation

- En RDD, on compare des observations proches du seuil, arguant que pour celles-ci, l'attribution au traitement est quasi-aléatoire.
- Plus on reste proche du seuil, plus cela risque d'être vrai.
- En même temps, plus on est proche du seuil, moins on a d'observations à utiliser \Rightarrow perte de puissance statistique.
- On est face à un arbitrage entre **biais** et **puissance**.

Hypothèses de la méthode: sharp design

Choix d'une fenêtre d'estimation

- Comment choisir alors la taille de la fenêtre qu'on utilise pour l'estimation ?
- On peut essayer de chercher une fenêtre **optimale**, basée sur un critère [erreur quadratique moyenne]
- Ou on peut essayer plusieurs fenêtres et garder celle qui caractérise le mieux la relation entre X et Y.
- En pratique, on présente souvent les résultats pour plusieurs choix de fenêtres afin de montrer leur **robustesse**.

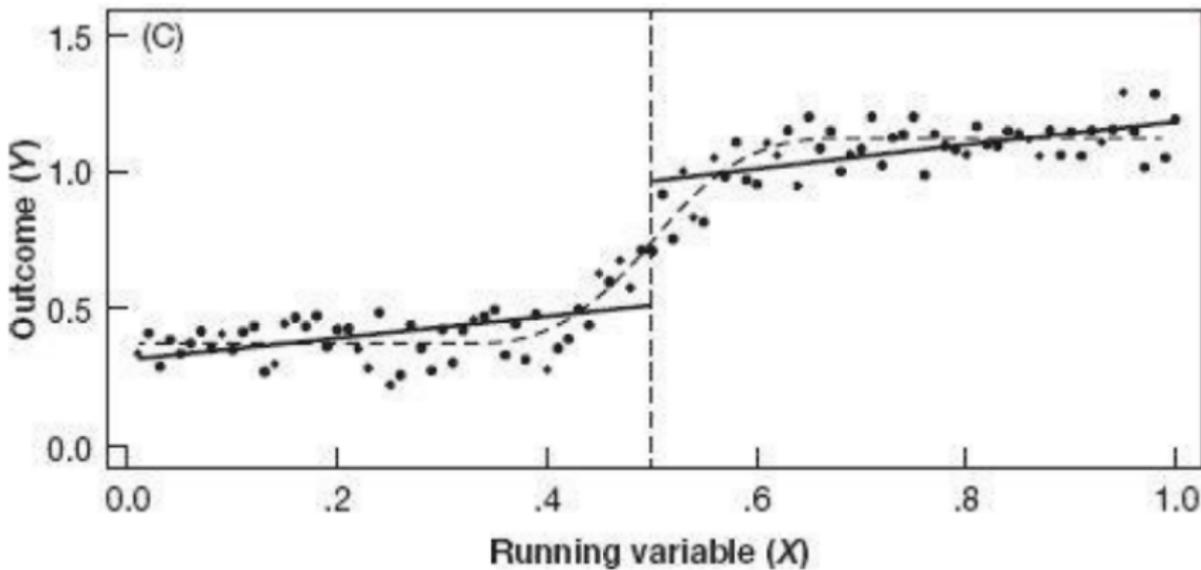
Hypothèses de la méthode: sharp design

Choix d'une forme fonctionnelle pour la relation entre X et l'outcome

- En cas de sharp RDD, toutes les observations traitées ont des valeurs de $X > c$, tandis que toutes les non-traitées ont des $X < c$.
- Cela veut dire que l'estimation repose sur l'**extrapolation**.
- Autrement dit, la forme fonctionnelle utilisée pour caractériser cette relation devient très importante.
 - Comment choisir cette forme fonctionnelle ?
 - Les graphiques sont très importants en RDD...

Hypothèses de la méthode: sharp design

La forme fonctionnelle est-elle bonne ?



Source: Qu (2021)

Hypothèses de la méthode: sharp design

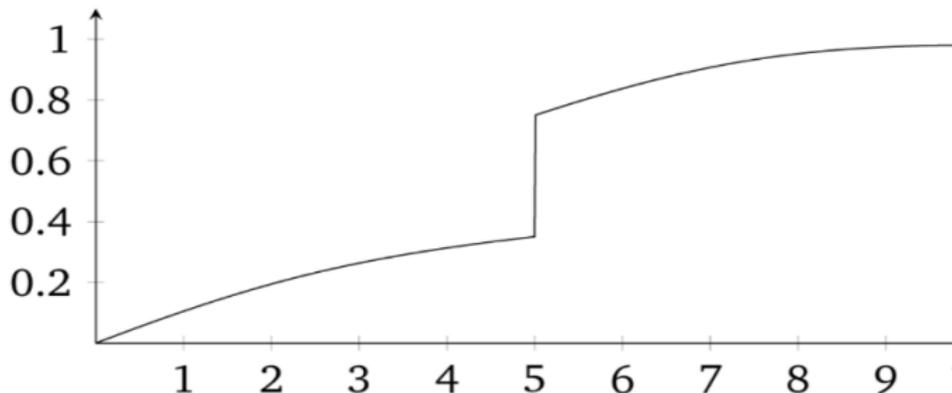
- Sur le graphique précédent, la forme fonctionnelle supposée est linéaire.
 - Conduit à la conclusion d'un effet positif du traitement.
- Or, on soupçonne qu'une autre forme fonctionnelle caractérise mieux la distribution des données.
 - Celle-ci ne fait pas apparaître de saut.
- **Recommandation:** utiliser un polynôme pour caractériser la relation entre X et Y.
 - Gelman & Imbens (2019): une forme quadratique est mieux que des polynômes d'ordre supérieur.
- $Y_i = \alpha + \beta D_i + \gamma_1(X_i - c) + \gamma_2(X_i - c)^2 + \epsilon_i$

Hypothèses de la méthode: fuzzy design

- Le fuzzy design reprend les hypothèses de la sharp design.
- Il y a des observation traitées et non-traitées des deux côtés du seuil. Le seuil ne détermine pas le traitement, mais fait varier la **probabilité** d'être traité.
 - Puisqu'il y a des personnes traitées en dessous du seuil et des personnes non-traitées en dessus, on ne peut pas juste comparer les gens en dessous et au-dessus pour avoir une idée de l'effet du traitement.
- Passage par les variables instrumentales.
- L'interprétation change un peu : l'effet est maintenant doublement local.
 - Basé sur les personnes à proximité du seuil.
 - Et correspondant à l'effet moyen du traitement pour le personnes dont le traitement change lorsqu'on dépasse le seuil.

Hypothèses de la méthode: fuzzy design

- Avec un fuzzy design, on a:
- $\lim_{X_i \rightarrow c_0} \Pr(D_i = 1 | X_i = c_0) \neq \lim_{c_0 \leftarrow X_i} \Pr(D_i = 1 | X_i = c_0)$
- La probabilité "saute" au seuil.



Hypothèses de la méthode: fuzzy design

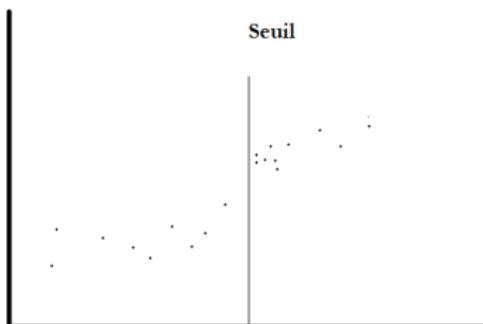
- On peut considérer que l'effet exact du traitement n'est pas si important que ça.
- Et ce que l'on veut savoir est ce qui se passe quand on franchit le seuil.
 - Estimation en "forme réduite".
 - Très courant.
- Mais si on ne se contente pas de ça ? Une possibilité est d'utiliser les "doubles" moindres carrés ordinaires:
 - 2SLS : on estime d'abord le traitement.
 - $D_i = \alpha + \beta_1 X_i + \beta_2 X_i^2 + \gamma Z_i + \epsilon_i$
 - Avec Z_i un indicateur qui vaut 1 quand $X > c$.

Hypothèses de la méthode: fuzzy design

- Une fois qu'on connaît α , β_1 , β_2 et γ , on crée la variable prédite \hat{D} .
- $\hat{D}_i = \alpha + \beta_1 X_i + \beta_2 X_i + \gamma Z_i$.
 - Et après on estime la régression à la slide 12 mais avec \hat{D} au lieu de D .
 - Mais pourquoi?
 - Le problème était que D n'est pas égal au fait d'être au dessus ou en dessous du seuil (il y en a des deux côtés).
 - Il ne s'agit plus d'une variation *exogène*. Le fait d'être traité est probablement lié à des choses qui sont elles-mêmes liées à l'outcome qu'on mesure.
- En utilisant \hat{D} , on utilise uniquement la variation en D qui vient du fait d'être au-dessus ou en-dessous du seuil (et on se débarrasse du risque d'une corrélation entre D et l'outcome).
 - On récupère alors le vrai effet du traitement, sur les personnes qui participent parce qu'ils sont just au-dessus plutôt que juste en dessous du seuil.

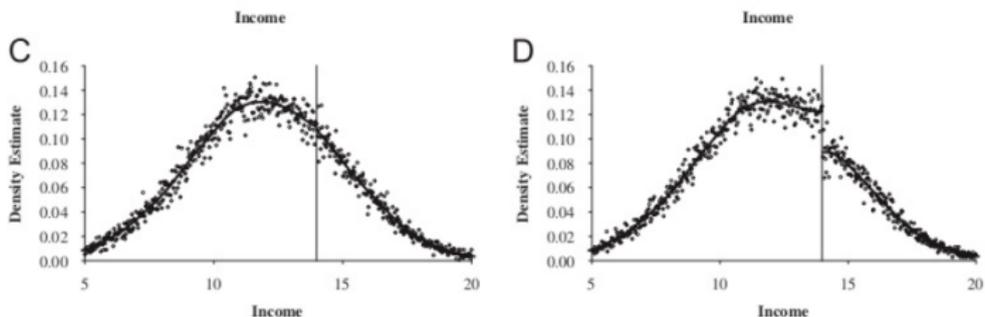
Retour sur l'hypothèse de manipulation

- Dans quel cas peut-on supposer que l'hypothèse de non-manipulation (avec précision) n'est pas satisfaite ?
 - Lorsque les valeurs sont autodéclarées et que le seuil est connu.
 - Possibilité de renouveler le score (refaire test, etc.).
- Le test de densité de McCrary (2008)
 - Intuition : dans la nature, pas de regroupement spontané des valeurs.



Retour sur l'hypothèse de manipulation

- Le test de McCrary regarde si la densité est différente au dessus et en dessous du seuil.
- Il est intuitivement simple, mais nécessite beaucoup de valeurs proche du seuil pour pouvoir être implémenté.



Source: McCrary (2008)

Exemple 1: qui choisit les programmes politiques

Lee, Moretti and Butler (2004)

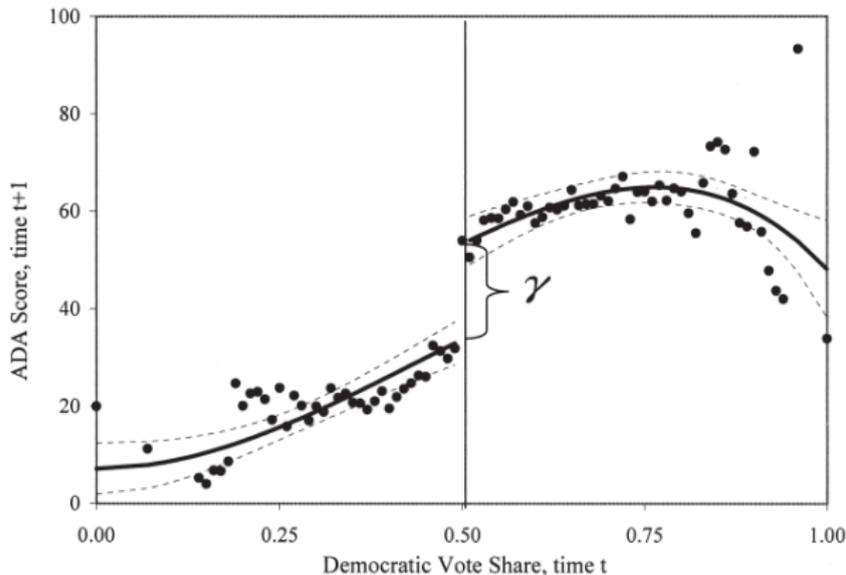
- Deux théories sur le rôle des élections dans la formation de programmes politiques.
 1. Selon l'**une**, les votants affectent les hommes politiques, qui choisissent des programmes pour capturer des votants.
 2. La compétition électorale fait bouger le programme des politiques vers le centre (l'électeur médian) \Rightarrow convergence.
 3. Selon l'**autre**, les politiques établissent des programmes et les votants élisent leur programme préféré.
 4. Les candidats ne sont pas crédibles dans leur modération de politique et vont juste implémenter leur politique préférée en cas de victoire \Rightarrow divergence.

Lee, Moretti and Butler (2004)

Comment tester cela?

- Ils utilisent des données sur les votes au Congrès américain en 1995-1996 et sur qui était au pouvoir dans chaque district en 1994 (démocrates ou républicains).
- Leur idée : si on pouvait randomiser le parti (D/R) qui détenait chaque district en 1994, on pourrait ensuite comparer les votes de 1995/1996 entre districts démocrates et républicains pour avoir une idée de l'effet causal du fait d'être un district républicain ou démocrate.
- D'après eux, cet effet est composé de deux effets:
 - Un effet dû au fait que puisqu'il y a un avantage à être titulaire, les votes dans les districts démocrates (1994) vont en 1995/1996 être plus libéraux (parce que l'élu a plus de chances d'être démocrate et donc d'avoir des idées libérales).
 - Un effet dû au fait que les politiques s'adaptent pour attirer l'électeur médian : un démocrate qui s'oppose à un titulaire républicain devrait voter moins libéral, afin de se rapprocher de l'électeur médian.

Lee, Moretti and Butler (2004)



Source: Lee, Moretti and Butler (2004)

Lee, Moretti and Butler (2004)

- L'outcome est un score ADA, qui varie de 0 à 100 (plus il est élevé plus les votes du représentant sont libéraux).
- L'effet total γ est bien positif et assez conséquent.
 - Jusque-là, rien de trop surprenant. Si c'est un candidat démocrate qui est élu, on s'attend à ce qu'il vote plus libéral.
 - Mais pourquoi ?
- La finesse de l'analyse réside dans la possibilité de décomposer l'effet total.
 - Les auteurs montrent en effet que l'effet 'elect' (le fait de voter pour un programme) peut être observé en comparant les votes des deux partis.
 - En enlevant cet effet de l'effet total, ils montrent qu'il y a peu de convergence vers l'électeur médian.

Exemple 2 Hoekstra (2009)

Hoekstra (2009) - cherche à mesurer l'effet sur le salaire du fait d'aller à l'Université d'Etat la plus prestigieuse.

Difficulté : les individus qui finissent dans les meilleures universités sont plus motivées, plus connectées, plus talentueux, plus riches que les autres.

Solution: utiliser des données sur les candidats et exploiter le seuil d'admission, qui dépendait des résultats dans deux examens nationaux.

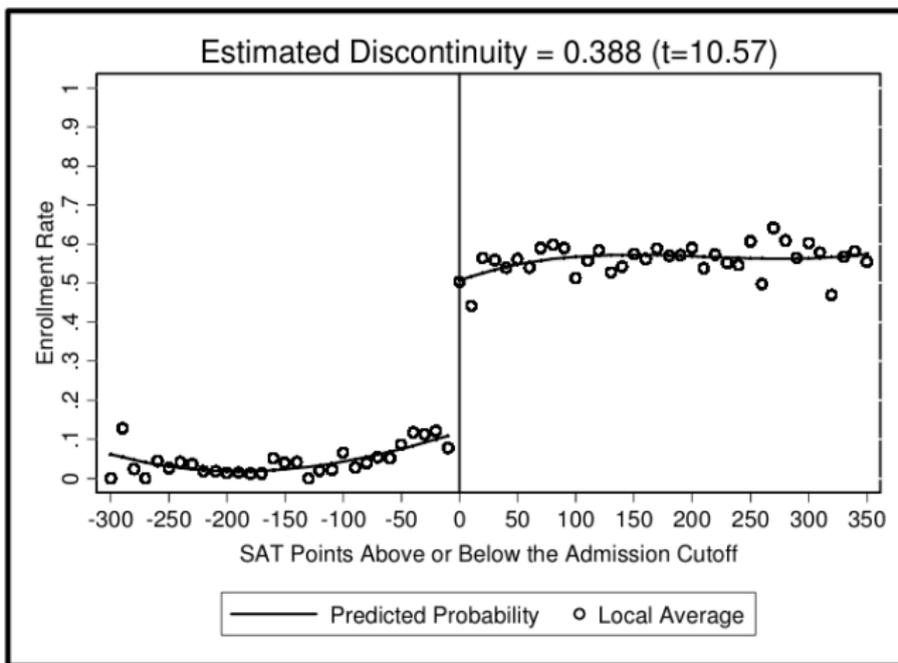
Hoekstra (2009)

GPA : Grade Point Average. Résultat composite sur les années de lycée (comprenant devoirs, examens, etc.).

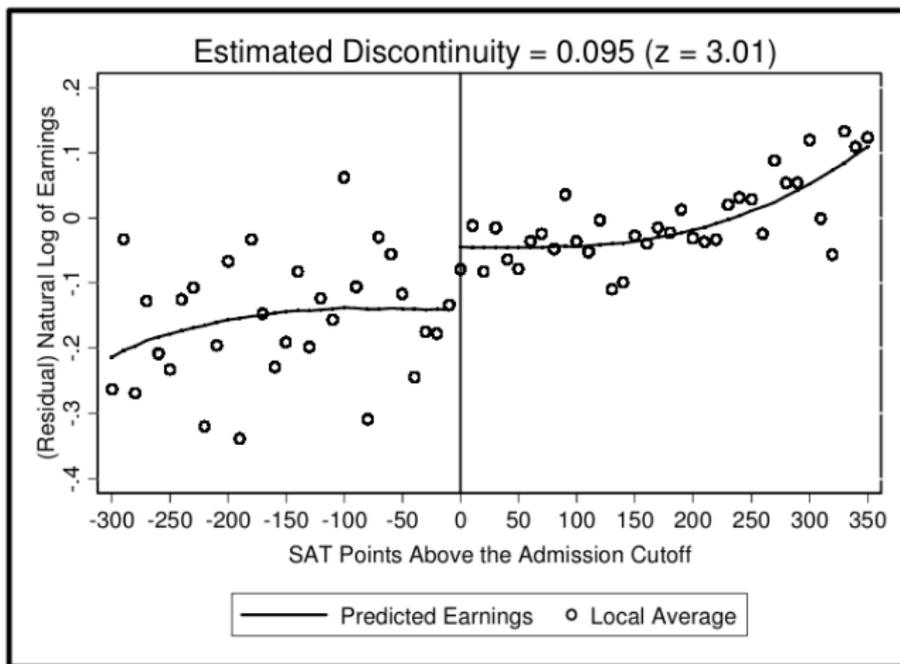
SAT : Résultat à un test standardisé de mathématiques, lecture critique et écriture.

Admission : pour chaque valeur de GPA, une note de SAT différente était nécessaire pour être admis. Par exemple, avec un GPA de 2.5, il fallait un SAT de 1300. Hoekstra convertit cela en un score individuel normalisé à 0 au seuil.

Hoekstra (2009)



Hoekstra (2009)



Wooclap

Wooclap