# La régression sur discontinuité

Introduction aux méthodes d’évaluation d’impact
des politiques publiques

Philippe De Vreyer

# Quand va-t-on employer cette méthode ?

# Conditions d’emploi de la méthode

- Quand le choix des personnes traitées et non traitées est effectué par application d’un critère d’éligibilité observable et qui ne peut être manipulé par les personnes traitées (ou non).

- Sous l’hypothèse que les personnes situées de par et d’autre de la frontière d’éligibilité au traitement, mais proches de celle-ci sont semblables et peuvent être comparées.

# Exemples

- Des variations dans le traitement du chômage peuvent s’opérer en fonction de l’âge ou du lieu d’habitation des chômeurs.
- La consommation d’alcool peut être autorisée à partir d’un certain âge.
- Les bourses d'études peuvent être réservées à des étudiants ayant obtenu une note minimum à un examen déterminé.
- …

# Principe

# Principe

- Utiliser des observations sur des individus à la marge du critère d'éligibilité pour évaluer l'efficacité du programme

- Hypothèse: les différences entre ceux qui sont de part et d'autre de la frontière, mais suffisamment proches, sont assez faibles pour ne pas biaiser l'estimation.

- Reproduire les conditions du tirage au hasard pour les personnes proches du seuil d'éligibilité

# Deux configurations possibles

- « Sharp design »: le critère d'éligibilité est appliqué de façon stricte. Le traitement, T, dépend alors de manière déterministe de la valeur prise par une variable de sélection S:

$$T_i = 1 \text{ si } S_i \geq S^*$$
$$T_i = 0 \text{ si } S_i < S^*$$

- « Fuzzy design »: la variable de sélection, S, ne fait qu'affecter la probabilité de participation. La probabilité d'être traité doit être plus importante lorsqu'on est éligible que lorsqu'on ne l'est pas:

$$P(T_i = 1 | S_i \geq S^*) > P(T_i = 1 | S_i < S^*)$$

# Exemple

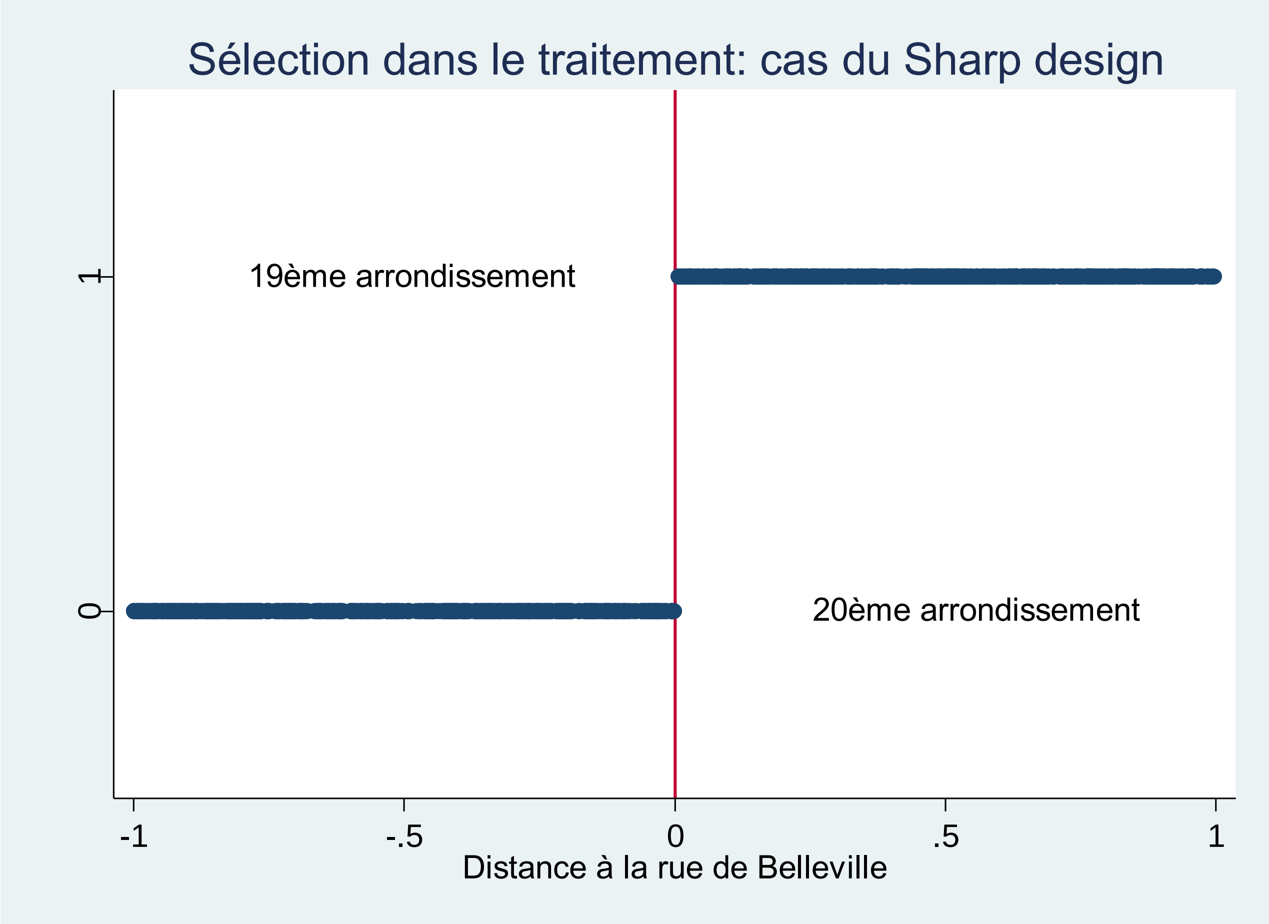
- A Paris, on désire évaluer l'impact d'une politique de distribution gratuite de livres de littérature dans les écoles primaires.

- Cette politique est supposée mise en place dans certains arrondissements uniquement.

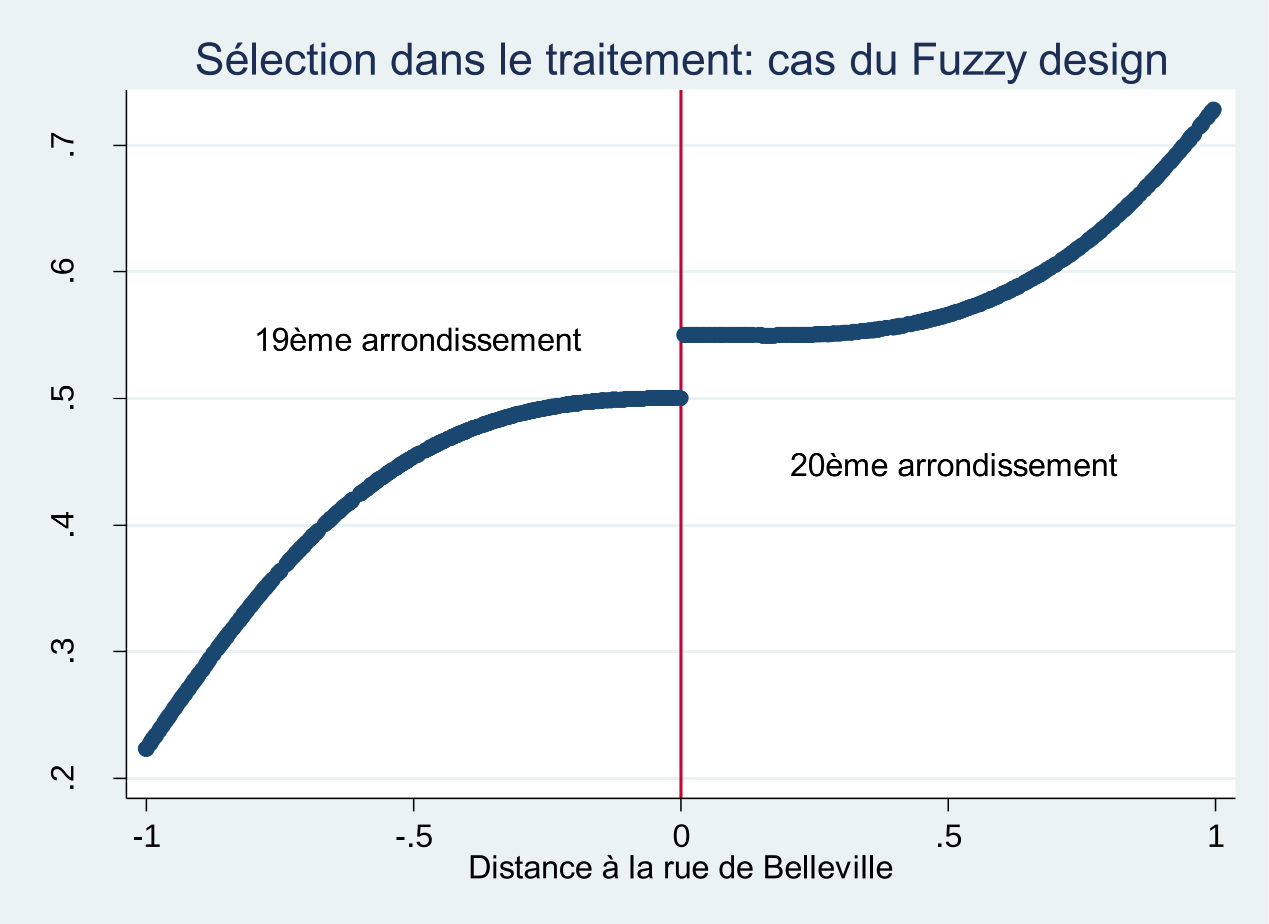
- Par exemple, supposons qu'elle est adoptée dans le 20ème arrondissement, mais pas dans le 19ème.


Av. Simón Bolivar
Rue Pradier
Rue Clavel
Rue Fessart
Rue Mélingue
Rue de la Villette
Villa de l'Adour
Centre D'Études
Supérieures Du Commerce
International
19ème
Rue de l'Atlas
Rue Rébeval
Rue Rébeval
Pl. Marcel Achard
Rue Jules Romains
Rue Rampal
Ecole Nationale
Supérieure D'Architecture
De Paris Belleville
Bureau De Poste
De Paris Simon Bolivar
Rue des Pyrénées
Pyrénées
Chapelle Notre-Dame
Du Bas Belleville
Rue Jouye-Rouve
Rue Piat
Rue Piat
Rue Lesage
Rue des Envierges
Maison De L'Air
Rue du Transvaal
Rue Ramponeau
20ème
Synagogue
Parc de Belleville
Parc De Belleville
10ème
Belleville
Bd de Belleville
11ème
Rue de Tourtille
Rue Bisson
Rue Vilin
Rue des Couronnes
100 m
© Mappy | 2016 TomTom


- *Sharp design* : les enfants du 19ème (et uniquement ceux-là) sont scolarisés dans le 19ème et ceux du 20ème (et uniquement ceux-là) sont scolarisés dans le 20ème.
- *Fuzzy design* : habiter sur le côté nord de la rue de Belleville (resp. le côté sud) accroît la probabilité d'être scolarisé dans le 19ème (resp. le 20ème).



Sélection dans le traitement: cas du Sharp design
19ème arrondissement
20ème arrondissement
0
1
-1
-.5
0
.5
1
Distance à la rue de Belleville

Sélection dans le traitement: cas du Fuzzy design
19ème arrondissement
20ème arrondissement
.2
.3
.4
.5
.6
.7
-1
-.5
0
.5
1
Distance à la rue de Belleville

# Théorie

# Identification de l'effet du traitement

- Notations:
- On peut écrire la valeur observée de la variable d'intérêt pour l'individu $i$ sous la forme:

$$Y_i = E\big(Y_i(0)\big) + \underbrace{\big(Y_i(1) - Y_i(0)\big)}.T_i + \underbrace{Y_i(0) - E(Y_i(0))}$$

$$Y_i = \alpha + \beta_i.T_i + \varepsilon_i$$

  - $\beta_i$ : effet du traitement sur l'individu $i$.

# Cas du « sharp design »

- Hypothèse: les individus situés de part et d'autre de la frontière, mais très proches de celle-ci, sont strictement comparables:
    - Pas de différence dans la valeur moyenne de $Y_i$ en dehors de l'effet du traitement
    - Réaction au traitement identique

- Autrement dit, dans l'équation: $Y_i = \alpha + \beta_i . T_i + \varepsilon_i$

(1) La composante inobservée de la variable d'intérêt, $\varepsilon_i$ vérifie:

$$\lim_{S \to S_-^*} E(\varepsilon_i | S_i = s) = \lim_{S \to S_+^*} E(\varepsilon_i | S_i = s)$$

(2) L'effet moyen du traitement $E(\beta_i | S_i = s)$ vérifie également:

$$\lim_{S \to S_-^*} E(\beta_i | S_i = s) = \lim_{S \to S_+^*} E(\beta_i | S_i = s)$$

Ni le terme d'erreur, ni l'effet du traitement ne connaissent de « saut » à la frontière d'éligibilité

- Effet moyen du traitement *à la frontière d'éligibilité*:

$$\beta = E(\beta_i | S_i = S^*) = \lim_{S \to S^*_+} E(Y_i | S_i = S) - \lim_{S \to S^*_-} E(Y_i | S_i = S)$$

Car en effet nous avons:

$$\lim_{S \to S^*_+} E(Y_i | S_i = S) = \alpha + E(\beta_i * 1 | S_i = S^*) + \lim_{S \to S^*_+} E(\varepsilon_i | S_i = s)$$

et

$$\lim_{S \to S^*_-} E(Y_i | S_i = S) = \alpha + E(\beta_i * 0 | S_i = S^*) + \lim_{S \to S^*_-} E(\varepsilon_i | S_i = s)$$

Et la différence entre ces deux limites conduit au résultat.

# Cas du « fuzzy design »

- Hypothèse maintenue: les autres déterminants de la variable d'intérêt, modélisés par le terme d'erreur $\varepsilon_i$, ne varient pas de façon discontinue aux alentours du seuil d'éligibilité, $S^*$:

$$\lim_{S \to S_-^*} E(\varepsilon_i | S_i = s) = \lim_{S \to S_+^*} E(\varepsilon_i | S_i = s)$$

- La différence $\lim_{S \to S_+^*} E(Y_i | S_i = S) - \lim_{S \to S_-^*} E(Y_i | S_i = S)$ est alors égale à:

$$\lim_{S \to S_+^*} E(\beta_i T_i | S_i = S) - \lim_{S \to S_-^*} E(\beta_i T_i | S_i = S)$$

- Deux possibilités:

  - L’effet du traitement est le même pour tous les individus au voisinage du seuil d’éligibilité

  - Au contraire cet effet varie d’individu en individu et a donc un caractère aléatoire.

- Premier cas: effet du traitement constant au voisinage du seuil d'éligibilité:

$$\beta_i = \beta \; pour \; tout \; i$$

On a alors:

$$\begin{aligned}
&\lim_{S \to S_+^*} E(Y_i|S_i = S) - \lim_{S \to S_-^*} E(Y_i|S_i = S) \\
&= \lim_{S \to S_+^*} E(\beta_i T_i|S_i = S) - \lim_{S \to S_-^*} E(\beta_i T_i|S_i = S) \\
&= \beta * \left(\lim_{S \to S_+^*} E(T_i|S_i = S) - \lim_{S \to S_-^*} E(T_i|S_i = S)\right)
\end{aligned}$$

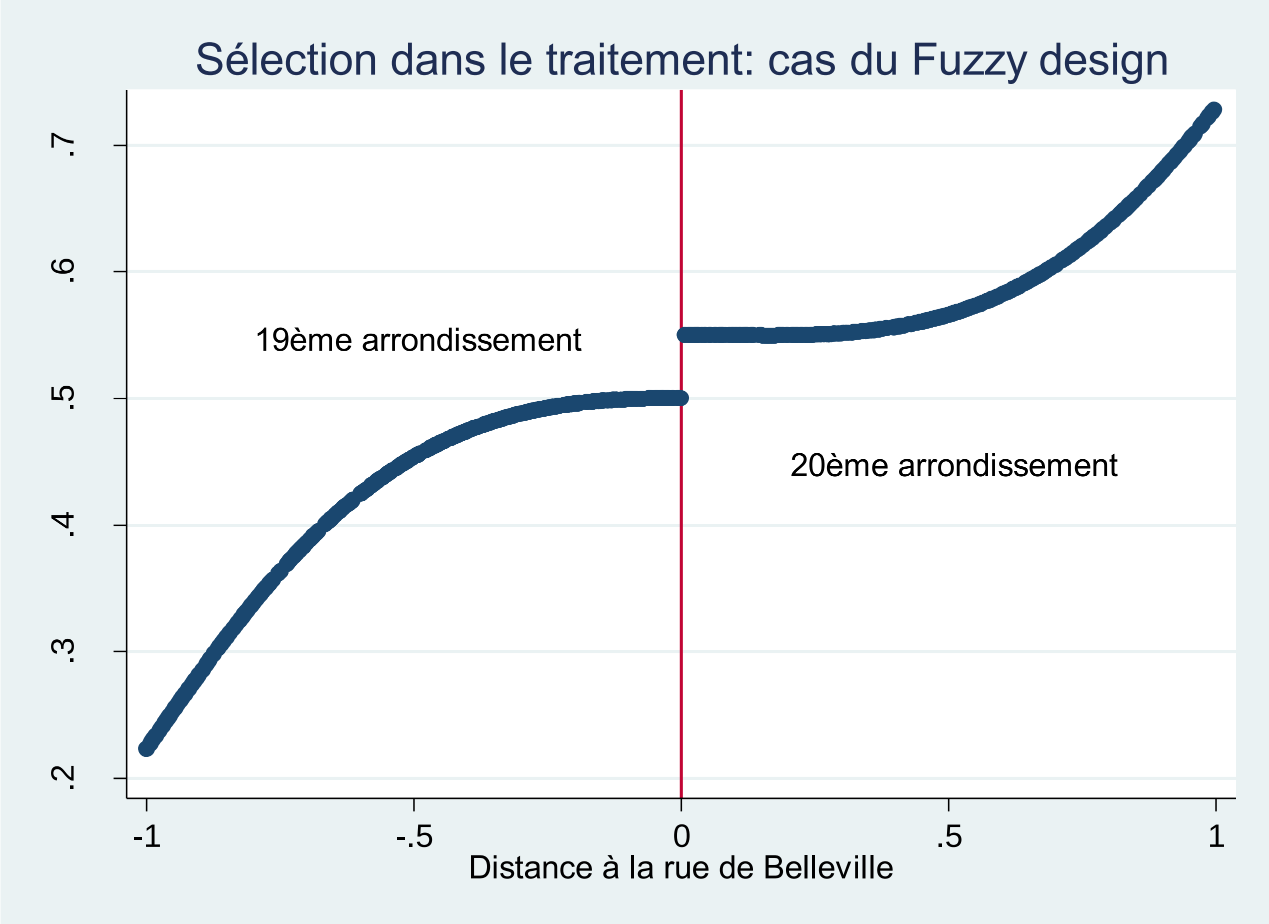
Et l'on obtient l'effet du traitement en calculant le ratio:

$$\beta = \frac{\lim_{S \to S_+^*} E(Y_i|S_i = S) - \lim_{S \to S_-^*} E(Y_i|S_i = S)}{\lim_{S \to S_+^*} E(T_i|S_i = S) - \lim_{S \to S_-^*} E(T_i|S_i = S)}$$

- Second cas: effet aléatoire du traitement au voisinage du seuil d’éligibilité.

  - L’effet du traitement est estimé de la même manière, mais l’interprétation est un peu différente.

  - Condition de validité du calcul: la probabilité d’être traité doit augmenter ou tout au moins ne pas diminuer lorsque l’on franchit le seuil.


Sélection dans le traitement: cas du Fuzzy design
19ème arrondissement
20ème arrondissement
.2
.3
.4
.5
.6
.7
-1
-.5
0
.5
1
Distance à la rue de Belleville

- **Effet aléatoire du traitement au voisinage du seuil d'éligibilité.**

  - L'effet du traitement est estimé de la même manière, mais l'interprétation est un peu différente.

  - Condition de validité du calcul: la probabilité d'être traité doit augmenter ou tout au moins ne pas diminuer lorsque l'on franchit le seuil.

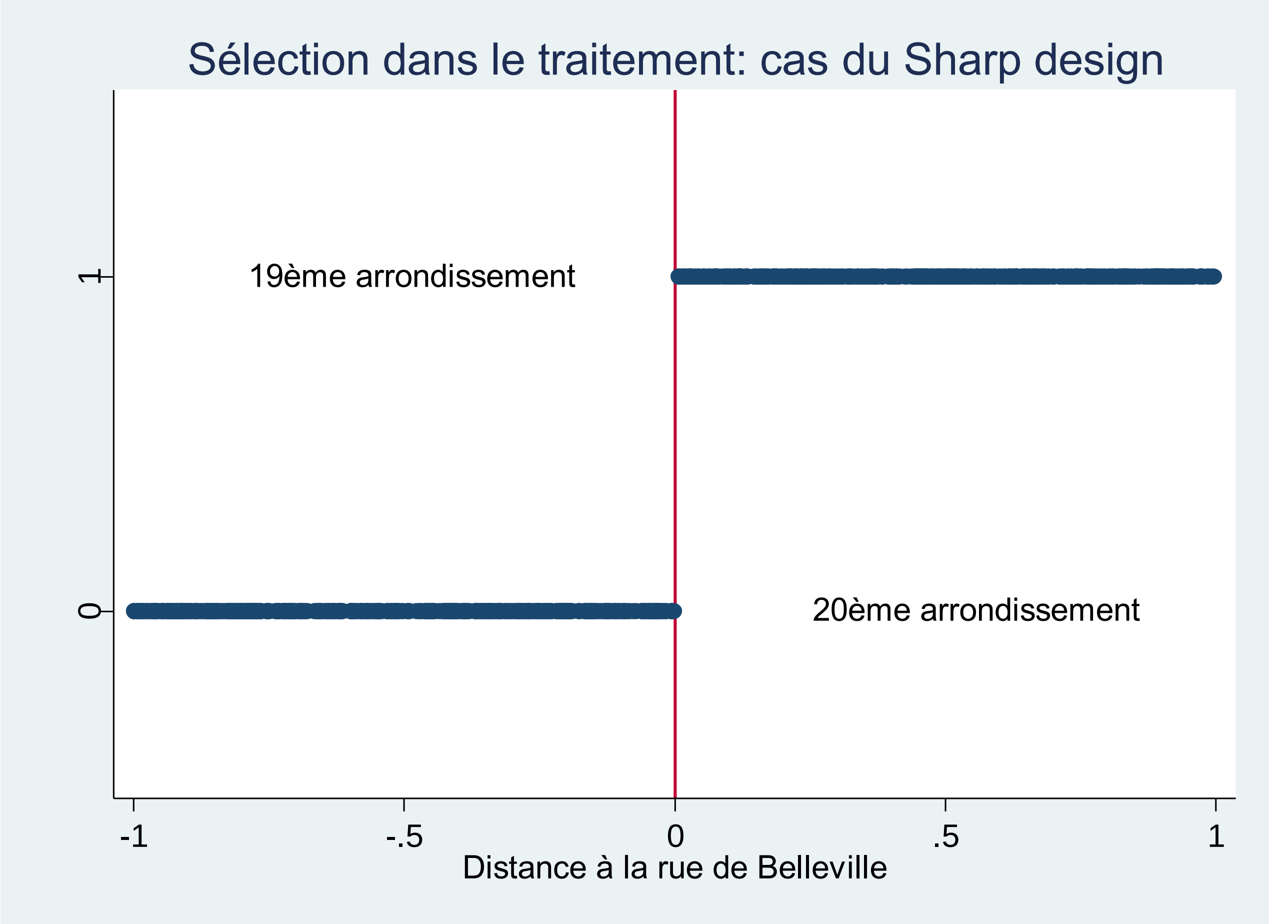
  - L'estimateur est de type LATE (Local Average Treatment Effect): il renseigne uniquement sur l'impact du traitement pour les individus pour lesquels franchir le seuil les conduit à être traité.

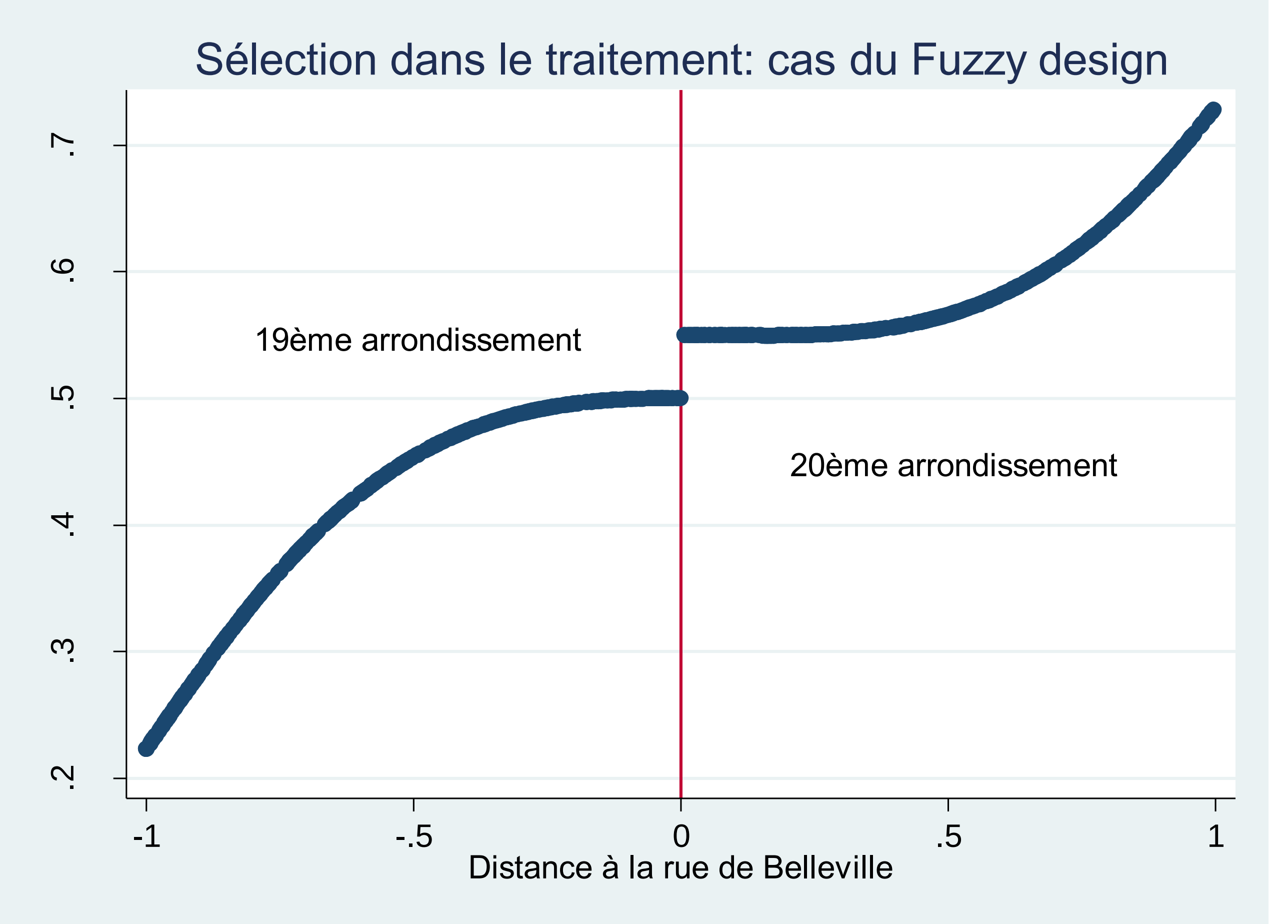
  - Exemple des livres distribués dans le 20ème: l'estimation du traitement renseigne sur l'effet de la politique pour les ménages qui *parce qu'ils habitent sur le côté sud de la rue de Belleville* décident de scolariser leur enfant dans le 20ème.

# Mise en oeuvre

# Mise en oeuvre

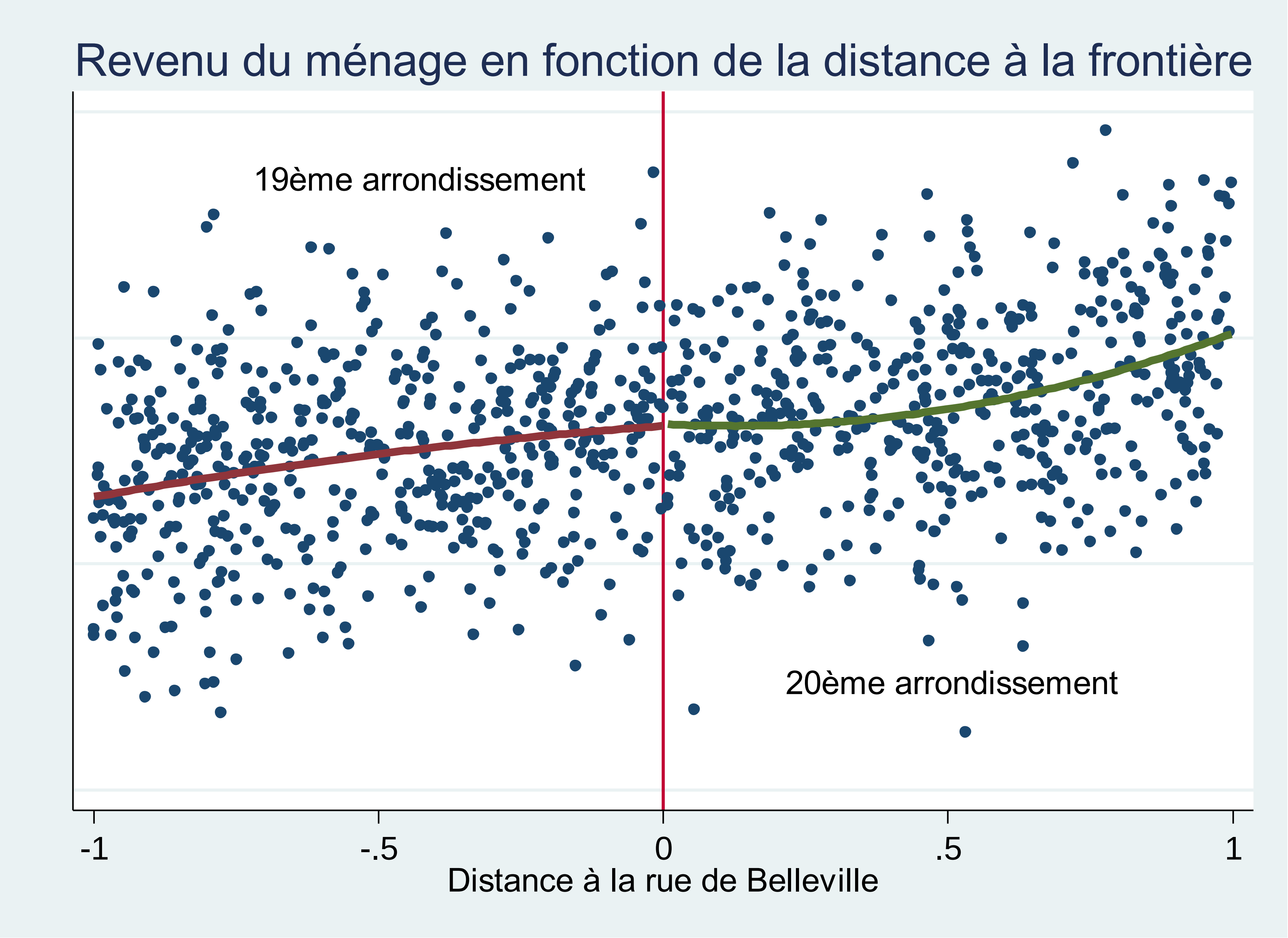
- Commencer par vérifier que la variable de traitement, T (distribuer ou non des livres) présente bien une discontinuité au point S=S* lorsque le traitement est appliqué.


Sélection dans le traitement: cas du Sharp design
1
0
19ème arrondissement
20ème arrondissement
-1
-.5
0
.5
1
Distance à la rue de Belleville


Sélection dans le traitement: cas du Fuzzy design
.7
.6
.5
.4
.3
.2
19ème arrondissement
20ème arrondissement
-1
-.5
0
.5
1
Distance à la rue de Belleville

# Mise en oeuvre

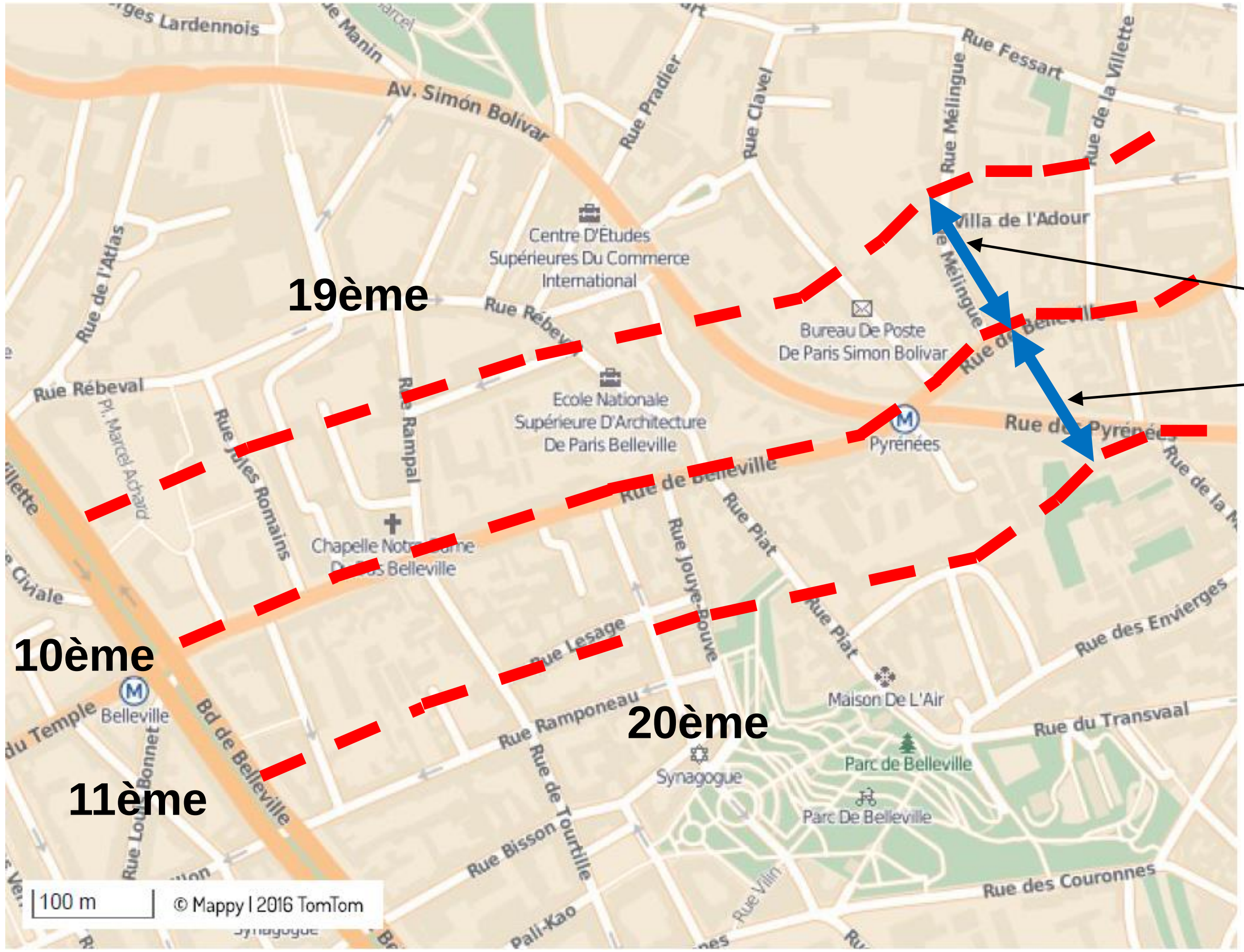
- Vérifier que la variable de traitement, T (distribuer ou non des livres) présente bien une discontinuité au point S=S* lorsque le traitement est appliqué.
- S'assurer que Y ne présente pas de discontinuité au point S=S* en l'absence du traitement.
- A défaut, vérifier que les autres déterminants de Y ne présentent pas de discontinuité aux alentours de S*.


Revenu du ménage en fonction de la distance à la frontière
19ème arrondissement
20ème arrondissement
-1
-.5
0
.5
1
Distance à la rue de Belleville

$$\beta = E(\beta_i|S_i = S^*) = \lim_{S \to S_+^*} E(Y_i|S_i = S) - \lim_{S \to S_-^*} E(Y_i|S_i = S)$$

$$\beta = \frac{\lim_{S \to S_+^*} E(Y_i|S_i = S) - \lim_{S \to S_-^*} E(Y_i|S_i = S)}{\lim_{S \to S_+^*} E(T_i|S_i = S) - \lim_{S \to S_-^*} E(T_i|S_i = S)}$$

- Choisir la largeur, 2h, de la bande autour de la frontière S=S* dans laquelle seront retenues les observations employées pour calculer l'estimateur.




Bande de largeur 2h.

# Le calcul en pratique

- Deux façons de procéder:

  - Avec un grand nombre d'observations disponibles *à proximité* du seuil d'éligibilité: estimation semi-paramétrique

  - Avec un nombre restreint d'observations à proximité du seuil: estimation paramétrique

# Estimation semi-paramétrique

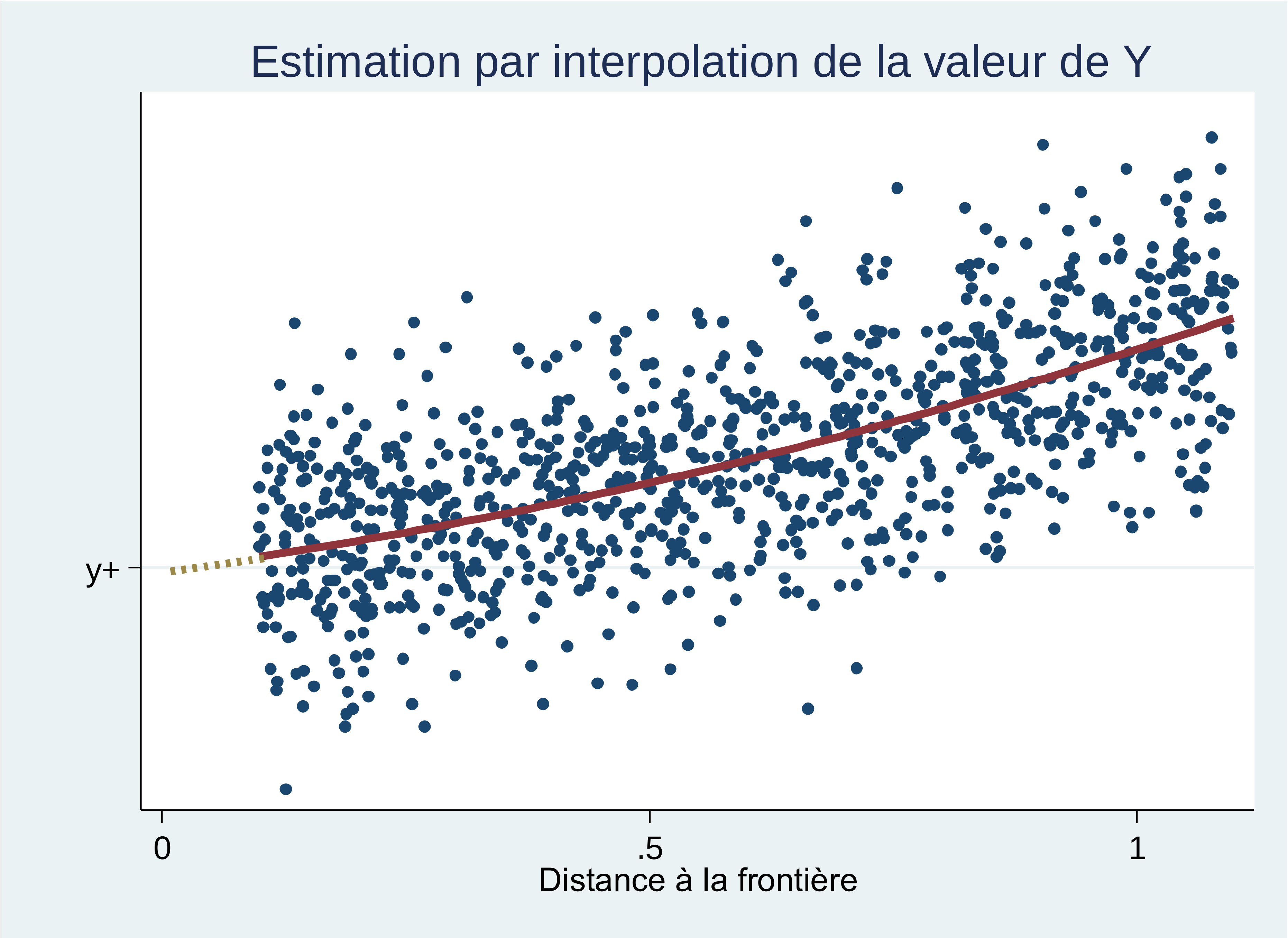
- « Sharp design »: $\beta = \lim_{S \to S_+^*} E(Y_i|S_i = S) - \lim_{S \to S_-^*} E(Y_i|S_i = S)$

- On cherche les solutions $(\widehat{y_+}, \widehat{b_1}, \dots, \widehat{b_p})$ de:

$$\min_{(y_+, b_1, \dots, b_p)} \sum_i 1_{\{S_i > S^*\}} k_h(S_i - S^*)(Y_i - y_+ - b_1(S_i - S^*) - \dots - b_p(S_i - S^*)^p)^2$$

L'estimation de $\lim_{S \to S_+^*} E(Y_i|S_i = S)$ est alors obtenue par:

$$\lim_{S \to S_+^*} E(Y_i\widehat{|S_i =} S) = \widehat{y_+}$$


Estimation par interpolation de la valeur de Y
y+
0
.5
1
Distance à la frontière

# Estimation semi-paramétrique

- « Sharp design »: $\beta = \lim_{S \to S_+^*} E(Y_i | S_i = S) - \lim_{S \to S_-^*} E(Y_i | S_i = S)$

- On cherche les solutions $(\widehat{y_+}, \widehat{b_1}, \dots, \widehat{b_p})$ de:

$$\min_{(y_+, b_1, \dots, b_p)} \sum_i 1_{\{S_i > S^*\}} k_h(S_i - S^*)(Y_i - y_+ - b_1(S_i - S^*) - \dots - b_p(S_i - S^*)^p)^2$$

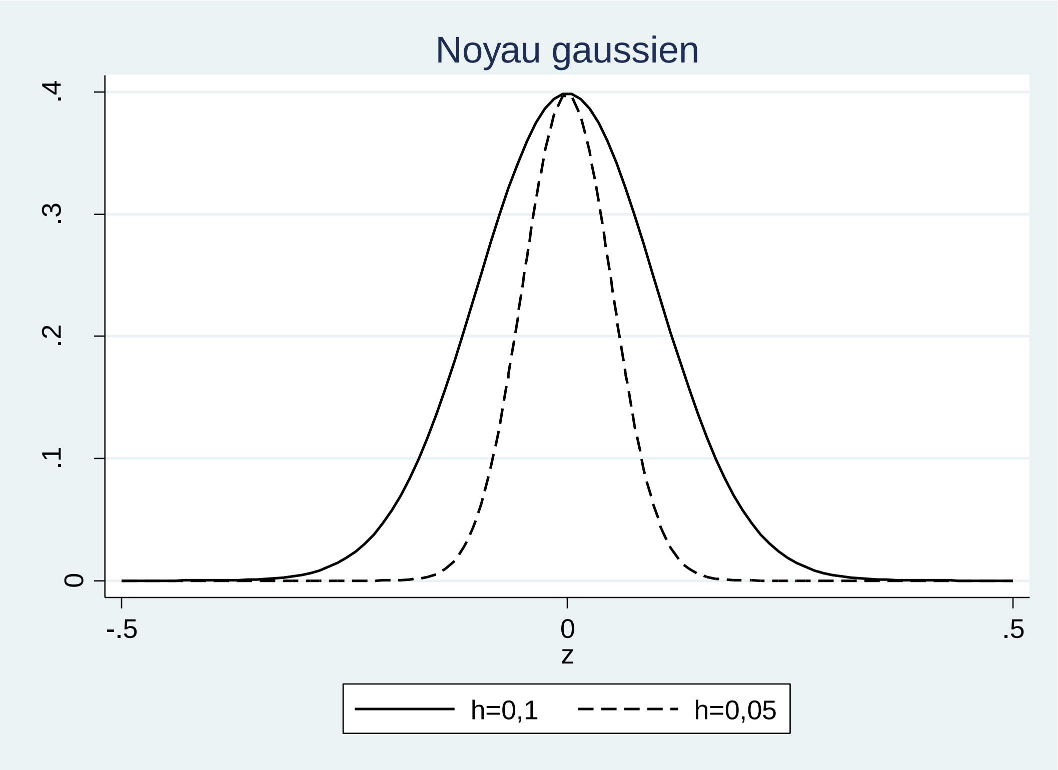
$k_h$: fonction « à noyau » paramétrée par h.

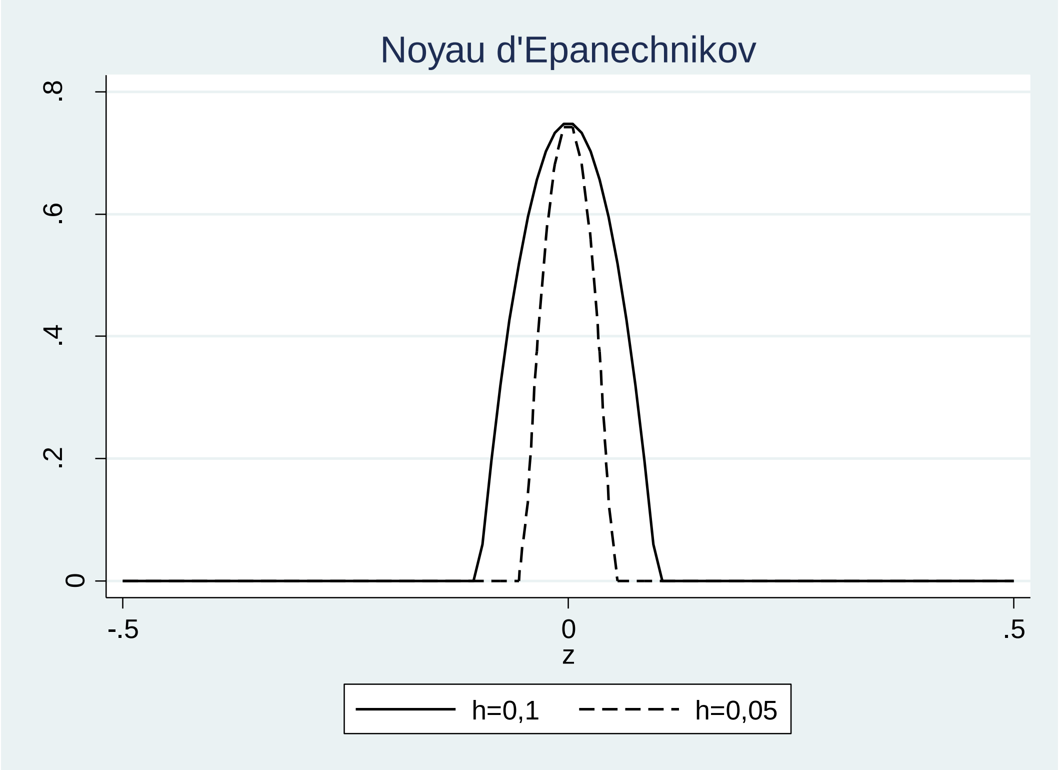
L'estimation de $\lim_{S \to S_+^*} E(Y_i | S_i = S)$ est alors obtenue par:

$$\lim_{S \to S_+^*} E(Y_i \widehat{| S_i = } S) = \widehat{y_+}$$

- Deux fonctions couramment employées:
  - Noyau gaussien $k_h(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2h^2}\right)$
  - Noyau d'Epanechnikov: $k_h(z) = \frac{3}{4}\left(1 - \frac{z^2}{h^2}\right) 1_{\{\frac{|z|}{h} \leq 1\}}$

où h est la taille de la « fenêtre » du noyau.



- Même calcul de l'autre côté de la frontière:

$$\lim_{S \to S_-^*} E(\widehat{Y_i | S_i = S}) = \widehat{y_-}$$

où $\widehat{y_-}$ est obtenu en recherchant:

$$\min_{(y_-, c_1, \ldots, c_p)} \sum_i 1_{\{S_i < S^*\}} k_h( S_i - S^*)(Y_i - y_- - c_1(S_i - S^*) - \cdots - c_p(S_i - S^*)^p)^2$$

L'estimation de l'impact du traitement est alors:

$$\beta = \widehat{y_+} - \widehat{y_-}$$

- « Fuzzy design »: $\beta = \dfrac{\lim\limits_{S \to S_+^*} E(Y_i | S_i = S) - \lim\limits_{S \to S_-^*} E(Y_i | S_i = S)}{\lim\limits_{S \to S_+^*} E(T_i | S_i = S) - \lim\limits_{S \to S_-^*} E(T_i | S_i = S)}$

- On cherche $\widehat{t_+}$ et $\widehat{t_-}$ solutions de:

$$\min_{(t_+, b_1, \ldots, b_p)} \sum_i 1_{\{S_i > S^*\}} k_h(S_i - S^*)(T_i - t_+ - b_1(S_i - S^*) - \cdots - b_p(S_i - S^*)^p)^2$$

et

$$\min_{(t_-, c_1, \ldots, c_p)} \sum_i 1_{\{S_i < S^*\}} k_h(S_i - S^*)(T_i - t_- - c_1(S_i - S^*) - \cdots - c_p(S_i - S^*)^p)^2$$

- On obtient l'estimation de l'impact du traitement en calculant le ratio:

$$\beta = \frac{\widehat{y_+} - \widehat{y_-}}{\widehat{t_+} - \widehat{t_-}}$$

# Estimation paramétrique

- « Sharp design »: on estime le modèle suivant par les moindres carrés ordinaires:

$$Y_i = m(S_i) + \beta . 1_{\{S_i > S^*\}} + \varepsilon_i$$

Fonction paramétrique de $S_i$ (par exemple polynôme: $a + bS_i$+c$S_i^2 + \cdots$)

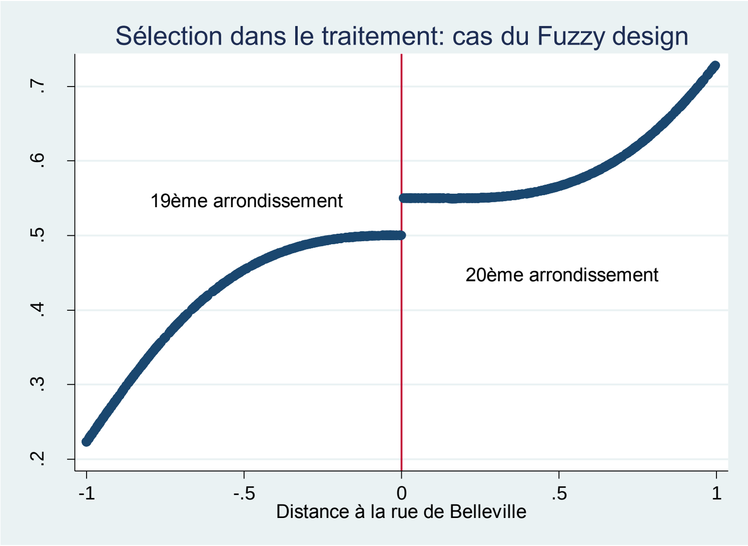
Effet du traitement

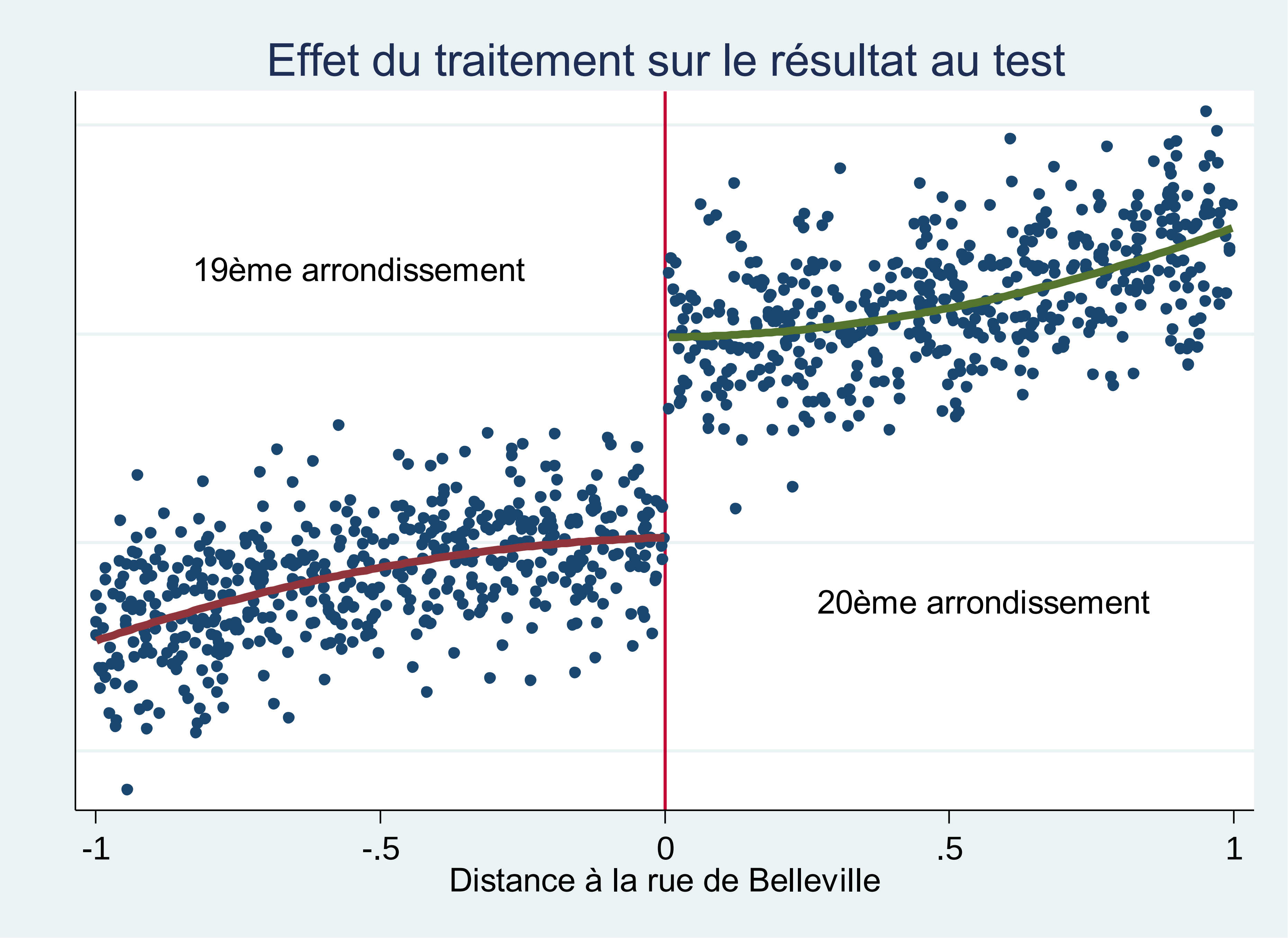
- « Fuzzy design »: on estime l'équation par doubles moindres carrés en utilisant $1_{\{S_i > S^*\}}$ est utilisé pour instrumenter $E(T_i|S_i)$.

- Concrètement on procède en deux étapes:

  - Etape 1, on régresse: $T_i = m(S_i) + \gamma. 1_{\{S_i > S^*\}} + u_i$

  - Etape 2, on régresse: $Y_i = m(S_i) + \beta. \widehat{T_i} + \varepsilon_i$




Effet du traitement sur le résultat au test
19ème arrondissement
20ème arrondissement
-1
-.5
0
.5
1
Distance à la rue de Belleville

# Limites et écueils

# Manipulation du seuil

- Hypothèse fondamentale: les personnes concernées par le programme ne sont pas en mesure de manipuler le seuil d'éligibilité.

- Dans le cadre d'un programme mis en place par une collectivité locale par exemple, une mairie ou un département, il ne faut pas que les individus choisissent leur lieu d'habitation en fonction de la politique testée.

# Limites

- La principale limite provient du fait que les résultats ne sont, stricto sensu, valables que pour les individus localisés à la frontière.

- La méthode de calcul paramétrique doit être employée avec précaution: les études sur le sujet indique qu'elle donne de mauvais résultats lorsque l'on emploie des observations trop éloignées de la frontière.

- Estimateur LATE (Local Average Treatment Effect) dans le cas du fuzzy design