Econométrie 2

Chapitre 6 : la méthode de régression par discontinuité

ENSAE 2021-2022

Michael Visser

CREST-ENSAE

Plan

Motivation

- Dans ce chapitre nous nous intéressons à l'estimation de l'effet d'une mesure (ou traitement) sur une variable Y.
- Soit P_i la variable indiquant si l'unité i (individu, ménage, firme, ville, etc.) est traité (dans ce cas $P_i = 1$) ou pas ($P_i = 0$).
- Nous adoptons le cadre d'analyse introduit par Rubin (voir Econométrie 1) :
  - Y_i^1 représente la variable expliquée lorsque i reçoit le traitement.
  - Y_i^0 représente la variable expliquée lorsque i ne reçoit pas le traitement.
- Y_i^0 et Y_i^1 sont les variables dépendantes contrefactuelles.
- Elles ne sont pas observées simultanément : $Y_i = Y_i^0$ si $P_i = 0$, ou $Y_i = Y_i^1$ si $P_i = 1$.

Motivation

- Donc
$$Y_i = P_i Y_i^1 + (1 - P_i) Y_i^0 = Y_i^0 + P_i (Y_i^1 - Y_i^0). \qquad (1)$$
- L'effet causal du traitement s'écrit
$$\alpha_i = Y_i^1 - Y_i^0.$$
- L'objectif ici est d'estimer l'effet moyen du traitement.
- La méthode de régression par discontinuité (RD) exploite le fait que l'assignation au traitement est complètement ou partiellement déterminée par une variable continue Z (en anglais : *forcing variable* ou *treatment-determining variable*).
- Par ailleurs, l'éligibilité au traitement est une fonction discontinue (de Z) en z_0.
- Deux cas de figure peuvent être distingués :

Motivation

- On parle de discontinuité franche (*Sharp Regression Discontinuity (SRD)*) lorsque i reçoit le traitement ssi Z_i est supérieur ou égal au seuil z_0 :
$$P_i = 1\{Z_i \geq z_0\}.$$
- On parle de discontinuité floue (*Fuzzy Regression Discontinuity (FRD)*) lorsque la probabilité que i recevoive le traitement (conditionnellement à $Z_i = z$), $P(P_i = 1|Z_i = z)$, est discontinue en $z = z_0$.
- Intuitivement, les unités i proches des deux côtés du seuil z_0 sont similaires $\Rightarrow$ identifier l'effet moyen du traitement en comparant les unités se situant juste au dessous et juste au-dessus du seuil.
- Exemples :
  - Effet des grandes écoles sur l'emploi, le revenu, etc. ($Z =$ résultat au concours).
  - Effet de l'aide financière aux ménages ($Z =$ revenu du ménage).

Motivation

- Exemples (suite) :
  - Elections : impact du sortant sur le vote ($Z =$% obtenu à l'élection précédente).
  - Effet de la taille de classe ($Z =$ nombre d'élèves d'un niveau donné dans l'école, e.g., règle de Maïmonide).
  - Effet d'une politique publique visant à réduire la pollution de l'air ($Z =$ niveau de pollution).
  - Effet du montant de l'allocation chômage ($Z =$ durée de chômage).
  - ... de nombreux autres exemples (voir par exemple Lee et Lemieux (2010)).

Plan

Identification

- Pour identifier l'effet moyen du traitement, il est nécessaire de faire une hypothèse de continuité locale.
- Hypothèse 1 : les espérances conditionnelles $E[Y_i^0|Z_i = z]$ et $E[Y_i^1|Z_i = z]$ sont continues en z.
- Cette hypothèse implique que les unités i juste au-dessous et juste au-dessus du seuil z_0 sont "comparables".
- En particulier, les moyennes de Y^0 et Y^1 sont similaires dans les deux groupes d'unités.
- Sous cette hypothèse on a

$$
\begin{aligned}
& \lim_{z \downarrow z_0} E[Y_i|Z_i = z] - \lim_{z \uparrow z_0} E[Y_i|Z_i = z] \qquad (2) \\
= \quad & \lim_{z \downarrow z_0} E[Y_i^1|Z_i = z] - \lim_{z \uparrow z_0} E[Y_i^0|Z_i = z] \\
= \quad & E[Y_i^1|Z_i = z_0] - E[Y_i^0|Z_i = z_0] \\
= \quad & E[Y_i^1 - Y_i^0|Z_i = z_0] \\
= \quad & E[\alpha_i|Z_i = z_0].
\end{aligned}
$$

Identification

- Les Figures 1 et 2 illustrent la stratégie d'identification.
- La Figure 1 montre la probabilité $P(P = 1|Z = z)$ en fonction de z. Cette probabilité vaut 0 pour tout $z < z_0 = 6$, et vaut 1 pour tout $z \geq z_0$.
- La Figure 2 montre trois espérances conditionnelles. Les deux courbes continues (partiellement en pointillés, partiellement solide) montrent $E[Y^1|Z = z]$ et $E[Y^0|Z = z]$ (la première étant au-dessus de la dexième).
- La courbe discontinue (solide) représente $E[Y|Z = z]$. En utilisant (1), on a

$$\begin{aligned} & E[Y|Z = z] \\ = \quad & E[Y^0|Z = z] + (E[Y^1|Z = z] - E[Y^0|Z = z])P, \end{aligned}$$

indiquant que $E[Y|Z = z]$ est discontinue en z_0 car P est une fonction discontinue en z_0.

Identification

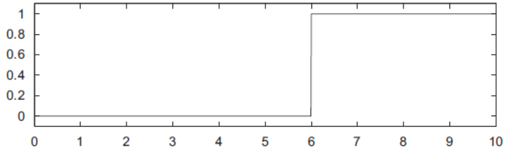

Figure 1 – Assignation du traitement

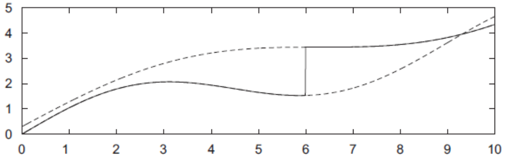

Figure 2 – $E[Y^0|Z=z]$, $E[Y^1|Z=z]$, et $E[Y|Z=z]$

Estimation

- D'après (2), $E[\alpha_i|Z_i = z_0] = \lim_{z\downarrow z_0} E[Y_i|Z_i = z] - \lim_{z\uparrow z_0} E[Y_i|Z_i = z]$.
- Pour estimer l'effet moyen du traitement il suffit donc d'estimer les espérances
$$\mu_l(z_0) \equiv \lim_{z\uparrow z_0} E[Y_i|Z_i = z]$$
et
$$\mu_r(z_0) \equiv \lim_{z\downarrow z_0} E[Y_i|Z_i = z].$$
- Hahn et al. (2001) proposent d'estimer $\mu_l(z_0)$ et $\mu_r(z_0)$ par une méthode de régression linéaire locale.
- Soit $K(u)$ une fonction de noyau (la fonction $K(\cdot)$ satisfait donc $\int_{-\infty}^{+\infty} K(u)du = 1$ et $K(u) = K(-u)$ pour tout u) et h une fenêtre.

Estimation

- On définit

$$(\hat{a}_l, \hat{b}_l) = \underset{a_l, b_l}{argmin} \sum_{i=1}^{n} [Y_i - a_l - b_l \cdot (Z_i - z_0)]^2 K\left(\frac{Z_i - z_0}{h}\right) 1\{Z_i < z_0\}$$

et

$$(\hat{a}_r, \hat{b}_r) = \underset{a_r, b_r}{argmin} \sum_{i=1}^{n} [Y_i - a_r - b_r \cdot (Z_i - z_0)]^2 K\left(\frac{Z_i - z_0}{h}\right) 1\{Z_i \geq z_0\}.$$

- Les espérances $\mu_l(z_0)$ et $\mu_r(z_0)$ peuvent alors être estimés par

$$\hat{\mu}_l(z_0) = \hat{a}_l \text{ et } \hat{\mu}_r(z_0) = \hat{a}_r.$$

- Dans la pratique on choisit fréquemment $K(u) = \frac{1}{2}1\{-1 \leq u \leq 1\}$ (noyau rectangulaire) ou $K(u) = \frac{1}{\sqrt{2\pi}} exp\left(-u^2/2\right)$ (noyau Gaussien).

Estimation

- L'inférence statistique et le choix optimal de la la fenêtre ne sont pas abordés dans ce cours.
- L'avantage de la méthode proposée par Hahn et al. (2001) : hypothèses peu restrictives sur le lien entre Y^0 et Z (linéarité locale entre les deux variables).
- Désavantage : le nombre d'observations n doit être grand pour que l'estimation et l'inférence statistique fonctionnent correctement.
- Une approche alternative consiste à imposer plus de structure au modèle :
- Tout d'abord en choisissant une spécification paramétrique (flexible) de $E[Y^0|Z=z]$.
- Dans la pratique on spécifie souvent cette espérance comme un polynome d'ordre p :

$$E[Y_i^0|Z_i=z_i]=\gamma_0+\sum_{j=1}^{p}\beta_{0j}z_i^j$$

Estimation

- Et donc

$$Y_i^0 = \gamma_0 + \sum_{j=1}^{p} \beta_{0j} z_i^j + \eta_i \tag{3}$$

où $E[\eta_i | Z_i = z_i] = 0$ par construction.

- Supposons ensuite que l'effet du traitement est constant : $Y_i^1 - Y_i^0 = \alpha_0$, pour tout all i
- En substituant (3) dans (1) on obtient

$$Y_i = \gamma_0 + \sum_{j=1}^{p} \beta_{0j} z_i^j + \alpha_0 P_i + \eta_i$$

où $E[\eta_i | Z_i = z_i, P_i] = E[\eta_i | Z_i = z_i] = 0$ (car P_i est une fonction déterministe de Z_i).

- En estimant ce modèle par les MCO on obtient un estimateur convergent des paramètres (en particulier α_0).

Plan

Identification

- Supposons, pour simplifier la preuve d'identification, que $Y_i^1 - Y_i^0 = \alpha_0$. On a alors le résultat d'identification suivant

$$\alpha_0 = \frac{\lim_{z \downarrow z_0} E[Y_i|Z_i = z] - \lim_{z \uparrow z_0} E[Y_i|Z_i = z]}{\lim_{z \downarrow z_0} E[P_i|Z_i = z] - \lim_{z \uparrow z_0} E[P_i|Z_i = z]} \tag{4}$$

- Preuve : en utilisant (1) on a

$$\begin{aligned} E[Y_i|Z_i = z_0 + e] - E[Y_i|Z_i = z_0 - e] \\ = \alpha_0\{E[P_i|Z_i = z_0 + e] - E[P_i|Z_i = z_0 - e]\} \\ + \{E[Y_i^0|Z_i = z_0 + e] - E[Y_i^0|Z_i = z_0 - e]\} \end{aligned}$$

où $e > 0$ est un nombre arbitrairement petit.

Grâce à l'hypothèse 1, le second terme entre {} tend vers zéro lorsque e tend vers zéro. Donc

$$\lim_{z \downarrow z_0} E[Y_i|Z_i = z] - \lim_{z \uparrow z_0} E[Y_i|Z_i = z]$$
$$= \alpha_0 \left\{ \lim_{z \downarrow z_0} E[P_i|Z_i = z] - \lim_{z \uparrow z_0} E[P_i|Z_i = z] \right\},$$

et on obtient alors (4).

- Les Figures 3 et 4 illustrent la stratégie d'identification.
- La Figure 3 montre que la probabilité $P(P = 1|Z = z)$ est discontinue en $z = z_0 = 6$.

Identification

- La Figure 4 montre $E[Y^1|P=1, Z=z]$ and $E[Y^0|P=0, Z=z]$ (courbes en pointillés), la première espérance étant au-dessus de la deuxième). Contrairement au cas SRD, ces espérances dépendent de P.
- La Figure 4 présente également (courbe solide) $E[Y|Z=z]$, où

$$
\begin{aligned}
E[Y|Z=z] = E[Y^0|P=0, Z=z] \cdot Pr(P=0|Z=z) \\
+ E[Y^1|P=1, Z=z] \cdot Pr(P=1|Z=z)
\end{aligned}
$$

Identification

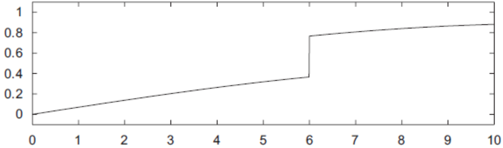

Figure 3 – Probabilité d'être traité

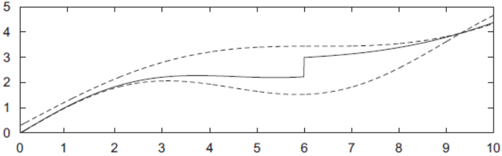

Figure 4 – $E[Y^0|Z = z, P = 0]$, $E[Y^1|Z = z, P = 1]$, et $E[Y|Z = z]$

Estimation

- Le traitement moyen α_0 peut être estimé non-paramétriquement en procédant comme dans le cas SRD.
- En effet, à la fois le numérateur et le dénominateur apparaissant dans (4) peuvent être estimés en utilisant la méthode de régression linéaire locale.
- Comme précédemment, une approche paramétrique a l'avantage de fonctionner correctement même si la taille de l'échantillon est relativement petite.
- Supposons, de nouveau, que $E[Y^0|Z = z]$ s'écrit comme un polynome d'ordre p.
- La variable Y^0 est donc toujours définie comme dans l'équation (3).

Estimation

- En utilisant (1), et le fait que $\alpha_0 = Y_i^1 - Y_i^0$ pour tout i, on obtient comme dans le cas SRD

$$Y_i = \gamma_0 + \sum_{j=1}^{p} \beta_{0j} z_i^j + \alpha_0 P_i + \eta_i \qquad (5)$$

- Par contre, P_i n'est plus une fonction déterministe de Z_i.
- Il y a potentiellement des variables inobservées qui déterminent à la fois P et Y, et dans ce cas $E[\eta_i | Z_i = z_i, P_i] \neq 0$.
- Une variable instrumentale naturelle pour P_i :

$$T_i \equiv 1\left\{Z_i \geq z_0\right\}.$$

- Les variables T_i et P_i sont positivement corrélées ; Par ailleurs, comme (5) inclut Z, Z^2, ..., Z^p, il y a vraisemblablement pas de corrélation entre T et η $\Rightarrow$ estimation par 2MC.

Plan

Méthode SRD : application

- Lee et Lemieux (2010) étudient l'avantage du sortant aux élections américaines de la Chambre des représentants.
- Plus précisement, $P_{it} = 1$ si le parti démocrate a gagné l'élection t dans la circonscription (*district*) i, et $P_{it} = 0$ sinon.
- $Y_{it+1} = \%$ des votes obtenu lors de l'élection $t+1$ par le candidat démocrate dans la circonsription i.
- Z_{it}=% des votes obtenu par le candidat démocrate moins % des votes obtenu par le candidat républicain, lors de l'élection t.
- Dans la plupart des circonscriptions il y a seulement 2 candidats (un démocrate et un républicain) : $P_{it} = 1\{Z_{it} \geq 0\}$.
- Objectif : estimer l'effet causal de P_{it} sur Y_{it+1}.
- Les données portent sur les élections de 1946-1998, et $n = 6558$.

Méthode SRD : application

- L'axe horizontal de la Figure 5 représente Z_{it} et l'axe verticale représente Y_{it+1}.
- Le support de Z a été divisé en 100 intervalles de largeur 0,01 et chaque point représente la moyenne de Y par intervalle.
- Le saut en $Z = 0$ suggère que l'effet causal vaut approximativement 0,08.

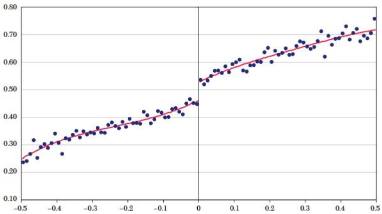

Figure 5 – Z_{it} et Y_{it+1}

Méthode SRD : application

- L'effet moyen du traitement, $E[\alpha_i | Z_i = 0]$, est estimé en utilisant le méthode de régression linéaire locale avec un noyau rectangulaire.
- Le tableau ci-dessous donne l'effet moyen estimé, c'est à dire $\hat{a}_r - \hat{a}_l$, et $se(\hat{a}_r - \hat{a}_l)$, pour différentes valeurs de la fenêtre h.
- Les estimations et leurs précisions diminuent lorsque h diminue $\Rightarrow$ arbitrage entre précision et biais.

Table 1 – Estimation de l'effet moyen du traitement

h	1	0,5	0,25	0,1	0,05	0,01
est. (*se*)	$0,347$ (0,003)	$0,257$ (0,004)	$0,179$ (0,004)	$0,125$ (0,006)	$0,096$ (0,009)	$0,088$ (0,015)
n	6558	4900	2763	1209	610	106

L'essentiel

- L'élégibilité à un traitement est une fonction discontinue (de Z) : potentiellement beaucoup d'applications.
- Discontinuité franche et discontinuité floue.
- Estimation par la méthode de régression linéaire locale.

Sources

- Angrist, J. et J.-F. Pischke (2009), *Mostly Harmless Econometrics*, Princeton University Press.
- Imbens, G. et T. Lemieux (2008), "Regression discontinuity designs : A guide to practice," *Journal of Econometrics*, 142, 615-635.
- Lee, D. et Lemieux (2010), "Regression discontinuity designs in economics," *Journal of Economic Literature*, 48, 281-355.
- Hahn, J., P. Todd et W. Van der Klaauw (2001), "Identification and estimation of treatment effects with a regression-discontinuity design," *Econometrica*, 69, 201-209.
- Van der Klaauw, W. (2008), "Regression-discontinuity analysis : a survey of recent developments in economics," *Labour*, 22, 219-245.