



Notes d'économétrie

Auteur :

Pierre ROUILLARD

Last update : 17 août 2023

Table des matières

1	Econo2	2
1.1	Interprétation du paramètre causal à estimer	2
1.2	Représentations linéaires & OLS	4
1.3	Modèle I.V - estimateur 2SLS/2MC	6
1.4	Méthode des moments généralisés	7
1.5	Données de panel	8
1.6	Modèles binaires	9
1.7	Modèles de censure et sélection	10
2	XXX	12
2.1	xxx	12

1 Econo2

1.1 Interprétation du paramètre causal à estimer

Selon le modèle considéré il est possible ou non d'avoir une interprétation quantitative directe et/ou qualitative du paramètre causal à estimer β_0 .

Définition de l'effet marginal de X_k sur Y : $\frac{\partial E[Y|X=x]}{\partial x_k}$

→ **Modèle linéaire** :

Comme toujours par la suite on considère l'analyse *toutes choses égales par ailleurs*, sur la population considérée...

Interprétation : la variable d'intérêt est Y et en l'absence de puissance ou d'interactions on peut interpréter quantitativement β_0 sur la variable d'intérêt. Le paramètre d'intérêt est l'effet marginal de X_k sur Y qui vaut bien β_{0k} lorsque X_k apparaît simplement dans le modèle. C'est justement pour cela qu'on peut bien interpréter directement quantitativement les coefficients de β_0 !

→ **Modèle binaire** :

$$E[Y|X] = P(Y = 1|X) = F(X'\beta_0)$$

$$E[Y|X] = F(X'\beta_0) \iff Y = \mathbb{1}(Y^* \geq s) : Y^* = X'\beta_0 + \varepsilon \quad \varepsilon \perp\!\!\!\perp X$$

Interprétation quantitative directe : la variable d'intérêt est Y et Y^* n'est qu'une variable latente qui n'a pas forcément de sens quantitatif précis. Le paramètre d'intérêt est l'effet marginal de X_k sur la variable d'intérêt, ici Y . Les coefficients de β_0 concernant Y^* on ne peut donc pas directement interpréter quantitativement ces derniers sur la variable d'intérêt Y . De plus, l'effet marginal de X_k sur Y est différent de β_{0k} : c'est pour cela qu'on ne peut avoir d'interprétation quantitative des coefficients de β_0 !

Interprétation qualitative : en revanche le signe de l'effet marginal de X_k sur Y , i.e. effet positif ou négatif sur $P(Y = 1|X)$, est donné par le signe de β_{0k} .

On peut en revanche comparer quantitativement le ratio des effets marginaux des variables i et j qui vaut $\widehat{\beta}_i/\widehat{\beta}_j$. L'effet sur la proba d'être ... de la variable i est <quantitativement> ... que l'effet de la variable $j \iff$ regarder le rapport $\widehat{\beta}_i/\widehat{\beta}_j$.

→ **Modèle de censure / Tobit1** :

1 seul mécanisme détermine la valeur de Y et si on observe la variable d'intérêt ou non. Deux cas sont à distinguer :

⇒ **Données censurées** : la variable d'intérêt est Y^* qui peut ne pas être observée au dessous d'un seuil causant un problème de censure. Le paramètre d'intérêt est l'effet marginal de X_k sur la variable d'intérêt Y^* , qui vaut bien β_{0k} lorsque X_k apparaît simplement dans

le modèle linéaire de Y^* . Ainsi, la variable Y^* ayant un sens quantitatif et malgré la censure liée aux problèmes d'observation on peut bien interpréter quantitativement β_0 sur la variable d'intérêt.

⇒ **Solution en coin** : la variable d'intérêt est bien Y alors que la variable Y^* est une variable latente potentiellement dépourvue de sens quantitatif. Typiquement un pb d'optimisation du consommateur où Y^* mesure l'utilité optimale (en nombre de biens) de consommation d'un bien donné : donc potentiellement négatif. Et Y représente le nombre d'unités effectivement consommées. Les coefficients de β_0 concernant Y^* qui n'as pas de sens quantitatif précis : on ne peut pas interpréter quantitativement les coefficients de β_0 sur la variable d'intérêt Y . Les paramètres d'intérêt sont les effets marginaux : le total $\frac{\partial E[Y|X=x]}{\partial x_k}$ (marge extensive et intensive) et $\frac{\partial E[Y|Y>0, X=x]}{\partial x_k}$ (marge intensive seulement). Ces paramètres sont tous les deux différents de β_{0k} ce qui explique le manque d'interprétation quantitative des coefficients de β_0 .

↪ **Modèle de sélection / Tobit2 :**

Ici on a bien deux processus différents : un qui détermine Y^* **et un autre** qui détermine si on observe cette valeur ou non i.e. modèle sur D .

Interprétation **quantitative directe** : il y a un problème d'observation des données, on observe $Y = D.Y^*$ mais la variable d'intérêt est bien Y^* (variable potentielle qui existe pour tous les *individus*). Par conséquent Y^* suivant un modèle linéaire, les paramètres d'intérêts sont les effets marginaux des variables explicatives sur la variable d'intérêt Y^* et les coefficients de β_0 sont toujours interprétables quantitativement.

1.2 Représentations linéaires & OLS

On note $Y \in \mathbf{R}$ la variable d'intérêt/dépendante et $X \in \mathbf{R}^K$ le vecteur des variables explicatives.

↪ Représentation non causale - Projection linéaire :

Sous conditions de moments¹, on a toujours par construction/définition de la représentation linéaire théorique (=projection linéaire orthogonale) l'orthogonalité des *résidus* de cette représentation non causale avec les régresseurs = pas une hypothèse mais une conséquence.

$Y = X' \cdot \tilde{\beta} + \tilde{\varepsilon}$, $E[X\tilde{\varepsilon}] = 0$ toujours définissable sous conditions de moments.

- $\hat{\beta}_{OLS}$ estime toujours $\tilde{\beta}$: $\hat{\beta}_{OLS} \xrightarrow{n \rightarrow +\infty} \tilde{\beta}$
- $X' \cdot \tilde{\beta}$ meilleure prédiction linéaire de Y par X : $\tilde{\beta}$ solution MSE.

$$\hat{\beta}_{OLS} \in \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^n (Y_i - X'_i \cdot \beta)^2 \quad \longleftrightarrow \quad \tilde{\beta} \in \underset{\beta}{\operatorname{argmin}} E[(Y - X' \cdot \beta)^2]$$

↪ Représentation causale :

La représentation causale fait intervenir le paramètre causal β_0 qu'on cherche à estimer : dans cette représentation le *terme d'erreur* n'est pas automatiquement orthogonal au régresseur.

$Y = X' \cdot \beta_0 + \varepsilon$, $E[X\varepsilon] \stackrel{?}{=} 0$

- β_0 paramètre causal à estimer.
- ε résidu : agrège les facteurs inobservés qui affectent Y.

Le terme d'erreur ε capte l'hétérogénéité inobservée, i.e capte les déterminants inobservés qui affectent la variable d'intérêt Y : deux individus avec les mêmes variables explicatives auront néanmoins la plupart du temps des variables expliquées différentes.

Avoir orthogonalité (→ indépendance) entre régresseurs et terme d'erreur est une hypothèse ! C'est l'hypothèse d'exogénéité.

↪ Lien :

Sans l'hypothèse d'exogénéité pour la représentation causale, les deux représentations diffèrent et l'estimateur OLS ne permet pas d'identifier le paramètre causal d'intérêt.

En revanche avec hypothèse d'exogénéité les deux représentations coïncident et $\tilde{\beta} = \beta_0$: $\hat{\beta}_{OLS}$ qui estime toujours $\tilde{\beta}$ est donc un estimateur consistant de β_0 .

1. $E[Y^2] < +\infty$, $E[\|X\|^2] < +\infty$ et $E[XX']$ inversible = de rang plein. En particulier : **any level of correlation between covariates except perfect colinearity** : composantes de X linéairement indépendantes mais *n'exclut pas* qu'elles soient corrélées. Si le modèle a une constante et une variable catégorielle il faut exclure une des modalités.

En dehors des expériences contrôlées les variables explicatives peuvent parfois être corrélées aux facteurs inobservables et pb d'endogénéité $E[X\varepsilon] \neq 0$.

↪ **Mémo OLS :**

⊙ *link-ols*

Représentation causale

Estimateur OLS - MCO

$$Y = X' \cdot \beta_0 + \varepsilon, \text{ avec } E[\varepsilon] = 0$$

- Conditions de moments
- $E[X\varepsilon] = 0$ hypothèse d'exogénéité

$$\Rightarrow \beta_0 \text{ est identifiable } \beta_0 = E[XX']^{-1}E[XY]$$

\Rightarrow L'estimateur $\hat{\beta}_{OLS}$ est consistant, sans biais & asymptotiquement normal

\Rightarrow Avec hypothèse homoscédasticité l'estimateur OLS est *BLUE*

On a :

$$\hat{\beta}_{OLS} = \left(\frac{1}{n} \sum_{i=1}^n X_i X_i' \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n X_i Y_i \right)$$

$$\sqrt{n}(\hat{\beta}_{OLS} - \beta_0) \xrightarrow{n \rightarrow +\infty} \mathcal{N}(0, \underline{E[XX']^{-1}E[\varepsilon^2 XX']E[XX']^{-1}})$$

Homoscédasticité : hypothèse sur la variance des résidus. L'erreur standard des MCO diffère en fonction de l'hypothèse !

- Faible $E[\varepsilon^2 XX'] = E[\varepsilon^2]E[XX']$
- Forte $E[\varepsilon^2|X] = \sigma^2 \leftrightarrow V[\varepsilon|X] = \sigma^2$

L'hypothèse d'homoscédasticité forte requiert que la variance des termes d'erreur soit la même pour chaque observation et indépendante de X . L'estimateur \hat{V} de la variance asymptotique de $\hat{\beta}_{OLS}$ est robuste à l'hétéroscédasticité. L'estimateur standard \tilde{V} n'est convergent que sous l'hypothèse d'homoscédasticité.

$$\begin{aligned} \underline{\text{Robust}} \quad \hat{V} &= \left(\frac{1}{n-k} \sum_{i=1}^n X_i X_i' \right)^{-1} \left(\frac{1}{n-k} \sum_{i=1}^n \hat{\varepsilon}_i^2 X_i X_i' \right) \left(\frac{1}{n-k} \sum_{i=1}^n X_i X_i' \right)^{-1} \\ \underline{\text{Standard}} \quad \tilde{V} &= \left(\frac{1}{n-k} \sum_{i=1}^n \hat{\varepsilon}_i^2 \right) \left(\frac{1}{n-k} \sum_{i=1}^n X_i X_i' \right)^{-1} \quad \text{sous hypothèse d'homoscédasticité} \end{aligned}$$

1.3 Modèle I.V - estimateur 2SLS/2MC

↻ *link-2sls*

On suppose $Y = X' \cdot \beta_0 + \varepsilon$ avec **problème d'endogénéité** $E[X\varepsilon] \neq 0$

Variables instrumentales : $X \in \mathbf{R}^K$, $Z \in \mathbf{R}^L$, avec $L \geq K$

1. Exogénéité : $E[Z\varepsilon] = 0$
2. Condition de rang $E[ZX']$ de rang $K \leftrightarrow$ Pertinence $Cov(Z, X^{(i)}) \neq 0 \quad \forall i$

Condition de rang donne qu'il existe Γ tq $\Gamma \cdot E[ZX']$ inversible. *Condition de rang* testable avec first step (significativité d'un des coeff. des vrais instruments) et *exogénéité* pas testable.

Si $L = K$ alors β_0 est *juste identifié*, si $L > K$ alors β_0 est *suridentifié*.

Z inclut toutes les variables exogènes. En pratique on ne régresse et remplace par la projection linéaire sur Z estimée que les variables endogènes.

Estimateur 2MC - 2SLS

1. Projection linéaire de X sur Z $\mapsto X^* = \Gamma Z$
où $\Gamma = E[XZ']E[ZZ']^{-1} = (\beta_{(1)/Z}^{OLS'}, \dots, \beta_{(K-1)/Z}^{OLS'})^T = \text{"}\beta_{X/Z}^{OLS}\text{"}$
2. Reg lin de Y sur $X^* \mapsto$ OLS estimé est $\hat{\beta}_{2SLS}$

$$Y = X' \cdot \beta_0 + \varepsilon$$

$$\beta_0 = E[\Gamma Z X']^{-1} E[\Gamma Z Y] = E[X^* X']^{-1} E[X^* Y]$$

$$\text{or } \langle z, x - p_Z(x) \rangle = 0 \quad \forall z \in Z \Rightarrow z = X^* \in \text{Vect}(Z) \quad E[X^* X^T] = E[X^* X^{*T}]$$

$$\beta_0 = E[X^* X^{*'}]^{-1} E[X^* Y]$$

Estimateur doubles moindres carrés : $\hat{\beta}_{2SLS} \xrightarrow{n \rightarrow +\infty} \beta_0$

$$\hat{\beta}_{2SLS} = \left(\frac{1}{n} \sum_{i=1}^n \hat{X}_i \hat{X}_i' \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n \hat{X}_i Y_i \right) \quad \text{où } \hat{X}_i = \hat{\Gamma} Z_i \quad \hat{\Gamma} = (\hat{\beta}_{(1)/Z}^{OLS'}, \dots, \hat{\beta}_{(K-1)/Z}^{OLS'})^T$$

\Rightarrow L'estimateur $\hat{\beta}_{2SLS}$ est convergent & asymptotiquement normal **mais** pas nécessairement sans biais.

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta_0) \xrightarrow{n \rightarrow +\infty} \mathcal{N}(0, \text{VA}(\hat{\beta}_{2SLS})) = \underline{E[X^* X^{*'}]^{-1} E[\varepsilon^2 X^* X^{*'}] E[X^* X^{*'}]^{-1}}$$

$\hookrightarrow \varepsilon$ estimé par $\hat{\varepsilon}_i = Y_i - \mathbf{X}_i' \cdot \hat{\beta}_{2SLS}$ et $\widehat{\text{VA}}(\hat{\beta}_{2SLS})$ **estimateur convergent robuste** (car ne repose pas sur des hypothèses d'homoscédasticité) de la variance asymptotique. Ce n'est pas l'estimateur qu'on obtient si l'on fait une régression de Y sur \hat{X} car $\hat{\varepsilon}_i \neq Y_i - \hat{X}_i' \cdot \hat{\beta}_{2SLS}$

1.4 Méthode des moments généralisés

☺ *link-gmm*

$(U_i = (Y_i, X_i, Z_i))$ l'ensemble des données sur i

On veut estimer $\theta_0 \in \mathbf{R}^K$ en utilisant : $E[g(U, \theta_0)] = 0 \quad g \mapsto \mathbf{R}^L, \quad L \geq K$

Estimateur GMM pas unique dans le cas $L > K$ car dépend de la matrice de pondération \widehat{W}_n choisie.

Hypothèse d'identification \Rightarrow GMM convergent et asymptotiquement normal.

Choix théorique $W_0 = H^{-1} = V(g(U, \theta_0))^{-1} = E[g(U, \theta_0).g(U, \theta_0)^T]^{-1}$ est **optimal** : inconnu car fonction de $\theta_0 \mapsto W_n = \widehat{W}_0$

GMM est optimal \mapsto on sous-entend qu'on choisit $W_n \xrightarrow{n \rightarrow +\infty} W_0 = H^{-1}$ *matrice de pondération optimale*. **Asymptotiquement efficace/optimal**

Cas exogène : $g(U, \beta) = X(Y - X'\beta)$, $E[g(U, \beta_0)] = 0 \quad L = K$

Tous les choix de \widehat{W}_n *matrice de pondération* (symétrique, définie potentiellement aléatoire) conduisent au même estimateur = MCO

Cas instrumental : $g(U, \beta) = Z(Y - X'\beta)$, $E[g(U, \beta_0)] = 0 \quad Z \in \mathbf{R}^L \quad L > K$

Différents $W_n \longleftrightarrow$ différents estimateurs.

- **Cas Homoscédastique :** l'estimateur GMM coïncide avec l'estimateur des 2MC (2SLS) \Rightarrow 2SLS est optimal dans le cas homoscédastique
- Sinon présence d'**hétéroscédastique** dans les résidus : GMM donne un nouvel estimateur \neq 2SLS qui lui est bien optimal.

1.5 Données de panel

☉ *link-panel*

Données de panel, exemple :

$$murder_{i,t} = \delta_0 + execution_{i,t} \beta_1 + unemp_{i,t} \beta_2 + \alpha_i + \varepsilon_{i,t} \quad \rightarrow \nu_{i,t} = \alpha_i + \varepsilon_{i,t}$$

- α_i effet individuel

Effet constant dans le temps sur la variable expliquée, *tx de meurtre*, qui peut varier entre états. Par exemple, $\alpha_i \leftrightarrow$ loi plus sévère dans l'état du Texas et moins en Virginie donc $\alpha_{Texas} \neq \alpha_{Virginie}$.

On peut donc avoir $E[X_{i,t} \alpha_i] \neq 0$: les décisions d'exécutions sont fortement liées à la sévérité des lois entre états donc $execution_{i,t}$ la variable explicative du nb d'exécutions est corrélée à la sévérité de la loi dans l'état considéré : $E[X_{i,t} \alpha_i] \neq 0 \rightarrow$ endogénéité à cause d'un biais de variable omise.

- $\varepsilon_{i,t}$ choc idiosyncratique

Agrège les facteurs inobservés variables dans le temps.

Hypothèses : on suppose toujours à minima l'exogénéité faible.

1. **Exogénéité des résidus** : $E[X_{i,t} \nu_{i,t}] = 0 \Leftrightarrow E[X_{i,t} \alpha_i] = 0$: **Random Effect**

Si $E[X_{i,t} \alpha_i] \neq 0$: **Fixed Effect**

2. **Exogénéité faible** : $E[X_{i,t} \varepsilon_{i,t'}] = 0 \quad \forall t' \geq t$

\hookrightarrow Variables explicatives non corrélées aux chocs futurs

3. **Exogénéité forte/stricte** : $E[X_{i,t} \varepsilon_{i,t'}] = 0 \quad \forall t' \neq t$

\hookrightarrow Variables explicatives non corrélées aux chocs passés et futurs

Exogénéité faible : taux d'exécution présent à priori non corrélé aux chocs futurs - non anticipables - du taux de meurtre.

Exogénéité forte : taux de meurtre passé et futur n'influencent pas le nombre d'exécutions présent : ce qui peut apparaître comme peu réaliste...

1.6 Modèles binaires

○ *link-binaire*

1.7 Modèles de censure et sélection

→ Modèle de censure : Tobit 1

↪ *link-tobit1*

Modèle Tobit.I

$$Y^* = X' \cdot \beta_0 + \sigma_0 \cdot \varepsilon$$

$\varepsilon|X \sim \mathcal{N}(0, 1) \implies \varepsilon \perp\!\!\!\perp X$ et $\varepsilon \sim \mathcal{N}(0, 1)$ le modèle suppose donc homoscélasticité.

On observe seulement $Y = \max(0, Y^*) = Y^* \cdot \mathbb{1}(Y^* > 0) \leftrightarrow$ **censure**

\implies Estimation par maximum de vraisemblance pour estimer les paramètres : $(\hat{\beta}, \hat{\sigma})_{MLE}$

Un seul mécanisme détermine la valeur de Y et si on observe la variable d'intérêt ou non.

Deux cas à distinguer : censure et solution en coin.

Régression de Y sur X par la méthode des MCO ne conduit pas à un estimateur convergent de β_0 en général.

La simple régression sur les données non censurées seules (ie Y^*) ne conduit pas à un estimateur convergent de β_0

Dans les deux cas on a un biais d'atténuation vers 0.

Interprétation causale des différents coefficients estimés si le modèle est bien spécifié et si le terme d'erreur suit bien une normale centrée réduite.

→ Modèle de sélection : Tobit 2

↪ *link-tobit2*

On s'intéresse à des situations où on observe Y seulement si $D = 1$. Ici on a bien deux processus différents : un qui détermine Y^* **et un autre** qui détermine si on observe cette valeur ou non i.e. modèle sur D . *Variable potentielle Y^* existe pour tous les individus.*

Modèle Tobit.II

$$Y^* = X' \cdot \beta_0 + \varepsilon \quad \text{distribution de } \varepsilon \text{ non spécifiée}$$

$Y = D \cdot Y^*$ est seulement observé.

Sélection exogène : $Y^* \perp\!\!\!\perp D|X \Leftrightarrow \varepsilon \perp\!\!\!\perp D|X$

Sélection exogène : conditionnellement aux variables explicatives observées la variable d'intérêt Y^* est indépendante de la variable de sélection D .

\implies la loi de $Y^*|X, D = 1$ est identique à celle de $Y^*|X \Rightarrow$ on peut **ignorer le problème de sélection**. Avec $E[\varepsilon|X] = 0$ l'estimateur OLS de Y sur X sur le sous-échantillon $\{i|D_i = 1\}$ converge vers β_0

Sélection endogène - Modèle de sélection généralisé Tobit.II $Y^* \not\perp D|X$

On dispose d'un instrument corrélé à D qui n'affecte pas directement Y^* (au moins une composante de Z exclue de X = vrai instrument qui n'a pas d'effet direct sur Y^*)

$$\begin{cases} Y^* = X' \cdot \beta_0 + \varepsilon & : \quad Y = D \cdot Y^* \text{ observé} \\ D = \mathbb{1}(Z' \cdot \gamma_0 + \eta \geq 0) \end{cases}$$

Hypothèses :

$$\begin{cases} (\varepsilon, \eta) \perp (X, Z) \rightarrow \text{homoscédasticité des résidus} \\ \eta \sim \mathcal{N}(0, 1) \\ \hookrightarrow \text{Probit de } D \text{ sur } Z \text{ bien spécifié !} \\ E[\varepsilon|\eta] = \delta_0 \cdot \eta \rightarrow \delta_0 = \text{Cov}(\varepsilon, \eta) \\ \hookrightarrow \text{Sélection endogène} \Leftrightarrow \varepsilon \text{ corrélé à } \eta \Leftrightarrow \delta_0 \neq 0 \\ \text{Sélection significativement endogène ?} \Leftrightarrow \text{regarder significativité dans 2.} \end{cases}$$

\Rightarrow **Méthode de Heckit** : \odot two-steps stata

1. Estimer le probit de D_i sur $Z_i \Rightarrow \hat{\gamma}$
 2. Régresser Y_i sur $(X_i, \lambda(Z_i' \cdot \hat{\gamma}))$ sur l'échantillon sélectionné $\{i | D_i = 1\} \Rightarrow \hat{\beta} \text{ et } \hat{\delta}$
- \Rightarrow Estimateurs $(\hat{\beta}, \hat{\gamma}, \hat{\delta})$ convergents et asymptotiquement normaux

Formellement : *sélection exogène* $Y^* \perp D|(X, Z) \Leftrightarrow \varepsilon \perp D|(\eta, Z)$

$\lambda = \frac{\varphi}{\phi} = \frac{\text{densité}}{\text{FdR}}$ inverse du ratio de Mills ; si $Z \sim \mathcal{N}(0, 1) \Rightarrow E[Z|Z > c] = \frac{\varphi(c)}{1 - \phi(c)}$

2 XXX

2.1 xxx