ÉCONOMÉTRIE (UGA, S2)

CHAPITRE 2: MODÈLE DE RÉGRESSION LINÉAIRE ET ESTIMATION PAR MCO

Michal W. Urdanivia*

*Université de Grenoble Alpes, Faculté d'Économie, GAEL,
e-mail: michal.wong-urdanivia@univ-grenoble-alpes.fr

29 janvier 2024

Contenu

1. Le modèle de régression linéaire

2. Trois méthodes pour calculer l'estimateur des MCO

3. Propriétés asymptotiques de l'estimateur des MCO

4. Identification des paramètres d'un modèle de régression linéaire

5. La notion de projection linéaire

Objectifs de ce chapitre

1. Rappeler les propriétés asymptotiques de l'estimateur des MCO

2. Donner « le point de vue des économètres » sur l'estimateur des MCO

L'estimateur des MCO est un cas particulier de la Méthode des Moments

3. Examiner un point important lié à l'identification des paramètres d'un modèle de régression :

Les conditions requises sur le PG de $\mathbf{x}_i$, dites ***conditions de rang***.

4. Introduire la notion de projection linéaire d'une variable aléatoire sur un vecteur de variables aléatoires

Et « s'entraîner » pour la suite.

Outline

1. Le modèle de régression linéaire

2. Trois méthodes pour calculer l'estimateur des MCO

3. Propriétés asymptotiques de l'estimateur des MCO

4. Identification des paramètres d'un modèle de régression linéaire

5. La notion de projection linéaire

1. Le modèle de régression linéaire

1. Le modèle de régression linéaire

On considère ici un modèle de forme linéaire général :

$$y_i = \mathbf{x}_i'\mathbf{a}_0 + u_i \quad \text{avec} \quad E[u_i] \equiv 0,$$

qui est un modèle de régression si on suppose que la condition d'exogénéité de $\mathbf{x}_i$ est vérifiée.

Définition. $\mathbf{x}_i$ ***est exogène par rapport à*** u_i si $E[u_i / \mathbf{x}_i] = E[u_i] \equiv 0$.

Définition. ***Modèle de régression linéaire :***

$$y_i = \mathbf{x}_i'\mathbf{a}_0 + u_i \quad \text{avec} \quad E[u_i / \mathbf{x}_i] = E[u_i] \equiv 0$$

Rmq. On utilise ici $E[u_i / \mathbf{x}_i] = 0$ plutôt que $E[\mathbf{x}_i u_i] = Cov[\mathbf{x}_i ; u_i] = \mathbf{0}$ pour définir l'exogénéité de $\mathbf{x}_i$. C'est la pratique usuelle car $E[u_i / \mathbf{x}_i] = 0$ est nécessaire pour les modèles de forme non linéaire.
Ca facilite également l'analyse de l'hétéroscédasticité potentielle des u_i.

On considèrera ici les deux cas :

Définition. ***Modèle avec termes d'erreur homoscédastiques (conditionnellement à*** $\mathbf{x}_i$***).*** Modèle tel que :

$$E\left[u_i^2 / \mathbf{x}_i\right] = V\left[u_i / \mathbf{x}_i\right] = V\left[u_i\right].$$

On définit alors le paramètre :

$$\sigma_0^2 \equiv V\left[u_i\right] = V\left[u_i / \mathbf{x}_i\right].$$

Définition. ***Modèle avec termes d'erreur hétéroscédastiques***
Modèle tel que :

$$E\left[u_i^2 / \mathbf{x}_i\right] = V\left[u_i / \mathbf{x}_i\right] \neq V\left[u_i\right].$$

Dans ce dernier cas, $V\left[u_i / \mathbf{x}_i\right]$ est potentiellement une fonction de $\mathbf{x}_i$. En toute rigueur on devrait parler de ***modèle à hétéroscédasticité conditionnelle*** (en $\mathbf{x}_i$) ***de forme inconnue***.

La forme linéaire est simple mais peut être utilisée pour représenter des formes de relation relativement complexes :

Modèle avec effets carrés

$$\begin{aligned} y_i &= \alpha_0 + b_{1,0}\tilde{q}_i + b_{2,0}(\tilde{q}_i)^2 + u_i \\ &= \alpha_0 + b_{1,0}\tilde{x}_{1,i} + b_{2,0}\tilde{x}_{2,i} + u_i \end{aligned} \quad \text{avec} \quad \tilde{x}_{1,i} \equiv \tilde{q}_i \text{ et } \tilde{x}_{2,i} \equiv (\tilde{q}_i)^2$$

Modèle quadratique : avec effets carrés et croisés

$$y_i = \alpha_0 + b_{1,0}\tilde{q}_{1,i} + b_{2,0}\tilde{q}_{2,i} + b_{3,0}(\tilde{q}_{1,i})^2 + b_{4,0}(\tilde{q}_{2,i})^2 + b_{5,0}(\tilde{q}_{1,i}\tilde{q}_{2,i}) + u_i$$

Modèle quadratique en* ln*: avec effets carrés et croisés

$$\begin{aligned} \ln w_i = \alpha_0 &+ b_{1,0}\ln\tilde{q}_{1,i} + b_{2,0}\ln\tilde{q}_{2,i} \\ &+ b_{3,0}(\ln\tilde{q}_{1,i})^2 + b_{4,0}(\ln\tilde{q}_{2,i})^2 + b_{5,0}(\ln\tilde{q}_{1,i}\ln\tilde{q}_{2,i}) + u_i \end{aligned}$$

C’est le ***modèle de base*** des économètres, comme il l’est pour les statisticiens. Ce qui compte c’est la ***linéarité dans les paramètres***.

Outline

1. Le modèle de régression linéaire

2. Trois méthodes pour calculer l'estimateur des MCO

3. Propriétés asymptotiques de l'estimateur des MCO

4. Identification des paramètres d'un modèle de régression linéaire

5. La notion de projection linéaire

2. Trois méthodes pour calculer l’estimateur des MCO

2. Trois méthodes pour calculer l'estimateur des MCO

L'estimateur des paramètres du modèle de régression linéaire :

$$y_i = \mathbf{x}_i'\mathbf{a}_0 + u_i = \alpha_0 + \tilde{\mathbf{x}}_i'\mathbf{b}_0 + u_i \quad \text{avec} \quad E\left[u_i / \mathbf{x}_i\right] = E\left[u_i\right] \equiv 0$$

est l'estimateur des MCO.

Cela a été en partie montré dans le chapitre précédent dans le cas simple, on montrera dans ce chapitre que :

- l'estimateur des MCO de $\mathbf{a}_0$ est convergent pour $\mathbf{a}_0$

et :

- l'estimateur des MCO est asymptotiquement normal

grâce à l'exogénéité de $\mathbf{x}_i$, *i.e.* grâce à la condition $E\left[u_i / \mathbf{x}_i\right] = E\left[u_i\right] \equiv 0$.

Trois approches sont disponibles pour calculer l'estimateur des MCO :

$$\boxed{\hat{\mathbf{a}}_N^{MCO} \equiv \left[N^{-1} \sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i' \right]^{-1} N^{-1} \sum_{i=1}^{N} \mathbf{x}_i y_i \,.}$$

(2a) L'approche par les Moindres Carrés (régression)
(2b) L'approche « directe »
(2c) L'approche par la Méthode des Moments (MM)

Remarques importantes

L'approche par la régression (2a) peut être utilisée en dehors de toute considération « statistique » sur la distribution des $(y_i, \mathbf{x}_i)$.
La ***régression*** est, en premier lieu, une ***technique d'ajustement***. Les propriétés statistiques de $\hat{\mathbf{a}}_N^{MCO}$ pouvant être examinées par la suite.

Les approches « directe » (2b) et par la MM (2c) sont directement ***fondées sur les propriétés statistiques*** des $(y_i, \mathbf{x}_i)$.

2.1. Approche par les Moindres Carrés

On a, par définition :

$$\hat{\mathbf{a}}_N^{MCO} \equiv \arg\min_{\mathbf{a}} \sum_{i=1}^{N} (y_i - \mathbf{a}'\mathbf{x}_i)^2 .$$

$\hat{\mathbf{a}}_N^{MCO}$ est donc la valeur de $\mathbf{a} \in \mathbb{R}^K$ qui minimise la somme des carrés des résidus du modèle, $\sum_{i=1}^{N} (y_i - \mathbf{a}'\mathbf{x}_i)^2$, ou de manière équivalente l'erreur quadratique moyenne $N^{-1}\sum_{i=1}^{N} (y_i - \mathbf{a}'\mathbf{x}_i)^2$ de prédiction des y_i par les $\mathbf{a}'\mathbf{x}_i$.

Interprétation. ***Ce critère de construction est un critère d'ajustement***.

On cherche la valeur de $\mathbf{a} \in \mathbb{R}^K$ telle que la prédiction de y_i par $\hat{y}_{i,N} \equiv \mathbf{x}_i'\hat{\mathbf{a}}_N^{MCO}$ soit telle que la somme des carrés des écarts de $\hat{y}_{i,N}$ à y_i, *i.e.* $\sum_{i=1}^{N} (y_i - \hat{y}_{i,N})^2$ soit la plus petite possible.

Les conditions du premier ordre (CO1) du programme de mininmisation :

$$\min_{\mathbf{a}} \sum_{i=1}^{N} (y_i - \mathbf{a}'\mathbf{x}_i)^2$$

sont données par :

$$\frac{\partial \sum_{i=1}^{N} (y_i - \mathbf{x}_i'\hat{\mathbf{a}}_N^{MCO})^2}{\partial \mathbf{a}} = -2\sum_{i=1}^{N} \mathbf{x}_i (y_i - \mathbf{x}_i'\hat{\mathbf{a}}_N^{MCO}) = \mathbf{0},$$

ce qui donne :

$$\sum_{i=1}^{N} \mathbf{x}_i (y_i - \mathbf{x}_i'\hat{\mathbf{a}}_N^{MCO}) = \mathbf{0} \quad \Leftrightarrow \quad \sum_{i=1}^{N} \mathbf{x}_i y_i - \sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i'\hat{\mathbf{a}}_N^{MCO} = \mathbf{0},$$

et donc :

$$\hat{\mathbf{a}}_N^{MCO} = \left[\sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i' \right]^{-1} \sum_{i=1}^{N} \mathbf{x}_i y_i = \left[N^{-1} \sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i' \right]^{-1} N^{-1} \sum_{i=1}^{N} \mathbf{x}_i y_i \,.$$

Rmq. La CO1 donne que $N^{-1}\sum_{i=1}^{N} \mathbf{x}_i (y_i - \mathbf{x}_i'\hat{\mathbf{a}}_N^{MCO}) = \mathbf{0}$, *i.e.* que les résidus de la régression $y_i - \mathbf{x}_i'\hat{\mathbf{a}}_N^{MCO}$ sont orthogonaux aux variables explicatives $\mathbf{x}_i$, ***par construction*** de $\hat{\mathbf{a}}_N^{MCO}$.

Définition. Soit $g(\mathbf{a}) : \mathbb{R}^K \to \mathbb{R}$, le ***gradient*** de $g(\mathbf{a})$ en **a** est donné par :

$$\mathbf{g}(\mathbf{a}) \equiv \frac{\partial g(\mathbf{a})}{\partial \mathbf{a}} \equiv \begin{bmatrix} \partial g(\mathbf{a}) / \partial a_1 \\ \partial g(\mathbf{a}) / \partial a_2 \\ \vdots \\ \partial g(\mathbf{a}) / \partial a_K \end{bmatrix}.$$

On a :

$$\textbf{\textit{Transposition}} : \frac{\partial g(\mathbf{a})}{\partial \mathbf{a}'} = \left[\frac{\partial g(\mathbf{a})}{\partial \mathbf{a}} \right]' = \left[\frac{\partial g(\mathbf{a})}{\partial a_1} \quad \frac{\partial g(\mathbf{a})}{\partial a_2} \quad \cdots \quad \frac{\partial g(\mathbf{a})}{\partial a_K} \right].$$

$$\textbf{\textit{Exemples}} : \frac{\partial(\mathbf{a}'\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial(\mathbf{x}'\mathbf{a})}{\partial \mathbf{x}} = \mathbf{a}, \quad \frac{\partial(\mathbf{x}'\mathbf{A}\mathbf{y})}{\partial \mathbf{x}} = \mathbf{A}\mathbf{y} = \frac{\partial(\mathbf{y}'\mathbf{A}\mathbf{x})}{\partial \mathbf{x}} \text{ et } \frac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}.$$

Définition. Soit $\mathbf{h}(\mathbf{a}):\mathbb{R}^K \to \mathbb{R}^P$, le ***Jacobien*** de $\mathbf{h}(\mathbf{a})$ en $\mathbf{a}$ est donné par :

$$\mathbf{H}(\mathbf{a}) \equiv \frac{\partial \mathbf{h}(\mathbf{a})}{\partial \mathbf{a}'} \equiv \begin{bmatrix} \partial h_1(\mathbf{a})/\partial \mathbf{a}' \\ \partial h_2(\mathbf{a})/\partial \mathbf{a}' \\ \vdots \\ \partial h_P(\mathbf{a})/\partial \mathbf{a}' \end{bmatrix}_{P\times K} = \begin{bmatrix} \partial h_1(\mathbf{a})/\partial a_1 & \partial h_1(\mathbf{a})/\partial a_2 & \cdots & \partial h_1(\mathbf{a})/\partial a_K \\ \partial h_2(\mathbf{a})/\partial a_1 & \partial h_2(\mathbf{a})/\partial a_2 & \cdots & \partial h_2(\mathbf{a})/\partial a_K \\ \vdots & \vdots & \ddots & \vdots \\ \partial h_P(\mathbf{a})/\partial a_1 & \partial h_P(\mathbf{a})/\partial a_2 & \cdots & \partial h_P(\mathbf{a})/\partial a_K \end{bmatrix}.$$

Transposition :

$$\frac{\partial \mathbf{h}(\mathbf{a})'}{\partial \mathbf{a}} = \left[\frac{\partial \mathbf{h}(\mathbf{a})}{\partial \mathbf{a}}\right]' = \begin{bmatrix} \partial h_1(\mathbf{a})/\partial a_1 & \partial h_2(\mathbf{a})/\partial a_1 & \cdots & \partial h_P(\mathbf{a})/\partial a_1 \\ \partial h_1(\mathbf{a})/\partial a_2 & \partial h_2(\mathbf{a})/\partial a_2 & \cdots & \partial h_P(\mathbf{a})/\partial a_2 \\ \vdots & \vdots & \ddots & \vdots \\ \partial h_1(\mathbf{a})/\partial a_K & \partial h_2(\mathbf{a})/\partial a_K & \cdots & \partial h_P(\mathbf{a})/\partial a_K \end{bmatrix}_{K\times P}.$$

Définition. Soit $g(\mathbf{a}):\mathbb{R}^K \to \mathbb{R}$, la ***Hessienne*** de $g(\mathbf{a})$ en $\mathbf{a}$ est donné par :

$$\mathbf{G}(\mathbf{a}) \equiv \frac{\partial g^2(\mathbf{a})}{\partial \mathbf{a} \partial \mathbf{a}'} \equiv \begin{bmatrix} \partial^2 g(\mathbf{a}) \big/ \partial a_1^2 & \partial^2 g(\mathbf{a}) \big/ \partial a_2 \partial a_1 & \cdots & \partial^2 g(\mathbf{a}) \big/ \partial a_K \partial a_1 \\ \partial^2 g(\mathbf{a}) \big/ \partial a_1 \partial a_2 & \partial^2 g(\mathbf{a}) \big/ \partial a_2^2 & \cdots & \partial^2 g(\mathbf{a}) \big/ \partial a_K \partial a_2 \\ \vdots & \vdots & \ddots & \vdots \\ \partial^2 g(\mathbf{a}) \big/ \partial a_1 \partial a_K & \partial^2 g(\mathbf{a}) \big/ \partial a_2 \partial a_K & \cdots & \partial^2 g(\mathbf{a}) \big/ \partial a_K^2 \end{bmatrix}_{K \times K}.$$

On a :

$$\frac{\partial g^2(\mathbf{a})}{\partial \mathbf{a} \partial \mathbf{a}'} = \frac{\partial}{\partial \mathbf{a}'} \left[\frac{\partial g(\mathbf{a})}{\partial \mathbf{a}} \right] = \frac{\partial}{\partial \mathbf{a}} \left[\frac{\partial g(\mathbf{a})}{\partial \mathbf{a}'} \right] = \frac{\partial \mathbf{g}(\mathbf{a})}{\partial \mathbf{a}'} = \frac{\partial \mathbf{g}(\mathbf{a})'}{\partial \mathbf{a}}.$$

Propriété 11.

La ***Hessienne*** de $g(\mathbf{a})$ en $\mathbf{a}$ est une matrice ***carrée*** et ***symétrique***.

Propriété 12. ***Hessienne et concavité/convexité***

La ***Hessienne*** de $g(\mathbf{a})$ en $\mathbf{a}$ est :

(i) définie négative si $g(\mathbf{a})$ est strictement concave en $\mathbf{a}$

(ii) semi-définie négative si $g(\mathbf{a})$ est concave en $\mathbf{a}$

(iii) définie positive si $g(\mathbf{a})$ est strictement convexe en $\mathbf{a}$

(iv) semi-définie positive si $g(\mathbf{a})$ est convexe en $\mathbf{a}$

La somme des carrés des résidus $\sum_{i=1}^{N}(y_i - \mathbf{a}'\mathbf{x}_i)^2$ est convexe en **a**, elle est strictement convexe en **a** ssi $\mathbf{X}'\mathbf{X} = \sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i'$ est inversible.

2.2. Approche « directe »

L'exogénéité des $\mathbf{x}_i$ dans le modèle de régression :

$$y_i = \mathbf{x}_i'\mathbf{a}_0 + u_i \quad \text{avec} \quad E\left[u_i / \mathbf{x}_i\right] = E\left[u_i\right] \equiv 0$$

donne que :

$$E\left[\mathbf{x}_i u_i\right] = Cov\left[\mathbf{x}_i ; u_i\right] = \mathbf{0}.$$

On a alors :

$$E\left[\mathbf{x}_i y_i\right] = E\left[\mathbf{x}_i \mathbf{x}_i'\mathbf{a}_0\right] + E\left[\mathbf{x}_i u_i\right] = E\left[\mathbf{x}_i \mathbf{x}_i'\right]\mathbf{a}_0 + \mathbf{0}$$

et donc :

$$\boxed{\mathbf{a}_0 = E\left[\mathbf{x}_i \mathbf{x}_i'\right]^{-1} E\left[\mathbf{x}_i y_i\right].}$$

Par la LGN et ses variantes, il est facile de construire un estimateur convergent de :

$$\mathbf{a}_0 = E\left[\mathbf{x}_i \mathbf{x}_i'\right]^{-1} E\left[\mathbf{x}_i y_i\right].$$

Il suffit en fait de « remplacer » les termes $E\left[\mathbf{x}_i \mathbf{x}_i'\right]$ et $E\left[\mathbf{x}_i y_i\right]$ par leurs contre-parties empiriques, *i.e.* des moyennes.

On obtient alors :

$$\hat{\mathbf{a}}_N \equiv \left[N^{-1} \sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i' \right]^{-1} N^{-1} \sum_{i=1}^{N} \mathbf{x}_i y_i = \hat{\mathbf{a}}_N^{MCO}.$$

De plus, on sait que :

$$\hat{\mathbf{a}}_N^{MCO} \xrightarrow[N \to +\infty]{p} \mathbf{a}_0 = E\left[\mathbf{x}_i \mathbf{x}_i'\right]^{-1} E\left[\mathbf{x}_i y_i\right],$$

par construction de $\hat{\mathbf{a}}_N = \hat{\mathbf{a}}_N^{MCO}$.

On a en fait ici exploité la forme de $E[\mathbf{x}_i y_i]$ telle qu'elle est donnée par le modèle :

$$E[\mathbf{x}_i y_i] = E[\mathbf{x}_i \mathbf{x}_i']\mathbf{a}_0 .$$

De fait, $E[\mathbf{x}_i y_i]$ n'est pas une covariance, mais c'est quand-même une ***mesure de co-variation*** entre $\mathbf{x}_i$ et y_i.

Shématiquement :

$E[\mathbf{x}_i y_i]$ ***« contient » la même information que*** $Cov[\mathbf{x}_i ; y_i]$.

$E[\mathbf{x}_i \mathbf{x}_i']$ ***« contient » la même information que*** $V[\mathbf{x}_i]$.

2.3. Méthode des Moments, condition d'orthogonalité et Principe d'analogie (ce qui sera utilisé dans toute la suite)

L'exogénéité des $\mathbf{x}_i$ dans le modèle de forme linéaire $y_i = \mathbf{x}_i'\mathbf{a}_0 + u_i$ donne que :

$$E\left[\mathbf{x}_i u_i\right] = \mathbf{0} \text{ car } E\left[u_i / \mathbf{x}_i\right] = 0 \quad \Rightarrow \quad E\left[\mathbf{x}_i u_i\right] = \mathbf{0}.$$

Définition. Par référence aux concepts d'algèbre linéaire la condition $E\left[\mathbf{x}_i u_i\right] = \mathbf{0}$ est appelée ***condition d'orthogonalité entre*** $\mathbf{x}_i$ ***et*** u_i.

Cette condition d'orthogonalité, ou de « non-corrélation », est une ***condition de moment***, c'est une espérance mathématique, qui peut être exploitée en tant que ***condition estimante*** pour $\mathbf{a}_0$ dans le cadre de la ***Méthode des Moments***.

En combinant $E\left[\mathbf{x}_i u_i\right] = \mathbf{0}$ ***et*** $u_i = y_i - \mathbf{x}_i'\mathbf{a}_0$ on obtient :

$$E\left[\mathbf{x}_i (y_i - \mathbf{x}_i'\mathbf{a}_0)\right] = \mathbf{0}.$$

Dans le cas de la Méthode des Moments, le principe d'analogie (de Goldberger) est la traduction formelle de l'idée suivante :

On supppose que $\boldsymbol{\beta}_0$ vérifie la condition de moment $E\left[\mathbf{h}(\mathbf{w}_i;\boldsymbol{\beta}_0)\right]=\mathbf{0}$, dit autrement on a :

$\boldsymbol{\beta}_0$ est la solution en $\boldsymbol{\beta}$ de $E\left[\mathbf{h}(\mathbf{w}_i;\boldsymbol{\beta})\right]=\mathbf{0}$ (***Problème théorique***).

Si on définit l'estimateur de $\boldsymbol{\beta}_0$, $\hat{\boldsymbol{\beta}}_N$ comme la contre-partie empirique du problème précédent, c'est-à-dire :

$\hat{\boldsymbol{\beta}}_N$ est la solution en $\boldsymbol{\beta}$ de $N^{-1}\sum_{i=1}^{N}\mathbf{h}(\mathbf{w}_i;\boldsymbol{\beta})=\mathbf{0}$ (***Problème empirique***)

alors l'estimateur de $\boldsymbol{\beta}_0$, $\hat{\boldsymbol{\beta}}_N$, aura de bonnes propriétés (as.) sous certaines conditions.

Rmq. Beaucoup d'estimateurs (MM, MMG, MV, …) sont construits sur ce principe.

Rmq. Les conditions requises pour le « bon fonctionnement » du principe d'analogie sont de deux ordres.

- Les ***conditions d'identification*** de $\boldsymbol{\beta}_0$ doivent être vérifiées, *i.e.* le critère d'estimation utilisé doit caractériser $\boldsymbol{\beta}_0$ de manière unique. Ici il faut $\boldsymbol{\beta}_0$ soit l'unique solution en $\boldsymbol{\beta}$ de $E\left[\mathbf{h}(\mathbf{w}_i;\boldsymbol{\beta})\right]=\mathbf{0}$. On peut alors espérer que $\hat{\boldsymbol{\beta}}_N$ soit l'unique solution du problème empirique.
- Les ***conditions de régularité*** doivent ensuite garantir que la solution du problème empirique, $\hat{\boldsymbol{\beta}}_N$, converge vers la solution du problème théorique $\boldsymbol{\beta}_0$. Ces conditions concernent $\mathbf{h}(.;.)$, en particulier sa continuité en ses arguments, et la distribution de $\mathbf{w}_i$ et des $\mathbf{h}(\mathbf{w}_i;\boldsymbol{\beta})$, au moins pour des valeurs de $\boldsymbol{\beta}$ au voisinage de $\boldsymbol{\beta}_0$. Schématiquement, il faut pouvoir utiliser la LGN et ses variantes.

Avec la condition d'orthogonalité $E\left[\mathbf{x}_i(y_i - \mathbf{x}_i'\mathbf{a}_0)\right] = \mathbf{0}$, on sait que :

$$\boxed{\mathbf{a}_0 \text{ est la solution en } \mathbf{a} \text{ de } E\left[\mathbf{x}_i(y_i - \mathbf{x}_i'\mathbf{a})\right] = \mathbf{0}.}$$

On suppose ici que $\mathbf{a}_0$ est l'unique solution de ce problème théorique.

L'***application du Principe d'analogie*** définit $\hat{\mathbf{a}}_N^{MM}$, l'estimateur de la MM de $\mathbf{a}_0$ fondé sur $E\left[\mathbf{x}_i(y_i - \mathbf{x}_i'\mathbf{a}_0)\right] = \mathbf{0}$ par :

$$\boxed{\hat{\mathbf{a}}_N^{MM} \text{ est la solution en } \mathbf{a} \text{ de } N^{-1}\sum_{i=1}^{N}\mathbf{x}_i(y_i - \mathbf{x}_i'\mathbf{a}) = \mathbf{0}_{K\times 1}.}$$

Ici, l'application directe du Principe d'analogie ne pose pas de problème car l'équation $N^{-1}\sum_{i=1}^{N}\mathbf{x}_i(y_i - \mathbf{x}_i'\mathbf{a}) = \mathbf{0}_{K\times 1}$ définit ***un système de K équations*** ($\dim \mathbf{x}_i = K$) ***linéaires*** (en **a**) ***à K inconnues*** (les K éléments de **a**).

L'estimateur $\hat{\mathbf{a}}_N^{MM}$ est ***caractérisé par l'équation*** :

$$\boxed{N^{-1}\sum_{i=1}^{N}\mathbf{x}_i\,(y_i-\mathbf{x}_i'\hat{\mathbf{a}}_N^{MM})=\mathbf{0}_{K\times 1}}$$

Après quelques manipulations, on obtient :

$$\boxed{\hat{\mathbf{a}}_N^{MM}=\left[N^{-1}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i'\right]^{-1}N^{-1}\sum_{i=1}^{N}\mathbf{x}_i y_i}$$

et donc :

$$\boxed{\hat{\mathbf{a}}_N^{MM}=\hat{\mathbf{a}}_N^{MCO}.}$$

En résumé, on dispose de 3 approches qui montrent que l'estimateur des MCO est *a priori* un bon estimateur de $\mathbf{a}_0$ dans le modèle de régression :

$$y_i = \mathbf{x}_i'\mathbf{a}_0 + u_i \quad \text{avec} \quad E\left[u_i / \mathbf{x}_i\right] = E\left[u_i\right] \equiv 0.$$

- La première, par les ***Moindres Carrés***, est une ***méthode d'ajustement***.

- Les deux autres, la méthode « directe » et la MM, utilisent des moments, $E\left[\mathbf{x}_i y_i\right] = E\left[\mathbf{x}_i\mathbf{x}_i'\right]\mathbf{a}_0$ et $E\left[\mathbf{x}_i(y_i - \mathbf{x}_i'\mathbf{a}_0)\right] = \mathbf{0}$. Elles sont ***directement fondées sur des propriétés statistiques du PG des*** $(y_i, \mathbf{x}_i)$. Dans les deux cas on utilise la ***condition d'orthogonalité*** $E\left[\mathbf{x}_i u_i\right] = \mathbf{0}$, ce qui illustre le rôle fondamental de l'***exogénéité de*** $\mathbf{x}_i$.

On retrouve l'importance de la ***condition d'orthogonalité*** $E\left[\mathbf{x}_i u_i\right] = \mathbf{0}$ pour démontrer que $\hat{\mathbf{a}}_N^{MCO}$ a de bonnes propriétés. C'est par exemple une condition de la convergence de $\hat{\mathbf{a}}_N^{MCO}$ dans un modèle linéaire $y_i = \mathbf{x}_i'\mathbf{a}_0 + u_i$.

Outline

1. Le modèle de régression linéaire

2. Trois méthodes pour calculer l'estimateur des MCO

3. Propriétés asymptotiques de l'estimateur des MCO

4. Identification des paramètres d'un modèle de régression linéaire

5. La notion de projection linéaire

3. Propriétés asymptotiques de l’estimateur des MCO

3. Propriétés asymptotiques de l’estimateur des MCO

L’objet de cette partie est triple :

- (A nouveau) Montrer que ***l’estimateur des MCO est convergent*** pour les paramètres d’un ***modèle de régression.***

- Montrer que l’estimateur des MCO est un estimateur ***asymptotiquement normal*** pour les paramètres d’un modèle de régression

- Montrer comment ***calculer la distribution as.*** de l’estimateur des MCO, dans les cas « homoscédastique » et « hétéroscédastique ».

On ne rappelle pas les propriétés à distance finie de l’estimateur des MCO (qui doivent maintenant être bien connue). Elles sont très importantes, même pour les économètres qui travaillent essentiellement « en asymptotique ».

Même si cela a déjà été fait, on démontre ici la propriété suivante.

Propriété 13.

Convergence de $\hat{\mathbf{a}}_N^{MCO}$ dans un modèle de régression

Soit $\{(y_i, \mathbf{x}_i); i = 1, 2, ..., N\}$ un échantillon de variables aléatoires réelles telles que :

$$y_i = \mathbf{x}_i'\mathbf{a}_0 + u_i \quad \text{avec} \quad E[u_i / \mathbf{x}_i] = E[u_i] \equiv 0.$$

L'estimateur des MCO de $\mathbf{a}_0$:

$$\hat{\mathbf{a}}_N^{MCO} \equiv \left[N^{-1} \sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i' \right]^{-1} N^{-1} \sum_{i=1}^{N} \mathbf{x}_i y_i$$

(*i*) existe avec une probabilité approchant 1

et :

(*ii*) est convergent, *i.e.* :

$$\hat{\mathbf{a}}_N^{MCO} \xrightarrow[N \to +\infty]{p} \mathbf{a}_0 .$$

On a :

$$\hat{\mathbf{a}}_N^{MCO} \equiv \left[N^{-1}\sum\nolimits_{i=1}^{N} \mathbf{x}_i\mathbf{x}_i' \right]^{-1} N^{-1}\sum\nolimits_{i=1}^{N} \mathbf{x}_i y_i \,.$$

Avec $y_i = \mathbf{x}_i'\mathbf{a}_0 + u_i$, on a :

$$\hat{\mathbf{a}}_N^{MCO} = \left[N^{-1}\sum\nolimits_{i=1}^{N} \mathbf{x}_i\mathbf{x}_i' \right]^{-1} N^{-1}\sum\nolimits_{i=1}^{N} \mathbf{x}_i (\mathbf{x}_i'\mathbf{a}_0 + u_i)$$

et, après simplifications :

$$\boxed{\hat{\mathbf{a}}_N^{MCO} = \mathbf{a}_0 + \left[N^{-1}\sum\nolimits_{i=1}^{N} \mathbf{x}_i\mathbf{x}_i' \right]^{-1} N^{-1}\sum\nolimits_{i=1}^{N} \mathbf{x}_i u_i \,.}$$

On sait, par la LGN que :

$$N^{-1}\sum\nolimits_{i=1}^{N} \mathbf{x}_i\mathbf{x}_i' \xrightarrow[N\to+\infty]{p} E\left[\mathbf{x}_i\mathbf{x}_i'\right] \quad \text{et} \quad N^{-1}\sum\nolimits_{i=1}^{N} \mathbf{x}_i u_i \xrightarrow[N\to+\infty]{p} E\left[\mathbf{x}_i u_i\right].$$

L'exogénéité des $\mathbf{x}_i$ donne également que :

$$E\left[\mathbf{x}_i u_i\right] = \mathbf{0}$$

On sait, par la propriété 9 que l'inverse de $N^{-1}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i'$ existe avec une probabilité approchant 1, et que :

$$\left[N^{-1}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i'\right]^{-1} \xrightarrow[N\to+\infty]{p} E\left[\mathbf{x}_i\mathbf{x}_i'\right]^{-1}.$$

Ceci donne, par la propriété 10, que :

$$\left[N^{-1}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i'\right]^{-1} N^{-1}\sum_{i=1}^{N}\mathbf{x}_i u_i \xrightarrow[N\to+\infty]{p} E\left[\mathbf{x}_i\mathbf{x}_i'\right]^{-1}\times\mathbf{0}=\mathbf{0}$$

et donc que :

$$\hat{\mathbf{a}}_N^{MCO} \xrightarrow[N\to+\infty]{p} \mathbf{a}_0+\mathbf{0}=\mathbf{a}_0$$

On démontre maintenant la propriété suivante.

Propriété 14.

Normalité asymptotique de $\hat{\mathbf{a}}_N^{MCO}$ dans un modèle de régression

Soit $\{(y_i, \mathbf{x}_i); i = 1, 2, ..., N\}$ un échantillon de variables aléatoires réelles telles que $y_i = \mathbf{x}_i' \mathbf{a}_0 + u_i$ et $E[u_i / \mathbf{x}_i] = E[u_i] \equiv 0$. L'estimateur des MCO de $\mathbf{a}_0$:

$$\hat{\mathbf{a}}_N^{MCO} \equiv \left[N^{-1} \sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i' \right]^{-1} N^{-1} \sum_{i=1}^{N} \mathbf{x}_i y_i$$

vérifie :

$$\sqrt{N} (\hat{\mathbf{a}}_N^{MCO} - \mathbf{a}_0) \xrightarrow[N \to +\infty]{L} \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_0)$$

avec :

$$\mathbf{\Sigma}_0 = E[\mathbf{x}_i \mathbf{x}_i']^{-1} E[u_i^2 \mathbf{x}_i \mathbf{x}_i'] E[\mathbf{x}_i \mathbf{x}_i']^{-1}, \text{ cas « hétéroscédastique » (général)}$$

et :

$$\mathbf{\Sigma}_0 = \sigma_0^2 E[\mathbf{x}_i \mathbf{x}_i']^{-1} \text{ cas « homoscédastique » où } \sigma_0^2 \equiv E[u_i^2 / \mathbf{x}_i] = E[u_i^2].$$

On utilise ici directement le fait que :

$$\hat{\mathbf{a}}_N^{MCO} = \mathbf{a}_0 + \left[N^{-1} \sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i' \right]^{-1} N^{-1} \sum_{i=1}^{N} \mathbf{x}_i u_i$$

qui implique que :

$$\sqrt{N}(\hat{\mathbf{a}}_N^{MCO} - \mathbf{a}_0) = \left[N^{-1} \sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i' \right]^{-1} \sqrt{N} \left[N^{-1} \sum_{i=1}^{N} \mathbf{x}_i u_i \right].$$

On sait, d'après la démonstration précédente, que l'inverse de $N^{-1} \sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i'$ existe avec une probabilité approchant 1, et que :

$$\left[N^{-1} \sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i' \right]^{-1} \xrightarrow[N \to +\infty]{p} E\left[\mathbf{x}_i \mathbf{x}_i' \right]^{-1}.$$

Le TCL implique que :

$$\sqrt{N} \left[N^{-1} \sum_{i=1}^{N} \mathbf{x}_i u_i \right] \xrightarrow[N \to +\infty]{L} \mathcal{N}\left(E\left[\mathbf{x}_i u_i \right], E\left[u_i^2 \mathbf{x}_i \mathbf{x}_i' \right] \right)$$

avec :

$$E\left[\mathbf{x}_i u_i \right] = \mathbf{0} \text{ et } V\left[\mathbf{x}_i u_i \right] = E\left[(\mathbf{x}_i u_i)(\mathbf{x}_i u_i)' \right] = E\left[u_i^2 \mathbf{x}_i \mathbf{x}_i' \right].$$

Avec un abus dans les notations on a :

$$\sqrt{N}(\hat{\mathbf{a}}_N^{MCO} - \mathbf{a}_0) = \left[N^{-1} \sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i' \right]^{-1} \sqrt{N} \left[N^{-1} \sum_{i=1}^{N} \mathbf{x}_i u_i \right]$$

$$\xrightarrow[N \to +\infty]{L} E\left[\mathbf{x}_i \mathbf{x}_i'\right]^{-1} \times \mathcal{N}\left(\mathbf{0}, E\left[u_i^2 \mathbf{x}_i \mathbf{x}_i'\right]\right)$$

et donc (en utilisant $V[\mathbf{x}] = \mathbf{\Omega} \Rightarrow V[\mathbf{Mx}] = \mathbf{M\Omega M}'$ et la symétrie de $E[\mathbf{x}_i \mathbf{x}_i']$) :

$$\sqrt{N}(\hat{\mathbf{a}}_N^{MCO} - \mathbf{a}_0) \xrightarrow[N \to +\infty]{L} \mathcal{N}\left(\mathbf{0}, E\left[\mathbf{x}_i \mathbf{x}_i'\right]^{-1} E\left[u_i^2 \mathbf{x}_i \mathbf{x}_i'\right] E\left[\mathbf{x}_i \mathbf{x}_i'\right]^{-1}\right).$$

La loi approchée de $\hat{\mathbf{a}}_N^{MCO}$ est donc donnée par :

$$\hat{\mathbf{a}}_N^{MCO} \sim_{app.} \mathcal{N}\left(\mathbf{a}_0, N^{-1} E\left[\mathbf{x}_i \mathbf{x}_i'\right]^{-1} E\left[u_i^2 \mathbf{x}_i \mathbf{x}_i'\right] E\left[\mathbf{x}_i \mathbf{x}_i'\right]^{-1}\right),$$

et on a besoin d'un estimateur de $\mathbf{\Sigma}_0 = E\left[\mathbf{x}_i \mathbf{x}_i'\right]^{-1} E\left[u_i^2 \mathbf{x}_i \mathbf{x}_i'\right] E\left[\mathbf{x}_i \mathbf{x}_i'\right]^{-1}$.

Propriété 15. ***Combinaisons de suites convergentes***

Soient $\{\mathbf{M}_N ; N = 1, 2, ...\}$ une suite de matrices aléatoires et, $\{\mathbf{m}_N ; N = 1, 2, ...\}$ et $\{\mathbf{w}_N ; N = 1, 2, ...\}$ deux suites de vecteurs aléatoires (de tailles conformes aux opérations présentées). Les termes $\mathbf{M}_0$ et $\mathbf{m}_0$ sont des termes réels. Le terme $\mathbf{w}$ est une variable aléatoire. On a alors :

$$(i)\quad \mathbf{m}_N \xrightarrow[N\to+\infty]{p} \mathbf{m}_0 \text{ et } \mathbf{w}_N \xrightarrow[N\to+\infty]{L} \mathbf{w} \quad \Rightarrow \quad \mathbf{m}_N + \mathbf{w}_N \xrightarrow[N\to+\infty]{L} \mathbf{m}_0 + \mathbf{w}$$

$$(ii)\quad \mathbf{M}_N \xrightarrow[N\to+\infty]{p} \mathbf{M}_0 \text{ et } \mathbf{w}_N \xrightarrow[N\to+\infty]{L} \mathbf{w} \quad \Rightarrow \quad \mathbf{M}_N \mathbf{w}_N \xrightarrow[N\to+\infty]{L} \mathbf{M}_0 \mathbf{w}$$

$$(iii)\quad \mathbf{M}_N \xrightarrow[N\to+\infty]{p} \mathbf{M}_0 \text{ et } \mathbf{w}_N \xrightarrow[N\to+\infty]{L} \mathbf{w} \quad \Rightarrow \quad \mathbf{w}'_N \mathbf{M}_N \mathbf{w}_N \xrightarrow[N\to+\infty]{L} \mathbf{w}' \mathbf{M}_0 \mathbf{w}$$

et :

$$(iv)\quad \mathbf{M}_N \xrightarrow[N\to+\infty]{p} \mathbf{0} \text{ et } \mathbf{w}_N \xrightarrow[N\to+\infty]{L} \mathbf{w} \quad \Rightarrow \quad \mathbf{M}_N \mathbf{w}_N \xrightarrow[N\to+\infty]{p} \mathbf{0}.$$

Interprétation. Les combinaisons convergent en loi, le terme le plus « lent » freinant l'ensemble. Le cas (*iv*) est une exception, le terme convergeant en probabilité vers **0** « absorbant » l'ensemble.

Il nous faut maintenant calculer un estimateur convergent de :

$$\mathbf{\Sigma}_0 = E\left[\mathbf{x}_i\mathbf{x}_i'\right]^{-1} E\left[u_i^2\mathbf{x}_i\mathbf{x}_i'\right] E\left[\mathbf{x}_i\mathbf{x}_i'\right]^{-1}.$$

On distingue deux cas, selon l'homo/hétéroscédasticité des u_i

Une simplification est possible dans le ***cas homoscédastique*** car on a $V\left[u_i / \mathbf{x}_i\right] = E\left[u_i^2 / \mathbf{x}_i\right] = \sigma_0^2$. La loi des conditionnements successifs donne alors que :

$$E\left[u_i^2\mathbf{x}_i\mathbf{x}_i'\right] = E\left[u_i^2\right] E\left[\mathbf{x}_i\mathbf{x}_i'\right] = \sigma_0^2 E\left[\mathbf{x}_i\mathbf{x}_i'\right]$$

et donc que :

$$\boxed{\mathbf{\Sigma}_0 = \sigma_0^2 E\left[\mathbf{x}_i\mathbf{x}_i'\right]^{-1}}$$

car :

$$\mathbf{\Sigma}_0 = E\left[\mathbf{x}_i\mathbf{x}_i'\right]^{-1} \sigma_0^2 E\left[\mathbf{x}_i\mathbf{x}_i'\right] E\left[\mathbf{x}_i\mathbf{x}_i'\right]^{-1} = \sigma_0^2 E\left[\mathbf{x}_i\mathbf{x}_i'\right]^{-1}.$$

Cas où les u_i sont homoscédastiques

On a :

$$\mathbf{\Sigma}_0 = \sigma_0^2 E\left[\mathbf{x}_i \mathbf{x}_i'\right]^{-1}$$

On utilise ici directement :

$$\left[N^{-1}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i'\right]^{-1} \xrightarrow[N\to+\infty]{p} E\left[\mathbf{x}_i\mathbf{x}_i'\right]^{-1}.$$

Avec $u_i = y_i - \mathbf{x}_i'\mathbf{a}_0$ on a :

$$\sigma_0^2 = E\left[(y_i - \mathbf{x}_i'\mathbf{a}_0)^2\right]$$

et donc :

$$\hat{\sigma}_N^2 \equiv N^{-1}\sum_{i=1}^{N}(y_i - \mathbf{x}_i'\hat{\mathbf{a}}_N^{MCO})^2 \xrightarrow[N\to+\infty]{p} \sigma_0^2 = E\left[(y_i - \mathbf{x}_i'\mathbf{a}_0)^2\right]$$

puisqu'on sait que $\hat{\mathbf{a}}_N^{MCO} \xrightarrow[N\to+\infty]{p} \mathbf{a}_0$.

Finalement :

$$\boxed{\begin{gathered}\hat{\mathbf{\Sigma}}_N \equiv \hat{\sigma}_N^2\left[N^{-1}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i'\right]^{-1} \xrightarrow[N\to+\infty]{p} \mathbf{\Sigma}_0 \\ \text{avec } \hat{\sigma}_N^2 \equiv N^{-1}\sum_{i=1}^{N}\hat{u}_{i,N}^2 \text{ et } \hat{u}_{i,N} \equiv y_i - \mathbf{x}_i'\mathbf{a}_N^{MCO}\end{gathered}}$$

Cas où les u_i sont hétéroscédastiques

On a :

$$\mathbf{\Sigma}_0 = E\left[\mathbf{x}_i\mathbf{x}_i'\right]^{-1} E\left[u_i^2\mathbf{x}_i\mathbf{x}_i'\right] E\left[\mathbf{x}_i\mathbf{x}_i'\right]^{-1}.$$

On utilise ici directement :

$$\left[N^{-1}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i'\right]^{-1} \xrightarrow[N\to+\infty]{p} E\left[\mathbf{x}_i\mathbf{x}_i'\right]^{-1}.$$

Avec $u_i = y_i - \mathbf{x}_i'\mathbf{a}_0$ on a :

$$E\left[u_i^2\mathbf{x}_i\mathbf{x}_i'\right] = E\left[(y_i - \mathbf{x}_i'\mathbf{a}_0)^2\,\mathbf{x}_i\mathbf{x}_i'\right]$$

et donc :

$$N^{-1}\sum_{i=1}^{N}(y_i - \mathbf{x}_i'\hat{\mathbf{a}}_N^{MCO})^2\,\mathbf{x}_i\mathbf{x}_i' \xrightarrow[N\to+\infty]{p} E\left[u_i^2\mathbf{x}_i\mathbf{x}_i'\right] = E\left[(y_i - \mathbf{x}_i'\mathbf{a}_0)^2\,\mathbf{x}_i\mathbf{x}_i'\right]$$

puisqu'on sait que $\hat{\mathbf{a}}_N^{MCO} \xrightarrow[N\to+\infty]{p} \mathbf{a}_0$.

Finalement :

$$\boxed{\begin{gathered}\hat{\mathbf{\Sigma}}_N^W \equiv \left[N^{-1}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i'\right]^{-1}\left[N^{-1}\sum_{i=1}^{N}\hat{u}_{i,N}^2\mathbf{x}_i\mathbf{x}_i'\right]\left[N^{-1}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i'\right]^{-1} \xrightarrow[N\to+\infty]{p} \mathbf{\Sigma}_0 \\ \text{avec } \hat{u}_{i,N} \equiv y_i - \mathbf{x}_i'\mathbf{a}_N^{MCO}\end{gathered}}$$

Propriété 16. ***Estimateurs de la variance asymptotique de $\hat{\mathbf{a}}_N^{MCO}$ dans un modèle de régression***

La variance asymptotique, $\mathbf{\Sigma}_0$, de l'estimateur des MCO de $\mathbf{a}_0$, $\hat{\mathbf{a}}_N^{MCO}$, peut être estimée par $\hat{\mathbf{\Sigma}}_N$ si $V\left[u_i / \mathbf{x}_i\right] = \sigma_0^2$ (cas homoscédastique) ou par $\hat{\mathbf{\Sigma}}_N^W$ (cas général, hétéroscédastique). En notant, $\hat{u}_{i,N} \equiv y_i - \mathbf{x}_i'\mathbf{a}_N^{MCO}$ le résidu, on a :

$$(i)\quad \hat{\mathbf{\Sigma}}_N^W \equiv \left[N^{-1}\sum_{i=1}^N \mathbf{x}_i\mathbf{x}_i'\right]^{-1}\left[N^{-1}\sum_{i=1}^N \hat{u}_{i,N}^2\mathbf{x}_i\mathbf{x}_i'\right]\left[N^{-1}\sum_{i=1}^N \mathbf{x}_i\mathbf{x}_i'\right]^{-1}$$

et :

$$(ii)\quad \hat{\mathbf{\Sigma}}_N \equiv \hat{\sigma}_N^2\left[N^{-1}\sum_{i=1}^N \mathbf{x}_i\mathbf{x}_i'\right]^{-1} \text{ avec } \hat{\sigma}_N^2 \equiv N^{-1}\sum_{i=1}^N \hat{u}_{i,N}^2 .$$

L'estimateur $\hat{\mathbf{\Sigma}}_N^W$ est dit « ***robuste à l'hétéroscédasticité*** » ou « ***de White*** ».

Rmq. Cet estimateur est très utilisé par les micro-économètres qui travaillent sur des phénomènes hétérogènes.

Outline

1. Le modèle de régression linéaire

2. Trois méthodes pour calculer l'estimateur des MCO

3. Propriétés asymptotiques de l'estimateur des MCO

4. Identification des paramètres d'un modèle de régression linéaire

5. La notion de projection linéaire

4. Identification des paramètres d’un modèle de régression linéaire

4. Identification des paramètres d’un modèle de régression linéaire

Lorsqu’on a utilisé l’approche « par la Méthode des Moments » on utilisé la condition d’orthogonalité :

$$E\left[\mathbf{x}_i (y_i - \mathbf{x}_i' \mathbf{a}_0)\right] = \mathbf{0}$$

pour calculer un estimateur de $\mathbf{a}_0$. Lorsqu’on a utilisé « l’approche directe », on a de fait utilisé cette équation pour montrer que :

$$E\left[\mathbf{x}_i y_i\right] = E\left[\mathbf{x}_i \mathbf{x}_i'\right]\mathbf{a}_0$$

et pour finalement écrire $\mathbf{a}_0$ sous la forme :

$$\mathbf{a}_0 = E\left[\mathbf{x}_i \mathbf{x}_i'\right]^{-1} E\left[\mathbf{x}_i y_i\right].$$

On a en fait négligé un point important : il faut pouvoir « inverser » $E\left[\mathbf{x}_i \mathbf{x}_i'\right]$.

On a pour l'instant considéré que $E\left[\mathbf{x}_i u_i\right]=\mathbf{0}$ était la condition d'identification de $\mathbf{a}_0$ parce qu'elle implique que :

$$E\left[\mathbf{x}_i(y_i-\mathbf{x}_i'\mathbf{a}_0)\right]=\mathbf{0}$$

Mais, si $E\left[\mathbf{x}_i u_i\right]=\mathbf{0}$ est nécessaire pour l'identification de $\mathbf{a}_0$ à partir des $(y_i,\mathbf{x}_i)$, elle n'est pas suffisante.

En fait, la condition $E\left[\mathbf{x}_i(y_i-\mathbf{x}_i'\mathbf{a}_0)\right]=\mathbf{0}$ doit être valide et :

$E\left[\mathbf{x}_i(y_i-\mathbf{x}_i'\mathbf{a}_0)\right]=\mathbf{0}$ doit déterminer $\mathbf{a}_0$ de ***manière unique***.

On montre dans la suite que la condition :

« $E\left[\mathbf{x}_i\mathbf{x}_i'\right]$ est inversible »

garantit ce point.

La condition $E\left[\mathbf{x}_i u_i\right]=\mathbf{0}$ est une ***condition d'identification « générale »***, *i.e.* par rapport au *modèle linéaire* $y_i=\mathbf{x}_i'\mathbf{a}_0+u_i$ et $E\left[u_i\right]\equiv 0$.

La condition « $E\left[\mathbf{x}_i\mathbf{x}_i'\right]$ est inversible » est une ***condition d'identification « interne »***, *i.e.* par rapport au *modèle de régression linéaire* $y_i=\mathbf{x}_i'\mathbf{a}_0+u_i$ et $E\left[u_i/\mathbf{x}_i\right]=E\left[u_i\right]\equiv 0$.

En fait, $\mathbf{a}_0$ est identifiée dans le modèle de régression linéaire (*i.e.* étant entendu que $E\left[\mathbf{x}_i(y_i-\mathbf{x}_i'\mathbf{a}_0)\right]=\mathbf{0}$ est valide) ***si et seulement si*** :

$$E\left[\mathbf{x}_i(y_i-\mathbf{x}_i'\mathbf{a})\right]=\mathbf{0} \quad\Leftrightarrow\quad \mathbf{a}=\mathbf{a}_0,$$

i.e., si $\mathbf{a}_0$ est l'unique solution en $\mathbf{a}$ de $E\left[\mathbf{x}_i(y_i-\mathbf{x}_i'\mathbf{a})\right]=\mathbf{0}$, *i.e.* si l'équation $E\left[\mathbf{x}_i(y_i-\mathbf{x}_i'\mathbf{a}_0)\right]=\mathbf{0}$ ***caractérise*** correctement $\mathbf{a}_0$.

Propriété 17. ***Unicité de la solution d'un système d'équations linéaires***

Soient **M** une matrice réelle ($P \times K$), **x** un vecteur réel ($K \times 1$) et **m** un vecteur réel ($P \times 1$). L'équation :

$$\mathbf{Mx} = \mathbf{m}$$

admet une solution unique en **x** si et seulement si $rang\mathbf{M} = \dim \mathbf{x} = K$ (ce qui suppose $P \geq K$)

Propriété 18. ***Rang d'une matrice carrée (semi-définie positive).***

Soit **M** une matrice réelle carrée de dimension *P*. Les conditions suivantes sont équivalentes :

$$\mathbf{M} \text{ est inversible} \iff rang\mathbf{M} = P$$

Soit **M** une matrice réelle semi-définie positive de dimension *P*. Les conditions suivantes sont équivalentes :

$$rang\mathbf{M} = P \iff \mathbf{M} \text{ est définie positive.}$$

En utilisant ces résultats on montre que $\mathbf{a}_0$ est l'unique solution en $\mathbf{a}$ de $E\left[\mathbf{x}_i(y_i - \mathbf{x}_i'\mathbf{a})\right] = \mathbf{0}$ si et seulement si $E\left[\mathbf{x}_i\mathbf{x}_i'\right]$ est inversible.

Propriété 19.

Identification de $\mathbf{a}_0$ dans un modèle de régression linéaire

Soit $\left\{(y_i, \mathbf{x}_i); i = 1, 2, ..., N\right\}$ un échantillon de variables aléatoires réelles telles que :

$$y_i = \mathbf{x}_i'\mathbf{a}_0 + u_i \quad \text{avec} \quad E\left[u_i / \mathbf{x}_i\right] = E\left[u_i\right] \equiv 0.$$

Le vecteur de paramètres $\mathbf{a}_0$ est identifiable si et seulement si :

$$rangE\left[\mathbf{x}_i\mathbf{x}_i'\right] = K \quad \Leftrightarrow \quad E\left[\mathbf{x}_i\mathbf{x}_i'\right] \text{ est inversible.}$$

Rmq. La condition $rangE\left[\mathbf{x}_i\mathbf{x}_i'\right]=K$ est l'« équivalent asymptotique » de la condition $rang\mathbf{X}=K$ utilisée à distance finie. L'interprétation de la ***condition de rang sur*** $E\left[\mathbf{x}_i\mathbf{x}_i'\right]$ n'est pas immédiate. Notons simplement pour l'instant que avec :

$$\mathbf{x}_i=\begin{bmatrix}1\\ \tilde{\mathbf{x}}_i\end{bmatrix} \quad \text{avec} \quad \tilde{\mathbf{x}}_i=\begin{bmatrix}\tilde{x}_{1,i}\\ \vdots\\ \tilde{x}_{K-1,i}\end{bmatrix}$$

$E\left[\mathbf{x}_i\mathbf{x}_i'\right]$ est inversible

$\Updownarrow$

(i) $V\left[\tilde{x}_{k,i}\right]>0$ pour tout $k=1,\ldots,K-1$

et

(ii) Les éléments de $\tilde{\mathbf{x}}_i$ ne sont pas parfaitement linéairement liés entre eux

$\Updownarrow$

$V\left[\tilde{\mathbf{x}}_i\right]$ est inversible.

Outline

1. Le modèle de régression linéaire

2. Trois méthodes pour calculer l'estimateur des MCO

3. Propriétés asymptotiques de l'estimateur des MCO

4. Identification des paramètres d'un modèle de régression linéaire

5. La notion de projection linéaire

5. La notion de projection linéaire

5. La notion de projection linéaire

La ***projection linéaire d'une variable aléatoire*** sur un vecteur de variables aléatoires est intimement liée au modèle de régression linéaire et à l'estimateur des MCO.

C'est une forme d'***espérance conditionnelle linéaire***.

1. Présentation de la projection linéaire et de ses principales propriétés

2. Ses liens avec le modèle de régression linéaire et les MCO

3. Une première utilisation : analyse de l'inversibilité de $E\left[\mathbf{x}_i\mathbf{x}_i'\right]$.

5.1. La projection linéaire de variables aléatoires et ses propriétés

Propriété 20. ***Projection linéaire de*** y_i ***sur*** $\mathbf{x}_i$

La projection linéaire de y_i sur $\mathbf{x}_i$, notée $EL\left[y_i/\mathbf{x}_i\right]$, est définie par :

$$EL\left[y_i/\mathbf{x}_i\right] \equiv \boldsymbol{\gamma}'\mathbf{x}_i \quad \text{où} \quad \boldsymbol{\gamma} \equiv \arg\min_{\mathbf{g}} E\left[(y_i - \mathbf{g}'\mathbf{x}_i)^2\right].$$

Le paramètre $\boldsymbol{\gamma}$ est unique si et seulement si $E\left[\mathbf{x}_i\mathbf{x}_i'\right]$ est inversible.

Interprétation.

- $EL\left[y_i/\mathbf{x}_i\right]$ est la meilleure prédiction ***linéaire*** de y_i par $\mathbf{x}_i$ au sens de l'espérance de l'erreur quadratique.
- La ***linéarité en*** $\mathbf{x}_i$ distingue $EL\left[y_i/\mathbf{x}_i\right]$ de l'espérance conditionnelle de y_i en $\mathbf{x}_i$, *i.e.* de $E\left[y_i/\mathbf{x}_i\right] \equiv \min_{m(.)} E\left[(y_i - m(\mathbf{x}_i))^2\right]$.

Propriété 21. ***Propriétés de la projection linéaire de*** y_i ***sur*** $\mathbf{x}_i$

La projection linéaire de y_i sur $\mathbf{x}_i$, notée $EL\left[y_i/\mathbf{x}_i\right] \equiv \boldsymbol{\gamma}'\mathbf{x}_i$, vérifie les propriétés suivantes :

(*i*) Le résidu de projection de y_i sur $\mathbf{x}_i$, $e_i \equiv y_i - \boldsymbol{\gamma}'\mathbf{x}_i$, vérifie $E\left[\mathbf{x}_i e_i\right] = \mathbf{0}$.

Si $E\left[\mathbf{x}_i\mathbf{x}_i'\right]$ est inversible alors :

(*ii*) $\boldsymbol{\gamma} = E\left[\mathbf{x}_i\mathbf{x}_i'\right]^{-1} E\left[\mathbf{x}_i y_i\right]$

Si de plus $\mathbf{x}_i \equiv (1, \tilde{\mathbf{x}}_i)$ alors :

(*iii*) $\boldsymbol{\gamma}'\mathbf{x}_i = \delta + \boldsymbol{\lambda}'\tilde{\mathbf{x}}_i$ avec $\boldsymbol{\lambda} = V\left[\tilde{\mathbf{x}}_i\right]^{-1} Cov\left[\tilde{\mathbf{x}}_i, y_i\right]$ et $\delta = E\left[y_i\right] - \boldsymbol{\lambda}'E\left[\tilde{\mathbf{x}}_i\right]$.

Eléments de démonstration des résultats (*i*) **et** (*ii*)

Les CO1 du programme de minimisation définissant $\boldsymbol{\gamma}$:

$$\min_{\mathbf{g}} E\left[(y_i - \mathbf{g}'\mathbf{x}_i)^2\right]$$

sont données par :

$$\frac{\partial E\left[(y_i - \mathbf{x}_i'\boldsymbol{\gamma})^2\right]}{\partial \mathbf{g}} = -2E\left[\mathbf{x}_i(y_i - \mathbf{x}_i'\boldsymbol{\gamma})\right] = \mathbf{0}.$$

On a alors (*i*) puisque :

$$E\left[\mathbf{x}_i(y_i - \mathbf{x}_i'\boldsymbol{\gamma})\right] = \mathbf{0} \quad \text{et} \quad e_i \equiv y_i - \boldsymbol{\gamma}'\mathbf{x}_i$$

donnent :

$$E\left[\mathbf{x}_i e_i\right] = \mathbf{0}.$$

La solution du programme de minimisation est unique si et seulement si $E\left[(y_i - \mathbf{g}'\mathbf{x}_i)^2\right]$ est strictement convexe pour tout **g**, ce qui est le cas si et seulement si la matrice :

$$E\left[\mathbf{x}_i\mathbf{x}_i'\right] = \frac{\partial^2 E\left[(y_i - \mathbf{x}_i'\mathbf{g})^2\right]}{\partial\mathbf{g}\partial\mathbf{g}'}$$

est définie positive, donc si et seulement si elle est inversible.

L'équation $E\left[\mathbf{x}_i(y_i - \mathbf{x}_i'\boldsymbol{\gamma})\right] = \mathbf{0}$ donnent alors (*ii*). On a :

$$E\left[\mathbf{x}_i(y_i - \mathbf{x}_i'\boldsymbol{\gamma})\right] = \mathbf{0} \quad \Leftrightarrow \quad E\left[\mathbf{x}_i y_i\right] = E\left[\mathbf{x}_i\mathbf{x}_i')\right]\boldsymbol{\gamma}$$

et finalement :

$$\boldsymbol{\gamma} = E\left[\mathbf{x}_i\mathbf{x}_i'\right]^{-1} E\left[\mathbf{x}_i y_i\right],$$

lorsque $E\left[\mathbf{x}_i\mathbf{x}_i'\right]$ est inversible.

Rmq. On retrouve ici la condition de rang de l'identification des paramètres d'un modèle de régression.

Interprétations et liens avec le modèle de régression linéaire et les MCO

Propriété 22. ***Projection linéaire et modèle de régression***

$$y_i = \mathbf{x}_i'\mathbf{a}_0 + u_i \text{ avec } E\left[u_i / \mathbf{x}_i\right] = E\left[u_i\right] \equiv 0$$
$$\Updownarrow$$
$$EL\left[y_i / \mathbf{x}_i\right] = E\left[y_i / \mathbf{x}_i\right] = \mathbf{a}_0'\mathbf{x}_i$$

Interprétation.

- La projection de y_i sur $\mathbf{x}_i$ ***coïncide*** avec la partie déterministe d'un modèle de y_i en $\mathbf{x}_i$ ***si et seulement si*** y_i suit un modèle de régression linéaire en $\mathbf{x}_i$.
- Il y a ***deux*** conditions ici :
  - la ***linéarité*** du modèle de y_i en $\mathbf{x}_i$

  <u>et</u>

  - l'***exogénéité*** des $\mathbf{x}_i$.

Propriété 23. ***Projection linéaire et décomposition***

On peut toujours écrire y_i sous la forme :

$$y_i = \boldsymbol{\gamma}'\mathbf{x}_i + e_i \quad \text{avec } E\left[\mathbf{x}_i e_i\right] = \mathbf{0}.$$

Il suffit de choisir $\boldsymbol{\gamma} \equiv \arg\min_{\mathbf{g}} E\left[(y_i - \mathbf{g}'\mathbf{x}_i)^2\right]$.

De plus il est toujours possible de calculer un estimateur convergent de $\boldsymbol{\gamma}$:

$$\hat{\boldsymbol{\gamma}}_N^{MCO} \xrightarrow[N\to\infty]{p} \boldsymbol{\gamma}$$

et donc de $EL\left[y_i / \mathbf{x}_i\right] = \boldsymbol{\gamma}'\mathbf{x}_i$ puisque :

$$\mathbf{x}_i'\hat{\boldsymbol{\gamma}}_N^{MCO} \xrightarrow[N\to\infty]{p} \boldsymbol{\gamma}'\mathbf{x}_i.$$

Il peut éventuellement être nécessaire d'éliminer des éléments linéairement redondants de $\mathbf{x}_i$.

Rmq. Il est important de noter que l'équation :

$$y_i = \boldsymbol{\gamma}'\mathbf{x}_i + e_i \text{ avec } \boldsymbol{\gamma} \equiv \arg\min_{\mathbf{g}} E\left[(y_i - \mathbf{g}'\mathbf{x}_i)^2\right]$$

ne définit pas un modèle de y_i ***en*** $\mathbf{x}_i$.

Cette équation définit une ***décomposition*** de y_i en la somme :

- de sa ***projection linéaire sur*** $\mathbf{x}_i$, $EL\left[y_i / \mathbf{x}_i\right] = \boldsymbol{\gamma}'\mathbf{x}_i$

et :

- du ***résidu de cette projection***, $e_i \equiv y_i - EL\left[y_i / \mathbf{x}_i\right]$ qui est orthogonal à $\mathbf{x}_i$ par construction.

Cette décomposition est un ***outil mathématique***.

Propriété 23. ***Projection linéaire et estimateur des MCO***

$$EL\left[y_i / \mathbf{x}_i\right] = \boldsymbol{\gamma}'\mathbf{x}_i \text{ et } y_i = \mathbf{x}_i'\mathbf{a}_0 + u_i \text{ avec } E\left[u_i\right] \equiv 0 \quad \Rightarrow \quad \hat{\mathbf{a}}_N^{MCO} \xrightarrow[N\to\infty]{p} \boldsymbol{\gamma}$$

Estimer les paramètres d'un modèle linéaire par les MCO fournit toujours un estimateur du paramètre de la projection linéaire de y_i sur $\mathbf{x}_i$:

$$\hat{\mathbf{a}}_N^{MCO} = \hat{\boldsymbol{\gamma}}_N^{MCO} \xrightarrow[N\to\infty]{p} \boldsymbol{\gamma}.$$

De fait, par la définition de $EL\left[y_i / \mathbf{x}_i\right] = \boldsymbol{\gamma}'\mathbf{x}_i$, le terme $\mathbf{x}_i'\hat{\boldsymbol{\gamma}}_N^{MCO}$ est ***le meilleur prédicteur de*** y_i ***linéaire en*** $\mathbf{x}_i$.

Mais $\hat{\mathbf{a}}_N^{MCO}$ n'est pas toujours un estimateur conv. du paramètre « causal » $\mathbf{a}_0$.

Identification d'effets causaux ≠ Ajustement ou prédiction

Les méthodes de régression ne fournissent des estimateurs « non biaisés » de paramètres causaux ***que si*** le modèle considéré est un modèle de régression

5.2. Analyse/interprétation des conditions de rang des matrices de variance

La condition nécessaire et suffisante d'identification du vecteur de paramètres d'un modèle de régression linéaire en $\mathbf{x}_i$ est donnée par :

$$\boxed{V\left[\tilde{\mathbf{x}}_i\right] \text{ est inversible} \quad \Leftrightarrow \quad rangE\left[\mathbf{x}_i\mathbf{x}_i'\right] = K \quad \Leftrightarrow \quad E\left[\mathbf{x}_i\mathbf{x}_i'\right] \text{ est inversible.}}$$

On va utiliser les projections linéaires pour montrer ce que ces conditions signifient « concrètement ».

Dans la suite on notera :

Définitions.

(i) Le vecteur $\tilde{\mathbf{x}}_{-k,i}$ est le vecteur $\tilde{\mathbf{x}}_i$ amputé de $\tilde{x}_{k,i}$

(ii) Le vecteur $\mathbf{x}_{-k,i}$ est le vecteur $\mathbf{x}_i$ amputé de $x_{k,i}$

$$\tilde{\mathbf{x}}_{-k,i} \equiv \begin{bmatrix} \tilde{x}_{1,i} \\ \\ \tilde{x}_{k-1,i} \\ \tilde{x}_{k+1,i} \\ \\ \tilde{x}_{K-1,i} \end{bmatrix} \quad \text{et} \quad \mathbf{x}_{-k,i} \equiv \begin{bmatrix} x_{1,i} \\ \\ x_{k-1,i} \\ x_{k+1,i} \\ \\ x_{K,i} \end{bmatrix}$$

Définition. ***Partie spécifique de*** $\tilde{x}_{k,i}$ ***dans*** $\tilde{\mathbf{x}}_i$

Une variable $\tilde{x}_{k,i}$ d'un vecteur $\tilde{\mathbf{x}}_i$ peut toujours être décomposée de la manière suivante :

$$\tilde{x}_{k,i} = EL\left[\tilde{x}_{k,i} \middle/ 1, \tilde{\mathbf{x}}_{-k,i}\right] + \tilde{e}_{k,i} \quad \text{où } \tilde{e}_{k,i} \equiv \tilde{x}_{k,i} - EL\left[\tilde{x}_{k,i} \middle/ 1, \tilde{\mathbf{x}}_{-k,i}\right].$$

On nomme ici le ***résidu de la projection*** de $\tilde{x}_{k,i}$ sur $(1, \tilde{\mathbf{x}}_{-k,i})$, le terme $\tilde{e}_{k,i}$, ***la partie « spécifique » de*** $\tilde{x}_{k,i}$ ***dans*** $\tilde{\mathbf{x}}_i$.

Avec :

$$EL\left[\tilde{x}_{k,i} \middle/ 1, \tilde{\mathbf{x}}_{-k,i}\right] \equiv \delta_k + \boldsymbol{\lambda}'_k \tilde{\mathbf{x}}_{-k,i},$$

on a :

$$\tilde{x}_{k,i} = \underbrace{\delta_k + \boldsymbol{\lambda}'_k \tilde{\mathbf{x}}_{-k,i}}_{\substack{\text{Partie de } \tilde{x}_{k,i} \\ \text{(potentiellement) corrélée à } (1, \tilde{\mathbf{x}}_{-k,i})}} + \underbrace{\tilde{e}_{k,i}}_{\text{Partie spécifique de } \tilde{x}_{k,i}}$$

Interprétation

(*i*) La partie spécifique de $\tilde{x}_{k,i}$ dans $\tilde{\mathbf{x}}_i$ et la partie de cette variable qui n'est pas linéairement liée aux autres variables de $\tilde{\mathbf{x}}_i$. En fait, on a :

$$\boxed{E\left[\tilde{\mathbf{x}}_{-k,i}\tilde{e}_{k,i}\right] = Cov\left[\tilde{\mathbf{x}}_{-k,i};\tilde{e}_{k,i}\right] = \mathbf{0}, \text{ par construction}}$$

(*ii*) La ***partie spécifique*** $\tilde{e}_{k,i}$ est ***la « part d'information » spécifique*** à $\tilde{x}_{k,i}$ dans un modèle linéaire en $\tilde{\mathbf{x}}_i$.

(*iii*) On verra dans la suite que ce sont les variations des $\tilde{e}_{\ell,i}$ qui « identifient » les $b_{\ell,0}$ dans le modèle :

$$y_i = \alpha_0 + \sum_{\ell=1}^{K-1} b_{\ell,0}\tilde{x}_{\ell,i} + u_i \quad \text{avec} \quad E\left[u_i\right] \equiv 0.$$

Propriété 25. ***Propriétés des parties spécifiques des éléments de $\tilde{\mathbf{x}}_i$.***

Avec :

$$\tilde{e}_{k,i} \equiv \tilde{x}_{k,i} - EL\left[\tilde{x}_{k,i} \middle/ 1, \tilde{\mathbf{x}}_{-k,i}\right] \text{ pour } k = 1,\ldots,K-1$$

on a :

(i) Espérance nulle et orthogonalité par rapport à $\tilde{\mathbf{x}}_{-k,i}$:

$$E\left[\tilde{e}_{k,i}\right] = 0 \text{ et } E\left[\tilde{\mathbf{x}}_{-k,i}\tilde{e}_{k,i}\right] = \mathbf{0}$$

(ii) Orthogonalité des parties spécifiques entre elles :

$$E\left[\tilde{e}_{k,i}\tilde{e}_{\ell,i}\right] = Cov\left[\tilde{e}_{k,i};\tilde{e}_{\ell,i}\right] = 0 \text{ si } k \neq \ell$$

(iii) Partie spécifique et multicolinéarité :

S'il existe $\mathbf{m}_k$ et η_k tels que $\tilde{x}_{k,i} = \eta_k + \mathbf{m}_k'\tilde{\mathbf{x}}_{-k,i}$ alors $\tilde{e}_{k,i} = 0$

Interprétations.

Résultat (i). Ce résultat est une conséquence de la définition des $\tilde{e}_{k,i}$ en tant que résidus de projections linéaires. Il détermine les deux autres résultats.

Résultat (ii). La partie spécifique d'une variable n'est pas corrélée à la partie spécifique d'une autre variable du vecteur (sinon elle ne serait pas spécifique !).

On a $Cov\left[\tilde{\mathbf{x}}_{-k,i};\tilde{e}_{k,i}\right]=\mathbf{0}$ par construction.

La variable $\tilde{x}_{\ell\neq k,i}$ étant un élément de $\tilde{\mathbf{x}}_{-k,i}$ on a $Cov\left[\tilde{x}_{\ell\neq k,i};\tilde{e}_{k,i}\right]=0$.

Le terme $\tilde{e}_{\ell\neq k,i}$ n'étant qu'une partie de $\tilde{x}_{\ell\neq k,i}$, on a donc $Cov\left[\tilde{e}_{\ell\neq k,i};\tilde{e}_{k,i}\right]=0$.

Résultat (iii). La partie spécifique de $\tilde{e}_{k,i}$ est nulle si $\tilde{x}_{k,i}$ s'exprime comme une fonction affine des éléments de $\tilde{\mathbf{x}}_{-k,i}$ (*multicolinéarité parfaite*)

Propriété 26. ***Conditions de rang sur*** $V\left[\tilde{\mathbf{x}}_i\right]$ et $E\left[\mathbf{x}_i\mathbf{x}_i'\right]$**.**

Avec $\mathbf{x}_i \equiv (1, \tilde{\mathbf{x}}_i)$ et $\tilde{e}_{k,i} \equiv \tilde{x}_{k,i} - EL\left[\tilde{x}_{k,i} / 1, \tilde{\mathbf{x}}_{-k,i}\right]$ pour $k = 1, \ldots, K-1$, les conditions suivantes sont équivalentes :

$$(i) \quad rang E\left[\mathbf{x}_i\mathbf{x}_i'\right] = K$$

$$(ii) \; E\left[\mathbf{x}_i\mathbf{x}_i'\right] \text{ est inversible}$$

$$(iii) \; V\left[\tilde{\mathbf{x}}_i\right] \text{ est inversible}$$

$$(iv) \; V\left[\tilde{e}_{k,i}\right] > 0 \text{ pour } k = 1, \ldots, K-1.$$

Interprétation. ***Chaque variable explicative doit avoir une <u>véritable</u> partie spécifique***.

5.3. Identification et efficacité des estimateurs des MCO dans le modèle de régression linéaire

Identification.

Par définition du modèle linéaire on a :

$$y_i = \mathbf{x}_i'\mathbf{a}_0 + u_i = \alpha_0 + \tilde{\mathbf{x}}_i'\mathbf{b}_0 + u_i = \alpha_0 + \sum_{k=1}^{K-1} b_{k,0}\tilde{x}_{k,i} + u_i$$

En notant :

$$\tilde{x}_{\ell,i} = \delta_\ell + \boldsymbol{\lambda}_\ell'\tilde{\mathbf{x}}_{-\ell,i} + \tilde{e}_{\ell,i} = \delta_\ell + \sum_{k=1,k\neq\ell}^{K-1} \lambda_{\ell k}\tilde{x}_{k,i} + \tilde{e}_{\ell,i},$$

par substitution, on a :

$$y_i = \alpha_0 + \delta_\ell b_{\ell,0} + \sum_{k=1,k\neq\ell}^{K-1} (b_{k,0} + \lambda_{\ell k} b_{\ell,0})\tilde{x}_{k,i} + b_{\ell,0}\tilde{e}_{\ell,i} + u_i.$$

Interprétation. Ce sont les variations de $\tilde{e}_{\ell,i}$ qui « identifient » $b_{\ell,0}$, le « reste » de $\tilde{x}_{\ell,i}$ ne sert à rien, *i.e.* $\boldsymbol{\lambda}_\ell'\tilde{\mathbf{x}}_{-\ell,i}$ n'est pas spécifique.

Dans la suite on notera :

$$\tilde{\mathbf{e}}_i \equiv \begin{bmatrix} \tilde{e}_{1,i} \\ \tilde{e}_{2,i} \\ \vdots \\ \tilde{e}_{K-1,i} \end{bmatrix}$$

On sait que $V\left[\tilde{\mathbf{e}}_i\right]$ ***est diagonale*** par construction puisque les éléments de $\tilde{\mathbf{e}}_i$ ne sont pas corrélés entre eux.

Propriété 27. « ***Anatomie » de l'estimateur des MCO***

$$b_{\ell,0} = V\left[\tilde{e}_{\ell,i}\right]^{-1} Cov\left[\tilde{e}_{\ell,i}; y_i\right] \text{ pour } \ell = 2,...,K \iff \mathbf{b}_0 = V\left[\tilde{\mathbf{e}}_i\right]^{-1} Cov\left[\tilde{\mathbf{e}}_i; y_i\right]$$

Interprétation

(*i*) L'équivalence entre les deux formulations est due à ce que la matrice $V\left[\tilde{\mathbf{e}}_i\right]$ est diagonale, par construction.

(*ii*) Le terme :

$$Cov\left[\tilde{e}_{\ell,i}; y_i\right] = Cov\left[\tilde{x}_{\ell,i} - EL\left[\tilde{x}_{\ell,i} / 1, \tilde{\mathbf{x}}_{-\ell,i}\right]; y_i\right]$$

est la covariance de $\tilde{x}_{\ell,i}$ et y_i « ***purgée*** » de la partie de $\tilde{x}_{\ell,i}$ corrélée aux autres éléments de $\tilde{\mathbf{x}}_i$.

(*iii*) On retrouve ici l'idée de ***covariance partielle*** associée à la régression. Ces covariances partielles montrant que les MCO mesurent des effets « ***toutes choses égales par ailleurs*** » des variables $\tilde{x}_{\ell,i}$.

(*iv*) Le nom de cette propriété est tiré de Angrist et Pischke (2008).

Efficacité

Dans le cas du modèle de régression linéaire à termes d'erreur homoscédastiques on a :

$$\sqrt{N}(\hat{\mathbf{a}}_N^{MCO} - \mathbf{a}_0) \xrightarrow[N\to+\infty]{L} \mathcal{N}\left(\mathbf{0}, \sigma_0^2 E\left[\mathbf{x}_i\mathbf{x}_i'\right]^{-1}\right),$$

où $\sigma_0^2 E\left[\mathbf{x}_i\mathbf{x}_i'\right]^{-1}$ est une mesure de la précision de $\hat{\mathbf{a}}_N^{MCO}$ lorsque N est très grand, *i.e.* une mesure de l'efficacité asymptotique de $\hat{\mathbf{a}}_N^{MCO}$.

La précision de $\hat{\mathbf{a}}_N^{MCO}$ est d'autant plus grande que :

(*i*) σ_0^2 est petit, *i.e.* les termes d'erreur du modèle sont « petits »

et :

(*ii*) $E\left[\mathbf{x}_i\mathbf{x}_i'\right]^{-1}$ est petite dans le pré-ordre des matrices définies positives, *i.e.* que $E\left[\mathbf{x}_i\mathbf{x}_i'\right]$ est grande dans ce pré-ordre. C'est le cas si les variances $V\left[\tilde{e}_{k,i}\right]$ sont « grandes » (/ échelles de mesure des variables).

En résumé :

Dans un modèle de régression linéaire :

$$y_i = \mathbf{x}_i'\mathbf{a}_0 + u_i \text{ avec } E\left[u_i / \mathbf{x}_i\right] = E\left[u_i\right] \equiv 0$$

On obtiendra une estimation précise de $\mathbf{a}_0$, par $\hat{\mathbf{a}}_N^{MCO}$, si :

(*i*) les termes d'erreur du modèle sont « petits », *i.e.* si le pouvoir explicatif de $\tilde{\mathbf{x}}_i$ pour y_i est grand (*spécification de la forme de* $\tilde{\mathbf{x}}_i'\mathbf{b}_0$),

(*ii*) les éléments de $\tilde{\mathbf{x}}_i$ sont peu corrélés entre eux, *i.e.* si les $\tilde{e}_{k,i}$ sont des « grandes » parts des $\tilde{x}_{\ell,i}$ (*absence de multicolinéarité*)

et :

(*iii*) les $\tilde{e}_{k,i}$ sont suffisamment variables (*pas de* $\tilde{x}_{\ell,i}$ *presque constante*).

Rmq. ***Plans d'expérience***.

La condition $rang E\left[\mathbf{x}_i \mathbf{x}_i'\right] = K$ ($\Leftrightarrow$ $E\left[\mathbf{x}_i \mathbf{x}_i'\right]$ inversible) en théorie asymptotique est analogue à la condition $rang \mathbf{X} = K$ à distance finie lorsque **X** est considérée comme « fixe ».

Des conditions analogues aux conditions (*ii*) et (*iii*) guident la construction des ***plans d'expérience « efficaces »***.

Par exemple en agronomie. Pour l'étude de l'effet de deux facteurs, $\tilde{x}_1$ et $\tilde{x}_2$, sur *y*, il faut :

(*i*) faire suffisamment varier les niveaux des deux facteurs (toutefois dans la limite de validité du modèle)

et :

(*ii*) ne pas faire varier les niveaux des facteurs systématiquement ensemble.

Rmq. ***Diagnostic de problèmes d'identification « empirique ».***

Sauf erreur grossière de spécification du modèle, **X** est toujours de plein rang colonne en pratique en économétrie.

Ceci-dit, il est possible que, sans l'être parfaitement, certains éléments de $\tilde{\mathbf{x}}_i$ soient ***fortement linéairement liés entre eux***. De même, sans être constants, certains éléments de $\tilde{\mathbf{x}}_i$ peuvent être ***très peu variables***.

Cela ne pose pas de problème en théorie, mais pose de ***sérieux problèmes en pratique*** :

- l'estimation de $\mathbf{a}_0$ par $\hat{\mathbf{a}}_N^{MCO}$ est alors ***impossible***

ou

- ***très peu précise*** (et ***très instable*** : retirer quelques observations peut modifier significativement la valeur estimée de $\mathbf{a}_0$).

On s’aperçoit de ce type de problème lorsqu’on programme le calcul de $\hat{\mathbf{a}}_N^{MCO}$ et que :

(*i*) le logiciel refuse d’inverser $\mathbf{X}'\mathbf{X} = \sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}'_i$ (rarement SAS !)

ou :

(*ii*) les valeurs estimées de $\hat{\mathbf{\Sigma}}_N^W$ ou de $\hat{\mathbf{\Sigma}}_N$ sont très grandes (les écarts-types estimés des estimateurs sont « énormes » (SAS : biaisé).

De manière générale, l’écart-type estimé de $\hat{\alpha}_N^{MCO}$ est « énorme » et :

(*i*) si seul celui de $\hat{b}_{k,N}^{MCO}$ est également « énorme », alors la variable $\tilde{x}_{k;i}$ varie trop peu,

(*ii*) si ceux de $\hat{b}_{k,N}^{MCO}$ et $\hat{b}_{\ell,N}^{MCO}$ (voire d’autres) sont également « énormes » alors, soit $\tilde{x}_{k;i}$ et $\tilde{x}_{\ell;i}$ varient trop peu, soit $\tilde{x}_{k;i}$ et $\tilde{x}_{\ell;i}$ sont fortement linéairement liées.

On peut affiner le diagnostic en analysant les relations entre $\tilde{x}_{k;i}$ et $\tilde{x}_{\ell;i}$ par les techniques de projection linéaire.

Rmq. ***Multi-colinéarité.***

Les problèmes de multi-colinéarité sont fréquents en économétrie. Mais les économètres n'utilisent pas procédures statistiques gérant ce problème (PLM, …).

- Par exemple, en économétrie de la consommation on des modèles de la forme :

$$Cons_i = \alpha_0 + b_{1,0}\, prix_{1,i} + b_{2,0}\, prix_{2,i} + b_{3,0} revenu_i + u_i$$

  avec $prix_{1,i}$ et $prix_{2,i}$ très liés. On en retire un des deux ou on abandonne.

- Autre exemple typique de problème d'identification : celui des effets de :

  âge – expérience – génération

- Ces problèmes posent souvent la question de la source de variation des variations explicatives.

Rmq. ***Multi-colinéarité et sources de variation des variables explicatives***

Données de panel, de pseudo-panel, ou « hiérarchisées » ou en « clusters » :

$$y_{it} = \alpha_0 + \alpha_{i,0}\tilde{d}_i + \alpha_{t,0}\tilde{d}_t + b_{1,0}\tilde{x}_{i,t} + b_{2,0}\tilde{x}_i + b_{3,0}\tilde{x}_t + u_{it} \ .$$

- Les $\alpha_{i,0}$ et $b_{2,0}$ pas identifiables séparément ($\tilde{d}_i = 1$ si i, $\tilde{d}_i = 0$ sinon)
- Les $\alpha_{t,0}$ et $b_{3,0}$ pas identifiables séparément ($\tilde{d}_t = 1$ si t, $\tilde{d}_t = 0$ sinon)

$$y_{it} = \alpha_0 + b_{1,0}\tilde{x}_{i,t} + b_{2,0}\tilde{x}_i + b_{3,0}\tilde{x}_t + u_{it} \ .$$

- Les $b_{1,0}$ et $b_{2,0}$ sont difficilement identifiables si (*a*) $\tilde{x}_{i,t}$ varie essentiellement avec *i* et (*b*) $i = 1,\ldots,I$ avec *I* « petit ».
- Les $b_{1,0}$ et $b_{3,0}$ sont difficilement identifiables si (*a*) $\tilde{x}_{i,t}$ varie essentiellement avec *t* et (*b*) $t = 1,\ldots,T$ avec *T* « petit ».

$$y_{it} = \alpha_0 + \alpha_{t,0}\tilde{d}_t + b_{1,0}\tilde{x}_{i,t} + b_{3,0}\tilde{x}_t + u_{it} \ .$$

- (a) Les $\alpha_{t,0}$ et $b_{3,0}$ pas identifiables et (b) les $\alpha_{t,0}$ et $b_{1,0}$ difficilement identifiables si (a) $\tilde{x}_{i,t}$ varie essentiellement avec t et (b) $t = 1,...,T$
- Exemple. y_{it} : production de i en t, $\tilde{x}_{i,t}$: prix agricoles, $\tilde{x}_t$: *trend* de progrès technique et les $\tilde{d}_t$ *dummies* annuelles (climats, chocs politiques) $\Rightarrow$ plus de « structure » (chocs politiques) ou d'info (climat) est nécessaire
- Si $\tilde{x}_{i,t}$: prix agricoles :
  - Variations de $\tilde{x}_{i,t}$ en t : évolution « générale » du marché (chocs politiques)
  - Variations de $\tilde{x}_{i,t}$ en i : spécificités du marché dans lequel i « opère » ou/et qualité du bien produit par i

  $\Rightarrow$ Si qualité problème : $\tilde{x}_{i,t}$ peut être endogène (voire le problème du calcul des indices de « volume » et de prix)

Rmq. ***Exogénéité.***

Aucun logiciel ne détectera si $E\left[u_i / \mathbf{x}_i\right] \neq 0$ dans le modèle linéaire :

$$y_i = \mathbf{x}_i' \mathbf{a}_0 + u_i \text{ avec } E\left[u_i\right] \equiv 0.$$

Si vous demandez à SAS le calcul de $\hat{\mathbf{a}}_N^{MCO}$, il le fera. Que $\hat{\mathbf{a}}_N^{MCO}$ soit biaisé ou pas, ce n'est pas son problème, c'est celui de l'économètre.

<u>C'est ce qu'on examine dans les chapitres suivants.</u>

Autres propriétés utiles des projections linéaires

Propriété 28. ***Projection sur les parties spécifiques des éléments de*** $\tilde{\mathbf{x}}_i$**.**
On suppose que $V\left[\tilde{\mathbf{x}}_i\right]$ est inversible et on note $\mathbf{x}_i'\boldsymbol{\gamma} \equiv \delta + \boldsymbol{\lambda}'\tilde{\mathbf{x}}_i \equiv EL\left[y_i / \mathbf{x}_i\right]$.
Avec :

$$\tilde{e}_{k,i} \equiv \tilde{x}_{k,i} - EL\left[\tilde{x}_{k,i} / 1, \tilde{\mathbf{x}}_{-k,i}\right] \text{ pour } k = 1,\ldots,K-1 \text{ et } \tilde{e}_{k,i} \equiv \begin{bmatrix} \tilde{e}_{1,i} \\ \vdots \\ \tilde{e}_{K-1,i} \end{bmatrix},$$

on a :

(*i*) $\boldsymbol{\gamma}'\mathbf{x}_i = \delta + \boldsymbol{\lambda}'\tilde{\mathbf{x}}_i$ avec $\boldsymbol{\lambda} = V\left[\tilde{\mathbf{x}}_i\right]^{-1} Cov\left[\tilde{\mathbf{x}}_i, y_i\right]$ et $\delta = E\left[y_i\right] - \boldsymbol{\lambda}'E\left[\tilde{\mathbf{x}}_i\right]$,

(*ii*) $\boldsymbol{\lambda} = V\left[\tilde{\mathbf{e}}_i\right]^{-1} Cov\left[\tilde{\mathbf{e}}_i ; y_i\right]$

(*iii*) $EL\left[y_i / 1, \tilde{\mathbf{e}}_i\right] = E\left[y_i\right] + \boldsymbol{\lambda}'\tilde{\mathbf{e}}_i = EL\left[y_i / \mathbf{x}_i\right]$

Cette propriété indique que la projection de y_i sur $\mathbf{x}_i$ ou sur $(1, \tilde{\mathbf{e}}_i)$ produit la même projection, car $(1, \tilde{\mathbf{e}}_i)$ est une combinaison linéaire des éléments de $\mathbf{x}_i$.

Résumé du Chapitre

(*i*) Le modèle de y_i linéaire en $\tilde{\mathbf{x}}_i$:

$$y_i = \alpha_0 + \tilde{\mathbf{x}}_i'\mathbf{b}_0 + u_i = \mathbf{x}_i'\mathbf{a}_0 + u_i \text{ avec } E[u_i] \equiv 0.$$

est un ***modèle de régression*** si et seulement si :

$$\boxed{E[u_i / \mathbf{x}_i] = 0.}$$

Cette condition doit être examinée théoriquement, par analyse du PG commun des $(y_i, \mathbf{x}_i)$.

(*ii*) Le paramètre $\mathbf{a}_0$ est ***identifiable*** dans ce modèle de régression si et seulement si :

$$\boxed{V[\tilde{\mathbf{x}}_i] \text{ est inversible } \Leftrightarrow \; rangE[\mathbf{x}_i\mathbf{x}_i'] = K \; \Leftrightarrow \; E[\mathbf{x}_i\mathbf{x}_i'] \text{ est inversible.}}$$

Cette condition peut être vérifiée empiriquement.

De fait, si les conditions précédentes sont satisfaites alors $y_i = \mathbf{x}_i'\mathbf{a}_0 + u_i$ avec $E[u_i] \equiv 0$ et $E[u_i / \mathbf{x}_i] = 0$ impliquent la validité de la ***condition d'orthogonalité*** :

$$\boxed{E[\mathbf{x}_i(y_i - \mathbf{x}_i'\mathbf{a}_0)] = \mathbf{0}_{K\times 1},}$$

qui est une condition de moment fondé sur l'exogénéité de $\mathbf{x}_i$.

(*iii*) Qui plus est on a :

$\mathbf{a}_0$ est l'***unique solution en*** $\mathbf{a}$ de l'équation $E[\mathbf{x}_i(y_i - \mathbf{x}_i'\mathbf{a})] = \mathbf{0}_{K\times 1}$.

(*iv*) Le ***principe d'analogie*** définit l'estimateur de la ***Méthode des Moments*** de $\mathbf{a}_0$, $\hat{\mathbf{a}}_N^{MM}$, à partir de la ***contre-partie empirique du problème théorique*** définissant $\mathbf{a}_0$:

$\hat{\mathbf{a}}_N^{MM}$ est solution de l'équation $N^{-1}\sum_{i=1}^{N}\mathbf{x}_i(y_i - \mathbf{x}_i'\mathbf{a}) = \mathbf{0}_{K\times 1}$.

(*v*) On montre alors que :

$$\boxed{\hat{\mathbf{a}}_N^{MM} = \hat{\mathbf{a}}_N^{MCO}.}$$

et que :

$$\boxed{\sqrt{N}\,(\hat{\mathbf{a}}_N^{MCO} - \mathbf{a}_0) \xrightarrow[N\to+\infty]{L} \mathcal{N}\left(\mathbf{0}, E\left[\mathbf{x}_i\mathbf{x}_i'\right]^{-1} E\left[u_i^2\mathbf{x}_i\mathbf{x}_i'\right] E\left[\mathbf{x}_i\mathbf{x}_i'\right]^{-1}\right).}$$

(*vi*) Dans le cas où les termes d'erreur du modèle de régression sont homoscédastiques, *i.e.* si $E\left[u_i^2 / \mathbf{x}_i\right] = V\left[u_i / \mathbf{x}_i\right] = \sigma_0^2$, alors on a :

$$\boxed{E\left[u_i^2\mathbf{x}_i\mathbf{x}_i'\right] = \sigma_0^2 E\left[\mathbf{x}_i\mathbf{x}_i'\right]}$$

et :

$$\boxed{\sqrt{N}\,(\hat{\mathbf{a}}_N^{MCO} - \mathbf{a}_0) \xrightarrow[N\to+\infty]{L} \mathcal{N}\left(\mathbf{0}, \sigma_0^2 E\left[\mathbf{x}_i\mathbf{x}_i'\right]^{-1}\right)}$$

On a également montré que si $E\left[\mathbf{x}_i u_i\right] \neq \mathbf{0}$, *i.e.* au moins un des éléments de $\tilde{\mathbf{x}}_i$ est endogène, alors :

(*i*) on a :

$$\boxed{\hat{\mathbf{a}}_N^{MCO} \xrightarrow[N \to \infty]{p} \boldsymbol{\gamma} \neq \mathbf{a}_0,}$$

i.e. l'estimateur des MCO de $\mathbf{a}_0$ est ***biaisé***.

et :

(*ii*) on sait que :

$$\boxed{\boldsymbol{\gamma} \equiv \arg\min_{\mathbf{g}} E\left[(y_i - \mathbf{g}'\mathbf{x}_i)^2\right],}$$

i.e. que $\hat{\mathbf{a}}_N^{MCO}$ converge vers $\boldsymbol{\gamma}$, le paramètre de la ***projection linéaire*** de y_i sur $\mathbf{x}_i$:

$$\boxed{EL\left[y_i / \mathbf{x}_i\right] = \boldsymbol{\gamma}'\mathbf{x}_i.}$$

- Dans tous les cas, en calculant l'estimateur des MCO de $\mathbf{a}_0$, $\hat{\mathbf{a}}_N^{MCO}$, on pourra calculer ***de bonnes prédictions*** (linéaires) ***de*** y_i ***par*** $\mathbf{x}_i$:

$$\mathbf{x}_i'\hat{\mathbf{a}}_N^{MCO} = \mathbf{x}_i'\hat{\boldsymbol{\gamma}}_N^{MCO}.$$

- Mais, on n'obtiendra une mesure satisfaisante de $\mathbf{a}_0$, l'effet causal recherché, ***que si*** $E\left[\mathbf{x}_i u_i\right] = \mathbf{0}$, *i.e.* si $\mathbf{x}_i$ est exogène par rapport à u_i.

C'est là la différence entre une ***logique d'ajustement*** et une ***logique d'identification***.

Remarques sur l'analyse des conditions d'identification

On a vu ici que si :

$$y_i = \mathbf{x}_i'\mathbf{a}_0 + u_i \quad \text{avec} \quad E\left[u_i / \mathbf{x}_i\right] = E\left[u_i\right] \equiv 0$$

alors :

$$E\left[\mathbf{x}_i(y_i - \mathbf{x}_i'\mathbf{a}_0)\right] = \mathbf{0} \quad \Leftrightarrow \quad E\left[\mathbf{x}_i y_i\right] = E\left[\mathbf{x}_i\mathbf{x}_i'\right]\mathbf{a}_0.$$

On a ensuite vu que :

$$E\left[\mathbf{x}_i(y_i - \mathbf{x}_i'\mathbf{a})\right] = \mathbf{0} \text{ admet une solution unique en } \mathbf{a}.$$

si et seulement si $rang E\left[\mathbf{x}_i\mathbf{x}_i'\right] = K$.

Cette solution étant ***nécessairement*** $\mathbf{a}_0$, on a en conclu que $rang E\left[\mathbf{x}_i\mathbf{x}_i'\right] = K$ est une ***condition nécessaire et suffisante*** d'***identification de*** $\mathbf{a}_0$ ***dans un modèle de régression linéaire***, *i.e.* en considérant que la condition d'orthogonalité $E\left[\mathbf{x}_i(y_i - \mathbf{x}_i'\mathbf{a}_0)\right] = \mathbf{0}$ est « acquise ».

La démarche suivie est en fait relativement simple.

1. On définit $\mathbf{a}_0$ comme la solution d'un ***problème théorique, dit « problème limite ».*** Dans le cas du modèle de régression linéaire, le problème théorique est :

$$\mathbf{a}_0 \text{ est solution en } \mathbf{a} \text{ de l'équation } E\left[\mathbf{x}_i(y_i - \mathbf{x}_i'\mathbf{a})\right] = \mathbf{0}$$

   Cette équation étant dérivée de la forme du modèle de y_i, $y_i = \mathbf{x}_i'\mathbf{a}_0 + u_i$ avec $E\left[u_i\right] \equiv 0$, et de l'exogénéité de $\mathbf{x}_i$ par rapport à u_i qui donne $E\left[\mathbf{x}_i u_i\right] = \mathbf{0}$.

2. On considère ensuite que $\mathbf{a}_0$ ***est identifié*** si $\mathbf{a}_0$ est la ***solution unique*** du ***problème théorique considéré***. Dans le cas du modèle de régression linéaire $\mathbf{a}_0$ ***est identifié ssi*** $E\left[\mathbf{x}_i\mathbf{x}_i'\right]$ est inversible.

La justification de cette démarche repose en fait sur celle de l'utilisation du ***principe d'analogie***.

3. De fait, ***le problème théorique*** considéré ***est*** également ***le problème limite*** (lorsque $N \to +\infty$) ***du problème empirique*** considéré pour le ***calcul de l'estimateur de*** $\mathbf{a}_0$, $\hat{\mathbf{a}}_N$.

4. La LGN et ses variantes permettent alors d'espérer que si le problème théorique identifie correctement $\mathbf{a}_0$, alors le problème empirique doit permettre de calculer un estimateur de $\mathbf{a}_0$, $\hat{\mathbf{a}}_N$, « proche » de $\mathbf{a}_0$ lorsque N est grand bien que fini.

D'une certaine manière, pour analyser les conditions d'identification de $\mathbf{a}_0$, on se place dans « ***le meilleur des cas*** », celui ***où on dispose d'un échantillon de taille infinie***, sachant qu'empiriquement on sera dans un situation proche si on dispose d'un grand échantillon (sous certaines conditions de régularité).

Pour le cas d'un modèle de régression linéaire :

Si N est grand et $rang E\left[\mathbf{x}_i\mathbf{x}_i'\right] = K$,

alors on n'a ***vraiment pas de chance*** si $N^{-1}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i'$ (et donc $\mathbf{X}'\mathbf{X} = \sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i'$) n'est pas inversible.