ÉCONOMÉTRIE APPROFONDIE

(Éléments de correction)

Mercredi 13 décembre 2023

EXERCICE 1. On suppose que les données sont générées par le modèle suivant :

$$y_i = x_{1,i}\beta_1 + \dots x_{K,i}\beta_K + \varepsilon_i$$

où $x_{k,i}$, pour $k = 1,\dots,K$ sont des variables explicatives déterministes, β_k, pour $k = 1,\dots,K$, sont des paramètres réels, ε_i une variable aléatoire centrée de variance σ_ε^2. **(1)** On obtient la représentation matricielle de ce modèle en concatément (verticalement) les observations. On pose : $Y = (y_1, y_2, \dots, y_N)'$, $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N)'$, $\beta = (\beta_1, \beta_2, \dots, \beta_K)$ et

$$X = \begin{pmatrix} x_{1,1} & x_{2,1} & \dots & x_{K,1} \\ x_{1,2} & x_{2,2} & \dots & x_{K,2} \\ \vdots & \vdots & & \vdots \\ x_{1,N} & x_{2,N} & \dots & x_{K,N} \end{pmatrix}$$

On peut alors réécrire le modèle sous la forme :

$$Y = X\beta + \varepsilon$$

(2) L'estimateur des MCO est le vecteur $\hat\beta$ qui minimise la somme des carrés des résidus :

$$\begin{aligned}\hat\beta &= \arg\min_{\{\beta\}} (Y - X\beta)'(Y - X\beta) \\ &= \arg\min_{\{\beta\}} Y'Y - \beta'X'Y - Y'X\beta + \beta'X'X\beta \\ &= \arg\min_{\{\beta\}} \beta'X'X\beta - 2\beta'X'Y\end{aligned}$$

La condition nécessaire d'optimalité est :

$$2X'X\hat\beta - 2X'Y = 0$$

$$\Leftrightarrow \hat\beta = (X'X)^{-1} X'Y$$

en supposant que la matrice $X'X$ est inversible (c'est le cas si, comme nous le supposons habituellement, X est une matrice de rang K). **(3)** $\hat\beta$ est un estimateur sans biais de β. Pour le montrer substituons le DGP dans l'expression de l'estimateur :

$$\begin{aligned}\hat\beta &= (X'X)^{-1} X'(X\beta + \varepsilon) \\ &= (X'X)^{-1} X'X\beta + (X'X)^{-1} X'\varepsilon \\ &= \beta + (X'X)^{-1} X'\varepsilon\end{aligned}$$

En supposant que les variables exogènes sont déterministes (sinon il faut recourrir au théorème des espérances itérées), on a donc :

$$\begin{aligned}\mathbb{E}\left[\hat\beta\right] &= \beta + \mathbb{E}\left[(X'X)^{-1} X'\varepsilon\right] \\ &= \beta + (X'X)^{-1} X'\mathbb{E}[\varepsilon] \\ &= \beta\end{aligned}$$

(4) La variance de $\hat\beta$ est définie par $\mathbb{V}\left[\hat\beta\right] = \mathbb{E}\left[\left(\hat\beta - \mathbb{E}\left[\hat\beta\right]\right)^2\right]$. En substituant l'expression de $\hat{\beta}$ en fonction de β (qui est aussi l'espérance de $\hat{\beta}$), il vient (en supposant que $\mathbb{E}\left[\varepsilon_i\varepsilon_j\right]=0$ pour tout $i\neq j$) :

$$\begin{aligned}\mathbb{V}\left[\hat{\beta}\right] &= \mathbb{V}\left[\left(X'X\right)^{-1}X'\varepsilon\right]\\ &= \left(X'X\right)^{-1}X'\mathbb{V}\left[\varepsilon\right]X\left(X'X\right)^{-1}\\ &= \left(X'X\right)^{-1}X'\sigma_\varepsilon^2 I_N X\left(X'X\right)^{-1}\\ &= \sigma_\varepsilon^2\left(X'X\right)^{-1}X'X\left(X'X\right)^{-1}\\ &= \sigma_\varepsilon^2\left(X'X\right)^{-1}\end{aligned}$$

Si la variance du vecteur ε est différente de $\sigma_\varepsilon^2 I_N$ (autocorrélation ou hétéroscédasticité) les matrices $X'X$ ne se simplifient plus et on obtient :

$$\mathbb{V}\left[\hat{\beta}\right] = \left(X'X\right)^{-1}X'\Sigma X\left(X'X\right)^{-1}$$

où Σ est la matrice de variance covariance de ε. **(5)** L'estimateur des MCO est une variable aléatoire, *i.e.* a une variance, car $\hat{\beta}$ est une fonction linéaire du vecteur aléatoire ε.

EXERCICE 2. On considère le modèle suivant :

$$y_i = \mu + \varepsilon_i$$

pour $i=1,\dots,N$ avec μ un paramètre réel et ε_i une variable aléatoire réelle d'espérance nulle et de variance σ_i^2 avec $\mathbb{E}\left[\varepsilon_i\varepsilon_j\right]=0$ si $i\neq j$. On suppose que $\lim_{N\to\infty}N^{-1}\sum_{i=1}^N\sigma_i^2=\bar{\sigma}^2<\infty$. Le modèle empirique est :

$$y_i = a + u_i$$

(1) On montre facilement que l'estimateur des MCO de a est la moyenne des y_i. Le modèle peut s'écrire sous forme matricielle :

$$Y = Xa + U$$

avec $Y=(y_1,y_2,\dots,y_N)'$, $U=(u_1,u_2,\dots,u_N)'$ et $X=(1,1,\dots,1)'$ un vecteur $N\times 1$. K'estimateur des MCO est :

$$\begin{aligned}\hat{a}_{\text{MCO}} &= (X'X)^{-1}X'Y\\ &= N^{-1}\sum_{i=1}^N y_i = \bar{y}\end{aligned}$$

(2) L'espérance $\hat{a}_{\text{MCO}}$, en substituant le processus générateur des données, est :

$$\begin{aligned}\mathbb{E}\left[\hat{a}_{\text{MCO}}\right] &= N^{-1}\sum_{i=1}^N\mathbb{E}[\mu+\varepsilon_i]\\ &= N^{-1}N\mu+0=\mu\end{aligned}$$

$\hat{a}_{\text{MCO}}$ est donc un estimateur sans biais de μ. **(3)** On suit la même démarche pour le calcul de la variance de $\hat{a}_{\text{MCO}}$ [1]. On a :

$$\begin{aligned}\mathbb{V}\left[\hat{a}_{\text{MCO}}\right] &= N^{-2}\mathbb{E}\left[\sum_{i=1}^N\sum_{j=1}^N\varepsilon_i\varepsilon_j\right]\\ &= N^{-2}\sum_{i=1}^N\sum_{j=1}^N\mathbb{E}\left[\varepsilon_i\varepsilon_j\right]\end{aligned}$$

comme les ε_i sont non corrélés, il vient :

$$\begin{aligned}\mathbb{V}\left[\hat{a}_{\text{MCO}}\right] &= N^{-2}\sum_{i=1}^N\sigma_i^2\\ &= \frac{\bar{\sigma}^2}{N}\end{aligned}$$

(4) La variance de $\hat{a}_{\text{MCO}}$ tend vers 0 quand N tend vers l'infini, l'estimateur tend donc, par la loi des frands nombres, en probabilité vers son espérance :

$$\hat{a}_{\text{MCO}}\xrightarrow[N\to\infty]{\text{proba}}\mu$$

1. En retenant de la question 2 que $\hat{a}_{\text{MCO}}=\mu+N^{-1}\sum_{i=1}^N$ et donc que $\hat{a}_{\text{MCO}}-\mathbb{E}\left[\hat{a}_{\text{MCO}}\right]=N^{-1}\sum_{i=1}^N\varepsilon_i$.

(5) Cet estimateur sans biais n'est pas efficace car la variance des résidus est spécifique à chaque observation (problème d'hétéroscédasticité). **(6)** De façon générale, l'estimateur des MCG est défini par :

$$\hat{a}_{\text{MCG}} = \left(X'\Sigma^{-1}X\right)^{-1} X'\Sigma^{-1}Y$$

où Σ est la variance du vecteur ε :

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & \dots & & 0 & \sigma_N^2 \end{pmatrix}$$

On a donc :

$$\hat{a}_{\text{MCG}} = \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \dots + \frac{1}{\sigma_N^2}\right)^{-1} \left(\frac{y_1}{\sigma_1^2} + \frac{y_2}{\sigma_2^2} + \dots + \frac{y_N}{\sigma_N^2}\right)$$

Il s'agit d'un estimateur sans biais de μ. En effet :

$$\begin{aligned} \mathbb{E}\left[\hat{a}_{\text{MCG}}\right] &= \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \dots + \frac{1}{\sigma_N^2}\right)^{-1} \sum_{i=1}^{N} \frac{\mathbb{E}[y_i]}{\sigma_i^2} \\ &= \left(\sum_{i=1}^{N} \frac{1}{\sigma_i^2}\right)^{-1} \sum_{i=1}^{N} \frac{\mu}{\sigma_i^2} \\ &= \mu \end{aligned}$$

(7) Pour calculer la variance de $\hat{a}_{\text{MCG}}$ on peut partir de la forumule obtenue en cours, $\left(X'\Sigma^{-1}X\right)^{-1}$, ou du cas particulier étudié ici :

$$\begin{aligned} \mathbb{V}\left[\hat{a}_{\text{MCG}}\right] &= \mathbb{E}\left[\left(\hat{a}_{\text{MCG}} - \mu\right)^2\right] \\ &= \left(\sum_{i=1}^{N} \frac{1}{\sigma_i^2}\right)^{-2} \mathbb{E}\left[\left(\sum_{i=1}^{N} \frac{\varepsilon_i}{\sigma_i^2}\right)^2\right] \\ &= \left(\sum_{i=1}^{N} \frac{1}{\sigma_i^2}\right)^{-2} \mathbb{E}\left[\sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\varepsilon_i}{\sigma_i^2}\frac{\varepsilon_j}{\sigma_j^2}\right] \\ &= \left(\sum_{i=1}^{N} \frac{1}{\sigma_i^2}\right)^{-2} \mathbb{E}\left[\sum_{i=1}^{N} \frac{\varepsilon_i^2}{\sigma_i^4}\right] \\ &= \left(\sum_{i=1}^{N} \frac{1}{\sigma_i^2}\right)^{-2} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \\ &= \left(\sum_{i=1}^{N} \frac{1}{\sigma_i^2}\right)^{-1} \end{aligned}$$

(8) La variance de $\hat{a}_{\text{MCO}}$ est plus grande que celle de $\hat{a}_{\text{MCG}}$. En effet :

$$\begin{aligned} & \mathbb{V}\left[\hat{a}_{\text{MCO}}\right] > \mathbb{V}\left[\hat{a}_{\text{MCO}}\right] \\ \Leftrightarrow\ & \frac{\bar{\sigma}^2}{N} > \left(\sum_{i=1}^{N} \frac{1}{\sigma_i^2}\right)^{-1} \\ \Leftrightarrow\ & \bar{\sigma}^2 > N\left(\sum_{i=1}^{N} \frac{1}{\sigma_i^2}\right)^{-1} \end{aligned}$$

On reconnaît gauche une moyenne arithmétique et à droite une moyenne harmonique. La moyenne harmonique est toujours plus petite que la moyenne arithmétique (voir le rappel). Donc la variance de l'estimateur des MCO est plus grande que la variance de l'estimateur des MCG. Comme la variance de l'estimateur des MCO converge vers 0 quand N tend vers l'infini, la variance de l'estimateur des MCG converge nécessairement vers 0 quand Ntend vers l'infini, et donc l'estimateur des MCG converge aussi en probabilité vers μ.

Rappel *Soient $\alpha_1, \ldots, \alpha_n$ des nombres réels positifs, alors la moyenne arithmétique des α_i est plus grande que la moyenne harmonique des α_i :*

$$\frac{\alpha_1 + \ldots + \alpha_n}{n} > \frac{n}{\frac{1}{\alpha_1} + \ldots + \frac{1}{\alpha_n}}$$

Exercice 3. Soit le modèle :

$$y_t = \beta x_t + \varepsilon_t$$

avec β un paramètre réel et

$$\varepsilon_t = \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + \nu_t$$

où ν_t est une variable aléatoire centrée de variance σ_ν^2 et les paramètres réels φ_1, φ_2 sont tels que les moments d'ordre 2 (la variance et la fonction d'autocorrélation) de ε_t sont bien définis. On peut montrer, cela fera l'objet d'un cours au second semestre que la fonction d'autocorrélation de ε, $\rho(k)$, est non nulle est tend vers zero quand k tend vers l'infini (la corrélation entre ε_t et ε_{t-k} se rapproche de zéro quand k devient assez grand). On suppose que les valeurs de φ_1 et φ_2 sont connues. **(1)** des MCO pour β n'est pas un estimateur efficace car les résidus du modèle sont autocorrélés. **(2)** On pose $\tilde{y}_t = y_t - \varphi_1 y_{t-1} - \varphi_2 y_{t-2}$ et $\tilde{x}_t = x_t - \varphi_1 x_{t-1} - \varphi_2 x_{t-2}$. En utilisant la définition de y on a :

$$y_{t-1} = \beta x_{t-1} + \varepsilon_{t-1}$$

et

$$y_{t-2} = \beta x_{t-2} + \varepsilon_{t-2}$$

ainsi :

$$\tilde{y}_t = y_t - \varphi_1 y_{t-1} - \varphi_2 y_{t-2}$$

$$\Leftrightarrow \tilde{y}_t = \beta x_t + \varepsilon_t - \beta\varphi_1 x_{t-1} - \varphi\varepsilon_{t-1} - \beta\varphi_2 x_{t-2} - \varphi_2\varepsilon_{t-2}$$

$$\Leftrightarrow \tilde{y}_t = \beta\tilde{x}_t + \varepsilon_t - \varphi_1\varepsilon_{t-1} - \varphi_2\varepsilon_{t-2}$$

$$\Leftrightarrow \tilde{y}_t = \beta\tilde{x}_t + \nu_t$$

Ici l'estimateur est efficace car ν_t n'est pas autocorrélé (il n'y a pas d'hétéroscédasticité car la variance est constante, elle ne dépend pas de t). **(3)** Il s'agit d'un estimateur des MCG. **(4)** Si les paramètres φ_1 et φ_2 ne sont pas connus, on peut les estimer dans une première étape. En estimant le modèle original, par les MCO on récupère les résidus estimés et on estime le modèle autorégressif :

$$\hat{\varepsilon}_t = \varphi_1 \hat{\varepsilon}_{t-1} + \varphi_2 \hat{\varepsilon}_{t-2} + \nu_t$$

On utilise alors $\hat{\varphi}_1$ et $\hat{\varphi}_2$ pour transformer les données comme décrit plus haut. On peut reprendre la procédure plusierus fois tant que les estimations pour β et les paramètres φ_1 et φ_2 changent.