Économétrie approfondie

(Exercices)

Stéphane Adjemian*

Le 8 juin 2025 à 16:24

Exercice 1 La nature génère des données avec le modèle suivant :

$$y_t = \mu_1 + \mu_2 \mathbb{1}_{\{t>\kappa\}}(t) + \varepsilon_t$$

avec μ_1 et μ_2 des paramètres réels, $\varepsilon_t \sim \text{iid}(0, \sigma^2_\varepsilon)$, κ un entier donné entre 1 and T. Le modèle empirique est :

$$y_t = c + u_t$$

On note $\mathcal{Y}_T = \{y_1, \dots, y_T\}$ l'échantillon produit par la nature et à disposition de l'économètre. **(1)** Donner l'expression de l'estimateur des MCO de c, noté $\hat{c}_T$. **(2)** Calculer l'espérance de $\hat{c}_T$, en explicitant le rapport avec les paramètres du modèle de la nature. **(3)** Calculer la variance de $\hat{c}_T$. (4) Quel est la limite en probabilité de $\hat{c}_T$?

Exercice 2 La nature génère des données avec le modèle suivant :

$$y_t = \mu_1 + \mu_2 \mathbb{1}_{\{t>[\kappa T]\}}(t) + \varepsilon_t$$

avec μ_1 et μ_2 des paramètres réels, $\varepsilon_t \sim \text{iid}(0, \sigma^2_\varepsilon)$, $\kappa \in]0,1[$, et $[x]$ représente la partie entière de x. Le modèle empirique est :

$$y_t = c + u_t$$

On note $\mathcal{Y}_T = \{y_1, \dots, y_T\}$ l'échantillon produit par la nature et à disposition de l'économètre. Reprendre les questions de l'exercice 1 avec ce nouveau DGP.

Exercice 3 Soit le modèle empirique :

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \dots + \beta_K x_{K,i} + u_i$$

L'échantillon est $\mathcal{Y}_T = \{y_1, \dots, y_N, x_1, \dots, x_N\}$. **(1)** Écrire le modèle matriciellement sous la forme $Y = X\beta + U$. **(2)** Montrer que l'estimateur des MCO est $\hat{\beta} = (X'X)^{-1}X'Y$. **(3)** Expliciter le contenu de $X'X$ et $X'Y$ en faisant apparaître les observations individuelles. **(4)** Dans le cas $K = 1$, sur la droite de l'équation ne reste que la constante et une variable explicative $x_{1,i}$, écrire

*Université du Mans. `stephane DOT adjemian AT univ DASH lemans DOT fr`

explicitement à partir des réponses aux questions (2) et (3) l'estimateur des MCO (pour la constante et la pente). En passant, montrer que le déterminant de $X'X$ est strictement positif dès lors que $N > 1$ et $x_1 \neq 0$. **(5)** Montrer que la somme des résidus est nulle par construction.

Exercice 4 La nature génère des données avec le modèle suivant :

$$y_t = \mu_0 + \mu_1 t + \varepsilon_t$$

avec μ_0 et μ_1 des paramètres réels, et $\varepsilon_t \sim \text{iid}(0, \sigma_\varepsilon^2)$. Le modèle empirique est :

$$y_t = c + u_t$$

On note $\mathcal{Y}_T = \{y_1, \dots, y_T\}$ l'échantillon produit par la nature et à disposition de l'économètre. **(1)** Donner l'expression de l'estimateur des MCO de c, noté $\hat{c}_T$. **(2)** Calculer l'espérance de $\hat{c}_T$. S'agit-il d'un estimateur non biaisé de μ_0 ?

Exercice 5 Soit la matrice carrée :

$$\mathcal{A} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

où A_{11} et A_{22} sont des matrices carrées, A_{11} et $A_{22} - A_{21}A_{11}^{-1}A_{12}$ sont des matrices de plein rang. **(1)** Montrer que :

$$\mathcal{A}^{-1} = \begin{pmatrix} A_{11}^{-1}\left(I + A_{12}BA_{21}A_{11}^{-1}\right) & -A_{11}^{-1}A_{12}B \\ -BA_{21}A_{11}^{-1} & B \end{pmatrix}$$

avec $B = \left(A_{22} - A_{21}A_{11}^{-1}A_{12}\right)^{-1}$. **(2)** Appliquer la formule dans le cas où A_{12} et A_{21} sont des matrices nulles.

Exercice 6 Soit le modèle, sous forme matricielle, suivant :

$$Y = X\beta + U$$

où Y et U sont des vecteurs $N \times 1$, X une matrice $N \times K$ et β un vecteur $K \times 1$. On partitionne la matrice des variables explicatives de la façon suivante :

$$X = \left(X_1 \vdots X_2\right)$$

où X_1 et X_2 sont des matrices $N \times K_1$ et $N \times K_2$ avec $K_1 \geq 1$, $K_2 \geq 1$ et $K_1 + K_2 = K$. On partitionne le vecteur des paramètres de la même façon β_1 contient les paramètres associés aux variables dans X_1, β_2 les paramètres associés aux variables dans X_2. **(1)** Donner les estimateurs des MCO de β_1 et β_2 dans le cas où les variables dans X_1 sont orthogonales à celles dans X_2 (c'est-à-dire dans le cas où $X_1'X_2 = 0$ et $X_2'X_1 = 0$). Interpréter. **(2)** Si les matrices X_1 et X_2 ne sont pas des matrices orthogonbales, donner une expression de $\hat{\beta}_1$ en la comparant à l'expression obtenue dans la question précédente. **(3)**

Montrer que le vecteur des résidus estimés peut s'écrire sous la forme $\hat{U} = MY$ (en donnant une expression pour la matrice M). **(4)** Montrer que la matrice M est idempotente.**(5)** Montrer que les matrices M et X sont orthogonales. **(6)** Montrer que la prédiction $\hat{Y} = X\hat{\beta}$ peut s'écrire sous la forme $\hat{Y} = HY$. La matrice H est-elle idempotente ? **(7)** Éliminer $\hat{\beta}_1$ dans l'expression de $\hat{\beta}_2$, obtenu en répondant à la question (2), en faisant apparaître le projecteur orthogonal $M_1 = I - X_1(X_1'X_1)^{-1}X_1$. Commenter.

Exercice 7 On cherche à illustrer à l'aide de simulations la convergence de l'estimateur des MCO. On considère le DGP suivant :

$$y_i = a + bx_i + \varepsilon_i$$

$i = 1, \dots, N$, où ε_i est un bruit blanc gaussien centré réduit, $a = 1$ et $b = \frac{1}{2}$ sont les vraies valeurs des paramètres. On commencera par supposer que la variable explicative est déterministe, c'est-à-dire que la réalisation des x_i est constante d'un échantillon à l'autre. **(1)** Pour $N = 10$ (taille de l'échantillon), simuler les $\{x_i\}_{i=1}^N$ avec une loi normale centrée sur 1 et de variance 2. puis dans une boucle 50000 échantillons $\{y_i\}_{i=1}^N$ en estimant pour chaque échantillon les paramètres α et β dans le modèle empirique :

$$y_i = \alpha + \beta x_i + u_i$$

Les 50000 valeurs obtenues pour $\widehat{\beta}$ devront être stockées dans un vecteur. Calculer la moyenne et la variance des $\widehat{\beta}$ obtenus. Que représentent ces quantités. **(2)** Reprendre la question (1) avec $N = 100$, $N = 1000$ et $N = 10000$. Comparer les moyennes et variances puis commenter. **(3)** Reprendre les questions (1) et (2) en considérant que la variable explicative n'est pas déterministe (pour former chaque échantillon il faut tirer une réalisation différente des $\{x_i\}_{i=1}^N$ dans une loi normale centrée sur 1 et de variance 2). **(4)** Reprendre la question (3) en considérant que la variance de la variable explicative est 0.2 puis 20 (plutôt que 2). Comment ce changement affecte les propriétés de l'estimateur de β ?

Exercice 8 On cherche à illustrer à l'aide de simulations la distribution de l'estimateur des MCO. On reprend le même cadre que dans l'exercice précédent, mais au lieu de calculer la moyenne et la variance on représente la distribution de l'estimateur de β en déviation à la vraie valeur du paramètre b, $\widehat{\beta} - \frac{1}{2}$. On peut utiliser des histogrammes pour représenter les distributions. Reprendre les mêmes scenarii que dans l'exercice précédent et représenter les distributions de $\widehat{\beta} - \frac{1}{2}$ puis $\sqrt{N}\left(\widehat{\beta} - \frac{1}{2}\right)$. Commenter.

Exercice 9 Soit un couple de variables aléatoires réelles (X_1, X_2) distribué selon une loi normale bivariée :

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right)$$

avec $(\mu_1, \mu_2) \in \mathbb{R}^2$, $(\sigma_1, \sigma_2) \in \mathbb{R}_+^2$ et $0 < \rho < 1$. **(0)** Comment s'interprète le paramètre ρ ? **(1)** Écrire la densité jointe du couple (X_1, X_2). **(2)** Montrer qu'il est possible de réécrire cette densité jointe comme le produit de la densité conditionnelle de $X_1|X_2$ et de la densité marginale de X_2. Montrer que ces densités sont gaussiennes. **(3)** Donner $\mathbb{E}[X_1|X_2]$ et $\mathbb{V}[X_1|X_2]$. **(4)** Soit X_2 une variable aléatoire normale $\mathcal{N}(\mu_2, \sigma_2^2)$. Quelle est la loi de $X_1 = a + bX_2 + \varepsilon$ si $X_1 \perp \varepsilon$ et $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$? **(5)** Écrire la densité de $X_1|X_2$. **(6)** L'utilisation du mot "régression" a été introduite par le statisticien Sir Francis Galton dans une étude de 1885 intitulée "Regression Toward Mediocrity in Hereditary Stature" (Régression vers la médiocrité dans la stature héréditaire). Il a montré que la taille des enfants issus de parents très petits ou très grands se rapprochait de la moyenne. En moyenne, l'enfant de parents de grandes tailles est plus petit. Discuter ce résultat à l'aide de vos réponses aux questions (2) et (3).

Exercice 10 On considère un modèle générateur des données de la forme :

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

avec une variable explicative (x) aléatoire (à chaque fois que l'économètre s'adresse à la nature pour obtenir un nouvel échantillon, celle-ci retourne des y **et** x différents). On suppose que ε est i.i.d. et que :

$$\mathbb{E}\left[\varepsilon_t | x_s\right] = 0$$
$$\mathbb{V}\left[\varepsilon_t | x_s\right] = \sigma_\varepsilon^2$$

pour tout t, s. Le modèle empirique est :

$$y_t = a + bx_t + \epsilon_t$$

(1) Écrire l'estimateur des MCO de b, noté $\hat{b}$. **(2)** Exprimer $\hat{b}$ en fonction de β. **(3)** Soit un couple de variables aléatoires U et V. Écrire la définition de l'espérance conditionnelle de $V|U$, que l'on notera $\Psi(U)$. **(4)** Montrer que l'espérance marginale de V est égale à l'espérance de $\Psi(U)$, en intégrant sur U. **(5)** Montrer que $\mathbb{E}\left[\varepsilon_t\right] = 0$. **(6)** Montrer que $\mathbb{E}\left[\varepsilon_t x_s\right] = 0$ pour tout t, s. **(7)** Montrer que ε_t est homoscédastique. **(8)** L'estimateur des MCO $\hat{b}$ est-il sans biais ?

Exercice 11 On considère un modèle générateur des données de la forme :

$$y_t = \alpha + \beta x_t + \varepsilon_t, \quad t = 1, \dots, T$$

avec une variable explicative (x) aléatoire. On suppose que la covariance de x et ε est non nulle. Le modèle empirique est :

$$y_t = a + bx_t + \epsilon_t$$

(1) L'estimateur des MCO de b, noté $\hat{b}$, est-il sans biais. **(2)** Calculer la limite en probabilité de l'estimateur des MCO. L'estimateur des MCO est-il convergent ? **(3)** On suppose que l'on dispose des réalisations d'une autre variable aléatoire $\{z_t, t = 1, \dots, T\}$. La variable aléatoire z est corrélée avec x mais non corrélée avec ε. On définit les estimateurs de a et b par le système suivant :

$$\left\{ \begin{array}{l} \frac{1}{T}\sum_{t=1}^{T}\left(y_t - \hat{a}_{\mathrm{IV}} - \hat{b}_{\mathrm{IV}} x_t\right) = 0 \\ \frac{1}{T}\sum_{t=1}^{T} z_t \left(y_t - \hat{a}_{\mathrm{IV}} - \hat{b}_{\mathrm{IV}} x_t\right) = 0 \end{array} \right.$$

Justifier la définition des estimateurs. **(4)** Résoudre de système d'équations en donnant les expressions de $\hat{a}_{\text{IV}}$ et $\hat{b}_{\text{IV}}$. **(5)** Montrer que $\hat{b}_{\text{IV}}$ converge en probabilité vers β. $\hat{b}_{\text{IV}}$ est l'estimateur à variable instrumentale de b. **(6)** Calculer la variance asymptotique de $\sqrt{T}\left(\hat{b}_{\text{IV}} - \beta\right)$. Comparer avec l'estimateur des MCO.

Exercice 12 On généralise l'esimateur à variable instrumentale de l'exercice précédent. Le modèle générateur des données est :

$$y_t = X_t\boldsymbol{\beta} + \varepsilon_t$$

où X_t est un vecteur $1 \times K$, $\boldsymbol{\beta}$ un vecteur $K \times 1$, ε_t une variable aléatoire iid. On suppose que les variables X_t sont corrélées avec ε_t, mais nous disposons de K instrument Z_t. Le modèle empirique est :

$$y_t = X_t\boldsymbol{b} + \epsilon_t$$

(1) Donner une expression matricielle de l'estimateur à variables instrumentales $\hat{\boldsymbol{b}}_{\text{IV}}$. **(2)** Montrer que cet estimateur est convergent. **(3)** Proposer un estimateur si le nombre d'instruments est strictment supérieur à K.