ÉCONOMÉTRIE APPROFONDIE

(Session de rattrapage — Éléments de correction)

12 juin 2026

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

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

où les $x_{i,k}$ sont des variables explicatives déterministes, les β_k des paramètres réels, et ε_i une variable aléatoire i.i.d. centrée de variance σ_ε^2.
(1) On obtient la représentation matricielle en concaténant (verticalement) les N observations. On pose $Y = (y_1, \ldots, y_N)'$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_N)'$, $\beta = (\beta_1, \ldots, \beta_K)'$ et

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

X est une matrice $N \times K$ (chaque ligne rassemble les K variables explicatives d'une observation, chaque colonne les N observations d'une variable). Le modèle s'écrit alors :

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

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

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

(le terme $Y'Y$ ne dépend pas de b et les scalaires $b'X'Y$ et $Y'Xb$ sont égaux). La condition du premier ordre (dérivation par rapport au vecteur b) est :

$$2X'X\hat{b} - 2X'Y = 0 \Leftrightarrow \hat{b} = (X'X)^{-1} X'Y$$

Pour que cet estimateur existe, il faut que la matrice $X'X$ soit inversible, c'est-à-dire que X soit de plein rang colonne ($\mathrm{rg}(X) = K$) : absence de colinéarité parfaite entre les variables explicatives et $N \geq K$. La condition du second ordre est satisfaite car $X'X$ est définie positive sous l'hypothèse de plein rang : le point critique est bien un minimum.

(3) $\hat{b}$ est sans biais. En substituant le DGP dans l'expression de l'estimateur :

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

Comme les variables explicatives sont déterministes :

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

puisque $\mathbb{E}[\varepsilon] = 0$.

(4) En utilisant $\hat{b} - \beta = (X'X)^{-1} X'\varepsilon$ et le fait que les ε_i sont i.i.d. ($\mathbb{V}[\varepsilon] = \sigma_\varepsilon^2 I_N$) :

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

(les régresseurs étant déterministes, $(X'X)^{-1} X'$ sort de l'espérance).

(5) Théorème de Gauss-Markov. Sous les hypothèses du modèle linéaire (régresseurs déterministes de plein rang, erreurs centrées, homoscédastiques et non corrélées, $\mathbb{V}[\varepsilon] = \sigma_\varepsilon^2 I_N$), l'estimateur des MCO est, parmi tous les estimateurs linéaires et sans biais de β, celui de variance minimale (*Best Linear Unbiased Estimator*).

Démonstration. Soit $\tilde{b} = CY$ un estimateur linéaire quelconque, avec C une matrice $K \times N$ déterministe. Posons $C = (X'X)^{-1}X' + D$, où D mesure l'écart aux MCO. En substituant le DGP :

$$\tilde{b} = CY = CX\beta + C\varepsilon$$

Pour que $\tilde{b}$ soit sans biais *quel que soit* β, il faut $\mathbb{E}[\tilde{b}] = CX\beta = \beta$, donc $CX = I_K$. Or :

$$CX = \left((X'X)^{-1}X' + D\right) X = I_K + DX$$

La condition de non-biais impose donc $DX = 0$. La variance de $\tilde{b}$ s'écrit alors :

$$\mathbb{V}[\tilde{b}] = \sigma_\varepsilon^2 \, CC'$$

avec

$$\begin{aligned} CC' &= \left((X'X)^{-1}X' + D\right)\left(X(X'X)^{-1} + D'\right) \\ &= (X'X)^{-1} + (X'X)^{-1} \underbrace{X'D'}_{=(DX)'=0} \\ &\quad + \underbrace{DX}_{=0}(X'X)^{-1} + DD' \\ &= (X'X)^{-1} + DD' \end{aligned}$$

d'où :

$$\mathbb{V}[\tilde{b}] = \sigma_\varepsilon^2(X'X)^{-1} + \sigma_\varepsilon^2 DD' = \mathbb{V}[\hat{b}] + \sigma_\varepsilon^2 DD'$$

La matrice DD' est semi-définie positive (pour tout vecteur a, $a'DD'a = \|D'a\|^2 \geq 0$). On a donc $\mathbb{V}[\tilde{b}] - \mathbb{V}[\hat{b}] \succeq 0$: la variance de tout autre estimateur linéaire sans biais dépasse (au sens des matrices semi-définies positives) celle des MCO. L'égalité n'est atteinte que si $D = 0$, c'est-à-dire $\tilde{b} = \hat{b}$. Les MCO sont donc BLUE.

EXERCICE 2. On considère le modèle générateur des données :

$$y_i = \beta x_i + \varepsilon_i, \qquad i = 1, \ldots, N,$$

avec x_i déterministe non nul, $\mathbb{E}[\varepsilon_i] = 0$, les ε_i indépendantes mais hétéroscédastiques : $\mathbb{V}[\varepsilon_i] = \sigma^2 w_i$, les poids $w_i > 0$ étant connus.
(1) En minimisant $\sum_i (y_i - bx_i)^2$, la condition du premier ordre donne l'estimateur des MCO :

$$\hat{b} = \frac{\sum_{i=1}^N x_i y_i}{\sum_{i=1}^N x_i^2}$$

Il est sans biais. En substituant le DGP :

$$\hat{b} = \frac{\sum_i x_i(\beta x_i + \varepsilon_i)}{\sum_i x_i^2} = \beta + \frac{\sum_i x_i \varepsilon_i}{\sum_i x_i^2}$$

et comme les x_i sont déterministes et $\mathbb{E}[\varepsilon_i] = 0$:

$$\mathbb{E}[\hat{b}] = \beta + \frac{\sum_i x_i \mathbb{E}[\varepsilon_i]}{\sum_i x_i^2} = \beta$$

L'absence de biais ne dépend pas de l'hétéroscédasticité : elle tient au caractère déterministe des régresseurs et à la nullité de l'espérance des erreurs.

(2) En partant de $\hat{b} - \beta = \dfrac{\sum_i x_i \varepsilon_i}{\sum_i x_i^2}$ et en utilisant l'indépendance des erreurs ($\mathbb{V}[\varepsilon_i] = \sigma^2 w_i$, covariances nulles) :

$$\begin{aligned}\mathbb{V}[\hat{b}] &= \frac{1}{\left(\sum_i x_i^2\right)^2} \mathbb{V}\left[\sum_i x_i \varepsilon_i\right] \\ &= \frac{\sum_i x_i^2 \, \mathbb{V}[\varepsilon_i]}{\left(\sum_i x_i^2\right)^2} = \frac{\sigma^2 \sum_i w_i x_i^2}{\left(\sum_i x_i^2\right)^2}\end{aligned}$$

On remarque que ce n'est plus la formule usuelle $\sigma^2 / \sum_i x_i^2$: utiliser cette dernière (comme le ferait un logiciel par défaut) conduirait à une estimation erronée de la variance, et donc à une inférence (tests, intervalles de confiance) invalide.

(3) Pour détecter l'hétéroscédasticité, on peut utiliser un test fondé sur les résidus des MCO $\hat{\epsilon}_i$:

- **Test de Breusch-Pagan** : on régresse les résidus au carré $\hat{\epsilon}_i^2$ sur les variables susceptibles d'expliquer la variance (ici x_i, x_i^2, ou les w_i). Sous l'hypothèse nulle d'homoscédasticité, ces variables ne doivent rien expliquer ; la statistique NR^2 de cette régression auxiliaire suit asymptotiquement un χ^2 dont le nombre de degrés de liberté est le nombre de régresseurs auxiliaires. Une valeur élevée conduit à rejeter l'homoscédasticité.
- **Test de White** : même principe, en ajoutant les carrés et produits croisés des régresseurs (test plus général, ne présupposant pas la forme de l'hétéroscédasticité).
- **Test de Goldfeld-Quandt** : on ordonne les observations selon une variable soupçonnée de gouverner la variance, on estime le modèle séparément sur les deux sous-échantillons extrêmes, et l'on compare les variances résiduelles par un test de Fisher.

(4) Non, l'estimateur des MCO n'est pas efficace. Le théorème de Gauss-Markov suppose des erreurs sphériques ($\mathbb{V}[\varepsilon] = \sigma^2 I_N$). Ici la matrice de variance-covariance des erreurs vaut $\mathbb{V}[\varepsilon] = \sigma^2 \operatorname{diag}(w_1, \ldots, w_N) \neq \sigma^2 I_N$: les MCO restent sans biais mais ne sont plus de variance minimale. Il existe un estimateur linéaire sans biais plus précis, l'estimateur des moindres carrés généralisés.

(5) Comme $w_i > 0$, on divise chaque observation par $\sqrt{w_i}$. En posant $\tilde{y}_i = y_i / \sqrt{w_i}$, $\tilde{x}_i = x_i / \sqrt{w_i}$ et $\tilde{\varepsilon}_i = \varepsilon_i / \sqrt{w_i}$, le modèle transformé :

$$\tilde{y}_i = \beta \, \tilde{x}_i + \tilde{\varepsilon}_i$$

a des erreurs homoscédastiques :

$$\mathbb{V}[\tilde{\varepsilon}_i] = \frac{1}{w_i}\,\mathbb{V}[\varepsilon_i] = \frac{\sigma^2 w_i}{w_i} = \sigma^2$$

On peut alors appliquer les MCO au modèle transformé (Gauss-Markov s'applique de nouveau) :

$$\hat{b}_{\text{MCG}} = \frac{\sum_i \tilde{x}_i \tilde{y}_i}{\sum_i \tilde{x}_i^2} = \frac{\sum_i x_i y_i / w_i}{\sum_i x_i^2 / w_i}$$

Il s'agit de l'estimateur des *moindres carrés pondérés* (MCP, cas particulier des MCG pour des erreurs hétéroscédastiques non corrélées) : chaque observation est pondérée par $1/w_i$, c'est-à-dire l'inverse de sa variance. Les observations les plus « bruitées » (grand w_i) reçoivent un poids plus faible.
(6) Le modèle transformé étant homoscédastique de variance σ^2 :

$$\mathbb{V}[\hat{b}_{\text{MCG}}] = \frac{\sigma^2}{\sum_i \tilde{x}_i^2} = \frac{\sigma^2}{\sum_i x_i^2 / w_i}$$

Comparons avec $\mathbb{V}[\hat{b}] = \sigma^2 \dfrac{\sum_i w_i x_i^2}{(\sum_i x_i^2)^2}$. En appliquant l'inégalité de Cauchy-Schwarz avec $a_i = x_i\sqrt{w_i}$ et $c_i = x_i/\sqrt{w_i}$:

$$\left(\sum_i x_i^2\right)^2 = \left(\sum_i a_i c_i\right)^2 \leq \left(\sum_i w_i x_i^2\right)\left(\sum_i \frac{x_i^2}{w_i}\right)$$

d'où :

$$\begin{aligned}\mathbb{V}[\hat{b}] = \sigma^2 \frac{\sum_i w_i x_i^2}{\left(\sum_i x_i^2\right)^2} &\geq \sigma^2 \frac{\sum_i w_i x_i^2}{\left(\sum_i w_i x_i^2\right)\left(\sum_i x_i^2 / w_i\right)} \\ &= \frac{\sigma^2}{\sum_i x_i^2 / w_i} = \mathbb{V}[\hat{b}_{\text{MCG}}]\end{aligned}$$

On a donc bien $\mathbb{V}[\hat{b}] \geq \mathbb{V}[\hat{b}_{\text{MCG}}]$, avec égalité si et seulement si les w_i sont tous égaux (homoscédasticité). Ceci confirme que les MCG/MCP sont efficaces quand les MCO ne le sont pas.
(7) Si les poids w_i sont inconnus, deux stratégies :

— **MCG réalisables (FGLS)** : on spécifie une forme paramétrique pour la variance, $w_i = h(z_i, \theta)$ (par exemple $\sigma_i^2 = \sigma^2 z_i^2$). On estime d'abord le modèle par les MCO, on récupère les résidus $\hat{\epsilon}_i$, on en déduit une estimation $\hat{\theta}$ (en régressant $\ln \hat{\epsilon}_i^2$ sur z_i, par exemple), puis on applique les MCP avec les poids estimés $\hat{w}_i$. L'estimateur obtenu n'est plus sans biais en échantillon fini (les poids dépendent des données) mais reste convergent et asymptotiquement efficace.

— **MCO + écarts-types robustes** : si l'on ne souhaite pas modéliser la variance, on conserve l'estimateur des MCO (sans biais et convergent) mais on corrige l'estimation de sa variance à l'aide de l'estimateur robuste à l'hétéroscédasticité de White :

$$\widehat{\mathbb{V}}[\hat{b}] = \frac{\sum_i x_i^2 \, \hat{\epsilon}_i^2}{\left(\sum_i x_i^2\right)^2}$$

ce qui rétablit la validité des tests, au prix d'une perte d'efficacité.

EXERCICE 3. Le (log) salaire de l'individu i est généré par :

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \varepsilon_i$$

où x_i est le nombre d'années d'études, z_i le talent (non observé), et ε_i une erreur i.i.d. centrée non corrélée à x_i et z_i. Les variables x_i et z_i sont aléatoires, centrées, de variances σ_x^2, σ_z^2, et corrélées : $\sigma_{xz} = \mathrm{cov}(x_i, z_i) \neq 0$. L'économètre n'observe pas z_i et estime $y_i = b_0 + b_1 x_i + u_i$.
(1) Le talent omis est rejeté dans le terme d'erreur du modèle empirique :

$$u_i = \beta_2 z_i + \varepsilon_i$$

(le modèle empirique « voit » comme une seule perturbation u_i la somme de l'effet du talent et du choc ε_i).
(2) En utilisant la bilinéarité de la covariance, et le fait que x_i est non corrélé à ε_i :

$$\begin{aligned} \mathrm{cov}(x_i, u_i) &= \mathrm{cov}(x_i, \beta_2 z_i + \varepsilon_i) \\ &= \beta_2 \, \mathrm{cov}(x_i, z_i) + \mathrm{cov}(x_i, \varepsilon_i) \\ &= \beta_2 \, \sigma_{xz} \neq 0 \end{aligned}$$

(dès lors que $\beta_2 \neq 0$ et $\sigma_{xz} \neq 0$). Le régresseur x_i est donc *endogène* : il est corrélé avec l'erreur du modèle empirique. La condition d'orthogonalité étant violée, l'estimateur des MCO de b_1 est biaisé et non convergent.
(3) L'estimateur des MCO de la pente s'écrit (variables centrées) :

$$\hat{b}_1 = \frac{\frac{1}{N}\sum_i x_i y_i}{\frac{1}{N}\sum_i x_i^2}$$

En substituant le DGP $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \varepsilon_i$ (on travaille sur les variables centrées, la constante disparaît) :

$$\hat{b}_1 = \beta_1 + \beta_2 \frac{\frac{1}{N}\sum_i x_i z_i}{\frac{1}{N}\sum_i x_i^2} + \frac{\frac{1}{N}\sum_i x_i \varepsilon_i}{\frac{1}{N}\sum_i x_i^2}$$

Par la loi des grands nombres, $\frac{1}{N}\sum_i x_i^2 \xrightarrow{p} \sigma_x^2$, $\frac{1}{N}\sum_i x_i z_i \xrightarrow{p} \sigma_{xz}$ et $\frac{1}{N}\sum_i x_i \varepsilon_i \xrightarrow{p} \mathrm{cov}(x_i, \varepsilon_i) = 0$. Par le théorème de Slutsky :

$$\mathrm{plim}_{N\to\infty}\, \hat{b}_1 = \beta_1 + \beta_2 \frac{\sigma_{xz}}{\sigma_x^2}$$

Interprétation. L'estimateur des MCO ne converge pas vers β_1 : il comporte un *biais de variable omise* égal à $\beta_2\, \sigma_{xz}/\sigma_x^2$. Le terme σ_{xz}/σ_x^2 est précisément le coefficient de la régression du talent z sur les études x : les MCO attribuent à l'éducation une partie de l'effet du talent. Ici $\beta_2 > 0$ (le talent augmente le salaire) et $\sigma_{xz} > 0$ (les plus talentueux étudient plus longtemps) : le biais est *positif*, on surestime le rendement de l'éducation. Le biais disparaît si $\beta_2 = 0$ (le talent n'influence pas le salaire) ou si $\sigma_{xz} = 0$ (talent et études non corrélés) : dans ces cas l'omission est sans conséquence.

(4) On dispose d'une variable g_i (l'éducation des parents) telle que $\mathrm{cov}(g_i, x_i) = \sigma_{gx} \neq 0$, $\mathrm{cov}(g_i, z_i) = 0$ et $\mathrm{cov}(g_i, \varepsilon_i) = 0$, donc $\mathrm{cov}(g_i, u_i) = \beta_2 \,\mathrm{cov}(g_i, z_i) + \mathrm{cov}(g_i, \varepsilon_i) = 0$. On peut alors construire l'estimateur des **variables instrumentales**, en instrumentant x_i par g_i :

$$\hat{b}_1^{\mathrm{VI}} = \frac{\frac{1}{N}\sum_i (g_i - \bar{g})(y_i - \bar{y})}{\frac{1}{N}\sum_i (g_i - \bar{g})(x_i - \bar{x})}$$

En substituant le DGP (variables centrées) :

$$\hat{b}_1^{\mathrm{VI}} = \beta_1 + \frac{\frac{1}{N}\sum_i g_i u_i}{\frac{1}{N}\sum_i g_i x_i}$$

Par la loi des grands nombres, $\frac{1}{N}\sum_i g_i u_i \xrightarrow{p} \mathrm{cov}(g_i, u_i) = 0$ et $\frac{1}{N}\sum_i g_i x_i \xrightarrow{p} \sigma_{gx} \neq 0$, d'où par le théorème de Slutsky :

$$\mathrm{plim}_{N\to\infty}\, \hat{b}_1^{\mathrm{VI}} = \beta_1 + \frac{0}{\sigma_{gx}} = \beta_1$$

L'estimateur des VI est donc convergent : l'instrument permet d'extraire la part de la variation des études x_i qui est orthogonale au talent omis, et de restaurer l'identification du rendement de l'éducation.

(5) Un bon instrument doit satisfaire deux conditions :

— **Pertinence** : l'instrument doit être corrélé avec le régresseur endogène, $\mathrm{cov}(g_i, x_i) = \sigma_{gx} \neq 0$ (sinon le dénominateur s'annule à la limite et l'estimateur n'est pas défini). C'est une condition vérifiable sur les données (force de la première étape).

— **Validité (exogénéité)** : l'instrument doit être non corrélé avec l'erreur du modèle, $\mathrm{cov}(g_i, u_i) = 0$, donc ici non corrélé à la fois avec le talent z_i et avec ε_i. L'éducation des parents ne doit influencer le salaire de l'individu qu'*à travers* sa propre éducation (hypothèse non testable en juste identification, à justifier économiquement).

Non, l'estimateur des VI n'est pas sans biais en échantillon fini : il s'écrit comme un rapport de variables aléatoires $\hat{b}_1^{\mathrm{VI}} = \beta_1 + (\frac{1}{N}\sum_i g_i u_i)/(\frac{1}{N}\sum_i g_i x_i)$, et l'espérance d'un rapport n'est pas le rapport des espérances (le numérateur et le dénominateur ne sont pas indépendants). Il est seulement *convergent* (sans biais asymptotiquement) : on renonce à l'absence de biais en petit échantillon pour corriger l'endogénéité.