ÉCONOMÉTRIE APPROFONDIE

(Éléments de correction)

Jeudi 18 décembre 2025

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

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

où $x_{i,k}$, pour $k = 1, \ldots, K$, sont des variables explicatives déterministes, β_k, pour $k = 1, \ldots, K$, sont des paramètres réels, ε_i une variable aléatoire i.i.d. centrée de variance σ_ε^2.

(1) On obtient la représentation matricielle de ce modèle en concaténant (verticalement) les N observations. On pose : $Y = (y_1, y_2, \ldots, y_N)'$, $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N)'$, $\beta = (\beta_1, \beta_2, \ldots, \beta_K)'$ 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 rassemble les N observations d'une variable explicative). On peut alors réécrire le modèle sous la forme :

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

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

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

(le terme $Y'Y$ ne dépend pas de b et les scalaires $b'X'Y$ et $Y'Xb$ sont égaux). La condition nécessaire d'optimalité (en dérivant par rapport au vecteur b) est :

$$\begin{aligned} 2X'X\hat{b} - 2X'Y = 0 \\ \Leftrightarrow \hat{b} = \left(X'X\right)^{-1} X'Y \end{aligned}$$

Pour que cet estimateur existe, il faut que la matrice $X'X$ soit inversible. C'est le cas si, et seulement si, X est de plein rang colonne ($\mathrm{rg}(X) = K$), ce qui suppose en particulier qu'il n'y a pas de colinéarité parfaite entre les variables explicatives et que $N \geq K$. La condition du second ordre est satisfaite car $X'X$ est semi-définie positive (définie positive sous l'hypothèse de plein rang), de sorte que le point critique est bien un minimum.

(3) $\hat{b}$ est un estimateur sans biais de β. Pour le montrer, substituons le processus générateur des données (DGP) dans l'expression de l'estimateur :

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

Comme les variables explicatives sont déterministes (sinon il faudrait recourir au théorème des espérances itérées), on a donc :

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

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

(4) La variance de $\hat{b}$ est définie par $\mathbb{V}\left[\hat{b}\right] = \mathbb{E}\left[\left(\hat{b} - \mathbb{E}\left[\hat{b}\right]\right)\left(\hat{b} - \mathbb{E}\left[\hat{b}\right]\right)'\right]$. En utilisant l'expression $\hat{b} - \beta = \left(X'X\right)^{-1} X'\varepsilon$ obtenue plus haut, et en notant que les ε_i sont i.i.d. (donc $\mathbb{V}[\varepsilon] = \sigma_\varepsilon^2 I_N$), il vient :

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

(les variables explicatives étant déterministes, la matrice $(X'X)^{-1} X'$ sort de l'espérance).

(5) L'estimateur des MCO est une variable aléatoire, et a donc une variance, parce que $\hat{b} = \beta + (X'X)^{-1} X'\varepsilon$ est une fonction (linéaire) du vecteur aléatoire ε. La matrice X étant déterministe, c'est uniquement l'aléa des erreurs, transmis par l'échantillon, qui rend $\hat{b}$ aléatoire. D'un échantillon (d'une réalisation de ε) à l'autre, la valeur de l'estimateur change : cette dispersion autour de β est mesurée par la variance.

(6) Pour montrer la convergence en probabilité, partons de :

$$\hat{b} = \beta + (X'X)^{-1} X'\varepsilon = \beta + \left(\frac{X'X}{N}\right)^{-1} \frac{X'\varepsilon}{N}$$

On suppose que la matrice des moments empiriques des régresseurs admet une limite finie et inversible :

$$\frac{X'X}{N} \xrightarrow[N\to\infty]{} Q$$

où Q est une matrice $K \times K$ définie positive. Par ailleurs, le terme $\frac{1}{N}X'\varepsilon$ est d'espérance nulle (les X sont déterministes et $\mathbb{E}[\varepsilon] = 0$), et sa variance tend vers zéro :

$$\mathbb{V}\left[\frac{X'\varepsilon}{N}\right] = \frac{\sigma_\varepsilon^2}{N}\frac{X'X}{N} \xrightarrow[N\to\infty]{} 0$$

Le vecteur $\frac{1}{N}X'\varepsilon$ converge donc en moyenne quadratique, et par conséquent en probabilité, vers 0 (c'est une application de la loi des grands nombres : $\frac{1}{N}X'\varepsilon$ est une moyenne d'erreurs centrées). Par continuité de l'inversion matricielle et du produit (théorème de Slutsky), il vient :

$$\hat{b} \xrightarrow[N\to\infty]{\text{proba}} \beta + Q^{-1} \times 0 = \beta$$

On peut aussi le voir directement : la variance de l'estimateur $\mathbb{V}[\hat{b}] = \sigma_\varepsilon^2 (X'X)^{-1} = \frac{\sigma_\varepsilon^2}{N}(X'X/N)^{-1}$ tend vers la matrice nulle quand $N \to \infty$. Comme $\hat{b}$ est sans biais et que sa variance s'annule à la limite, $\hat{b}$ converge en moyenne quadratique, donc en probabilité, vers β : l'estimateur des MCO est convergent.

EXERCICE 2. On considère le modèle générateur des données :

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

avec β un vecteur colonne de paramètres et x_i un vecteur ligne de variables explicatives déterministes. Les erreurs sont normalement distribuées :

$$\varepsilon = (\varepsilon_1, \ldots, \varepsilon_N)' \sim \mathcal{N}(\mathbf{0}, \Omega)$$

avec Ω symétrique et définie positive. Sous forme matricielle, $Y = X\beta + \varepsilon$, et le modèle empirique est $y_i = x_i b + \epsilon_i$.
(1) Oui, l'estimateur des MCO $\hat{b} = (X'X)^{-1} X'Y$ est sans biais. En substituant le DGP :

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

et comme X est déterministe et $\mathbb{E}[\varepsilon] = 0$:

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

L'absence de biais ne dépend pas de la structure de Ω : elle tient uniquement à l'exogénéité (ici la nature déterministe) des régresseurs et à la nullité de l'espérance des erreurs.
(2) En reprenant $\hat{b} - \beta = (X'X)^{-1} X'\varepsilon$ et en utilisant $\mathbb{V}[\varepsilon] = \Omega$:

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

Contrairement au cas sphérique ($\Omega = \sigma^2 I_N$), la matrice Ω ne se simplifie plus avec les $X'X$: on obtient une expression « en sandwich ».
(3) Non, l'estimateur des MCO n'est pas efficace dans la classe des estimateurs linéaires et sans biais. Le théorème de Gauss-Markov ne s'applique en effet que lorsque la matrice de variance-covariance des erreurs est scalaire ($\Omega = \sigma^2 I_N$, homoscédasticité et absence d'autocorrélation). Ici Ω est quelconque (symétrique définie positive) : il existe un autre estimateur linéaire sans biais, l'estimateur des moindres carrés généralisés (MCG, théorème d'Aitken), dont la variance est plus faible (la différence des deux matrices de variance est semi-définie positive). Les MCO ne sont donc pas BLUE.
(4) Comme Ω est symétrique définie positive, son inverse l'est aussi et admet une factorisation $\Omega^{-1} = P'P$ (par exemple de Cholesky), avec P une matrice $N \times N$ inversible. En prémultipliant le modèle par P :

$$\underbrace{PY}_{\tilde{Y}} = \underbrace{PX}_{\tilde{X}}\beta + \underbrace{P\varepsilon}_{\tilde{\varepsilon}}$$

le modèle transformé a des erreurs sphériques :

$$\mathbb{V}[\tilde{\varepsilon}] = P\,\mathbb{V}[\varepsilon]\,P' = P\Omega P' = P(P'P)^{-1}P' = I_N$$

On peut donc appliquer les MCO au modèle transformé : l'estimateur obtenu est alors efficace (théorème de Gauss-Markov). C'est l'estimateur des

MCG :

$$\begin{aligned}\hat{b}_{\text{MCG}} &= \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{Y} \\ &= \left(X'P'PX\right)^{-1}X'P'PY \\ &= \left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}Y\end{aligned}$$

Sa variance s'obtient comme celle d'un estimateur des MCO appliqué au modèle transformé (dont les erreurs ont pour variance I_N) :

$$\mathbb{V}\left[\hat{b}_{\text{MCG}}\right] = \left(\tilde{X}'\tilde{X}\right)^{-1} = \left(X'\Omega^{-1}X\right)^{-1}$$

On vérifie qu'il s'agit bien d'un estimateur sans biais :

$$\hat{b}_{\text{MCG}} = \beta + \left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}\varepsilon$$

d'espérance β.

(5) Si Ω est inconnue, on ne peut pas l'estimer librement : elle comporte $N(N+1)/2$ paramètres distincts, soit plus que le nombre d'observations. Il faut donc en spécifier la structure, c'est-à-dire faire dépendre Ω d'un petit nombre de paramètres : $\Omega = \Omega(\theta)$ (par exemple une forme d'hétéroscédasticité $\sigma_i^2 = h(z_i, \theta)$, ou une structure d'autocorrélation des erreurs). On procède alors en deux étapes (MCG *réalisables*, FGLS) :

- estimer le modèle par les MCO, récupérer les résidus $\hat{\epsilon}_i$ et en déduire une estimation $\hat{\theta}$ des paramètres de Ω, d'où une matrice estimée $\hat{\Omega} = \Omega(\hat{\theta})$;
- calculer l'estimateur des MCG réalisables en remplaçant Ω par $\hat{\Omega}$:

$$\hat{b}_{\text{FGLS}} = \left(X'\hat{\Omega}^{-1}X\right)^{-1}X'\hat{\Omega}^{-1}Y$$

Cet estimateur n'est en général plus sans biais en échantillon fini (car $\hat{\Omega}$ dépend des données, donc de ε), mais il reste convergent et asymptotiquement équivalent aux MCG si $\hat{\Omega}$ converge vers Ω. On peut éventuellement itérer la procédure jusqu'à convergence des estimations.

(6) Oui, l'estimation par maximum de vraisemblance est possible puisque la loi des erreurs est spécifiée (gaussienne). Comme $Y \sim \mathcal{N}(X\beta, \Omega)$, la log-vraisemblance est :

$$\ln L(\beta, \Omega) = -\frac{N}{2}\ln 2\pi - \frac{1}{2}\ln|\Omega| - \frac{1}{2}(Y - X\beta)'\,\Omega^{-1}(Y - X\beta)$$

À Ω donnée, maximiser $\ln L$ par rapport à β revient à minimiser la forme quadratique $(Y - X\beta)' \, \Omega^{-1} \, (Y - X\beta)$. La condition du premier ordre

$$X'\Omega^{-1} \, (Y - X\beta) = 0$$

redonne exactement l'estimateur des MCG :

$$\hat{\beta}_{\text{MV}} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y = \hat{b}_{\text{MCG}}$$

Lorsque Ω est inconnue (mais paramétrée par θ), on maximise conjointement $\ln L$ par rapport à β et θ. Sous normalité, le maximum de vraisemblance et les MCG coïncident donc pour la partie β.

EXERCICE 3. Sur le marché d'un bien, la demande et l'offre à la date t sont données par le modèle structurel :

$$\begin{cases} y_t^d = \alpha_0 + \alpha_1 P_t + \alpha_2 R_t + u_t \\ y_t^o = \gamma_0 + \gamma_1 P_t + \gamma_2 W_t + v_t \end{cases}$$

où P_t est le prix, R_t le revenu des consommateurs, W_t les coûts de production, u_t et v_t des chocs i.i.d., centrés, de variances σ_u^2 et σ_v^2 et non corrélés. Les échanges ont lieu à l'équilibre ($y_t^d = y_t^o \equiv y_t$). Les coûts de production sont orthogonaux au choc de demande ($\text{cov}(W_t, u_t) = 0$) et le revenu est orthogonal aux deux chocs.

(1) Signes attendus. Pour la demande : $\alpha_1 < 0$ (la demande décroît avec le prix) et $\alpha_2 > 0$ (la demande croît avec le revenu, bien normal). Pour l'offre : $\gamma_1 > 0$ (l'offre croît avec le prix) et $\gamma_2 < 0$ (une hausse des coûts de production réduit l'offre). Les constantes α_0 et γ_0 n'ont pas de signe attendu a priori.

(2) À l'équilibre, $y_t^d = y_t^o$, donc le prix doit être tel que :

$$\alpha_0 + \alpha_1 P_t + \alpha_2 R_t + u_t = \gamma_0 + \gamma_1 P_t + \gamma_2 W_t + v_t$$

En regroupant les termes en P_t :

$$(\alpha_1 - \gamma_1) P_t = (\gamma_0 - \alpha_0) + \gamma_2 W_t - \alpha_2 R_t + v_t - u_t$$

soit, comme $\alpha_1 \neq \gamma_1$ (les pentes d'offre et de demande sont de signes opposés) :

$$P_t = \frac{(\gamma_0 - \alpha_0) + \gamma_2 W_t - \alpha_2 R_t + v_t - u_t}{\alpha_1 - \gamma_1}$$

Le prix d'équilibre est une fonction (affine) des chocs structurels u_t et v_t : c'est donc une variable aléatoire. C'est la *forme réduite* de l'équation du

prix. (En substituant P_t dans l'une des équations, on obtiendrait de même la quantité d'équilibre y_t, elle aussi aléatoire et corrélée aux deux chocs.)
(3) Non, ce n'est pas une bonne idée. Le régresseur P_t de l'équation de demande est corrélé avec le terme d'erreur u_t de cette même équation. En effet, comme u_t et v_t ne sont pas corrélés :

$$\operatorname{cov}(P_t, u_t) = \operatorname{cov}\left(\frac{v_t - u_t}{\alpha_1 - \gamma_1}, u_t\right) = \frac{-\sigma_u^2}{\alpha_1 - \gamma_1} = \frac{\sigma_u^2}{\gamma_1 - \alpha_1} \neq 0$$

Le prix est donc une variable *endogène* : un choc de demande u_t déplace la courbe de demande, ce qui modifie le prix d'équilibre. La condition d'orthogonalité entre régresseur et erreur n'est pas vérifiée, et l'estimateur des MCO de l'équation de demande est biaisé et non convergent (biais de simultanéité / d'endogénéité). En revanche, le revenu R_t, supposé orthogonal aux chocs, n'est pas problématique.
(4) Il faut recourir à une méthode de variables instrumentales (VI), estimée par les doubles moindres carrés (2SLS). Pour instrumenter le prix P_t, il faut une variable corrélée avec P_t mais orthogonale au choc de demande u_t. Les coûts de production W_t sont un instrument idéal :
— W_t est corrélé avec P_t, car W_t apparaît dans la forme réduite du prix (avec le coefficient $\gamma_2/(\alpha_1 - \gamma_1) \neq 0$, l'instrument est pertinent) ;
— W_t est orthogonal à u_t par hypothèse (validité de l'instrument) : les coûts de production n'affectent la demande qu'à travers le prix.

L'équation de demande $y_t = \alpha_0 + \alpha_1 P_t + \alpha_2 R_t + u_t$ contient une variable endogène (P_t) et une variable exogène incluse (R_t, qui s'instrumente elle-même). On dispose d'un instrument exclu (W_t) : le modèle est *juste identifié*. Concrètement, en notant $Z_t = (1, R_t, W_t)$ le vecteur des instruments et $X_t = (1, P_t, R_t)$ les régresseurs, l'estimateur des VI est :

$$\hat{\alpha}_{\mathrm{VI}} = (Z'X)^{-1} Z'Y$$

(ou, de façon équivalente, on régresse d'abord P_t sur $(1, R_t, W_t)$ pour obtenir $\hat{P}_t$ — première étape — puis on régresse y_t sur $(1, \hat{P}_t, R_t)$ — seconde étape). Cet estimateur est convergent car les instruments sont valides et pertinents : $\frac{1}{T} Z'u \xrightarrow{\text{proba}} 0$ et $\frac{1}{T} Z'X$ converge vers une matrice inversible.
(5) Non, l'estimateur des variables instrumentales n'est pas sans biais en échantillon fini. Il s'écrit comme un rapport de variables aléatoires :

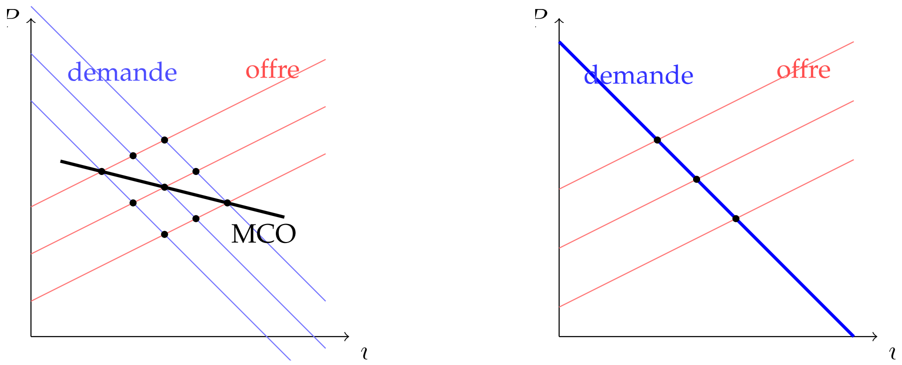
$$\hat{\alpha}_{\mathrm{VI}} = \alpha + (Z'X)^{-1} Z'u$$

et l'espérance de ce rapport n'est pas le rapport des espérances : en général,

$$\mathbb{E}\left[(Z'X)^{-1} Z'u\right] \neq 0$$

car $Z'X$ et $Z'u$ ne sont pas indépendants. L'estimateur des VI est seulement *convergent* (sans biais asymptotiquement) : c'est précisément parce qu'on accepte de renoncer à l'absence de biais en petit échantillon que l'on peut corriger le problème d'endogénéité.

Illustration graphique. On représente les couples (y_t, P_t) effectivement observés (les équilibres de marché) dans le plan quantité-prix. À chaque date, l'équilibre est l'intersection d'une courbe de demande et d'une courbe d'offre, toutes deux déplacées par leurs chocs respectifs.



(a) Les deux courbes se déplacent.

(b) Seule l'offre se déplace (via W).

Sur le panneau **(a)**, la demande est déplacée par les chocs u_t et l'offre par v_t (et W_t) : les équilibres forment un nuage de points dont la dispersion ne reproduit ni la courbe de demande ni la courbe d'offre. La droite des MCO de y sur P ajustée sur ce nuage a une pente comprise entre celle de la demande et celle de l'offre (c'est une moyenne des deux pentes structurelles, pondérée par les variances des chocs) : elle ne restitue donc pas la pente α_1 de la demande. C'est l'expression graphique du biais de simultanéité.

Sur le panneau **(b)**, on isole la variation du prix provoquée par l'instrument W_t : lorsque seule l'offre se déplace (la demande restant stable, u_t orthogonal à W_t), les équilibres successifs *glissent le long de la courbe de demande*, qui devient alors identifiable. C'est exactement ce que réalise l'estimateur des variables instrumentales : il n'exploite que la part des fluctuations du prix imputable à W_t, qui est orthogonale au choc de demande, pour estimer de façon convergente la pente de la demande.