L3 Econometrie **Contrôle Continu N2**

Linas Tarasonis **le 13 avril, 2012**

Les exercices sont indépendants. Le barème est indicatif. L'utilisation de documents, calculatrices, téléphones portables ou tout autre appareil éléctronique n'est pas autorisée. Les réponses devront être soigneusement argumentées et justifiées. Vous pouvez laisser les résultats sous la forme de fractions.

Question 1 (5 points)

On considère le modèle $y_n = b_0 + b_1 x_{n1} + ... + b_p x_{np} + \epsilon_n$ pour lequel on dispose de N observations, et que l'on note aussi sous la forme $y = X\beta + \epsilon$. On suppose que les ϵ_n sont i.i.d. de loi $N(0, \sigma^2)$.

i. (1 point) Donner la formule pour l'estimateur des MCO de β, $\hat{\beta}$.

Solution:

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

ii. (1 point) Montrer que c'est un estimateur sans biais de β.

Solution:

$$\mathbb{E}[\hat{\beta}] = \mathbb{E}[(X'X)^{-1}X'y] = (X'X)^{-1}X'\mathbb{E}[y] = (X'X)^{-1}X'\mathbb{E}[X\beta + \epsilon]$$

et puisque $\mathbb{E}[\epsilon] = 0$, il vient:

$$\mathbb{E}[\hat{\beta}] = (X'X)^{-1}X'\mathbb{E}[X\beta] = (X'X)^{-1}X'X\beta = \beta.$$

iii. (2 points) Quelle est la loi de $\hat{\beta}$?

Solution:

$$Var(\hat{\beta}) = Var((X'X)^{-1}X'y) = (X'X)^{-1}X'Var(y)X(X'X)^{-1}$$

or $Var(y) = Var(X\beta + \epsilon) = Var(\epsilon) = \sigma^2 I_n$, donc:

$$Var(\hat{\beta}) = \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1}.$$

D'où

$$\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2(X'X)^{-1}).$$

iv. (1 point) Qu'appelle-t-on "résidus estimés" ? Donnez la formule permettant de calculer le vecteur des résidus estimés.

Solution: Les résidus estimés sont définis par:

$$\hat{\epsilon} = y - \hat{y}.$$

La formule permettant de calculer ce vecteur est la suivante:

$$\hat{\epsilon} = y - X\hat{\beta} = y - X(X'X)^{-1}X'y.$$

Question 2 (15 points)

Une étude statistique a été menée sur un groupe de 149 personnes pour analyser leurs niveau d'études. Le modèle suivant a été proposé:

$$educ_n = b_0 + b_1 sibs_n + b_2 meduc_n + b_3 feduc_n + \epsilon_n$$

dans lequel *educ* est le nombre d'années d'études effectuées par l'individu n, *sibs* est le nombre d'enfants de mêmes parents, *meduc* est le nombre d'années d'études effectuées par la mère de l'individu n, et $feduc$ est le nombre d'années d'études effectuées par le père de l'individu n. Les résultats suivants ont été obtenus:

$$\hat{\beta} = \begin{pmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \end{pmatrix} = \begin{pmatrix} 9 \\ -0.10 \\ 0.15 \\ 0.20 \end{pmatrix}$$

et la matrice de variance-covariance estimée de $\hat{\beta}$ est

$$\hat{V}\hat{\beta} = \begin{pmatrix} 3.0 & 1.5 & 0.5 & 2 \\ 1.5 & 0.05 & -0.05 & -0.1 \\ 0.5 & -0.05 & 0.05 & 0.75 \\ 2 & -0.1 & 0.75 & 0.1 \end{pmatrix}$$

(a) (1 point) Est-ce que le nombre d'enfants de mêmes parents (*sibs*) a un impact attendu? Expliquer.

Solution: Un enfant supplémentaire de mêmes parents baisse le nombre d'années d'études effectuées par l'individu n par -0.1 an. Cet impact est attendu sous l'hypothèses que les études sont coûteuses et les parents les financent.

(b) (1 point) De combien il faut augmenter le nombre d'enfants de mêmes parents (*sibs*) pour réduire $\hat{educ}$ de 1 an à niveau d'études des parents donné?

Solution: Il faut augmenter le nombre d'enfants de mêmes parents (*sibs*) par 10 ($-0.1 \times 10 = -1$).

(c) (1 point) Expliquer soigneusement comment on doit interpréter le coefficient de *meduc*.

Solution:

$$b_2 = \frac{\partial educ_n}{\partial meduc_n}$$

et donc $b_2 = 0.15$ veut dire qu'une année supplémentaire d'études effectuées par la mère de l'individu n augmente le nombre d'années d'études effectuées par l'individu n de 0.15 ans, i.e. ≈ 1.8 mois.

(d) (1 point) Supposons que un Homme A n'a pas de frères et soeurs et que son père et sa mère ont effectué 12 ans d'études. Un Homme B n'a pas de frères et soeurs non plus, mais son père et sa mère ont effectué 16 ans d'études. Quelle est la différence de niveau d'études ateint entre un Homme A et un Homme B prédite par le modèle?

Solution: Le niveau d'études esperé par un Homme A:

$$\mathbb{E}[educ_A | sibs_A = 0, meduc_A = feduc_A = 12] = 9 - 0.1 \times 0 + 0.15 \times 12 + 0.2 \times 12 = 13.2$$

Le niveau d'études esperé par un Homme B:

$$\mathbb{E}[educ_B | sibs_B = 0, meduc_B = feduc_B = 16] = 9 - 0.1 \times 0 + 0.15 \times 16 + 0.2 \times 16 = 14.6$$

La différence prédite est donc:

$$\mathbb{E}[educ_A | sibs_A = 0, meduc_A = feduc_A = 12] - \mathbb{E}[educ_B | sibs_B = 0, meduc_B = feduc_B = 16] = -1.4$$

i.e. 1 an et 3 mois.

(e) (2 points) On suppose que les résidus sont i.i.d. de $N(0,\sigma^2)$. Testez la significativité des différents coefficients au seuil 1%.

Solution: Test de significativité de b_i: $H_0 : b_i = 0$ contre $H_1 : b_i \neq 0$ pour $i \in \{0,1,2,3\}$. On a vu que la statistique du test est $t_{\hat{b}_i} = \frac{\hat{b}_i}{\hat{\sigma}_{\hat{b}_i}} \sim \mathcal{T}(N-4)$. La région critique est determinée par $\mathbb{P}_{H_0}(|t_{\hat{b}_i}| > k) = 1\%$. On en déduit que $k = t_{0.995}(145) \approx 2.61$.

- $|t_{\hat{b}_0}| = |\frac{\hat{b}_0}{\hat{\sigma}_{\hat{b}_0}}| = |\frac{9}{\sqrt{3}}| = |\frac{9}{1.7}| > 2.61$: b_0 est significatif.
- $|t_{\hat{b}_1}| = |\frac{\hat{b}_1}{\hat{\sigma}_{\hat{b}_1}}| = |\frac{-0.1}{\sqrt{0.05}}| = |\frac{-0.1}{0.2}| < 2.61$: b_1 n'est pas significatif.
- $|t_{\hat{b}_2}| = |\frac{\hat{b}_2}{\hat{\sigma}_{\hat{b}_2}}| = |\frac{0.15}{\sqrt{0.05}}| = |\frac{0.15}{0.2}| < 2.61$: b_2 n'est pas significatif.
- $|t_{\hat{b}_3}| = |\frac{\hat{b}_3}{\hat{\sigma}_{\hat{b}_3}}| = |\frac{0.20}{\sqrt{0.1}}| = |\frac{0.2}{0.3}| < 2.61$: b_3 n'est pas significatif.

(f) (2 points) Testez, au seuil 1%, l'hypothèse $H_0 : b_0 \geq 10$ contre l'hypothèse $H1 : b_0 < 10$.

Solution: Le test a même région de critique que $H_0' : b_0 = 10$ contre $H1 : b_0 < 10$ (cf. cours). On refuse H_0' et donc H_0 si $\frac{\hat{b}_0 - 10}{\hat{\sigma}_{\hat{b}_0}} < k$ avec k determiné par $\mathbb{P}_{H_0'}(\frac{\hat{b}_0-10}{\hat{\sigma}_{\hat{b}_0}} < k) = 1\%$ cad $\mathbb{P}_{H_0'}(T < k) = 1\%$ pour $T \sim \mathcal{T}(N-4)$. Comme $\mathbb{P}_{H_0'}(T < k) = \mathbb{P}_{H_0'}(T > -k)$, on en déduit que $-k = t_{0.99}(145) = 2.35$. Comme $\frac{\hat{b}_0-10}{\hat{\sigma}_{\hat{b}_0}} = \frac{9-10}{1.7} \approx -0.6$ on accepte H_0' et donc H_0.

(g) i. (1 point) Quelle est la loi de $\hat{b}_0 + 3\hat{b}_1$?

Solution:

$$\hat{b}_0 + 3\hat{b}_1 \sim \mathcal{N}(\mathbb{E}[\hat{b}_0 + 3\hat{b}_1], Var(\hat{b}_0 + 3\hat{b}_1)).$$

Comme

$$\mathbb{E}[\hat{b}_0 + 3\hat{b}_1] = b_0 + 3b_1$$

et

$$Var(\hat{b}_0 + 3\hat{b}_1) = Var(\hat{b}_0) + 9Var(\hat{b}_1) + 6Cov(\hat{b}_0, \hat{b}_1) = \sigma^2(w_{11} + 9w_{22} + 6w_{12}) = \sigma^2\lambda$$

où w_{ij} est le terme général de matrice $(X'X)^{-1}$. Et donc:

$$\hat{b}_0 + 3\hat{b}_1 \sim \mathcal{N}(b_0 + 3b_1, \sigma^2\lambda).$$

ii. (1 point) Calculez la variance estimée de $\hat{b}_0 + 3\hat{b}_1$.

Solution:

$$\hat{Var}(\hat{b}_0 + 3\hat{b}_1) = \hat{\sigma}^2(w_{11} + 9w_{22} + 6w_{12}) = \hat{\sigma}^2\lambda = 3 + 9 * 0.05 + 6 * 1.5 = 12.45$$

iii. (2 points) Construisez le test de l'hypothèse $H_0 : b_0 + 3b_1 = 0$ contre l'hypothèse $H_1 : b_0 + 3b_1 \neq 0$ et effectuez ce test au seuil 5%.

Solution: On rejette H_0 si $|\hat{b}_0 + 3\hat{b}_1| > K$ avec K déterminée par $\mathbb{P}_{H_0}(|\hat{b}_0 + 3\hat{b}_1| > k) = 5\%$. Sous H_0:

$$\hat{b}_0 + 3\hat{b}_1 \sim \mathcal{N}(0, \sigma^2\lambda).$$

Alors

$$\frac{\hat{b}_0 + 3\hat{b}_1}{\sigma\sqrt{\lambda}} \sim \mathcal{N}(0,1).$$

Et donc

$$\frac{\hat{b}_0 + 3\hat{b}_1}{\hat{\sigma}\sqrt{\lambda}} \sim \mathcal{T}(145).$$

On rejette H_0 si $\frac{|\hat{b}_0+3\hat{b}_1|}{\hat{\sigma}\sqrt{\lambda}} > k$ avec la région critique determinée par $\mathbb{P}_{H_0}(\frac{|\hat{b}_0+3\hat{b}_1|}{\hat{\sigma}\sqrt{\lambda}} > k) = 5\%$. On en déduit que $k = t_{0.975} = 1.97$. Comme $\frac{|8.7|}{\sqrt{12.45}} \approx 2.5 > 1.97$, on conclut $H_1 : \hat{b}_0 + 3\hat{b}_1 \neq 0$.

(h) (3 points) On veux tester en plus si l'impact du nombre d'années d'études effectuées par le père dépend de sexe de l'individu n. Proposez le modèle pour effectuer ce test.[1]

Solution: Il faut estimer un modèle suivant:

$$educ_n = b_0 + b_1 sibs_n + b_2 meduc_n + b_3 feduc_n + b_4 feduc_n \times female_n + \epsilon_n$$

L'impact du nombre d'années d'études effectuées par le père est

$$\frac{\partial educ_n}{\partial feduc_n} = b_3 + b_4 \times female_n.$$

Et donc le test $H_0 : b_4 = 0$ contre $H_1 : b_4 \neq 0$ nous dira si l'impact du nombre d'années d'études effectuées par le père dépend de sexe de l'individu n.

[1] Soit $female_n$ une variable égale à 1 si l'individu n est une femme.