Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2014 – 2015

Econométrie II

Contrôle continu n°1, mars 2015

Examen de 1h30. Tout document ou calculatrice est interdit.

(Sur 30 points) Soit $(Y_i)_{1\leq i\leq n}$ une famille de variables aléatoires observées définie par:

$$Y_i = \theta_0 + \theta_1 Z_i + \varepsilon_i \qquad \text{pour tout } i \in \{1,\dots,n\}, \qquad \text{où:} \tag{1}$$

$\theta = {}^t(\theta_0,\theta_1)$ est un vecteur composé de 2 réels inconnus, les $(Z_i)_{1\leq i\leq n}$ sont n variables aléatoires indépendantes et identiquement distribuées (v.a.i.i.d.) de même loi que Z_0 et vérifiant $0 < \mathrm{var}(Z_0) = \sigma_Z^2 < \infty$ et $m_Z = \mathbb{E} Z_0$, l'échantillon $(Z_1,\cdots,Z_n)$ étant observé. La suite $(\varepsilon_i)_i$ est une suite non observée de v.a.i.i.d. de loi centrée et de variance $\sigma_\varepsilon^2 > 0$, indépendante de $(Z_i)_{1\leq i\leq n}$. Dans la suite, on note $\mathbf{1}$ le vecteur de $\mathbf{R}^n$ constitué uniquement de 1, $[\mathbf{1}]$ la droite vectorielle de $\mathbf{R}^n$ engendrée par $\mathbf{1}$ et $[\mathbf{1}]^\perp$ son orthogonal dans $\mathbf{R}^n$.

1. On note $Y = (Y_i)_{1\leq i\leq n}$, $Z = (Z_i)_{1\leq i\leq n}$ et $\varepsilon = (\varepsilon_i)_{1\leq i\leq n}$. Ecrire le modèle (1) sous une forme vectorielle $Y = X\theta + \varepsilon$ en précisant ce qu'est la matrice X **(1pt)**.
2. Montrer que pour tout $C \in \mathbf{R}$, $\mathbb{P}(Z_0 = C) < 1$ **(1.5pts)**. En déduire que $\mathbb{P}\big(\mathrm{Rang}(X) = 2\big) \underset{n\to\infty}{\longrightarrow} 1$ **(1.5pts)**. On supposera désormais que $\mathrm{Rang}(X) = 2$.
3. Soit $\widehat{\theta}$ l'estimateur par moindres carrés de θ. Donner l'expression de $\sqrt{n}\,(\widehat{\theta} - \theta)$ en fonction de n, X et ε **(1pt)**. Montrer que: $n\,({}^tX\,X)^{-1} \underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}} Q$, où $Q = \frac{1}{\sigma_Z^2}\begin{pmatrix} \sigma_Z^2 + m_Z^2 & -m_Z \\ -m_Z & 1 \end{pmatrix}$ **(2.5pts)**. Montrer que le vecteur aléatoire $\frac{1}{\sqrt{n}}\,{}^t\big(\sum_{i=1}^n \varepsilon_i\ ,\ \sum_{i=1}^n \varepsilon_i Z_i\big)$ tend en loi vers une loi gaussienne centrée de dimension 2 dont on précisera la matrice de covariance **(3pts)**. En utilisant le Lemme de Slutski, en déduire que $\sqrt{n}\big(\widehat{\theta} - \theta\big) \underset{n\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}_2\big(0\,,\,\sigma_\varepsilon^2 Q\big)$ **(2.5pts)**.
4. On note respectivement $P_{[\mathbf{1}]}$ et $P_{[\mathbf{1}]^\perp}$ les matrices de projection orthogonale sur $[\mathbf{1}]$ et sur $[\mathbf{1}]^\perp$. Montrer que $Y - P_{[\mathbf{1}]}Y = \theta_1 P_{[\mathbf{1}]^\perp} Z + P_{[\mathbf{1}]^\perp}\varepsilon$ **(1.5pts)**. En déduire que $\|Y - P_{[\mathbf{1}]}Y\|^2 = \theta_1^2 \|P_{[\mathbf{1}]^\perp} Z\|^2 + \|P_{[\mathbf{1}]^\perp}\varepsilon\|^2 + 2\theta_1 U_n$ où l'on précisera U_n **(1.5pts)**. Montrer que $\frac{1}{n-1}\,\mathbb{E}(U_n^2) = \sigma_\varepsilon^2(m_Z^2 + \sigma_Z^2)$ (on pourra utiliser la trace de matrice...) **(3pts)**. En déduire que $\frac{1}{n-1}\,\|Y - P_{[\mathbf{1}]}Y\|^2 \underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}} \theta_1^2\sigma_Z^2 + \sigma_\varepsilon^2$ **(3pts)**.
5. Rappeler l'expression du coefficient de détermination $\mathbf{R}^2$ **(0.5pts)**. En utilisant les questions précédentes, montrer que $R^2 \underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}} \dfrac{\theta_1^2\,\sigma_Z^2}{\theta_1^2\,\sigma_Z^2 + \sigma_\varepsilon^2}$ **(2.5pts)**. Que pensez-vous de R^2 comme statistique de test pour un test d'adéquation du modèle linéaire **(1pt)**?
6. On se place dans une situation de sélection de modèle, où le modèle est (1) avec $\theta_1 \neq 0$, mais on ne sait pas si la variable Z doit être considérée dans le modèle. On utilise le critère BIC. Montrer, en utilisant les résultats précédents, que la probabilité que le critère BIC choisisse le modèle avec Z plutôt que le modèle sans Z tend vers 1 lorsque $n \to \infty$ **(4pts)**.

Proof.

1. On a X qui est la matrice $X = \big(\mathbf{1}\; Z\big)$.
2. S'il existe $C_0 \in \mathbf{R}$, tel que $\mathbb{P}(Z_0 = C_0) = 1$, alors $\mathbb{E}Z_0 = C_0$ et $\mathrm{var}(Z_0) = 0$, ce qui contredit les hypothèses.
   Si $\mathrm{Rang}(X) < 2$, alors Z est colinaire à $\mathbf{1}$ donc $\mathbb{P}(\mathrm{Rang}(X) < 2) = \mathbb{P}(Z_1 = Z_j\ \forall j = 1,\cdots,n)$. Mais $\mathbb{P}(Z_1 = Z_j\ \forall j = 1,\cdots,n) = \mathbb{E}\Big(\mathbb{P}(Z_1 = Z_j\ \forall j = 1,\cdots,n \mid Z_1)\Big)$. D'après ce qui précède, $\mathbb{P}\Big(Z_i = Z_1 \text{ pour tout } i = 1,\cdots,n \mid Z_1\Big) = \Big(\mathbb{P}(Z_0 = Z_1 \mid Z_1)\Big)^{n-1} \underset{n\to\infty}{\longrightarrow} 0$ puisque $\mathbb{P}(Z_0 = Z_1 \mid Z_1) < 1$. D'où $\mathbb{P}\Big(\mathrm{Rang}(X) = 2\Big) \underset{n\to\infty}{\longrightarrow} 1$.

3. On a $\widehat{\theta} = ({}^t X\, X)^{-1}\, {}^t X\, Y = \theta + ({}^t X\, X)^{-1}\, {}^t X\, \varepsilon$, d'où $\sqrt{n}\big(\widehat{\theta} - \theta\big) = \sqrt{n}\, ({}^t X\, X)^{-1}\, {}^t X\, \varepsilon$.
   On a $\frac{1}{n}\, {}^t X\, X = \begin{pmatrix} 1 & \frac{1}{n}\sum_{i=1}^n Z_i \\ \frac{1}{n}\sum_{i=1}^n Z_i & \frac{1}{n}\sum_{i=1}^n Z_i^2 \end{pmatrix}$. Mais d'après la loi forte des grands nombres, comme les Z_i sont des v.a.i.i.d. de variance finie, $\frac{1}{n}\sum_{i=1}^n Z_i \xrightarrow[n\to\infty]{p.s.} m_Z$ et $\frac{1}{n}\sum_{i=1}^n Z_i \xrightarrow[n\to\infty]{p.s.} \mathbb{E} Z_0^2 = \sigma_Z^2 + m_Z^2$. D'où $\frac{1}{n}\, {}^t X\, X \xrightarrow[n\to\infty]{p.s.} \begin{pmatrix} 1 & m_Z \\ m_Z & \sigma_Z^2 + m_Z^2 \end{pmatrix}$. Or on sait que pour n assez grand, X est de rang 2, donc la matrice est inversible et par continuité de l'inversion on a donc $\Big(\frac{1}{n}\, {}^t X\, X\Big)^{-1} = n\, ({}^t X\, X)^{-1} \xrightarrow[n\to\infty]{p.s.} \begin{pmatrix} 1 & m_Z \\ m_Z & \sigma_Z^2 + m_Z^2 \end{pmatrix}^{-1} = Q$. La convergence presque-sûre entraînant la convergence en probabilité on a donc le résultat.
   Soit la suite de vecteurs $\big({}^t(\varepsilon_i, \varepsilon_i Z_i)\big)_i$. Cette suite est une suite de vecteurs aléatoires indépendants et identiquement distribués. Ces vecteurs sont centrés puisque $\mathbb{E}[\varepsilon_i] = 0$ et $\mathbb{E}(\varepsilon_i Z_i) = \mathbb{E}(\varepsilon_i)\mathbb{E}(Z_i) = 0$ par l'indépendance. De plus, sa matrice de covariance existe et vaut $\begin{pmatrix} \mathrm{var}(\varepsilon_0) & \mathrm{cov}(\varepsilon_0, \varepsilon_0 Z_0) \\ \mathrm{cov}(\varepsilon_0, \varepsilon_0 Z_0) & \mathrm{var}(\varepsilon_0 Z_0) \end{pmatrix} = \sigma_\varepsilon^2 \begin{pmatrix} 1 & m_Z \\ m_Z & \sigma_Z^2 + m_Z^2 \end{pmatrix}$ car ε_0 est indépendant de Z_0 et centré, d'où $\mathrm{cov}(\varepsilon_0, \varepsilon_0 Z_0) = \mathbb{E}(\varepsilon_0^2 Z_0) = \mathbb{E}(\varepsilon_0^2)\mathbb{E}(Z_0) = \sigma_\varepsilon^2 m_Z$ et $\mathrm{var}(\varepsilon_0 Z_0) = \mathbb{E}(\varepsilon_0^2 Z_0^2) = \mathbb{E}(\varepsilon_0^2)\mathbb{E}(Z_0^2) = \sigma_\varepsilon^2(m_Z^2 + \sigma_Z^2)$. D'après le Théorème de la limite centrale, on en déduit que
   $$\sqrt{n}\,\Big(\frac{1}{n}\sum_{i=1}^n \begin{pmatrix} \varepsilon_i \\ Z_i \end{pmatrix}\Big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}_2\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma_\varepsilon^2 \begin{pmatrix} 1 & m_Z \\ m_Z & \sigma_Z^2 + m_Z^2 \end{pmatrix}\Big).$$
   On a $\sqrt{n}\,\Big(\frac{1}{n}\sum_{i=1}^n \begin{pmatrix} \varepsilon_i \\ Z_i \end{pmatrix}\Big) = \frac{1}{\sqrt{n}}\, {}^t X\, \varepsilon$. D'où $\sqrt{n}\big(\widehat{\theta} - \theta\big) = n\, ({}^t X\, X)^{-1} \times \sqrt{n}\,\Big(\frac{1}{n}\sum_{i=1}^n \begin{pmatrix} \varepsilon_i \\ Z_i \end{pmatrix}\Big)$. Comme $n\, ({}^t X\, X)^{-1}$ converge en probabilité vers une matrice constante et comme $\sqrt{n}\,\Big(\frac{1}{n}\sum_{i=1}^n \begin{pmatrix} \varepsilon_i \\ Z_i \end{pmatrix}\Big)$ converge en loi vers la loi gaussienne précédente, on peut utiliser le Lemme de Slutsky et la loi limite de $\sqrt{n}\big(\widehat{\theta} - \theta\big)$ est $Q \times \mathcal{N}_2\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma_\varepsilon^2 \begin{pmatrix} 1 & m_Z \\ m_Z & \sigma_Z^2 + m_Z^2 \end{pmatrix}\Big)$. On remarque que cette matrice de covariance limite est $\sigma_\varepsilon^2 Q^{-1}$. D'où le résultat final, la loi limite de $\sqrt{n}\big(\widehat{\theta} - \theta\big)$ étant $\mathcal{N}_2\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, Q \times \sigma_\varepsilon^2\, Q^{-1} \times {}^t Q\Big) = \mathcal{N}_2\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma_\varepsilon^2\, Q\Big)$.

4. On a $X\theta = \theta_0 \mathbf{1} + \theta_1 Z$. D'où $P_{[\mathbf{1}]} X\theta = \theta_0 \mathbf{1} + \theta_1 P_{[\mathbf{1}]} Z$. Ainsi, $Y - P_{[\mathbf{1}]} Y = (X\theta - P_{[\mathbf{1}]} X\theta) + P_{[\mathbf{1}]^\perp}\varepsilon = \theta_1 P_{[\mathbf{1}]^\perp} Z + P_{[\mathbf{1}]^\perp}\varepsilon$.
   En utilisant le produit scalaire euclidien classique de $\mathbf{R}^n$ tel que $< U, V >= {}^t UV$, on a $\|Y - P_{[\mathbf{1}]} Y\|^2 = \|\theta_1 P_{[\mathbf{1}]^\perp} Z + P_{[\mathbf{1}]^\perp}\varepsilon\|^2 = \theta_1^2 \|P_{[\mathbf{1}]^\perp} Z\|^2 + \|P_{[\mathbf{1}]^\perp}\varepsilon\|^2 + 2 < \theta_1 P_{[\mathbf{1}]^\perp} Z, P_{[\mathbf{1}]^\perp}\varepsilon >$. On a donc $U_n =< P_{[\mathbf{1}]^\perp} Z, P_{[\mathbf{1}]^\perp}\varepsilon >= {}^t Z P_{[\mathbf{1}]^\perp}\varepsilon$.
   On a $\mathbb{E}(U_n^2) = \mathbb{E}({}^t\varepsilon P_{[\mathbf{1}]^\perp} Z\, {}^t Z P_{[\mathbf{1}]^\perp}\varepsilon) = \mathbb{E}\big(\mathrm{trace}({}^t\varepsilon P_{[\mathbf{1}]^\perp} Z\, {}^t Z P_{[\mathbf{1}]^\perp}\varepsilon)\big)$ puisque l'on considère un réel. Comme $\mathrm{Trace}(AB) = \mathrm{Trace}(BA)$ et du fait de l'indépendance entre ε et Z, on a $\mathbb{E}(U_n^2)\ \mathrm{trace}\big(\mathbb{E}(\varepsilon\varepsilon^t)\, \mathbb{E}(Z Z^t P_{[\mathbf{1}]^\perp})\big) = \sigma_\varepsilon^2 \mathbb{E}(\ \mathrm{trace} Z Z^t P_{[\mathbf{1}]^\perp})) = \sigma_\varepsilon^2 \times (\sigma_Z^2 + m_Z^2) \times (n-1)$. D'où le résultat.
   L'inégalité de Markov implique que pour tout $\varepsilon > 0$, $\mathbb{P}(\frac{1}{n-1}|U_n| \geq \varepsilon) \leq \frac{1}{n-1}\, \mathbb{E}(\frac{1}{n-1} U_n^2)(\varepsilon)^{-2}$, donc d'après ce qui précède, on en déduit que $\frac{1}{n-1}|U_n| \xrightarrow[n\to+\infty]{\mathcal{P}} 0$. Par suite, $\frac{1}{n-1}\, \|Y - P_{[\mathbf{1}]} Y\|^2 = \frac{1}{n-1}\, (\theta_1^2 \|P_{[\mathbf{1}]^\perp} Z\|^2 + \|P_{[\mathbf{1}]^\perp}\varepsilon\|^2) + 2\frac{\theta_1}{n-1} U_n$. Un rapide calcul permet d'écrire que $\frac{1}{n-1}\, \|P_{[\mathbf{1}]^\perp} Z\|^2 = \frac{1}{n-1}\, \sum_{i=1}^n (Z_i - \overline{Z}_n)^2 \xrightarrow[n\to+\infty]{\mathcal{P}} \sigma_Z^2$ et $\frac{1}{n-1}\, \|P_{[\mathbf{1}]^\perp}\varepsilon\|^2 \xrightarrow[n\to+\infty]{\mathcal{P}} \sigma_\varepsilon^2$. En conséquence, en utilisant le Lemme Slutsky, on en déduit que $\frac{1}{n-1}\, \|Y - P_{[\mathbf{1}]} Y\|^2 \xrightarrow[n\to+\infty]{\mathcal{P}} \theta_1^2\, \sigma_Z^2 + \sigma_\varepsilon^2$.

5. On a $R^2 = 1 - \frac{\|Y - X\widehat{\theta}\|^2}{\|Y - P_{[\mathbf{1}]} Y\|^2}$.
   On a $\|Y - X\widehat{\theta}\|^2 = \|P_{[X]^\perp}\varepsilon\|^2$. Or $\mathbb{E}\big(\|Y - X\widehat{\theta}\|^2\big) = \mathbb{E}(\mathrm{Trace}({}^t\varepsilon P_{[X]^\perp}\varepsilon)) = \mathrm{Trace}(\mathbb{E}(\varepsilon^t\varepsilon P_{[X]^\perp}))$ par les propriétés de la trace. En utilisant l'indépendance entre ε et Z (donc X), on a $\mathbb{E}(\varepsilon^t\varepsilon P_{[X]^\perp}) = \mathbb{E}(\varepsilon^t\varepsilon)\mathbb{E}(P_{[X]^\perp}) = \sigma_\varepsilon^2 \mathbb{E}(P_{[X]^\perp})$. Grâce au fait que $\mathrm{Trace}(P_{[X]^\perp}) = n-2$, on a $\mathbb{E}\big(\|Y - X\widehat{\theta}\|^2\big) = (n-2)\sigma_\varepsilon^2$. Si $V_n = \|\varepsilon\|^2 - \|P_{[X]^\perp}\varepsilon\|^2$, on sait que $V_n \geq 0$ d'après les propriétés de la projection orthogonale, d'où d'après l'Inégalité de Markov, $\mathbb{P}\big(\frac{1}{n} V_n \geq 0\big) \leq \frac{1}{n\varepsilon}\, (n\sigma_\varepsilon^2 - (n-2)\sigma_\varepsilon^2) \xrightarrow[n\to\infty]{} 0$. On en déduit que $\frac{1}{n-1}\, \|Y - X\widehat{\theta}\|^2 \xrightarrow[n\to+\infty]{\mathcal{P}} \sigma_\varepsilon^2$.
   En écrivant que $R^2 = 1 - \frac{\frac{1}{n-1}\, \|Y - X\widehat{\theta}\|^2}{\frac{1}{n-1}\, \|Y - P_{[\mathbf{1}]} Y\|^2}$, on montre ainsi par le Lemme de Slutsky que $R^2 \xrightarrow[n\to+\infty]{\mathcal{P}} 1 - \frac{\sigma_\varepsilon^2}{\theta_1^2\, \sigma_Z^2 + \sigma_\varepsilon^2} = \frac{\theta_1^2\, \sigma_Z^2}{\theta_1^2\, \sigma_Z^2 + \sigma_\varepsilon^2}$.
   Il est clair que la limite du coefficient R^2 dépend du rapport bruit et signal, soit $\sigma_\varepsilon/\sigma_Z$. Plus celui-ci tend est grand, plus le R^2 tend vers 0, plus celui-ci est roche de 0, plus le R^2 tend vers 1. Pourtant, dans les 2 cas, le modèle linéaire est le vrai modèle! Il n'est donc pas possible d'utiliser le R^2 pour valider le modèle.

6. Pour le modèle avec Z, on a $BIC(Z) = n\log(\|P_{[X]^\perp} Y\|^2) + 2\log n$ et pour le modèle sans Z et uniquement $\mathbf{1}$, on a $BIC(\mathbf{1}) = n\log(\|P_{[\mathbf{1}]^\perp} Y\|^2) + \log n$. Par suite,
   $$\mathbb{P}\big(BIC(Z) \leq BIC(\mathbf{1})\big) = \mathbb{P}\Big(\log(\|P_{[X]^\perp} Y\|^2) - \log(\|P_{[\mathbf{1}]^\perp} Y\|^2) \leq -\frac{\log n}{n}\Big) = \mathbb{P}\big(R^2 \geq 1 - \exp(-\log n/n)\big) \geq \mathbb{P}\big(R^2 \geq \log n/n\big),$$
   car pour $x \geq 0$, $1 - e^{-x} \leq x$. Comme $\log n/n \xrightarrow[n\to\infty]{} 0$ et d'après la question précédente, si $\theta_1 \neq 0$, on a $\mathbb{P}\big(BIC(Z) \leq BIC(\mathbf{1})\big) = \mathbb{P}(\text{Choisir le modèle avec } Z) \xrightarrow[n\to\infty]{} 1$.

$\square$