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 $\sqrt{n}\big(\widehat{\theta}-\theta\big)$ 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)**.