Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2013 − 2014

Econométrie II

Contrôle continu n°1, février 2014

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

1. **(Sur 23 points)** Soit $(Y_i)_{1\leq i\leq n}$ une famille de variables aléatoires définie par:

$$Y_i = \theta_0 + \sum_{k=1}^{p} \theta_k \, Z_i^{(k)} + \varepsilon_i \qquad \text{pour tout } i \in \{1,\dots,n\}, \qquad \text{où:} \tag{1}$$

$\theta = {}^t(\theta_0,\theta_1,\dots,\theta_p)$ est un vecteur composé de $p+1$ réels inconnus, pour $1\leq j\leq p$, les $(Z_i^{(j)})_{1\leq i\leq n}$ sont p familles de réels connues et la matrice $X=(Z_i^{(j)})_{1\leq i\leq n,\ 0\leq j\leq p}$, avec par définition $Z_i^{(0)}=1$ pour tout $i=1,\cdots,n$. On suppose que X est de rang $p+1\leq n$. La suite $(\varepsilon_i)_i$ est une suite de v.a.i.i.d. de loi centrée et de variance $\sigma^2>0$. On note $H=(H_{ij})_{ij}=X\left({}^tX\,X\right)^{-1}{}^tX$.

   (a) On note $Y=(Y_i)_{1\leq i\leq n}$ et $\varepsilon=(\varepsilon_i)_{1\leq i\leq n}$. Ecrire le modèle (1) sous une forme vectorielle.

   (b) Rappeler l'expression de l'estimateur $\widehat{\theta}$ de θ par moindres carrés ordinaires en fonction de X et Y.

   (c) On note $\widehat{Y}=X\,\widehat{\theta}$, $\widehat{\varepsilon}=(\widehat{\varepsilon}_i)_{1\leq i\leq n}=Y-\widehat{Y}$ et $\widehat{\sigma}^2=\frac{1}{n-p-1}\sum_{i=1}^n\widehat{\varepsilon}_i^2$. Ecrire $\widehat{\varepsilon}$ en fonction de H et de ε. En déduire $\mathbb{E}(\widehat{\varepsilon})$ et $\mathrm{cov}(\widehat{\varepsilon})$ en fonction de H et de σ^2.

   (d) Soit le vecteur colonne J_i tel que $J_i={}^t(0,\cdots,0,1,0,\cdots,0)$, le 1 se situant en $i$ème position. Calculer ${}^tJ_i\widehat{\varepsilon}$. Ecrire $\mathrm{var}(\widehat{\varepsilon}_i)$ en fonction de J_i, H et σ^2. En déduire que $\mathrm{var}(\widehat{\varepsilon}_i)=(1-H_{ii})\sigma^2$ (on pourra utiliser la trace).

   (e) Lorsque ε est de loi gaussienne, quelle est la loi de $\widehat{W}_i=\frac{\widehat{\varepsilon}_i^2}{(1-H_{ii})\widehat{\sigma}^2}$ (justifier...)? Si $\max_{1\leq i\leq n}|H_{ii}| \underset{n\to\infty}{\longrightarrow} 0$ et si ε est de loi quelconque, quelle est la loi limite de $\widehat{W}_i$?

   (f) Si $p=1$ et $Z_i^{(1)}=\alpha^i$ pour $1\leq i\leq n$ avec $\alpha\in\mathbf{R}$, calculer $\widehat{\theta}$ et H. A-t-on $\max_{1\leq i\leq n}|H_{ii}| \underset{n\to\infty}{\longrightarrow} 0$?

2. **(Sur 15 points)** On note $\widehat{\theta}^{(-i)}$ l'estimateur de θ lorsque l'on n'utilise pas Y_i et on note $\widehat{Y}^{(-i)}=(\widehat{Y}_j^{(-i)})_{1\leq j\leq n}=X\,\widehat{\theta}^{(-i)}$ avec $X^{(-i)}$ la matrice X dans laquelle on a remplacé la $i$ème ligne par des 0. On note également la matrice $X^{(i)}=X-X^{(-i)}$, et $[X]$ (respectivement, $[X^{(-i)}]$ et $[X^{(i)}]$) le sev de $\mathbf{R}^n$ engendré par X, soit $\{X\theta,\ \theta\in\mathbf{R}^{p+1}\}$ (respectivement engendré par $X^{(-i)}$ et par $X^{(i)}$). Dans $\mathbf{R}^n$, on considère le produit scalaire $<\cdot,\cdot>$ tel que pour $U,V\in\mathbf{R}^n$, $<U,V>={}^tUV$ et $\|\cdot\|$ sa norme associée. On appelle distance de Cook pour la $i$ème observation, la statistique

$$\widehat{D}_i=\frac{\sum_{j=1}^n(\widehat{Y}_j-\widehat{Y}_j^{(-i)})^2}{(p+1)\widehat{\sigma}^2}.$$

   (a) Exprimer $\widehat{\theta}^{(-i)}$ en fonction de $X^{(-i)}$ et Y. Montrer que $[X^{(i)}]$ et $[X^{(-i)}]$ sont orthogonaux.

   (b) Montrer que $((p+1)\widehat{\sigma}^2)\widehat{D}_i={}^t(\widehat{\theta}-\widehat{\theta}^{(-i)}){}^tXX(\widehat{\theta}-\widehat{\theta}^{(-i)})$.

   (c) Soit A une matrice carrée de taille $p+1$, inversible et soit U et V deux vecteurs de $\mathbf{R}^{p+1}$. On suppose que $A+U{}^tV$ est inversible. Montrer que ${}^tUV$ et ${}^tUA^{-1}V$ sont des réels. Vérifier que

$$(A+U{}^tV)^{-1}=A^{-1}-\frac{A^{-1}U{}^tVA^{-1}}{1+{}^tUA^{-1}V}.$$

   (d) En supposant que $1-{}^tX_i({}^tXX)^{-1}X_i>0$, en déduire que $({}^tX^{(-i)}X^{(-i)})^{-1}=({}^tXX)^{-1}+\dfrac{({}^tXX)^{-1}X_i{}^tX_i({}^tXX)^{-1}}{1-{}^tX_i({}^tXX)^{-1}X_i}$, avec X_i le vecteur de $\mathbf{R}^{p+1}$ constitué de la $i$ème ligne de X. En déduire que $({}^tX^{(-i)}X^{(-i)})^{-1}X_i=\dfrac{({}^tXX)^{-1}X_i}{1-H_{ii}}$.

   (e) Montrer que ${}^tXX^{(i)}={}^tX^{(i)}X^{(i)}$. En déduire que $\left({}^tX^{(-i)}X^{(-i)}\right)(\widehat{\theta}-\widehat{\theta}^{(-i)})={}^tX^{(i)}(Y-X^{(i)}\widehat{\theta})$.

   (f) Déduire de tout ceci que

$$\widehat{D}_i=\frac{H_{ii}}{(p+1)(1-H_{ii})}\,\frac{\widehat{\varepsilon}_i^2}{(1-H_{ii})\widehat{\sigma}^2}.$$

   Lorsque les (ε_i) sont gaussiens, quelle est la loi de $\widehat{D}_i$?