Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2020 − 2021

Econométrie II

Contrôle continu n°1, mars 2021

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

On suppose que pour n et p deux entiers tels que $n \geq p+2$, on observe $Y = {}^t(Y_i)_{1\leq i\leq n}$ défini par:

$$Y_i = \theta_0 + \theta_1 X_i^{(1)} + \cdots + \theta_p X_i^{(p)} + \varepsilon_i \qquad \text{pour tout } i = 1, \cdots, n, \tag{1}$$

avec une famille connue de réels $(X_i^{(j)})_{1\leq i\leq n,\, 1\leq j\leq p}$, et telle que $X = \begin{pmatrix} 1 & X_1^{(1)} & \cdots & X_1^{(p)} \\ \vdots & \vdots & & \vdots \\ 1 & X_n^{(1)} & \cdots & X_n^{(p)} \end{pmatrix}$ soit une matrice de rang $p+1$, $\theta = {}^t(\theta_j)_{0\leq j\leq p}$ un vecteur de nombres réels inconnus et $\varepsilon = {}^t(\varepsilon_i)_{1\leq i\leq n}$ un vecteur d'erreur de loi $\mathcal{N}(0\,,\,\sigma^2 I_n)$ non observé, où $\sigma^2 > 0$ est inconnu et I_n est la matrice identité de taille n.
Pour M une matrice de taille $(n, p+1)$, on note $[M] = \{M\,\alpha,\ \alpha \in \mathbf{R}^{p+1}\}$ et P_A la matrice de projection orthogonale sur un sous-espace A de $\mathbf{R}^n$.

1. Pour A et B deux sous-espaces vectoriels de $\mathbf{R}^n$, montrer que $P_A\, P_B = P_B\, P_A = P_{A\cap B}$ **(1.5pts)**.

2. Ecrire le modèle sous une forme matricielle, et rappeler l'expression du vecteur $\widehat{Y} = (\widehat{Y}_i)_{1\leq i\leq n} = X\,\widehat{\theta}$ obtenu par moindres carrés. Démontrer que la loi de $\widehat{Y}$ est $\mathcal{N}_n(X\,\theta\,,\,\sigma^2\, X({}^tX\,X)^{-1}\,{}^tX)$ **(2.5pts)**.

3. Démontrer que $\frac{1}{n}\sum_{i=1}^n \widehat{Y}_i = \frac{1}{n}\sum_{i=1}^n Y_i$ **(1.5pts)**.

4. Rappeler en la justifiant la loi de $\widehat{\sigma}^2$ estimateur non biaisé de σ^2 **(1pt)**.

5. On veut tester si l'observation Y_n n'est pas une donnée aberrante, alors que les autres Y_i ne le sont pas. On considère ainsi le problème de test H_0: ε_n suit la loi $\mathcal{N}(0\,,\,\sigma^2)$, contre H_1: ε_n ne suit pas la loi $\mathcal{N}(0\,,\,\sigma^2)$. Soit $X_{(n)}$ la matrice X où l'on remplace la dernière ligne par des 0 et $E_n = \mathrm{Vect}({}^t(0, \cdots, 0, 1))$.

   (a) Montrer que $E_n \subset [X_{(n)}]^\perp$ **(1pt)**.

   (b) Montrer que $\widehat{\sigma}^2_{(n)}$ l'estimateur non biaisé de σ^2 à partir de l'échantillon privé de l'individu n s'écrit $\frac{1}{n-p-2}\,\|P_{E_n^\perp}\, P_{[X_{(n)}]^\perp}\varepsilon\|^2$ **(3pts)**.

   (c) Montrer que $\widehat{\varepsilon}_n = Y_n - \widehat{Y}_n = (0, \cdots, 0, 1)\, P_{E_n}\, P_{[X]^\perp}\varepsilon$ **(2pts)**.

   (d) En déduire que $\varepsilon'_n = \dfrac{Y_n - \widehat{Y}_n}{\left(\widehat{\sigma}^2_{(n)}\,(1 - p_{nn})\right)^{1/2}}$ suit une loi de student à $n-p-2$ degrés de liberté, où $P_{[X]} = (p_{ij})_{1\leq i,j\leq n}$. Comment utiliser ε'_n pour accepter ou non H_0? **(2pts)**.

6. On suppose que l'on observe $X_{n+1} = (1, X_{n+1}^{(1)}, \dots, X_{n+1}^{(p)})$, vecteur de réels connus, et l'on veut prédire Y_{n+1}.

   (a) Montrer qu'une prédiction de Y_{n+1} est $\widetilde{Y}_{n+1} = X_{n+1}({}^t X\, X)^{-1\,t} X\, Y$ et montrer que $\mathbb{E}(\widetilde{Y}_{n+1}) = \mathbb{E}(Y_{n+1})$ **(2pts)**.

   (b) Déterminer la loi de $\widetilde{Y}_{n+1}$, puis montrer que $\dfrac{\widetilde{Y}_{n+1} - \mathbb{E}(Y_{n+1})}{\left(\widehat{\sigma}^2\, X_{n+1}({}^t X\, X)^{-1\,t} X_{n+1}\right)^{1/2}}$ suit une loi que l'on précisera **(2pts)**.

   (c) En déduire un intervalle de prédiction à 95% pour Y_{n+1}. Sans l'hypothèse gaussienne sur ε, quel intervalle aurait-on pu obtenir et sous quelles hypothèses? **(2pts)**

7. On se place dans le simple cadre où $p = 1$, et sans l'hypothèse de gaussianité de ε.

   (a) Ecrire $\widehat{\theta}$ en fonction des $x_i = X_i^{(1)}$ et des Y_i (on utilisera les notations $\overline{x}_n = \frac{1}{n}(x_1 + \cdots + x_n)$ et $\overline{Y}_n = \frac{1}{n}(Y_1 + \cdots + Y_n)$) **(1pt)**.

   (b) Donner une hypothèse suffisante pour que $\widehat{\theta}$ satisfasse un théorème de la limite centrale que l'on précisera **(1.5pts)**.

   (c) Donner alors explicitement la statistique du test de Fisher global du modèle ainsi que sa limite **(1pt)**.

   (d) Donner ensuite l'intervalle asymptotique de prédiction à 95% pour Y_{n+1} en fonction de x_{n+1}. Que se passe-t-il quand $x_{n+1} \to \infty$? **(1.5pts)**.