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 $\mathcal{N}_n(X\,\theta\,,\,\sigma^2\, X({}^tX\,X)^{-1\,t}X)$ **(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 $\varepsilon_n' = \dfrac{Y_n - \widehat{Y}_n}{\big(\widehat{\sigma}^2_{(n)}\,(1-p_{nn})\big)^{1/2}}$ 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 $X_{n+1} = (1, X_{n+1}^{(1)}, \ldots, X_{n+1}^{(p)})$, vecteur de réels connus, et l'on veut prédire Y_{n+1}.
   (a) Montrer qu'une prédiction de Y_{n+1} est $\widetilde{Y}_{n+1} = X_{n+1}({}^tX\,X)^{-1}\,{}^tX\,Y$ 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 $\dfrac{\widetilde{Y}_{n+1} - \mathbb{E}(Y_{n+1})}{\left(\widehat{\sigma}^2\, X_{n+1}({}^tX\,X)^{-1}\,{}^tX_{n+1}\right)^{1/2}}$ 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)**.

Proof. 1. Pour $U \in A \cap B$, on a $P_A U = U$ puisque $U \in A \cap B$ donc $U \in A$, $P_B U = U$ et également $P_{A\cap B} U = P_A P_B U = P_B P_A U$. Pour $V \in (A \cap B)^\perp$, on a $P_{A\cap B} V = 0$ par définition. Mais $(A \cap B)^\perp = A^\perp + B^\perp$, donc $V = P_{A^\perp} V + P_{B^\perp} V$. Mais $P_A(P_{A^\perp} V + P_{B^\perp} V) = P_A P_{B^\perp} V$. D'où $P_B P_A V = P_B = P_A P_{B^\perp} V = 0$. Et la même chose pour $P_A P_B V = 0$. D'où $P_{A\cap B} = P_A P_B = P_B P_A$.

2. On a $Y = Z\theta + \varepsilon$ et $\widehat{Y} = P_{[X]} Y = X({}^tX\,X)^{-1}\,{}^tX\,Y$.
   On a $Y = X\theta + \varepsilon$, d'où $\widehat{Y} = P_{[X]} X\theta + P_{[X]}\varepsilon = X\theta + P_{[X]}\varepsilon$ car $X\theta \in [X]$. On sait que ε est un vecteur gaussien, comme $P_{[X]}$ est une matrice de réels, alors $P_{[X]}\varepsilon$ est aussi un vecteur gaussien. Et comme $X\theta$ est un vecteur de réels, alors $\widehat{Y}$ est bien un vecteur gaussien. Il est définit par son espérance et sa matrice de variance-covariance. $\mathbb{E}[\widehat{Y}] = X\theta + \mathbb{E}[P_{[X]}\varepsilon] = X\theta$ car $\mathbb{E}[\varepsilon] = 0$. Et $\mathrm{cov}(\widehat{Y}) = \mathrm{cov}(P_{[X]}\varepsilon) = P_{[X]}\mathrm{cov}(\varepsilon)\,{}^tP_{[X]}$. On sait que $\mathrm{cov}(\varepsilon) = \sigma^2 I_n$, donc $\mathrm{cov}(\widehat{Y}) = \sigma^2\, P_{[X]}\,{}^tP_{[X]} = \sigma^2\, P_{[X]}$ car ${}^tP_{[X]} = P_{[X]}$ pour une projection orthogonale et $P_{[X]} P_{[X]} = P_{[X]}$ car projection. D'où le résultat.

3. Soit le vecteur $\mathbb{I} = {}^t(1, \ldots, 1)$. Alors $\frac{1}{n}\sum_{i=1}^n \widehat{Y}_i = \frac{1}{n}\,{}^t\mathbb{I}\,\widehat{Y} = \frac{1}{n}\,{}^t\mathbb{I}\,P_{[X]} Y = \frac{1}{n}\,{}^t(P_{[X]}\mathbb{I})\,Y = \frac{1}{n}\,{}^t\mathbb{I}\,Y = \overline{Y}$ car $\mathbb{I} \in [X]$ donc $P_{[X]}\mathbb{I} = \mathbb{I}$.

4. On a $\widehat{\sigma}^2 = \frac{1}{n-(p+1)}\left\|Y - \widehat{Y}\right\|^2 = \frac{1}{n-(p+1)}\left\|P_{[X]^\perp} Y\right\|^2$. D'après Cochran, $\left\|P_{[X]^\perp} Y\right\|^2 \stackrel{\mathcal{L}}{\sim} \sigma^2\,\chi^2(n - (p+1))$, d'où $\widehat{\sigma}^2 \stackrel{\mathcal{L}}{\sim} \frac{\sigma^2}{n-(p+1)}\,\chi^2(n - (p+1))$.

5. (a) $[X_{(n)}]$ est généré par $p + 1$ vecteurs de $\mathbf{R}^n$ ayant tous 0 pour dernière coordonnée. Donc le produit scalaire euclidien classique de ces vecteurs et de E_n est nul: E_n est orthogonal à chacun de ces vecteurs, donc $E_n \subset [X_{(n)}]^\perp$.
   (b) Si l'on effectue une régression par moindres carrés par rapport au $(n - 1)$ premières données, on obtient un vecteur de prédiction $\widehat{Y}^{(n)} = P_{[X_{(n)}]} Y$ et un vecteur de résidus $\widehat{\varepsilon}^{(n)} = Y - P_{[X_{(n)}]} Y = P_{[X_{(n)}]^\perp}\varepsilon$ mais il faut enlever la n-ième composante qui vaut Y_n. Aussi peut-on appliquer une projection sur $E_n^\perp$ puisque $P_{[E_n]^\perp}\widehat{\varepsilon}^{(n)} = \widehat{\varepsilon}^{(n)} - P_{[E_n]}\widehat{\varepsilon}^{(n)}$ et $P_{[E_n]}\widehat{\varepsilon}^{(n)} = {}^t(0, \ldots, 0, Y_n)$. Donc d'appliquer $P_{[E_n]^\perp}$ à $\widehat{\varepsilon}^{(n)}$ transforme sa n-ième composante en 0, et les $n - 1$ autres composantes sont inchangées. Or l'estimateur $\widehat{\sigma}^2_{(n)}$ vaut $\widehat{\sigma}^2_{(n)} = \frac{1}{n-1-(p+1)}\sum_{i=1}^{n-1}\left(\widehat{\varepsilon}_i^{(n)}\right)^2$. Et cette somme est exactement égale à $\frac{1}{n-p-2}\,\|P_{E_n^\perp}\,P_{[X_{(n)}]^\perp}\varepsilon\|^2$.

(c) Il est clair que $\widehat{\varepsilon} = P_{[X]^\perp}\varepsilon$. Donc $P_{E_n}\widehat{\varepsilon} = P_{[X]^\perp}\varepsilon = {}^t(0,\dots,0,\widehat{\varepsilon}_n)$, puisque projeter sur E_n c'est annuler les $(n-1)$ premières composantes d'un vecteur de $\mathbf{R}^n$. Donc $(0,\dots,0,1)P_{E_n}\,P_{[X]^\perp}\varepsilon$ représente le produit scalaire entre ${}^t(0,\dots,0,\widehat{\varepsilon}_n)$ et ${}^t(0,\dots,0,1)$, soit $\widehat{\varepsilon}_n$.

(d) On a $\widehat{\varepsilon}_n \overset{\mathcal{L}}{\sim} \mathcal{N}(0\,,\sigma^2(1-p_{nn}))$ car $\widehat{\varepsilon} \overset{\mathcal{L}}{\sim} \mathcal{N}_n(0\,,\sigma^2(I_n - P_{[X]}))$. Donc $\frac{\widehat{\varepsilon}_n}{(\sigma^2(1-p_{nn}))^{1/2}} \overset{\mathcal{L}}{\sim} \mathcal{N}(0,1)$. Or $\varepsilon' = \frac{\widehat{\varepsilon}_n}{(\sigma^2(1-p_{nn}))^{1/2}}\Big(\frac{\widehat{\sigma}^2_{(n)}}{\sigma^2}\Big)^{-1/2}$. De plus, grâce au résultat vu précédemment mais appliqué au $(n-1)$ premières données, $\widehat{\sigma}^2_{(n)} \overset{\mathcal{L}}{\sim} \frac{\sigma^2}{n-p-2}\,\chi^2(n-p-2)$. Ainsi, il ne nous reste qu'à montrer que $\widehat{\varepsilon}_n$ est indépendant de $\widehat{\sigma}^2_{(n)}$ pour obtenir le fait que $\varepsilon'_n \overset{\mathcal{L}}{\sim} t(n-p-2)$. Mais $\widehat{\varepsilon}_n = (0,\dots,0,1)P_{E_n}\,P_{[X]^\perp}\varepsilon$ et $\widehat{\sigma}^2_{(n)} = \frac{1}{n-p-2}\,\|P_{E_n^\perp}\,P_{[X_{(n)}]^\perp}\varepsilon\|^2$. Or les sous-espaces vectoriels de $\mathbf{R}^n$, $E_n \cap [X]^\perp$ (numérateur) et $E_n^\perp \cap [X_{(n)}]^\perp$ (dénominateur) sont orthogonaux, donc d'après Cochran, les vecteurs aléatoires dans $\mathbf{R}^n$ constitués par les projections orthogonales sur ses sev sont indépendantes. D'où le résultat.
Pour accepter H_0 avec un risque α, il suffira de vérifier que ε'_n est dans un intervalle $[q_{\alpha/2}, q_{1-\alpha/2}]$ où q_β désigne le quantile β de la loi $t(n-p-2)$.

6. (a) On a $\widehat{\theta} = ({}^tX\,X)^{-1\,t}XY$. On prédira donc Y_{n+1} à l'aide de $\widehat{\theta}$ et X_{n+1} par la formule $X_{n+1}\widehat{\theta}$ soit $\widetilde{Y}_{n+1}$. $\mathbb{E}\big[\widetilde{Y}_{n+1}\big] = \mathbb{E}\big[X_{n+1}\widehat{\theta}\big] = \mathbb{E}\big[X_{n+1}\theta\big] + \mathbb{E}\big[X_{n+1}{}^tX\,X)^{-1\,t}X\varepsilon\big] = X_{n+1}\theta + 0$ car $\mathbb{E}[\varepsilon] = 0$. Et $\mathbb{E}[Y_{n+1}] = \mathbb{E}[X_{n+1}\theta + \varepsilon] = X_{n+1}\theta + 0$. D'où le résultat.

(b) $\widetilde{Y}_{n+1}$ est gaussien car Y est un vecteur gaussien et $\widetilde{Y}_{n+1} = X_{n+1}({}^tX\,X)^{-1\,t}XY$. D'où

$$\widetilde{Y}_{n+1} \overset{\mathcal{L}}{\sim} \mathcal{N}\Big(X_{n+1}\theta\,,\,X_{n+1}({}^tX\,X)^{-1\,t}X\mathrm{cov}(\varepsilon)^t\big(X_{n+1}({}^tX\,X)^{-1\,t}X\big)\Big) = \mathcal{N}\Big(X_{n+1}\theta\,,\,\sigma^2\,X_{n+1}({}^tX\,X)^{-1\,t}X_{n+1})\Big).$$

On en déduit que $\frac{\widetilde{Y}_{n+1}-\mathbb{E}(Y_{n+1})}{\big(\sigma^2\,X_{n+1}({}^tX\,X)^{-1\,t}X_{n+1}\big)^{1/2}} \overset{\mathcal{L}}{\sim} \mathcal{N}(0,1)$. On a vu que $\widehat{\sigma}^2 \overset{\mathcal{L}}{\sim} \frac{\sigma^2}{n-(p+1)}\,\chi^2(n-(p+1))$. Il reste à montrer que $\widehat{\sigma}^2$ et $\widetilde{Y}_{n+1}$ sont indépendants, ce qui est le cas car on sait d'après le cours que $\widehat{\sigma}^2$ et $\widehat{\theta}$ sont indépendantes (Cochran). Ainsi

$$\frac{\widetilde{Y}_{n+1}-\mathbb{E}(Y_{n+1})}{\big(\widehat{\sigma}^2\,X_{n+1}({}^tX\,X)^{-1\,t}X_{n+1}\big)^{1/2}} \overset{\mathcal{L}}{\sim} t(n-(p+1)).$$

(c) On en déduit qu'un intervalle de prédiction à 95% pour Y_{n+1} est

$$\big[\widetilde{Y}_{n+1} - q_{0.975}\big(\widehat{\sigma}^2\,X_{n+1}({}^tX\,X)^{-1\,t}X_{n+1}\big)^{1/2}, \widetilde{Y}_{n+1} + q_{0.975}\big(\widehat{\sigma}^2\,X_{n+1}({}^tX\,X)^{-1\,t}X_{n+1}\big)^{1/2}\big],$$

avec $q_{0.975}$ le quantile à 97.5% de la loi $t(n-p-1)$.
Si on rajoute l'hypothèse $\max_{1\le i\le n}\big\{|p_{ii}|\big\} \underset{n\to\infty}{\longrightarrow} 0$, alors $\widehat{\sigma}^2 \underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}} \sigma^2$ et $\widehat{\theta}$ donc $\widetilde{Y}_{n+1}$ est asymptotiquement gaussien. Et on aura

$$\frac{\widetilde{Y}_{n+1}-\mathbb{E}(Y_{n+1})}{\big(\widehat{\sigma}^2\,X_{n+1}({}^tX\,X)^{-1\,t}X_{n+1}\big)^{1/2}} \underset{n\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}(0,1).$$

7. (a) On a $\widehat{\theta} = {}^t\Big(\overline{y} - \dfrac{\overline{\sigma}_{xy}}{\overline{\sigma}_x^2}\,\overline{x}\,,\,\dfrac{\overline{\sigma}_{xy}}{\overline{\sigma}_x^2}\Big)$.

(b) On calcule $P_{[X]}$ dans ce cas, qui vaut $X({}^tX\,X)^{-1\,t}X$. Or ici $({}^tX\,X)^{-1} = \frac{1}{n\,\overline{\sigma}_x^2}\begin{pmatrix}\overline{x^2} & -\overline{x}\\ -\overline{x} & 1\end{pmatrix}$ et après calculs on obtient pour terme général $\Big(\frac{1}{n\,\overline{\sigma}_x^2}\big(\overline{x^2}-(x_i+x_j)\overline{x}+x_ix_j\big)\Big)_{1\le i,j\le n}$. La condition pour obtenir la normalité est donc

$$\max_{1\le i\le n}\Big\{\frac{1}{n\,\overline{\sigma}_x^2}\,\big|\overline{x^2} - 2\,x_i\,\overline{x} + x_i^2\big|\Big\} = \frac{1}{n} + \max_{1\le i\le n}\Big\{\frac{(\overline{x}-x_i)^2}{n\,\overline{\sigma}_x^2}\Big\} \underset{n\to\infty}{\longrightarrow} 0.$$

(c) La statistique de Fisher est obtenue pour $C = {}^t(0,1)$ et ainsi $\widehat{F} = \dfrac{1}{\widehat{\sigma}^2}\,{}^t\widehat{\theta}_1\,\big(C\,({}^tX\,X)^{-1\,t}C\big)^{-1}\widehat{\theta}_1$. Après simplifications, on obtient:

$$\widehat{F} = \frac{n\,\overline{\sigma}_x^2}{\widehat{\sigma}^2}\Big(\frac{\overline{\sigma}_{xy}}{\overline{\sigma}_x^2}\Big)^2 = n\,\frac{(\overline{\sigma}_{xy})^2}{\widehat{\sigma}^2\,\overline{\sigma}_x^2} \underset{n\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \chi^2(1).$$

□