Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2024 – 2025

Econométrie II

Contrôle continu n°1, mars 2025

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

On suppose que l'on a observé $Y_1, \dots, Y_n$, n réels avec $n \in \mathbf{N}^*$. On suppose qu'il existe des réels observés $(x_i)_{1 \leq i \leq n}$, avec $0 < x_1 < x_2 < \cdots < x_n$, ainsi qu'une fonction $g : \mathbf{R} \to \mathbf{R}$, inconnue, tels que

$$Y_i = g(x_i) + \varepsilon_i \quad \text{pour tout } i \in \{1, \dots, n\},$$

avec (ε_i) des variables gaussiennes indépendantes de même loi $\mathcal{N}(0\,,\,\sigma^2)$, où $\sigma^2 > 0$ est inconnue. On notera par la suite pour $U, V \in \mathbf{R}^n$, le produit scalaire $< U, V >= {}^tUV$ et la norme $\|U\|^2 =< U, U >$.

1. Un peu d'algèbre pour commencer... On note X_j le vecteur colonne $(x_i^j)_{1 \leq i \leq n}$. Pour $1 \leq p < n$, montrer que $(X_0, X_1, \dots, X_p)$ forme une famille de rang $p+1$ **(2pts)**. En déduire qu'il existe des polynômes Q_j de degré j pour $j = 0, 1, \dots, p$, tels que la famille $(X'_0, X'_1, \dots, X'_p)$ soit une famille orthonormale, en notant X'_j le vecteur colonne $(Q_j(x_i))_{1 \leq i \leq n}$ **(2pts)**. Donner explicitement Q_0 et Q_1 **(1pt)**.

2. On définit
$$\big(\widehat{a}_0, \widehat{a}_1, \widehat{a}_2, \widehat{a}_3\big) = \mathop{\arg\min}_{(a_0,a_1,a_2,a_3) \in \mathbf{R}^4} \sum_{i=1}^n \big(Y_i - a_0 - a_1 x_i - a_2 x_i^2 - a_3 x_i^3\big)^2.$$
Démontrer que pour $n \geq 4$, il existe une matrice M de taille $(4, n)$ (justifier son existence!) dépendant des x_i telle que ${}^t\big(\widehat{a}_0, \widehat{a}_1, \widehat{a}_2, \widehat{a}_3\big) = M\,Y$ où $Y = {}^t\big(Y_1, \dots, Y_n\big)$ **(2pts)**.

3. Pour $n \geq 4$, on définit pour $x \in \mathbf{R}$, $\widehat{P}(x) = \widehat{a}_0 + \widehat{a}_1\,x + \widehat{a}_2\,x^2 + \widehat{a}_3\,x^3$. Pour $x \in \mathbf{R}$ fixé, déterminer la loi de $\widehat{P}(x)$ **(2pts)**.

4. On suppose uniquement dans cette question que $n = 4$. Démontrer que:
$$\widehat{P}(x) = \sum_{i=1}^4 Y_i \, \frac{\prod_{j=1, j \neq i}^4 (x - x_j)}{\prod_{j=1, j \neq i}^4 (x_i - x_j)} \quad \textbf{(2pts)}.$$
En déduire alors $\widehat{a}_3$ **(1pt)**.

5. On suppose ici que g est polynôme de degré 2. Proposer un estimateur $\widehat{\sigma}_n^2$ de σ^2 et donner sa loi **(1pt)**. Proposer une statistique de test (avec sa loi) pour tester si g est bien un polynôme de degré 2 contre l'hypothèse que g est un polynôme de degré 3 **(2pts)**.

6. En reprenant la question 1., montrer que $\big(\widehat{P}(x_i)\big)_{1 \leq i \leq n} = \sum_{j=0}^3 < Y, X'_j > X'_j$ **(2pts)**. Montrer que pour effectuer le même test qu'à la question 5., on peut utiliser la statistique de test suivante dont on précisera la loi:
$$\widehat{T} = \frac{< Y, X'_3 >^2}{\frac{1}{n-4} \big\| Y - \sum_{j=0}^3 < Y, X'_j > X'_j \big\|^2} \quad \textbf{(3pts)}.$$

7. On suppose ici que $g(x) = x^4$. En reprenant la question 1., montrer que:
$$X_4 = \sum_{j=0}^4 < X_4, X'_j > \, X'_j \quad \textbf{(1pt)}.$$
En déduire que $\sum_{i=1}^n \big(Y_i - \widehat{P}(x_i)\big)^2 = \big\| < X_4, X'_4 > X'_4 + \varepsilon - \sum_{j=0}^3 < \varepsilon, X'_j > X'_j \, \big\|^2$ **(2pts)** et en déduire que $\mathbb{E}\big[\widehat{\sigma}_n^2\big] = \sigma^2 + \frac{1}{n-4} \, < X_4, X'_4 >^2$ **(2pts)**.

8. Pour pouvoir mieux approcher la fonction g quand celle-ci n'est pas un polynôme de $\mathbf{R}^3[X]$ (ensemble des polynômes de degré inférieur ou égal à 3), on découpe l'ensemble des $(x_1,\dots,x_n)$ en m sous-ensembles et on détermine des polynômes de degré 3 sur chaque ensemble. Dans la suite nous considérons seulement le cas où $m=2$, pour lequel on supposera que $n=2q$ avec $q\geq 4$. Ainsi, on définit:

$$\big(\widehat P_1,\widehat P_2\big)=\mathop{\arg\min}_{P_1,P_2\in\mathbf{R}^3[X]}\Big\{\sum_{i=1}^{q}\big(Y_i-P_1(x_i)\big)^2+\sum_{i=q+1}^{2q}\big(Y_i-P_2(x_i)\big)^2\Big\}. \qquad (1)$$

Démontrer que ${}^t\big(\widehat P_1(x_1),\dots,\widehat P_1(x_q),\widehat P_2(x_{q+1}),\dots,\widehat P_2(x_{2q})\big)=Z\,\widehat\theta$ où $\widehat\theta=\arg\min_{\theta\in\mathbf{R}^8}\big\|Y-Z\,\theta\big\|^2$ et Z est une matrice que l'on précisera **(2pts)**. Montrer que le vecteur ${}^t\big(\widehat P_1(x_1),\dots,\widehat P_1(x_q)\big)$ peut s'écrire sous la forme $N\,(Y_i)_{1\leq i\leq q}$ avec N une matrice que l'on précisera **(1pt)**.

9. Lors de la question précédente, il est très probable qu'entre $\widehat P_1(x_q)$ et $\widehat P_2(x_{q+1})$ il y ait un saut conséquent et un changement de variations. Pour éviter cela, on va remplacer dans (1) la condition $P_1,P_2\in\mathbf{R}^3[X]$ par la condition $P_1,P_2\in\mathbf{R}^3[X]$, $P_1(x_q)=P_2(x_q)$ et $P_1'(x_q)=P_2'(x_q)$ (la dérivée). En écrivant $P_1(x)=a_0+a_1x+a_2x^2+a_3x^3$ et $P_2(x)=b_0+b_1x+b_2x^2+b_3x^3$, montrer que cette nouvelle condition implique le système:

$$\left\{\begin{array}{lcl} b_0 &=& a_0-(a_2-b_2)x_q^2-2(a_3-b_3)x_q^3\\ b_1 &=& a_1+2(a_2-b_2)x_q+3(a_3-b_3)x_q^2\end{array}\right. \qquad \textbf{(2pts)}.$$

Si l'on note $\widehat Q_1$ et $\widehat Q_2$ la solution de (1) sous la nouvelle condition, en déduire que

$${}^t\big(\widehat Q_1(x_1),\dots,\widehat Q_1(x_q),\widehat Q_2(x_{q+1}),\dots,\widehat Q_2(x_{2q})\big)\;=\;\widetilde Z\,\widehat\theta\quad\text{avec}\quad\widehat\theta\;=\;\mathop{\arg\min}_{\theta\in\mathbf{R}^6}\big\|Y\;-\;Z'\,\theta\big\|^2\quad\textbf{(4pts)}$$

où $\widetilde Z$ est la matrice

$$\widetilde Z=\begin{pmatrix} 1 & x_1 & x_1^2 & x_1^3 & 0 & 0\\ 1 & x_2 & x_2^2 & x_2^3 & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 1 & x_q & x_q^2 & x_q^3 & 0 & 0\\ 1 & x_{q+1} & x_q(2x_{q+1}-x_q) & x_q^2(3x_{q+1}-2x_q) & (x_{q+1}-x_q)^2 & x_{q+1}^3+2x_q^3-3x_q^2x_{q+1}\\ 1 & x_{q+2} & x_q(2x_{q+2}-x_q) & x_q^2(3x_{q+2}-2x_q) & (x_{q+2}-x_q)^2 & x_{q+2}^3+2x_q^3-3x_q^2x_{q+2}\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 1 & x_{2q} & x_q(2x_{2q}-x_q) & x_q^2(3x_{2q}-2x_q) & (x_{2q}-x_q)^2 & x_{2q}^3+2x_q^3-3x_q^2x_{2q}\end{pmatrix}.$$

Si $g\in\mathbf{R}_3[X]$, déterminer le risque quadratique $\mathbb{E}\big[\big\|G-Z'\widehat\theta\big\|^2\big]$ avec $G=(g(x_i))_{1\leq i\leq 2q}$ **(2pts)**.