Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2017 – 2018

Econométrie II

Contrôle continu n°1, avril 2018

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

Soit une fonction $f : [0,1] \to \mathbf{R}$ connue, de classe $\mathcal{C}^1$ sur $[0,1]$ et telle que $f(0) \neq f(1)$. On dispose d'une variable $Y = {}^t(Y_1, \cdots, Y_n)$ observée pour n individus, émanant du modèle:

$$Y_i = \theta_0 + \theta_1 \, f\big(\frac{i}{n}\big) + \varepsilon_i \quad \text{pour } i = 1, \ldots, n,$$

où $\theta = {}^t(\theta_0, \theta_1)$ est un vecteur de paramètres inconnus et $\varepsilon = {}^t(\varepsilon_1, \cdots, \varepsilon_n)$ est un vecteur gaussien composé de n v.a.i.i.d. centrées de variance σ_ε^2 inconnue. On note $\|\cdot\|$ la norme euclidienne classique de $\mathbf{R}^n$.

1. Montrer que $\Big|\dfrac{1}{n}\displaystyle\sum_{i=1}^n f\big(\frac{i}{n}\big) - \int_0^1 f(t)dt\Big| \leq \dfrac{1}{2n} \sup_{x\in[0,1]} |f'(x)|$ pour tout $n \in \mathbf{N}$ **(2.5 pts)**. En déduire $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n f\big(\frac{i}{n}\big)$ **(0.5 pts)**.

2. On effectue la régression linéaire par moindres carrés de Y. Donner une expression matricielle de $\widehat{\theta} = {}^t(\widehat{\theta}_0, \widehat{\theta}_1)$ **(1 pt)** en justifiant son existence pour n suffisamment grand **(2 pts)**, puis l'expression exacte de $\widehat{\theta}_0$ et de $\widehat{\theta}_1$ **(1 pt)**. Montrer le théorème de la limite centrale
$$\sqrt{n}\big(\widehat{\theta} - \theta\big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(0\,, \frac{\sigma_\varepsilon^2}{\int_0^1 f^2(x)dx - (\int_0^1 f(x)dx)^2} \begin{pmatrix} \int_0^1 f^2(x)dx & -\int_0^1 f(x)dx \\ -\int_0^1 f(x)dx & 1 \end{pmatrix}\right) \qquad \textbf{(2 pts)}.$$

3. Rappeler l'expression exacte (non matricielle) d'un estimateur $\widehat{\sigma}_\varepsilon^2$ non biaisé de σ_ε^2 **(1 pt)** dont on précisera la loi **(0.5 pts)**.

4. Donner l'expression exacte (non matricielle) de la statistique $\widehat{T}$ de Student permettant de tester $H_0 : \ \theta_1 = 0$ contre $H_1 : \ \theta_1 \neq 0$ **(1 pt)**. Lorsque $n \to \infty$ déterminer en fonction de $\widehat{\theta}_1$, $\widehat{\sigma}_\varepsilon^2$, n et f l'expression de la p-value de ce test **(1 pt)**.

5. On note ici $F = {}^t\big(f(1/n), \ldots, f(1)\big)$, $\mathbf{1} = {}^t(1, \ldots, 1)$, $[\mathbf{1}]$ le sous-espace vectoriel de $\mathbf{R}^n$ engendré par la droite vectorielle $\mathbf{1}$ et P_A la matrice de la projection orthogonale sur un sous-espace vectoriel A de $\mathbf{R}^n$. Enfin, pour $U = {}^t(U_1, \ldots, U_n) \in \mathbf{R}^n$, on note $\overline{U}_n = \frac{1}{n}(U_1 + \cdots + U_n)$ et $\overline{U} = \overline{U}_n \, \mathbf{1}$.
   (a) Montrer que $\dfrac{1}{n}\big\|Y - \overline{Y}\big\|^2 = \dfrac{\theta_1^2}{n}\|F - \overline{F}\|^2 + \dfrac{2\theta_1}{n} {}^tF \, P_{[1]^\perp}\varepsilon + \dfrac{1}{n}\|P_{[1]^\perp}\varepsilon\|^2$ **(1.5 pts)**.
   (b) Montrer que $\dfrac{1}{n}\|F - \overline{F}\|^2 \xrightarrow[n\to\infty]{} \displaystyle\int_0^1 f^2(t)dt - \Big(\int_0^1 f(t)dt\Big)^2$ **(1.5 pts)**.
   (c) Montrer que $\dfrac{1}{n}\|P_{[1]^\perp}\varepsilon\|^2 \xrightarrow[n\to+\infty]{\mathcal{P}} \sigma_\varepsilon^2$ **(1 pt)**.
   (d) Déterminer les espérance **(1 pt)** et variance **(2 pts)** de $\dfrac{2}{n} {}^tF \, P_{[1]^\perp}\varepsilon$. En déduire que $\dfrac{2}{n} {}^tF \, P_{[1]^\perp}\varepsilon \xrightarrow[n\to+\infty]{\mathcal{P}} 0$ **(1 pt)**.
   (e) Déduire de tout ce qui précède que le coefficient $\widehat{R}^2$ vérifie $\widehat{R}^2 \xrightarrow[n\to+\infty]{\mathcal{P}} 1 - \dfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \theta_1^2\big(\int_0^1 f^2(t)dt - \big(\int_0^1 f(t)dt\big)^2\big)}$ **(1.5 pts)**. Quelle est sa limite lorsque $\sigma_\varepsilon^2 \to \infty$ **(0.5 pts)**?

6. Déduire de ce qui précède que si $\theta_1 \neq 0$, alors $\mathbb{P}\big(\widehat{BIC}(m_0) \leq \widehat{BIC}(m_1)\big) \xrightarrow[n\to\infty]{} 0$, où m_0 est le modèle sans F et m_1 le modèle avec F **(3 pts)**. Qu'en déduire quant à la probabilité de choisir un faux modèle avec le critère BIC **(0.5 pts)**?

Proof. 1. On écrit $\left|\frac{1}{n}\sum_{i=1}^n f\left(\frac{i}{n}\right)-\int_0^1 f(t)dt\right| \leq \sum_{i=1}^n \int_{(i-1)/n}^{i/n}|f(i/n)-f(x)|\,dx$ et $|f(i/n)-f(x)|\leq \sup_{x\in[0,1]}|f'(x)|\times(i/n-x)$ pour $x\in[(i-1)/n,i/n]$ par inégalité des accroissements finis.
On a donc $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n f\left(\frac{i}{n}\right)=\int_0^1 f(x)dx$.

2. On a $\widehat{\theta}=({}^tX\,X)^{-1}\,{}^tX\,Y$, où X est la matrice $(n,2)$ avec des 1 sur la première colonne et les $f(i/n)$ sur la deuxième. La matrice X est de rang 2 dès que n est sufficamment grand car $f(0)\neq f(1)$ et comme f est continue, il existe n tel que $f(1/n)$ est suffisamment proche de $f(0)$ pour être également différent de $f(1)$. Ainsi quand $f(1/n)\neq f(1)$ la seconde colonne de X n'est plus colinéaire à la première.
Comme les ε_i sont gaussiennes, que $\widehat{\theta}$ est sans biais, on a $(\widehat{\theta}-\theta)\overset{\mathcal{L}}{\sim}\mathcal{N}(0\,,\,\sigma_\varepsilon^2({}^tXX)^{-1})$. Après calcul, on montre que
$$n\,({}^tXX)^{-1}\underset{n\to\infty}{\longrightarrow}\frac{1}{\int_0^1 f^2(x)dx-(\int_0^1 f(x)dx)^2}\begin{pmatrix}\int_0^1 f^2(x)dx & -\int_0^1 f(x)dx\\ -\int_0^1 f(x)dx & 1\end{pmatrix}.$$
On utilise alors le Lemme de Slutsky pour conclure.

3. On a $\widehat{\sigma}_\varepsilon^2=\frac{1}{n-2}\,\|Y-X\widehat{\theta}\|^2=\frac{1}{n-2}\,\sum_{i=1}^n\big(Y_i-\widehat{\theta}_0-\widehat{\theta}_1 f(i/n)\big)^2$.
En utilisant Cochran, on montre que $\widehat{\sigma}_\varepsilon^2\overset{\mathcal{L}}{\sim}\frac{\sigma_\varepsilon^2}{n-2}\,\chi^2(n-2)$.

4. La statistique de Student est: $\widehat{T}=\dfrac{\widehat{\theta}_1}{\widehat{\sigma}_\varepsilon\,\left(\sum_{i=1}^n f^2(i/n)-\left(\sum_{i=1}^n f^2(i/n)\right)^2\right)^{-1/2}}$.
Lorsque $n\to\infty$, la loi limite de $\widehat{T}$ est une loi normale centrée réduite. Ainsi la $p-value$ asymptotique du test est $\mathbb{P}\big(\mathcal{N}(0,1)>|\widehat{T}|\big)$.

5. On note d'abord que $\overline{U}=P_{[1]}U$ et $U-\overline{U}=P_{[1]^\perp}U$.
   (a) On a $\overline{Y}=\theta_0\mathbf{1}+\theta_1\overline{F}+P_{[1]}\varepsilon$. Ainsi $Y-\overline{Y}=\theta_1(F-\overline{F})+P_{[1]^\perp}\varepsilon$, d'où le résultat.
   (b) On a $\frac{1}{n}\|F-\overline{F}\|^2=\frac{1}{n}\sum_{i=1}^n f^2(i/n)-\big(\frac{1}{n}\sum_{i=1}^n f(i/n)\big)^2\underset{n\to\infty}{\longrightarrow}\int_0^1 f^2(t)dt-\big(\int_0^1 f(t)dt\big)^2$.
   (c) D'après le résultat du cours, on a $\widehat{\sigma}_\varepsilon^2\underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}}\sigma_\varepsilon^2$, donc d'après Slutsky, $\frac{1}{n}\|P_{[1]^\perp}\varepsilon\|^2=\frac{n-2}{n}\,\widehat{\sigma}_\varepsilon^2\underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}}\sigma_\varepsilon^2$.
   (d) Comme $\mathbb{E}(\varepsilon)=0$, on a $\mathbb{E}\big(\frac{2}{n}{}^tF\,P_{[1]^\perp}\varepsilon\big)=0$.
   De plus, $\mathrm{var}\big(\frac{2}{n}{}^tF\,P_{[1]^\perp}\varepsilon\big)=\frac{4}{n^2}\mathbb{E}\big(({}^tF\,P_{[1]^\perp}\varepsilon)^2\big)=\frac{4}{n^2}\mathbb{E}\big({}^tF\,P_{[1]^\perp}\varepsilon\,{}^t\varepsilon\,P_{[1]^\perp}F\big)=\frac{4\sigma_\varepsilon^2}{n^2}\,{}^tF\,P_{[1]^\perp}F=\frac{4\sigma_\varepsilon^2}{n}\,\frac{1}{n}\|F-\overline{F}\|^2$. Par la question précédente, on déduit que $\mathrm{var}\big(\frac{2}{n}{}^tF\,P_{[1]^\perp}\varepsilon\big)\underset{n\to\infty}{\longrightarrow}0$.
   En utilisant l'Inégalité de Bienaymé-Tchebychev, on en déduit que $\frac{2}{n}{}^tF\,P_{[1]^\perp}\varepsilon\underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}}0$.
   (e) Grâce à ce qui précède et avec le Lemme de Slutsky on a $\frac{1}{n}\,\|Y-\overline{Y}\|^2\underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}}\sigma_\varepsilon^2+\theta_1^2\big(\int_0^1 f^2(t)dt-\big(\int_0^1 f(t)dt\big)^2\big)$. De plus $\frac{1}{n}\,\|Y-\widehat{Y}\|^2\underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}}\sigma_\varepsilon^2$. Comme $\widehat{R}^2=1-\frac{\frac{1}{n}\,\|Y-\widehat{Y}\|^2}{\frac{1}{n}\,\|Y-\overline{Y}\|^2}$, par continuité de la division en dehors de 0, on en déduit le résultat.
   Si $\sigma_\varepsilon^2\to\infty$, on en déduit que $\widehat{R}^2\underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}}0$, ce qui est logique car le bruit devient de plus en plus important.

6. On a $\widehat{BIC}(m_0)=n\log(\frac{1}{n}\,\|Y-\overline{Y}\|^2)+2\log n$ et $\widehat{BIC}(m_1)=n\log(\frac{1}{n}\,\|Y-\widehat{Y}\|^2)+3\log n$. On en déduit que
$$\mathbb{P}\big(\widehat{BIC}(m_0)\leq\widehat{BIC}(m_1)\big)=\mathbb{P}\Big(\exp(\frac{\log n}{n})\,\frac{\frac{1}{n}\,\|Y-\widehat{Y}\|^2}{\frac{1}{n}\,\|Y-\overline{Y}\|^2}\geq 1\Big).$$
Comme $\exp(\frac{\log n}{n})\underset{n\to\infty}{\longrightarrow}1$ et $\frac{\frac{1}{n}\,\|Y-\widehat{Y}\|^2}{\frac{1}{n}\,\|Y-\overline{Y}\|^2}\underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}}\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\theta_1^2\big(\int_0^1 f^2(t)dt-\big(\int_0^1 f(t)dt\big)^2\big)}<1$, on en déduit que
$$\exp(\frac{\log n}{n})\,\frac{\frac{1}{n}\,\|Y-\widehat{Y}\|^2}{\frac{1}{n}\,\|Y-\overline{Y}\|^2}\underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}}\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\theta_1^2\big(\int_0^1 f^2(t)dt-\big(\int_0^1 f(t)dt\big)^2\big)}<1$$
et donc $\mathbb{P}\big(\widehat{BIC}(m_0)\leq\widehat{BIC}(m_1)\big)\underset{n\to\infty}{\longrightarrow}0$.
De ceci, comme le seul modèle faux est m_0, en déduit que la probabilité de choisir un modèle faux avec le BIC tend vers 0 quand $n\to\infty$.

$\square$