Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2016 – 2017

Econométrie II

Contrôle continu n°1, mars 2017

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

1. **(Sur 20 points)** On dispose d'une variable $Y = {}^t(Y_1, \cdots, Y_n)$ observée pour n individus ainsi que de $p+1$ variables potentiellement explicatives observées $X^{(j)}$ avec $1 \leq j \leq p$, tel que $X^{(j)} = {}^t(X_1^{(j)}, \cdots, X_n^{(j)})$. On suppose également que la matrice $Z = (\mathbb{I}, X^{(1)}, \cdots, X^{(p)})$ de taille $(n, p+1)$ est de rang $p+1$, avec $\mathbb{I} = {}^t(1, \cdots, 1)$, et qu'il existe $\theta = {}^t(\theta_0, \theta_1, \cdots, \theta_p)$ et $\theta_{p+1} \neq 0$ tels que:
$$Y = Z\,\theta + \theta_{p+1} X^{(p+1)} + \varepsilon,$$
où $X^{(p+1)} = {}^t(X_1^{(p+1)}, \cdots, X_n^{(p+1)})$ est un vecteur composé de n v.a.i.i.d. absolument continues par rapport à la mesure de Lebesgue et de variance σ_{p+1}^2 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, $X^{(p+1)}$ et ε étant indépendants. On note $\|\cdot\|$ la norme euclidienne classique de $\mathbf{R}^n$.

   (a) Montrer que presque sûrement la famille des vecteurs $(\mathbb{I}, X^{(1)}, \cdots, X^{(p+1)})$ est de rang $p+2$ (on pourra montrer que ces vecteurs sont presque sûrement linéairement indépendants) **(2 pts)**.

   (b) On effectue la régression linéaire multiple par moindres carrés de Y par rapport aux variables $X^{(j)}$ avec $1 \leq j \leq p+1$. Donner une expression matricielle de $\widehat{Y}$ **(1 pt)**. Démontrer qu'il existe une matrice carrée M de taille n telle que $\|Y - \widehat{Y}\|^2 = \|M\varepsilon\|^2 \leq \|\varepsilon\|^2$ **(2 pts)**, puis que $\mathbb{E}\big(\|\varepsilon\|^2 - \|M\varepsilon\|^2\big) = (p+2)\sigma_\varepsilon^2$ (utiliser l'espérance conditionnelle par rapport à $X^{(p+1)}$ car M est aléatoire!) **(3 pts)**. Après avoir déterminé la limite de $\frac{1}{n}\|\varepsilon\|^2$, en déduire que $\frac{1}{n}\|Y - \widehat{Y}\|^2 \xrightarrow[n\to+\infty]{\mathcal{P}} \sigma_\varepsilon^2$ **(2 pts)**. Déterminer ainsi qu'un équivalent asymptotique (en probabilité) du critère BIC, noté $\widehat{BIC}_{p+1}$, pour ce modèle de régression est $n\log(\sigma_\varepsilon^2)$ **(1 pt)**.

   (c) Rappeler quelle est la statistique du test de Student sur la significativité de $X^{(p+1)}$ **(0.5 pts)**. Quelle est, sous l'hypothèse H_0 (que l'on précisera), sa loi conditionnelle sachant $X^{(p+1)}$ **(0.5 pts)**? Si on suppose que, sachant $X^{(p+1)}$, l'estimateur du vecteur $\gamma = {}^t({}^t\theta, \theta_{p+1})$ est convergeant dans $\mathbb{L}^2$, montrer qu'asymptotiquement le test de Student décide la significativité de $X^{(p+1)}$ quelque soit l'erreur de première espèce α fixée **(3 pts)**.

   (d) On effectue maintenant une régression linéaire multiple par moindres carrés de Y par rapport aux variables $X^{(j)}$ avec $1 \leq j \leq p$. Donner l'expression matricielle de $\widehat{Y}$ en utilisant la matrice Z **(0.5 pts)**. On note $m = \mathbb{E}(X_1^{(p+1)})$ et $U_i = X_i^{(p+1)} - m$ pour $1 \leq i \leq n$. Montrer que $Y - \widehat{Y} = (I_n - H)(\theta_{p+1}U + \varepsilon)$ avec $U = {}^t(U_1, \cdots, U_n)$, I_n la matrice identité de taille n et H une matrice que l'on précisera **(1.5 pts)**. En déduire que $n\log(\theta_{p+1}^2\sigma_{p+1}^2 + \sigma_\varepsilon^2)$ est un équivalent asymptotique (en probabilité) du critère BIC, noté $\widehat{BIC}_p$, pour ce modèle de régression **(2 pts)**.

   (e) Démontrer qu'avec le critère BIC, la probabilité de choisir le vrai modèle avec $p+1$ variables plutôt que le sous-modèle avec p variables tend asymptotiquement vers 1 **(1 pt)**.

Proof. (a) Cela revient à montrer que pour tout $\theta \in (\mathbf{R}^*)^{p+1}$ et $\theta_{p+1} \in \mathbf{R}^*$, $\mathbb{P}\big(Z\,\theta + \theta_{p+1}X^{(p+1)} = 0\big) = 0$. Comme $X^{(p+1)}$ suit une loi absolument continue par rapport à la mesure de Lebesgue, s'il existe $(a, b) \in \mathbf{R}^2$ tel que $a + bX^{(p+1)} = 0$ alors nécessairement $(a, b) = (0, 0)$ presque sûrement. On en déduit donc que $\mathbb{P}\big(Z\,\theta + \theta_{p+1}X^{(p+1)} = 0\big) \neq 0$ entraîne nécessairement $\theta_{p+1} = 0$ et $Z\,\theta = 0$. Mais comme Z est supposée de rang $p+1$ alors $\theta = 0$. D'où la conclusion.

(b) En posant X la matrice $(\mathbb{I}, X^{(1)}, \cdots, X^{(p+1)})$, on a $\widehat{Y} = X({}^tXX)^{-1}\,{}^tXY$.
Il est clair que $Y - \widehat{Y} = (I_n - X({}^tXX)^{-1}\,{}^tX)\varepsilon$ d'où $M = I_n - X({}^tXX)^{-1}\,{}^tX$ matrice de projection orthogonale sur $[X]^\perp$. On a donc $\|Y - \widehat{Y}\|^2 = \|M\varepsilon\|^2$.
De plus comme M est une matrice de projection orthogonale, on sait que $\|M\varepsilon\|^2 + \|(I_n - M)\varepsilon\|^2 = \|\varepsilon\|^2$ par Pythagore donc $\|M\varepsilon\|^2 \leq \|\varepsilon\|^2$.
Toujours d'après Pythagore, on a $(\|\varepsilon\|^2 - \|M\varepsilon\|^2) = \|(I_n - M)\varepsilon\|^2$. Comme $(I_n - M)$ est la matrice de projection orthogonale

sur $[X]$ qui est un sev de $\mathbf{R}^n$ de dimension $p+2$, et comme ε est un vecteur centré tel que $\mathrm{var}(\varepsilon) = \sigma_\varepsilon^2 I_n$, on en déduit que $\mathbb{E}\big(\|(I_n - M)\varepsilon\|^2 \mid X^{(p+1)}\big) = \mathbb{E}\big({}^t\varepsilon(I_n - M)\varepsilon \mid X^{(p+1)}\big)$. Après le passage par la trace et le fait que $\mathrm{Trace}(AB) = \mathrm{Trace}(BA)$, on obtient que $\mathbb{E}\big(\|(I_n - M)\varepsilon\|^2 \mid X^{(p+1)}\big) = \mathbb{E}\big(\mathrm{Trace}\big((I_n - M)\varepsilon{}^t\varepsilon\big) \mid X^{(p+1)}\big) = \sigma_\varepsilon^2 \mathbb{E}\big(\mathrm{Trace}\big((I_n - M)\big) \mid X^{(p+1)}\big)$ car $X^{(p+1)}$ et ε sont indépendants. Mais comme quelque soit $X^{(p+1)}$, $I_n - M$ est une matrice de projection orthogonale sur un sev de dimension $p+2$, on en déduit que $\mathbb{E}\big(\mathrm{Trace}\big((I_n - M)\big) \mid X^{(p+1)}\big) = (p+2)$, en utilisant le fait que $\mathbb{E}\big(\|(I_n - M)\varepsilon\|^2\big) = \mathbb{E}\big(\mathbb{E}\big(\|(I_n - M)\varepsilon\|^2 \mid X^{(p+1)}\big)\big)$, on en déduit le résultat.
On utilise enfin l'inégalité de Markov pour $\frac{1}{n}\left(\|\varepsilon\|^2 - \|M\varepsilon\|^2\right)$ qui est une variable positive, d'espérance tendant vers 0, d'où la convergence en probabilité vers vers 0. On en conclut la limite en probabilité $\frac{1}{n}\|M\varepsilon\|^2$ vers σ_ε^2.
Pour finir, on utilise comme expression du BIC: $\widehat{BIC}_{p+1} = n\,\log(\frac{1}{n}\|M\varepsilon\|^2) + (p+2)\log(n)$ et on en déduit que $\widehat{BIC}_1 \simeq n\,\log(\sigma_\varepsilon^2)$ en probabilité.

(c) La statistique de test est $\widehat{T} = \dfrac{\widehat{\theta}_{p+1}}{\sqrt{\frac{1}{n-p-2}\|M\varepsilon\|^2\,{}^tC({}^tXX)^{-1}C}}$, où ${}^tC = (0, 0, \cdots, 0, 1)$ de taille $(p+2)$.
On sait que sous H_0, c'est-à-dire si $\theta_{p+1} = 0$, la loi de $\widehat{T}$ lorsque les variables exogènes sont déterministes est une loi $t(n-p-2)$. Il est donc de même pour la loi de $\widehat{T}$ conditionnellement à $X^{(p+1)}$.
Conditionnellement à $X^{(p+1)}$, l'estimateur $\widehat{\gamma}$ est sans biais et de matrice de variance de $\sigma_\varepsilon^2({}^tXX)^{-1}$. Donc si $\widehat{\gamma}$ est convergeant dans $\mathbb{L}^2$ conditionnellement à $X^{(p+1)}$ alors $\sigma_\varepsilon^2({}^tXX)^{-1} \xrightarrow[n\to+\infty]{\mathcal{P}} 0$. De même on en déduit que ${}^tC({}^tXX)^{-1}C \xrightarrow[n\to+\infty]{\mathcal{P}} 0$.
Par suite, comme $\theta_{p+1} \neq 0$, on peut écrire que conditionnellement à $X^{(p+1)}$, on a $\dfrac{\widehat{\theta}_{p+1} - \theta_{p+1}}{\sqrt{\frac{1}{n-p-2}\|M\varepsilon\|^2\,{}^tC({}^tXX)^{-1}C}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)$.
Or $\left|\dfrac{\theta_{p+1}}{\sqrt{\frac{1}{n-p-2}\|M\varepsilon\|^2\,{}^tC({}^tXX)^{-1}C}}\right| \xrightarrow[n\to+\infty]{\mathcal{P}} \infty$ puisque $\theta_{n+1} \neq 0$ et $\frac{1}{n-p-2}\|M\varepsilon\|^2 \xrightarrow[n\to+\infty]{\mathcal{P}} \sigma_\varepsilon^2$. On en conclut que $|\widehat{T}| \xrightarrow[n\to+\infty]{\mathcal{P}} \infty$, donc on choisira toujours H_1.

(d) On a $\widehat{Y} = Z\widehat{\theta} = Z({}^tZZ)^{-1}{}^tZY$.
On a $H = Z({}^tZZ)^{-1}{}^tZ$ matrice de la projection orthogonale sur le sev engendré par Z. Comme $\mathbf{1}$ appartient à Z, $HY = Z\theta + m\theta_{p+1}\mathbf{1} + H(\theta_{p+1}U + \varepsilon)$, d'où $(I_n - H)Y = (I_n - H)(\theta_{p+1}U + \varepsilon)$.
On peut utiliser les résultats obtenus plus haut, avec un sev de projection de dimension $(p+1)$ et où on remplace ε par $\theta_{p+1}U + \varepsilon$.
Par conséquent, $\frac{1}{n}\|Y - \widehat{Y}\|^2 \xrightarrow[n\to+\infty]{\mathcal{P}} \theta_{p+1}^2\sigma_{p+1}^2 + \sigma_\varepsilon^2$.
Donc $\widehat{BIC}_p \simeq n\,\log(\theta_{p+1}^2\sigma_{p+1}^2 + \sigma_\varepsilon^2)$ en probabilité.

(e) On en déduit que comme $\theta_{p+1} \neq 0$, alors $\mathbb{P}(\widehat{BIC}_p/\widehat{BIC}_{p+1} > 1) \xrightarrow[n\to\infty]{} 1$.

□

2. **(Sur 6 points)** Exercice avec le logiciel R:

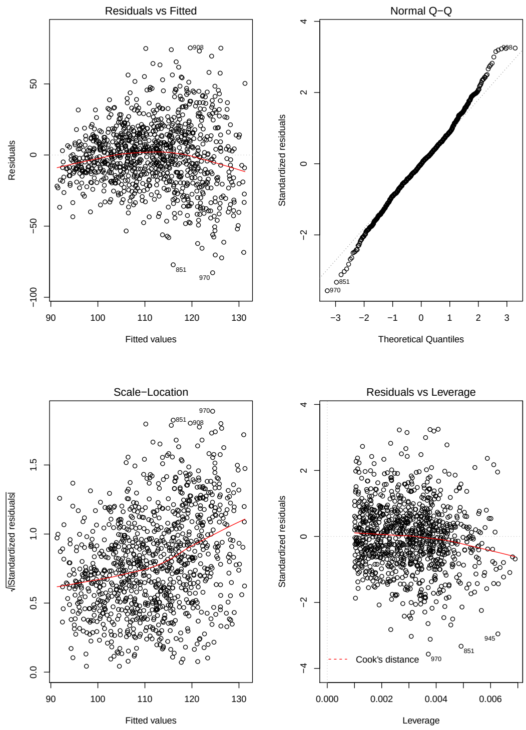
(a) On compile les commandes suivantes:

```
n=500;i=c(1:n)
X1=i; X2=runif(n,3,10); X3=log(i); epsilon=rnorm(n)
Y=40+10*X3+2*X2+i^0.5*epsilon
reg1=lm(Y~X1+X2)
summary(reg1)
par(mfrow=c(2,2));  plot(reg1)
```

Les résultats numériques et graphiques sont les suivants:

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 84.900889   2.783674  30.500   <2e-16 ***
X1           0.029195   0.002546  11.467   <2e-16 ***
X2           1.862415   0.364945   5.103    4e-07 ***

Residual standard error: 23.24 on 997 degrees of freedom
Multiple R-squared:  0.1371,Adjusted R-squared:  0.1353
F-statistic: 79.17 on 2 and 997 DF,  p-value: < 2.2e-16
```



Question 1: Que peut-on conclure quant à la régression effectuée? Est-ce surprenant?

(b) On poursuit avec les commandes qui suivent:

```
BX=boxcox(Y~X1+X2,plotit = TRUE,lambda = seq(-3,3))
ind=which(BX$y==max(BX$y))
(lambda=BX$x[ind])
```

Question 2: On obtient comme résultat `> [1] 0.5151515`*. Qu'a-t-on fait et qu'obtient-on? Pouvait-on s'y attendre?*

(c) Les commandes suivantes:

```
Z=Y^0.5; reg2=lm(Z~X1+X2+X3)
summary(reg2)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.9998093  0.3343708  20.934  < 2e-16 ***
X1          -0.0002147  0.0002418  -0.888    0.375
X2           0.0874513  0.0168894   5.178 2.71e-07 ***
X3           0.5138798  0.0708845   7.250 8.38e-13 ***

Residual standard error: 1.075 on 996 degrees of freedom
Multiple R-squared:  0.1732,Adjusted R-squared:  0.1707
F-statistic: 69.56 on 3 and 996 DF,  p-value: < 2.2e-16
```

Question 3: Expliquer ce qui a été fait et pourquoi. A-t-on gagné par rapport à la première régression? Quel devrait être la démarche suivante?