Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2018 – 2019

Econométrie II

Contrôle continu n°2, avril 2019

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

1. **(Sur 27 points)** Soit $(Y_i)_{1\leq i\leq n}$ une famille de variables aléatoires définie par:

$$Y_i = \theta_0 + \sum_{k=1}^{p} \theta_k \, X_i^{(k)} + \varepsilon_i \qquad \text{pour tout } i \in \{1,\dots,n\}, \qquad \text{où:} \tag{1}$$

- $\theta = {}^t(\theta_0, \theta_1, \dots, \theta_p)$ est un vecteur composé de $p+1$ réels inconnus.
- pour $1 \leq k \leq p$, les $(X_i^{(k)})_{1\leq i\leq n}$ sont p familles de réels connues telles que la matrice $X = \begin{pmatrix} 1 & X_1^{(1)} & \cdots & X_1^{(p)} \\ 1 & X_2^{(1)} & \cdots & X_2^{(p)} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & X_n^{(1)} & \cdots & X_n^{(p)} \end{pmatrix}$ soit une matrice de rang $p+1$ (on suppose que $p < n$).
- pour $0 < \tau < 1$, connu, et $n_1 = [n\tau]$ (partie entière de $n\tau$), avec $(\xi_i)_{i\in\mathbf{N}}$ une suite de v.a.i.i.d. de loi gaussienne centrée réduite, $\varepsilon_i = \sigma_1 \, \xi_i$ pour $1 \leq i \leq n_1$ et $\varepsilon_i = \sigma_2 \, \xi_i$ pour $n_1 + 1 \leq i \leq n$ avec $0 < \sigma_1 \neq \sigma_2 > 0$.

Par ailleurs, pour tout $k \in \mathbf{N}^*$, et pour M une matrice carrée de taille k, on note $\|M\| = \sup_{U\in(\mathbf{R}^k)^*} \dfrac{\|M\,U\|}{\|U\|}$, où pour $V \in \mathbf{R}^k$, $\|V\|^2 = {}^tV\,V$. On rappelle que pour A et B deux matrices carrées de taille k, $\|A\,B\| \leq \|A\|\,\|B\|$.

(a) On note $Y = (Y_i)_{1\leq i\leq n}$ et $\varepsilon = (\varepsilon_i)_{1\leq i\leq n}$. Ecrire le modèle (1) sous une forme vectorielle, en précisant la loi de l'erreur ε, en précisant $\Sigma = \mathrm{cov}(\varepsilon)$ **(1pt)**.

(b) Donner l'expression de la log-vraisemblance de $(Y_1, \cdots, Y_n)$ **(1pt)**. En supposant σ_1^2 et σ_2^2 connus, déterminer sous forme matricielle l'estimateur $\widehat{\theta}$ par maximum de vraisemblance **(2pts)**. En supposant θ connu, déterminer explicitement les estimateurs $\widehat{\sigma}_1^2$ et $\widehat{\sigma}_2^2$ de σ_1^2 et σ_2^2 (respectivement) par maximum de vraisemblance **(2pts)**. Expliquer pourquoi une expression explicite de l'estimateur par maximum de vraisemblance de θ, σ_1^2 et σ_2^2 n'est pas possible à obtenir dans le cas général où θ, σ_1^2 et σ_2^2 sont inconnus **(1pt)**.

(c) On va estimer ces paramètres en 2 temps. On commence par estimer θ par moindres carrés ordinaires, et on note $\widehat{\theta}$ cet estimateur. Rappeler son expression, donner son espérance et sa matrice de covariance **(2pts)**.

(d) On note Z_1 (respectivement Z_2) la matrice de taille $(n, p+1)$ dont les lignes de 1 à n_1 (respectivement, de n_1+1 à n) sont celles de X, et les lignes de n_1+1 à n (respectivement, de 1 à n_1) sont constituées de 0. Montrer que pour tout $U \in \mathbf{R}^n$, $\|{}^tX\,X\,U\|^2 = \|{}^tZ_1 Z_1\,U\|^2 + \|{}^tZ_2\,Z_2\,U\|^2$ **(2pts)**. En déduire que $\max\big(\|{}^tZ_1\,Z_1\|\,,\,\|{}^tZ_2\,Z_2\|\big) \leq \|{}^tX\,X\|$ **(1pt)**.

(e) Exprimer ${}^tX\,\Sigma\,X$ en fonction de σ_1, σ_2, Z_1 et Z_2 **(1.5pts)**. En utilisant la question précédente montrer que $\|\mathrm{cov}(\widehat{\theta})\| \leq (\sigma_1^2+\sigma_2^2)\|({}^tX\,X)^{-1}\|$ et donner une condition pour que $\widehat{\theta} \xrightarrow[n\to+\infty]{\mathcal{P}} \theta$ **(3.5pts)**.

(f) Soit $\widehat{\varepsilon} = Y - X\,\widehat{\theta} = (\widehat{\varepsilon}_i)_{1\leq i\leq n}$ et $\widehat{\sigma}_1^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}\widehat{\varepsilon}_i^2$. Montrer que $\frac{1}{n_1}\sum_{i=1}^{n_1}\varepsilon_i^2 \underset{n\to+\infty}{\xrightarrow{p.s.}} \sigma_1^2$ **(1pt)**. En notant $P_{[X]}$ la matrice de la projection orthogonale sur le sev engendré par X, montrer que $\sum_{i=1}^{n_1}((P_{[X]}\varepsilon)_i)^2 \leq \|P_{[X]}\varepsilon\|^2$ **(1pt)** puis que $\mathbb{E}(\|P_{[X]}\varepsilon\|^2) \leq (\sigma_1^2+\sigma_2^2)\text{Trace}(P_{[X]})$ **(2pts)**. En déduire que $\frac{1}{n_1}\sum_{i=1}^{n_1}((P_{[X]}\varepsilon)_i)^2 \underset{n\to+\infty}{\xrightarrow{\mathcal{P}}} 0$ **(2pts)**, et enfin que $\widehat{\sigma}_1^2 \underset{n\to+\infty}{\xrightarrow{\mathcal{P}}} \sigma_1^2$ **(3pts)**.

(g) Dans un deuxième temps, on déduit de ce qui précède un estimateur par moindres carrés pseudo-généralisés $\widetilde{\theta}$ de θ. Donner l'expresssion de cet estimateur **(1pt)** et montrer sa convergence **(2pts)**.

Proof. (a) On a $Y = X\theta + \varepsilon$ avec $\varepsilon = \mathcal{N}(0,\Sigma)$ avec Σ une matrice diagonale avec σ_1^2 (n_1 premiers termes) et σ_2^2 ($n - n_1$ premiers termes).

(b) La log-vraisemblance de $(Y_1,\cdots,Y_n)$ est $-\frac{n}{2}\log(2\pi) - \frac{n_1}{2}\log(\sigma_1^2) - \frac{n-n_1}{2}\log(\sigma_2^2) - \frac{1}{2}{}^t(Y - X\theta)\Sigma^{-1}(Y - X\theta)$.
Si θ est inconnu mais les σ_i^2 connus, cela revient à minimiser ${}^t(Y - X\theta)\Sigma^{-1}(Y - X\theta)$, donc $\widehat{\theta} = ({}^tX\Sigma^{-1}X)^{-1}\,{}^tX\Sigma^{-1}Y$ (estimateur par MCG).
Si les σ_i^2 sont inconnus mais θ connu, cela revient à minimiser $n_1\log(\sigma_1^2) + (n - n_1)\log(\sigma_2^2) + \frac{1}{\sigma_1^2}\sum_{i=1}^{n_1}(Y_i - (X\theta)_i)^2 +$ $+\frac{1}{\sigma_2^2}\sum_{i=n_1+1}^{n}(Y_i - (X\theta)_i)^2$. A l'aide des dérivées partielles, on trouve facilement que $\widehat{\sigma}_1^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}(Y_i - (X\theta)_i)^2$ et $\widehat{\sigma}_2^2 = \frac{1}{n-n_1}\sum_{i=n_1+1}^{n}(Y_i - (X\theta)_i)^2$.
Si tous les paramètres sont inconnus, la minimisation de $\widehat{\theta} = ({}^tX\Sigma^{-1}X)^{-1}\,{}^tX\Sigma^{-1}Y$ dépend de σ_1^2 et σ_2^2 comme l'expression de $\widehat{\sigma}_1^2$ dépend aussi de θ, on tombe sur une équation vérifiée par $\widehat{\sigma}_1^2$ et σ_2^2 qui n'a pas de solution explicite.

(c) On a $\widehat{\theta} = ({}^tXX)^{-1}\,{}^tXY$, $\mathbb{E}\widehat{\theta} = \theta$ et $\text{cov}(\widehat{\theta}) = ({}^tXX)^{-1}\,{}^tX\Sigma X({}^tXX)^{-1}$.

(d) Un rapide calcul montre que ${}^tX\,X = {}^tZ_1\,Z_1 + {}^tZ_2\,Z_2$. Or $\|{}^tX\,X\,U\|^2 = {}^t\big({}^tZ_1\,Z_1 + {}^tZ_2\,Z_2\big)\big({}^tZ_1\,Z_1 + {}^tZ_2\,Z_2\big)U$. Or $Z_2\,{}^tZ_1 = Z_1\,{}^tZ_2 = 0$ donc

$$\|{}^tX\,X\,U\|^2 = {}^t\big({}^tZ_1\,Z_1{}^tZ_1\,Z_1 + {}^tZ_2\,Z_2{}^tZ_2\,Z_2\big)U = \|{}^tZ_1Z_1\,U\|^2 + \|{}^tZ_2\,Z_2\,U\|^2. \tag{2}$$

Pour tout $U \in (\mathbf{R}^n)^*$, d'après ce qui précède, $\|{}^tZ_1Z_1\,U\|/\|U\| \leq \|{}^tX\,X\,U\|/\|U\|$. On en déduit donc que $\|{}^tZ_1Z_1\| \leq \|{}^tX\,X\|$. Le même raisonnement peut être fait pour aboutir à $\|{}^tZ_2Z_2\| \leq \|{}^tX\,X\|$. D'où le résultat.

(e) On a facilement ${}^tX\,\Sigma\,X = \sigma_1^2{}^tZ_1\,Z_1 + \sigma_2^2{}^tZ_2\,Z_2$.
On a $\|\text{cov}(\widehat{\theta})\| \leq \|({}^tX\,X)^{-1}\| \times \|{}^tX\,\Sigma\,X\,({}^tX\,X)^{-1}\|$ d'après la formule de la norme du produit de matrice. Mais d'après l'inégalité triangulaire, $\|{}^tX\,\Sigma\,X\,({}^tX\,X)^{-1}\| \leq \sigma_1^2\|{}^tZ_1\,Z_1\,({}^tX\,X)^{-1}\| + \sigma_2^2\|{}^tZ_2\,Z_2\,({}^tX\,X)^{-1}\|$. En utilisant le raisonnement de la relation (2), avec $U = ({}^tX\,X)^{-1}Y$ on aboutit à

$$\|Y\|^2 = \|{}^tZ_1Z_1\,({}^tX\,X)^{-1}Y\|^2 + \|Z_2\,Z_2\,({}^tX\,X)^{-1}Y\|^2. \tag{3}$$

On a donc $\|{}^tZ_1Z_1\,({}^tX\,X)^{-1}\| \leq 1$ et ainsi $\|{}^tX\,\Sigma\,X\,({}^tX\,X)^{-1}\| \leq (\sigma_1^2+\sigma_2^2)$ et donc $\|\text{cov}(\widehat{\theta})\| \leq (\sigma_1^2+\sigma_2^2)\|({}^tX\,X)^{-1}\|$.
On peut utiliser l'Inégalité de Bienaymé-Tchebytchev, et $\mathbb{P}\big(\|\widehat{\theta} - \theta\| > \varepsilon\big) \leq \frac{1}{\varepsilon^2}\,\mathbb{E}(\|\widehat{\theta} - \theta\|^2)$ et comme $\mathbb{E}(\|\widehat{\theta} - \theta\|^2) =$ $\text{Trace}(\text{cov}(\widehat{\theta})) \leq (p+1)\|\text{cov}(\widehat{\theta})\| \underset{n\to\infty}{\longrightarrow} 0$, on en déduit que $\mathbb{P}\big(\|\widehat{\theta} - \theta\| > \varepsilon\big) \underset{n\to\infty}{\longrightarrow} 0$, d'où le résultat.

(f) On peut appliquer la loi forte des grands nombres à $(\varepsilon_i^2)_{1\leq[n\tau]}$, qui est une suite de vaiid d'espérance $\sigma_1^2 < \infty$, d'où la convergence presque sûre de $\widehat{\sigma}_1^2$.
On a $\sum_{i=1}^{n_1}\big((P_{[X]}\varepsilon)_i\big)^2 \leq \sum_{i=1}^{n}\big((P_{[X]}\varepsilon)_i\big)^2 = \|P_{[X]}\varepsilon\|^2$.
On a $\mathbb{E}\big(\|P_{[X]}\varepsilon\|^2\big) = \mathbb{E}\big({}^t\varepsilon P_{[X]}\varepsilon\big) = \mathbb{E}\big(\text{Trace}({}^t\varepsilon P_{[X]}\varepsilon)\big) = \mathbb{E}\big(\text{Trace}(P_{[X]}\varepsilon{}^t\varepsilon)\big) = \text{Trace}(P_{[X]}\,\Sigma)$. Mais $\text{Trace}(P_{[X]}\,\Sigma) =$ $\text{Trace}(P_{[X]}\,(\sigma_1^2+\sigma_2^2)I_n) - \text{Trace}(P_{[X]}\,\Sigma')$ où Σ' est la matrice diagonale avec d'abord n_1 fois σ_2^2 puis $n - n_1$ fois σ_1^2. Or $\text{Trace}\big(P_{[X]}\,\Sigma'\big) \geq 0$ car on peut encore écrire cette trace comme $\mathbb{E}\|P_{[X]}\varepsilon'\|^2$ où les $\varepsilon' \sim \mathcal{N}(0,\Sigma')$. D'où $\text{Trace}\big(P_{[X]}\,\Sigma\big) \leq \text{Trace}\big(P_{[X]}\,(\sigma_1^2+\sigma_2^2)I_n\big) = (\sigma_1^2+\sigma_2^2)(p+1)$ (car $\text{Trace}(P_{[X]}) = p+1$).
On a $\widehat{\sigma}_1^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}\big(\varepsilon_i - (P_{[X]}\varepsilon)_i\big)^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}\varepsilon_i^2 - \frac{2}{n_1}\sum_{i=1}^{n_1}\varepsilon_i(P_{[X]}\varepsilon)_i) + \frac{1}{n_1}\sum_{i=1}^{n_1}\big((P_{[X]}\varepsilon)_i\big)^2$. Le premier terme tend en probabilité vers σ_1^2, le troisième terme est positif avec une espéance qui tend vers 0, donc tend en probabilité vers 0 et le second terme, avec l'Inégalité de Cauchy-Schwartz tend aussi en probabilité vers 0. On en déduit que le tout tend vers σ_1^2 en probabilité.

(g) On considère $\widehat{\Sigma}$ la matrice diagonale avec $\widehat{\sigma}_1^2$ sur les n_1 premiers termes et $\widehat{\sigma}_2^2 = \frac{1}{n-n_1}\sum_{i=n_1+1}^{n}\varepsilon_i^2$ sur les $(n - n_1)$ autres termes. L'estimateur par moindres carrés pseudo-généralisés sera: $\widetilde{\theta} = ({}^tX\widehat{\Sigma}^{-1}X)^{-1}\,{}^tX\widehat{\Sigma}^{-1}Y$.

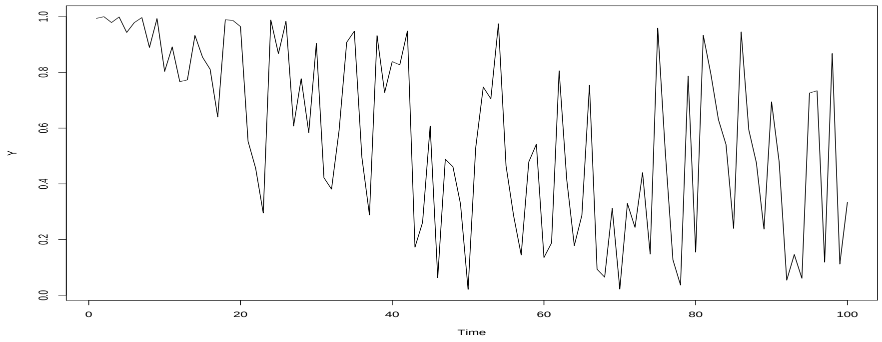
□

2. **(Sur 7 points)** Exercice avec le logiciel R:

(a) On lance les commandes suivantes:

```
n=100; th0=-8; th1=2; sigma=2
X1=log(c(1:n))
Y=1/(1+exp(th0+th1*X1+sigma*rnorm(n)))
plot.ts(Y)
```

Le graphique obtenu est le suivant:



Question 1: Donner le modèle vérifié par Y. Peut-on ainsi escompter les paramètres du modèle? **(1pt)**

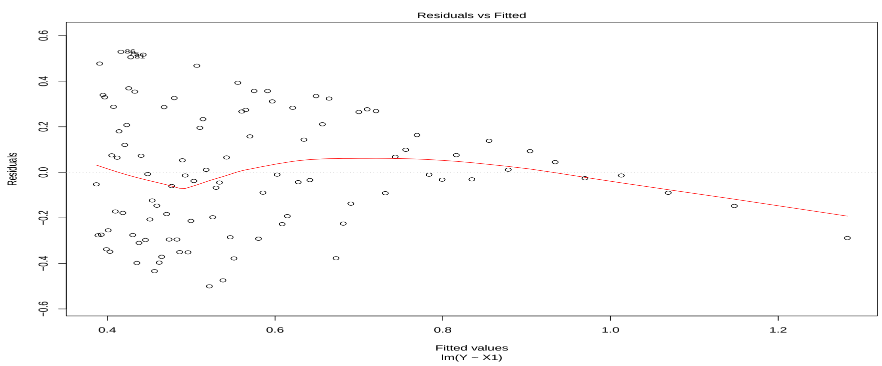
(b) On poursuit avec les commandes:

```
reg1=lm(Y~X1)
summary(reg1)
plot(reg1)
```

Les résultats numériques et graphiques sont les suivants:

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.28248    0.10716  11.968  < 2e-16 ***
X1          -0.19449    0.02855  -6.811 7.85e-10 ***

Residual standard error: 0.2637 on 98 degrees of freedom
Multiple R-squared:  0.3213,Adjusted R-squared:  0.3144
F-statistic: 46.39 on 1 and 98 DF,  p-value: 7.855e-10
```



Question 2: Expliquez ce qui a été fait. Que peut-on conclure quant à la régression effectuée et était-ce escompté? **(1.5pts)**

(c) On poursuit avec les commandes qui suivent:

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

Voici le résultat numérique:

```
> lambda
[1] 0.8787879
```

Question 3: Qu'a-t-on fait et qu'obtient-on? Qu'en conclure? **(1pt)**

(d) Voici les dernières commandes:

```
T=as.numeric(Y>1/2)
reg2=glm(T~X1,family=binomial(link="probit"),na.action=na.pass)
summary(reg2)
```

Les résultats numériques et graphiques sont les suivants:

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   4.0143     0.9618   4.174 3.00e-05 ***
X1           -1.0098     0.2436  -4.146 3.38e-05 ***

    Null deviance: 136.66  on 99  degrees of freedom
Residual deviance: 110.71  on 98  degrees of freedom
AIC: 114.71
```

Question 4: Expliquer ce qui a été fait. Montrer mathématiquement pourquoi on a obtenu ces estimations et pourquoi on ne pouvait pas escompter identifier les différents paramètres du modèle initial **(3.5pts)**

Proof. (a) Le modèle s'écrit $Y_i = (1 + \exp(\theta_0 + \theta_1 \log(i) + \varepsilon_i))^{-1}$ pour $i = 1, \ldots, 100$, où (ε_i) est une famille de vaiid de loi $\mathcal{N}(0, \sigma^2)$, avec $\theta_0 = -8$, $\theta_1 = 2$ et $\sigma = 2$.
Ce n'est pas un modèle linéaire, mais c'est un modèle paramétrique, donc une estimation par maximum de vraisemblance pour permettre l'estimation des paramètres.

(b) On effectue une régression linéaire par moindres carrés de Y en fonction de $X1$ pour les 100 individus.
Si le test global de Fisher ou le test de Student sur la significativité de $X1$ valident le modèle, le $R2$ est faible et surtout le graphe des résidus en fonction des valeurs prédites est très mauvais: une forme se dégage clairement. On n'est pas surpris car le modèle n'est pas du tout linéaire et on a utilisé un modèle linéaire pour le représenter. De plus toutes les valeurs des Y_i sont dans l'intervalle $[0, 1]$ ce qui rend tout à fait inappropriée l'estimation par régression.

(c) On effectue une transformation de Box-Cox pour diminuer l'hétéroscédasticité du modèle. Le paramètre obtenu, 0.88 est très proche de 1, ce qui signifie que l'on préfère garder Y plutôt que d'utiliser une puissance de Y.

(d) On utilise une nouvelle variable T qui vaut 1 si $Y > 1/2$ et 0 sinon. On effectue ensuite une régression logistique de T en fonction de $X1$ avec la fonction de lien PROBIT.
Les résultats sont très bons mais ne permettent pas d'identifier exactement le modèle. En effet, si on regarde $\mathbb{P}(T = 1) = \mathbb{P}(Y > 1/2) = \mathbb{P}(\theta_0 + \theta_1 \log(i) + \varepsilon_i < 0)$. En utilisant la fonction de répartition F d'une loi $\mathcal{N}(0, 1)$, cela revient à calculer $F(-\frac{\theta_0}{\sigma} - \frac{\theta_1}{\sigma} \log(i))$. En utilisant la fonction PROBIT on estime donc exactement ce modèle, mais les paramètres estimés sont $-\frac{\theta_0}{\sigma}$ et $-\frac{\theta_1}{\sigma}$: le modèle n'est pas identifiable avec 3 paramètres mais seulement avec 2. Et on comprend que le premier paramètre estimé soit 4.0143 très proche de $-(-8)/2 = 4$.

$\square$