Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2014 – 2015

Econométrie II

Contrôle continu n°2, avril 2015

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

1. **(Sur 22 points)** Pour n et p deux entiers tels que $n \geq p+1$, soit $Y = {}^t(Y_i)_{1\leq i\leq n}$ le vecteur colonne de variables aléatoires représentant une variable quantitative (le nombre de personne infectées par épidémie) en différents lieux d'un pays et défini par:

$$Y_i = \theta_0 + \theta_1 X_i^{(1)} + \cdots + \theta_1 X_i^{(p)} + \varepsilon_i \qquad \text{pour tout } i = 1, \cdots, n, \tag{1}$$

avec:

- $(X_i^{(j)})_{1\leq i\leq n,\, 1\leq j\leq p}$ étant connues et telles 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 $p+1$;
- $\theta = (\theta_j)_{0\leq j\leq p}$ un vecteur de nombres réels inconnus;
- $\varepsilon = {}^t(\varepsilon_i)_{1\leq i\leq n}$ un vecteur d'erreur. On suppose qu'il y a une interférence spatiale entre les erreurs, qu'il existe une matrice connue $W = (W_{ij})_{1\leq i,j\leq n}$ vérifiant $W_{ii} = 0$ et $\sum_{j=1}^n |W_{ij}| \leq 1$ pour tout $i = 1, \cdots, n$, pas forcément symétrique, donnant l'influence entre ces erreurs et ainsi:

$$\varepsilon = \lambda\, W\, \varepsilon + \xi,$$

avec ξ un vecteur gaussien de loi $\mathcal{N}_n(0, \sigma^2 I_n)$, I_n la matrice identité, $\sigma^2 > 0$ et $\lambda \in \mathbf{R}$, deux réels inconnus.

(a) **(Facultatif, sur 6 points)** Pour A une matrice carrée d'ordre n, on note la norme matricielle $\|A\| = \sup\limits_{Z\in\mathbf{R}^n} \dfrac{\|AZ\|}{\|Z\|}$, où $\|Z\|^2 = \sum_{i=1}^n Z_i^2$, avec $Z = (Z_i)_{1\leq i\leq n}$. Montrer que $\|W\| \leq \sqrt{n}$ **(2pts)**, puis que pour tout $\lambda \in]-1/\sqrt{n}, 1/\sqrt{n}[$, $\|\lambda^N W^N\| \underset{N\to\infty}{\longrightarrow} 0$ **(1.5pt)**. Calculer $(I_n - \lambda\, W)(I_n + \sum_{k=1}^N \lambda^k W^k)$ pour $N \in \mathbf{N}^*$ **(1pt)**, et en déduire que $I_n - \lambda\, W$ est une matrice inversible pour tout $\lambda \in]-1/\sqrt{n}, 1/\sqrt{n}[$ **(1.5pts)**.

(b) Montrer que pour tout $\lambda \in]-1/\sqrt{n}, 1/\sqrt{n}[$ ε peut s'écrire en fonction de $\lambda\, W$, I_n et ξ **(1pt)**. En déduire la loi de Y **(1pt)**, puis que la log-vraisemblance de Y s'écrit **(1pt)**:

$$\log\big(L_{\theta,\lambda,\sigma^2}(Y)\big) = \log(|\det(I_n - \lambda\, W)|) - \frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\,{}^t(Y - X\theta)(I_n - \lambda\,{}^tW)(I_n - \lambda W)(Y - X\theta).$$

(c) Si on suppose que λ est connu, déterminer les expressions des estimateurs par maximum de vraisemblance $(\widehat{\theta}_\lambda, \widehat{\sigma}^2_\lambda)$ de (θ, σ^2) **(1.5pt)**. Quelle est la loi de $\widehat{\theta}$ **(1pt)**?

(d) On suppose maintenant que λ est inconnu. Exprimez $\log(|\det(I_n - \lambda\, W)|)$ en fonction de P_W le polynôme caractéristique de la matrice W **(1pt)**. Soit les estimateurs par maximum de vraisemblance $(\widehat{\theta}, \widehat{\lambda}, \widehat{\sigma}^2)$ de $(\theta, \lambda, \sigma^2)$. Exprimer $\widehat{\theta}$ en fonction de $\widehat{\lambda}$, $\widehat{\sigma}^2$ en fonction de $\widehat{\theta}$ et $\widehat{\lambda}$, et une relation liant $\widehat{\theta}$, $\widehat{\lambda}$ et $\widehat{\sigma}^2$ en fonction de P'_W **(2pts)**.

(e) On veut tester s'il y a réellement une interaction spatiale entre les données. Proposer une statistique de test, dépendant de $\widehat{\theta}_0, \widehat{\sigma}^2_0, \widehat{\theta}, \widehat{\lambda}$ et $\widehat{\sigma}^2$, permettant de tester H_0: $\lambda = 0$ contre H_1: $\lambda \neq 0$ **(2pts)**.

(f) On se place dans le cadre où une observation (par exemple Y_1) a une influence très forte sur toutes les autres et ainsi on pose $W = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix}$. Montrer que $I_n - \lambda W$ est inversible pour tout $\lambda \in \mathbf{R}$. Déterminer alors les équations explicites permettant de calculer $(\widehat{\theta}, \widehat{\lambda}, \widehat{\sigma}^2)$ **(2pts)**. Dans le cas où θ est connu, montrer que $\widehat{\lambda}$ converge vers λ **(3.5pts)**.

Proof. (a) On a $WZ = (\sum_{j=1}^n W_{ij} Z_j)_{1\leq j \leq n}$. Mais d'après Cauchy-Schwarz, $(\sum_{j=1}^n W_{ij} Z_j)^2 \leq (\sum_{j=1}^n W_{ij}^2)(\sum_{j=1}^n Z_j^2)$ donc $\|WZ\|^2 \leq \left(\sum_{i=1} \sum_{j=1}^n W_{ij}^2\right)\|Z\|^2$ et comme $|W_{ij}| \leq 1$, $\sum_{1\leq j\leq n} W_{ij}^2 \leq \sum_{1\leq j \leq n} |W_{ij}| \leq 1$, d'où $\|WZ\|^2 \leq n\,\|Z\|^2$ ce qui donne le résultat puisque $\frac{\|WZ\|^2}{\|Z\|^2} \leq n$ donc $\frac{\|WZ\|}{\|Z\|} \leq \sqrt{n}$ pour tout Z.

Il est clair que pour toute mat rice A, $\|C\,A\| = |C|\|A\|$ avec $C \in \mathbf{R}$ et $\|A^p Z\| = \|A(A^{p-1}Z)\| \leq \|A\|\|A^{p-1}Z\| \leq \cdots \leq \|A\|^p\|Z\|$ par itération. Donc $\|\lambda^N W^N\| \leq (|\lambda|\|W\|)^N$ et comme $|\lambda|\|W\| < 1$ on a bien $\|\lambda^N W^N\| \underset{N\to\infty}{\longrightarrow} 0$.

On a par somme télescopique, $(I_n - \lambda W)(I_n + \sum_{j=1}^N \lambda^j W^j) = I_n - \lambda^{N+1} W^{N+1}$.

On déduit de ce qui précède que $\|I_n - (I_n - \lambda W)(I_n + \sum_{j=1}^N \lambda^j W^j)\| \underset{N\to\infty}{\longrightarrow} 0$, donc $(I_n - \lambda W)(I_n + \sum_{j=1}^\infty \lambda^j W^j) = I_n$ et ainsi $(I_n - \lambda W)$ est inversible de matrice inverse $(I_n + \sum_{j=1}^\infty \lambda^j W^j)$.

(b) On a $\varepsilon = \lambda W \varepsilon + \xi$ donc $\varepsilon = (I_n - \lambda W)^{-1}\xi$.
De ceci, on déduit que $Y = X\theta + (I_n - \lambda W)^{-1}\xi$ donc comme $\xi \sim \mathcal{N}_n(0, \sigma^2\, I_n)$, on a $Y \sim \mathcal{N}_n\big(X\theta\,,\,\sigma^2\,(I_n - \lambda W)^{-1}(I_n - \lambda^t W)^{-1}\big)$.
Cela implique la vraisemblance de Y:

$$L_{\theta,\lambda,\sigma^2}(Y) = \frac{1}{(2\pi)^{n/2}}\,\frac{1}{\sigma^{2n}(\det((I_n - \lambda W)^{-1}(I_n - \lambda^t W)^{-1}))^{1/2}}\exp\Big(-\frac{1}{2\sigma^2}{}^t(Y - X\theta)(I_n - \lambda^t W)(I_n - \lambda W)(Y - X\theta)\Big),$$

et ainsi on obtient la log-vraisemblance donnée.

(c) Dans le cas où λ, on se retrouve dans le cadre des MCG et on obtient:

$$\begin{cases} \widehat{\theta} &= \left({}^tX(I_n - {}^t\lambda W)(I_n - \lambda W)X\right)^{-1}\,{}^tX(I_n - \lambda^t W)(I_n - \lambda W)Y \\ \widehat{\sigma}^2 &= \frac{1}{n}\,{}^t(Y - X\widehat{\theta})(I_n - \lambda^t W)(I_n - \lambda W)(Y - X\widehat{\theta}) \end{cases}$$

On a alors $\widehat{\theta} \sim \mathcal{N}_{p+1}\Big(\theta\,,\,\sigma^2\big({}^tX(I_n - {}^t\lambda W)(I_n - \lambda W)X\big)^{-1}\Big)$.

(d) On a $\log(|\det(I_n - \lambda\, W)|) = \log(|\lambda|^n|\det(\lambda^{-1}I_n - W)|) = n\log(|\lambda|) + \log(|P_W(\lambda^{-1})|)$.
Pour obtenir un maximum local de la log-vraisemblance, on différentie et ainsi:

$$\begin{cases} \widehat{\theta} &= \left({}^tX(I_n - {}^t\widehat{\lambda} W)(I_n - \widehat{\lambda} W)X\right)^{-1}\,{}^tX(I_n - \lambda^t W)(I_n - \lambda W)Y \\ \widehat{\sigma}^2 &= \frac{1}{n}\,{}^t(Y - X\widehat{\theta})(I_n - \widehat{\lambda}^t W)(I_n - \widehat{\lambda} W)(Y - X\widehat{\theta}) \\ \frac{\widehat{\sigma}^2}{\widehat{\lambda}^2}\left(n\widehat{\lambda} - P'_W(1/\widehat{\lambda})\right) &= {}^t(Y - X\widehat{\theta})(I_n - \widehat{\lambda} W)(Y - X\widehat{\theta}) \end{cases}$$

(e) On peut utiliser un test du rapport de vraisemblance pour tester si $\lambda = 0$. Ainsi on définit: $\widehat{T} = \frac{\sup_{\lambda=0,\theta,\sigma^2} L_{\theta,\lambda,\sigma^2}(Y)}{\sup_{\lambda,\theta,\sigma^2} L_{\theta,\lambda,\sigma^2}(Y)}$. En utilisant ce qui précède, on obtient:

$$\widehat{T} = \frac{L_{\widehat{\theta}_0,0,\widehat{\sigma}_0^2}(Y)}{L_{\widehat{\theta},\widehat{\lambda},\widehat{\sigma}^2}(Y)}.$$

(f) Il est clair que dans ce cas $\det(I - \lambda W) = 1$, donc $\widehat{\lambda}$ est solution de

$${}^t(Y - X\widehat{\theta})(I_n - \widehat{\lambda} W)(Y - X\widehat{\theta}) = 0.$$

Ceci permet d'écrire que $\widehat{\lambda} = \dfrac{{}^t(Y - X\widehat{\theta})(Y - X\widehat{\theta})}{{}^t(Y - X\widehat{\theta})W(Y - X\widehat{\theta})}$.

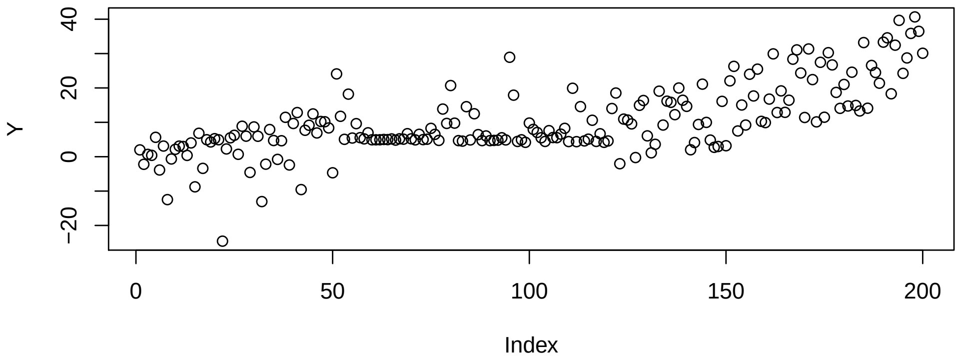
$\square$

2. **(Sur 10 points)** Exercice de TP utilisant le logiciel R

   (a) On a tapé les commandes suivantes:

```
Y=0; i=1:200;
Z1=i; Z2=rnorm(200,0)^2
epsilon=runif(200,-3,3)
Y[1:50]=(3+0.2*Z1[1:50]-5*Z2[1:50]+epsilon[1:50])
Y[51:120]=(50-3*sqrt(Z1[51:120])+10*Z2[51:120]^2+epsilon[51:120])^0.5
Y[121:200]=(-30+0.3*Z1[121:200]+4*epsilon[121:200])
plot(Y)
```

   On obtient la figure:



   Questions 1: Quelle est la loi de `Z2`*? Quelle est la loi de* `epsilon` *et sa variance? Formaliser le modèle généré. A-t-on simulé un modèle linéaire?*

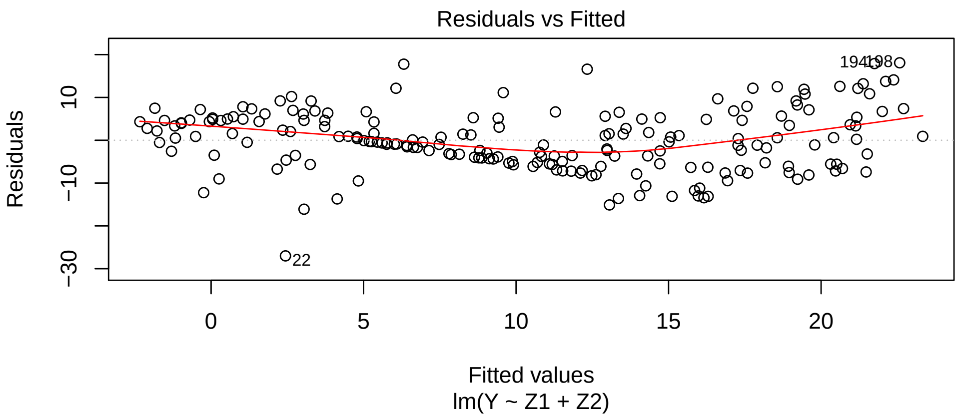
   (b) On tape ensuite les commandes:

```
reg=lm(Y~Z1+Z2)
summary(reg)
plot(reg)
```

   Voici les résultats et un des graphes obtenus:

```
    Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.15502    0.99535  -2.165   0.0316 *
Z1           0.12480    0.00808  15.445   <2e-16 ***
Z2          -0.81316    0.32420  -2.508   0.0129 *

Residual standard error: 6.596 on 197 degrees of freedom
Multiple R-squared:  0.5559,    Adjusted R-squared:  0.5514
F-statistic: 123.3 on 2 and 197 DF,  p-value: < 2.2e-16
```



   Questions 2: Qu'a-t-on fait avec ces commandes? Que représentent explicitement (en fonction des variables (Y_i), $(Z1_i)$ et $(Z2_i)$) les valeurs numériques -0.81316, 0.00808, -2.165, 0.5559 et $2.2e-16$? Que peut-on en conclure?

(c) On a ensuite tapé les commandes:

```
Q=c(); M=c()
for (j in c(3:195))
{for (k in c(j+3:200))
{reg1=lm(Y[1:j]~Z1[1:j]+Z2[1:j])
reg2=lm(Y[(j+1):k]~Z1[(j+1):k]+Z2[(j+1):k])
reg3=lm(Y[(k+1):200]~Z1[(k+1):200]+Z2[(k+1):200])
M=cbind(M,c(j,k))
Q=c(Q,sum(reg1$res^2)+sum(reg2$res^2)+sum(reg3$res^2))}}
i0=(Q==min(Q)); Indice=c(M[1,i0],M[2,i0]); Indice
```

Voici le résultat:

```
[1]  50 121
```

Questions 3: Qu'a-t-on fait avec ces commandes et pourquoi l'a-t-on fait? Que conclure de ce résultat?

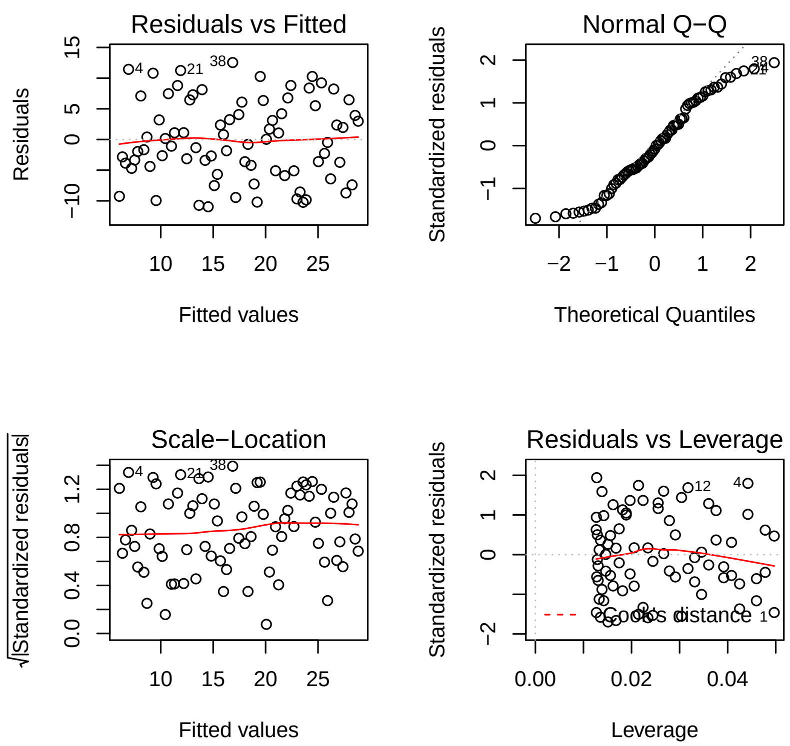
(d) On a enfin tapé les commandes:

```
library(MASS)
Y3=Y[122:200]; Z13=Z1[122:200]; Z23=Z2[122:200]
ZZ=as.data.frame(cbind(Z13,Z23));
y.lm=lm(Y3~.,data=ZZ);
y.bic=stepAIC(y.lm,k=log(79)); plot(lm(Y3~Z13))
```

Voici les résultats:

```
Start:  AIC=306.66
Y3 ~ Z13 + Z23
       Df Sum of Sq    RSS    AIC
- Z23   1      16.4 3262.7 302.69
<none>              3246.3 306.66
- Z13   1    3518.9 6765.1 360.30

Step:  AIC=302.69
Y3 ~ Z13
       Df Sum of Sq    RSS    AIC
<none>              3262.7 302.69
- Z13   1      3503 6765.7 355.93
```



Questions 4: Qu'a-t-on fait avec ces commandes et qu'a-t-on obtenu? Qu'est-ce que le 79 apparaissant dans une commande? Est-on satisfait du résultat obtenu et expliquer pourquoi on pouvait s'y attendre?