Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2015 – 2016

Econométrie II

Contrôle continu n°2, avril 2016

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

1. **(Sur 29 points)** Pour n et p deux entiers tels que $n \geq p+1$, soit $Y = {}^t(Y_i)_{1\leq i\leq n}$ défini par:

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

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 $p+1$;
- $\theta = {}^t(\theta_j)_{0\leq j\leq p}$ un vecteur de nombres réels inconnus;
- $\varepsilon = {}^t(\varepsilon_i)_{1\leq i\leq n}$ un vecteur d'erreur non observé tel que $\mathbb{E}(\varepsilon) = 0$ et $\mathrm{cov}(\varepsilon) = \Sigma$, où Σ est une matrice définie positive inconnue de valeurs propres $0 < \lambda_1 \leq \cdots \leq \lambda_n$ (inconnues).

(a) On effectue une régression par moindres carrés ordinaires pour obtenir un estimateur $\widehat{\theta}$. Montrer que $\widehat{\theta}$ est un estimateur sans biais de θ et $\mathrm{cov}(\widehat{\theta}) = \big({}^tX\,X\big)^{-1}\,{}^tX\,\Sigma\,X\,\big({}^tX\,X\big)^{-1}$ **(1pt)**.

(b) Pour une matrice carrée $A = (a_{ij})_{1\leq i,j\leq m}$ où $m \in \mathbf{N}^*$, soit la norme matricielle définie par

$$\|A\|_2 = \sup_{U\in\mathbf{R}^m,\ U\neq 0} \frac{\|A\,U\|_2}{\|U\|_2} \quad \text{où} \quad \|Z\|_2^2 = {}^tZ\,Z \quad \text{pour tout } Z \in \mathbf{R}^m.$$

Montrer que si A est une matrice symétrique, alors $\|A\|_2 = \max\big(\big|\text{Valeurs propres de } A\big|\big)$ **(2pts)**.

(c) Montrer que ${}^tX\,X$ est une matrice définie positive et on notera $0 < \mu_1 \leq \cdots \leq \mu_{p+1}$ ses valeurs propres **(1.5pts)**. Montrer que ${}^tV\,\Sigma\,V \leq \lambda_n \|V\|_2^2$ pour tout $V \in \mathbf{R}^n$ **(2pts)**, et en déduire que ${}^tW\,\mathrm{cov}(\widehat{\theta})\,W \leq \lambda_n\,\|X\,\big({}^tX\,X\big)^{-1}\,W\|_2^2$ pour tout $W \in \mathbf{R}^{p+1}$ **(1.5pts)**, puis que ${}^tW\,\mathrm{cov}(\widehat{\theta})\,W \leq \frac{\lambda_n}{\mu_1}\|W\|_2^2$ **(2pts)**. En déduire une condition suffisante de convergence de $\widehat{\theta}$ vers θ quand $n \to \infty$ **(1pt)**.

(d) On se place désormais dans le cadre d'hétéroscédasticité suivant: les (ε_i) sont indépendantes et $\mathrm{var}(\varepsilon_i) = \sigma_i^2$, avec $\underline{\sigma}^2 \leq \sigma_i^2 \leq \overline{\sigma}^2$ pour tout $i = 1, \ldots, n$ et $\frac{1}{n}\sum_{i=1}^n \sigma_i^2 \underset{n\to\infty}{\longrightarrow} \sigma^2$ où $0 < \underline{\sigma}^2 \leq \sigma^2 \leq \overline{\sigma}^2$ sont trois réels fixés inconnus. Montrer que $\widehat{\theta} \xrightarrow[n\to+\infty]{\mathcal{P}} \theta$ si $\mu_1 \underset{n\to\infty}{\longrightarrow} +\infty$ **(1pt)**.

(e) On considère l'estimateur de la variance par MCO $\widehat{\sigma}^2 = \frac{1}{n-(p+1)}\,{}^t(Y - X\,\widehat{\theta})\,(Y - X\,\widehat{\theta})$. Montrer que $\mathbb{E}(\widehat{\sigma}^2) = \frac{1}{n-(p+1)}\big(\mathrm{Tr}(\Sigma) - \mathrm{Tr}(H\,\Sigma)\big)$ où vous exprimerez $H = (H_{ij})_{1\leq i,j\leq n}$ en fonction X **(2pts)**. Montrer que pour tout $i = 1, \ldots, n$, $H_{ii} = \sum_{k=1}^n H_{ik}^2 \geq 0$ **(1pt)**, puis que $\underline{\sigma}^2(p+1) \leq \mathrm{Tr}(H\,\Sigma) \leq \overline{\sigma}^2(p+1)$ **(2pts)**. En déduire que $\mathbb{E}(\widehat{\sigma}^2) \underset{n\to\infty}{\longrightarrow} \sigma^2$ **(1pt)**.

(f) Montrer que $\frac{1}{n-(p+1)}\sum_{i=1}^n \varepsilon_i^2 - \widehat{\sigma}^2 \geq 0$ **(1.5pts)**, et $\frac{1}{n-(p+1)}\sum_{i=1}^n \varepsilon_i^2 \xrightarrow[n\to+\infty]{\mathcal{P}} \sigma^2$ **(2.5pts)**. En déduire que $\widehat{\sigma}^2 \xrightarrow[n\to+\infty]{\mathcal{P}} \sigma^2$ **(2pts)**.

(g) On va montrer que les tests de Student peuvent être valables dans un tel cadre. On suppose que $p = 1$, n pair, et $X_i^{(1)} = 1$, $\sigma_i^2 = \overline{\sigma}^2$ pour $i = 1, \ldots, n/2$, $X_i^{(1)} = 0$, $\sigma_i^2 = \underline{\sigma}^2$ pour $i = n/2+1, \ldots, n$, avec $\underline{\sigma}^2 < \overline{\sigma}^2$. Donner l'expression exacte de $\widehat{\theta}$, de $\mathrm{cov}(\widehat{\theta})$, de σ^2 **(3.5pts)**. Rappeler l'expression matricielle du test de Student $\widehat{T}_1$ relatif à la significativité de $X^{(1)}$ **(0.5pts)**. En déduire que $\widehat{T}_1 = \sqrt{n}\,\frac{\widehat{\theta}_1}{2\,\sigma}$ **(1pt)**, puis montrer que $\widehat{T}_1 \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0\,,\,1)$ **(1pt)**.

Proof. (a) On a $\widehat{\theta} = ({}^tX\,X)^{-1}\,{}^tX\,Y = \theta + ({}^tX\,X)^{-1}\,{}^tX\,\varepsilon$. Comme $\mathbb{E}(\varepsilon) = 0$, on a bien $\mathbb{E}(\widehat{\theta}) = \theta$.
On a $\mathrm{cov}(\widehat{\theta}) = \mathbb{E}\big(({}^tX\,X)^{-1}\,{}^tX\,\varepsilon\,{}^t\varepsilon\,X\,({}^tX\,X)^{-1}\big) = ({}^tX\,X)^{-1}\,{}^tX\,\mathbb{E}(\varepsilon\,{}^t\varepsilon)\,X\,({}^tX\,X)^{-1} = ({}^tX\,X)^{-1}\,{}^tX\,\Sigma\,X\,({}^tX\,X)^{-1}$.

(b) A est symétrique donc diagonalisable. Il en est de même pour A^2, dont les valeurs propres sont le carré de celles de A. Ainsi $\|A\,U\|_2^2 = {}^tU\,A^2\,U = {}^t(P\,U)\,D^2\,(P\,U)$ où P est une matrice orthogonale et D une matrice diagonale. D'où $\sup_{U\in\mathbf{R}^m,\ U\neq 0} \frac{\|A\,U\|_2^2}{\|U\|_2^2} = \sup_{U'\in\mathbf{R}^m,\ U'\neq 0} \frac{\|D\,U'\|_2^2}{\|U'\|_2^2} = \sup_{(u_i)_i} \frac{\sum_{i=1}^m vp_i^2 u_i^2}{\sum_{i=1}^m u_i^2} \leq \max(vp_i^2)$ (où vp_i désigne les valeurs propres de A).

(c) ${}^tX\,X$ est une matrice symétrique donc diagonalisable. De plus ${}^tU^tX\,X\,U = \|X\,U\|_2^2$, donc la forme quadratique $U \to {}^tU^tX\,X\,U$ est positive. Comme ${}^tX\,X$ est inversible, elle ne peut pas avoir 0 pour valeur propre, donc elle est définie positive.
Comme précédemment, on écrit ${}^tV\,\Sigma\,V = {}^t(P\,V)\,D\,(P\,V)$ où P est diagonale et D la matrice diagonale constituée des λ_i. D'où le résultat.
On a également ${}^tW\,\mathrm{cov}(\widehat{\theta})\,W = {}^tW\,\mathrm{cov}(\widehat{\theta})\,W = ({}^tX\,X)^{-1\,t}X\,\Sigma\,X\,({}^tX\,X)^{-1}\,W = {}^tW'\,\Sigma\,W'$ avec $W' = X\,({}^tX\,X)^{-1}\,W$. En utilisant ce qui précède, on a bien ${}^tW'\,\Sigma\,W' \leq \lambda_n\,\|W'\|_2^2$. Mais $\|W'\|_2^2 = {}^tW({}^tX\,X)^{-1\,t}X\,X\,({}^tX\,X)^{-1}\,W = {}^tW({}^tX\,X)^{-1}\,W \leq \max(|\text{Valeur propre de }({}^tX\,X)^{-1}|)\,\|W\|_2^2 = 1/\mu_1\,\|W\|_2^2$. D'où le résultat.
On a donc $\|(\mathrm{cov}(\widehat{\theta}))^{1/2}\|_2^2 \leq \frac{\lambda_n}{\mu_1}$, d'où $\|\mathrm{cov}(\widehat{\theta})\|_2 \leq \frac{\lambda_n}{\mu_1}$. Une condition suffisante de convergence est donc $\frac{\lambda_n}{\mu_1} \underset{n\to\infty}{\longrightarrow} 0$.

(d) Dans ce cas, $\lambda_n \leq \overline{\sigma}^2$ donc la condition précédente devient bien $\mu_1 \to \infty$.

(e) On a $\mathbb{E}(\widehat{\sigma}^2) = \frac{1}{n-(p+1)}\,\mathbb{E}\Big[\mathrm{Tr}({}^t(P_{[X]^\perp}\varepsilon)P_{[X]^\perp}\varepsilon)\Big] = \frac{1}{n-(p+1)}\,\mathbb{E}\Big[\mathrm{Tr}(P_{[X]^\perp}\,P_{[X]^\perp}\,\varepsilon\,{}^t\varepsilon)\Big]$ du fait que $\mathrm{Tr}(A\,B) = \mathrm{Tr}(B\,A)$. Avec le fait que $P_{[X]^\perp}\,P_{[X]^\perp} = P_{[X]^\perp} = I_n - P_{[X]}$ et $\mathbb{E}(\varepsilon\,{}^t\varepsilon) = \Sigma$, on en déduit que $\mathbb{E}(\widehat{\sigma}^2) = \frac{1}{n-(p+1)}\,\mathrm{Tr}((I_n - P_{[X]})\Sigma) = \frac{1}{n-(p+1)}\,\big(\mathrm{Tr}(\Sigma) - \mathrm{Tr}(H\,\Sigma)\big)$ avec $H = P_{[X]} = X({}^tX\,X)^{-1\,t}X$.
Comme $H = {}^tH$ et $H^2 = H = {}^tH\,H$, la matrice ${}^tH\,H$ a pour terme général $\sum_{k=1}^n H_{ki}H_{kj}$, donc sur la diagonale, pour $i = j$, on obtient $H_{ii} = \sum_{k=1}^n H_{ki}H_{ki}$.
On a $\mathrm{Tr}(H\,\Sigma) = \sum_{i=1}^n \sigma_i^2 H_{ii}$. Comme d'après ce qui précède, $H_{ii} \geq 0$, et $\underline{\sigma}^2 \leq \sigma_i^2 \leq \overline{\sigma}^2$, on a $\underline{\sigma}^2\sum_{i=1}^n H_{ii} \leq \mathrm{Tr}(H\,\Sigma) \leq \overline{\sigma}^2\sum_{i=1}^n H_{ii}$. Mais on sait que $\sum_{i=1}^n H_{ii} = \mathrm{Tr}(H) = p+1$ (matrice de projection orthogonale sur un sous-espace vectoriel de dimension $p+1$), d'où le résultat.

(f) On sait que $\|\varepsilon\|_2^2 \geq \|P_{[X]^\perp}\varepsilon\|_2^2 = {}^t(Y - X\,\widehat{\theta})\,(Y - X\,\widehat{\theta})$, d'où le résultat.
Soit $Z_n = \frac{1}{n}\sum_{k=1}^n \varepsilon_k^2$. Alors pour $u \in \mathbf{R}$, $\mathbb{E}(e^{i\,u\,Z_n}) = \prod_{k=1}^n \mathbb{E}(e^{i\,u\,\varepsilon_k^2/n})$, du fait de l'indépendance des (ε_k). On en déduit que $\log(\mathbb{E}(e^{i\,u\,Z_n})) = \sum_{k=1}^n \log(\mathbb{E}(e^{i\,u\,\varepsilon_k^2/n}))$. Mais grâce à un DL en 0, $\mathbb{E}(e^{i\,u\,\varepsilon_k^2/n}) = 1 + iu\,\sigma_k^2/n(1+o(1))$, d'où $\sum_{k=1}^n \log(\mathbb{E}(e^{i\,u\,\varepsilon_k^2/n})) = \sum_{k=1}^n \log(1+iu\,\sigma_k^2/n(1+o(1))) = i\,u\,\frac{1+o(1)}{n}\sum_{k=1}^n \sigma_k^2 \underset{n\to\infty}{\longrightarrow} i\,u\,\sigma^2$. Par suite, $\mathbb{E}(e^{i\,u\,Z_n}) \underset{n\to\infty}{\longrightarrow} e^{i\,u\,\sigma^2}$, qui est la fonction caractéristique de la masse de Dirac en σ^2. D'où $Z_n \underset{n\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \sigma^2$, soit $Z_n \underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}} \sigma^2$, et avec Slutsky, $\frac{1}{n-(p+1)}\sum_{k=1}^n \varepsilon_k^2 \underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}} \sigma^2$.
La variable $S_n = \frac{1}{n-(p+1)}\sum_{i=1}^n \varepsilon_i^2 - \widehat{\sigma}^2$ est positive, et son espérance est $\frac{1}{n-(p+1)}\,\mathrm{Tr}(H\Sigma) \underset{n\to\infty}{\longrightarrow} 0$. Donc d'après l'Inégalité de Markov, $S_n \underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}} 0$ et comme $\frac{1}{n-(p+1)}\sum_{i=1}^n \varepsilon_i^2 \underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}} \sigma^2$, on en déduit que $\widehat{\sigma}^2 \underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}} \sigma^2$.

(g) Dans ce cas, ${}^tX\,X = \frac{n}{2}\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$, d'où $({}^tX\,X)^{-1} = \frac{2}{n}\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$.
On en déduit que $\widehat{\theta} = ({}^tX\,X)^{-1\,t}X\,Y = \frac{2}{n}\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^{n/2} Y_i \end{pmatrix} = \begin{pmatrix} \frac{2}{n}\sum_{i=n/2+1}^n Y_i \\ \frac{2}{n}\Big(\sum_{i=1}^{n/2} Y_i - \sum_{i=n/2+1}^n Y_i\Big) \end{pmatrix}$.
Par ailleurs, ${}^tX\,\Sigma\,X = \frac{n}{2}\begin{pmatrix} \overline{\sigma}^2 + \underline{\sigma}^2 & \overline{\sigma}^2 \\ \overline{\sigma}^2 & \overline{\sigma}^2 \end{pmatrix}$ d'où $\mathrm{cov}(\widehat{\theta}) = \frac{2}{n}\begin{pmatrix} \underline{\sigma}^2 & -\underline{\sigma}^2 \\ -\underline{\sigma}^2 & \underline{\sigma}^2 + \overline{\sigma}^2 \end{pmatrix}$.
Enfin, on facilement $\sigma^2 = \frac{1}{2}\,(\underline{\sigma}^2 + \overline{\sigma}^2)$. Il est clair que $\widehat{\theta}_1$ est asymptotiquement gaussien car H est telle que $\max(|H_{ii}|) \underset{n\to\infty}{\longrightarrow} 0$.
La variance de $\widehat{\theta}_1$ est $\frac{2}{n}\,(\underline{\sigma}^2 + \overline{\sigma}^2) = \frac{4}{n}\,\sigma^2$. Or on sait que $\widehat{\sigma}^2 \underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}} \sigma^2$, donc $\widehat{T} = \frac{\sqrt{n}\,\widehat{\theta}_1}{2\,\widehat{\sigma}}$, statistique de Student, converge bien en loi vers une $\mathcal{N}(0,1)$.

□

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

(a) On a tapé les commandes suivantes:

```
n=100; Y=0; i=1:100;
Z1=i; Z2=5*rnorm(n,0)
Z3=runif(n,0,20); Z4=4*cos(2*pi*i/12)
epsilon=2*rnorm(n,0)^2
Y=10-3*Z3+4*sqrt(i+3)+5*epsilon
reg1=lm(Y~Z1+Z2+Z3+Z4)
summary(reg1)
```

On obtient les résultats:

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 30.52028    3.52702   8.653 1.24e-13 ***
Z1           0.30467    0.04661   6.537 3.12e-09 ***
Z2           0.13032    0.28060   0.464    0.643
Z3          -2.84563    0.22257 -12.785  < 2e-16 ***
Z4          -0.79840    0.48118  -1.659    0.100
```

```
Residual standard error: 13.44 on 95 degrees of freedom
Multiple R-squared:  0.6933,    Adjusted R-squared:  0.6804
F-statistic: 53.68 on 4 and 95 DF,  p-value: < 2.2e-16
```

Questions 1: Quelle est la loi de `epsilon`*? Ecrire Y sous la forme $Y = X\,\theta + \varepsilon$ avec X dépendant des Z_i et $\mathbb{E}(\varepsilon) = 0$. Que pensez-vous des résultats de cette régression? Pouvait-on s'y attendre?*

Proof. La loi de `epsilon` est 2 fois un Chi-deux à 1 degré de liberté **(0.5pts)**.
Le modèle s'écrit $Y = \theta_0 + \theta_1\,\sqrt{Z_1+3} + \theta_3\,Z_3 + \varepsilon$, avec $\varepsilon_i = 5(epsilon - 2)$ car $\mathbb{E}(epsilon) = 2$, d'où $\theta_0 = 20$, $\theta_1 = 4$ et $\theta_3 = -3$ **(1.5pts)**.
Les résultats indiquent que $Z2$ et $Z4$ sont non significatives, ce qui n'est pas surprenant car elles ne sont pas dans le modèle **(0.5pts)**. □

(b) On tape ensuite les commandes:

```
library(MASS)
ZZ=as.data.frame(cbind(Z1,Z2,Z3,Z4));
y.lm=lm(Y~.,data=ZZ);
y.bic=stepAIC(y.lm,k=log(n));
```

Voici les résultats:

```
Start:  AIC=537.54
Y ~ Z1 + Z2 + Z3 + Z4

       Df Sum of Sq   RSS    AIC
- Z2    1      39.0 17198 533.16
- Z4    1     497.3 17656 535.79
<none>              17159 537.54
- Z1    1    7718.8 24878 570.08
- Z3    1   29524.2 46683 633.02

Step:  AIC=533.16
Y ~ Z1 + Z3 + Z4

       Df Sum of Sq   RSS    AIC
- Z4    1     515.2 17713 531.50
<none>              17198 533.16
- Z1    1    7710.1 24908 565.59
- Z3    1   31283.9 48482 632.19

Step:  AIC=531.5
Y ~ Z1 + Z3

       Df Sum of Sq   RSS    AIC
<none>              17713 531.50
- Z1    1    7551.5 25264 562.41
- Z3    1   30880.0 48593 627.82
```

Questions 2: Qu'a-t-on fait avec ces commandes et que peut-on en conclure? Pouvait-on s'y attendre?

Proof. On a effectué une sélection de modèle à l'aide du critère BIC et on a obtenu un modèle ne contenant que $Z1$ et $Z3$. Ce n'est pas surprenant puisque ce sont les seules variables intervenant dans le modèle **(1pt)**. □

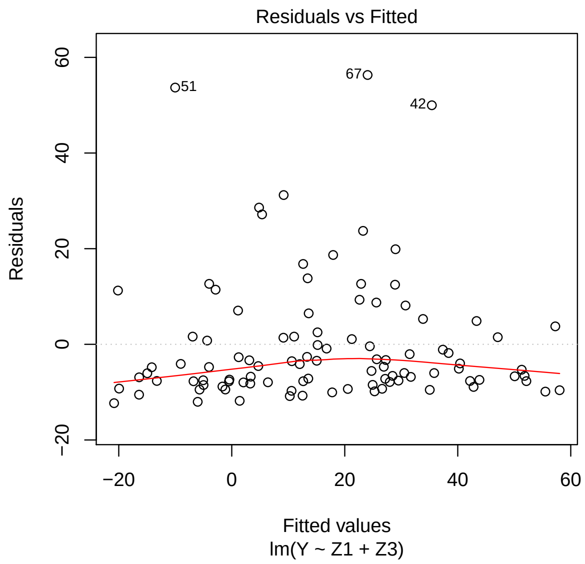
(c) De nouvelles commandes sont tapées:

```
summary(y.bic)
plot(y.bic)
```

Voici les résultats et un des graphes obtenus:

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 30.57434    3.52651   8.670 9.93e-14 ***
Z1           0.30105    0.04681   6.431 4.79e-09 ***
Z3          -2.84153    0.21851 -13.004  < 2e-16 ***

Residual standard error: 13.51 on 97 degrees of freedom
Multiple R-squared:  0.6834,    Adjusted R-squared:  0.6768
F-statistic: 104.7 on 2 and 97 DF,  p-value: < 2.2e-16
```



Questions 3: Qu'a-t-on fait avec ces commandes? Que signifie précisément la valeur numérique 30.57434 *et pouvait-on s'y attendre? Dire également en termes mathématiques ce que sont* 0.30105*,* −2.84153*,* 0.21851*,* −13.004 *et* 4.79e − 09*. Au final que pensez-vous des résultats obtenus?*

Proof. On a effectué une régression linéaire par moindres carrés de Y par rapport à $Z1$ et $Z3$, variables issus de la sélection de modèle BIC **(0.5pts)**.
30.57434 représente l'estimation de α_0 par moindres carrés dans le modèle $Y_i = \alpha_0 + \alpha_1\, Z1 + \alpha_3\, Z3 + \varepsilon_i$. Notons ici que la valeur théorique de α_0 est $\theta_0 + 4 * \sqrt{3} \simeq 27$, donc on n'est pas surpris d'avoir obtenu 30.57434 **(1pt)**.
0.30105 et −2.84153 sont les estimateurs respectifs $\widehat{\alpha}_1$ et $\widehat{\alpha}_3$, où $\widehat{\alpha} = ({}^tZ\,Z)^{-1}\,{}^tZ\,Y$, Z étant la matrice de 3 colonnes contenant des 1, $Z1$ et $Z3$ **(0.5pts)**.
0.21851 représente une estimation de l'écart-type de $\widehat{\alpha}_3$, obtenue en prenant la racine carrée du terme $(3,3)$ de la matrice $\widehat{\sigma}^2\,({}^tZ\,Z)^{-1}$ **(0.5pts)**.
−13.004 représente la valeur du test de Student de significativité de $Z3$, test de nullité de α_3, statistique qui vaut $\frac{\widehat{\alpha}_3}{\widehat{\sigma}\,\sqrt{{}^tC_3({}^tZ\,Z)^{-1}C_3}}$, avec ${}^tC_3 = (0,0,1)$ **(0.5pts)**.
$4.79e-09$ représente la p-value du test précédent, c'est-à-dire la probabilité que $|T| > 13.004$ lorsque T suit une loi de Student à 97 degrés de liberté **(0.5pts)**.
Si toutes les variables sont significatives, le graphe des $\widehat{\varepsilon}$ en fonction des $\widehat{Y}$ indique une forme particulière qui nous indique un problème de modèle. Cela n'est pas surprenant puisque ce n'est pas $Z1$ qui intervient mais plutôt $\sqrt{3 + Z1}$ **(0.5pts)**. □

(d) On a enfin tapé les commandes:

```
Q=matrix(0,50,50)
for (a in c(1:50))
{ for (b in c(1:50))
{ Z1a=Z1^(a/10); Z3b=Z3^(b/10)
reg=lm(Y~Z1a+Z3b)
Q[a,b]=sum(reg$res^2) }}
which(Q==min(Q),arr.ind = TRUE)
```

Voici le résultat:

```
  row col
[1,]   6  10
```

Questions 4: Qu'a-t-on fait avec ces commandes et qu'a-t-on obtenu? Pouvait-on s'attendre au résultat obtenu? Si vous deviez prédire Y_i pour $i = 101$, quelles commandes R *utiliseriez vous?*

Proof. Ici on a cherché les valeurs de α et β qui optimise le modèle $Y_i = \theta_0 + \theta_1\, Z1_i^{\alpha} + \theta_3\, Z3_i^{\beta} + \varepsilon_i$ au sens de la somme des résidus au carré (donc qui maximise le R^2). On effectue des régressions pour α et β variant dans $\{0.1, 0.2, \ldots, 4.9, 5\}$. On obtient $\widehat{\alpha} = 0.6$ et $\widehat{\beta} = 1$ **(1pt)**. On pouvait s'attendre à de tels résultats puisque le modèle simulé l'est avec $\alpha \simeq 0.5$ (on considère $\sqrt{Z1 + 3}$) et $Z3$ **(0.5pts)**.
Pour la prédiction, on pourrait utiliser les commandes (on remplace $Z3$ par son espérance, donc 10): **(1pt)**

```
Z1a=Z1^0.6
y.reg=lm(Y~Z1a+Z3)
y.pred=y.reg$coeff[1]+ y.reg$coeff[2]*101^0.6+ y.reg$coeff[3]*10
```

□