Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2013 – 2014

Econométrie II

Examen final, mai 2014

Examen de 3h00. Tout document ou calculatrice est interdit.

1. **(18 points)** Soit une expérience médicale consistant à mesurer le taux d'insuline (variable T) en fonction du sexe du patient (variable S) et de la prise ou non d'un médicament (variable M). On suppose que l'on a ainsi $2n$ individus (où $n \in \mathbf{N}$) et avec $(\varepsilon_i)_{i\in\mathbf{N}}$ une suite de variables aléatoires centrées indépendantes de même loi possédant un moment d'ordre 4, on a:

$$T_i = \theta_0 + \theta_1 S_i + \theta_2 M_i + \varepsilon_i \qquad \text{pour } i = 1, \cdots, 2n, \text{ avec} \tag{1}$$

- $S_i = 1$ (femme) pour i pair et $S_i = 0$ (homme) pour i impair;
- $M_i = 1$ (prise du médicament) pour $i = 1, \cdots, n$, $M_i = 0$ (prise d'un placébo) pour $i = n+1, \cdots, 2n$;
- $\theta = {}^t(\theta_0, \theta_1, \theta_2) \in \mathbf{R}^3$ est le vecteur des paramètres inconnus.

(a) Déterminer explicitement l'estimateur $\widehat{\theta}$ par moindres carrés ordinaires de θ pour $n \geq 2$ (on obtient notamment $\widehat{\theta}_1 = \frac{1}{n}\sum_{i=1}^n (Y_{2i-1} - Y_{2i})$ et $\widehat{\theta}_2 = \frac{1}{n}\sum_{i=1}^n (Y_i - Y_{n+i})$) **(3pts)**.

(b) Montrer que cet estimateur est convergent quand $n \to \infty$ **(2pts)** et donner le théorème de la limite centrale qu'il vérifie **(1pt)**.

(c) Déterminer un estimateur de $\sigma^2 = \mathrm{var}(\varepsilon_0)$ **(1pt)**. Déterminer un théorème de la limite centrale vérifié par cet estimateur lorsque $n \to \infty$ **(1pt)**.

(d) Proposer un test (expliciter la statistique de test) $\widehat{T}$ pour décider si $\theta_1 = \theta_2$ **(2pts)**. Quel est le comportement asymptotique de $\widehat{T}$ lorsque $n \to \infty$ **(1pt)**?

(e) On suppose que le vrai modèle est bien (1) mais qu'on ne le sait pas. On effectue d'abord une régression du modèle comportant S mais pas M et on obtient ainsi $\widetilde{\theta}$. Démontrer qu'asymptotiquement $\widetilde{\theta}$ ne tend pas vers θ **(3pts)**. Démontrer qu'asymptotiquement le critère BIC de ce sous-modèle a une probabilité tendant vers 0 d'être inférieur au critère BIC du vrai modèle **(2pts)**.

(f) On suppose maintenant que la variance de $(\varepsilon_{n+1}, \cdots, \varepsilon_{2n})$ n'est plus σ^2 mais $2\sigma^2$. Déterminer l'expression de l'estimateur par moindres carrés généralisés de θ ainsi que sa matrice de covariance asymptotique **(2pts)**.

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

(a) On a tapé les commandes suivantes:

```
k=c(1:100);  v=runif(100);  X=4+3*rnorm(100);  Y=log(v*k)*sqrt(k)+2*X
```

Questions 1: Quelle est la loi de X? On supposera que les valeurs de X sont connues, mais pas celles de u. Quel modèle suit formellement Y? Est-ce un modèle linéaire? Si c'est le cas, l'écrire exactement sous la forme habituelle **(2pts)**.

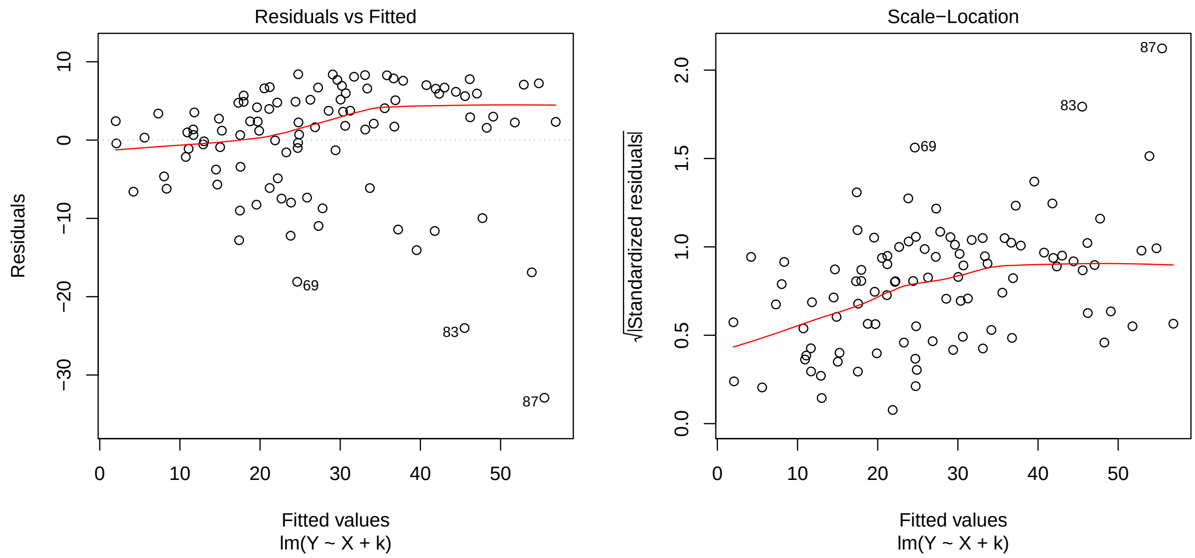
(b) On a ensuite tapé les commandes:

```
reg1=lm(Y~X+k); summary(reg1); plot(reg1)
```

Voici les résultats et une partie des graphes obtenus:

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.70971    1.72023   0.413    0.681
X            2.06024    0.24904   8.273 7.05e-13 ***
k            0.36965    0.02652  13.940  < 2e-16 ***

Residual standard error: 7.574 on 97 degrees of freedom
Multiple R-squared: 0.7573,     Adjusted R-squared: 0.7523
F-statistic: 151.3 on 2 and 97 DF,  p-value: < 2.2e-16
```



Questions 2: Qu'a-t-on fait ici? Dîtes exactement ce que représentent les résultats numériques suivants: 0.70971, 0.02652, 8.273, 7.574, 0.7523 *et* 151.3*. S'attendait-on aux résultats* 2.06024 *et* 0.36965*? Décrire les 2 graphes en expliquant ce qu'ils représentent et les conclusions à en tirer. Au final, que pensez-vous de cette régression? Que voudrait-on faire pour améliorer cette régression?* **(4pts)**

(c) On tape ensuite les commandes:

```
Z=cbind(1+0*k,X,sqrt(k)*log(k));   Sig=diag(1/sqrt(k))
Theta2=solve(t(Z)%*%Sig%*%Z)%*%t(Z)%*%Sig%*%Y;   Theta2
```

Voici les résultats:

```
  -2.5266115
X  2.1811195
   0.8097249
```

Questions 3: Qu'a-t-on fait ici et pourquoi? Est-on satisfait des résultats obtenus? **(2pts)**

(d) Enfin on a tapé les commandes:

```
YY=(Y>20)*1; YY;  kk=sqrt(k)*log(k)
logitY=glm(YY~X+kk,family=binomial(link="probit"),na.action=na.pass)
summary(logitY)
library(lmtest); lrtest(logitY,1:2)
```

Voici ce que l'on obtient:

```
YY
  [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 1 0 1 0 0 1 1 0 0 1 1 1 1 0 1 1
 [38] 1 1 0 0 1 1 1 1 0 1 1 1 1 0 1 1 0 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1
 [75] 0 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.58638    1.03427  -4.434 9.23e-06 ***
X            0.48411    0.13154   3.680 0.000233 ***
kk           0.14704    0.03018   4.873 1.10e-06 ***

Likelihood ratio test

Model 1: YY ~ X + kk
Model 2: YY ~ 1
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   3 -26.911
2   1 -64.104 -2 74.385  < 2.2e-16 ***
```

Questions 4: Expliquez ce qui a été fait. Pouvait-on s'y attendre? Que conclure du test? Pouvait-on s'y attendre? **(2pts)**