

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

**Econométrie II**

Examen final, mai 2017

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

1. **(Sur 20 points)** Pour  $n$  et  $p$  deux entiers tels que  $n \geq p + 1$ , on observe  $Y = {}^t(Y_i)_{1 \leq i \leq n}$  défini par:

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

avec une famille connue de réels  $(X_i^{(j)})_{1 \leq i \leq n, 1 \leq j \leq p}$ , telle que  $X = \begin{pmatrix} 1 & X_1^{(1)} & \dots & X_1^{(p)} \\ \vdots & \vdots & & \vdots \\ 1 & X_n^{(1)} & \dots & X_n^{(p)} \end{pmatrix}$  soit une matrice de

rang  $p + 1$ ,  $\theta = {}^t(\theta_j)_{0 \leq j \leq p}$  un vecteur de nombres réels inconnus et  $\varepsilon = {}^t(\varepsilon_i)_{1 \leq i \leq n}$ , où les  $(\varepsilon_i)$  sont des v.a.i.i.d. centrées non observées, avec  $\text{var}(\varepsilon_1) = \sigma^2 > 0$ , inconnue.

**Notations:** Pour  $m \in \mathbf{N}^*$ ,  $I_m$  est la matrice identité de taille  $m$ . Pour  $M$  une matrice réelle quelconque,  ${}^tM$  est la transposée de  $M$ . Pour  $u$  un vecteur colonne quelconque dans  $\mathbf{R}^m$ ,  $\|u\|^2 = {}^tu u$ .

L'exercice propose une méthode pour estimer  $\theta$ , appelée *régression ridge*, en déterminant pour  $\lambda \geq 0$  fixé:

$$\tilde{\theta}(\lambda) = \underset{\theta \in \mathbf{R}^{p+1}}{\text{Arg min}} \|Y - X \theta\|^2 + \lambda \sum_{i=0}^p \theta_i^2. \quad (2)$$

- (a) Déterminer  $\tilde{\theta}(0)$  en fonction de  $X$  et  $Y$  **(0.5 pts)**.  
 (b) Que peut-on dire de  $\lim_{\lambda \rightarrow \infty} \tilde{\theta}(\lambda)$  **(1.5 pts)**?  
 (c) Montrer que  ${}^tX X$  est une matrice définie positive (on pourra considérer une forme quadratique...) **(1.5 pts)**.  
 En déduire que pour tout  $\lambda \geq 0$ ,  ${}^tX X + \lambda I_{p+1}$  est inversible **(0.5 pts)**.  
 (d) En utilisant la différentiation, démontrer que pour tout  $\lambda \geq 0$  fixé,

$$\tilde{\theta}(\lambda) = ({}^tX X + \lambda I_{p+1})^{-1} {}^tX Y$$

(on vérifiera que  $\tilde{\theta}(\lambda)$  est bien un minimum et qu'il est unique) **(2 pts)**.

- (e) Montrer que si  $p + 1 > n$  alors  $\tilde{\theta}(0)$  existe dès que  $\lambda > 0$  alors que l'estimateur de  $\theta$  par moindres carrés ordinaires n'existe pas **(1.5pts)**.  
 (f) Déterminer  $\mathbb{E}(\tilde{\theta}(\lambda))$  **(1pt)**. Pour quelles valeurs de  $\lambda$ ,  $\tilde{\theta}(\lambda)$  est-il sans biais **(0.5 pts)**? Déterminer  $\text{var}(\tilde{\theta}(\lambda))$  **(1pt)**.  
 (g) On rappelle que  $R(\tilde{\theta}(\lambda)) = \mathbb{E}(\|\tilde{\theta}(\lambda) - \theta\|^2)$  est le risque quadratique de  $\tilde{\theta}(\lambda)$ . Montrer que  $R(\tilde{\theta}(\lambda)) = \|\mathbb{E}(\tilde{\theta}(\lambda)) - \theta\|^2 + \text{Trace}(\text{var}(\tilde{\theta}(\lambda)))$  **(1pt)**. En déduire  $R(\tilde{\theta}(0))$  **(0.5pts)** et plus généralement  $R(\tilde{\theta}(\lambda))$  **(1pt)**.  
 (h) Dans cette question uniquement on suppose  ${}^tX X = n I_{p+1}$ . Montrer que dans ce cas  $R(\tilde{\theta}(0)) = \sigma^2(p + 1)/n$  **(0.5pts)** et  $R(\tilde{\theta}(\lambda)) = (\lambda^2 \|\theta\|^2 + n \sigma^2(p + 1))/(n + \lambda^2)$  **(1pt)**. Montrer qu'il existe toujours une unique valeur de  $\lambda$  notée  $\lambda_0 > 0$  minimisant  $R(\tilde{\theta}(\lambda))$  **(1pt)**. Dans le cas gaussien, en déduire qu'il existe un estimateur de plus petit risque quadratique que l'estimateur par maximum de vraisemblance **(1pt)**. Comment cela est-il possible **(0.5pts)**?  
 (i) On désire choisir  $\lambda$  ayant de bonnes propriétés. Démontrer que  $\hat{Y} = X \tilde{\theta}(\lambda) = H_\lambda Y$ , avec  $H_\lambda$  une matrice à préciser. Démontrer que dans le cas des moindres carrés ordinaires, le nombre de paramètres estimés est  $\text{Trace}(H_0)$  **(1pt)**. Pour  $\lambda > 0$ , on désire calculer  $\text{Trace}(H_\lambda)$ . Montrer que si l'on note  $\lambda_i$  les  $p + 1$  valeurs propres (pas forcément distinctes!) de  ${}^tX X$ , alors  $\text{Trace}(H_\lambda) = \sum_{i=1}^{p+1} \frac{\lambda_i}{\lambda + \lambda_i}$  **(2pts)**. Pour choisir un  $\hat{\lambda}$ , on minimisera en  $\lambda$  un critère BIC, où le nombre de paramètres estimés est remplacé par  $\text{Trace}(H_\lambda)$ . Quelle est sa formule **(0.5pts)**?

## 2. (Sur 8 points) Exercice de TP utilisant le logiciel R

(a) Soit la suite de commandes suivantes:

```
n=100
X1=rnorm(n,-2,1); X2=rnorm(n,mean=X1,sd=0.1)
X3=c(1:n); X4=1/(1+c(1:n)/n); X5=4*cos(c(1:n)/6)
eps=runif(n,-3,3)
Y=7+X1+2*X2+4*log(1+3*X3)-5*X4+eps
reg1=lm(Y~X1+X2+X3+X4+X5)
summary(reg1)
```

Voici le résultat obtenu:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	69.49816	7.52430	9.236	7.73e-15 ***
X1	0.01829	2.13650	0.009	0.993187
X2	3.09389	2.15568	1.435	0.154540
X3	-0.13701	0.03889	-3.523	0.000662 ***
X4	-57.56289	8.03334	-7.165	1.72e-10 ***
X5	-0.03816	0.07005	-0.545	0.587217

Residual standard error: 1.95 on 94 degrees of freedom  
 Multiple R-squared: 0.8822, Adjusted R-squared: 0.8759  
 F-statistic: 140.8 on 5 and 94 DF, p-value: < 2.2e-16

*Questions I.1: Ecrire formellement les différentes variables et le modèle. Réécrire le modèle sans X1. Que pensez vous des résultats obtenus par la régression? Est-ce surprenant? (2.5pts)*

(b) On tape ensuite les commandes:

```
library(MASS)
reg2=stepAIC(lm(Y~X1+X2+X3+X4+X5),k=log(n),direction="both")
summary(reg2); plot(reg2)
```

Voici ce que l'on obtient à la fin des résultats:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	69.18644	7.38600	9.367	3.40e-15 ***
X2	3.10791	0.21278	14.606	< 2e-16 ***
X3	-0.13503	0.03799	-3.554	0.000591 ***
X4	-57.25583	7.89632	-7.251	1.05e-10 ***

Residual standard error: 1.933 on 96 degrees of freedom  
 Multiple R-squared: 0.8818, Adjusted R-squared: 0.8781  
 F-statistic: 238.8 on 3 and 96 DF, p-value: < 2.2e-16

*Questions I.2: Qu'a-t-on fait en tapant ces commandes? Que concluez vous quant aux résultats obtenus? Connaissant le modèle, pensez-vous qu'une transformation de Box-Cox pourrait améliorer les résultats? (1.5pts)*

(c) On tape ensuite les commandes:

```
BX=boxcox(X3~Y+X4,plotit = TRUE,lambda = seq(-3,3))
ind=which(BX$y==max(BX$y))
lambda=BX$x[ind]
lambda
X33=log(X3)
summary(lm(Y~X2+X33+X4))
```

Voici ce que l'on obtient à la fin des résultats:

```
> lambda
[1] 0.3333333
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	10.6914	4.8465	2.206	0.0298 *
X2	3.2813	0.1882	17.440	< 2e-16 ***
X33	4.1036	0.6010	6.828	7.81e-10 ***
X4	-3.5649	4.0079	-0.889	0.3760

Residual standard error: 1.687 on 96 degrees of freedom

Multiple R-squared: 0.91, Adjusted R-squared: 0.9072

F-statistic: 323.5 on 3 and 96 DF, p-value: < 2.2e-16

*Questions I.3: Qu'a-t-on fait en tapant ces commandes? Connaissant le modèle, est-ce surprenant que cette transformation de Box-Cox améliore les résultats? (1.5pts)*

- (d) On effectue maintenant ce qui suit, sachant que la fonction `optim` permet une minimisation par la méthode du gradient conjugué:

```
NLLS=function(theta){sum((Y-theta[1]-theta[2]*X2-theta[3]*log(1+theta[4]*X3)-theta[5]*X4)^2)}
NLE=optim(c(0,0,0,1,0),NLLS,method = "CG")
NLE$par
theta=NLE$par
Res=Y-theta[1]-theta[2]*X2-theta[3]*log(1+theta[4]*X3)-theta[5]*X4
1-sum(Res^2)/sum((Y-mean(Y))^2)
```

Et on obtient:

```
> NLE$par
[1] 1.2714965 3.2877026 4.8831616 2.1319011 0.4787979
> 1-sum(Res^2)/sum((Y-mean(Y))^2)
[1] 0.9092672
```

*Questions I.4: Qu'a-t-on fait en tapant ces commandes? Le minimum obtenu est-il celui auquel on s'attendait? Que concluez vous à partir des résultats obtenus? (2.5pts)*