Exercice 1 - Série 5

On reprend les données de l'exemple vu en classe

$n=20$

Variables dans le modèle	SSE	SSR
$\emptyset$	10	0
X_1	5	5
X_2	9	1
X_3	8	2
X_1, X_2	4	6
X_1, X_3	3.9	6.1
X_2, X_3	8.5	1.5
X_1, X_2, X_3	3.8	6.2

a) Utiliser le test de Fisher pour tester la validité globale du modèle (niveau=95%)

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \beta_3 X_{t3} + \varepsilon_t$$

b) Utiliser le test partiel de Fisher pour tester si $H_0: \beta_1 = \beta_2 = 0$
$H_1: \beta_1 \neq 0$ ou $\beta_2 \neq 0$

dans le modèle $Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \beta_3 X_{t3} + \varepsilon_t$

c) Appliquer les algorithmes "forward" et "stepwise" afin de sélectionner un modèle "optimal"

Solution 1 - Série 5

a) $H_0: \beta_1 = \beta_2 = \beta_3 = 0$

H_1: Au moins un β_i $(i=1,2,3)$ est $\neq 0$.

$$F = \frac{MSR}{MSE} = \frac{SSR/3}{SSE/(20-(3+1))} = \frac{6.2/3}{3.8/16} = 8.702$$

Puisque $F > F_{0.05}(3,16) = 3.24$, on rejette H_0

$\Rightarrow$ La régression est significative au global

b) $H_0: \beta_1 = \beta_2 = 0$; $M_0 \equiv Y_t = \beta_0 + \beta_3 X_{t3} + \varepsilon_t$

$H_1: \beta_1 \neq 0$ ou $\beta_2 \neq 0$; $M_1 \equiv Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \beta_3 X_{t3} + \varepsilon_t$

$$F^* = \frac{(8 - 3.8) / \left([20-(1+1)] - [20-(3+1)]\right)}{3.8 / (20-(3+1))}$$

$$F^* = \frac{4.2/2}{3.8/16} = 8.842$$

Puisque $F^* > F_{0.05}(2,16) = 3.63$, on rejette H_0

$\Rightarrow$ On rejette le modèle "réduit" $(Y_t = \beta_0 + \beta_3 X_{t3} + \varepsilon_t)$

$\Rightarrow$ On ne peut pas exclure "simultanément" les variables X_1 et X_2 du modèle!

c) Forward:

1) Début : $Y_t = \beta_0 + \varepsilon_t$ $\qquad SSE = 10$

2) Meilleure variable à ajouter (= générant la plus grande diminution du SSE) : $X_1 \Rightarrow \Delta = 10-5 = 5$

$X_2 \Rightarrow \Delta = 10-9 = 1$

$X_3 \Rightarrow \Delta = 10-8 = 2$

Choix : X_1

3) $H_0: Y_t = \beta_0 + \varepsilon_t$ ... modèle "réduit"

$H_1: Y_t = \beta_0 + \beta_1 X_{t1} + \varepsilon_t$ --- modèle "augmenté"

$$F^* = \frac{(10-5)/1}{5/(20-(1+1))} = \frac{5/1}{5/18} = 18$$

Puisque $F^* > F_{0.05}(1, 18) = 4.41$, on rejette H_0

$\Rightarrow$ On accepte H_1

$\Rightarrow$ On préfère le modèle $Y_t = \beta_0 + \beta_1 X_{t1} + \varepsilon_t$

$\Rightarrow$ On inclut X_1 dans le modèle !

2) Meilleure variable à ajouter : $X_2 \Rightarrow \Delta = 5-4 = 1$

$X_3 \Rightarrow \Delta = 5-3.9 = 1.1$

Choix $= X_3$

3) $H_0: Y_t = \beta_0 + \beta_1 X_{t1} + \varepsilon_t$

$H_1: Y_t = \beta_0 + \beta_1 X_{t1} + \beta_3 X_{t3} + \varepsilon_t$

$$F^* = \frac{(5-3.9)/1}{3.9/(20-(2+1))} = \frac{1.1/1}{3.9/17} = 4.795$$

Puisque $F^* > F_{0.05}(1,17) = 4.45$, on rejette H_0

$\Rightarrow \ldots \Rightarrow$ on inclut X_3 dans le modèle

2) Meilleure variable à ajouter : X_2 (seul choix!)

3) $H_0: Y_t = \beta_0 + \beta_1 X_{t1} + \beta_3 X_{t3} + \varepsilon_t$

$H_1: Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \beta_3 X_{t3} + \varepsilon_t$

$$F^* = \frac{(3.9-3.8)/1}{3.8/(20-(3+1))} = \frac{0.1/1}{3.8/16} = 0.421$$

Puisque $F^* < F_{0.05}(1,16) = 4.49$; on accepte H_0

$\Rightarrow$ X_2 n'est pas incluse dans le modèle!

$$\boxed{\text{Modèle final: } Y_t = \beta_0 + \beta_1 X_{t1} + \beta_3 X_{t3} + \varepsilon_t}$$

Stepwise:

1) Début: $Y_t = \beta_0 + \varepsilon_t$; $SSE = 10$

2) Meilleure variable à ajouter : X_1

3) $F^* = \dfrac{(10-5)/1}{5/18} = 18 > F_{0.05}(1,18) = 4.41 \Rightarrow X_1$ entre ds modèle

4) Pire variable à exclure : X_1 (un seul choix!)

5) $F^* = \dfrac{(10-5)/1}{5/18} = 18 > F_{0.05}(1,18) = 4.41$ $\Rightarrow$ on rejette le modèle "réduit" de H_0

$\Rightarrow X_1$ non exclut du modèle

2) Meilleure variable à ajouter : X_3

3) $F^* = \dfrac{(5-3.9)/1}{3.9/17} = 4.795 > F_{0.05}(1,17) \Rightarrow X_3$ entre ds modèle

4) Pire variable à exclure dans le modèle $Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t3} + \varepsilon_t$

$X_1 \Rightarrow |\Delta| = |3.9 - 8| = 4.1$

$X_3 \Rightarrow |\Delta| = |3.9 - 5| = 1.1$

Choix: X_3

$$F^* = \frac{(5-3.9)/1}{3.9/17} = 4.795 > F_{0.05}(1,17) \quad \left[\begin{array}{l} H_0: Y_t = \beta_0 + \beta_1 X_{t1} + \varepsilon_t \\ H_1: Y_t = \beta_0 + \beta_1 X_{t1} + \beta_3 X_{t3} + \varepsilon_t \end{array}\right.$$

$\Rightarrow$ On rejette H_0

$\Rightarrow$ On rejette le modèle "simplifié" M_0

$\Rightarrow$ On ne peut pas exclure X_3

$\Rightarrow$ Modèle final : $Y_t = \beta_0 + \beta_1 X_{t1} + \beta_3 X_{t3} + \varepsilon_t$!