Solution série n°2 **Econométrie I**

1. **Mettre le modèle sous forme matricielle en précisant ses dimensions**

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{14} \end{pmatrix} = \begin{pmatrix} 12 \\ 14 \\ \vdots \\ 21 \end{pmatrix} \quad X = \begin{pmatrix} 1 & X_{1\,1t} & X_{1\,2t} & X_{1\,3t} \\ 1 & X_{2\,1t} & X_{2\,2t} & X_{2\,3t} \\ & & \vdots & \\ 1 & X_{14\,1t} & X_{14\,2t} & X_{14\,3t} \end{pmatrix} = \begin{pmatrix} 1 & 2 & 45 & 121 \\ 1 & 1 & 43 & 132 \\ & & \vdots & \\ 1 & 7 & 29 & 180 \end{pmatrix}$$

$$a = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{14} \end{pmatrix}$$

$$Y_{(14\ 1)} = a_{(4\ \ 1)}\, X_{(14\ \ 4)} + \ \varepsilon_{(14\ \ 1)}$$

2. **Le calcul de la matrice inverse $(\boldsymbol{X'X})^{-1}$**

$$X'X = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 2 & 1 & \dots & 7 \\ 45 & 43 & \dots & 29 \\ 121 & 132 & \dots & 180 \end{pmatrix} * \begin{pmatrix} 1 & 2 & 45 & 121 \\ 1 & 1 & 43 & 132 \\ & & \vdots & \\ 1 & 7 & 29 & 180 \end{pmatrix} = \begin{pmatrix} 14 & 85 & 532 & 2094 \\ 85 & 631 & 3126 & 13132 \\ 532 & 3126 & 20666 & 78683 \\ 2094 & 13132 & 78683 & 317950 \end{pmatrix}$$

$$(X'X)^{-1} = \frac{1}{\det(X'X)} * Co(X'X)$$

$$\det(X'X) = 1427069483$$

$$Co(X'X) = \begin{pmatrix} 28782057869 & 21499749 & -330295538 & -108706778 \\ 21499749 & 18844306 & 1704494 & -1341716 \\ -330295538 & 1704494 & 5187478 & 821169 \\ -108706778 & -1341716 & 821169 & 572626 \end{pmatrix}$$

Donc $(X'X)^{-1} = \begin{pmatrix} 20{,}16864 & 0{,}015065 & -0{,}23145 & -0{,}07617 \\ 0{,}015065 & 0{,}013204 & 0{,}001194 & -0{,}00094 \\ -0{,}23145 & 0{,}001194 & 0{,}003635 & 0{,}000575 \\ -0{,}07617 & -0{,}00094 & 0{,}000575 & 0{,}000401 \end{pmatrix}$

3. **Estimation des paramètres du modèle**

$$\hat{a} = (X'X)^{-1}(X'Y)$$

$$(X'Y) = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 2 & 1 & \dots & 7 \\ 45 & 43 & \dots & 29 \\ 121 & 132 & \dots & 180 \end{pmatrix} * \begin{pmatrix} 12 \\ 14 \\ 10 \\ \dots \\ 21 \end{pmatrix} = \begin{pmatrix} 248 \\ 1622 \\ 9202 \\ 37592 \end{pmatrix}$$

$$\hat{a} = (X'X)^{-1}(X'Y) = \begin{pmatrix} 20{,}16864 & 0{,}015065 & -0{,}23145 & -0{,}07617 \\ 0{,}015065 & 0{,}013204 & 0{,}001194 & -0{,}00094 \\ -0{,}23145 & 0{,}001194 & 0{,}003635 & 0{,}000575 \\ -0{,}07617 & -0{,}00094 & 0{,}000575 & 0{,}000401 \end{pmatrix} * \begin{pmatrix} 248 \\ 1622 \\ 9202 \\ 37592 \end{pmatrix}$$

$$= \begin{pmatrix} 32{,}89132 \\ 0{,}8019 \\ -0{,}38136 \\ -0{,}03713 \end{pmatrix}$$

Donc : $\hat{a}_0 = 32{,}89$; $\hat{a}_1 = 0{,}8$; $\hat{a}_2 = -0{,}38$; $\hat{a}_3 = -0{,}03$

$\hat{Y} = 32{,}89 + 0{,}8X_1 - 0{,}38X_2 - 0{,}03X_3$

4. L'estimation de la variance de l'erreur et les écart-types de chacun des coefficients :

D'abord nous calculons la variance de l'erreur :

L'erreur est égale à : $e_t = Y_t - \hat{Y}$

$$e_t = Y_t - (\hat{a}_0 + \hat{a}_1 X_1 + \hat{a}_2 X_2 + \hat{a}_3 X_3)$$

$$e_t = Y_t - (32{,}89 + 0{,}8X_1 - 0{,}38X_2 - 0{,}03X_3)$$

t	Y_t	$\hat{Y}$	e_t	ε_t^2
1	12	12,84	-0,84	0,71
2	14	12,39	1,61	2,58
3	10	13,18	-3,18	10,11
4	16	13,39	1,61	2,58
5	14	17,7	-3,70	13,67
6	19	17,88	1,12	1,26
7	21	22,2	-1,2	1,44
8	19	18,86	0,14	0,02
9	21	16,51	4,49	20,14
10	16	18,76	-2,76	7,63
11	19	17,92	1,08	1,17
12	21	21,9	-0,9	0,81
13	25	22,71	2,29	5,27
14	21	20,76	0,24	0,06
			0	67,45

$$V(\delta_\varepsilon) = \hat{\delta}_\varepsilon^2 = \frac{\sum \varepsilon_t^2}{n-k-1} = \frac{\sum (Y_t - \hat{Y})^2}{n-k-1} = \frac{67{,}45}{10} = 6{,}745$$

Avec : n: nombre d'observations (=14) ; k: nombre de variables (=3)

La matrice variance-covariance est donnée comme suit : $\hat{\Omega}_{\hat{a}} = \hat{\delta}_\varepsilon^2 \, (X'X)^{-1}$

$$\hat{\Omega}_{\hat{a}} = 6{,}745 \begin{pmatrix} 20{,}16864 & 0{,}015065 & -0{,}23145 & -0{,}07617 \\ 0{,}015065 & 0{,}013204 & 0{,}001194 & -0{,}0094 \\ -0{,}23145 & 0{,}001194 & 0{,}003635 & 0{,}000575 \\ -0{,}07617 & -0{,}0094 & 0{,}000575 & 0{,}000401 \end{pmatrix}$$

Les variances des coefficients de régression se trouvent sur la 1ère diagonale

$\hat{\delta}^2_{a0} = 6{,}745 * 20{,}168 = 136{,}03$ donc $\hat{\delta}_{a0} = \sqrt{136{,}03} = 11{,}66$

$\hat{\delta}^2_{a1} = 6{,}745 * 0{,}013 = 0{,}087$ donc $\hat{\delta}_{a1} = \sqrt{0{,}087} = 0{,}29$

$\hat{\delta}^2_{a2} = 6{,}745 * 0{,}0036 = 0{,}024$ donc $\hat{\delta}_{a2} = \sqrt{0{,}024} = 0{,}15$

$\hat{\delta}^2_{a3} = 6{,}745 * 0{,}0004 = 0{,}0026$ donc $\hat{\delta}_{a3} = \sqrt{0{,}0026} = 0{,}05$

5. **Le calcul de coefficient de détermination R^2 et le coefficient de détermination corrigé $\bar{R}^2$:**

$\sum \varepsilon_t^2 = \sum(Y_t - \hat{Y})^2 = SCR = 67{,}45$

$SCT = \sum(Y_t - \bar{Y})^2 = 226{,}86$

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} = 1 - \frac{\sum(Y_t - \hat{Y})^2}{\sum(Y_t - \bar{Y})^2} = 1 - \frac{67{,}45}{226{,}86} = 0{,}702$$

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}(1 - R^2) = 1 - \frac{14}{10}(1 - 0{,}702) = 0{,}613$$

6. **Test de significativité globale du modèle**

Les hypothèses : $H_0: a_0 = a_1 = a_2 = a_3 = 0$ contre $H_1: Il\ existe\ au\ moins\ un\ paramètre\ \neq 0$

$$Fc = \frac{\sum(\hat{Y} - \bar{Y})^2/k}{\sum(Y_t - \hat{Y})^2/_{n-k-1}} = \frac{R^2/_k}{(1 - R^2)/_{n-k-1}} = \frac{0{,}702/_3}{(1 - 0{,}702)/_{10}} = 7{,}878$$

La statistique de Ficher tablé : $F^{5\%}_{(k,n-k-1)} = F^{5\%}_{(3,10)} = 3{,}71$

$Fc > F^{5\%}_{(3,10)}$ nous rejetons l'hypothèse H_0 et nous acceptons l'hypothèse H_1. Le modèle est globalement significatif.

7. **Test de significativité des variables explicatives du modèle :**

- Test de significativité pour a_1 :

Les hypothèses : $H_0: a_1 = 0$ contre $H_1: a_1 \neq 0$

Sous l'hypothèse H_0 :

$Tc = \frac{|a_1 - 0|}{\delta_{a_1}} = \frac{|0{,}8 - 0|}{0{,}29} = 2{,}75$

La statistique de student tablé $t^{\alpha/2}_{n-k-1} = t^{0{,}05/2}_{10} = 2{,}228$

Donc $Tc > t_{10}^{0,05/2}$ nous rejetons l'hypothèse H_0 et nous acceptons l'hypothèse H_1. Le paramètre a_1 est significativement différent de 0. Donc la variable X_1 contribue significativement à l'explication de la variable endogène (Y).

- Test de significativité pour a_2 :

Les hypothèses : $H_0: a_2 = 0$ contre $H_1: a_2 \neq 0$

Sous l'hypothèse H_0 :

$Tc = \frac{|a_2 - 0|}{\delta_{a_2}} = \frac{|-0,38-0|}{0,15} = 2,53$

La statistique de student tablé $t_{n-k-1}^{\alpha/2} = t_{10}^{0,05/2} = 2,228$

Donc $Tc > t_{10}^{0,05/2}$ nous rejetons l'hypothèse H_0 et nous acceptons l'hypothèse H_1. Le paramètre a_2 est significativement différent de 0. Donc la variable X_2 contribue significativement à l'explication de la variable endogène (Y).

- Test de significativité pour a_3 :

Les hypothèses : $H_0: a_3 = 0$ contre $H_1: a_3 \neq 0$

Sous l'hypothèse H_0 :

$Tc = \frac{|a_3 - 0|}{\delta_{a_3}} = \frac{|-0,03-0|}{0,05} = 0,6$

La statistique de student tablé $t_{n-k-1}^{\alpha/2} = t_{10}^{0,05/2} = 2,228$

Donc $Tc < t_{10}^{0,05/2}$ nous acceptons l'hypothèse H_0. Le paramètre a_3 n'est pas significativement différent de 0. Donc la variable X_3 ne contribue pas à l'explication de la variable endogène (Y).

8. L'intervalle de confiance pour chaque paramètre

$IC_{\widehat{ai}}: [\widehat{a_i} - \left(\delta_{ai} * t_{n-k-1}^{\frac{\alpha}{2}}\right); \widehat{a_i} + \left(\delta_{a_i} * t_{n-k-1}^{\frac{\alpha}{2}}\right)]$

$IC_{\widehat{a_0}}: [\widehat{a_0} - \left(\delta_{a_0} * t_{n-k-1}^{\frac{\alpha}{2}}\right); \widehat{a_0} + \left(\delta_{a_0} * t_{n-k-1}^{\frac{\alpha}{2}}\right)]$

$IC_{\widehat{a_0}}: [32,89 - (11,66 * 2,228); 32,89 + (11,66 * 2,228)]$

$IC_{\widehat{a_0}}: [6,911;\ \ 58,868]$

$0 \notin IC_{\widehat{a_0}}$

$IC_{\widehat{a_1}}: [\widehat{a_1} - \left(\delta_{a_1} * t_{n-k-1}^{\frac{\alpha}{2}}\right); \widehat{a_1} + \left(\delta_{a_1} * t_{n-k-1}^{\frac{\alpha}{2}}\right)]$

$IC_{\widehat{a_1}}: [0,8 - (0,29 * 2,228); 0,8 + (0,29 * 2,228)]$

$IC_{\widehat{a_1}}$: [0,153; 1,446]

$0 \notin IC_{\widehat{a_1}}$

$IC_{\widehat{a_2}}: [\widehat{a_2} - \left(\delta_{a_2} * t_{n-k-1}^{\frac{\alpha}{2}}\right); \widehat{a_2} + \left(\delta_{a_2} * t_{n-k-1}^{\frac{\alpha}{2}}\right)]$

$IC_{\widehat{a_2}}: [-0{,}38 - (0{,}15 * 2{,}228); -0{,}38 + (0{,}15 * 2{,}228)]$

$IC_{\widehat{a_2}}: [-0{,}71;\ -0{,}04]$

$0 \notin IC_{\widehat{a_2}}$

$IC_{\widehat{a_3}}: [\widehat{a_3} - \left(\delta_{a_3} * t_{n-k-1}^{\frac{\alpha}{2}}\right); \widehat{a_3} + \left(\delta_{a_3} * t_{n-k-1}^{\frac{\alpha}{2}}\right)]$

$IC_{\widehat{a_3}}: [-0{,}03 - (0{,}005 * 2{,}228); -0{,}03 + (0{,}005 * 2{,}228)]$

$IC_{\widehat{a_3}}: [-0{,}14;\ 0{,}08]$

$0 \in IC_{\widehat{a_3}}$

Les résultats se concordent avec le test de significativité des paramètres (Question 7).

9. Le calcul de la prévision et son intervalle de confiance au seuil de signification de 95% (α=0,05)

Nous ne pouvons pas faire la prévision avec le modèle estimé précédemment car le paramètre $\widehat{a_3}$ n'est pas significatif donc la variable X_3ne contribue pas à l'explication de la variable endogène Y. Donc pour calculer la prévision il est nécessaire de ré-estimer le modèle avec les deux variables significatives uniquement X_1 et X_2

Le nouveau modèle à ré-estimer s'écrit : $\hat{Y} = \widehat{a_0} + \widehat{a_1}X_1 + \widehat{a_2}X_2$

Les résultats de l'estimation sont donnés comme suit :

$$\hat{a} = (X'X)^{-1}(X'Y) = \begin{pmatrix} 25{,}84213 \\ 0{,}714895 \\ -0{,}32811 \end{pmatrix}$$

Donc : $\hat{a}_0 = 25{,}84$; $\hat{a}_1 = 0{,}714$; $\hat{a}_2 = -0{,}328$;

$\hat{Y} = 25{,}84 + 0{,}714X_1 - 0{,}328X_2$

$R^2 = 0{,}687$; $\hat{\delta}_\varepsilon^2 = 6{,}44$

Le test de significativité globale effectué sur ce nouveau modèle (de la même manière que la question 6) montre $Fc = 12{,}102 > F_{(3,11)}^{5\%} = 3{,}98$ Le modèle est globalement significative.

Les tests de significativité partiels sur les paramètres du modèle (effectué de la même manière que la question 7) montrent $Tc_{a1} = 2{,}684$ et $Tc_{a2} = 2{,}438$ supérieur à la statistique tablé $t_{11}^{5\%} = 2{,}201$ donc les paramètres estimés a_1 et a_2 sont significativement différent de 0.

Nous pouvons donc effectuer la prévision pour la 15ème période à partir de ce nouveau modèle estimé.

$\hat{Y} = 25{,}84 + 0{,}714X_1 - 0{,}328X_2$

$\widehat{Y_{15}} = 25{,}84 + 0{,}714\ (3) - 0{,}328(24) = 20{,}11$

L'intervalle de confiance pour la prévision est donné comme suit :

$IC_{\widehat{Y_{15}}}: [\widehat{Y_{15}} - \left(\hat{\delta}_{e_{15}} * t_{n-k-1}^{\frac{\alpha}{2}}\right); \widehat{Y_{15}} + \left(\hat{\delta}_{e_{15}} * t_{n-k-1}^{\frac{\alpha}{2}}\right)]$

Avec $\hat{\delta}_{e_{15}} = \sqrt{\hat{\delta}_{\varepsilon}^2[V'(X'X)^{-1}V_{15} + 1]} \quad ; V = \begin{pmatrix} 1 \\ 3 \\ 24 \end{pmatrix}$

Donc $\hat{\delta}_{e_{15}}{}^2 = 6{,}44 \left[(1\ 3\ 24) \begin{pmatrix} 5{,}707687 & -0{,}16341 & -0{,}12221 \\ -0{,}16341 & 0{,}11001 & 0{,}002541 \\ -0{,}12221 & 0{,}002541 & 0{,}002809 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ 24 \end{pmatrix} + 1 \right] = 12{,}53$

$\hat{\delta}_{e_{15}} = \sqrt{12{,}53} = 3{,}54$

$IC_{\widehat{Y_{15}}}: [20{,}25 - (3{,}54 * 2{,}201);\ \ 20{,}25 + (3{,}54 * 2{,}201)]$

$IC_{\widehat{Y_{15}}}: [12{,}26;\ \ 27{,}84]$