CHAPITRE II : MODELE LINEAIRE DE REGRESSION MULTIPLE (M.R.M)

Les M.R.M. sont du type :

$$Y_t = a_1 X_{1t} + a_2 X_{2t} + \ldots + a_p X_{pt} + U_t$$

$Y_{t:}$ Variable endogène, aléatoire à cause de l'introduction de U_t

X_{1t} … X_{pt} Sont les observations à chaque période t des variables exogènes

a_1X_{1t}… a_pX_{pt} Est la partie déterministe ou systématique ou explicative du modèle.

U_t est la partie aléatoire du modèle.

HYPOTHESES DANS LE M.R.M :

Hypothèse 1 :

<u>Le modèle est correctement spécifié</u>

Autrement dit, Il faut que les variables explicatives retenues soient les «meilleures» sans omission d'autres variables, la vraie relation soit une relation linéaire dans ou par rapport aux paramètres à estimer et enfin la variable aléatoire intervienne de manière additive.

Hypothèse 2 :

<u>Les</u> Y_t *<u>et les</u>* X_{it} *<u>sont des grandeurs numériques observées sans erreur.</u>*

E(Ut) = 0 quelque soient Xit.

Pour i = 1…p

Hypothèse 3 :

Hypothèse d'homoscédasticité

Ut est distribuée selon une loi indépendante de t et des Xit , pour t=1,…n et i = 1,….p

$$V(U_t) = E(U_t^2) = \delta_u^2$$

est une quantité finie

Hypothèse 4 :

Indépendance des erreurs

$$Cov(Ut, U_{t'}) = 0$$

Hypothèse 5 :

La loi de distribution de l'aléa est une loi gaussienne de moyenne nulle et l'écart-type fini.

Hypothèse 6 :

Hypothèse sur les variables exogènes:

Absence de colinéarité des variables $X_{1t} \ldots X_{pt}$ et E (vecteur unité).

Hypothèse 7 :

On n'introduit pas de restriction sur les estimateurs.
Ils peuvent être positifs, négatifs ou nuls.

DETERMINATION ET PROPRIETES DES ESTIMATEURS :

Détermination du vecteur des estimateurs

- **<u>1°) calcul de â par les M.C.O :</u>**
- La méthode consiste à chercher les paramètres $\hat{a}_i$ tels que :

$$\sum_{i=1}^{n} (\hat{u}_t)^2$$

soit minimum.

- Revenons à hypothèse 4, concernant $\hat{U}$

$$X_1'\hat{U} = X_2'\hat{U} = \ldots = X_j'\hat{U} = \ldots = X_{p-1}'\hat{U} = E'\hat{U} = 0$$

X’ est le transposé de X

$$X'.(\overbrace{Y - X\hat{a}}^{\hat{U}}) = 0$$

$$(X'Y) - (X'X.\hat{a}) = 0$$

$$X'Y = X'.X\hat{a}$$

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

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

Propriétés de â

$$\hat{a} = (X'X)^{-1}X'(Xa+U)$$

$$\hat{a} = (X'X)^{-1}(X'Xa+X'U)$$

$$\hat{a} = (X'X)^{-1}(X'X).a+(X'X).X'U$$

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

$$E(\hat{a}) = E\left[a+(X'X)^{-1}X'U\right]$$

$$E(\hat{a}) = E(a)+(X'X)^{-1}X'E(U) \quad ; \quad \text{E(U) étant egal à 0}$$

$$E(\hat{a}) = a \quad \text{donc a est un estimateur sans biais de a.}$$

$$V(\hat{a}) = \delta_u^2 (X'X)^{-1}$$

On montre qu'une condition nécessaire et suffisante pour que â soit un estimateur convergent de a est que les vecteurs variables exogènes ne tendent pas à être colinéaires quand n tend vers l'infini.

Autrement dit H6 reste valable quand n tend vers l'infini

â est BLUE si et seulement si:
(Best Linear Unbiaised Estimator)
Meilleur estimateur linéaire sans biais

$E(\hat{a})=a \quad \forall n$ **Lorsque n tend vers l'infini**

$V(\hat{a})$ **tend vers 0 quand n tend vers l'infini**

Détermination de S^2 estimateur de δ_u^2

On sait que: $$S^2 = \frac{\sum \hat{u}_t^2}{n-p}$$ **Avec** $$\sum \hat{u}_t^2 = \hat{u}'\hat{u}$$

S^2 est un **estimateur** sans biais de δ_u^2

p est le nombre de paramètres à estimer
n est le nombre d'observations

LE COEFFICIENT DE CORRELATION MULTIPLE

$$R^2 = \frac{\sum_{t=1}^{n}(Y_t - \overline{Y})^2 - \sum_{t=1}^{n}(Y_t - X_t \hat{a})^2}{\sum_{t=1}^{n}(Y_t - \overline{Y})^2}$$

Tests d’hypothèses dans le M.R.M

1/ Le test de Fisher (F)

û'û suit une loi khi-deux (n-p)

Si a= 0 ^y'^y suit la loi khi-deux p -→Ces deux khi-deux sont indépendants en probabilité.

Or, on sait que le rapport de 2 khi-deux est un test de FISHER.
Donc,
la variable F qui suit une loi de Fisher est:

$$F=\frac{\frac{\hat{Y}'\hat{Y}}{p}}{\frac{\hat{u}'\hat{u}}{n-p}}=\frac{\hat{Y}'\hat{Y}}{p}\times\frac{n-p}{\hat{u}'\hat{u}}$$

tend vers un $F_{p;n-p}$ ddl

-***Solution du Test :***

On compare le F théorique càd le F lu sur la table de distribution de Fisher et le F calculé à partir de nos observations:

2 cas peuvent se présenter :

*** F calculé > F théorique ; on rejette Ho. Cela veut dire que les variables retenues sont explicatives.**

*** F calculé < F théorique ; non rejet de Ho .**

***Exemple* : Soit un modèle à 5 variables exogènes (y compris la constante) et suppose qu'il a été estimé sur 25 observations. Supposons enfin que le résultat obtenu est :**

$$\frac{\hat{Y}'\hat{Y}}{5} \times \frac{20}{\hat{u}'\hat{u}} = 3.0$$

La valeur lue sur la table du F de Fisher pour 5 et 20 d.d.l

Et 2,71 au seuil de 5% et 4,10 au seuil de 1%

Nous allons rejeter au seuil de 5% (car 3>2,71) mais on ne pourra pas la rejeter au seuil de 1% (car 3<4,10)..

2/Test Student

suit une loi de STUDENT à (n-p) ddl

Solution du test :

$$P\left[-t_{\alpha}.\hat{\delta} a_k < \hat{a}_k < +t_{\alpha}.\hat{\delta} a_k\right] = 0.95$$

2 cas possibles :

* appartient à l'intervalle alors on ne refuse pas H0

$$\left[-t_{\alpha} . \hat{\delta} a_k \, , +t_{\alpha} . \hat{\delta} a_k \right]$$

* n'appartient pas à cet intervalle alors on refuse H0

EXEMPLE

$$M_t = e + aI_t + bC_t + U_t$$

pour n = 22, on a

$$M_t = 15.8 + 3.07 I_t + 1.92 C_t$$
$$(10.53) \quad (0.81) \quad (1.25)$$

$$T = \frac{15.8}{10.53}$$

suit une loi de Student à 19 ddl,
le t lu sur la table à 5% est 2,093.
Tc = 1,5< TT donc non rejet de H_0 .
!!!!!(n'est pas important)

$$T = \frac{3.07}{0.81} > 2$$

donc le coefficient est différent de 0 l'investissement explique le niveau des importations.

$$T = \frac{1.92}{1.25} < 2$$

donc selon l'échantillon, la consommation n'explique pas le niveau des importations.

<u>3</u>°) Test de DURBIN & WATSON : Problème de l’autocorrélation des résidus

Il y a autocorrélation des résidus quand le modèle est mal spécifié

Dans le cas de l'omission d'une variable significative ou l'inclusion d'une variable non significative, les résidus ne sont plus distribués de façon aléatoire autour du Y mais reflètent le comportement des variables omis ou inclus.

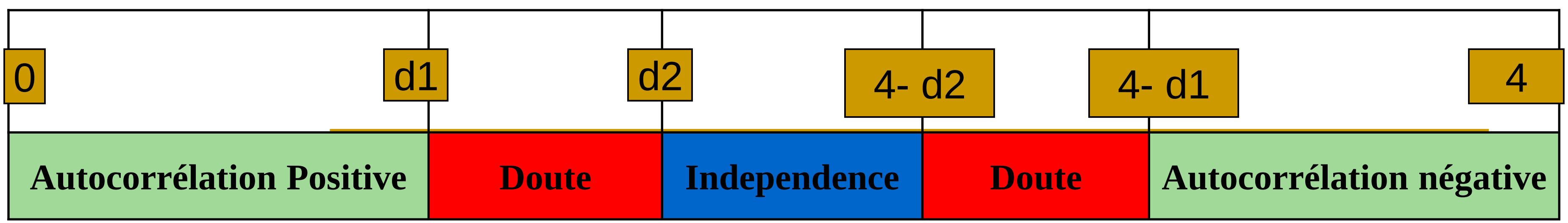
Pour s'assurer du risque d'autocorrélation des erreurs, on procède au test de Durbin & Watson.

$$\hat{d} = \frac{\sum (\hat{u}_t - \hat{u}_{t-1})^2}{\sum \hat{u}_t^2}$$

on peut déduire

$$\hat{d} = 2 - 2\frac{\sum (\hat{u}_t \hat{u}_{t-1})}{\sum \hat{u}_t^2}$$

Tableau de décisions



.

Durbin & Watson ont déterminé 2 valeurs d_1 et d_2 fonction de n et de p.
Donc on calcule ^d à partir de sa formule et donner résultats de l'estimation ; ensuite, nous serons face à plusieurs éventualités :

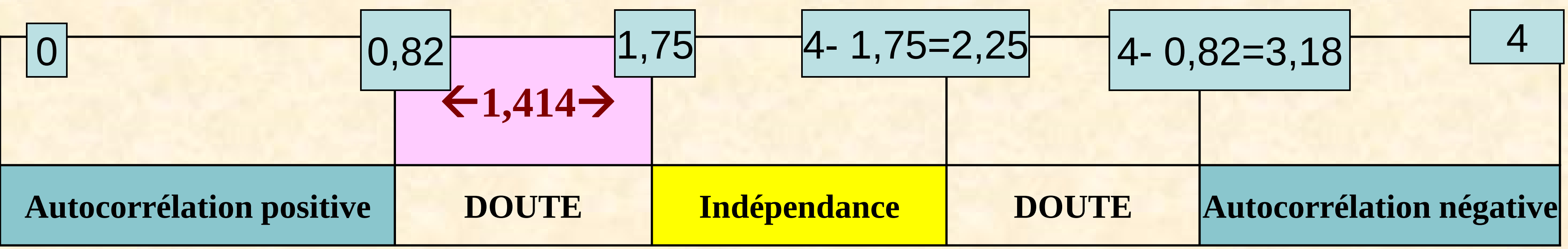
- **$\hat{d} < d_1$ on rejette H0**
- **$d1 < \hat{d} < d_2$ doute**
- **$\hat{d} > d_2$ non rejet H0 , absence d'autocorrélation des résidus.**

<u>*Exemple :*</u>

Log Mt = 0,469 Log Ct + 0,034 Log FBCFt + 0,471 Log Xt – 0,993

^DW = 1,414

n = 15 et p = 3 d_1=0,82 et d_2=1,75



donc 0,82 < ^DW < 1,75 il y a doute

APPLICATION

MODELES MULTIPLES

Ex1

- **Soit le modèle suivant concernant le cas d'une banque. On veut étudier l'impact trois variables (dépenses publicitaires X1, les crédits X2 et les dépôts X3) sur le résultat net de cette banque (Y).** (Y, X1, X2 et X3 sont en million de dh)
- Le tableau ci-dessous regroupe les variables à étudier:

Année	Y	X1	X2	X3
1989	12	2	45	121
1990	14	1	43	132
1991	10	3	43	154
1992	16	6	47	145
1993	14	7	42	129
1994	19	8	41	156
1995	21	8	32	132
1996	19	5	33	147
1997	21	5	41	128
1998	16	8	38	163
1999	19	4	32	161
2000	21	9	31	172
2001	25	12	35	174
2002	21	7	29	180

$$Y_t = a_o + a_1 X_{1t} + a_2 X_{2t} + a_3 X_{3t} + U_t$$

- **Mettre le modèle sous sa forme matricielle.**
- **Estimer les paramètres du modèle**
- **Calculer le R^2**

Les données de base sous forme matricielle sont:

$$
Y = \begin{pmatrix} 12 \\ 14 \\ 10 \\ 16 \\ 14 \\ 19 \\ 21 \\ 19 \\ 21 \\ 16 \\ 19 \\ 21 \\ 25 \\ 21 \end{pmatrix}
\xrightarrow{et}
X = \begin{pmatrix}
1 & 2 & 45 & 121 \\
1 & 1 & 43 & 132 \\
1 & 3 & 43 & 154 \\
1 & 6 & 47 & 145 \\
1 & 7 & 42 & 129 \\
1 & 8 & 41 & 156 \\
1 & 8 & 32 & 132 \\
1 & 5 & 33 & 147 \\
1 & 5 & 41 & 128 \\
1 & 8 & 38 & 163 \\
1 & 4 & 32 & 161 \\
1 & 9 & 31 & 172 \\
1 & 12 & 35 & 174 \\
1 & 7 & 29 & 180
\end{pmatrix}
\xrightarrow{et}
A = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix}
\xrightarrow{et}
U = \begin{pmatrix} U1 \\ U2 \\ U3 \\ U4 \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ U14 \end{pmatrix}
$$

On sait que â= (X'X)$^{-1}$X'Y

$$X'X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 1 & 3 & & & & & & & & & & & 7 \\ 45 & 43 & 43 & & & & & & & & & & & 29 \\ 121 & 132 & 154 & & & & & & & & & & & 180 \end{pmatrix} x \begin{pmatrix} 1 & 2 & 45 & 121 \\ 1 & 1 & 43 & 132 \\ 1 & 3 & 43 & 154 \\ 1 & & & \\ 1 & & & \\ 1 & & & \\ 1 & & & \\ 1 & & & \\ 1 & & & \\ 1 & & & \\ 1 & & & \\ 1 & & & \\ 1 & & & \\ 1 & 7 & 29 & 180 \end{pmatrix}$$

Donc X'X est égal

$$\begin{pmatrix} 14 & 85 & 532 & 2094 \\ 85 & 631 & 3126 & 13132 \\ 532 & 3126 & 20666 & 78683 \\ 2094 & 13132 & 78683 & 317950 \end{pmatrix}$$

$(X'X)^{-1}$ est :

$$\begin{pmatrix} 20.16864 & 0.015065 & -0.23145 & -0.07617 \\ 0.015065 & 0.013204 & 0.001194 & -0.00094 \\ -0.23145 & 0.001194 & 0.003635 & 0.000575 \\ -0.07617 & -0.00094 & 0.000575 & 0.000401 \end{pmatrix}$$

X’Y est égale à

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 1 & 3 & & & & & & & & & & & 7 \\ 45 & 43 & 43 & & & & & & & & & & & 29 \\ 121 & 132 & 154 & & & & & & & & & & & 180 \end{pmatrix} * \begin{pmatrix} 12 \\ 14 \\ 10 \\ 16 \\ 14 \\ 19 \\ 21 \\ 19 \\ 21 \\ 16 \\ 19 \\ 21 \\ 25 \\ 21 \end{pmatrix} = \begin{pmatrix} 248 \\ 1622 \\ 9202 \\ 37592 \end{pmatrix}$$

Â est don est égal à

$$(X'X)^{-1}X'Y = \begin{pmatrix} 32,89132 \\ 0,8019 \\ -0,38136 \\ -0,03713 \end{pmatrix}$$

$\hat{a}_0 = 32{,}89, \hat{a}_1 = 0{,}80,\ \hat{a}_2 = -0{,}38$ et $\hat{a}_3 = -0{,}03$

- Donc

$$Y_t = 32{,}89 + 0{,}8X_{1t} - 0{,}38X_{2t} - 0{,}03X_{3t} + U_t$$

Sachant que

$$R^2 = 1 - \frac{\sum e_t^2}{\sum (y_t - \bar{y})^2}$$

Calcule de R^2

$$1 = Année, 2 = Y_t, 3 = \hat{Y}_t, 4 = e_t,$$

$$5 = e_t^2, 6 = y_t - \bar{y}_t \; et \; 7 = \sum (y_t - \bar{y}_t)^2$$

1	2	3	4	5	6	7
1989	12	12,84	-0,84	0,7056	-5,7142857	32,653
1990	14	12,39	1,61	2,5921	-3,7142857	13,796
1991	10	13,18	-3,18	10,1124	-7,7142857	59,51
1992	16	14,39	1,61	2,5921	-1,7142857	2,9388
1993	14	17,7	-3,7	13,69	-3,7142857	13,796
1994	19	17,88	1,12	1,2544	1,28571429	1,6531
1995	21	22,2	-1,2	1,44	3,28571429	10,796
1996	19	18,86	0,14	0,0196	1,28571429	1,6531
1997	21	16,51	4,49	20,1601	3,28571429	10,796
1998	16	18,76	-2,76	7,6176	-1,7142857	2,9388
1999	19	17,92	1,08	1,1664	1,28571429	1,6531
2000	21	21,9	-0,9	0,81	3,28571429	10,796
2001	25	22,71	2,29	5,2441	7,28571429	53,082
2002	21	20,76	0,24	0,0576	3,28571429	10,796
Moyenne	**17,71**					
Somme	**248**	**248**	**0**	**67,46**		**226.86**

- Donc

R^2= 1-(67,45/226,86) = 0,702

- **1/Calculer l’estimation de la variance de l’erreur ainsi que les écarts types de chacun des coefficients.**

- **2/Les variables exogènes sont-elles significativement contributives pour expliquer la variable endogène**

- **1/On sait que:**

$$\hat{\sigma}_{\varepsilon}^{2} = \frac{\sum_{t=1}^{t=14} e_t^2}{n-k-1} = \frac{67,45}{14-3-1} = \frac{67,45}{10} = 6,745$$

La matrice des variances et covariances estimées des coefficients nous est donnée par:

$$\hat{\Omega}_{\hat{a}} = \hat{\sigma}_{\varepsilon}^{2}(X'X)^{-1}$$

$$\hat{\Omega}_{\hat{a}} = \hat{\sigma}_{\varepsilon}^{2}(X'X)^{-1}$$

$$= 6,745x\begin{pmatrix} 20,16 & 0,015 & -0,2314 & -0,076 \\ 0,015 & 0,0132 & 0,0012 & -0,00094 \\ -0,2314 & 0,0012 & 0,0036 & 0,00057 \\ -0,076 & -0,00094 & -0,0006 & 0,0004 \end{pmatrix}$$

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

$$\hat{\sigma}^2_{\hat{a}0} = 6,745x20,17 = 136,04 \xrightarrow{alors} \hat{\sigma}_{\hat{a}0} = 11,66$$

$$\hat{\sigma}^2_{\hat{a}1} = 6,745x0,013 = 0,084 \xrightarrow{alors} \hat{\sigma}_{\hat{a}1} = 0,29$$

$$\hat{\sigma}^2_{\hat{a}2} = 6,745x0,0036 = 0,024 \xrightarrow{alors} \hat{\sigma}_{\hat{a}2} = 0,15$$

$$\hat{\sigma}^2_{\hat{a}3} = 6,745x0,0004 = 0,0026 \xrightarrow{alors} \hat{\sigma}_{\hat{a}3} = 0,05$$

2/ Pour cela nous devons calculer

$$\left| \frac{\hat{a}_i}{\hat{\sigma}_{\hat{a}_i}} \right| = t_{\hat{a}\,i}$$

- **Pour $\hat{a}_1$, $t_{\hat{a}1}$= 0,80/0,29= 2,75>2**

X1 est pertinente

- **Pour $\hat{a}_2$, $t_{\hat{a}2}$= va(-0,38/0,15)= 2,53>2**

X2 est pertinente.

- **Pour $\hat{a}_3$, $t_{\hat{a}3}$= va(-0,03/0,05)=0,60<2→ X3 n’est pas contributive à l’explication de Y**