

# **ECONOMETRIE (\*)**

**Hélène Hamisultane**

## **I/ QU'EST CE QUE L'ECONOMETRIE ?**

## **II/ LE MODELE DE REGRESSION SIMPLE**

II.1/ Méthode d'estimation des Moindres Carrés Ordinaires (MCO)

II.2/ Hypothèses et propriétés des estimateurs des MCO

II.3/ Critère de jugement de la qualité de l'ajustement d'un modèle :  $R^2$

## **III/ LE MODELE DE REGRESSION MULTIPLE**

III.1/ Méthode d'estimation des Moindres Carrés Ordinaires (MCO)

III.2/ Hypothèses et propriétés des estimateurs des MCO

III.3/ Critère de jugement de la qualité de l'ajustement d'un modèle :  $R^2$ ,  $R_c^2$ ,  $s$

III.4/ Utilisation de variables indicatrices pour la correction des valeurs anormales et détection des valeurs anormales.

III.5/ Prévision

## **IV/ LES TESTS**

IV.1/ Test de significativité d'un coefficient : test de student

IV.2/ Test de significativité global : test de Fisher

IV.3/ Test de normalité des erreurs

IV.4/ Tests d'autocorrélation : Durbin-Watson et Box-Pierce

IV.5/ Test d'hétéroscédasticité : test de White

IV.6/ Test de stabilité : test de Chow

IV.7/ Test de colinéarité : test de Belsley Khu Welsh

## **V/ VIOLATION DES HYPOTHESES**

V.1/ Méthode des Moindres Carrés Généralisés (MCG)

V.2/ Autocorrélation des erreurs d'ordre 1 et MCG : méthode de Cochrane-Orcutt

V.3/ Hétéroscédasticité et MCG

---

(\*) Document inspiré de l'ouvrage de Bourbonnais (2000), Econométrie, Dunod.

## VI/ LES MODELES DYNAMIQUES

VI.1/ Modèle autorégressif : critères de Akaike, Schwarz et « h » de Durbin

VI.2/ Modèle autorégressif à retards échelonnés

### BIBLIOGRAPHIE :

- Bourbonnais R. (2000), *Econométrie*, DUNOD.
- Johnston J. et Dinardo J. (1999), *Méthodes Econométriques*, Economica.

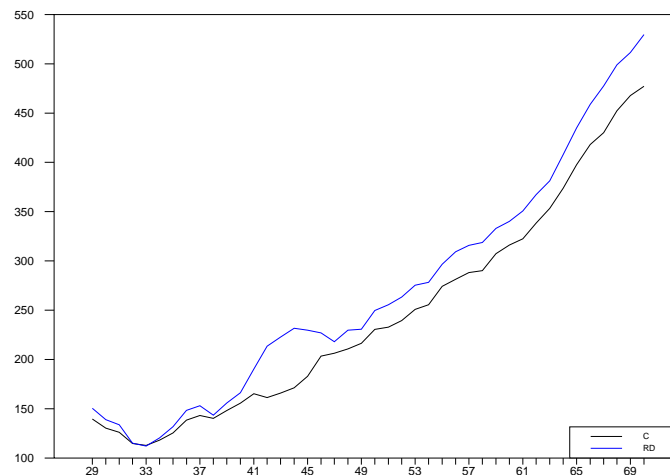
\* \* \*

## I/ QU'EST CE QUE L'ECONOMETRIE ?

La démarche économétrique consiste à représenter à l'aide d'équations le comportement d'un phénomène observé et à estimer les coefficients des équations en recourant à l'historique du phénomène et ceci dans le but de le comprendre, de l'expliquer, de le reproduire et de le prévoir.

Admettons que nous constatons le fait économique suivant :

**Figure 1 :** Revenu disponible et Consommation des ménages au cours du temps

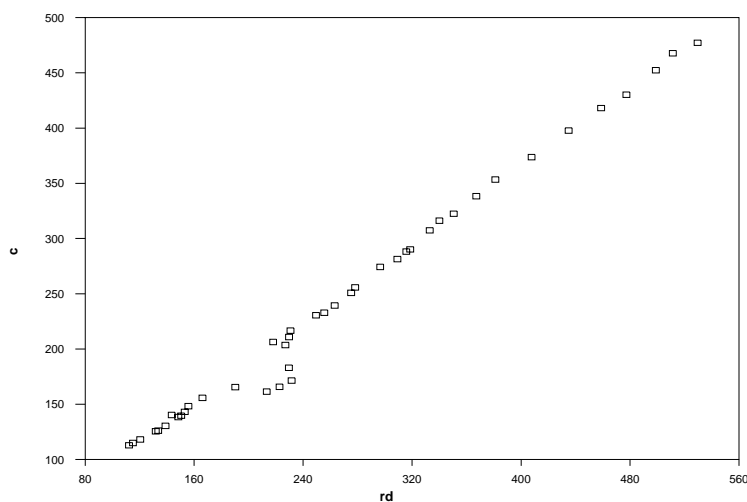


On observe que les 2 courbes évoluent pratiquement dans le même sens : elles augmentent et diminuent simultanément.

On peut penser qu'il y a un lien entre ces 2 variables. On peut en effet penser que la consommation C des ménages est influencée par le revenu disponible RD. Lorsque le revenu augmente, la consommation s'accroît.

En mettant en abscisse le revenu disponible et en ordonnée la consommation des ménages, on obtient le graphique suivant :

**Figure 2 :** Consommation des ménages en fonction du revenu disponible



On s'aperçoit que les points forment une droite. On peut supposer qu'elle a pour équation :

$$C_t = a_1 RD_t + a_0$$

où  $C_t$  et  $RD_t$  désignent respectivement la consommation et le revenu disponible à l'instant  $t$ .

A partir de cette droite (ou dit modèle<sup>(1)</sup>), des données recueillies sur la consommation et le revenu disponible des ménages au fil du temps et de la théorie économétrique que nous présenterons ci-après, on peut déterminer la valeur des paramètres  $a_1$  et  $a_0$ . La connaissance de ces valeurs nous permettra d'une part de mesurer l'influence de la variable explicative ( $RD_t$ ) sur la variable à expliquer ( $C_t$ ) et d'autre part de prévoir l'évolution de la variable endogène. En connaissant l'évolution future de la consommation des ménages, une entreprise peut par exemple envisager d'augmenter ou non sa production.

## II/ LE MODELE DE REGRESSION SIMPLE

Soit le modèle suivant :

$$y_t = a_1 x_t + a_0 .$$

---

<sup>(1)</sup> Il s'agit ici d'un modèle en série temporelle dans lequel les variables évoluent au cours du temps. Il existe aussi les modèles en coupe instantanée dans lesquels les variables représentent des phénomènes observés au même instant.

On parle de modèle de régression simple car le modèle ne comporte qu'une seule variable explicative qui est  $x_t$ . Lorsque le modèle comporte plusieurs variables explicatives, on parlera de modèle de régression multiple.

On cherche à estimer les coefficients  $a_1$  et  $a_0$  de cette droite dans le but de reproduire le phénomène économique observé.

On n'étudiera que l'estimation des modèles linéaires (les droites) à une ou plusieurs variables. Il existe des modèles non linéaires (à seuil<sup>(2)</sup> par exemple) dont l'étude ne sera pas abordés ici.

### Notations :

Le modèle à estimer s'écrit :

$$y_t = a_1 x_t + a_0 + \varepsilon_t$$

avec par exemple  $t = 1980, 1981, \dots, 2004$  (qui peut être remplacé par un nombre :  $t = 1, 2, \dots, T$ )

où  $t$  est la date à laquelle on observe la valeur de  $y_t$  et de  $x_t$  et  $\varepsilon_t$  est une variable aléatoire représentant l'erreur de spécification dont les caractéristiques seront précisées au cours de l'énoncé des hypothèses du modèle. On introduit la variable  $\varepsilon_t$  pour marquer le fait que toute modélisation d'un phénomène ne peut pas être parfaite.

Une fois que les coefficients sont estimés, le modèle va s'écrire :

$$\hat{y}_t = \hat{a}_1 x_t + \hat{a}_0$$

ou encore

$$y_t = \hat{a}_1 x_t + \hat{a}_0 + e_t$$

où  $\hat{a}_1$  et  $\hat{a}_0$  désignent les valeurs estimées des paramètres  $a_1$  et  $a_0$ ,  $e_t = y_t - \hat{y}_t$  est appelé le résidu du modèle.  $e_t$  est l'estimateur de l'erreur  $\varepsilon_t$  que l'on ne connaît pas.

## **II.1/ Méthode d'estimation des Moindres Carrés Ordinaires (MCO)**

Comment estimer  $a_1$  et  $a_0$  pour reproduire au mieux le phénomène économique observé ?

La technique des Moindres Carrés Ordinaire (MCO) apporte une réponse au problème posé.

On doit estimer  $a_1$  et  $a_0$  de façon à minimiser la distance au carré entre chaque point observé  $y_t$  et chaque point  $\hat{y}_t$  donné par la droite  $\hat{y}_t = \hat{a}_1 x_t + \hat{a}_0$ .

---

<sup>(2)</sup> Se reporter à l'ouvrage de Lardic et Mignon (2002), Econométrie des Séries Temporelles Macroéconomiques et Financières, Economica.

Soit  $e = y_t - \hat{y}_t$  l'écart entre ces deux mesures, la méthode ou technique des MCO consiste à rechercher les valeurs de  $a_0$  et  $a_1$  de façon à minimiser la quantité suivante :

$$\text{Min} \sum_{t=1}^T e_t^2 = \text{min} \sum_{t=1}^T (y_t - \hat{y}_t)^2 = \text{min} \sum_{t=1}^T (y_t - \hat{a}_1 x_t - \hat{a}_0)^2 = \text{min} S$$

où  $T$  désigne le nombre d'observations pour les variables  $y_t$  et  $x_t$ .

Les conditions nécessaires du 1<sup>er</sup> ordre pour obtenir un optimum pour  $S$  sont :

- $\frac{\partial S}{\partial \hat{a}_0} = 0 \Leftrightarrow \sum_{t=1}^T -2(y_t - \hat{a}_1 x_t - \hat{a}_0) = 0 \rightarrow$  dite "équation normale"

$$\Leftrightarrow \sum_{t=1}^T y_t - \sum_{t=1}^T \hat{a}_1 x_t - \sum_{t=1}^T \hat{a}_0 = 0$$

$$\Leftrightarrow \frac{\sum_{t=1}^T y_t}{T} - \frac{\sum_{t=1}^T \hat{a}_1 x_t}{T} - \frac{T \hat{a}_0}{T} = 0$$

$$\Leftrightarrow \bar{y} - \hat{a}_1 \bar{x} - \hat{a}_0 = 0$$

$\rightarrow$  la droite d'ajustement  $\hat{y}_t$  passe par le point moyen  $(\bar{x}, \bar{y})$

$$\Leftrightarrow \boxed{\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x}}$$

$\rightarrow$  estimateur de  $a_0$  par les MCO

- $\frac{\partial S}{\partial \hat{a}_1} = 0 \Leftrightarrow \sum_{t=1}^T -2x_t(y_t - \hat{a}_1 x_t - \hat{a}_0) = 0 \rightarrow$  "équation normale"

$$\Leftrightarrow -\sum_{t=1}^T x_t y_t + \sum_{t=1}^T \hat{a}_1 x_t^2 + \sum_{t=1}^T \hat{a}_0 x_t = 0$$

$$\Leftrightarrow -\frac{\sum_{t=1}^T x_t y_t}{T} + \frac{\hat{a}_1 \sum_{t=1}^T x_t^2}{T} + \frac{\hat{a}_0 \sum_{t=1}^T x_t}{T} = 0$$

$$\Leftrightarrow -\frac{\sum_{t=1}^T x_t y_t}{T} + \frac{\hat{a}_1 \sum_{t=1}^T x_t^2}{T} + (\bar{y} - \hat{a}_1 \bar{x}) \bar{x} = 0 \text{ en utilisant l'expression de l'estimateur de } a_0$$

$$\Leftrightarrow -\frac{\sum_{t=1}^T x_t y_t}{T} + \hat{a}_1 \left( \frac{\sum_{t=1}^T x_t^2}{T} - \bar{x}^2 \right) + \bar{y} \bar{x} = 0$$

$$\Leftrightarrow \hat{a}_1 = \frac{\frac{\sum_{t=1}^T x_t y_t}{T} - \bar{y} \bar{x}}{\left( \frac{\sum_{t=1}^T x_t^2}{T} - \bar{x}^2 \right)} = \frac{\sum_{t=1}^T x_t y_t - T \bar{y} \bar{x}}{\left( \frac{\sum_{t=1}^T x_t^2 - T \bar{x}^2}{T} \right)} = \frac{\sum_{t=1}^T x_t y_t - T \bar{y} \bar{x}}{\sum_{t=1}^T x_t^2 - T \bar{x}^2}$$

$$\Leftrightarrow \hat{a}_1 = \frac{\sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^T (x_t - \bar{x})^2} \rightarrow \text{estimateur de } a_1 \text{ par les MCO}$$

car on a  $\sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y}) = \sum_{t=1}^T (x_t y_t - x_t \bar{y} - \bar{x} y_t + \bar{y} \bar{x})$

$$= \sum_{t=1}^T x_t y_t - \sum_{t=1}^T x_t \bar{y} - \sum_{t=1}^T \bar{x} y_t + \sum_{t=1}^T \bar{y} \bar{x}$$

$$= \sum_{t=1}^T x_t y_t - T \cdot \frac{\sum_{t=1}^T x_t}{T} \cdot \bar{y} - \bar{x} \cdot \frac{\sum_{t=1}^T y_t}{T} \cdot T + T \bar{y} \bar{x}$$

$$= \sum_{t=1}^T x_t y_t - T \bar{x} \bar{y} - \bar{x} \bar{y} T + T \bar{y} \bar{x}$$

$$= \sum_{t=1}^T x_t y_t - \bar{x} \bar{y} T.$$

Les conditions suffisantes du 2nd ordre pour obtenir un minimum pour S sont :

La fonction S est convexe car on a

$$\frac{\partial^2 S}{\partial \hat{a}_1^2} > 0 ; \quad \begin{vmatrix} \frac{\partial^2 S}{\partial \hat{a}_1^2} & \frac{\partial^2 S}{\partial \hat{a}_1 \partial \hat{a}_0} \\ \frac{\partial^2 S}{\partial \hat{a}_0 \partial \hat{a}_1} & \frac{\partial^2 S}{\partial \hat{a}_0^2} \end{vmatrix} > 0.$$

L'optimum trouvé est donc un minimum.

## II.2/ Hypothèses et propriétés des estimateurs des MCO

Les hypothèses liées à l'erreur  $\varepsilon_t$  sont :

H<sub>1</sub> :  $x_t$  est une variable certaine (non aléatoire)  
 $\Rightarrow \text{Cov}(x_t, \varepsilon_t) = 0 \quad \forall t$  : la variable explicative et l'erreur sont indépendantes.

H<sub>2</sub> :  $E(\varepsilon_t) = 0 \quad \forall t$  : l'erreur est d'espérance nulle.

H<sub>3</sub> :  $\text{Var}(\varepsilon_t) = E(\varepsilon_t^2) - (E(\varepsilon_t))^2 = E(\varepsilon_t^2) = \sigma_\varepsilon^2 \quad \forall t$  car on a supposé  $E(\varepsilon_t) = 0$   
 $\Rightarrow$  la variance de l'erreur est constante (soit homoscedasticité de l'erreur).

H<sub>4</sub> :  $\text{Cov}(\varepsilon_t, \varepsilon_{t'}) = E(\varepsilon_t \cdot \varepsilon_{t'}) - E(\varepsilon_t) \cdot E(\varepsilon_{t'}) = E(\varepsilon_t \cdot \varepsilon_{t'}) = 0$  car on a  $E(\varepsilon_t) = 0 \quad \forall t \neq t'$   
 $\Rightarrow$  les erreurs sont non corrélées.

Ces hypothèses permettent aux estimateurs d'obtenir les bonnes propriétés suivantes :

1/ les estimateurs sont sans biais <sup>(3)</sup> :  $E(\hat{a}_1) = a_1$  et  $E(\hat{a}_0) = a_0$  ;

2/ les estimateurs sont convergents <sup>(4)</sup> :  $\lim_{T \rightarrow \infty} \text{Var}(\hat{a}_1) = 0$  et  $\lim_{T \rightarrow \infty} \text{Var}(\hat{a}_0) = 0$  .

Théorème de Gauss-Markov :

Les estimateurs des MCO ont la plus petite variance parmi les estimateurs linéaires sans biais. On dit que ce sont des estimateurs BLUE (Best Linear Unbiased Estimator).

---

<sup>(3)</sup> Voir la démonstration à l'aide du modèle de régression multiple qui est plus simple à mener.

$$\text{Var}(\hat{a}_1) = \frac{\sigma_\varepsilon^2}{\sum_{t=1}^T (x_t - \bar{X})^2} \quad \text{et} \quad \text{Var}(\hat{a}_0) = \sigma_\varepsilon^2 \left( \frac{1}{T} + \frac{\bar{X}^2}{\sum_{t=1}^T (x_t - \bar{X})^2} \right)$$

(pour la démonstration : se reporter à l'ouvrage de Dormont, Introduction à l'économétrie, Montchrestien).

### II.3/ Critère de jugement de la qualité de l'ajustement d'un modèle

Soit la décomposition suivante :

$$\sum_{t=1}^T (y_t - \bar{y})^2 = \sum_{t=1}^T (\hat{y}_t - \bar{y})^2 + \sum_{t=1}^T e_t^2$$

$$\text{SCT} = \text{SCE} + \text{SCR}$$

où SCT = somme des carrés totale ou variabilité totale de  $y_t$ , SCE = somme des carrés expliquée ou variabilité expliquée par  $\hat{y}_t$ , SCR = somme des carrés des résidus ou variabilité des résidus.

Il vient l'équation suivante appelée équation d'analyse de la variance :

$$\frac{\sum_{t=1}^T (y_t - \bar{y})^2}{T} = \frac{\sum_{t=1}^T (\hat{y}_t - \bar{y})^2}{T} + \frac{\sum_{t=1}^T e_t^2}{T}$$

soit

$$\text{Var}(y) = \text{Var}(\hat{y}) + \text{Var}(e).$$

A partir de l'équation d'analyse de la variance, on va construire le critère du  $R^2$  (ou coefficient de détermination) pour juger de la qualité d'un ajustement.

Le  $R^2$  est donné par le rapport suivant :

$$R^2 = \frac{\text{SCE}}{\text{SCT}} = \frac{\sum_{t=1}^T (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^T (y_t - \bar{y})^2} = \frac{\sum_{t=1}^T (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^T (y_t - \bar{y})^2} = 1 - \frac{\sum_{t=1}^T e_t^2}{\sum_{t=1}^T (y_t - \bar{y})^2}.$$



On a  $\bar{\hat{y}} = \bar{y}$  car on a  $\sum_{t=1}^T e_t = 0$  lorsque le modèle comporte une constante <sup>(5)</sup>.

Plus SCE est proche de SCT, meilleur est l'ajustement du nuage de points par la droite des MCO.

Le  $R^2$  est compris entre 0 et 1 ( $0 \leq R^2 \leq 1$ ) : plus il est proche de 1, meilleur est l'ajustement.

### III/ LE MODELE DE REGRESSION MULTIPLE

Le modèle de régression multiple est une généralisation du modèle de régression simple. Il comporte plusieurs variables explicatives.

Soit le modèle de régression multiple suivant qui comporte k variables explicatives :

$$y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \dots + a_{k-1} x_{(k-1)t} + \varepsilon_t$$

Pour  $t = 1, 2, \dots, T$ , on a

---

<sup>(5)</sup> Démonstration de  $\sum_{t=1}^T e_t = 0$  :

$$\text{On a } e_t = y_t - \hat{y}_t \Leftrightarrow e_t = y_t - \hat{a}_1 x_t - \hat{a}_0.$$

$$\text{En sommant les } e_t, \text{ on obtient } \sum_{t=1}^T e_t = \sum_{t=1}^T (y_t - \hat{a}_1 x_t - \hat{a}_0) = \sum_{t=1}^T y_t - \hat{a}_1 \sum_{t=1}^T x_t - T \hat{a}_0.$$

$$\text{Or on a } \hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x}. \text{ Il vient alors : } \sum_{t=1}^T e_t = \sum_{t=1}^T y_t - \hat{a}_1 \sum_{t=1}^T x_t - T(\bar{y} - \hat{a}_1 \bar{x})$$

$$= T \frac{\sum_{t=1}^T y_t}{T} - \hat{a}_1 T \frac{\sum_{t=1}^T x_t}{T} - T \bar{y} + T \hat{a}_1 \bar{x}$$

$$= T \bar{y} - \hat{a}_1 T \bar{x} - T \bar{y} + T \hat{a}_1 \bar{x} = 0.$$

Démonstration de  $\bar{\hat{y}} = \bar{y}$  :

$$\text{On a } e_t = y_t - \hat{y}_t \Rightarrow \frac{\sum_{t=1}^T e_t}{T} = \frac{\sum_{t=1}^T y_t}{T} - \frac{\sum_{t=1}^T \hat{y}_t}{T} = 0 \text{ puisque } \sum_{t=1}^T e_t = 0.$$

D'où  $\bar{y} = \bar{\hat{y}}$ .

$$\begin{cases} y_1 = a_0 + a_1 x_{11} + a_2 x_{21} + \dots + a_{k-1} x_{(k-1)1} + \varepsilon_1 \\ y_2 = a_0 + a_1 x_{12} + a_2 x_{22} + \dots + a_{k-1} x_{(k-1)2} + \varepsilon_2 \\ \dots \\ y_T = a_0 + a_1 x_{1T} + a_2 x_{2T} + \dots + a_{k-1} x_{(k-1)T} + \varepsilon_T \end{cases}$$

Pour alléger cette écriture, on va écrire ce système d'équations sous forme matricielle :

$$Y_{(T,1)} = X_{(T,k)} a_{(k,1)} + \varepsilon_{(T,1)}$$

où on a

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix}, X = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{(k-1)1} \\ 1 & x_{12} & x_{22} & \dots & x_{(k-1)2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1T} & x_{2T} & \dots & x_{(k-1)T} \end{pmatrix}, a = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{k-1} \end{pmatrix} \text{ et } \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix}$$

### III.1/ Méthode d'estimation des Moindres Carrés Ordinaires (MCO)

Soit le modèle général suivant :

$$Y_{(T,1)} = X_{(T,k)} a_{(k,1)} + \varepsilon_{(T,1)}$$

Afin d'estimer le vecteur  $a$  des coefficients, on applique la méthode des MCO qui consiste toujours à minimiser la somme des carrés des résidus, soit :

$$\text{Min} \sum_{t=1}^T e_t^2 = \min e'e = \min (Y - X\hat{a})'(Y - X\hat{a}) = \min S$$

où  $e'$  est le transposé du vecteur  $e$ .

La fonction  $S$  est minimale si on a :

$$\frac{\partial S}{\partial \hat{a}} = -2X'Y + 2X'X\hat{a} = 0 \quad \Leftrightarrow \quad \hat{a} = (X'X)^{-1}X'Y$$

$$\begin{aligned} \text{En effet, on a } S &= (Y - X\hat{a})'(Y - X\hat{a}) = (Y' - \hat{a}'X')(Y - X\hat{a}) = Y'Y - Y'X\hat{a} - \hat{a}'X'Y + \hat{a}'X'X\hat{a} \\ S &= Y'Y - (\hat{a}'X'Y)' - \hat{a}'X'Y + \hat{a}'X'X\hat{a} \\ S &= Y'Y - 2\hat{a}'X'Y + \hat{a}'X'X\hat{a} \end{aligned}$$

car le transposé d'un scalaire est un scalaire :  $(\hat{a}'X'Y)' = \hat{a}'X'Y$ .

En effet, on a  $S = \sum_{t=1}^T e_t^2$  qui est un scalaire, donc  $S = Y'Y - Y'X\hat{a} - \hat{a}'X'Y + \hat{a}'X'X\hat{a}$  est un scalaire avec  $(Y'Y)_{(1,1)}$ ,  $(Y'X\hat{a})_{(1,1)}$ ,  $(\hat{a}'X'Y)_{(1,1)}$  et  $(\hat{a}'X'X\hat{a})_{(1,1)}$ .

Les équations issues de la relation  $-X'Y + X'X\hat{a} = 0$  sont appelées équations normales.

On voit que l'on ne peut obtenir l'estimateur  $\hat{a}$  de  $a$  que si  $(X'X)$  est inversible. Lorsqu'il y a colinéarité des variables explicatives, la matrice  $(X'X)^{-1}$  n'est pas inversible !

### III.2/ Hypothèses et propriétés des estimateurs des MCO

$H_1$  :  $E(\varepsilon) = 0 \Rightarrow E(\varepsilon_t) = 0 \forall t$  : l'erreur est d'espérance nulle.

$H_2$  :  $X$  est une matrice composée de variables certaines (non aléatoires).

$\Rightarrow \text{Cov}(x_{it}, \varepsilon_t) = 0 \forall t$  et  $i$  : pas de corrélation entre la variable explicative  $x_{it}$  et l'erreur  $\varepsilon_t$ .

$H_3$  :  $\text{Rg}(X) = k$  et  $T > k$

(le nb d'observations  $T$  doit être supérieur au nb de variables explicatives  $k$  car on a

$\text{Rg}(X_{(T,k)}) \leq \text{Min}(T,k)$ . Si  $T > k$ , on a alors  $\text{Rg}(X) = k$  qui est vérifiée )

$\Rightarrow$  il n'existe pas de colinéarité stricte des  $k$  variables explicatives.

$H_4$  :  $V_\varepsilon = E[(\varepsilon - E(\varepsilon))(\varepsilon - E(\varepsilon))'] = E(\varepsilon \varepsilon') = \sigma_\varepsilon^2 I$  car on a  $E(\varepsilon) = 0$  ( $H_1$ )

$V_\varepsilon =$  matrice des variances-covariances des erreurs  $\varepsilon$

où

$$V_\varepsilon = E(\varepsilon \varepsilon') = \begin{pmatrix} \text{Var}(\varepsilon_1) & \text{Cov}(\varepsilon_1, \varepsilon_2) & \dots & \text{Cov}(\varepsilon_1, \varepsilon_T) \\ \text{Cov}(\varepsilon_2, \varepsilon_1) & \text{Var}(\varepsilon_2) & \dots & \text{Cov}(\varepsilon_2, \varepsilon_T) \\ \dots & \dots & \dots & \dots \\ \text{Cov}(\varepsilon_T, \varepsilon_1) & \dots & \dots & \text{Var}(\varepsilon_T) \end{pmatrix} = \begin{pmatrix} E(\varepsilon_1 \varepsilon_1) & E(\varepsilon_1 \varepsilon_2) & \dots & E(\varepsilon_1 \varepsilon_T) \\ E(\varepsilon_2 \varepsilon_1) & E(\varepsilon_2 \varepsilon_2) & \dots & E(\varepsilon_2 \varepsilon_T) \\ \dots & \dots & \dots & \dots \\ E(\varepsilon_T \varepsilon_1) & \dots & \dots & E(\varepsilon_T \varepsilon_T) \end{pmatrix}$$

car on a  $\text{Var}(\varepsilon_1) = E(\varepsilon_1^2) - (E(\varepsilon_1))^2 = E(\varepsilon_1^2)$  car  $E(\varepsilon_1) = 0$

et  $\text{Cov}(\varepsilon_1, \varepsilon_2) = E(\varepsilon_1 \varepsilon_2) - E(\varepsilon_1).E(\varepsilon_2) = E(\varepsilon_1 \varepsilon_2)$  car  $E(\varepsilon_1) = E(\varepsilon_2) = 0$  d'après hypothèse  $H_1$  postulée.

Lorsqu'il n'y a pas autocorrélation ni hétéroscédasticité ( $\neq$  homoscélasticité) des erreurs, on a :

$$V_\varepsilon = E(\varepsilon \varepsilon') = \begin{pmatrix} \sigma_\varepsilon^2 & 0 & \dots & 0 \\ 0 & \sigma_\varepsilon^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \sigma_\varepsilon^2 \end{pmatrix} = \sigma_\varepsilon^2 I \quad (I \text{ est une matrice identité})$$

car on a  $\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2 \forall t$  (homoscélasticité des erreurs) et  $\text{Cov}(\varepsilon_t, \varepsilon_{t'}) = 0 \forall t \neq t'$  (non

autocorrélation des erreurs).

$H_5$  :  $\frac{X'X}{T}$  tend vers une matrice finie non singulière (inversible ou régulière).

Ces hypothèses permettent aux estimateurs d'obtenir les bonnes propriétés suivantes :

1/ les estimateurs sont sans biais <sup>(6)</sup> :  $E(\hat{a}) = a$  ;

2/ les estimateurs sont convergents :  $\lim_{T \rightarrow \infty} V_{\hat{a}} = 0$  .

En effet, on a <sup>(7)</sup>

$$V_{\hat{a}} = \sigma_{\varepsilon}^2 (X'X)^{-1}$$

ou encore

$$V_{\hat{a}} = \sigma_{\varepsilon}^2 \frac{1}{T} \left( \frac{1}{T} (X'X) \right)^{-1}$$

On a  $\lim_{T \rightarrow \infty} V_{\hat{a}} = 0$  si l'hypothèse  $H_5$  suivante est vérifiée :  $\frac{1}{T} (X'X)$  tend vers une matrice finie, définie positive et inversible lorsque  $T$  tend vers  $\infty$  .

On calcule  $V_{\hat{a}} = \sigma_{\varepsilon}^2 (X'X)^{-1}$  à l'aide de l'estimateur de  $\sigma_{\varepsilon}^2$  qui s'écrit comme suit :

$$\hat{\sigma}_{\varepsilon}^2 = s^2 = \frac{\text{SCR}}{T-k} = \frac{\sum_{t=1}^T e_t^2}{T-k}$$

(pour la démonstration : voir Dormont, Introduction à l'économétrie, Montchrestien).

<sup>(6)</sup> On a  $\hat{a} = (X'X)^{-1}X'Y = (X'X)^{-1}X'(Xa + \varepsilon) = (X'X)^{-1}X'Xa + (X'X)^{-1}X'\varepsilon = a + (X'X)^{-1}X'\varepsilon$ .  
D'où  $E(\hat{a}) = E(a + (X'X)^{-1}X'\varepsilon) = a + (X'X)^{-1}X'E(\varepsilon) = a$  car on a  $E(\varepsilon) = 0$  d'après  $H_1$ .

<sup>(7)</sup>  $V_{\hat{a}} = E[(\hat{a} - E(\hat{a}))(\hat{a} - E(\hat{a}))'] = E[(\hat{a} - a)(\hat{a} - a)']$  car  $E(\hat{a}) = a$ .  
Or  $\hat{a} = (X'X)^{-1}X'Y = (X'X)^{-1}X'(Xa + \varepsilon) = a + (X'X)^{-1}X'\varepsilon$ . D'où  $\hat{a} - a = (X'X)^{-1}X'\varepsilon$ .  
Il vient alors :  
 $V_{\hat{a}} = E[(X'X)^{-1}X'\varepsilon((X'X)^{-1}X'\varepsilon)']$   
 $= E[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}]$   
 $= (X'X)^{-1}X'E[\varepsilon\varepsilon']X(X'X)^{-1}$  car  $X$  est une matrice composée de variables certaines d'après  $H_2$   
 $= \sigma_{\varepsilon}^2 (X'X)^{-1}X'X(X'X)^{-1}$  car  $E[\varepsilon\varepsilon'] = \sigma_{\varepsilon}^2 I$  d'après  $H_4$   
 $= \sigma_{\varepsilon}^2 (X'X)^{-1}$   
 $\sigma_{\varepsilon}^2 (X'X)^{-1}$ .

### III.3/ Critère de jugement de la qualité de l'ajustement d'un modèle

Comme pour le modèle de régression simple, on a la décomposition suivante :

$$\sum_{t=1}^T (y_t - \bar{y})^2 = \sum_{t=1}^T (\hat{y}_t - \bar{\hat{y}})^2 + \sum_{t=1}^T e_t^2$$

$$\text{SCT} = \text{SCE} + \text{SCR}$$

où SCT = somme des carrés totale ou variabilité totale de  $y_t$ , SCE = somme des carrés expliquée ou variabilité expliquée par  $\hat{y}_t$ , SCR = somme des carrés des résidus ou variabilité des résidus.

D'où

$$\frac{\sum_{t=1}^T (y_t - \bar{y})^2}{T} = \frac{\sum_{t=1}^T (\hat{y}_t - \bar{\hat{y}})^2}{T} + \frac{\sum_{t=1}^T e_t^2}{T}$$

$$\text{Var}(y) = \text{Var}(\hat{y}) + \text{Var}(e).$$

Comme pour le modèle de régression simple, on va construire le critère du  $R^2$  (ou coefficient de détermination) à partir de l'équation d'analyse de la variance, d'où

$$R^2 = \frac{\text{SCE}}{\text{SCT}} = \frac{\sum_{t=1}^T (\hat{y}_t - \bar{\hat{y}})^2}{\sum_{t=1}^T (y_t - \bar{y})^2} = \frac{\sum_{t=1}^T (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^T (y_t - \bar{y})^2} = 1 - \frac{\sum_{t=1}^T e_t^2}{\sum_{t=1}^T (y_t - \bar{y})^2}.$$

On a comme pour le modèle de régression simple, on a :

$$\bar{\hat{y}} = \bar{y} \quad \text{et} \quad \sum_{t=1}^T e_t = 0 \quad (\text{lorsque le modèle comporte une constante}).$$

Le coefficient de détermination corrigé :

Le  $R^2$  ne permet de comparer que des modèles ayant le même nombre de variables explicatives, le même nombre d'observations et la même forme (on ne peut pas comparer un modèle simple avec un modèle en log).

Lorsque l'on ajoute des variables explicatives supplémentaires dans un modèle, le  $R^2$  a tendance à augmenter sans qu'il y ait forcément amélioration du modèle. C'est pourquoi, lorsque l'on veut comparer des modèles qui n'ont pas le même nombre de variables explicatives, on utilise le  $R^2$  corrigé pour s'affranchir du biais :

$$R_c^2 = 1 - (1 - R^2) \frac{T-1}{T-k}.$$

Remarque :

En général, lorsque les modèles n'ont pas le même nombre de variables explicatives, on utilise pour comparer les modèles le critère du  $s = \hat{\sigma}_e = \sqrt{\frac{e'e}{T-k}}$  où  $e'$  est le transposé du vecteur des résidus du modèle estimé,  $T$  désigne le nombre d'observations et  $k$  le nombre de variables explicatives. Le meilleur modèle est celui qui a le  $s$  le plus petit.

Test de Box-Cox : (voir le doc. "pratique de l'économétrie par Eviews")

Le test de Box-Cox permet de comparer un modèle simple à un modèle en log.

### III.4/ Utilisation de variables indicatrices (ou variables muettes ou dummies)

Une variable indicatrice est une variable explicative particulière qui est composée de 0 et de 1.

On peut l'utiliser dans 3 cas :

- Correction des valeurs anormales ;
- En tant que variable qualitative (Ex : pour la variable sexe de l'individu) ;
- Correction de la saisonnalité.

On n'étudiera ici que la correction des valeurs anormales.

Correction des valeurs anormales (ou points aberrants) :

La présence d'un point aberrant dans une série temporelle a pour conséquence l'autocorrélation des erreurs, il faut donc l'éliminer. Si on a une série temporelle, on ne peut pas le supprimer directement de la série (on aurait un « trou » dans la série). Il faut utiliser une variable indicatrice.

Cette variable aura une valeur égale à 1 pour la date à laquelle on observe le point aberrant et une valeur égale à 0 pour toutes les autres dates de la série temporelle, c'est-à-dire on a :

$$I_{82.3} = \begin{cases} 1 & \text{si } t = 1982.3 \text{ pour le 3}^{\text{ème}} \text{ trimestre de l'année 1982} \\ 0 & \text{sinon.} \end{cases}$$

Lorsque l'on estime le modèle suivant avec la variable indicatrice  $I_{82.3}$  :

$$y_t = a_1 x_t + a_2 I_{82.3} + a_0 + \varepsilon_t$$

cela revient à estimer le modèle suivant sans le point aberrant en  $t = 1982.3$

$$y_t = a_1 x_t + a_0 + v_t$$

où  $\varepsilon_t$  et  $v_t$  désignent l'erreur de chacun des 2 modèles.

Détection des valeurs anormales : construction de l'intervalle de confiance pour les résidus :

Soit le modèle suivant :

$$Y = Xa + \varepsilon$$

Le modèle estimé s'écrit :

$$\hat{Y} = X\hat{a} + e$$

avec  $e = Y - \hat{Y} = Y - PY$  et  $P = X(X'X)^{-1}X'$  est une matrice de projection orthogonale.

$$\begin{aligned} \text{On a } e &= (I - X(X'X)^{-1}X')Y = (I - X(X'X)^{-1}X')(Xa + \varepsilon) = Xa + \varepsilon - X(X'X)^{-1}X'(Xa + \varepsilon) \\ &= Xa + \varepsilon - X(X'X)^{-1}X'Xa - X(X'X)^{-1}X'\varepsilon \\ &= Xa + \varepsilon - Xa - X(X'X)^{-1}X'\varepsilon \\ &= (I - X(X'X)^{-1}X')\varepsilon \\ &= M\varepsilon \quad \text{où } M \text{ est une matrice de projection orthogonale.} \end{aligned}$$

On obtient :

$$E(e) = E(M\varepsilon) = ME(\varepsilon) = 0 \quad \text{car } E(\varepsilon) = 0.$$

$$\text{Var}(e) = E[(e - E(e))(e - E(e))'] = E(e \cdot e') \quad \text{car } E(e) = 0$$

$$\begin{aligned} &= E(M\varepsilon \cdot \varepsilon' M') \\ &= M E(\varepsilon \varepsilon') M' \\ &= M \text{Var}(\varepsilon) M' \\ &= \sigma_\varepsilon^2 M M' \\ &= \sigma_\varepsilon^2 M^2 \quad \text{car } M \text{ est une matrice de projection orthogonale, on a alors } M^2 = M \text{ et } M^2 = M \\ &= \sigma_\varepsilon^2 M \\ &= \sigma_\varepsilon^2 (I - X(X'X)^{-1}X') \end{aligned}$$

D'où, on a

$$\text{Var}(e_t) = \sigma_\varepsilon^2 (1 - h_{tt}).$$

Si on a  $\varepsilon \sim N(0, \sigma_\varepsilon^2 I)$ , on a alors :

$$e_t \sim N(0, \text{Var}(e_t)), \quad e_t \text{ suit aussi la loi normale car } e = M\varepsilon.$$

On obtient alors :

$$\frac{e_t}{\sqrt{\text{Var}(e_t)}} = \frac{e_t}{\sigma_\varepsilon \sqrt{1 - h_{tt}}} \sim N(0, 1).$$

Si on remplace  $\sigma_e$  par son estimateur  $\hat{\sigma}_e = s = \sqrt{\frac{SCR}{T-k}}$ , le rapport  $\frac{e_t}{s\sqrt{1-h_{tt}}}$  suit alors la loi de Student avec  $(T-k)$  degrés de liberté :

$$\frac{e_t}{s\sqrt{1-h_{tt}}} \sim S(T-k).$$

Lorsque  $T$  grand ( $T > 30$ ), on a  $\frac{e_t}{s\sqrt{1-h_{tt}}}$  qui suit la loi normale centrée et réduite, on peut alors écrire :

$$\text{Prob}\left(-1,96 \leq \frac{e_t}{s\sqrt{1-h_{tt}}} \leq 1,96\right) = 1 - \alpha$$

où 1,96 représente la valeur critique de la loi normale centrée et réduite pour un risque de 1<sup>ère</sup> espèce  $\alpha$  (en générale,  $\alpha = 5\%$ ).

D'où

$$\text{Prob}(-1,96 s\sqrt{1-h_{tt}} \leq e_t \leq 1,96 s\sqrt{1-h_{tt}}) = 95\%.$$

Comme  $h_{tt}$  est petit et négligeable, on obtient :

$$-1,96s \leq e_t \leq 1,96s$$

qui est approché par

$$-2s \leq e_t \leq 2s.$$

Les résidus qui se trouvent en dehors de cet intervalle sont des points aberrants.

On ne peut enlever un point aberrant (à l'aide de la variable indicatrice) que si ce point aberrant a une explication économique autrement il faut le conserver dans l'échantillon de données.

### III.5/ Prévisions

Lorsque les coefficients du modèle ont été estimés, il est possible de faire une prévision à un horizon  $h$ .

Soit un modèle estimé sur la période  $t = 1, \dots, T$  :

$$\hat{y}_t = \hat{a}_0 + \hat{a}_1 x_{1t} + \hat{a}_2 x_{2t} + \dots + \hat{a}_{k-1} x_{(k-1)t},$$

si la valeur des variables explicatives  $x_{1(T+1)}, x_{2(T+1)}, \dots, x_{(k-1)(T+1)}$  est connue en  $T+1$ , la prévision de  $\hat{y}_{T+1}$  est donnée par :



$$\hat{y}_{T+1} = \hat{a}_0 + \hat{a}_1 x_{1(T+1)} + \hat{a}_2 x_{2(T+1)} + \dots + \hat{a}_{k-1} x_{(k-1)(T+1)}.$$

L'erreur de prévision en T+1 est donnée par :

$$e_{T+1} = y_{T+1} - \hat{y}_{T+1}.$$

Cette erreur est sans biais car on a  $E(e_{T+1}) = 0$ .

Est ce que cet écart  $e_{T+1} = y_{T+1} - \hat{y}_{T+1}$  est acceptable ? Pour répondre à cette question, on va construire un intervalle de confiance (ou de prévision) pour  $e_{T+1}$ .

Pour un modèle avec une seule variable explicative , on a :

$$- t_{T-2}^{\alpha/2} \hat{\sigma}_\varepsilon \sqrt{\frac{1}{T} + \frac{(x_{T+1} - \bar{x})^2}{\sum_{t=1}^T (x_t - \bar{x})^2} + 1} \leq y_{T+1} - \hat{y}_{T+1} \leq + t_{T-2}^{\alpha/2} \hat{\sigma}_\varepsilon \sqrt{\frac{1}{T} + \frac{(x_{T+1} - \bar{x})^2}{\sum_{t=1}^T (x_t - \bar{x})^2} + 1}$$

$$\text{car } V(e_{T+1}) = \hat{\sigma}_\varepsilon^2 \left( \frac{1}{T} + \frac{(x_{T+1} - \bar{x})^2}{\sum_{t=1}^T (x_t - \bar{x})^2} + 1 \right)$$

(pour la démonstration : voir Dormont, Introduction à l'économétrie, Montchrestien)

où  $t_{T-2}^{\alpha/2}$  représente la valeur critique de la loi de Student pour un risque de  $\alpha$  % et (T-2) degrés de liberté.

D'où l'intervalle de prévision suivant :

$$\hat{y}_{T+1} - t_{T-2}^{\alpha/2} \hat{\sigma}_\varepsilon \sqrt{\frac{1}{T} + \frac{(x_{T+1} - \bar{x})^2}{\sum_{t=1}^T (x_t - \bar{x})^2} + 1} \leq y_{T+1} \leq \hat{y}_{T+1} + t_{T-2}^{\alpha/2} \hat{\sigma}_\varepsilon \sqrt{\frac{1}{T} + \frac{(x_{T+1} - \bar{x})^2}{\sum_{t=1}^T (x_t - \bar{x})^2} + 1}.$$

Pour un modèle avec plusieurs variables explicatives, on a :

$$- t_{T-k}^{\alpha/2} \hat{\sigma}_\varepsilon \sqrt{X'_{T+1}(X'X)^{-1}X_{T+1} + 1} \leq y_{T+1} - \hat{y}_{T+1} \leq + t_{T-k}^{\alpha/2} \hat{\sigma}_\varepsilon \sqrt{X'_{T+1}(X'X)^{-1}X_{T+1} + 1}$$

$$\text{car } V(e_{T+1}) = \hat{\sigma}_\varepsilon^2 [ X'_{T+1}(X'X)^{-1}X_{T+1} + 1 ].$$

D'où :

$$\hat{y}_{T+1} - t_{T-k}^{\alpha/2} \hat{\sigma}_\varepsilon \sqrt{X'_{T+1}(X'X)^{-1}X_{T+1} + 1} \leq y_{T+1} \leq \hat{y}_{T+1} + t_{T-k}^{\alpha/2} \hat{\sigma}_\varepsilon \sqrt{X'_{T+1}(X'X)^{-1}X_{T+1} + 1}.$$

La vraie valeur  $y_{T+1}$  est contenue dans cet intervalle.

Lorsque l'on dispose de la valeur observée  $y_{T+1}$ , on peut vérifier si le modèle que l'on a estimé est bon ou non en regardant si cette valeur appartient ou non à l'intervalle de confiance. Si cette valeur n'appartient pas à l'intervalle de confiance, le modèle estimé n'est pas bon.

#### IV/ LES TESTS

##### IV.1/ Test de significativité d'un coefficient : test de Student

Pour savoir si une variable joue un rôle explicatif dans un modèle, on effectue un test de Student ou test de significativité du coefficient de la variable explicative.

Pour faire un test de Student, il faut vérifier au préalable que les erreurs suivent une loi normale :

$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2).$$

Posons d'abord les hypothèses du test de Student :

soit le modèle général suivant :

$$y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \dots + a_{k-1} x_{(k-1)t} + \varepsilon_t \text{ pour } t = 1, 2, \dots, T$$

on a

$$\begin{cases} H_0 : a_i = 0 \text{ où } i = 0, 1, \dots, (k-1) \Rightarrow \text{le coefficient n'est pas significatif} \\ H_1 : a_i \neq 0 \Rightarrow \text{le coefficient est significatif} \end{cases}$$

La statistique de test est :

$$t = \frac{\hat{a}_i - a_i}{\hat{\sigma}_{\hat{a}_i}} \sim S(T-k).$$

La statistique de test suit la loi de Student à  $(T-k)$  degrés de liberté car les erreurs du modèle suivent une loi normale.

Sous  $H_0$  vraie, on a

$$t = \frac{\hat{a}_i}{\hat{\sigma}_{\hat{a}_i}} \sim S(T-k).$$

La règle de décision est la suivante :

Si  $|t| > t^*$  où  $t^*$  est la valeur critique de la table de Student pour un risque fixé et un nombre de degré de liberté égal à  $(T-k)$

$\Rightarrow$  on rejette  $H_0$  et on accepte  $H_1$  : le coefficient est significativement différent de zéro et la variable joue un rôle explicatif dans le modèle.

Remarque :

Lorsque la taille d'échantillon est grande ( $T > 30$ ), on peut comparer  $|t|$  directement avec le seuil critique de la loi normale centrée et réduite qui est 1,96 (pour un risque de 5%) car d'après le théorème central limite, la loi de Student tend vers une loi normale lorsque  $T$  est suffisamment grand.

Donc , si  $|t| > 1,96 \Rightarrow$  on rejette  $H_0$  et on accepte  $H_1$  : le coefficient est significatif et la variable joue un rôle explicatif dans le modèle.

Si le coefficient n'est pas significativement différent de zéro, il faut enlever la variable explicative correspondante du modèle (à condition que le critère du  $s$  n'augmente pas ! Il arrive que nous puissions nous tromper sur la non significativité d'une variable en présence d'une colinéarité des variables explicatives qui entraîne des  $t$  de Student relativement faibles nous conduisant à rejeter à tort certaines variables explicatives. C'est pourquoi il faut examiner la valeur du  $s$  après le retrait des variables jugées non significatives. Une hausse de la valeur du  $s$  indique que la variable retirée était en fait contributive à l'explication de la variable endogène).

**IV.2/ Test de significativité global (de plusieurs coefficients) : test de Fisher**

Le test de Fisher permet de tester la significativité de l'ensemble des coefficients d'un modèle.

Soit le modèle général :

$$y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \dots + a_{k-1} x_{(k-1)t} + \varepsilon_t \text{ pour } t = 1, 2, \dots, T.$$

Les hypothèses du test de Fisher sont les suivantes :

$$\left\{ \begin{array}{l} H_0 : a_1 = a_2 = \dots = a_{k-1} = 0 \text{ (la constante } a_0 \text{ est non nul)} \\ \quad \Rightarrow \text{l'ensemble des coefficients du modèle est non significatif} \\ H_1 : \text{il existe } \underline{\text{au moins un}} \text{ coefficient non nul.} \end{array} \right.$$

La statistique de test sous  $H_0$  vraie est :

$$f = \frac{(SCR_c - SCR_{nc}) / (dl_c - dl_{nc})}{SCR_{nc} / dl_{nc}} \sim F(dl_c - dl_{nc}, dl_{nc}) = F(p, q)$$

où  $SCR_c = SCR$  du modèle contraint (modèle lorsque  $H_0$  est vérifiée)

$SCR_{nc}$  = SCR du modèle non contraint (modèle lorsque  $H_1$  est vérifiée)

$dl_c$  = degré de liberté du modèle contraint =  $T - 1$  (car il n'y a qu'une seule variable explicative qui est non nul qui est le terme constant  $a_0$ )

$dl_{nc}$  = degré de liberté du modèle non contraint =  $T - k$  (car il y a  $k$  variables explicatives au maximum dans le modèle).

La règle de décision est la suivante :

Si  $f > f^*(p,q)$  où  $f^*(p,q)$  est la valeur donnée par la table de Fisher pour  $p$  et  $q$  donnés et pour un risque fixé

⇒ On accepte  $H_1$  : il existe au moins un coefficient non nul.

Ce test est peu utilisé car lorsqu'il indique qu'il y a au moins un coefficient non nul, il ne précise pas lesquels. Il est moins précis que le test de Student.

### IV.3/ Test de normalité des erreurs

Avant d'effectuer le test de Student, il faut effectuer un test de normalité afin de vérifier que les erreurs sont gaussiennes.

Soit le modèle suivant :

$$Y = Xa + \varepsilon.$$

Les hypothèses du test sont les suivantes :

$$\begin{cases} H_0 : \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \\ H_1 : \text{les erreurs ne suivent pas une loi normale.} \end{cases}$$

Une loi normale a un coefficient de symétrie (ou skewness) égal à 0 et un coefficient d'aplatissement (kurtosis) égal à 3.

Le coefficient de symétrie est donné par :

$$\alpha_3 = \frac{\mu_3}{\sigma_\varepsilon^3}$$

où  $\mu_3$  est un moment d'ordre 3 et  $\sigma_\varepsilon$  est l'écart-type de l'erreur.

Le coefficient d'aplatissement est donné par :

$$\alpha_4 = \frac{\mu_4}{\sigma_\varepsilon^4}$$

où  $\mu_4$  est un moment d'ordre 4.

Tester les deux hypothèses précédentes revient en fait à tester les quatre hypothèses suivantes :

$$\begin{cases} H_0 : \alpha_3 = 0 \Leftrightarrow \mu_3 = 0 \\ H_1 : \alpha_3 \neq 0 \Leftrightarrow \mu_3 \neq 0 \end{cases} \quad \text{et} \quad \begin{cases} H_0 : \alpha_4 = 3 \Leftrightarrow \mu_4 = 3\sigma_\varepsilon^4 \\ H_1 : \alpha_4 \neq 3 \Leftrightarrow \mu_4 \neq 3\sigma_\varepsilon^4 \end{cases}$$

On a une loi normale que si on a  $\alpha_3 = 0$  et  $\alpha_4 = 3$ .

Les statistiques de test sont calculées de la façon suivante :

$$\alpha_3 = \frac{\mu_3}{s^3} \quad \text{et} \quad \alpha_4 = \frac{\mu_4}{s^4}$$

$$\text{avec } s = \sqrt{\frac{SCR}{T-k}}, \quad \mu_3 = \frac{1}{T} \sum_{t=1}^T (e_t - \bar{e})^3 \quad \text{et} \quad \mu_4 = \frac{1}{T} \sum_{t=1}^T (e_t - \bar{e})^4.$$

Sous  $H_0$  vraie, on a :

$$\alpha_3 \sim N\left(0, \frac{3!}{T}\right) \quad \text{et} \quad \alpha_4 \sim N\left(3, \frac{4!}{T}\right).$$

D'où

$$\frac{\alpha_3 - 0}{\sqrt{\frac{3!}{T}}} = \frac{\left(\frac{\mu_3}{s^3}\right) - 0}{\sqrt{\frac{3!}{T}}} \sim N(0,1) \quad \text{et} \quad \frac{\alpha_4 - 3}{\sqrt{\frac{4!}{T}}} = \frac{\left(\frac{\mu_4}{s^4}\right) - 3}{\sqrt{\frac{4!}{T}}} \sim N(0,1)$$

On compare ensuite les valeurs de ces ratios à 1,96 (valeur critique de la loi normale centrée et réduite pour un risque de 5% et pour un nombre d'observations grand ( $T > 30$ )).

La règle de décision est la suivante :

si l'une des valeurs de ces 2 ratios  $> 1,96 \Rightarrow$  On accepte  $H_1$  : les erreurs ne suivent pas une loi normale. Les erreurs n'obéissent à une loi normale que si les valeurs des 2 ratios  $< 1,96$ .

#### IV.4/ Tests d'autocorrélation

##### a) Test d'autocorrélation des erreurs d'ordre 1 : test de Durbin-Watson

Soit le modèle général :

$$y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \dots + a_{k-1} x_{(k-1)t} + \varepsilon_t \quad \text{pour } t = 1, 2, \dots, T$$

avec  $\varepsilon_t = \rho \varepsilon_{t-1} + v_t$ ,  $|\rho| < 1$ ,  $v_t \sim N(0, \sigma_v^2)$  et  $\text{cov}(v_t, v_{t'}) = 0$  pour  $t \neq t'$ .

Les hypothèses du test sont les suivantes :

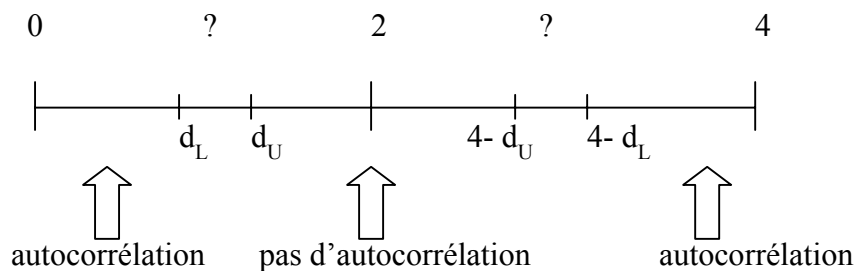
$$\begin{cases} H_0 : \rho = 0 & \Rightarrow \text{les erreurs ne sont pas autocorrélées} \\ H_1 : \varepsilon_t = \rho \varepsilon_{t-1} + v_t & \Rightarrow \text{les erreurs sont autocorrélées d'ordre 1} \end{cases}$$

La statistique de test est :

$$DW = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T (e_t - \bar{e})^2} = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2}$$

où  $e_t$  sont les résidus du modèle général estimé et on a  $\bar{e} = 0$  car il y a un terme constant dans le modèle.

De par sa construction, le DW varie entre 0 et 4. Nous avons les cas suivants selon les valeurs que peut prendre le DW :



Les valeurs  $d_L$  et  $d_U$  sont déterminées à partir de la table de Durbin et Watson en fonction de la taille de l'échantillon et du nombre de variables explicatives pour un risque fixé.

Lorsque nous nous trouvons dans la zone d'incertitude où apparaît un point d'interrogation (dans l'intervalle  $[d_L, d_U]$  ou dans l'intervalle  $[4-d_U, 4-d_L]$ ), nous choisissons comme hypothèse celle qui est la plus fâcheuse, c'est-à-dire  $H_1$ .

#### Conditions d'utilisation du test de DW :

- Le modèle doit comporter un terme constant (car les tables de DW sont tabulées pour des modèles comportant un terme constant, cependant il existe des tables pour des modèles sans terme constant).
- La variable à expliquer ne doit pas figurer parmi les variables explicatives (en tant que variable retardée). Si c'est le cas, on doit utiliser la statistique du « h » de Durbin (voir le chapitre VI sur les modèles autorégressifs).

### Causes de l'autocorrélation des erreurs :

- Mauvaise spécification du modèle
- Instabilité des coefficients
- Présence de points aberrants
- Oubli d'une variable explicative importante
- Véritable autocorrélation  $\Rightarrow$  on utilise dans ce cas la méthode d'estimation des MCG.

### b) Test d'autocorrélation des erreurs d'ordre supérieur : test de Box-Pierce :

Soit le modèle suivant :

$$y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \dots + a_{k-1} x_{(k-1)t} + \varepsilon_t \text{ pour } t = 1, 2, \dots, T$$

avec une autocorrélation des erreurs d'ordre K (  $K > 1$  ) :

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \dots + \rho_K \varepsilon_{t-K} + v_t \quad \text{où } v_t \sim N(0, \sigma_v^2).$$

Les hypothèses de ce test sont les suivantes :

$$\begin{cases} H_0 : \rho_1 = \rho_2 = \dots = \rho_K = 0 \\ H_1 : \text{il existe au moins un } \rho_i \text{ significativement différent de } 0. \end{cases}$$

Pour effectuer ce test, on a recours à la statistique Q qui est donnée par :

$$Q = n \sum_{k=1}^K \hat{\rho}_k^2$$

où n est le nombre d'observations et  $\hat{\rho}_k^2$  est le coefficient d'autocorrélation d'ordre k des résidus estimés  $e_t$ .

Sous l'hypothèse  $H_0$  vraie, Q suit la loi du Khi-deux avec K degrés de liberté :

$$Q = n \sum_{k=1}^K \hat{\rho}_k^2 \sim \chi^2(K).$$

La règle de décision est la suivante :

si  $Q > k^*$  où  $k^*$  est la valeur donnée par la table du Khi-Deux pour un risque fixé et un nombre K de degrés de liberté

$\Rightarrow$  On rejette  $H_0$  et on accepte  $H_1$  (autocorrélation des erreurs).

#### IV.5/ Test d'hétéroscédasticité

Il existe plusieurs tests pour détecter l'hétéroscédasticité des erreurs : test de Goldfeld-Quandt, test de White, test de Breusch-Pagan et test de Park-Glejser.

Nous n'étudierons ici que le test de White.

##### Test de White :

Le test de White permet de tester plusieurs variables explicatives censées être responsables de l'hétéroscédasticité des erreurs. Lorsqu'il y a hétéroscédasticité, la variance de l'erreur est liée aux valeurs de la variable explicative responsable de l'hétéroscédasticité.

Le test de White prend en compte toutes les variables explicatives du modèle ainsi que leur carré et leur produit (de deux variables explicatives).

Soit par exemple le modèle suivant :

$$C_t = a_0 + a_1 Y_t + \varepsilon_t \quad \text{pour } t = 1, 2, \dots, T$$

les hypothèses du test s'écrivent :

$$\begin{cases} H_0 : V(\varepsilon_t) = \alpha_0 \text{ et } \alpha_1 = \alpha_2 = 0 & \Rightarrow \text{homoscédasticité des erreurs} \\ H_1 : V(\varepsilon_t) = \alpha_0 + \alpha_1 Y_t + \alpha_2 Y_t^2 & \Rightarrow \text{hétéroscédasticité des erreurs} \end{cases}$$

La statistique de test sous  $H_0$  vraie s'exprime de la manière suivante :

$$T R^2 \sim \chi^2(2) \quad (\text{le } 2 \text{ se réfère aux } 2 \text{ contraintes } \alpha_1 = \alpha_2 = 0)$$

où  $T$  est le nombre d'observations et  $R^2$  est le coefficient de détermination de la régression suivante :

$$e_t^2 = \alpha_0 + \alpha_1 Y_t + \alpha_2 Y_t^2$$

où  $e_t$  est le résidu du modèle estimé  $C_t = \hat{a}_0 + \hat{a}_1 Y_t + e_t$ .

La règle de décision du test est la suivante :

Si  $T R^2 > k^*$  où  $k^*$  est la valeur donnée par la table du Khi-Deux pour un risque et un nombre de degrés de liberté fixés (ici 2 car on a 2 contraintes  $\alpha_1 = \alpha_2 = 0$ )

$\Rightarrow$  On rejette  $H_0$  et on accepte  $H_1$  (il y a hétéroscédasticité des erreurs).

##### Remarque :

Lorsqu'il y a des variables indicatrices dans le modèle, on ne les prend pas en compte dans le test car elles ne peuvent pas être responsables de l'hétéroscédasticité (car ce sont des variables dont les valeurs valent 1 et 0 qui ne sont pas liées à la variance de l'erreur).



### Causes de l'hétéroscédasticité des erreurs :

L'hétéroscédasticité est un problème qui est en général spécifique aux modèles en coupe instantanée qui s'écrivent par exemple de la manière suivante :

$$C_i = a_0 + a_1 Y_i \quad \text{pour } i=1, \dots, 20$$

où  $C_i$  = consommation pour le pays  $i$  en 1997 et  $Y_i$  = revenu pour le pays  $i$  en 1997.

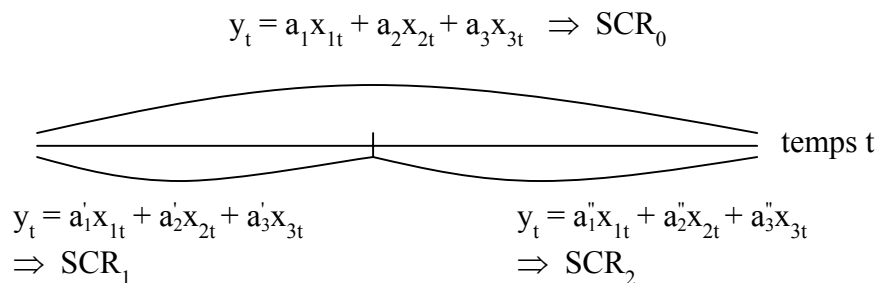
On étudie la relation entre les variables pour un ensemble d'individus (personne, pays, véhicules, ...) et pour une même date.

#### **IV.6/ Test de stabilité : test de Chow**

Le test de Chow permet de savoir si un modèle est stable ou non sur une période donnée. Un modèle instable a ses coefficients qui varient durant la période considérée.

Avant d'effectuer un test de Chow, il faut toujours vérifier qu'il n'y a pas une hétéroscédasticité des erreurs. En présence d'une hétéroscédasticité, le test de Chow rejette à tort l'hypothèse  $H_0$  de stabilité des coefficients.

Le test de Chow se construit comme un test de Fisher. Les hypothèses du test sont les suivantes pour le cas ci-dessous :



$$\left\{ \begin{array}{l} H_0 : a_1 = a_1' = a_1'' \\ \quad a_2 = a_2' = a_2'' \rightarrow \text{le modèle est stable} \\ \quad a_3 = a_3' = a_3'' \\ \Rightarrow \text{modèle contraint avec } SCR_c = SCR_0 \\ \\ H_1 : a_1' \neq a_1'' \\ \quad a_2' \neq a_2'' \rightarrow \text{le modèle est instable} \\ \quad a_3' \neq a_3'' \\ \Rightarrow \text{modèle non contraint avec } SCR_{nc} = SCR_1 + SCR_2 \end{array} \right.$$

où  $SCR_c$  = somme des carrés des résidus du modèle contraint et  $SCR_{nc}$  = somme des carrés des résidus du modèle non contraint.

La statistique de test sous  $H_0$  vraie s'écrit :

$$f = \frac{(SCR_c - SCR_{nc}) / (dl_c - dl_{nc})}{SCR_{nc} / dl_{nc}} \sim F(dl_c - dl_{nc}, dl_{nc}) = F(p, q)$$

$$\Leftrightarrow f = \frac{(SCR_0 - (SCR_1 + SCR_2)) / (dl_c - dl_{nc})}{(SCR_1 + SCR_2) / dl_{nc}} \sim F(dl_c - dl_{nc}, dl_{nc}) = F(p, q).$$

La règle de décision est la suivante :

Si  $f > f^*(p, q) \Rightarrow$  On accepte  $H_1$  : le modèle est instable.

#### IV.7/ Test de colinéarité : test de Belsley Kuh Welsh

Lorsqu'un coefficient n'est pas significativement différent de 0, il convient de l'éliminer et de ré-estimer les coefficients du modèle.

Les causes de la non significativité sont :

- \_ soit une absence de corrélation avec la variable à expliquer
- \_ soit une colinéarité trop élevée entre les variables explicatives.

Une colinéarité stricte (par exemple  $x_1 = 2x_2$ ) entraîne un message ERROR de l'ordinateur car  $(X'X)^{-1}$  ne peut être calculé,  $X'X$  n'est pas inversible. En effet, on a  $\det(X'X) = 0$ .

Dans le cas d'une colinéarité non stricte ( $x_1 \approx 2x_2$ ), le  $\det(X'X) \rightarrow 0$  ce qui va impliquer des valeurs numériques de  $(X'X)^{-1}$  qui seront très grandes puisque le déterminant tend vers 0.

Il en résulte que les estimateurs des coefficients tendront à avoir des variances élevées. En effet, on a  $V_{\hat{a}} = \sigma_e^2 (X'X)^{-1}$  et les t de Student des coefficients estimés de  $\hat{a}$  vont .

Dans ce cas, les variables deviennent non significatives alors qu'elles peuvent expliquer la variable endogène.

Il convient donc de faire un test de colinéarité (test de Belsley Kuh Welsh) dans le cas où on aurait plusieurs variables et des t de Student petits.

La colinéarité des variables explicatives est très fréquente dans les modèles à retards échelonnés.

#### Principe du test :

On regarde pour faire le test de colinéarité l'indice de conditionnement :

$$I_k = \frac{d_{\max}}{d_{\min}} = \frac{\sqrt{\lambda_{\max}}}{\sqrt{\lambda_{\min}}}$$

où  $d$  est la valeur singulière de  $X_{(T,k)}$  et  $\lambda$  est la valeur propre de la matrice  $X'X$ .

On a

$$X'X = V\Lambda V'$$

où  $V$  est la matrice orthogonale des vecteurs propres de  $X'X$  et  $\Lambda$  désigne la matrice diagonale des valeurs propres  $\lambda$  de  $X'X$  car  $X'X$  est une matrice carrée et symétrique ( $(X'X) = (X'X)'$ ) et on peut donc la diagonaliser.

De ce fait, on peut aussi écrire

$$X_{(T,k)} = UDV'$$

où  $U$  est telle que  $U'U = I$  et  $D$  est la matrice diagonale des valeurs singulières  $d$  de  $X$ .

Si on remplace cette expression dans celle de  $X'X$ , on obtient

$$X'X = (UDV')'(UDV') = VD'U'UDV' = VD'DV' = VD^2V'.$$

En procédant par identification, on obtient alors

$$D^2 = \Lambda.$$

On a bien

$$d_i = \sqrt{\lambda_i}.$$

L'indice de conditionnement va nous indiquer si nous avons ou non une relation entre les variables explicatives.

Belsley Kuh Welsch ont défini plusieurs indices pour repérer plusieurs relations de colinéarité :

Pour  $d_0 > d_1 > \dots > d_{k-1}$ , on a

$$I_0 = \frac{d_0}{d_0} = 1, \quad I_1 = \frac{d_0}{d_1} > I_0, \quad I_2 = \frac{d_0}{d_2} > I_1, \quad \dots, \quad I_{k-1} = \frac{d_0}{d_{k-1}} = \frac{d_{\max}}{d_{\min}}.$$

On commence par l'étude de l'indice le plus grand qui est  $I_{k-1}$ .

Si on a une relation de colinéarité, on a la valeur propre  $\lambda_i \rightarrow 0$ ,  $d_i \rightarrow 0$  et  $I_i = \frac{d_0}{d_i}$  grand.

Si  $I_{k-1} < 30$ , il n'y a pas de problème de colinéarité

Si  $30 < I_{k-1} < 100$ , il y a un problème moyen de colinéarité (non négligeable)

Si  $I_{k-1} > 100$ , il y a un gros problème de colinéarité.

Dans le cas d'une colinéarité ( $I_{k-1} > 30$ ), on analyse les autres indices pour savoir s'il y a une autre relation de colinéarité.

S'il y a colinéarité, on regardera ensuite le tableau de décomposition de la variance pour déterminer les variables qui sont liées entre elles (pour une part de la variance > 0,4 ⇒ la variable est responsable de la colinéarité). On analyse le tableau de décomposition de la variance que si on a détecté une colinéarité en regardant les indices I.

Tableau de décomposition de la variance :

Valeurs singulières	$V(\hat{a}_0)$	$V(\hat{a}_1)$	...	$V(\hat{a}_{(k-1)})$	indices
$d_0$	$\Pi_{00}$	$\Pi_{10}$			$I_0 = \frac{d_0}{d_0} = 1$
$d_1$	$\Pi_{01}$	$\Pi_{11}$			$I_1 = \frac{d_0}{d_1}$
...	...	...			...
$d_{k-1}$	$\Pi_{0(k-1)}$	$\Pi_{1(k-1)}$			$I_{k-1} = \frac{d_0}{d_{k-1}}$

$$\text{On a } V_{\hat{a}} = \sigma_{\varepsilon}^2 (X'X)^{-1} = \sigma_{\varepsilon}^2 VD^{-2}V' \Rightarrow V(\hat{a}_1) = \sigma_{\varepsilon}^2 \sum_{n=0}^{k-1} \frac{v_{in}^2}{d_n^2} = \sigma_{\varepsilon}^2 \sum_{n=0}^{k-1} \varphi_{in}$$

où  $\sum_{n=0}^{k-1} \frac{v_{in}^2}{d_n^2}$  = somme de tous les éléments de  $n=0, \dots, k-1$  de la ligne  $i$  de la matrice  $VD^{-2}V'$ .

$$\text{On a } \Pi_{in} = \frac{\varphi_{in}}{\varphi_i} \quad \text{où } \varphi_i = \sum_{n=0}^{k-1} \varphi_{in} \quad \text{pour } i = 0, \dots, k-1.$$

$$\text{On a } V_{\hat{a}} = \sigma_{\varepsilon}^2 (X'X)^{-1} = \sigma_{\varepsilon}^2 VD^{-2}V'$$

$$\text{car } X'X = VD^2V'.$$

$$\text{D'où } (X'X)^{-1} = (VD^2V')^{-1} = (V')^{-1}D^{-2}V^{-1}.$$

Or  $V$  est une matrice orthogonale d'où  $V' = V^{-1}$  et on a  $(X'X)^{-1} = VD^{-2}V'$ .

Comment remédier à la colinéarité ?

On regroupe les variables colinéaires pour ne former qu'une seule variable :

- On peut faire une combinaison linéaire :  $y_t = a_0 + a_1(x_t + z_t)$
- On peut faire une différence première :  $y_t = a_0 + a_1(x_t - x_{t-1})$

On peut aussi supprimer une (ou plusieurs) des variables colinéaires.

## V/ VIOLATION DES HYPOTHESES

Lorsque les erreurs du modèle de base ne respectent pas les hypothèses, notamment lorsqu'elles sont :

- (i) autocorrélées ;
- (ii) hétéroscédastiques ;

les estimateurs des coefficients à partir de la méthode des MCO n'ont pas les bonnes propriétés (estimateurs de variance non minimale).

C'est pourquoi en présence d'une autocorrélation et/ou d'une hétéroscédasticité des erreurs, il faut recourir à une autre méthode d'estimation qui est celle des Moindres Carrés Généralisés (MCG).

Cette méthode est à utiliser seulement après avoir vérifié que le non respect des hypothèses ne provient pas des causes citées plus haut (absence d'une variable explicative importante, présence d'un point aberrant,...).

### V.1/ Méthode des Moindres Carrés Généralisés (MCG)

Soit le modèle linéaire générale suivant :

$$Y_{(T,1)} = X_{(T,k)} a_{(k,1)} + \varepsilon_{(T,1)}$$

où la matrice des variances-covariances s'écrit

$$V_{\varepsilon} = E(\varepsilon \varepsilon') = \begin{pmatrix} \text{Var}(\varepsilon_1) & \text{Cov}(\varepsilon_1, \varepsilon_2) & \dots & \text{Cov}(\varepsilon_1, \varepsilon_T) \\ \text{Cov}(\varepsilon_2, \varepsilon_1) & \text{Var}(\varepsilon_2) & \dots & \text{Cov}(\varepsilon_2, \varepsilon_T) \\ \dots & \dots & \dots & \dots \\ \text{Cov}(\varepsilon_T, \varepsilon_1) & \dots & \dots & \text{Var}(\varepsilon_T) \end{pmatrix} = \begin{pmatrix} E(\varepsilon_1 \varepsilon_1) & E(\varepsilon_1 \varepsilon_2) & \dots & E(\varepsilon_1 \varepsilon_T) \\ E(\varepsilon_2 \varepsilon_1) & E(\varepsilon_2 \varepsilon_2) & \dots & E(\varepsilon_2 \varepsilon_T) \\ \dots & \dots & \dots & \dots \\ E(\varepsilon_T \varepsilon_1) & \dots & \dots & E(\varepsilon_T \varepsilon_T) \end{pmatrix}$$

$$V_{\varepsilon} \neq \begin{pmatrix} \sigma_{\varepsilon}^2 & 0 & \dots & 0 \\ 0 & \sigma_{\varepsilon}^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \sigma_{\varepsilon}^2 \end{pmatrix} = \sigma_{\varepsilon}^2 I \quad \text{lorsqu'il y a autocorrélation ou/et hétéroscédasticité}$$

des erreurs  $\varepsilon$ .

L'estimateur des MCG est donné par :

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

Cet estimateur est appelé aussi estimateur de Aitken.

On remarque que lorsque les hypothèses sont vérifiées, on retrouve l'estimateur des MCO :

$$\hat{a} = (X'V_{\varepsilon}^{-1}X)^{-1} (X'V_{\varepsilon}^{-1}Y) = \left( X' \left( \frac{1}{\sigma_{\varepsilon}^2} I \right) X \right)^{-1} \left( X' \left( \frac{1}{\sigma_{\varepsilon}^2} I \right) Y \right) = (X'X)^{-1} X'Y .$$

Les propriétés de cet estimateur sont les suivantes :

- Il est biaisé mais asymptotiquement sans biais (si T est grand).

En effet, on a

$$\begin{aligned} \hat{a} &= (X'V_{\varepsilon}^{-1}X)^{-1} (X'V_{\varepsilon}^{-1}Y) \\ &= (X'V_{\varepsilon}^{-1}X)^{-1} X'V_{\varepsilon}^{-1}(Xa + \varepsilon) \\ &= (X'V_{\varepsilon}^{-1}X)^{-1} X'V_{\varepsilon}^{-1}Xa + (X'V_{\varepsilon}^{-1}X)^{-1} X'V_{\varepsilon}^{-1} \varepsilon \\ &= Ia + (X'V_{\varepsilon}^{-1}X)^{-1} X'V_{\varepsilon}^{-1} \varepsilon \end{aligned}$$

$$D'où E(\hat{a}) = E(a + (X'V_{\varepsilon}^{-1}X)^{-1} X'V_{\varepsilon}^{-1} \varepsilon) = a + E((X'V_{\varepsilon}^{-1}X)^{-1} X'V_{\varepsilon}^{-1} \varepsilon).$$

Comme  $V_{\varepsilon} \neq \sigma_{\varepsilon}^2 I$ , on ne peut pas le sortir des parenthèses et on a donc :

$$E((X'V_{\varepsilon}^{-1}X)^{-1} X'V_{\varepsilon}^{-1} \varepsilon) \neq 0 \quad \text{et} \quad E(\hat{a}) \neq a.$$

- Il a une variance minimale.

## V.2/ Autocorrélation des erreurs et MCG

Lorsque l'on a une autocorrélation des erreurs, on a  $cov(\varepsilon_t, \varepsilon_{t'}) \neq 0$  pour  $t \neq t'$ .

Supposons que l'on ait une autocorrélation des erreurs d'ordre 1, le modèle linéaire va s'écrire :

$$Y = Xa + \varepsilon$$

avec  $\varepsilon_t = \rho \varepsilon_{t-1} + v_t \quad |\rho| < 1$ ,  $v_t \sim N(0, \sigma_v^2)$  et  $cov(v_t, v_{t'}) = 0$  pour  $t \neq t'$ .

On a

$$\text{Var}(\varepsilon_t) = \sigma_{\varepsilon}^2 = \text{Var}(\rho\varepsilon_{t-1} + v_t) = \rho^2\text{Var}(\varepsilon_{t-1}) + \text{Var}(v_t) + 2 \text{cov}(\rho\varepsilon_{t-1}, v_t).$$

Or ces éléments peuvent s'écrire :

$$\text{Var}(\varepsilon_{t-1}) = \sigma_{\varepsilon}^2, \quad \text{Var}(v_t) = \sigma_v^2 \quad \text{et} \quad \text{cov}(\varepsilon_{t-1}, v_t) = E(\varepsilon_{t-1} \cdot v_t) - E(\varepsilon_{t-1}) \cdot E(v_t) = 0$$

car  $E(\varepsilon_t) = 0 \quad \forall t$ ,  $E(v_t) = 0 \quad \forall t$  et  $E(\varepsilon_{t-1} \cdot v_t) = E(\varepsilon_{t-1}) \cdot E(v_t)$  du fait que  $\varepsilon_{t-1}$  et  $v_t$  sont indépendants.

D'où

$$\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2 = \rho^2 \sigma_\varepsilon^2 + \sigma_v^2$$

et on a  $\sigma_\varepsilon^2 (1 - \rho^2) = \sigma_v^2$  et donc  $\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2 = \frac{\sigma_v^2}{1 - \rho^2}$ .

On remarque que :

$$\begin{aligned} \varepsilon_t &= \rho \varepsilon_{t-1} + v_t \\ &= \rho(\rho \varepsilon_{t-2} + v_{t-1}) + v_t = \rho^2 \varepsilon_{t-2} + \rho v_{t-1} + v_t \quad \text{car on a } \varepsilon_{t-1} = \rho \varepsilon_{t-2} + v_{t-1} \\ &= \rho^2(\rho \varepsilon_{t-3} + v_{t-2}) + \rho v_{t-1} + v_t = \rho^3 \varepsilon_{t-3} + \rho^2 v_{t-2} + \rho v_{t-1} + v_t \end{aligned}$$

...

$$\varepsilon_t = \rho^s \varepsilon_{t-s} + \sum_{i=0}^{s-1} \rho^i v_{t-i}.$$

De plus, on observe que :

$$\text{Cov}(\varepsilon_t, \varepsilon_{t-s}) = E(\varepsilon_t \cdot \varepsilon_{t-s}) - E(\varepsilon_t) \cdot E(\varepsilon_{t-s}) = E(\varepsilon_t \cdot \varepsilon_{t-s}) \quad \text{car on a } E(\varepsilon_t) = 0 \quad \forall t.$$

D'où

$$\begin{aligned} \text{Cov}(\varepsilon_t, \varepsilon_{t-s}) &= E(\varepsilon_t \cdot \varepsilon_{t-s}) = E\left[\left(\rho^s \varepsilon_{t-s} + \sum_{i=0}^{s-1} \rho^i v_{t-i}\right) \varepsilon_{t-s}\right] = \rho^s E(\varepsilon_{t-s}^2) + \sum_{i=0}^{s-1} \rho^i E(v_{t-i} \cdot \varepsilon_{t-s}) \\ &= \rho^s \sigma_\varepsilon^2 = \frac{\rho^s \sigma_v^2}{1 - \rho^2} \end{aligned}$$

car on a  $E(\varepsilon_t) = 0 \quad \forall t$  et donc  $E(\varepsilon_{t-s}^2) = E(\varepsilon_{t-s}^2) - (E(\varepsilon_{t-s}))^2 = \text{Var}(\varepsilon_{t-s}) = \sigma_\varepsilon^2$  et  $E(v_{t-i} \cdot \varepsilon_{t-s}) = E(v_{t-i}) \cdot E(\varepsilon_{t-s}) = 0$  car  $E(\varepsilon_t) = 0 \quad \forall t$  et  $v_t$  et  $\varepsilon_t$  sont indépendants.

La matrice  $V_\varepsilon$  des variances-covariances va alors s'écrire :

$$\begin{aligned} V_\varepsilon &= \begin{pmatrix} \text{Var}(\varepsilon_1) & \text{Cov}(\varepsilon_1, \varepsilon_2) & \dots & \text{Cov}(\varepsilon_1, \varepsilon_T) \\ \text{Cov}(\varepsilon_2, \varepsilon_1) & \text{Var}(\varepsilon_2) & \dots & \text{Cov}(\varepsilon_2, \varepsilon_T) \\ \dots & \dots & \dots & \dots \\ \text{Cov}(\varepsilon_T, \varepsilon_1) & \dots & \dots & \text{Var}(\varepsilon_T) \end{pmatrix} = \begin{pmatrix} \sigma_\varepsilon^2 & \text{Cov}(\varepsilon_1, \varepsilon_2) & \dots & \text{Cov}(\varepsilon_1, \varepsilon_T) \\ \text{Cov}(\varepsilon_2, \varepsilon_1) & \sigma_\varepsilon^2 & \dots & \text{Cov}(\varepsilon_2, \varepsilon_T) \\ \dots & \dots & \dots & \dots \\ \text{Cov}(\varepsilon_T, \varepsilon_1) & \dots & \dots & \sigma_\varepsilon^2 \end{pmatrix} \\ &= \begin{pmatrix} \sigma_\varepsilon^2 & \sigma_\varepsilon^2 \rho & \dots & \sigma_\varepsilon^2 \rho^{T-1} \\ \sigma_\varepsilon^2 \rho & \sigma_\varepsilon^2 & \dots & \sigma_\varepsilon^2 \rho^{T-2} \\ \dots & \dots & \dots & \dots \\ \sigma_\varepsilon^2 \rho^{T-1} & \dots & \dots & \sigma_\varepsilon^2 \end{pmatrix} = \frac{\sigma_v^2}{1 - \rho^2} \begin{pmatrix} 1 & \rho & \dots & \rho^{T-1} \\ \rho & 1 & \dots & \rho^{T-2} \\ \dots & \dots & \dots & \dots \\ \rho^{T-1} & \dots & \dots & 1 \end{pmatrix} \neq \begin{pmatrix} \sigma_\varepsilon^2 & 0 & \dots & 0 \\ 0 & \sigma_\varepsilon^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \sigma_\varepsilon^2 \end{pmatrix} = \sigma_\varepsilon^2 I. \end{aligned}$$

Lorsque l'on utilise la méthode des MCO sur un modèle qui comporte une autocorrélation des erreurs d'ordre 1, on a :

- Un estimateur  $\hat{a} = (X'X)^{-1} X'Y$  qui est toujours sans biais :  $E(\hat{a}) = a$
- Un estimateur  $\hat{a} = (X'X)^{-1} X'Y$  qui n'a plus la plus petite variance.

On va estimer un modèle qui comporte une autocorrélation d'ordre 1 à l'aide de la méthode des MCG car on a vu que l'estimateur des MCG était non biaisé asymptotiquement et avait la plus petite variance.

Il y a deux façons d'utiliser la méthode des MCG :

1/ On applique directement la méthode des MCG au modèle de base :

On a le modèle de base suivant :

$$Y = Xa + \varepsilon$$

avec  $E(\varepsilon \varepsilon') = V_\varepsilon \neq \sigma_\varepsilon^2 I$ .

L'estimateur des MCG s'écrit :

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

$$\text{avec } V_\varepsilon = \sigma_\varepsilon^2 \begin{pmatrix} 1 & \rho & \dots & \rho^{T-1} \\ \rho & 1 & \dots & \rho^{T-2} \\ \dots & \dots & \dots & \dots \\ \rho^{T-1} & \dots & \dots & 1 \end{pmatrix} = \frac{\sigma_\varepsilon^2}{1-\rho^2} \begin{pmatrix} 1 & \rho & \dots & \rho^{T-1} \\ \rho & 1 & \dots & \rho^{T-2} \\ \dots & \dots & \dots & \dots \\ \rho^{T-1} & \dots & \dots & 1 \end{pmatrix}$$

$$\text{et } V_\varepsilon^{-1} = \frac{1}{\sigma_\varepsilon^2} \begin{pmatrix} 1 & -\rho & \dots & 0 & 0 \\ -\rho & 1+\rho^2 & \dots & 0 & 0 \\ 0 & -\rho & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 1+\rho^2 & -\rho \\ 0 & 0 & \dots & -\rho & 1 \end{pmatrix}.$$

2/ On transforme le modèle de base et on applique la méthode des MCO. On aboutit à un résultat équivalent à celui donné en 1/.

On recherche une transformation matricielle P telle que le modèle  $PY = PXa + P\varepsilon$  ait ses erreurs non corrélées, c'est-à-dire :

$$\begin{aligned} E((P\varepsilon)(P\varepsilon)') &= E(P\varepsilon \varepsilon' P') = PE(\varepsilon \varepsilon')P' = PV_\varepsilon P' = P\sigma_\varepsilon^2 MP' = \sigma_\varepsilon^2 PMP' = \sigma_\varepsilon^2 P(P^{-1}(P')^{-1})P' \\ &= \sigma_\varepsilon^2 P(P'P)^{-1}P' = \sigma_\varepsilon^2 I. \end{aligned}$$

Dans ce cas, on peut estimer le modèle par la méthode des MCO :

$$\hat{a} = ((PX)'PX)^{-1}(PX)'PY = (X'P'PX)^{-1}X'P'PY.$$

En comparant l'estimateur des MCO avec celui des MCG, on constate que:

$$V_\varepsilon^{-1} = P'P \text{ et } V_\varepsilon = \sigma_\varepsilon^2 M = \sigma_\varepsilon^2 (P'P)^{-1}.$$



Comme on sait que

$$V_{\varepsilon} = \sigma_{\varepsilon}^2 \begin{pmatrix} 1 & \rho & \dots & \rho^{T-1} \\ \rho & 1 & \dots & \rho^{T-2} \\ \dots & \dots & \dots & \dots \\ \rho^{T-1} & \dots & \dots & 1 \end{pmatrix} = \frac{\sigma_v^2}{1-\rho^2} \begin{pmatrix} 1 & \rho & \dots & \rho^{T-1} \\ \rho & 1 & \dots & \rho^{T-2} \\ \dots & \dots & \dots & \dots \\ \rho^{T-1} & \dots & \dots & 1 \end{pmatrix} = \sigma_v^2 \begin{pmatrix} \frac{1}{1-\rho^2} & \rho & \dots & \rho^{T-1} \\ \rho & \frac{1}{1-\rho^2} & \dots & \rho^{T-2} \\ \dots & \dots & \dots & \dots \\ \rho^{T-1} & \dots & \dots & \frac{1}{1-\rho^2} \end{pmatrix}$$

$$\text{et } V_{\varepsilon}^{-1} = \frac{1}{\sigma_v^2} \begin{pmatrix} 1 & -\rho & \dots & 0 & 0 \\ -\rho & 1+\rho^2 & \dots & 0 & 0 \\ 0 & -\rho & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1+\rho^2 & -\rho \\ 0 & 0 & \dots & -\rho & 1 \end{pmatrix} = P'P.$$

Or on peut également écrire  $V_{\varepsilon}^{-1} = \frac{1}{\sigma_v^2} P'P$  puisqu'on a :

$$\hat{a} = \left[ X' \left( \frac{1}{\sigma_v^2} P'P \right) X \right]^{-1} X' \left( \frac{1}{\sigma_v^2} P'P \right) Y = (X'P'PX)^{-1} X'P'PY.$$

On peut alors déduire que :

$$P = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 & 0 \\ -\rho & 1 & 0 & \dots & 0 & 0 \\ 0 & -\rho & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{pmatrix} \quad \text{et} \quad P' = \begin{pmatrix} \sqrt{1-\rho^2} & -\rho & 0 & \dots & \dots & 0 \\ 0 & 1 & -\rho & \dots & \dots & 0 \\ 0 & 0 & 1 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \dots & 1 \end{pmatrix}.$$

On transforme alors le modèle de base en le multipliant par la matrice P :

$$PY = PXa + P\varepsilon$$

$$\text{où } P = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 & 0 \\ -\rho & 1 & 0 & \dots & 0 & 0 \\ 0 & -\rho & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{pmatrix}.$$

On obtient alors :

$$\begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 & 0 \\ -\rho & 1 & 0 & \dots & 0 & 0 \\ 0 & -\rho & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} \\ = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 & 0 \\ -\rho & 1 & 0 & \dots & 0 & 0 \\ 0 & -\rho & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{pmatrix} \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{(k-1)1} \\ 1 & x_{12} & x_{22} & \dots & x_{(k-1)2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1T} & x_{2T} & \dots & x_{(k-1)T} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{k-1} \end{pmatrix} + \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 & 0 \\ -\rho & 1 & 0 & \dots & 0 & 0 \\ 0 & -\rho & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix}$$

D'où

$$\begin{pmatrix} (\sqrt{1-\rho^2}) y_1 \\ y_2 - \rho y_1 \\ \vdots \\ y_n - \rho y_{n-1} \end{pmatrix} = \begin{pmatrix} \sqrt{1-\rho^2} & (\sqrt{1-\rho^2}) x_{11} & \dots \\ 1-\rho & x_{21} - \rho x_{11} & \dots \\ \vdots & \vdots & \ddots \\ 1-\rho & x_{n1} - \rho x_{(n-1)1} & \dots \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{k-1} \end{pmatrix} + \begin{pmatrix} (\sqrt{1-\rho^2}) \varepsilon_1 \\ \varepsilon_2 - \rho \varepsilon_1 \\ \vdots \\ \varepsilon_n - \rho \varepsilon_{n-1} \end{pmatrix}$$

On obtient l'équation suivante :

$$y_t - \rho y_{t-1} = a_0(1-\rho) + a_1(x_{1t} - \rho x_{1(t-1)}) + \dots + a_{k-1}(x_{(k-1)t} - \rho x_{(k-1)(t-1)}) + (\varepsilon_t - \rho \varepsilon_{t-1})$$

$$\Leftrightarrow dy_t = b_0 + a_1 dx_{1t} + \dots + a_{k-1} dx_{(k-1)t} + v_t$$

car on a  $\varepsilon_t = \rho \varepsilon_{t-1} + v_t$   $|\rho| < 1$ ,  $v_t \sim N(0, \sigma_v^2)$ .

Le terme aléatoire  $v_t$  répond aux hypothèses de base de la méthode des MCO, nous pouvons donc utiliser la méthode des MCO pour estimer ce modèle transformé.

On remarque que pour estimer le modèle transformé, il nous faut connaître  $\rho$ .

Plusieurs procédures permettent d'estimer à la fois  $\rho$  et les autres paramètres du modèle transformé : estimation directe de  $\rho$ , estimation itérative de  $\rho$  et des autres paramètres (méthode de Cochrane-Orcutt), méthode du balayage (méthode de Hildreth-Lu) et méthode du maximum de vraisemblance. La méthode de Cochrane-Orcutt est la plus répandue (dans les logiciels) et la méthode de Hildreth-Lu donne des résultats relativement similaires à ceux de la méthode de Cochrane-Orcutt.

### Estimation itérative de $\rho$ et des autres paramètres : méthode de Cochrane-Orcutt :

Soit le modèle de base suivant :

$$Y = Xa + \varepsilon$$

on estime ce modèle par la méthode des MCO et on constate une autocorrélation des erreurs d'ordre 1 en analysant la statistique de DW. On envisage alors d'estimer le modèle transformé :

$$PY = PXa + P\varepsilon$$

$$\text{où } P = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 & 0 \\ -\rho & 1 & 0 & \dots & 0 & 0 \\ 0 & -\rho & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{pmatrix}$$

afin d'éliminer l'autocorrélation des erreurs. On aboutit à l'équation suivante à estimer :

$$y_t - \rho y_{t-1} = a_0(1-\rho) + a_1(x_{1t} - \rho x_{1(t-1)}) + \dots + a_{k-1}(x_{(k-1)t} - \rho x_{(k-1)(t-1)}) + (\varepsilon_t - \rho \varepsilon_{t-1})$$

On va utiliser la méthode de Cochrane-Orcutt :

#### **Etape 1 :** initialisation de $\rho$

Détermination de la première valeur de  $\rho$

- par une régression directe des résidus  $e_t$  sur  $e_{t-1}$  du modèle de base :

$$\rho = \frac{\sum_{t=2}^T e_t e_{t-1}}{\sum_{t=1}^T e_t^2} \quad \text{avec } e = Y - Xa$$

- à l'aide de la statistique de DW donnée suite à la régression du modèle de base :

$$\rho = 1 - \frac{DW}{2}$$

Soit

$$\rho_0 = \rho$$

#### **Etape 2 :** régression sur les quasi-différences à l'aide de la méthode des MCO

$$y_t - \rho_0 y_{t-1} = a_0(1-\rho_0) + a_1(x_{1t} - \rho_0 x_{1(t-1)}) + \dots + a_{k-1}(x_{(k-1)t} - \rho_0 x_{(k-1)(t-1)}) + v_t$$

Les paramètres estimés sont alors  $a_1, a_2, \dots, a_{k-1}$  et  $a_0 = \frac{b_0}{(1-\rho_0)}$ .

**Etape 3** : réestimation de  $\rho$

A partir des nouveaux résidus d'estimation  $e_t^1$  nous recalculons une nouvelle valeur de  $\rho$  soit  $\rho_1$  :

$$e_t^1 = y_t - a_0 - a_1 x_{1t} - \dots - a_{k-1} x_{(k-1)t}$$

et

$$\rho_1 = \frac{\sum_{t=2}^T e_t^1 e_{t-1}^1}{\sum_{t=1}^T (e_t^1)^2}$$

**Etape 4** : régression sur les quasi-différences par la méthode des MCO :

$$y_t - \rho_1 y_{t-1} = a_0(1-\rho_1) + a_1(x_{1t} - \rho_1 x_{1(t-1)}) + \dots + a_{k-1}(x_{(k-1)t} - \rho_1 x_{(k-1)(t-1)}) + v_t.$$

Puis nous calculons un nouveau résidu  $e_t^2$  à partir de la nouvelle estimation des coefficients, ce qui nous permet d'obtenir un  $\rho_2$ . On réitère ainsi de suite l'opération jusqu'à la stabilité des coefficients estimés.

Remarque :

La méthode des MCG ne permet de traiter que l'autocorrélation des erreurs d'ordre 1. Pour les autocorrélations d'un ordre supérieur, on résout le problème en utilisant les modèles autorégressifs.

**V.3/ Hétéroscédasticité et MCG**

Lorsque l'on a une hétéroscédasticité des erreurs, on a  $\text{Var}(\varepsilon_t) \neq \sigma_\varepsilon^2 \forall t \Rightarrow$  la variance de l'erreur n'est plus constante.

Le modèle linéaire va alors s'écrire :

$$Y = Xa + \varepsilon$$

avec  $V_\varepsilon = E(\varepsilon \varepsilon')$  =

$$\begin{pmatrix} \text{Var}(\varepsilon_1) & \text{Cov}(\varepsilon_1, \varepsilon_2) & \dots & \text{Cov}(\varepsilon_1, \varepsilon_T) \\ \text{Cov}(\varepsilon_2, \varepsilon_1) & \text{Var}(\varepsilon_2) & \dots & \text{Cov}(\varepsilon_2, \varepsilon_T) \\ \dots & \dots & \dots & \dots \\ \text{Cov}(\varepsilon_T, \varepsilon_1) & \dots & \dots & \text{Var}(\varepsilon_T) \end{pmatrix} = \begin{pmatrix} \sigma_{\varepsilon,1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{\varepsilon,2}^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \sigma_{\varepsilon,T}^2 \end{pmatrix}$$

$$\neq \begin{pmatrix} \sigma_{\varepsilon}^2 & 0 & \dots & 0 \\ 0 & \sigma_{\varepsilon}^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \sigma_{\varepsilon}^2 \end{pmatrix} = \sigma_{\varepsilon}^2 I.$$

On suppose que la variance de l'erreur est liée aux valeurs de la variable explicative  $x_{1i}$  :

$$\sigma_{\varepsilon,i}^2 = \sigma_{\varepsilon}^2 \cdot x_{1i} \quad \text{où } i = 1, 2, \dots, T.$$

On a alors :

$$V_{\varepsilon} = E(\varepsilon \varepsilon') = \sigma_{\varepsilon}^2 \begin{pmatrix} x_{11} & 0 & \dots & 0 \\ 0 & x_{12} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & x_{1T} \end{pmatrix}.$$

Lorsque l'on utilise la méthode des MCO sur un modèle qui comporte de l'hétéroscédasticité, on a les mêmes conséquences qu'avec le modèle comportant une autocorrélation des erreurs, c'est-à-dire :

- L'estimateur  $\hat{a} = (X'X)^{-1} X'Y$  est toujours sans biais :  $E(\hat{a}) = a$
- L'estimateur  $\hat{a} = (X'X)^{-1} X'Y$  n'a plus la plus petite variance.

On va donc utiliser la méthode des MCG pour estimer le modèle comportant une hétéroscédasticité.

Il y a deux façons d'utiliser la méthode des MCG :

1/ On applique directement la méthode des MCG au modèle initial :

On a le modèle initial suivant :

$$Y = Xa + \varepsilon$$

avec  $E(\varepsilon \varepsilon') = V_{\varepsilon} \neq \sigma_{\varepsilon}^2 I$ .

L'estimateur des MCG s'écrit :

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

$$\text{avec } V_{\varepsilon} = E(\varepsilon \varepsilon') = \sigma_{\varepsilon}^2 \begin{pmatrix} x_{11} & 0 & \dots & 0 \\ 0 & x_{12} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & x_{1T} \end{pmatrix} \quad \text{et} \quad V_{\varepsilon}^{-1} = \frac{1}{\sigma_{\varepsilon}^2} \begin{pmatrix} \frac{1}{x_{11}} & 0 & \dots & 0 \\ 0 & \frac{1}{x_{12}} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \frac{1}{x_{1T}} \end{pmatrix}$$

2/ On transforme le modèle originel et on applique la méthode des MCO. On aboutit à un résultat équivalent à celui donné en 1/.

On recherche une transformation matricielle P telle que le modèle  $PY = PXa + P\varepsilon$  ait une variance des erreurs qui est constante, c'est-à-dire :

$$E((P\varepsilon)(P\varepsilon)') = E(P\varepsilon \varepsilon' P') = PE(\varepsilon \varepsilon')P' = PV_{\varepsilon}P' = P\sigma_{\varepsilon}^2 MP' = \sigma_{\varepsilon}^2 PMP' = \sigma_{\varepsilon}^2 P(P^{-1}(P')^{-1})P' = \sigma_{\varepsilon}^2 P(P'P)^{-1}P' = \sigma_{\varepsilon}^2 I.$$

Dans ce cas, on peut estimer le modèle par la méthode des MCO :

$$\hat{a} = ((PX)'PX)^{-1}(PX)'PY = (X'P'PX)^{-1}X'P'PY.$$

En comparant l'estimateur des MCO avec celui des MCG, on a :

$$V_{\varepsilon}^{-1} = P'P \text{ et } V_{\varepsilon} = \sigma_{\varepsilon}^2 M = \sigma_{\varepsilon}^2 (P'P)^{-1}.$$

$$\text{Or on a } V_{\varepsilon} = E(\varepsilon \varepsilon') = \sigma_{\varepsilon}^2 \begin{pmatrix} x_{11} & 0 & \dots & 0 \\ 0 & x_{12} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & x_{1T} \end{pmatrix} = \sigma_{\varepsilon}^2 M = \sigma_{\varepsilon}^2 (P'P)^{-1}$$

$$\text{et } V_{\varepsilon}^{-1} = \frac{1}{\sigma_{\varepsilon}^2} \begin{pmatrix} \frac{1}{x_{11}} & 0 & \dots & 0 \\ 0 & \frac{1}{x_{12}} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \frac{1}{x_{1T}} \end{pmatrix} = P'P.$$

Similairement au cas de l'autocorrélation des erreurs, on peut écrire également  $V_{\varepsilon}^{-1} = \frac{1}{\sigma_{\varepsilon}^2} P'P$ .

On déduit alors :

$$P = \begin{pmatrix} \frac{1}{\sqrt{x_{11}}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{x_{12}}} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \frac{1}{\sqrt{x_{1T}}} \end{pmatrix} \text{ et } P' = \begin{pmatrix} \frac{1}{\sqrt{x_{11}}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{x_{12}}} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \frac{1}{\sqrt{x_{1T}}} \end{pmatrix}.$$

On peut transformer le modèle de base en le multipliant par la matrice P :

$$PY = PXa + P\varepsilon .$$

On obtient alors :

$$\begin{pmatrix} \frac{1}{\sqrt{x_{11}}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{x_{12}}} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \frac{1}{\sqrt{x_{1T}}} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{x_{11}}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{x_{12}}} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \frac{1}{\sqrt{x_{1T}}} \end{pmatrix} \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{(k-1)1} \\ 1 & x_{12} & x_{22} & \dots & x_{(k-1)2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1T} & x_{2T} & \dots & x_{(k-1)T} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{k-1} \end{pmatrix} + \begin{pmatrix} \frac{1}{\sqrt{x_{11}}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{x_{12}}} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \frac{1}{\sqrt{x_{1T}}} \end{pmatrix} \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix}$$

D'où

$$\begin{pmatrix} \frac{y_1}{\sqrt{x_{11}}} \\ \frac{y_2}{\sqrt{x_{12}}} \\ \vdots \\ \frac{y_T}{\sqrt{x_{1T}}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{x_{11}}} & \frac{x_{11}}{\sqrt{x_{11}}} & \dots & \frac{x_{(k-1)1}}{\sqrt{x_{11}}} \\ \frac{1}{\sqrt{x_{12}}} & \frac{x_{12}}{\sqrt{x_{12}}} & \dots & \frac{x_{(k-1)2}}{\sqrt{x_{12}}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{\sqrt{x_{1T}}} & \frac{x_{1T}}{\sqrt{x_{1T}}} & \dots & \frac{x_{(k-1)T}}{\sqrt{x_{1T}}} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{k-1} \end{pmatrix} + \begin{pmatrix} \frac{\varepsilon_1}{\sqrt{x_{11}}} \\ \frac{\varepsilon_2}{\sqrt{x_{12}}} \\ \vdots \\ \frac{\varepsilon_T}{\sqrt{x_{1T}}} \end{pmatrix} .$$

D'où l'équation suivante :

$$\frac{y_t}{\sqrt{x_{1t}}} = a_0 \frac{1}{\sqrt{x_{1t}}} + a_1 \frac{x_{1t}}{\sqrt{x_{1t}}} + a_2 \frac{x_{2t}}{\sqrt{x_{1t}}} + \dots + a_{k-1} \frac{x_{(k-1)t}}{\sqrt{x_{1t}}} + \frac{\varepsilon_t}{\sqrt{x_{1t}}} .$$

On remarque que  $\text{Var}\left(\frac{\varepsilon_t}{\sqrt{x_{1t}}}\right) = \frac{1}{x_{1t}} \text{Var}(\varepsilon_t) = \frac{1}{x_{1t}} \sigma_{\varepsilon,t}^2 = \frac{1}{x_{1t}} \sigma_{\varepsilon}^2$ .  $x_{1t} = \sigma_{\varepsilon}^2$ .

On peut donc appliquer la méthode des MCO directement sur le modèle transformé pour trouver les valeurs des coefficients  $a_0, a_1, \dots, a_{k-1}$  (puisque le modèle transformé a ses erreurs qui sont homoscédastiques).

## VI/ LES MODELES DYNAMIQUES

### a) Modèle autorégressif :

On a vu que lorsqu'il y avait oublié d'une variable explicative importante dans le modèle de régression, il y avait l'apparition d'une autocorrélation des erreurs.

On a vu également que face à une autocorrélation d'ordre 1, on pouvait traiter le problème en utilisant la méthode d'estimation des MCG.

En présence d'une autocorrélation d'ordre supérieur, nous estimerons le modèle autorégressif suivant :

$$y_t = b_1 y_{t-1} + b_2 y_{t-2} + \dots + b_h y_{t-h} + b_0 + \varepsilon_t$$

ou encore

$$y_t = \sum_{i=1}^h b_i y_{t-i} + b_0 + \varepsilon_t$$

### Détermination du nombre de retards :

Plusieurs critères possibles :

- Critère de Akaike :

On choisit le nombre  $h$  qui minimise la fonction d'Akaike qui est donnée par :

$$AIC(h) = \text{Ln}\left(\frac{SCR_h}{T}\right) + \frac{2h}{T}$$

avec  $SCR_h$  = somme des carrés des résidus pour le modèle à  $h$  retards,  $T$  = nombre d'observations,  $\text{Ln}$  = logarithme népérien.

- Critère de Schwarz :

On choisit le nombre  $h$  qui minimise la fonction de Schwarz qui est donnée par :

$$SC(h) = \text{Ln}\left(\frac{SCR_h}{T}\right) + \frac{h \text{Ln} T}{T}.$$

### Estimation du modèle autorégressif :

On constate ici que l'hypothèse  $H_2$  qui stipule la non dépendance des variables explicatives avec l'erreur  $\varepsilon$  n'est pas satisfaite car les variables  $y_{t-1}, y_{t-2}, \dots, y_{t-h}$  dépendent de  $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-h}$ . Les variables explicatives sont alors aléatoires et les estimateurs des MCO sont biaisés.



Cependant, ils peuvent être asymptotiquement sans biais s'il n'y a pas autocorrélation des erreurs. S'il y a autocorrélation des erreurs, nous pouvons utiliser la méthode d'estimation des MCG (pour une autocorrélation d'ordre 1) et la méthode des variables instrumentales (pour une autocorrélation d'ordre supérieur).

### Test d'autocorrélation d'ordre 1 : le « h » de Durbin :

Soit le modèle autorégressif suivant :

$$y_t = b_1 y_{t-1} + b_0 + \varepsilon_t.$$

Lorsque l'on a un modèle autorégressif, le test de Durbin et Watson est biaisé. Il faut dans ce cas utiliser la statistique de test suivante :

$$h = \rho \sqrt{\frac{T}{1 - Ts_{b_1}^2}}$$

où  $\rho = 1 - \frac{DW}{2}$  (DW est la statistique de Durbin et Watson calculée à partir du modèle autorégressif), T est le nombre d'observations et  $s_{b_1}^2$  est la variance estimée du coefficient estimé  $b_1$  du modèle autorégressif.

Cette statistique de test est appelée aussi le « h » de Durbin.

Les hypothèses du test sont :

$$\begin{cases} H_0 : h = 0 \Leftrightarrow \rho = 0 : \text{pas d'autocorrélation des erreurs} \\ H_1 : h \neq 0 \Leftrightarrow \rho \neq 0 \end{cases}$$

La règle de décision est la suivante :

Si  $|h| \leq t^{\alpha/2}$  où  $t^{\alpha/2}$  représente la valeur critique de la loi normale pour un risque fixé  $\alpha$   
 $\Rightarrow$  on accepte  $H_0$  (indépendance des erreurs).

On remarque que si  $Ts_{b_1}^2 \geq 1$ , on ne peut pas calculer la statistique h. Dans ce cas, on fait comme si l'hypothèse la plus gênante étaient réalisées : l'hypothèse  $H_1$  avec autocorrélation des erreurs et on cherche à estimer ensuite la modèle avec cette autocorrélation des erreurs.

### b) Modèle autorégressif à retards échelonnés :

Si l'autocorrélation des erreurs n'a pas disparu avec le passage du modèle statique au modèle autorégressif, nous pouvons utiliser le modèle autorégressif à retards échelonnés qui s'écrit :

$$y_t = \sum_{i=1}^h b_i y_{t-i} + \sum_{j=1}^k a_{1j} x_{1(t-j)} + \sum_{p=1}^n a_{2p} x_{2(t-p)} + \dots + \varepsilon_t$$

Ce modèle comporte des retards sur la variable endogène et sur toutes les variables exogènes.