

## Chapitre 7 : Les modèles à décalages temporels

### 7.1. Introduction

En économie, la dépendance d'une variable  $Y$  (la variable dépendante) par rapport à une ou plusieurs autres variables  $X$  (les variables explicatives) est rarement instantanée. Très souvent,  $Y$  répond à  $X$  après un certain décalage (laps) de temps. Un tel laps est appelé retard.

Dans l'analyse de régression impliquant des données temporelles, si le modèle de régression comporte non seulement des valeurs présentes mais aussi décalées (passées) des variables explicatives (les  $X$ ), il est appelé *modèle à retards échelonnés* ; s'il comprend une ou plusieurs valeurs de la variable dépendante parmi ses variables explicatives, il est nommé *modèle autorégressif*. D'où :

$y_t = b_0 + a_0x_t + a_1x_{t-1} + a_2x_{t-2} + \varepsilon_t \dots$  (7.1) est un modèle à retards échelonnés,

alors que :

$y_t = b_0 + a_0x_t + a_1y_{t-1} + \varepsilon_t \dots$  (7.2) est un exemple de modèle autorégressif.

Les modèles autorégressifs sont également appelés modèles dynamiques puisqu'ils décrivent le passage temporel (le laps) de la variable dépendante en liaison avec sa (ses) valeur (s) passée (s).

Les modèles autorégressifs et à retards échelonnés sont très utilisés en analyse économétrique vu le rôle central que joue ces retards entre variables en termes de causalité en économie, à court et à long terme.

Parmi les causes des retards, on peut citer des causes psychologiques, technologiques et institutionnelles...etc.

### 7.2. Modèles à retards échelonnés

#### 7.2.1. Formulation générale

Nous postulons que la variable endogène dépend des valeurs prises pour une variable exogène à des périodes antérieures, le modèle est :

$$y_t = b_0 + a_0x_t + a_1x_{t-1} + a_2x_{t-2} + \dots + a_kx_{t-k} + \varepsilon_t \dots \dots \dots (7.1)$$

$$= b_0 + \sum_{j=0}^k a_j x_{t-j} + \varepsilon_t \text{ avec, } a_0 > a_1 > a_2 > \dots > a_k .$$

L'effet de la variable explicative diminue avec le temps :

- lorsque la longueur du retard est spécifiée. On parle de modèle à retards échelonnés *fini*.
- Si, on ne précise pas la longueur du retard, c'est-à-dire de combien de temps on souhaite remonter dans le passé, on parle alors de modèle à retards échelonnés *infini*.

### 7.2.2. Détermination du nombre de retards

Le nombre de retard du modèle général (7.1) est inconnu, il existe des critères statistiques pour le déterminer.

#### a) Test de Fisher

On teste l'hypothèse de nullité des coefficients de régression pour les retards supérieurs à  $k$ .

La formulation des hypothèses est que lorsque l'on teste, d'une manière descendante, une valeur de  $k$  comprise entre 0 et  $M$  :  $0 < k < M$

$$H_0^1: k = M - 1 \rightarrow a_M = 0 \text{ contre } H_1^1: k = M \rightarrow a_M \neq 0$$

$$H_0^2: k = M - 2 \rightarrow a_{M-1} = 0 \text{ contre } H_1^2: k = M - 1 \rightarrow a_{M-1} \neq 0$$

⋮

$$H_0^i: k = M - i \rightarrow a_{M-i+1} = 0 \text{ contre } H_1^i: k = M - i + 1 \rightarrow a_{M-i+1} \neq 0$$

Chacune de ses hypothèses fait l'objet du test de Fisher :

$$F_i^* = \frac{(SCR_{M-i} - SCR_{M-i+1})/1}{(SCR_{M-i+1})/(n - M + i - 3)}$$

Que l'on compare au Fisher lu dans la table à 1 et  $(n - M + i - 3)$  degré de liberté.

Dès que, pour un seuil donné, le  $F$  empirique est supérieur au  $F$  lu, nous rejetons l'hypothèse  $H_0^j$  est la procédure est terminée.

La valeur du retard est à  $M - i + 1$  ;  $k = M - i + 1$

Sachant que chaque décalage entraîne la perte d'une donnée.

#### b) Critère de Akaike (AIC)

Il consiste à retenir comme valeur de  $k$  celle qui minimise la fonction d'Akaike donnée par :

$$AIC(k) = \ln\left(\frac{SCR_k}{n}\right) + \frac{2k}{n}$$

avec :

$SCR_k$  : somme des carrés des résidus pour le modèle à  $k$  retards ;

$n$  : nombre d'observations disponibles (chaque retard entraîne la perte d'une observation) ;

$\ln$  : logarithme népérien.

### c) Critère de Schwarz (SC)

Méthode très proche de la précédente qui consiste à retenir la valeur de  $k$  qui minimise la fonction de Schwarz :

$$SC(k) = \ln\left(\frac{SCR_k}{n}\right) + \frac{k \ln n}{n}$$

## 7.3. Estimation des modèles à retards échelonnés : l'approche de KOYCK

### 7.3.1. Cas de modèle à retards échelonnés infini

#### a) l'approche de KOYCK

Koyck a proposé une méthode d'estimation des modèles à retards échelonnés. Supposons qu'on débute avec le modèle à retards échelonnés infini de la forme :

$$y_t = b_0 + a_0 x_t + a_1 x_{t-1} + a_2 x_{t-2} + \dots + \varepsilon_t \dots \dots \dots (7.2)$$

En supposant que les  $a$  ont tous le même signe, Koyck suppose qu'ils diminuent selon une progression géométrique :

$$a_k = a_0 \lambda^k \text{ avec } k = 0, 1, 2, \dots, \dots \dots \dots (7.3)$$

Où  $\lambda$  ; telle que  $0 < \lambda < 1$ , est appelé *le taux de décroissance* du retard échelonné, et où  $(1 - \lambda)$  est appelé *la vitesse d'ajustement*.

La formule (7.3) suppose que tout coefficient  $a$  successif est numériquement inférieur à chaque  $a$  qui le précède (cette affirmation tient à  $\lambda < 1$ ).

Ce qui implique que lorsqu'on remonte le temps, l'effet du retard sur  $y_t$  devient progressivement plus faible.

La valeur du coefficient du retard  $a_k$ , dépend, outre  $a_0$ , de la valeur de  $\lambda$ .

- Si  $\lambda$  est proche de 1, le taux de décroissance de  $a_k$  est plus lent, les valeurs de  $x_t$  les moins lointaines exerceront une influence notable sur  $y_t$ .
- Si  $\lambda$  est proche de 0, le taux de décroissance de  $a_k$  est plus rapide, l'influence des valeurs de  $x_t$  les moins lointaines s'épuisera rapidement.

#### ➤ Transformation de Koyck

Le modèle à retard infini s'écrit :

$$y_t = b_0 + a_0 x_t + a_1 x_{t-1} + a_2 x_{t-2} + \dots + \varepsilon_t$$

$$\Leftrightarrow y_t = b_0 + a_0 x_t + a_0 \lambda x_{t-1} + a_0 \lambda^2 x_{t-2} + \dots + \varepsilon_t \dots \dots \dots (7.4)$$

Ainsi présenté, le modèle n'est pas prêt pour une estimation aisée puisque il ya un nombre important (infini) de paramètres à estimer, le paramètre  $\lambda$  est sous forme non linéaire, ajoutant à cela, le problème de multicollinéarité entre les variables décalées. Donc, la méthode des MCO ne peut s'appliquer à un tel modèle.

Koyck suggère de décaler l'équation précédente d'une période.

$$y_{t-1} = b_0 + a_0x_{t-1} + a_0\lambda x_{t-2} + a_0\lambda^2 x_{t-3} + \dots + \varepsilon_{t-1} \dots \dots \dots (7.5)$$

On multiplie l'équation précédente par  $\lambda$ , on obtient :

$$\lambda y_{t-1} = \lambda b_0 + a_0\lambda x_{t-1} + a_0\lambda^2 x_{t-2} + a_0\lambda^3 x_{t-3} + \dots + \lambda \varepsilon_{t-1} \dots \dots \dots (7.6)$$

On soustrait (7.6) de (7.4), on obtient :

$$y_t = b_0(1 - \lambda) + a_0x_t + \lambda y_{t-1} + \vartheta_t \dots \dots \dots (7.7)$$

où :  $\vartheta_t = \varepsilon_t - \varepsilon_{t-1}$ . On se ramène vers un modèle appelé, modèle autorégressif.

En comparant (7.7) et (7.2), on observe la simplification opérée par Koyck. Alors qu'auparavant, il fallait estimer  $b_0$  et un nombre infini de  $a$ , il ne faut estimer maintenant que trois inconnus,  $b_0$ ,  $a_0$  et  $\lambda$ , il n ya pas donc de raison à s'attendre à de la multicollinéarité.

### b) Modèle de Solow (distribution de pascal)

Dans ce cas, les coefficients du modèle sont distribués selon :

$$a_i = (1 - \lambda)^{r+1} C_{r+i}^i \lambda^i \text{ où } C_{r+i}^i \text{ est le coefficient du binôme de Newton, } r \text{ et } \lambda \text{ sont deux paramètres avec } 0 < \lambda < 1 \text{ et } r \in N$$

Le modèle général s'écrit ;

$$y_t = b_0 + \sum_{i=0}^{+\infty} (1 - \lambda)^{r+1} C_{r+i}^i \lambda^i x_{t-i} + \varepsilon_t \dots \dots \dots (7.8)$$

### 7.3.2. Cas de modèles à retards échelonnés fini (l'approche d'Almon ou polynomial)

Considérons le modèle à retards échelonnés fini :

$$y_t = b_0 + a_0x_t + a_1x_{t-1} + a_2x_{t-2} + \dots + a_kx_{t-k} + \varepsilon_t$$

L'utilisation des MCO peut se révéler impossible du fait de la multicollinéarité entres variables explicatives décalées. L'utilisation d'un théorème mathématique de Weierstrass, permet à Almon de supposer que  $a_i$  peut être approximé par un polynôme de degré adéquat en  $i$ , la longueur du décalage.

$$a_i = \alpha_0 + \alpha_1 i + \alpha_2 i^2 + \dots + \alpha_m i^m = \sum_{j=0}^m \alpha_j i^j$$

On suppose que  $m$  (le degré du polynôme) est inférieur à  $k$  (la longueur maximale de décalage).

Pour un polynôme de second degré :

$$y_t = b_0 + \sum_{i=0}^k (\alpha_0 + \alpha_1 i + \alpha_2 i^2) x_{t-i} + \varepsilon_t$$

$$= b_0 + \alpha_0 \sum_{i=0}^k x_{t-i} + \alpha_1 \sum_{i=0}^k i x_{t-i} + \alpha_2 \sum_{i=0}^k i^2 x_{t-i} + \varepsilon_t \dots \dots \dots (7.9)$$

Donc, nous avons la séquence des coefficients :

$$i = 0, a_0 = \alpha_0$$

$$i = 1, a_1 = \alpha_0 + \alpha_1 + \alpha_2$$

$$i = 2, a_2 = \alpha_0 + 2\alpha_1 + 4\alpha_2$$

$$i = 3, a_3 = \alpha_0 + 3\alpha_1 + 9\alpha_2$$

$$\Rightarrow \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 2^2 \\ 1 & 3 & 3^2 \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{pmatrix}$$

$$\Rightarrow a = H \times \alpha$$

Pour généraliser cette formule pour  $k$  retards et un polynôme de degré  $m$  :

$$\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ \vdots \\ a_k \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 1 \\ 1 & 2 & 2^2 & \dots & 2^m \\ 1 & 3 & 3^2 & \dots & 3^m \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & k & k^2 & \dots & k^m \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix} \dots \dots \dots (7.10)$$

Le modèle initial  $Y = X a + \varepsilon$  permet d'écrire  $Y = X H \alpha + \varepsilon = Z \alpha + \varepsilon$

avec :  $Z = XH$  représente la matrice des observations des nouvelles variables explicatives.

En définissant :  $Z_{0t} = \sum_{i=0}^k X_{t-i}$ ,  $Z_{1t} = \sum_{i=0}^k i X_{t-i}$ ,  $Z_{2t} = \sum_{i=0}^k i^2 X_{t-i}$

On peut écrire (7.9) comme suit :

$$Y_t = b_0 + \alpha_0 Z_{0t} + \alpha_1 Z_{1t} + \alpha_2 Z_{2t} + \varepsilon_t \dots \dots \dots (7.11)$$

Ainsi, on estime par les MCO l'équation (7.11), une fois estimé les  $\hat{\alpha}$  à partir de (7.11), les  $\hat{a}$  d'origine peuvent être obtenus (estimés) à partir des relations (7.10).

### Exemple :

Une fois spécifier  $m$  et  $k$ , les  $Z$  peuvent être facilement construits. Par exemple si  $m=2$  et  $k=5$ , les  $Z$  sont :

$$Z_{0t} = \sum_{i=0}^5 X_{t-i} = (X_t + X_{t-1} + X_{t-2} + X_{t-3} + X_{t-4} + X_{t-5})$$

$$Z_{1t} = \sum_{i=0}^5 i X_{t-i} = (X_{t-1} + 2X_{t-2} + 3X_{t-3} + 4X_{t-4} + 5X_{t-5})$$

$$Z_{2t} = \sum_{i=0}^5 i^2 X_{t-i} = (X_{t-1} + 4X_{t-2} + 9X_{t-3} + 16X_{t-4} + 25X_{t-5})$$

## 7.4. Les modèles linéaires autorégressifs

### 7.4.1. Formulation générale

Dans ce type de modèles temporels, la variable endogène  $y_t$  dépend de  $k$  variables exogènes  $x_{1t}, x_{2t}, \dots, x_{kt}$  à la période  $t$  et des valeurs que prend la variable  $y_t$  pendant les périodes précédentes  $y_{t-1}, y_{t-2}, \dots, y_{t-h}$ . Soit la formulation :

$$y_t = b_1 y_{t-1} + b_2 y_{t-2} + \dots + b_h y_{t-h} + a_0 + a_1 x_{1t} + a_2 x_{2t} + \dots + a_k x_{kt} + \varepsilon_t \dots \quad (7.12)$$

Dans ce modèle, l'hypothèse d'indépendance entre les variables explicatives et l'erreur, n'est pas satisfaite car les variables  $y_{t-1}, y_{t-2}, \dots, y_{t-h}$  qui dépendent de  $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-h}$  sont aléatoires puisque  $y_{t+1}$  est fonction de  $y_t$  qui dépend de  $\varepsilon_t$ ,  $E(y_{t+1} \varepsilon_t) \neq 0$ .

**Cas particulier** : le modèle autorégressif d'ordre (1)

$$y_t = b_1 y_{t-1} + a_0 + a_1 x_{1t} + a_2 x_{2t} + \dots + a_k x_{kt} + \varepsilon_t$$

Le modèle est stable si  $|b_1| < 1$ .

### 7.4.2. Test d'autocorrélation et méthodes d'estimation

Dans le cas de modèle autorégressif, le test de Durbin et Watson a une puissance réduite et est biaisé. C'est pourquoi il convient d'utiliser une autre statistique, celle de « h » de Durbin.

#### a) Test d'autocorrélation des erreurs

La statistique « h » de Durbin est donnée par :

$$h = \hat{\rho} \sqrt{\frac{n}{1 - n \hat{\sigma}_{b_1}^2}} \quad \text{avec} \quad \hat{\rho} = 1 - \frac{DW}{2}$$

DW : est la statistique de Durbin et Watson calculée sur le modèle (7.12)

$n$  : est le nombre d'observations.

$\hat{\sigma}_{b_1}^2$  : est la variance estimée du coefficient  $b_1$  à partir du modèle (7.12).

Cette statistique « h » est distribuée asymptotiquement comme variable centrée et réduite. Ainsi, il ya équivalence entre les deux tests d'hypothèses suivants :

$H_0$  " $\rho = 0$ " contre  $H_1$  " $\rho \neq 0$ " et le test :  $H_0$  " $h = 0$ " contre  $H_1$  " $h \neq 0$ "

- Si  $|h| \leq t^{\alpha/2}$ , nous acceptons l'hypothèse  $H_0$  d'indépendance des erreurs.

$t^{\alpha/2}$  : valeur issue de la loi normale pour un test bilatéral au seuil  $\alpha$ .

### b) Estimation dans le cas d'autocorrélation des erreurs

Dans le modèle (7.12),  $\varepsilon_t = \rho\varepsilon_{t-1} + \vartheta_t$ . Il s'agit d'un modèle à autocorrélation des erreurs, nous avons vu au chapitre 5 que la transformation en quasi-différences premières permet de lever l'autocorrélation des erreurs d'ordre 1.

On peut utiliser les méthodes déjà développées en Chapitre 5 pour l'estimation des paramètres du modèle.

## 7.5. Modèles dynamiques

Les modèles à décalages temporels sont nombreux dans la littérature, nous allons présenter dans ce cas intéressants dans ce qui suit :

### 7.5.1. Modèle d'ajustement partiel

Afin de faire face à une augmentation de la demande, une entreprise cherche à se doter des moyens de production supplémentaires. Cependant, cet investissement ne peut pas être réalisé immédiatement et demande un certain temps d'ajustement.

- Lors de l'augmentation des prix du pétrole, les consommateurs et les institutions réagissent avec retard à cette modification des prix.

La formulation de ce modèle est la suivante :

On distingue la valeur désirée  $y_t^*$  de la variable à expliquer  $y_t$ , et la valeur vraie de cette même variable. Le niveau désiré de la variable à expliquer est fonction de la variable explicative  $x_t$  :

$$y_t^* = a_0 + a_1x_t + \varepsilon_t \dots\dots\dots (7.13)$$

On ne peut pas appliquer les MCO puisque nous n'avons aucune mesure de  $y_t^*$ . En revanche, nous connaissons les valeurs de  $y_t$  et nous pouvons simplifier une relation entre  $y_t^*$  et  $y_t$  :

$$y_t - y_{t-1} = \lambda(y_t^* - y_{t-1}) \dots\dots\dots(7.14)$$

avec  $0 \leq \lambda \leq 1$  appelé le coefficient d'ajustement.

La relation (7.14) représente l'ajustement progressif qui s'opère entre la valeur désirée et la valeur observée.

Ce modèle peut se ramener au modèle de Koyck :

On remplace (7.13) dans (7.14) on obtient :

$$y_t - y_{t-1} = \lambda(a_0 + a_1x_t + \varepsilon_t - y_{t-1})$$

$$\Rightarrow y_t = (1 - \lambda) y_{t-1} + \lambda a_0 + \lambda a_1 x_t + \lambda \varepsilon_t$$

$\Rightarrow y_t = b_1 y_{t-1} + c_0 + c_1 x_t + \vartheta_t$ , dont on peut estimer les paramètres par les MCO avec :

$$b_1 = (1 - \lambda), c_0 = \lambda a_0 \text{ et } c_1 = \lambda a_1$$

### 7.5.2. Les modèles d'anticipations adaptatives

Dans cette spécification, les valeurs de la variable  $y_t$  sont fonction des valeurs, non pas observées d'une variable explicative, mais des valeurs attendues, telles que :

$$y_t = a_0 + a_1 x_t^* + \varepsilon_t \dots \dots \dots (7.15)$$

où :  $x_t^*$  est la valeur prévue de la variable explicative  $x_t$ .

**Exemple :** La production d'une entreprise en fonction des valeurs prévues des ventes.

On n'a pas la mesure de la variable  $x_t^*$  ; pour lever cette difficulté, nous devons poser une hypothèse concernant la fonction de la variable  $x_t^*$ , celle des anticipations adaptatives.

$$x_t^* - x_{t-1}^* = \lambda(x_t - x_{t-1}^*) \dots \dots \dots (7.16)$$

avec  $0 \leq \lambda \leq 1$  appelé le coefficient d'anticipation.

L'équation (7.16) peut s'écrire ;

$$x_t^* = \lambda x_t + (1 - \lambda)x_{t-1}^*$$

En développant cette formule, il vient ::

$$x_t^* = \lambda x_t + \lambda(1 - \lambda)x_{t-1} + \lambda(1 - \lambda)^2 x_{t-2} + \lambda(1 - \lambda)^3 x_{t-3} + \dots$$

En remplaçant  $x_t^*$  par son expression, dans (7.15)

$$y_t = a_0 + a_1 \lambda \sum_{i=0}^{+\infty} (1 - \lambda)^i x_{t-i} + \varepsilon_t$$

qui est un modèle à retards échelonnés dont une transformation de Koyck permet de mettre sous forme autorégressive :

$$y_t = \lambda a_0 + \lambda a_1 x_t + (1 - \lambda)y_{t-1} + (\varepsilon_t - (1 - \lambda)\varepsilon_{t-1})$$

Nous pouvons estimer ce modèle autorégressif à autocorrélations des erreurs, et ainsi, en déduire une estimation des paramètres :  $\lambda$ ,  $a_0$  et  $a_1$ .

**Références bibliographiques**

1. Brigitte Dormont, « Introduction à l'économétrie » Ed Montchrestien, Paris 1999.
2. Domadar N. Gujarati, « Econométrie ». Ed de boeck, Bruxelles, 2004.
3. Régis Bourbonnais, « Econométrie, manuel et exercices corrigés ». Ed Dunod, Paris 2015.
4. Régis Bourbonnais, « Exercices pédagogiques d'économétrie ». Ed Economica Paris 2015.
5. William Greene, « Econometric analysis ». Ed Pearson, New York, 2018.
6. William Greene, « Économétrie. Annexes : exercices et corrigés. » Ed Pearson, New York, 2006.