Séance 2
Autocorrélation des erreurs et hétéroscédasticité et Test de normalité

Contenu :

Définition et causes

Détection de l'autocorrélation

Tests usuels d'autocorrélation:

Test des runs, Durbin et Watson, Breusch-Godfrey, Box-Pierce, Ljung-Box

Hétéroscédasticité : définition et tests

Test de normalité

1. Définition et causes

Nous sommes en présence d'une autocorrélation des erreurs lorsque les erreurs sont liées par un processus de reproduction ou processus à mémoire (par comparaison au processus purement aléatoire).

L'autocorrélation des erreurs peut être observée pour plusieurs raisons :
- absence d'une variable explicative importante;
- une mauvaise spécification du modèle (relations entre variables explicatives et la variable endogène sont de type non linéaire – logarithme, différences premières…-);
- un lissage par moyenne mobile ou interpolation des données crée une autocorrélation artificielle des erreurs.

2. Détection de l'autocorrélation

La détection d'une éventuelle dépendance des erreurs ne peut s'effectuer qu'à partir de l'analyse des résidus (eux seuls sont connus).
L'examen visuel graphique des résidus permet le plus souvent de détecter un processus de reproduction des erreurs lorsque :
- les résidus sont, pendant plusieurs périodes consécutifs, soit positifs, soit négatifs (autocorrélation positive)
- Les résidus sont alternés (autocorrélation négative)

Cependant, le plus souvent, l'analyse graphique est délicate d'interprétation, car le dessin des résidus ne présente pas de caractéristiques toujours évidentes. D'où la nécessité de recours aux tests statistiques plus significatifs.

3. Tests usuels d'autocorrélation

a- Test de Durbin et Watson (1950) et (1951)

Le test DW permet de détecter une autocorrélation des erreurs d'ordre 1 selon la forme :
$\varepsilon_t = \rho\varepsilon_{t-1} + v_t$ avec v_t = BB (0, σ^2_v)

Le test d'hypothèse est le suivant :
H0 : $\rho = 0$
H1 : $\rho \neq 0$

On calcule la statistique de DW :

$$DW = \frac{\sum_{t=2}^{n}\left(e_t - e_{t-1}\right)^2}{\sum_{t=1}^{n} e_t^2}$$ où $\boldsymbol{e_t}$ sont les résidus de l'estimation du modèle

La statistique DW, de par sa construction, varie entre 0 et 4.

0	d1	d2	2	4-d2	4-d1	4
ρ > 0	doute ?	ρ = 0	ρ = 0	doute ?	ρ < 0	

La table de DW donnent les valeurs critiques de la statistique au seuil de 5% en fonction de la taille de l'échantillon ***n*** et du nombre de variables explicatives ***k***.

Remarques

- Le test DW est uniquement valable pour détecter des autocorrélations d'ordre 1.
- La variable endogène ne doit pas figurer parmi les variables explicatives (en tant que variable retardée).

b- Test non paramétriques des runs (changements de signes)

Les tests de changement de signe ne dépendent pas de la forme de distribution des rendements.
Ce test examine la fréquence de caractères répétitifs dans une chronique. Le nombre *R* de « *runs* » est le nombre de passage du signe « + » au signe « - » et inversement. Si les observations sont indépendantes (donc non corrélées), il est improbable d'avoir un nombre faible de « *runs* ».
Dans le cas d'une indépendance, le nombre attendu de « *runs* » est égal à *m* :

$$m = \frac{1}{n}\left(n(n+1) - \sum_{i=1}^{3} n_i^2\right)$$ où n_i est le nombre de runs respectivement positifs, négatifs ou nuls (i =1, 2, 3).

La variance de *m* est :

$$\sigma_m^2 = \frac{1}{n^2(n-1)}\left(\left(\sum_{i=1}^{3} n_i^2\right)\left(\sum_{i=1}^{3} n_i^2 + n(n+1)\right) - 2n\sum_{i} n_i^3 - n^3\right)$$

Pour *n* grand, la distribution est considérée comme normale et on démontre que :

$$z = \frac{R + 0{,}5 - m}{\sigma_m} \rightarrow N(0{,}1)$$

L'hypothèse d'indépendance (H0) des observations est refusée si $z > 1{,}96$ à 95% de chances.
La technique du test pour valider l'absence d'autocorrélation consiste alors à comparer le nombre attendu de runs pour un nombre donné d'observations dans un contexte purement aléatoire au nombre réel de runs obtenus pour l'échantillon envisagé.

c- Test de Breusch-Godfrey (1978)

Ce test, fondé sur un test de Fisher de nullité de coefficients ou de multiplicateur de Lagrange, permet de tester une autocorrélation d'ordre supérieur à 1 et reste valide en présence de la variable dépendante décalée en tant que variable explicative.
Une autocorrélation des erreurs d'un ordre ***p*** s'écrit :

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \ldots + \rho_p \varepsilon_{t-p} + v_t$$

Le modèle général à erreurs autocorrélées d'ordre p s'écrit :

$$y_t = a_1 x_{1t} + a_2 x_{2t} + \cdots + a_k x_{kt} + a_0 + \rho_1 \varepsilon_{t-1} + \ldots + \rho_p \varepsilon_{t-p} + v_t$$

Ce test est mené en trois étapes :
- Estimation par les MCO du modèle et calcul du résidu e_t, puisque les erreurs sont inconnues, le test porte sur les résidus.
- Estimation par les MCO de l'équation intermédiaire :

$$e_t = a_1 x_{1t} + a_2 x_{2t} + \cdots + a_k x_{kt} + a_0 + \rho_1 e_{t-1} + \ldots + \rho_p e_{t-p} + v_t$$

Soit n le nombre d'observations disponibles (chaque décalage entraîne la perte d'une observation) pour estimer les paramètres du modèle et R^2 (coefficient de détermination).

- Test d'hypothèses sur l'équation intermédiaire :

L'hypothèse H0 d'absence d'autocorrélation des erreurs à tester est :
H0 : $\rho_1 = \rho_2 = \ldots = \rho_p = 0$ contre H1 : il existe un ρ_i différent de 0.

Deux possibilités :
- soit le test classique de Fisher de nullité des coefficients
- soit la statistique LM qui est distribuée selon un χ^2 à p degrés de libertés :

  Si $n*R^2 > \chi^2(p)$ lu dans la table au seuil α, on rejette l'hypothèse H0 d'indépendance des erreurs

d- Test de Box-Pierce (1970)

Aussi appelé test portmanteau, ce test permet d'identifier les processus de bruit blanc.
Si x_t est un bruit blanc, alors $cov(x_t, x_{t-k})=0$ ou $\rho_k=0$, $\forall k$, d'où le test d'hypothèses :

Test d'hypothèse :
- H0 : $\rho_1 = \rho_2 = \ldots = \rho_k = 0$
- H1 : il existe au moins un $\rho_i \neq 0$ (significativement différent de 0)

Statistique Q de Box-Pierce :

$$Q = n \sum_{k=1}^{h} r_k^2$$

où : h = nombre de retards, r_k = autocorrélation empirique d'ordre k,
n = nombre d'observations

La statistique Q est distribuée de manière asymptotique comme un χ^2 (khi-deux) à h degrés de liberté. On rejette l'hypothèse de bruit blanc, au seuil α, si la statistique Q est supérieure au χ^2 lu dans la table au seuil (1-α) et h degrés de liberté.

e- Test de Ljung-Box (1978)

On peut utiliser une autre statistique, dont les propriétés asymptotiques sont meilleures, dérivée de la précédente (Statistique de Box-Pierce) qui est le Q' de Ljung et Box (1978)

$Q' = n(n+2)\sum_{k=1}^{h}\frac{r_k^2}{n-k}$ qui est aussi distribué selon un χ^2 à h degrés de liberté et dont les règles de décision sont identiques au test précédent.

Exemple :
Soit la chronique x_t suivante :

t	x_t
1	10
2	3
3	-1
4	3
5	2
6	5
7	3
8	2
9	-1
10	3

1- Calculer les 3 premiers termes de la FAC (r_1, r_2 et r_3)
2- Calculer la statistique Q' de Ljung-Box pour h = 3, la série x_t est-elle un bruit blanc?

Réponse :

1- Calcul des 3 premiers termes de la FAC (r_1, r_2 et r_3)

$$r_k = \frac{\sum_{t=k+1}^{n}\left(x_t-\bar{x}_1\right)\left(x_{t-k}-\bar{x}_2\right)}{\sqrt{\sum_{t=k+1}^{n}\left(x_t-\bar{x}_1\right)^2\sum_{t=k+1}^{n}\left(x_{t-k}-\bar{x}_2\right)^2}}$$

$$\bar{x}_1 = \frac{1}{n-k}\sum_{t=k+1}^{n}x_t \qquad \bar{x}_2 = \frac{1}{n-k}\sum_{t=k+1}^{n}x_{t-k}$$

r_1 = 0,021; **r_2 = -0,502** **r_3 = -0,348**

2- Calcul de la statistique Q' de Ljung-Box pour $h = 3$

$$Q' = n(n+2)\sum_{k=1}^{h}\frac{r_k^2}{n-k}$$

$$= 10*12\left[\frac{(0{,}021)^2}{9}+\frac{(-0{,}502)^2}{8}+\frac{(-0{,}348)^2}{7}\right] = 5{,}86$$

Cette statistique est à comparer χ^2 lu dans la table à 3 degrés de libertés :
$Q' = 5{,}86 < 7{,}815$

Conclusion :

On accepte l'hypothèse H0 de nullité des coefficients d'autocorrélation. La série x_t est donc un bruit blanc.