

VI - Modélisation selon un modèle linéaire standard Normal

Modèle $(\mathcal{Y}, \mathcal{X}, \mathcal{P}, \Theta)$ où $\mathcal{Y} = \mathbb{R}^T$, $\mathcal{X} = \mathbb{R}^{kT}$

H1 $E(y) = \beta_0 e + \beta_1 x_1 + \dots + \beta_k x_k = X\beta$

$$\beta \in \Theta = \mathbb{R}^{k+1}$$

H2 $V(y) = \sigma^2 I_T$

H3N y vecteur Normal

72

On désire expliquer les variations de y (observé T fois) en fonction de k variables « explicatives », x_1, \dots, x_k (observées simultanément avec y).

Si on postule un **modèle linéaire standard Normal**, c'est l'estimateur MCO qui est le meilleur.

1. Il faut tester du mieux possible la validité d'un tel modèle, qui se traduit par

$$u = y - X\beta \sim N(0 ; \sigma^2 I_T)$$

Nous effectuerons des **Tests de spécification**, utilisant les résidus : test de H1, tests de H2

2. Si le modèle est acceptable, des **tests de contraintes linéaires** pourront alors vérifier s'il est possible de faire l'économie de quelque(s) paramètre(s).

VI-1. Tests de spécification : tests de bruit blanc

- VI - 1.a. Test de H1
 - test des séquences
- VI - 1.b. Tests de H2
 - Durbin et Watson
 - Portmanteau
 - Hétéroscédasticité (Golfeld-Quandt)

73

Il s'agit ici de tester la validité du modèle : l'équation

$$y = X\beta + u$$

définit simplement les perturbations u . Ce sont les hypothèses sur les propriétés de la loi de u (sachant X :rappelons que X est considérée comme certaine) qu'il faut examiner :

- hypothèse d'ordre 1 : H1 $E(u) = 0$
- hypothèse d'ordre 2 : H2 $V(u) = \sigma^2 I_T$

Le problème est ici que l'on n'observe jamais les perturbations u , puisque le paramètre β est inconnu.

Les tests utiliseront le résidu de l'estimation MCO.

Rappel :

$$\hat{\beta} = (X'X)^{-1}X'y \text{ estimateur MCO de } \beta$$

$$X\hat{\beta} = X(X'X)^{-1}X'y = P_X y \text{ ajustement de } y$$

$$\hat{u} = y - X\hat{\beta} \text{ résidu d'ajustement}$$

$$\hat{u} = (I - P_X)y = (I - P_X)u$$

VI-1.a. Test des séquences

+ - - - + + - - - - + + + - - + - - - + +

- Chaque résidu est remplacé par son signe
- H_0 : les u_t sont i.i.d. d'espérance nulle
- N = nbre d'observations, n_1 = nbre de '+', n_2 = nbre de '-' et R = nbre de séquences
- Statistique : $U = \frac{R-m}{S}$

$$\text{où } m = \frac{2n_1n_2}{N} + 1 \text{ et } S^2 = \frac{2n_1n_2(2n_1n_2 - N)}{N^2(N-1)}$$

- Sous H_0 , si n_1 ou $n_2 > 20$, $U \sim N(0;1)$
- Refus de H_0 si $|U| > A$

74

Le test des séquences est appliqué aux résidus de l'ajustement MCO.

L'hypothèse de base est « les u_t sont i.i.d. centrés », que l'on traduit par « les u_t forment un Bruit Blanc indépendant », qui est en particulier fausse lorsque H_1 ou H_2 sont fausses.

Dans un modèle avec constante, nous ne pourrions tester l'hypothèse d'espérance nulle des perturbations : par construction, les résidus sont de moyenne nulle.

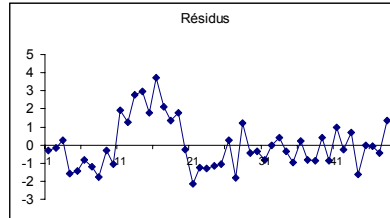
Le test des séquences, ou 'test des suites' détecte surtout les erreurs de spécification sur H_1 (oubli de variable, ou non-linéarité de $E(y)$ en fonction de X).

L'hypothèse de bruit blanc est refusée si le nombre de séquences est soit trop grand, soit trop petit. La statistique utilisée est asymptotiquement normale centrée réduite : la loi limite $N(0;1)$ est utilisée lorsque le nombre de signes + ou celui des signes - est supérieur à 20.

Ce test donne l'occasion d'une réflexion sur le choix du seuil du test : l'hypothèse de base n'est pas celle dont le rejet à tort a les conséquences les plus fâcheuses. Elle est seulement celle sous laquelle on connaît la loi de la statistique. Dans ce cas, on peut choisir un seuil élevé, afin d'augmenter la puissance.

Une règle usuelle est de refuser l'hypothèse de Bruit Blanc si $|U| > 1$. Ce test est de seuil 31,7%.

VI-1.a. Test des séquences



- $N = 48$ observations
- $n_1 = 18$, $n_2 = 30$ et $R = 20$
- Pour un seuil de 5%, on refuse l'hypothèse de Bruit Blanc si $|U| > 2$
- ici, $U_{\text{observé}} = -1,1$: H_1 est acceptable au seuil de 5%

75

Le graphique des résidus montre que sur 48 observations, il y a 18 signes + et 30 signes -, répartis en 20 séquences.

--+-----+++++++-----+-+-----+--+--+--+-----+

La statistique observée est égale à -1.1 n'est pas dans la région critique du test : ce test au seuil 5% ne permet pas de rejeter le modèle linéaire qui a été estimé.

Le graphique montre pourtant la présence d'un comportement suspect en première période ($t < 20$) : cette « bulle » n'est pas significative au seuil de 5%.

Si on calcule $\text{PROB} = \Pr(|N(0;1)| > 1.1) = 27.13\%$, on voit que cette valeur est significative pour un seuil du test plus élevé (de l'ordre de 30%).

En particulier, le test usuel de seuil 31,7% (de région critique $|U| > 1$) refuse ici l'hypothèse de bruit blanc.

VI-1.b. Tests de H2

- On suppose que H1 est vraie : $E(y) = X\beta$
- On ne peut tester directement
 $H2 = \{V(y) = \sigma^2 I_T\}$ contre $K2 = \{V(y) \neq \sigma^2 I_T\}$,
 car cela correspond à $T(T+1)/2 - 1$ contraintes à tester simultanément à partir de T observations.
 On distingue deux types de tests de H2, selon l'alternative envisagée.
- tests d'homoscédasticité :
 $\{V(y_t) = \sigma_t^2, \text{cov}(y_{t+h}, y_t) = 0 \text{ si } h \neq 0\}$
- tests de non-corrélation des perturbations :
 $\{ \text{cov}(y_{t+h}, y_t) = \gamma(h) \}$

76

La variance du vecteur y est la matrice des covariances des composantes entre elles. Nous ne perdons aucune généralité en notant

$$\text{cov}(y_{t+h}, y_t) = \gamma(h, t)$$

La variance de y s'écrit alors, en notant t le numéro de la colonne, et dans chaque colonne, $t+h$ est le numéro de la ligne (h est la différence entre le numéro de la ligne et celui de la colonne).

$$V(y) = \begin{bmatrix} \gamma(0,1) & \gamma(-1,2) & \dots & \dots & \gamma(2-T, T-1) & \gamma(1-T, T) \\ \gamma(1,1) & \gamma(0,2) & \gamma(-1,3) & \dots & \gamma(3-T, T-1) & \gamma(2-T, T) \\ \gamma(2,1) & \gamma(1,2) & \gamma(0,3) & \ddots & \dots & \gamma(3-T, T) \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \gamma(T-2,1) & \gamma(T-3,2) & \dots & \ddots & \gamma(0, T-1) & \gamma(-1, T) \\ \gamma(T-1,1) & \gamma(T-2,2) & \gamma(T-3,3) & \dots & \gamma(1, T-1) & \gamma(0, T) \end{bmatrix}$$

Les tests d'homoscédasticité prennent pour acquis (on ne le teste pas) que $\gamma(h, t) = 0$ dès que h est différent de 0, et on note alors $\gamma(0, t) = \sigma_t^2$.

Les tests d'absence d'autocorrélations des perturbations prennent pour acquis que $\gamma(h, t)$ est indépendant de t (« stationnarité au second ordre »), noté alors $\gamma(h)$.

Test de Durbin et Watson

- Modèle structurel :

H1 $E(y) = X\beta$, $\beta \in \mathbb{R}^{k+1}$

les X sont certaines , $e \in \text{Im}(X)$

HDW $\text{Var}(y_t) = \gamma(0)$, $\text{cor}(y_{t+h}, y_t) = \rho^{|h|}$

où $-1 < \rho < 1$

H3N $y \sim$ loi Normale dans \mathbb{R}^T .

- Hypothèse de base $H_0 = \{\rho = 0\}$

- Statistique utilisée $DW = \frac{\sum_{t=2}^T (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^T (\hat{u}_t)^2} \# 2(1-r)$

77

Les tests de non corrélation des perturbations se situent dans un cadre dit « stationnaire au second ordre » : la suite des perturbations (u_t) vérifie les conditions :

$$E(u_t) = 0 \quad \text{cov}(u_{t+h}, u_t) = \gamma(h) \text{ indépendant de } t.$$

$\gamma(0)$ est la variance commune des u_t , et $\text{cor}(u_{t+h}, u_t) = \gamma(h) / \gamma(0)$ notées $\rho(h)$

La variance du vecteur y est alors :

$$V(y) = \begin{bmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(T-2) & \gamma(T-1) \\ \gamma(1) & \gamma(0) & \ddots & \ddots & \gamma(T-2) \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \gamma(T-2) & \ddots & \ddots & \gamma(0) & \gamma(1) \\ \gamma(T-1) & \gamma(T-2) & \dots & \gamma(1) & \gamma(0) \end{bmatrix} = \gamma(0) \begin{bmatrix} 1 & \rho(1) & \dots & \rho(T-2) & \rho(T-1) \\ \rho(1) & 1 & \ddots & \ddots & \rho(T-2) \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho(T-2) & \ddots & \ddots & 1 & \rho(1) \\ \rho(T-1) & \rho(T-2) & \dots & \rho(1) & 1 \end{bmatrix}$$

Dans ce cadre, l'hypothèse de bruit blanc des perturbations est la nullité de tous les $\rho(h)$, pour $h = 1, \dots, T-1$.

L'hypothèse HDW correspond en fait à ce que l'on appelle un processus « Autorégressif d'ordre 1 », qui permet de réduire l'hypothèse de nullité de tous les $\gamma(h)$ (h différent de zéro) par la nullité d'un seul paramètre, ρ .

En fait, le test de DW ne teste que l'hypothèse $\rho(1) = 0$, qui est une des conséquences de l'hypothèse de bruit blanc.

Test de Durbin et Watson (suite)

- Sous H_0 , la loi de DW est symétrique sur l'intervalle $[0; 4]$. Cette loi dépend de X .
- On décide que $\rho \neq 0$ (test bilatéral) si $\{DW < A \text{ ou } DW > 4 - A\}$ où A dépend du seuil α et de X
- Les tables de Durbin et Watson fournissent, pour un seuil α donné, et en fonction de T et k , des valeurs d_L et d_U telles que
$$d_L < A(\alpha, X) < d_U$$

78

Puisque $DW \approx 2(1 - r)$, où r est un estimateur de $\rho(1) = \rho$, un DW proche de 0 correspond à ρ proche de 1, et un DW proche de 4 correspond à un ρ proche de -1.

Lorsque $\rho = 0$, DW est centré autour de 2.

La loi étant symétrique, la région critique du test bilatéral correspond aux probabilités $\alpha/2$ pour chacune des régions $\{DW < A\}$ et $\{DW > 4 - A\}$

On remplace la condition $\{DW > 4 - A\}$ par $\{4 - DW < A\}$.

On considère donc $Z = \min(DW, 4 - DW)$, c'est-à-dire celle des deux qui est plus petite que 2 (elles sont symétriques l'une de l'autre par rapport à 2)

Si $Z < d_L$, on décide que $\rho \neq 0$, si $Z > d_U$, on ne peut refuser $\rho = 0$

Si $d_L < Z < d_U$, la procédure ne permet pas de conclure.

Si on doit faire des tests de contraintes sur le coefficient β , on prendra d_U comme valeur critique, et on décidera par exemple de refuser $\rho = 0$ dès que $Z < d_U$. Le test ainsi mis en œuvre n'est pas de seuil α , mais de seuil $\alpha^* > \alpha$.

Noter que le test de Durbin et Watson ne s'applique que lorsque les X sont certaines (pas de variables y retardées comme variables explicatives), et lorsque la constante appartient à l'espace engendré par les X_i .

Le nombre k représente alors le nombre de variables explicatives autres que la constante.

Test de Durbin et Watson : exemple

- On a régressé y sur x et z avec constante, en utilisant 48 données.
- Les valeurs lues dans la table de DW sont, pour un test bilatéral de seuil 10%, $n = 48$ et $k = 2$: $d_L = 1.448$ $d_U = 1.626$
- $DW_{\text{observé}} = 0.918 < d_L$
- Conclusion : au seuil de 10%, on refuse $\{\rho = 0\}$

79

Les résidus sont ceux dont le graphique a été montré précédemment.

Compte tenu de la « bulle » observée, il est possible que H_0 soit refusée à cause d'une mauvaise spécification de H_1 (changement de structure, ou points aberrants), plutôt que pour des autocorrélations.

Historiquement, il y a eu un choc possible entre $t = 19$ et $t = 20$.

Nous verrons plus loin comment tester le changement de structure et trouver un modèle plus adapté aux données et à l'historique connu des observations.

Test du portmanteau

- Modèle structurel :

$$H1 \quad E(y) = X\beta, \beta \in \mathbb{R}^{k+1}$$

X certaine de rang $k+1$, $e \in \text{Im}(X)$

$$HP \quad \text{var}(u_t) = \gamma(0), \text{cor}(u_{t+h}, u_t) = \gamma(h)$$

$$H3N \quad y \sim \text{loi Normale dans } \mathbb{R}^T.$$

- Hypothèse de base $H_0 = \{\gamma(h) = 0, h \neq 0\}$
- Statistique utilisée

$$Q(H) = T \sum_{h=1}^H [\hat{\rho}(h)]^2 \quad \text{où} \quad \hat{\rho}(h) = \frac{\sum_{t=1}^{T-h} (\hat{u}_{t+h} \hat{u}_t)}{\sum_{t=1}^T (\hat{u}_t)^2}, \quad h = 1, \dots, H$$

80

On choisit un H compatible avec les données : on ne doit pas dépasser $T/4$, et il faut tenir compte des autocorrélations de type saisonnier (la même époque de l'année dernière, ou des années dernières).

Si les données sont mensuelles, on prendra H supérieur à un multiple de 12, si elles sont trimestrielles, on le prendra supérieur à un multiple de 4.

Certains logiciels fournissent systématiquement $Q(6)$, $Q(12)$, $Q(18)$ et $Q(24)$.

La supériorité de portmanteau sur DW est qu'il permet de détecter des autocorrélations saisonnières que DW n'aurait pas détectées.

L'inconvénient de portmanteau est qu'il concerne surtout l'analyse directe d'une suite de variables aléatoires, et qu'il n'est valable que pour de grandes valeurs de T .

Le mot **portmanteau** est un mot anglais signifiant « grosse valise ». Issu du français, il nous est revenu avec une faute d'orthographe et un sens particulier qui nous convient pour ce test qui a vocation à tenir compte **globalement** de toutes les autocorrélations des perturbations (et pas seulement de $\rho(1)$).

Test portmanteau, suite

- Sous H_0 , $Q(H) \sim \chi^2(H)$ lorsque T tend vers l'infini.
- Pour un seuil α choisi, on refuse H_0 :
 - si $Q(H) > A_\alpha$ lu dans la table du chi-deux.ou
 - si $\text{PROB} = P(\chi^2(H) > Q(H)_{\text{observé}}) < \alpha$

81

Si on a constaté la présence d'autocorrélations, il faut faire une estimation dite « Moindres Carrés Généralisés », en modélisant les perturbations de façon à rendre compte des autocorrélations constatées. Les modèles utilisés sont du type « ARMA » :

$$u_t - \varphi_1 u_{t-1} - \dots - \varphi_p u_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

où

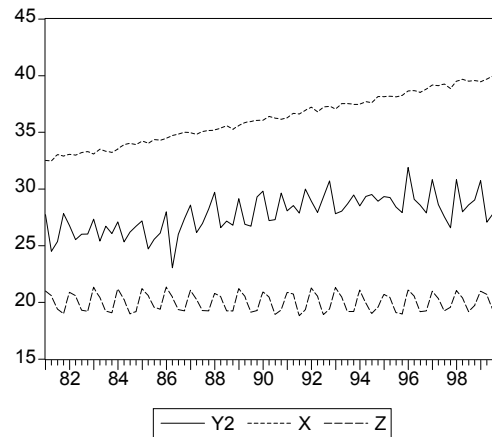
Les ε_t forment un Bruit Blanc centré de variance σ^2

$$\text{cov}(\varepsilon_t, \varepsilon_s) = 0 \text{ dès que } s < t$$

Dans ce cas, le test d'absence d'autocorrélations des ε_t sera fait en utilisant la même statistique $Q(H)$, qui suit alors sous l'hypothèse de Bruit Blanc (des ε_t) un chi-deux à $H-m$ degrés de liberté, où m est le nombre de paramètres φ et θ qui ont été estimés.

L'exemple qui suit montre la faiblesse du test de DW et le danger qu'il y a à négliger la présence d'autocorrélations (les tests de significativité des coefficients ne sont plus valables)

Exemple : Y // e, X, Z



82

Données trimestrielles, de 1981:1 à 1999:4

Au vu des observations, on pourrait penser que la variable à expliquer, Y2, dépend de Z, dont elle semble reprendre les variations saisonnières.

Lorsqu'on régresse Y2 sur X et Z avec constante, on obtient le modèle (A), dans lequel

- le test de DW ne détecte pas d'autocorrélations des perturbations
- les tests de Student sur les coefficients, au seuil de 10% conduisent à considérer X et Z comme explicatives.

Lorsqu'on examine toutes les autocorrélations, on constate qu'il existe une autocorrélation d'ordre 4, et les Q-stat sont :

Sample: 1981:1 1999:4 Included observations: 76

| h | r(h) | Q-Stat | Prob | |
|---|--------|--------|-------|-------------|
| 1 | 0.019 | 0.0285 | 0.866 | |
| 2 | -0.158 | 2.0237 | 0.364 | |
| 3 | 0.091 | 2.6915 | 0.442 | |
| 4 | 0.593 | 31.605 | 0.000 | Refus de Ho |
| 5 | -0.087 | 32.232 | 0.000 | |
| 6 | -0.146 | 34.048 | 0.000 | |
| 7 | 0.081 | 34.612 | 0.000 | |
| 8 | 0.434 | 51.054 | 0.000 | |
| 9 | -0.122 | 52.376 | 0.000 | |

Sortie des calculs effectués par « EViews », Modèle (A)

Dependent Variable: Y2 Method: Least Squares
Sample: 1981:1 1999:4 Included observations: 76

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|----------|-------------|------------|-------------|---------------|
| C | 4.135430 | 4.346438 | 0.951453 | 0.3445 |
| X | 0.486838 | 0.065705 | 7.409398 | 0.0000 |
| Z | 0.301791 | 0.176956 | 1.705463 | 0.0924 |

| | | | |
|---------------------------|----------------|-----------------------|----------|
| R-squared | 0.437997 | Mean dependent var | 27.81582 |
| Adjusted R-squared | 0.422599 | S.D. dependent var | 1.653400 |
| S.E. of regression | 1.25636 | Akaike info criterion | 3.332999 |
| Sum squared resid | 115.2275 | Schwarz criterion | 3.425002 |
| Log likelihood | -123.6540 | F-statistic | 28.44625 |
| Durbin-Watson stat | 1.94404 | Prob(F-statistic) | 0.00000 |

83

Pour les test individuels des coefficients, **au seuil de 10%** :

pour C : $0.3445 > 0.10$ 4.13 n'est pas significativement différent de 0

pour X : $0.0000 < 0.10$ 0.487 est significativement différent de 0

pour Z : $0.0924 < 0.10$ 0.302 est significativement différent de 0

En fait, la présence constatée d'autocorrélation nous interdit de faire ces tests, et les « conclusions » que nous venons d'écrire ne sont pas valables.

Si on fait l'estimation en introduisant le modèle $u_t = \varphi u_{t-4} + \varepsilon_t$

qui ne contient qu'un coefficient φ sur le retard « 4 », on obtient le modèle (B) suivant.

Sortie des calculs effectués par « EViews », Modèle (B)

| | | | | |
|------------------------|-------------|---------------------------|-----------------------|--------|
| Dependent Variable: Y2 | | Method: Least Squares | | |
| Sample: 1981:1 1999:4 | | Included observations: 76 | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | 11.83632 | 9.268425 | 1.277058 | 0.2057 |
| X | 0.404961 | 0.157360 | 2.573466 | 0.0121 |
| Z | 0.064694 | 0.350871 | 0.184382 | 0.8542 |
| AR (4) | 0.682923 | 0.093953 | 7.268803 | 0.0000 |
| R-squared | 0.669774 | Mean dependent var | 27.81582 | |
| Adjusted R-squared | 0.656014 | S.D. dependent var | 1.653400 | |
| S.E. of regression | 0.969724 | Akaike info criterion | 2.827585 | |
| Sum squared resid | 67.70626 | Schwarz criterion | 2.950256 | |
| Log likelihood | -103.4482 | F-statistic | 48.67744 | |
| Durbin-Watson stat | 1.934224 | Prob(F-statistic) | 0.00000 ⁸⁴ | |

On constate que la variable Z n'est pas explicative lorsqu'on fait un test correct.

Le test de DW ne voit toujours rien, mais maintenant il n'y a vraiment plus rien à voir : le calcul des autocorrélations des résidus donne

| | | | |
|-----------------------|--------|---------------------------|-------|
| Sample: 1981:1 1999:4 | | Included observations: 76 | |
| h | r(h) | Q-Stat | Prob |
| 1 | 0.011 | 0.0100 | |
| 2 | -0.035 | 0.1059 | 0.745 |
| 3 | 0.112 | 1.1298 | 0.568 |
| 4 | -0.104 | 2.0180 | 0.569 |
| 5 | -0.079 | 2.5392 | 0.638 |
| 6 | -0.081 | 3.0941 | 0.685 |

Ces probabilités correspondent à des lois de chi-deux à (H-1) degrés de liberté, car il y a 1 paramètre estimé dans le modèle de génération des perturbations. Aucune n'est inférieure à 10% : l'hypothèse de Bruit Blanc des ε_t ne peut donc pas être refusée.

VI-1.c. Tests d'homoscédasticité

Sur données individuelles, on dispose de plusieurs tests d'homoscédasticité qui diffèrent selon la façon dont la variance varie entre les individus.

- Modèle structurel :

$$H1 \quad E(y) = X\beta, \beta \in \mathbb{R}^p$$

X est certaine, de rang p

$$K2^* \quad V(y_i) = \sigma_i^2, \text{cov}(y_i, y_j) = 0 \text{ si } i \neq j$$

$$H3N \quad y \sim \text{loi Normale dans } \mathbb{R}^N.$$

- Hypothèse de base $H_0 = \{ \sigma_n^2 = \sigma^2 \}$

85

Ces tests s'appliquent chaque fois que l'hypothèse d'absence de corrélations est acceptable.

C'est le cas des **données individuelles**, mais peut aussi se rencontrer dans les cas de **changement de structure** entre deux périodes : si, dans chaque période, on peut admettre l'absence d'autocorrélations, alors on pourra l'admettre sur la totalité des observations.

Test de Golfeld-Quandt

- Ce test suppose que $\sigma_i^2 = f(z_i)$,
où f est ↗ non précisée, et z est observable.
- Trier les N observations par z ↗
Soit $m \# N/3$. On effectue séparément les
MCO-régressions pour les m premières et les
 m dernières données. La statistique utilisée
est le rapport SCR_2/SCR_1 des SCR obtenues.
- Sous H_0 $F = \frac{SCR_2}{SCR_1} \approx FISHER(m - p, m - p)$
- On refuse H_0 si $F > A$

86

La variable z peut ou non faire partie des p variables explicatives du modèle. L'essentiel est qu'elle soit observable pour chaque observation, afin de pouvoir ordonner les données selon les valeurs croissantes de z .

Exemple : la variance de la production est fonction croissante de la taille des entreprises.

Golfeld-Quandt (suite)

- Pour un seuil α donné, on lit A_α dans la table de Fisher, et on refuse l'hypothèse H_0 d'homoscédasticité des perturbations si $F_{\text{observé}} > A_\alpha$.
- Variante : on calcule $\text{PROB} = P\{F(m-p, m-p) > F_{\text{observé}}\}$, et on refuse H_0 si $\text{PROB} < \alpha$.

87

Exemple : Une fonction de production de type Cob-Douglas : $Y = C K^a L^b$ donne, en passant aux logarithmes, le modèle linéaire suivant :

$$E(y_i) = c + a \ln K_i + b \ln L_i \text{ sur 18 entreprises.}$$

On peut faire l'hypothèse d'absence de corrélation entre les différentes entreprises, mais on craint que la variance de la production soit fonction croissante de la taille L de l'entreprise.

On réordonne les entreprise suivant les tailles croissantes .

On choisit $m = 7$.

Sur les 7 premières observations, $\text{SCR1} = 0.266979$

Sur les 7 dernières observations, $\text{SCR2} = 0.358569$

Sous H_0 , $\text{SCR2}/\text{SCR1} \sim F(4, 4)$

Pour $\alpha = 5\%$, on lit : $P(F[4,4] > 6.39) = 5\%$

On observe $F_{\text{observé}} = 1.343$ qui n'est pas supérieur à 6.39

Conclusion : au seuil de 5%, on ne peut refuser l'hypothèse d'homoscédasticité des productions.

Variante : $\text{PROB} = P\{F(4,4) > 1.343\} = 0.391 > 0.05$

(donnée par la fonction $=\text{LOI.F}(1.343;4;4)$ de Excel)

conduit à la même conclusion.

VI-2. Tests de contrainte linéaire sur les coefficients

- Modèle structurel = **modèle linéaire standard normal** $y \sim N(X\beta ; \sigma^2 I_T)$, $\beta \in \mathbb{R}^p$
- Hypothèse testée : $C\beta = 0$. où C (r, p) est donnée, de rang r (au programme : $r = 1$)
- $\{X\beta ; C\beta = 0\}$ est un sous espace vectoriel de $\text{Im}(X)$, de dimension $(p-r)$.
- Régression non contrainte : SCR_1
- Régression contrainte : SCR_0

88

$V_0 = \{X\beta ; C\beta = 0\}$ est un sous espace vectoriel de $V(X) = \{X\beta ; \beta \in \mathbb{R}^p\}$.

Si C est de rang r (r contraintes linéairement indépendantes), V_0 est de dimension $(p-r)$.

Il est très facile de faire la régression sous contrainte, en construisant $p-r$ variables engendrant V_0 et en régressant y sur ces $p-r$ variables.

Exemple 1 : soit $y \sim N(a x + b z + c e ; \sigma^2 I_{18})$, $(a, b, c) \in \mathbb{R}^3$

- Contrainte $\{a = 0\}$: $V_0 = \{b z + c e ; (b, c) \in \mathbb{R}^2\}$ engendré par z et e .
- Contrainte $\{b = 0\}$: $V_0 = \{a x + c e ; (a, c) \in \mathbb{R}^2\}$ engendré par x et e .
- Contrainte $\{a = 0 ; b = 0\}$: $V_0 = \{c e ; c \in \mathbb{R}\}$ engendrée par e .
- Contrainte $\{a - b = 0\}$: $V_0 = \{b(x + z) + c e ; (b, c) \in \mathbb{R}^2\}$ engendré par $w = x + z$ et e .

On sait déjà tester une contrainte de nullité d'un coefficient. **En faisant éventuellement un changement de variable, toute contrainte scalaire ($r = 1$) peut se ramener à une contrainte de nullité d'un coefficient.**

Exemple : la contrainte $a - b = 0$ dans le modèle précédent.

$$a x + b z + c e = (a - b) x + b (z+x) + c e$$

Les nouvelles variables sont x , $w = (z+x)$ et e . Tester dans cette régression de y sur $(x w e)$ la nullité du coefficient de x , cela revient à tester l'égalité de a et b .

Exemple 2 : $y \sim N(a x + b z + c w + d e ; \sigma^2 I_{18})$, $(a, b, c, d) \in \mathbb{R}^4$

Contraintes $\{a - b + d = 0 ; 2a + c - d = 0\}$

Exemple

$y \sim N(a x + b z + c w + d e ; \sigma^2 I_N)$, $(a, b, c, d) \in \mathbb{R}^4$

- 2 Contraintes

$$\{a - b + d = 0 ; 2a + c - d = 0\} : \begin{pmatrix} 1 & -1 & 0 & 1 \\ 2 & -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = 0$$

entraînent $d = b - a$ et $c = b - 3a$

- modèle contraint

$$a x + b z + c w + d e = a (x - 3w - e) + b (z + w + e)$$

- modèle non contraint

$$a x + b z + c w + d e = a (x - 3w - e) + b (z + w + e) + (a - b + d)(w + e) + (2a + c - d) w$$

89

L'hypothèse H1 du modèle linéaire a été écrite sous la forme :

$$E(y) = X \beta, \beta \in \mathbb{R}^p$$

Cette propriété est en fait entièrement déterminée par la donnée du sous espace vectoriel $V = \{X \beta, \beta \in \mathbb{R}^p\}$: si on choisit une autre base de V , cela ne change pas l'ensemble des lois qui appartiennent au modèle.

Ainsi : $V = \{a x + b z + c w + d e, (a, b, c, d) \in \mathbb{R}^4\}$

et $V' = \{a' v_1 + b' v_2 + c' v_3 + d' v_4, (a', b', c', d') \in \mathbb{R}^4\}$

$$\text{où } v_1 = (x - 3w - e), v_2 = (z + w + e), v_3 = (w + e) \text{ et } v_4 = w$$

sont identiques, le changement de base correspondant au changement de coordonnées $a' = a, b' = b, c' = a - b + d, d' = 2a + c - d$.

Lorsqu'on calcule l'ajustement par les MCO, l'ajustement de y est le point de V qui est à la distance minimum de y . Cet ajustement ne change pas si on change de base, seules changent ses coordonnées. Restent donc inchangés dans un changement de base de V :

$$\hat{y} = P_V y, SCT = \|y - \bar{y}e\|^2, SCE = \|\hat{y} - \bar{y}e\|^2$$

$$\hat{u} = y - \hat{y}, SCR = \|\hat{u}\|^2$$

Nous considérons donc dans $\mathbb{R}^T V = \{X \beta ; \beta \in \mathbb{R}^p\}$, et $V_0 = \{X \beta ; C \beta = 0\}$ le sous espace contraint. Nous avons $u/\sigma \sim N(0, I_T)$.

L'ensemble des vecteurs orthogonaux à V s'appelle l'espace orthogonal à V dans \mathbb{R}^T , et nous le notons V^\perp . Il est de dimension $T - p$.

De même, V_0^\perp est de dimension $T - (p - r)$.

Test de contrainte linéaire $C\beta = 0$

(hors programme licence)

- Statistique utilisée : $F = \frac{(SCR_0 - SCR_1)/r}{SCR_1/(T-p)}$
- Sous H_0 $F \approx FISHER(r, T-p)$
- Au seuil α , refus de H_0 si $F > A_\alpha$.
ou si $PROB = P\{F[r, T-p] > F_{\text{observé}}\} < \alpha$

90

Deux résultats importants sur les lois des vecteurs normaux :

Soit $z \sim N(0; I_n)$, et V un sous espace vectoriel de \mathbb{R}^n , de dimension p ;
on note P_V l'opérateur de projection orthogonale sur V . On a alors :

$$\|P_V z\|^2 \sim \chi^2(p)$$

• Soit $z \sim N(0; I_n)$, V_1 et V_2 deux sous espaces vectoriels de \mathbb{R}^n .

Si $V_1 \perp V_2$, alors $P_{V_1} z$ et $P_{V_2} z$ sont indépendants en probabilité.

Ici :

$$\hat{y} = P_V y, \quad \hat{u} = (y - \hat{y}) \perp V, \quad \hat{u} = y - P_V y = X\beta + u - P_V X\beta - P_V u = P_{V^\perp} u$$

$$\tilde{y} = P_{V_0} y, \quad \tilde{u} = (y - \tilde{y}) \perp V_0, \quad \tilde{u} = y - P_{V_0} y = X\beta + u - P_{V_0} X\beta - P_{V_0} u$$

$$\tilde{u} = P_{V_0^\perp} u \text{ si } P_{V_0} X\beta = X\beta, \text{ c'est à dire si } C\beta = 0 \text{ (} X\beta \in V_0 \text{)}$$

On retrouve bien que $SCR/\sigma^2 = \|P_{V^\perp} u/\sigma\|^2 \sim \chi^2(T-p)$.

De même pour SCR_0 , sous H_0 , avec un degré de liberté de $T - (p-r)$. Mais les deux résidus ne sont pas indépendants. Il faut faire intervenir deux vecteurs orthogonaux.

$$\tilde{u} = y - \tilde{y} = y - \hat{y} + \hat{y} - \tilde{y} = \hat{u} + (\hat{y} - \tilde{y})$$

$$(\hat{y} - \tilde{y}) \in V, \text{ et } \hat{u} \perp V \Rightarrow (\hat{y} - \tilde{y}) \perp \hat{u}$$

$$\Rightarrow \|\tilde{u}\|^2 = \|\hat{u}\|^2 + \|\hat{y} - \tilde{y}\|^2 \Rightarrow \|\hat{y} - \tilde{y}\|^2 = SCR_0 - SCR_1 \text{ indépendant de } SCR_1$$

$$(\hat{y} - \tilde{y}) = P_V \tilde{u} = P_V P_{V_0^\perp} u = P_{V \cap V_0^\perp} u, \text{ où } V \cap V_0^\perp \text{ est de dimension } p - (p-r) = r$$

$$\left\| \frac{\hat{y} - \tilde{y}}{\sigma} \right\|^2 = \frac{SCR_0 - SCR_1}{\sigma^2} \approx \chi^2(r) \text{ est indépendant de } \frac{SCR_1}{\sigma^2} \approx \chi^2(T-p)$$

VI-2.a. Test de Fisher global (au programme de la licence)

- Modèle linéaire avec constante :

$$y \sim N(X\beta + c e ; \sigma^2 I_N), \beta \in \mathbb{R}^k \quad (\text{donc } p = k+1)$$

- les valeurs observées et les valeurs ajustées ont la même moyenne empirique
- Formule de décomposition de la variance

$$SCT = \sum_{n=1}^N (y_n - \bar{y})^2, SCE = \sum_{n=1}^N (\hat{y}_n - \bar{y})^2, SCR = \sum_{n=1}^N (y_n - \hat{y}_n)^2$$

$$SCT = SCE + SCR$$

91

Sous H_0 , la seule variable restante est la constante e : régresser y sur e , c'est simplement calculer la moyenne empirique des y_n , les résidus de régression sont les valeurs centrées empiriquement, et la SCR_0 est la somme des carrés des écarts à la moyenne des y_n , appelée SCT (somme des carrés totale)

Les valeurs observées ont même moyenne empirique que les valeurs ajustées. La somme des carrés des écarts des valeurs ajustées à leur moyenne empirique est appelée SCE (somme des carrés expliquée).

On se rappelle que $SCT = SCE + SCR$

Le coefficient de détermination R^2 qui le rapport entre les sommes de carrés expliquée et totale mesure la part de la variance des y_n qui est expliquée par les variables explicatives du modèle.

$$R^2 = SCE/SCT.$$

Le Test global de Fisher teste la nullité de β_1

$$F = \frac{SCE/(p-1)}{SCR/(T-p)}$$

• $H_0 : \{ \beta = 0 \}$ contre $H_1 : \{ \text{au moins un } \beta_j \neq 0 \}$

• Statistique utilisée

$$SCR_0 = SCT, \text{ et } SCR_0 - SCR_1 = SCT - SCR = SCE$$

• Sous H_0 , $F \sim \text{FISHER } [p-1, N-p]$

• Refus de H_0 si $F > A$

92

Sous H_0 , la seule variable restante est la constante e : régresser y sur e , c'est simplement calculer la moyenne empirique des y_t , les résidus de régression sont les valeurs centrées empiriquement, et la SCR_0 est la somme des carrés des écarts à la moyenne des y_t , appelée SCT.

On se rappelle que $SCT = SCE + SCR$ et que le coefficient de détermination R^2 est $R^2 = SCE/SCT$.

On peut donc réécrire F en fonction de R^2 :

$$F = \frac{R^2}{1-R^2} \frac{T-p}{p-1}$$

Le test s'appelle test de FISHER Global, car il teste globalement la significativité de toutes les variables. Le rejet de H_0 signifie qu'au moins une des variables est explicative.

C'est la F-statistique fournie dans tous les listings de régression, avec PROB égale à $P(F[p-1, T-p] > F_{\text{observé}})$.

La règle est alors de refuser H_0 si $\text{PROB} < \alpha$ choisi.

VI-2.b. Cas d'une contrainte scalaire

Soit à tester $\{a = 0\}$ contre $\{a \neq 0\}$, où a est une des coordonnées du paramètre β .

- statistique de Student : $t = \frac{a}{\hat{\sigma}_a}$

- décision $\{a \neq 0\}$ si

$$|t_{\text{observé}}| > t_\alpha \text{ où } P[|STUDENT(T-p)| > t_\alpha]$$

- expression équivalente : décision $\{a \neq 0\}$ si $p.value = P[|STUDENT(T-p)| > t_{\text{observé}}] < \alpha$

93

Le test de student a déjà été présenté dans le chapitre des tests.

C'est la statistique de Student qui est fournie dans les listings.

La « p-value » (ou « PROB ») donnée dans ces listings correspond au test bilatéral.

comparaison avec le test de Fisher

hors programme licence

- Test de Student : décision $\{a \neq 0\}$ si $\left| \frac{a}{\hat{\sigma}_a} \right| > t_a$
où $P\{|STUDENT(T-p)| > t_\alpha\} = \alpha$
 - Test de Fisher : décision $\{a \neq 0\}$ si $\frac{(SCR_0 - SCR)/1}{SCR/(T-p)} > A_\alpha$
où $P\{FISHER(1,T-p) > A_\alpha\} = \alpha$
 - Or $A_\alpha = (t_\alpha)^2$ et $\left(\frac{a}{\hat{\sigma}_a}\right)^2 = \frac{SCR_0 - SCR}{SCR/(T-p)}$
- Les deux tests sont donc identiques

94

$$FISHER(1,T-p) = (STUDENT(T-p))^2$$

$$P\{|STUDENT(T-p)| > t_\alpha\} = P\{[STUDENT(T-p)]^2 > [t_\alpha]^2\} = \alpha$$

entraîne bien $A_\alpha = (t_\alpha)^2$.

On démontre également l'égalité sur les statistiques utilisées. Les deux tests sont bien identiques. Le test de Student est plus fréquemment utilisé, car il permet de faire des tests unilatéraux.

C'est la statistique de Student qui est fournie dans les listings. La « PROB » donnée dans ces listings correspond au test bilatéral.

VI-2.c. Test d'hétérogénéité (hors programme licence)

- Modèle structurel : $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ et $u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ où $y_1, u_1 \in R^{n_1}$
 $y_2, u_2 \in R^{n_2}$
- $X = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix}$ où $\begin{cases} \text{format}(X_1) = (n_1, p) \\ \text{format}(X_2) = (n_2, p) \end{cases}$
- $\beta^* = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$ où $\beta_1 \in R^p$ et $\beta_2 \in R^p$
- H1 $E(y) = X\beta^*$, $\beta^* \in R^{2p}$
 X_1 et X_2 sont certaines, de rang p
- H2* $V(y_1) = \sigma_1^2 I_{n_1}$, $V(y_2) = \sigma_2^2 I_{n_2}$,
 $\text{cov}(y_1, y_2) = 0$
- H3N $y \sim$ loi Normale dans R^N .

95

Ce modèle correspond au cas où les $N = n_1 + n_2$ observations sont issues de deux populations différentes (ou de deux périodes différentes, on parle alors de changement de structure plutôt que d'hétérogénéité, avec n_1 observations d'un côté et n_2 observations de l'autre. Le coefficient de chacune des p variables explicatives dépend de la population où sont observées y et les x_j .

Les N égalités $y = X\beta + u$ peuvent aussi s'écrire en deux égalités vectorielles, de dimensions respectives n_1 et n_2 :

$$y_1 = X_1\beta_1 + u_1 \sim N(X_1\beta_1 ; \sigma_1^2 I_{n_1})$$

$$y_2 = X_2\beta_2 + u_2 \sim N(X_2\beta_2 ; \sigma_2^2 I_{n_2})$$

Dans chaque population, le modèle est linéaire standard Normal, mais pour qu'il le soit pour l'ensemble des données, il faut de plus que les variances soient égales (on admet a priori que y_1 et y_2 soient non corrélés et forment un seul vecteur normal de dimension N)

Avant de faire un test d'égalité des coefficients sur les deux populations, il faut faire le test d'égalité des variances.

(i) test d'égalité des variances :

- (1) MCO sur population 1

$$y_1 = X_1 \hat{\beta}_1 + \hat{u}_1, \hat{\sigma}_1^2 = \frac{SCR1}{n_1 - p}$$

- (2) MCO sur population 2

$$y_2 = X_2 \hat{\beta}_2 + \hat{u}_2, \hat{\sigma}_2^2 = \frac{SCR2}{n_2 - p}$$

- SCR1 et SCR2 sont indépendantes

$$F = \frac{SCR1 / (n_1 - p)}{SCR2 / (n_2 - p)} \underset{\text{si } \{\sigma_1 = \sigma_2\}}{\approx} FISHER((n_1 - p), (n_2 - p))$$

- On décide que $\{\sigma_1 \neq \sigma_2\}$ si $F > A$ ou $1/F > B$

96

Voir plus haut le test d'égalité des variances. On calcule A et B pour que la région critique soit de probabilité α (en attribuant $\alpha/2$ à chaque moitié de la région critique) : dans la table de FISHER correspondant à $\alpha/2$, on lit A qui correspond à un $F(n_1 - p, n_2 - p)$ et B qui correspond à un $F(n_2 - p, n_1 - p)$.

(ii) test de $\{\beta_1 = \beta_2\}$

- H_0 est fautive dès que l'une des variables x_j change de coefficient d'une population à l'autre.
- Régression sous contrainte : $y = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \beta + u$
SCR₀ à (N - p) degrés de liberté
- Régression non contrainte : $y = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + u$
SCR = SCR₁ + SCR₂
à (N - 2p) degrés de liberté
- On décide que $\beta_1 \neq \beta_2$ si $F > A$:
$$F = \frac{(SCR_0 - SCR_1 - SCR_2)/p}{(SCR_1 + SCR_2)/(N - 2p)} \Big|_{\beta_1 = \beta_2} \approx FISHER(p, N - 2p)$$

97

Le modèle structurel non contraint suppose ici que tous les coefficients de toutes les variables explicatives sont susceptibles de dépendre de la population. Dans ce cas, la forme spéciale (« bloc diagonale ») de X permet de calculer les résidus et la SCR sans être obligé de construire les vecteurs colonnes de X. Il suffit en effet d'utiliser les SCR obtenues dans les MCO séparés sur chaque population.

Il est indispensable que les variances soient les mêmes pour les deux populations.

Il y a p contraintes scalaires à tester simultanément (β est à p dimensions). Le degré de liberté du numérateur est bien égal à $N - p - (N - 2p) = p$.

Ce test peut se généraliser facilement au cas de plus de deux catégories d'observations.

Le test qui suit montre un modèle dans lequel il n'y a *a priori* que certaines variables dont le coefficient peut changer selon la catégorie des observations. La différence avec le cas ici traité est qu'il faudra construire effectivement la matrice X correspondant au cas non contraint (qui n'est plus bloc diagonale puisque certaines variables y interviennent sous une seule colonne avec un coefficient valable pour toutes les catégories).

VI-2.d. Test de composante saisonnière

(hors programme)

Modèle structurel :

$$e = Q_1 + Q_2 + Q_3 + Q_4$$

$$Q_{1t} = \begin{cases} 1 & \text{si } t \text{ est un 1er trimestre} \\ 0 & \text{sinon} \end{cases} \quad \dots \quad Q_{4t} = \begin{cases} 1 & \text{si } t \text{ est un 4ème trimestre} \\ 0 & \text{sinon} \end{cases}$$

- H1 (Q1 Q2 Q3 Q4 X) de rang $p = 4+k$
 $E(y) = c_1 Q_1 + c_2 Q_2 + c_3 Q_3 + c_4 Q_4 + X\alpha$, $\alpha \in \mathbb{R}^k$
- H2 $V(y) = \sigma^2 I_T$
- H3N $y \sim$ loi Normale dans \mathbb{R}^T .

Hypothèse testée : $H_0 = \{c_1 = c_2 = c_3 = c_4\}$

98

Sur données mensuelles, ce sont M_1, M_2, \dots, M_{12} indicatrices de chaque mois qu'il faut faire intervenir.

L'équation écrite en H1 signifie que le terme constant dépend du trimestre auquel la date t correspond : voir le TD1.

L'hypothèse H_0 signifie que ce terme constant est indépendant du trimestre (absence de composante saisonnière trimestrielle).

Remarque : au lieu d'utiliser les 4 indicatrices de trimestre, on pourrait faire intervenir le vecteur constant e , accompagné de trois des indicatrices, Q_2, Q_3 et Q_4 **par exemple**. L'ajustement MCO de y est inchangé, les résidus également, ainsi que la SCR, car $L(Q_1, Q_2, Q_3, Q_4, X) = L(e, Q_2, Q_3, Q_4, X)$.

Seuls les coefficients sont changés, suite au changement de base.

$$\begin{aligned} E(y) &= c_1 Q_1 + c_2 Q_2 + c_3 Q_3 + c_4 Q_4 + X\alpha \\ &= c_1 e + (c_2 - c_1) Q_2 + (c_3 - c_1) Q_3 + (c_4 - c_1) Q_4 + X\alpha \end{aligned}$$

Cette base permet de tester individuellement les contraintes

$$\{c_2 = c_1\}, \{c_3 = c_1\} \text{ et } \{c_4 = c_1\}.$$

Le test global des trois contraintes ne dépend pas de la base choisie.

Test de $\{c_1 = c_2 = c_3 = c_4\}$

- Régression non contrainte :
 $y = c_1 Q1 + c_2 Q2 + c_3 Q3 + c_4 Q4 + X\alpha + u$
SCR à $T - k - 4$ degrés de liberté
- Régression contrainte :
 $y = c e + X\alpha + u$
SCRo à $T - k - 1$ degrés de liberté
- il y a $r = 3$ contraintes
- Refus de H_0 si $F > A$, où :

$$F = \frac{(SCR_o - SCR)/3}{SCR/(T - k - 4)} \underset{H_0}{\approx} FISHER(3, T - k - 4)$$

99

Le **refus de H_0** signifie qu'au moins deux constantes sont différentes parmi les quatre constantes trimestrielles, et donc qu'**il y a un effet saisonnier**.

On peut vérifier que $r = T - k - 1 - (T - k - 4)$: le nombre de contraintes est bien la différence entre les degrés de liberté des SCR contrainte et non contrainte.

Conclusion : choix d'un modèle

- Ne garder que les modèles pour lesquels l'hypothèse du modèle linéaire standard est acceptable
- Pour chaque modèle, calculer la valeur du critère de qualité
$$AIC = \ln(\hat{\sigma}^2) + 2p \frac{\ln(\ln T)}{T}$$
- Choisir le modèle dont le critère est minimum

100

Le critère de choix correspond au désir de minimiser la variance résiduelle, mais avec le moins de paramètres possible.

La pénalité due au nombre p de paramètres doit tendre vers zéro lorsque T tend vers l'infini, mais pas trop vite.

La fonction $\ln(\ln T)/T$ est préférable à la fonction $1/T$ qui tend trop vite vers 0 (et qui correspond au critère « BIC »).