TESTS D'HYPOTHÈSE LINÉAIRE

Les ***tests d'hypothèse linéaire*** d'un modèle économétrique appartiennent au vaste domaine des ***tests statistiques***. On en rappelle le principe général.

● Une ou plusieurs grandeurs pour lesquelles on dispose d'observations sont supposées suivre un certain modèle, dit ***modèle de base*** ou ***alternatif*** et noté **Ha** (par exemple la taille d'un adulte est modélisée par une loi normale).
● On veut tester une hypothèse plus particulière, dite hypothèse $\mathbf{H_0}$, inspirée par la réflexion ou l'examen des données (par exemple la *taille moyenne* des hommes et celle des femmes sont-elles égales ?)
● Pour ce faire, on détermine une certaine grandeur, ou ***statistique***, dérivée des grandeurs initiales et calculable sur les observations, dont on connaît - au moins de manière approximative - la loi de distribution lorsque H_0 est vraie, (par exemple la différence normalisée des moyennes empiriques des hommes et des femmes examinés suit une loi de Student si l'hypothèse est correcte).
● On détermine la zone des valeurs les plus improbables pour cette loi de distribution, dite ***zone de rejet***, pour une probabilité totale choisie, dite ***niveau de risque*** (souvent 5%). La zone complémentaire est la ***zone d'acceptation*** et sa probabilité est le ***niveau de confiance,***
● Enfin, si la valeur, calculée sur les données, de la statistique choisie est dans la zone de rejet, on rejette l'hypothèse H_0 jugée trop improbable au niveau de risque choisi et on conserve l'hypothèse alternative H_a qui n'est pas remise en doute. Sinon, on accepte H_0.

HYPOTHÈSES LINÉAIRES

Hypothèse linéaire

Soit un modèle linéaire, une ***hypothèse linéaire*** (ou ***restriction linéaire)*** est un ensemble de une, ou plusieurs, conditions du premier degré, portant sur les coefficients.

Exemples : à partir du modèle de départ à quatre variables explicatives (incluant la constante), et d'aléa ε :

$$Y = a.X + b.Z + c.T + d + \varepsilon$$

on donne différents exemples d'hypothèses linéaires.

- $b = 0$ (une condition)
- $b = c$ (une condition)
- $b = -1$ et $c = 0$ (deux conditions)
- $a + c = 1$ et $d = 0$ (deux conditions)
- $a = b = c = 0$ (trois conditions)

Modèle transformé

Le modèle initial peut s'écrire sous une forme plus simple, traduisant algébriquement l'hypothèse linéaire.

Exemples : on donne les modèles transformés du modèle initial dans les cas précédents.

- $Y = a.X + c.T + d + \varepsilon$ (trois explicatives)
- $Y = a.X + b.(Z+T) + d + \varepsilon$ (trois explicatives)
- $Y+Z = a.X + d + \varepsilon$ (deux explicatives)
- $Y-T = a.(X-T) + b.Z + \varepsilon$ (deux explicatives)
- $Y = d + \varepsilon$ (une explicative)

Le nombre des variables explicatives est diminué du nombre de conditions élémentaires.

Le modèle transformé (par l'hypothèse linéaire) est un cas particulier du modèle initial.

Problème

Lorsqu'on estime le modèle initial par les mco, les estimations des coefficients peuvent satisfaire de manière plus ou moins approchée l'hypothèse linéaire envisagée, alors que les coefficients estimés sur le modèle transformé, par construction, vérifient exactement l'hypothèse linéaire.

On veut tester, ou éprouver, si l'hypothèse linéaire est acceptable, c'est à dire si le modèle transformé est correct (auquel cas, ses coefficients estimés seront retenus).

TEST DE FISHER D'UNE HYPOTHÈSE LINÉAIRE

Notations

Elles sont empruntées au domaine de la théorie des tests statistiques rappelée plus-haut.

- On note $\mathbf{H_0}$ l'hypothèse linéaire que l'on veut tester, et le modèle transformé correspondant.
- On note $\mathbf{H_a}$ le modèle initial (ou alternatif), supposé à priori correct.

Conditions

On suppose que le modèle satisfait aux hypothèses des mco, avec aléa normal (c'est à dire que les aléas ε_i sont indépendants et de même loi normale $\boldsymbol{N}(0,s)$).

Principe

Sous les conditions précédentes, on montre que, si l'hypothèse H_0 est vraie, la quantité, notée F :

$$\mathbf{F} = \frac{\dfrac{\mathbf{SCR_0 - SCR_a}}{\mathbf{dl_0 - dl_a}}}{\dfrac{\mathbf{SCR_a}}{\mathbf{dl_a}}}$$

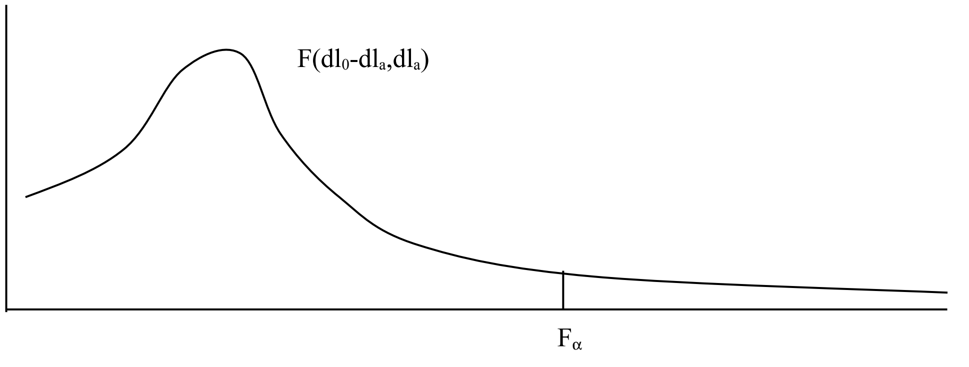
suit une ***loi de Fisher*** (dite aussi parfois de ***Fisher-Snedecor***) : $\mathbf{F(dl_0-dl_a,dl_a)}$, à dl_0-dl_a et dl_a degrés de liberté. SCR_a note la somme des carrés des résidus de la régression par les mco du modèle H_a, et dl_a le nombre de degrés de liberté, c'est à dire le nombre d'observations diminué du nombre d'explicatives (N-k), SCR_0 et dl_0 notent les quantités correspondantes du modèle H_0, et dl_0-dl_a est donc le nombre de conditions élémentaires.

Test de Fisher

Les lois de Fisher F(p,q), dérivées de la loi normale, étant tabulées, pour tester au niveau de risque: α, le modèle H_0, on examine la position de la quantité F calculée par rapport au seuil de rejet de niveau de risque α: F_α, pour la loi $F(dl_0-dl_a,dl_a)$.

- Si $F \leq F_\alpha$ on admet l'hypothèse linéaire et le modèle H_0 au risque α.

- Si $F > F_\alpha$ on juge la valeur obtenue de F trop improbable et on rejette H_0 au risque α (et on retient le modèle initial H_a).



Exemple : soit le modèle de demande de thé de Ceylan aux États-Unis:

$$\ln Q = a + b.\ln P_C + c.\ln P_I + d.\ln P_B + f.\ln R + \varepsilon$$

où Q note les importations de thé de Ceylan, P_C le prix du thé de Ceylan, P_I le prix du thé d'Inde, P_B le prix du café du Brésil, R le revenu national et ε l'aléa.

La régression par les mco donne l'équation estimée:

$$\underset{(2{,}000)}{\ln Q = 2{,}837} - \underset{(0{,}987)}{1{,}481}.\ln P_C + \underset{(0{,}690)}{1{,}181}.\ln P_I + \underset{(0{,}134)}{0{,}186}.\ln P_B + \underset{(0{,}370)}{0{,}257}.\ln R$$

avec N = 22 et SCR = 0,4277

On veut tester l'hypothèse: b = -1 et c = 0 (c'est à dire d'élasticité unitaire des quantités au prix et conjointement d'absence d'influence du prix du thé d'Inde). L'estimation du modèle transformé donne:

$$\ln Q + \ln P_C = \underset{(0{,}820)}{-0{,}738} + \underset{(0{,}155)}{0{,}199}.\ln P_B + \underset{(0{,}165)}{0{,}261}.\ln R \quad \text{et } SCR = 0{,}6788$$

on calcule: $F = \dfrac{(0{,}6788 - 0{,}4277)/2}{0{,}4277/(22-5)} = 4{,}99$

pour une loi de Fisher F(2,17), on lit d'autre part dans la table le seuil de rejet au risque 5%: $F_{0,05} = 3,59$. On rejette donc au risque 5% l'hypothèse linéaire envisagée.

CAS PARTICULIERS DU TEST DE FISHER

Test de nullité d'un coefficient

La nullité d'un coefficient est une hypothèse linéaire particulière (premier exemple), cependant le test de Fisher d'une telle hypothèse est inutile car il est mathématiquement équivalent au test de Student de significativité d'un coefficient, déjà présenté.

Test de nullité de tous les coefficients sauf la constante

Ce test radical de significativité minimum de la régression (cinquième exemple) est appelé *test F*, et il est calculé systématiquement par les logiciels économétriques. Concrètement, il conduit presque toujours au rejet de l'hypothèse testée et présente donc un intérêt limité.

Test de stabilité d'une partie des coefficients

On présente la méthode sur un exemple.

Soit un modèle temporel, expliquant Y par les variables X, Z, W et la constante, on suppose que les coefficients de X et de Z ont pu varier entre les deux sous-périodes: I et II.

Le modèle alternatif peut s'écrire, en dédoublant les variables X et Z:

$$H_a: \quad Y = a_I.X_I + a_{II}.X_{II} + b_I.Z_I + b_{II}.Z_{II} + c.W + d + \varepsilon$$

où X_I vaut X sur la sous-période I et 0 sur la sous-période II, et X_{II} l'inverse (soit $X_{II} = X - X_I$), de même pour Z_I et Z_{II}. L'aléa est ε.

L'hypothèse linéaire est: $a_I = a_{II}$ et $b_I = b_{II}$, et le modèle transformé:

$$H_0: \quad Y = a.X + b.Z + c.W + d + \varepsilon$$

Si N est le nombre d'observations, le test de H_0 revient à

comparer: $$F = \frac{(SCR_0 - SCR_a)/2}{SCR_a/(N-6)}$$ à une loi F(2,N-6).

Test de stabilité de tous les coefficients (test de Chow)

L'estimation du modèle alternatif, après dédoublement de l'ensemble des variables, revient en fait à faire deux régressions séparément sur les deux sous-périodes, tandis que le modèle traduisant la stabilité est le modèle simple sur toute la période.

Avec des notations naturelles, N observations et k explicatives, le test revient à comparer la quantité:

$$F = \frac{\dfrac{SCR_0 - (SCR_I + SCR_{II})}{k}}{\dfrac{SCR_I + SCR_{II}}{N-2k}}$$ à une loi F(k,N-2k).

<u>Exemple</u> : l'étude d'une fonction de consommation aux États-Unis a donné les trois régressions :

pour 1929-41: C = 282,56 + 0,69328.R où SCR = 3611,39
(39,6) (0,035) et dl = 11

pour 1946-70: C = 63,09 + 0,88286.R où SCR = 7399,05
(22,2) (0,011) et dl = 23

pour l'ensemble: C = 85,06 + 0,87104.R où SCR = 19839,62
(13,9) (0,008) et dl = 36

Pour tester la stabilité, on compare :

$$F = \frac{(19838{,}62 - 11010{,}44)/2}{11010{,}44/34} = 13{,}63$$ à une loi F(2,34)

le seuil de rejet au risque 5% est de l'ordre de 3,28 et on rejette donc l'hypothèse de stabilité.

On vient de voir comment effectuer « à la main » un test d'hypothèse linéaire (en régressant explicitement le modèle transformé puis en examinant la quantité F calculée à partir des résidus des deux régressions) ; les logiciels économétriques, tels SAS, Eviews ou Gretl, dispensent l'utilisateur de ces opérations : il suffit d'indiquer dans une syntaxe appropriée la ou les contraintes à tester, le logiciel calcule alors la valeur de F et indique à quel seuil d'acceptation/rejet elle correspond.

Il faut d'autre part garder à l'esprit que les tests de Fisher ne sont corrects qu'autant que les hypothèses des mco, avec normalité des aléas, sont satisfaites. Il convient donc en principe de s'en assurer par l'examen préalable des résidus.

Il est également important de noter que le rejet d'une hypothèse: H_0, à la suite d'un test de Fisher d'hypothèse linéaire, ne constitue aucunement une confirmation *a posteriori* de l'hypothèse initiale : H_a. Si celle-ci était douteuse, elle le reste: on teste H_0 contre H_a, **en supposant H_a vraie**.

----===ooOoo===----

(02.10.2008)