Problème de régression factice cointégration et modèle à correction d'erreur

I- Non stationnarité et régression factice

Lorsqu'on a étudié les modèles de régression simple et multiple en utilisant des séries temporelles on a supposé implicitement que les variables explicatives sont stat. Cependant, dans la pratique, de nombreuses séries macro-économique et financières sont non stationnaires et si on applique les méth habituelles de l'économétrie à des séries non stat, plusieurs prob se posent parmis lesquels, le prob de régression factice ou fallacieuse. ("spurious regression") mise en avant par Granger et Newbold (1974)

Des régressions factices sont souvent caractérisé par une R^2 élevée et une DW très faible indiquant une forte autocorr des erreurs d'autant plus que, le test de significativité individ des paramètres peuvent induire en erreur.

Exemples de régressions factices :

1/ Régression du taux de mortalité infantile en Egypte (MOR) sur le revenu des fermiers américains (REV) et sur l'offre de monnaie en ~~[illegible]~~ France (M), données annuelles 1971 - 1990.

$$\hat{MOR}_t = \underset{(26,65)}{179,9} - \underset{(-2,32)}{0,29\,REV_t} - 0,04\,M_t$$

t de student calculé

2/ Régression des exportations aux États-Unis (EXP) sur l'espérance de vie des hommes en Australie (VIE) données annuelles 1960 - 1990

$$\hat{EXP}_t = \underset{(-16,70)}{-2943} + \underset{(17,75)}{45,79\,VIE}$$

avec : $R^2 = 0,916$ et $DW = 0,36$

ces exemples sont ainsi illustratifs de régression fallacieuse c'à'd de régression dénué de sens. [1/(aucun sens)]

Cela provient essentiellement de la non stationnarité des différentes séries entrant en jeu dans les regressions. On montrera par ailleurs 2 caractéristiques communes ou 2 exp de regressions, d'une part, le coef de détermination est très elevé (>0,9) et d'autre part, la valeur de la stat de DW est très faible montrant une forte auto-corrélation des 2. Ces 2 carat sont symptomatiques de regression falacieuse. Une procedure très freq utilisée pour éviter le prob de reg factice consiste alors à différencier les séries non stat afin de les stationnariser et de pouvoir appliquer les méthodes habituelles de l'économétrie cependant, cette opération de diff a pour limite essentiel de masquer les prop de long terme des séries etudiées puisque les relations entre les var en niveaux ne sont plus considérés. La théorie de cointégration permet de pallier ce problème en offrant la possibilité de spécifier des relations stables à long terme tout en analysant conjointement la dynamique de ct des variables en qst.

II – Le concept de cointégration :

1) Définition :

Si x_t et y_t sont deux séries intégrées d'ordre d $I(d)$ alors en général la combinaison linéaire de z_t tel que :

$$z_t = y_t - \alpha x_t \text{ est aussi } I(d)$$

Cependant, il est possible que z_t ne soit pas $I(d)$ mais $z_t \sim I(d-b)$ avec b étant un entier positif $(d \geq b > 0)$. En d'autres termes z_t est intégrée d'un ordre inférieur à l'ordre d'intégration des variables considérés, dans ce cas x_t et y_t sont cointingrées et on note $(x_t, y_t) \sim CI(d, b)$

α est le param de cointégration et le vecteur $(1, -\alpha)$ c'est le vecteur de cointégration

On considère maintenant le cas le plus fréquemment utilisé qui correspond à $d = b = 1$, dans ce cas deux séries non stationnaires et intégrées d'ordre 1 sont cointégrées s'il existe une combinaison linéaire stationnaire c-à-d $I(0)$ de ces deux séries

si $x_t \sim I(1)$ et $y_t \sim I(1)$, alors si $z_t \sim I(0)$, y_t et x_t sont cointegrées

2) Test de cointégration :

On va se focaliser sur l'algo en deux étapes d'Engle et Granger (1987)

1– étape Il s'agit de tester l'ordre d'integ des deux variables.

Une cond necessaire de cointégration

et que les séries doivent être intégrées de même ordre

si les séries ne sont pas intégrées de m^e ordre, elles ne sont pas cointégrées. ~~Ile~~

Il convient donc de déterminer soigneusement l'ordre d'integ d ~~que~~ pour cela on utilise les tests de racine unité tq ~~A~~DF et ADF.

si les séries étudiées ne sont pas intégrées de m^e ordre, la procédure est arrêtée et il n'y a pas de risque de cointégration

Etape 2 : Estimer la relation de long terme entre y_t et x_t

si la condition nécessaire est vérifiée, on estime par les MCO la relation de long terme entre les variables tq

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

Pour que la relation de cointégration soit acceptée le résidu de l'estimation $\hat{\varepsilon}_t$ doit être stationnaire

résidu : $\hat{\varepsilon}_t = y_t - \hat{\alpha} - \hat{\beta} x_t$

$\Rightarrow \hat{\varepsilon}_t \sim I(0)$: stationnaire

La stationnarité des résidus est testée à l'aide du test DF ou ADF

Cependant puisque le test porte sur les résidus estimés et non pas sur les vrais résidus de la relation de cointégration, on ne peut plus utiliser les tables de DF dans ce cas.

Table d'Engle et Yoo (1987)

Table de MacKinnon (1991)

=> voir drive !

Ces tables dépendent du nombre d'obs T et du nombre de variables N (la variable à expl y + les var explicatives les x figurant dans la regression)

Règle de décision :

si les résidus sont non stationnaires, la relation estimée est une régression factice.

- si les résidus sont stationnaires, la relation estimée est une relation de cointégration, et on peut alors estimer un modèle à correction d'erreurs.

III - Le modèle à correction d'erreurs ECM et l'approche en deux étapes d'Engle et Granger :

1) Le théorème de représentation de Granger (1981) :

[a] Une des propriétés fondamentales des

séries cointégrés et qu'elles peuvent être modélisées sous la forme d'un modèle à correction d'erreurs, ce résultat a été démontré dans le cadre de théorème de Granger ~~(1987)~~ de représentation valable pour les séries cointégrées d'ordre 1 $CI(1,1)$, de tels modèles permettent de modéliser les ajustements qui conduisent à une situation d'équilibre de long terme il s'agit ainsi de modèles dynamiques qui intègrent à la fois les évolutions du long terme et du court terme

Enoncé du théorème:
si y_t et x_t sont $I(1)$ et cointégrés, alors il existe une représentation à correction d'erreur de la forme

$$\Delta y_t = \gamma_0 + \gamma_1 \Delta x_t + c\hat{z}_{t-1} + \varepsilon_t$$

ou bien:

$$\Delta y_t = \gamma_0 + \gamma_1 \Delta x_t + c\underbrace{\left[y_{t-1} - \hat{\alpha} - \hat{\beta} x_{t-1}\right]}_{z_{t-1}}$$

avec $\hat{z}_{t-1}$ = résidu d'estimation de la relation de cointégration

$$\Rightarrow \hat{z}_{t-1} = y_{t-1} - \hat{\alpha} - \hat{\beta} x_{t-1}$$

$$\Rightarrow \text{ECM}: \Delta y = \gamma_0 + \gamma_1 \Delta x_t + c(y_{t-1} - \hat{\alpha} - \hat{\beta} x_{t-1}) + u_t$$

* Définition des termes du modèle d'équilibre

• $(y_{t-1} - \hat{\alpha} - \hat{\beta} x_{t-1})$: la relation de long terme. Il s'agit de la relation de cointégration.

• $\gamma_1 \Delta x_t$: il s'agit de la composante dynamique de court terme du modèle.

• c : il représente la vitesse d'ajustement de la variable endogène vers l'équilibre de long terme.

2) Estimation et validation de (modèle) l'ECM : méthode en deux étapes d'Engle et Granger (1987) :

L'estimation de l'ECM peut se faire en deux étapes selon l'approche d'Engle et Granger cette méthode est valable pour des séries $CI(1,1)$

* 1ère étape:

Estimation de la relation de long terme (LT) par les MCO

$$y_t = \alpha + \beta x_t + z_t$$

où z_t est le terme d'erreur

- Si les variables sont cointégrés, on passe à la seconde étape

* 2ème étape: Estimation par les MCO de l'ECM

$\Delta y_t = \gamma_0 + \gamma_1 \Delta x_t + c\,\hat{z}_{t-1} + u_t$

l'estimation de l'ECM ne pose pas de problèmes particuliers et nécessite simplement le recours à une technique des moindres carrés ordinaires puisque la procédure en deux étapes conduit à une estimation convergente des coeff du modèle qui peuvent s'interpréter de manière classique

* Validation de l'ECM:

  * si $c \leq 0$: force de rappel de l'ECM
  * si $c \geq 0$: force de répulsion et le déséquilibre est encore persistant

le coefficient c doit être significatif et négatif dans le cas contraire il convient de rejeter une spécification de type ECM puisque dans ce cas il n'existe pas de phénomène de retour à l'équilibre.

Le modèle ECM permet ainsi d'intégrer les fluctuations de ~~long~~ court terme au cours de l'équilibre de long terme donné par la relation de cointégration il décrit ainsi un processus d'ajustement et combine deux types de variables:

- les variables en différence première et donc stationnaire qui représente les fluctuations de court terme.
- les variables en niveau (ici $\hat{z}_{t-1}$) qui est la combinaison linéaire stationnaire de variables non stationnaires qui assure la prise en compte du long terme.

Il est à noter que le modèle ECM peut contenir des variables en différences premières retardées (comme variables explicatives) et aussi la variable à expliquer en différence première retardée

IV - La cointégration dans le cas de N variables:

Jusque là on a étudié le concept de cointégration en considérant uniquement deux variables y_t et x_t (autrement dit en considérant une seule variable explicative) mais dans la réalité il peut y avoir plus qu'une seule variable explicative donc on peut étudier la cointégration entre N variables c.à.d la variable à expliquer y et les k variables explicatives x donc on peut avoir (N-1) relations de cointégration et non plus une seule.

Dans un modèle économétrique à k variables explicatives :

$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_k x_{kt} + \varepsilon_t$

- Si les variables (y_t et x_{kt}) sont toutes $I(1)$ alors il existe une possibilité de cointégration entre ces variables.

- Dans ce cas :

* La relation de LT :

$$\hat{\varepsilon}_t = y_t - \hat{\beta}_0 - \hat{\beta}_1 x_{1t} - \dots - \hat{\beta}_k x_{kt}$$

si $\hat{\varepsilon}_t \rightsquigarrow I(0)$, alors il existe une relation de cointégration sinon, la relation estimée est une régression factice.

l'ECM s'écrit alors comme suit :

$$\Delta y_t = \gamma_0 + \gamma_1 \Delta x_{1t} + \gamma_2 \Delta x_{2t} + \dots + \gamma_k \Delta x_{kt} + c\,\hat{\varepsilon}_{t-1} + u_t$$

- Si les variables sont de même ordre d'integ $I(1)$ par exemple, l'existence d'un seul vecteur de cointégration est possible

$\Rightarrow$ vecteur de cointeg : $(1, -\beta_0, -\beta_1, -\beta_k)$

En revanche si les séries ne sont pas toutes intégrées du même ordre nous pouvons être certain que le vecteur de cointégration n'est pas unique il peut donc y avoir $N-1$ vecteurs de cointégration ou k vecteurs. Le nombre de vecteur de cointégration est appelé le rang de la cointégration s'il existe plusieurs vecteurs de cointégration, on ne peut pas appliquer l'approche en deux étapes d'Engle et Granger ni pour tester l'existence d'une relation de cointégration ni pour estimer le modèle à correction d'erreur. Dans ce cas nous devons faire appel à l'approche vectorielle à correction d'Erreur (VECM) et pour tester la relation de cointégration l'existence de on doit recourir aux tests de Johansen (1988)

- Test de la trace
- test de la valeur propre maximale

* Remarque :

Dans ce cours, nous ne faisons que citer les techniques vectorielles de cointégration et de mécanisme à correction d'erreur, le lecteur intéressé peut se référer à Greene (2005) et Bourbonnais (2009)