MODELE A CORRECTION D'ERREUR
LA METHODE D'ENGEL ET GRANGER

1 Cas simple:

1.1 Présentation

Prenons un modèle dit 'de *long terme*' avec deux variables y, x toutes deux I(1)

$$y = ax + c + \epsilon$$

Ce modèle ramène à un problème de cointégration. Si il y a cointégration l'erreur est stationnaire et on peut faire les MCO dans ce modèle , l'équation de long terme existe donc (s'il n'y a pas de cointégration il n'existe pas de relation de long terme entre ces variables). On construit le modèle de court terme correspondant autorégressif et à retards échelonnés avec l'hypothèse u_t BB.
Supposons qu'un seul retard r soit suffisant pour que u_t soit un BB

$$y_t = \beta_0 x_t + \beta_1 x_{t-1} + \alpha_1 y_{t-1} + c_0 + v_t$$

- Lien entre l'équation de court terme et l'équation de long terme

  si t$\rightarrow \infty$ $y_t \simeq y_{t-1}$ et $x_t \simeq x_{t-1}$ le modèle ci-dessus peut s'écrire

$$\begin{aligned} y_t &= \beta_0 x_t + \beta_1 x_t + \alpha_1 y_t + c_0 + u_t \\ (1-\alpha_1) y_t &= (\beta_0 + \beta_1) x_t + c_0 + u_t \\ y_t &= \frac{(\beta_0 + \beta_1)}{(1-\alpha_1)} x_t + \frac{c_0}{(1-\alpha_1)} + \frac{u_t}{(1-\alpha_1)} \end{aligned}$$

  On retrouve l'équation de long terme avec

$$a = \frac{(\beta_0 + \beta_1)}{(1-\alpha_1)} \quad \text{et} \quad c = \frac{c_0}{(1-\alpha_1)} \tag{1}$$

  On peut regrouper les termes du modèle de court terme

$$\begin{aligned} y_t &= \beta_0 (x_t - x_{t-1}) + \beta_0 x_{t-1} + \beta_1 x_{t-1} + \alpha_1 y_{t-1} + c_0 + u_t \\ y_t - y_{t-1} &= -y_{t-1} + \beta_0 (x_t - x_{t-1}) + \beta_0 x_{t-1} + \beta_1 x_{t-1} + \alpha_1 y_{t-1} + c_0 + u_t \\ \Delta y_t &= \beta_0 \Delta x_t - (1-\alpha_1) y_{t-1} + (\beta_0 + \beta_1) x_{t-1} + c_0 + u_t \\ \Delta y_t &= \beta_0 \Delta x_t - (1-\alpha_1)(y_{t-1} - \frac{(\beta_0 + \beta_1)}{1-\alpha_1} x_{t-1} - \frac{c_0}{1-\alpha_1}) + u_t \\ \Delta y_t &= \beta_0 \Delta x_t - (1-\alpha_1)(y_{t-1} - a x_{t-1} - c) + u_t \end{aligned}$$

1.2 Définition du Modèle à correction d'erreur MCE

Ce modèle de court terme s'écrit donc : écarts de y en fonction des écarts de x et des erreurs du modèle de long terme en $(t-1)$

$$\Delta y_t = \beta_0 \Delta x_t - (1-\alpha_1)\epsilon_{t-1} + u_t$$

- Comme y et x sont I(1) , leurs écarts sont I(0). Les erreurs ϵ_{t-1} doivent donc être stationnaires pour que les MCO soient applicables à ce modèle , ce qui conduit à avoir un modèle de long terme cointégré. Ce modèle de court terme est dit modèle à correction d'erreurs.

Pour qu'un tel modèle existe il faut donc $(1-\alpha_1) \neq 0$.

1.3 Remarque:

Montrons que si le modèle de long terme n'est pas cointégré alors $(1-\alpha_1) = 0$ et alors on obtient le modèle classique (passer en écarts si les variables sont I(1) :

$$\Delta y_t = \beta_0 \Delta x_t + u_t$$

Démonstration: pour tester la cointégration du modèle de long terme on utilise la méthode de D. F. en construisant

$$\Delta\epsilon_t = \gamma\epsilon_{t-1} + v_t$$

et montrons que $\gamma = 1-\alpha_1$ .Si on reprend l'équation

$$\Delta y_t = \beta_0 \Delta x_t - (1-\alpha_1)(y_{t-1} - a x_{t-1} - c) + u_t \tag{2}$$

Elle peut s'écrire en retranchant $a\Delta x_t$

$$\begin{aligned}
\Delta y_t - a\Delta x_t &= -a\Delta x_t + \beta_0\Delta x_t - (1-\alpha_1)(y_{t-1} - a x_{t-1} - c) + u_t \\
\Delta(y_t - a x_t - c) &= (\beta_0 - a)\Delta x_t - (1-\alpha_1)(\epsilon_{t-1}) + u_t \\
\Delta(\epsilon_t) &= -(1-\alpha_1)(\epsilon_{t-1}) + (\beta_0 - a)\Delta x_t + u_t \\
\text{avec} \quad (\beta_0 - a)\Delta x_t + u_t &= v_t
\end{aligned}$$

on retrouve donc $\gamma = 1-\alpha_1$
Donc si il n'y a pas cointégration le test D. F. décide H_0 et donc $\gamma = 0$ soit $1-\alpha_1 = 0$

1.4 Conclusion:

- Si les variables I(1) sont cointégrées, il existe entre elles une relation de long terme. On peut alors construire une équation de court terme en prenant les écarts des variables et l'erreur du modèle de long terme en (t-1). c'est le modèle à correction d'erreur.

- Si les variables I(1) ne sont pas cointégrées, le modèle de court terme (CT) se réduit à un modèle sur les écarts des variables. C'est le seul modèle que l'on puisse construire avec ces variables car il ne contient alors que des variables I(0) les erreurs par conséquent étant aussi I(0).

2 Cas plus général

2.1 Présentation

Soit un modèle dit 'de *long terme*' avec ces variables y, x, w toutes I(1).

$$y = ax + bw + c + \epsilon \tag{3}$$

Ce modèle ramène à un problème de cointégration. Si il y a cointégration, l'erreur est stationnaire et on peut faire les MCO dans ce modèle , l'équation de long terme existe donc (s'il n'y a pas de cointégration il n'existe pas de relation de long terme entre ces variables).
Qu'il y ait cointégration ou non on peut toujours construire un modèle de *court terme*.
Soit le modèle autorégressif et à retards échelonnés où l'on met autant de retards r qu'il faut pour que u_t soit un BB

$$y_t = \sum_{i=0}^{r} \beta_i x_{t-i} + \sum_{i=1}^{r} \alpha_i y_{t-i} + \sum_{i=0}^{r} \gamma_i w_{t-i} + c_0 + u_t$$

2.2 On va traiter le cas particulier $r = 2$

$$y_t = \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \gamma_0 w_t + \gamma_1 w_{t-1} + \gamma_2 w_{t-2} + c_0 + u_t$$

Sans changer l'équation on peut regrouper les termes de la façon suivante:

$$\begin{aligned}
\Delta y_t &= \beta_0 \Delta x_t + (\beta_0 + \beta_1 + \beta_2) x_{t-1} - \beta_2 \Delta x_{t-1} + \gamma_0 \Delta w_t + (\gamma_0 + \gamma_1 + \gamma_2) w_{t-1} \\
&\quad -\gamma_2 \Delta w_{t-1} + (\alpha_1 + \alpha_2 - 1) y_{t-1} - \alpha_2 \Delta y_{t-1} + c_0 + u_t \\
\Delta y_t &= \beta_0 \Delta x_t + (\beta_0 + \beta_1 + \beta_2) x_{t-1} - \beta_2 \Delta x_{t-1} + \gamma_0 \Delta w_t + (\gamma_0 + \gamma_1 + \gamma_2) w_{t-1} \\
&\quad -\gamma_2 \Delta w_{t-1} + (\alpha_1 + \alpha_2 - 1) y_{t-1} - \alpha_2 \Delta y_{t-1} + c_0 + u_t
\end{aligned}$$

En regroupant les coefficients en $t - 1$ qui ne sont pas en écart et le terme constant :

$$\begin{aligned}
\Delta y_t &= \beta_0 \Delta x_t + \gamma_0 \Delta w_t - (1 - \alpha_1 - \alpha_2)(y_{t-1} - a x_{t-1} - b w_{t-1} - c) \\
&\quad -\beta_2 \Delta x_{t-1} - \alpha_2 \Delta y_{t-1} - \gamma_2 \Delta w_{t-1} + u_t
\end{aligned}$$

avec $a = (\beta_0 + \beta_1 + \beta_2)/(1 - \alpha_1 - \alpha_2)$ et $b = (\gamma_0 + \gamma_1 + \gamma_2)/(1 - \alpha_1 - \alpha_2)$

$$\begin{aligned}
\Delta y_t &= \beta_0 \Delta x_t + \gamma_0 \Delta w_t - (1 - \alpha_1 - \alpha_2)(\epsilon_{t-1}) \\
&\quad -\beta_2 \Delta x_{t-1} - \alpha_2 \Delta y_{t-1} - \gamma_2 \Delta w_{t-1} + u_t
\end{aligned} \tag{4}$$

2.3 Dans le cas r général

On obtient le résultat suivant

$$\Delta y_t = \beta_0 \Delta x_t + \gamma_0 \Delta w_t - (1 - \sum_{1}^{r} \alpha_i)(\epsilon_{t-1}) + \sum_{i=1}^{r-1} a_i \Delta x_{t-i} + \sum_{i=1}^{r-1} b_i \Delta y_{t-i} + \sum_{i=1}^{r-1} c_i \Delta w_{t-i} + u_t \tag{5}$$

$$\Delta y_t = \sum_{i=1}^{r-1} b_i \Delta y_{t-i} - (1 - \sum_{1}^{r} \alpha_i)(\epsilon_{t-1}) + \sum_{i=0}^{r-1} a_i \Delta x_{t-i} + \sum_{i=0}^{r-1} c_i \Delta w_{t-i} + u_t \tag{6}$$

comme y, x et w sont I(1) leurs écart sont I(0) , pour que le modèle existe il faut donc l'une des deux conditions

- soit que ϵ soit stationnaire c'est-à-dire que les variables soient cointégrées
- soit que $(1 - \sum_1^l \alpha_i) = 0$ qui est équivalent à dire que les variables ne sont pas cointégrées.

Voici une écriture plus générale qui peut s'appliquer à un nombre plus important de variables explicatives

avec $\overrightarrow{h_t} = \begin{pmatrix} x_t \\ w_t \end{pmatrix}$ $\quad \overrightarrow{m} = \begin{pmatrix} \beta_0 \\ \gamma_0 \end{pmatrix}$ $\quad \overrightarrow{A_1} = \begin{pmatrix} \beta_0 + \beta_1 + \beta_2 \\ \gamma_0 + \gamma_1 + \gamma_2 \end{pmatrix}$ $\quad \overrightarrow{A_2} = \begin{pmatrix} -\beta_2 \\ -\alpha_2 \\ -\gamma_2 \end{pmatrix}$

$\overrightarrow{z} = \begin{pmatrix} \Delta x_t \\ \Delta y_t \\ \Delta w_t \end{pmatrix}$

$$\Delta y_t = \overrightarrow{m}'\overrightarrow{\Delta h_t} - (1 - \alpha_1 - \alpha_2)(y_{t-1} - B'\overrightarrow{h_{t-1}}) + A_2'\overrightarrow{\Delta z_{t-1}} + c_0 + u_t$$

avec $B' = \frac{1}{(1-\alpha_1-\alpha_2)} A_1'$
en généralisant ce résultat à r retards on obtient

$$\Delta y_t = \overrightarrow{m}'\overrightarrow{\Delta h_t} + \gamma(y_{t-1} - B'\overrightarrow{h_{t-1}} + cte) + \sum_{i=1}^{r-1} A_i^t \overrightarrow{\Delta z_{t-i}} + u_t \tag{7}$$

si y x et w sont I(1) leur écarts sont stationnaires, pour que les erreurs u soient stationnaires il faut que $(y_{t-1} - B'\overrightarrow{h_{t-1}} + cte)$ soit stationnaire c'est-à-dire que les variables y,x et z soient cointégrées.

2.4 *Comment construire une équation à correction d'erreur*

1. estimer le modèle de long terme

$$y_t = a_1 x_t + a_2 w_t + a_3 + \epsilon_t$$

2. construire le test de cointégration de D.F. à l'aide des résidus de ce modèle (voir chapitre sur la cointégration)
3. construire le modèle de court terme

$$y_t = \sum_{i=0}^{r} \beta_i x_{t-i} + \sum_{i=1}^{r} \alpha_i y_{t-i} + \sum_{i=0}^{r} \gamma_i w_{t-i} + c_0 + u_t \tag{8}$$

en choisissant le nombre r de retards pour que u_t soit un BB

4. transformer cette équation à fin de faire intervenir l'erreur du modèle de long terme

$$\Delta y_t = \beta_0 \Delta x_t + \gamma_0 \Delta w_t - (1 - \sum_1^r \alpha_i)(\epsilon_{t-1}) + \sum_{i=1}^{r-1} a_i \Delta x_{t-i} + \sum_{i=1}^{r-1} b_i \Delta y_{t-i} + \sum_{i=1}^{r-1} c_i \Delta w_{t-i} + u_t \tag{9}$$

- s'il n'y a pas cointégration alors

$$\gamma = (1 - \sum_1^r \alpha_i) = 0$$

  et l'équation devient un simple modèle en écart.

- s'il y a cointégration alors on estime la variable ϵ_{t-1} par les résidus de l'équation de long terme en t-1 et on estime ce modèle à correction d'erreur. C'est ce qu'on nomme la méthode en deux temps. On peut aussi estimer globalement (8) en estimant donc par des méthodes non linéaires car on aura alors des produits de coefficients à estimer

$$\begin{aligned} \Delta y_t &= \beta_0 \Delta x_t + \gamma_0 \Delta w_t - (1 - \sum_1^r \alpha_i)(y_{t-1} - a_1 x_{t-1} - a_2 w_{t-1} - a_3) + \sum_{i=1}^{r-1} a_i \Delta x_{t-i} + \sum_{i=1}^{r-1} b_i \Delta y_t \\ &+ \sum_{i=1}^{r-1} c_i \Delta w_{t-i} + u_t \end{aligned}$$

  Bien sur la première méthode est la plus utilisée.

2.5 Présence de séries I(0)

Elles peuvent alors

- si elles font parti de l'équation de long terme, apparaitre à la fois dans l'équation de long terme et dans les séries en écart exactement comme une série I(1)
- soit aussi venir enrichir le modèle de variables n'intervenant pas dans l'équation de long terme. On peut ainsi mettre le temps par exemple ou toute autre série exogène.. Ainsi, lorsque ces variables sont I(1) il faut prendre leur écart avec d'éventuels retards. Elles ne joueront aucun rôle si $t \to \infty$ car leur écart tend vers 0.

$$\Delta y_t = \beta_0 \Delta x_t + \gamma_0 \Delta w_t - (1 - \sum_1^r \alpha_i)(\epsilon_{t-1}) + \sum_{i=1}^{r-1} a_i \Delta x_{t-i} + \sum_{i=1}^{r-1} b_i \Delta y_{t-i} + \sum_{i=1}^{r-1} c_i \Delta w_{t-i} + u_t + \text{ variables I(0)}$$

2.6 Présence de séries I(1) ou I(2) supplémentaires

D'autres séries, autres que celles figurant dans l'équation de long terme peuvent enrichir le modèle (9).
Mais bien sur elles doivent être présentes sous leur forme I(0) c'est-à-dire en écart si elles sont I(1) ou en écart d'écart si elles sont I(2). On constate bien alors qu'elles ne jouent pas de rôle dans le long terme car si t$\to \infty$ alors les écarts tendent vers 0.

3 Exemple:

Reprenons la consommation des ménages aux USA (voir les résultats dans la révision de la cointégration). Toutes les variables sont I(1) nous avons montré qu'il y a cointégration entre les variables CM, RD, TCHO et SP. On peut donc construire le modèle de court terme à correction d'erreurs.

3.1 Construction du modèle de court terme

On recherche le nombre de retards r qui conduit à un modèle sans autocorrélation des erreurs

$$CM_t = \sum_{i=1}^{r} a_i CM_{t-i} + \sum_{i=0}^{r} b_i RD_{t-i} + \sum_{i=0}^{r} c_i TCHO_{t-i} + \sum_{i=0}^{r} d_i SP_{t-i} + cte + u_t$$

En faisant des essais ou en utilisant un logiciel spécifique on trouve qu'avec r=4 on obtient le résultat souhaité, les erreurs u sont non autocorrélées ni d'ordre 1, ni d'ordre >1.

```
Linear Regression - Estimation by Least Squares
Dependent Variable CM
Quarterly Data From 1955:02 To 2000:04
Usable Observations    183      Degrees of Freedom   163
Centered R**2     0.999891      R Bar **2   0.999878
Uncentered R**2   0.999983      T x R**2     182.997
Mean of Dependent Variable      3209.9207650
Std Error of Dependent Variable 1391.0463403
Standard Error of Estimate        15.3397356
Sum of Squared Residuals        38355.120620
Regression F(19,163)              78762.2171
Significance Level of F           0.00000000
Log Likelihood                    -748.74763
Durbin-Watson Statistic             1.959350

   Variable                     Coeff       Std Error      T-Stat     Signif
*******************************************************************************
1.  Constant                   5.59296934   6.55928195      0.85268  0.39508788
2.  CM{1}                      0.78575570   0.07911383      9.93196  0.00000000
3.  CM{2}                      0.29395194   0.10351572      2.83968  0.00509142
4.  CM{3}                      0.06451453   0.10445415      0.61763  0.53767792
5.  CM{4}                     -0.27478001   0.07886643     -3.48412  0.00063425
6.  RD                         0.23748705   0.04147666      5.72580  0.00000005
7.  RD{1}                     -0.10485490   0.05528704     -1.89655  0.05965568
8.  RD{2}                     -0.06962233   0.05600285     -1.24319  0.21558207
9.  RD{3}                     -0.03645838   0.05521304     -0.66032  0.50997931
10. RD{4}                      0.08600369   0.04684147      1.83606  0.06817084
11. TCHO                     -23.74081284   4.78370511     -4.96285  0.00000174
12. TCHO{1}                   31.81047552   8.61563525      3.69218  0.00030292
13. TCHO{2}                   -0.78054896   9.26140058     -0.08428  0.93293740
14. TCHO{3}                   -9.39910466   8.31771261     -1.13001  0.26013179
15. TCHO{4}                    3.31633567   4.48845178      0.73886  0.46105485
16. SP                         0.23732372   0.06161362      3.85181  0.00016822
17. SP{1}                     -0.06368870   0.11246133     -0.56632  0.57195758
18. SP{2}                     -0.01772643   0.11715106     -0.15131  0.87991623
19. SP{3}                     -0.17857706   0.11718688     -1.52387  0.12948010
20. SP{4}                      0.07418708   0.08311711      0.89256  0.37340819
```

3.2 Modèle à correction d'erreur (MCE)

$$\Delta CM_t = \sum_{i=1}^{4-1} a_i \Delta CM_{t-i} - (1 - \sum_{1}^{4} \alpha_i)(\epsilon_{t-1}) + \sum_{i=0}^{4-1} b_i \Delta RD_{t-i} + \sum_{i=0}^{4-1} c_i \Delta TCHO_{t-i} + \sum_{i=0}^{4-1} d_i \Delta SP_{t-i} + cte + u_t$$

```
Linear Regression - Estimation by Least Squares
Dependent Variable DCM
Quarterly Data From 1955:01 To 2000:04
Usable Observations    183      Degrees of Freedom   166
 Total Observations    184      Skipped/Missing        1
Centered R**2     0.631066      R Bar **2   0.595506
Uncentered R**2   0.837586      T x R**2     153.278
Mean of Dependent Variable      27.369945355
Std Error of Dependent Variable 24.338490375
Standard Error of Estimate      15.479241639
Sum of Squared Residuals        39774.749008
Regression F(16,166)                 17.7466
Significance Level of F           0.00000000
Log Likelihood                    -752.07313
Durbin-Watson Statistic             1.965060

   Variable                     Coeff       Std Error      T-Stat     Signif
*******************************************************************************
1.  Constant                   7.73565179   2.70286717      2.86202  0.00475193
2.  DCM{1}                    -0.05012782   0.07869576     -0.63698  0.52501374
3.  DCM{2}                     0.24905417   0.07632611      3.26303  0.00133830
4.  DCM{3}                     0.30329827   0.07662395      3.95827  0.00011174
5.  RES{1}                    -0.13811715   0.04095709     -3.37224  0.00092774
6.  DRD                        0.24744898   0.04154404      5.95631  0.00000001
7.  DRD{1}                     0.01498708   0.04968223      0.30166  0.76328974
8.  DRD{2}                    -0.06072125   0.04889108     -1.24197  0.21599946
9.  DRD{3}                    -0.09420841   0.04673122     -2.01596  0.04541659
10. DTCHO                    -25.16058920   4.64329585     -5.41869  0.00000021
11. DTCHO{1}                   9.12829877   5.44900539      1.67522  0.09577322
12. DTCHO{2}                   7.71753251   5.39199148      1.43130  0.15422554
13. DTCHO{3}                  -0.64700738   4.38051768     -0.14770  0.88275789
14. DSP                        0.23739451   0.06150048      3.86004  0.00016210
15. DSP{1}                     0.12577658   0.07235373      1.73836  0.08400222
16. DSP{2}                     0.10519576   0.07369767      1.42740  0.15534420
17. DSP{3}                    -0.08264005   0.07479583     -1.10488  0.27081260
```

Le résidu du modèle de long terme RE{1} en t-1 est très significatif.