

# Une présentation simplifiée de l'approche de Johansen La cointégration dans un cadre multivarié

Gilbert Colletaz

3 avril 2017

## Table des matières

<b>1</b>	<b>Les corrélations canoniques</b>	<b>2</b>
<b>2</b>	<b>Le calcul des corrélations canoniques</b>	<b>2</b>
<b>3</b>	<b>Approche de Johansen et corrélations canoniques</b>	<b>4</b>
3.1	Les enseignements d'une ré-écriture de type Dickey-Fuller d'un processus autorégressif vectoriel . . . . .	4
3.2	corrélations canoniques et rang de l'espace de cointégration . .	7
3.3	La sélection de $\hat{r}$ : le test de la trace et le test $\lambda$ -max . . . . .	8
<b>4</b>	<b>L'identification des matrices <math>\alpha</math> et <math>\beta</math></b>	<b>9</b>
<b>5</b>	<b>La présence de termes déterministes</b>	<b>12</b>
5.1	un exemple illustrant l'attention qui doit être portée à la position des termes déterministes dans le VECM . . . . .	13
5.2	Les cinq cas types de prise en compte de variables déterministes	14
5.3	Test de rang dans VARMAX en présence d'une constante . . . .	17
<b>6</b>	<b>Les tests sur <math>\alpha</math> et <math>\beta</math></b>	<b>17</b>
6.1	tests sur $\beta$ . . . . .	18
6.2	tests sur $\alpha$ . . . . .	18
6.2.1	Trois définitions de l'exogénéité . . . . .	19
6.2.2	Exogénéité faible dans le VECM . . . . .	20
<b>7</b>	<b>La causalité selon Granger dans un système cointégré</b>	<b>22</b>
<b>8</b>	<b>Exemples</b>	<b>23</b>
8.1	Un premier exemple sur séries stationnaires . . . . .	23
8.2	Un second exemple sur séries I(1) . . . . .	24

## 1 Les corrélations canoniques

On a deux ensembles de variables  $X$  et  $Y$  respectivement en nombre  $n_X$  et  $n_Y$ . On les suppose stationnaires et centrées de sorte que :

$$\begin{pmatrix} E[X_t X_t^\top] & E[X_t Y_t^\top] \\ E[Y_t X_t^\top] & E[Y_t Y_t^\top] \end{pmatrix} = \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix} \quad (1)$$

Soit encore  $n = \min(n_X, n_Y)$  et deux matrices  $A(n, n_X)$  et  $B(n, n_Y)$ . On définit également deux vecteurs de taille  $n$  comme :

$$u_t = AX_t \quad (2)$$

$$v_t = BY_t \quad (3)$$

Ainsi, le  $i^{\text{ème}}$  élément de  $u_t$  (resp.  $v_t$ ) est une combinaison linéaire des observations de  $X$  (resp.  $Y$ ) du temps  $t$  avec des poids donnés dans la  $i^{\text{ème}}$  ligne de  $A$  (resp. de  $B$ ). On cherche alors à trouver des pondérations telles que les combinaisons linéaires au sein de chacun des 2 groupes soient orthogonales entre elles, de variance unitaire et telles que la corrélation entre les combinaisons linéaires de rang  $i$  construites avec  $X$  d'une part et  $Y$  d'autre part soit maximale pour  $i = 1, \dots, n$  et accessoirement, de sorte à ce que ces corrélations soient classées par ordre décroissant. Les corrélations en question sont les corrélations canoniques. On a donc une structure de la forme :

$$E[u_t u_t^\top] = A \Sigma_{XX} A^\top = I \quad (4)$$

$$E[v_t v_t^\top] = B \Sigma_{YY} B^\top = I \quad (5)$$

$$E[u_t v_t^\top] = R = \begin{pmatrix} \rho_1 & 0 & \dots & 0 \\ 0 & \rho_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \rho_n \end{pmatrix} \quad (6)$$

## 2 Le calcul des corrélations canoniques

On va rechercher une estimation de la première corrélation  $\rho_1$ . Simultanément, on trouvera également les coefficients des combinaisons linéaires associées, c'est à dire les deux premières lignes de  $A$  et  $B$  notée par la suite  $a^\top$  et  $b^\top$ . Soit  $u_{1,t} = a^\top X_t$  et  $v_{1,t} = b^\top Y_t$ , on a donc :

$$\rho_1 = \text{cor}(u_{1,t}, v_{1,t}) = \frac{a^\top \Sigma_{XY} b}{\sqrt{a^\top \Sigma_{XX} a} \sqrt{b^\top \Sigma_{YY} b}} = a^\top \Sigma_{XY} b \quad (7)$$

Il s'agit de trouver  $a$  et  $b$  de sorte à maximiser cette corrélation. Le lagrangien associé s'écrit donc :

$$\mathcal{L}(a, b, \kappa_1, \kappa_2) = a^\top \Sigma_{XY} b - \frac{1}{2} \kappa_1 (a^\top \Sigma_{XX} a - 1) - \frac{1}{2} \kappa_2 (b^\top \Sigma_{YY} b - 1) \quad (8)$$

Par suite :

$$\frac{\delta \mathcal{L}}{\delta a} = 0 = \Sigma_{XY}b - \kappa_1 \Sigma_{XX}a \quad (9)$$

$$\frac{\delta \mathcal{L}}{\delta b} = 0 = \Sigma_{YX}a - \kappa_2 \Sigma_{YY}b \quad (10)$$

$$(11)$$

d'où :

$$a^\top \Sigma_{XY}b = \kappa_1 a^\top \Sigma_{XX}a = \kappa_1 \quad (12)$$

$$b^\top \Sigma_{YX}a = \kappa_2 b^\top \Sigma_{YY}b = \kappa_2 \quad (13)$$

et donc :  $\kappa_1 = \kappa_2 = \text{cor}(u_{1,t}, v_{1,t})$ .

En prémultipliant la première condition d'ordre 1 du problème de maximisation ci-dessus par  $\Sigma_{YX}\Sigma_{XX}^{-1}$ , et en utilisant la deuxième condition ainsi que la dernière égalité, il vient :

$$\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}b = \kappa_1 \Sigma_{YX}a = \kappa_1^2 \Sigma_{YY}b \quad (14)$$

Avec des opérations symétriques, on montre également que :

$$\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}a = \kappa_1^2 \Sigma_{XX}a \quad (15)$$

Si on note  $\lambda_1 = \kappa_1^2$  et que l'on multiplie (14) par  $\Sigma_{YY}^{-1}$  et (15) par  $\Sigma_{XX}^{-1}$ , on obtient finalement

$$\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}b = \lambda_1 b, \text{ et} \quad (16)$$

$$\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}a = \lambda_1 a \quad (17)$$

On reconnaît alors une structure connue :  $\lambda_1$ , qui est donc le carré de la première corrélation canonique, est la valeur propre d'une certaine matrice et  $a$  (ou  $b$ ) est le vecteur propre associé. En pratique, pour trouver les corrélations canoniques, il suffit donc de résoudre un problème de recherche de valeurs propres et de prendre leurs racines carrées. Entre les deux systèmes précédents, on utilisera plutôt celui dont la dimension est la plus faible, sachant que l'on pourra ensuite passer sans difficulté de l'un à l'autre système de vecteurs propres en utilisant l'une des deux équations de transition suivantes (encore tirées de (14) et (15)) :

$$b = \frac{1}{\sqrt{\lambda}} \Sigma_{YY}^{-1} \Sigma_{YX} a, \text{ ou} \quad (18)$$

$$a = \frac{1}{\sqrt{\lambda}} \Sigma_{XX}^{-1} \Sigma_{XY} b \quad (19)$$

### 3 Approche de Johansen et corrélations canoniques

Dans un premier temps, on discute de la réécriture de type Dickey-Fuller d'un processus vectoriel autorégressif d'ordre  $p$ . Cela permet de mettre en évidence la matrice sur laquelle peut être révélé le rang de cointégration, c'est à dire le nombre de combinaisons linéaires  $I(0)$  orthogonales entre elles que l'on peut construire à partir de  $k$  variables supposées  $I(1)$ . Par la suite, on présente l'intérêt de considérer les corrélations canoniques pour identifier ce rang de cointégration, ce qui doit faciliter la compréhension, au moins intuitive de l'approche de Johansen. Enfin les deux tests proposés par cet auteur permettant la sélection du rang de cointégration seront présentés.

#### 3.1 Les enseignements d'une ré-écriture de type Dickey-Fuller d'un processus autorégressif vectoriel

On part d'une écriture autorégressive d'ordre  $p$  sur un vecteur constitué des réalisations au temps  $t$  de  $k$  variables<sup>1</sup> :

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \dots + \Phi_p y_{t-p} + u_t \quad (20)$$

dans laquelle  $u_t$  est un bruit blanc vectoriel.

L'écriture équivalente de type Dickey-Fuller est :

$$\Delta y_t = \Pi y_{t-1} + \Phi_1^* \Delta y_{t-1} + \dots + \Phi_{p-1}^* \Delta y_{t-p+1} + u_t \quad (21)$$

avec  $\Pi = \sum_{i=1}^p \Phi_i - I$  et  $\Phi_i^* = -\sum_{j=i+1}^p \Phi_j$ .

On rappelle qu'on ne considère ici que des variables qui, si elles sont intégrées, le sont au plus à l'ordre 1, et dans ce cas  $\Delta y_t = y_t - y_{t-1}$  est intégré d'ordre 0. Il ressort, de l'écriture précédente, que 3 cas doivent être considérés en fonction du rang de la matrice  $\Pi$  :

1. Rang( $\Pi$ ) =  $k$   
chacune des composantes du vecteur  $y$  est stationnaire en niveau puisqu'elles s'écrivent comme combinaison linéaire de variables  $I(0)$ . En effet, on a :

$$y_{t-1} = \Pi^{-1}(\Delta y_t - \Phi_1^* \Delta y_{t-1} + \dots - \Phi_{p-1}^* \Delta y_{t-p+1} - u_t) \quad (22)$$

Dans ce premier cas, un VAR( $p$ ) sur  $y_t$  est évidemment adapté pour réaliser une analyse des relations dynamiques dans le système étudié.

2. Rang( $\Pi$ ) = 0, soit encore  $\Pi = 0$   
L'équation (21) devient alors :

$$\Delta y_t = \Phi_1^* \Delta y_{t-1} + \dots + \Phi_{p-1}^* \Delta y_{t-p+1} + u_t \quad (23)$$

1. On omet pour l'instant la présence de termes déterministes. leur inclusion sera discutée par la suite.

Le système est dans ce second cas gouverné par un VAR d'ordre  $p-1$  écrit sur les différences premières des variables initiales.

3.  $0 < \text{rang}(\Pi) = r < k$

C'est le cas le plus intéressant. Lorsque toutes les constituantes de  $y$  sont  $I(1)$ , il signifie que ces variables sont cointégrées et qu'il existe  $r$  relations de cointégration. La présence de cointégration est en effet nécessaire pour assurer la cohérence de l'écriture (21) : à gauche du signe = on a une expliquée qui est  $I(0)$ . Il est donc nécessaire d'avoir un objet qui soit également  $I(0)$  à droite. Or cet objet est construit comme une somme de composantes  $I(0)$ ,  $\Phi_1^* \Delta y_{t-1} + \dots + \Phi_{p-1}^* \Delta y_{t-p+1} + u_t$ , et de combinaisons linéaires non nulles des niveaux des variables initiales,  $\Pi y_{t-1}$ . On ne peut donc retrouver un objet  $I(0)$  que si ces combinaisons linéaires sont elles-mêmes stationnaires, ce qui, par définition, signifie la présence de relations de cointégration au sein du système  $y_t$ .

Maintenant, si  $\Pi_{(k \times k)}$  est de rang  $r$ , alors elle peut toujours s'écrire comme

$$\Pi = \alpha \beta^\top \quad (24)$$

où  $\alpha_{(k \times r)}$  et  $\beta_{(k \times r)}$  sont 2 matrices de constantes<sup>2</sup> également de rang( $r$ ). Ainsi :

$$\Pi y_{t-1} = [\alpha \beta^\top] y_{t-1} = \alpha [\beta^\top y_{t-1}] \sim I(0) \quad (25)$$

Comme  $\alpha$  est une matrice de constante, l'intégration d'ordre zéro de  $\alpha [\beta^\top y_{t-1}]$  est impliquée par celle de  $\beta^\top y_{t-1}$  : la matrice  $\beta$  contiendrait donc en colonne les coefficients des relations de cointégration. Comme  $\beta$  possède  $r$  colonnes, cela signifie que nous sommes en présence de  $r$  relations de cointégration. On sait également que  $\beta^\top y_{t-1}$  s'interprète souvent comme l'écart en  $t-1$  à une position d'équilibre de long terme, les relations de cointégration étant précisément chargées de décrire les  $r$  relations d'équilibre de long terme existant entre les variables en niveau. Dans ces conditions, la matrice  $\alpha_{(k \times r)}$  contient les coefficients de ces  $r$  écarts pour chacune des  $k$  équations, et précise donc la vitesse avec laquelle chacune des variables s'ajuste à ces déséquilibres de long terme.

Un exemple permet d'illustrer les paragraphes précédents. Soit un système de 3 variables obéissant à un VAR(2),  $\mathbf{y}_t = (y_{1,t}, y_{2,t}, y_{3,t})^\top$  et supposons l'existence de 2 relations de cointégration. Les écritures suivantes sont alors impliquées par cette structure :

1. Le modèle VAR(2)

---

2. Comme nous allons le voir, ces matrices  $\alpha$  et  $\beta$  ne sont pas uniques ce qui posera des problèmes d'identification.

$$\mathbf{y}_t = \Phi_1 \mathbf{y}_{t-1} + \Phi_2 \mathbf{y}_{t-2} + \mathbf{u}_t, \text{ soit encore :}$$

$$\begin{aligned} y_{1,t} &= {}_1\Phi_{11}y_{1,t-1} + {}_1\Phi_{12}y_{2,t-1} + {}_1\Phi_{13}y_{3,t-1} \\ &\quad + {}_2\Phi_{11}y_{1,t-2} + {}_2\Phi_{12}y_{2,t-2} + {}_2\Phi_{13}y_{3,t-2} + u_{1,t} \\ y_{2,t} &= {}_1\Phi_{21}y_{1,t-1} + {}_1\Phi_{22}y_{2,t-1} + {}_1\Phi_{23}y_{3,t-1} \\ &\quad + {}_2\Phi_{21}y_{1,t-2} + {}_2\Phi_{22}y_{2,t-2} + {}_2\Phi_{23}y_{3,t-2} + u_{2,t} \\ y_{3,t} &= {}_1\Phi_{31}y_{1,t-1} + {}_1\Phi_{32}y_{2,t-1} + {}_1\Phi_{33}y_{3,t-1} \\ &\quad + {}_2\Phi_{31}y_{1,t-2} + {}_2\Phi_{32}y_{2,t-2} + {}_2\Phi_{33}y_{3,t-2} + u_{3,t}, \end{aligned} \quad \text{avec } \mathbf{u}_t = (u_{1,t}, u_{2,t}, u_{3,t})^\top$$

et

$$\Phi_1 = \begin{pmatrix} {}_1\Phi_{11} & {}_1\Phi_{12} & {}_1\Phi_{13} \\ {}_1\Phi_{21} & {}_1\Phi_{22} & {}_1\Phi_{23} \\ {}_1\Phi_{31} & {}_1\Phi_{32} & {}_1\Phi_{33} \end{pmatrix} \quad \text{et} \quad \Phi_2 = \begin{pmatrix} {}_2\Phi_{11} & {}_2\Phi_{12} & {}_2\Phi_{13} \\ {}_2\Phi_{21} & {}_2\Phi_{22} & {}_2\Phi_{23} \\ {}_2\Phi_{31} & {}_2\Phi_{32} & {}_2\Phi_{33} \end{pmatrix}$$

## 2. L'écriture de Dickey-Fuller associée

$$\Delta \mathbf{y}_t = \Pi \mathbf{y}_{t-1} + \Phi_1^* \Delta \mathbf{y}_{t-1} + \mathbf{u}_t, \text{ avec } \Phi_1^* = -\Phi_2, \text{ soit encore :}$$

$$\begin{aligned} \Delta y_{1,t} &= \pi_{11}y_{1,t-1} + \pi_{12}y_{2,t-1} + \pi_{13}y_{3,t-1} \\ &\quad - {}_2\Phi_{11}\Delta y_{1,t-1} - {}_2\Phi_{12}\Delta y_{2,t-1} - {}_2\Phi_{13}\Delta y_{3,t-1} + u_{1,t} \\ \Delta y_{2,t} &= \pi_{21}y_{1,t-1} + \pi_{22}y_{2,t-1} + \pi_{23}y_{3,t-1} \\ &\quad - {}_2\Phi_{21}\Delta y_{1,t-1} - {}_2\Phi_{22}\Delta y_{2,t-1} - {}_2\Phi_{23}\Delta y_{3,t-1} + u_{2,t} \\ \Delta y_{3,t} &= \pi_{31}y_{1,t-1} + \pi_{32}y_{2,t-1} + \pi_{33}y_{3,t-1} \\ &\quad - {}_2\Phi_{31}\Delta y_{1,t-1} - {}_2\Phi_{32}\Delta y_{2,t-1} - {}_2\Phi_{33}\Delta y_{3,t-1} + u_{3,t}, \end{aligned}$$

et,

$$\Pi = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{pmatrix}$$

## 3. L'écriture de $\Pi = \alpha\beta^\top$ avec

$$\alpha_{(3 \times 2)} = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{pmatrix} \quad \text{et} \quad \beta_{(3 \times 2)} = \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \beta_{31} & \beta_{32} \end{pmatrix}$$

## 4. Les deux termes d'erreurs $I(0)$ donnés par les deux relations de cointégration

$$\beta^\top \mathbf{y}_{t-1} = \begin{pmatrix} \beta_{11}^\top y_{1,t-1} \\ \beta_{21}^\top y_{2,t-1} \end{pmatrix} = \begin{pmatrix} \beta_{11}y_{1,t-1} + \beta_{21}y_{2,t-1} + \beta_{31}y_{3,t-1} \\ \beta_{12}y_{1,t-1} + \beta_{22}y_{2,t-1} + \beta_{32}y_{3,t-1} \end{pmatrix}$$

où  $\beta_{.i}, i = 1, 2$  est la  $i^{\text{me}}$  colonne de  $\beta$ .

## 5. Le modèle à correction d'erreur, ou VECM dont l'écriture est identique à celle de Dickey-Fuller avec seulement substitution de $\alpha\beta^\top$ à $\Pi$

$$\Delta \mathbf{y}_t = \alpha \beta^\top \mathbf{y}_{t-1} + \Phi_1^* \Delta \mathbf{y}_{t-1} + \mathbf{u}_t, \text{ soit encore :}$$

$$\begin{aligned} \Delta y_{1,t} &= \alpha_{11} \left[ \beta_{1,1}^\top y_{t-1} \right] + \alpha_{12} \left[ \beta_{2,1}^\top y_{t-1} \right] \\ &\quad - {}_2\Phi_{1,1} \Delta y_{1,t-1} - {}_2\Phi_{1,2} \Delta y_{2,t-1} - {}_2\Phi_{1,3} \Delta y_{3,t-1} + u_{1,t} \\ \Delta y_{2,t} &= \alpha_{21} \left[ \beta_{1,1}^\top y_{t-1} \right] + \alpha_{22} \left[ \beta_{2,1}^\top y_{t-1} \right] \\ &\quad - {}_2\Phi_{2,1} y_{1,t-1} - {}_2\Phi_{2,2} y_{2,t-1} - {}_2\Phi_{2,3} y_{3,t-1} + u_{2,t} \\ \Delta y_{3,t} &= \alpha_{31} \left[ \beta_{1,1}^\top y_{t-1} \right] + \alpha_{32} \left[ \beta_{2,1}^\top y_{t-1} \right] \\ &\quad - {}_2\Phi_{3,1} \Delta y_{1,t-1} - {}_2\Phi_{3,2} \Delta y_{2,t-1} - {}_2\Phi_{3,3} \Delta y_{3,t-1} + u_{3,t} \end{aligned}$$

### 3.2 corrélations canoniques et rang de l'espace de cointégration

L'attention se porte donc naturellement sur la matrice  $\Pi$ . Dans cette logique, les augmentations en  $\Delta y_{t-i}$  sont relatives aux fluctuations de court terme et apparaissent comme des termes de nuisances au regard de la dynamique de long terme qui cherche à traduire l'impact de  $y_{t-1}$  sur  $\Delta y_t$ , d'où l'idée de purger ces deux termes de ces nuisances en considérant les régressions suivantes :

$$\Delta y_t = A_1^* \Delta y_{t-1} + \dots + A_{p-1}^* \Delta y_{t-p+1} + R_{0t} \quad (26)$$

$$y_{t-1} = B_1^* \Delta y_{t-1} + \dots + B_{p-1}^* \Delta y_{t-p+1} + R_{1t} \quad (27)$$

En considérant les résidus de ces régressions, on obtient un modèle plus simple que l'écriture initiale (rappelez-vous le théorème de Frisch-Waugh) :

$$R_{0t} = \Pi R_{1t} + u_t \quad (28)$$

Dans ce système, les résidus  $R_{0t}$  sont afférents aux différences premières des variables  $y$  alors que les résidus  $R_{1t}$  sont afférents aux niveaux de ces variables. On rappelle aussi que la corrélation entre une  $I(1)$  et une  $I(0)$  est asymptotiquement nulle. En conséquence, on est donc face à deux ensembles de variables tels que, en l'absence de cointégration, les corrélations intergroupes sont nulles, ce qui implique que la corrélation existant entre une combinaison linéaire construite avec les variables du premier groupe et une combinaison linéaire des variables du deuxième groupe doit être nulle. En d'autres termes les corrélations canoniques sont nulles. En revanche, si on peut construire une combinaison linéaire avec  $R_{1t}$  corrélée avec une combinaison linéaire construite sur  $R_{0t}$  c'est que la première est  $I(0)$  et que cette combinaison linéaire est une relation de cointégration. Comme les corrélations canoniques sont les corrélations maximales que l'on peut construire au moyen de combinaisons linéaires de ce type, vous devez comprendre que le nombre de coefficients de corrélation canonique non nuls est également le nombre de relations de cointégration, c'est à dire le rang de  $\Pi$ . Il suffit donc de trouver le nombre  $r$  de valeurs propres non nulles, leurs vecteurs propres associés composant alors la matrice  $\beta$  dans l'écriture  $\Pi = \alpha \beta^\top$ .

Naturellement il y a une difficulté : même si  $r$  valeurs propres sont non nulles, et donc si  $\lambda_i = 0, i = r + 1, \dots, k$ , on sait que  $\text{Prob}[\hat{\lambda}_i = 0] = 0$ . En d'autres termes nous devons être en mesure de tester cette nullité.

### 3.3 La sélection de $\hat{r}$ : le test de la trace et le test $\lambda$ -max

On suppose que les valeurs propres estimées sont classées par ordre décroissant :  $\hat{\lambda}_1 > \hat{\lambda}_2 > \dots > \hat{\lambda}_k$ . Johansen montre que l'expression de la logvraisemblance sous l'hypothèse d'un rang de cointégration égal à  $r$  est :

$$\hat{\ell} = C - \frac{T}{2} \sum_{i=1}^r \log(1 - \hat{\lambda}_i) \quad (29)$$

A partir de là, on peut imaginer de mettre en place aisément des tests de type LRT. Par exemple, si l'hypothèse  $H_0$  est que le rang de  $\Pi$  est égal à  $r_0$  versus  $H_1$  :  $\Pi$  est de rang  $r_1 > r_0$ , on aura :

$$LRT = -2 \left( \hat{\ell}[\text{rang}(\Pi) = r_0] - \hat{\ell}[\text{rang}(\Pi) = r_1] \right) = -T \sum_{i=r_0+1}^{r_1} \log(1 - \hat{\lambda}_i) \quad (30)$$

Le test précédent est connu sous le nom de test de la trace (*trace test*). Attention sa distribution asymptotique n'est pas un chi-2 pour la raison habituelle : sous l'hypothèse nulle, il est construit avec des variables non stationnaires (dans le cas précédent, sous  $H_0$  il y a  $(r_1 - r_0)$  racines unitaires).

Un second test, connu sous l'appellation  $\lambda$ -max, est aussi proposé par Johansen. Il confronte  $H_0$  : "il y a  $r$  relations de cointégration" à  $H_1$  : "il y a  $r+1$  relations de cointégration", et il vient dans ce cas :

$$LRT = -2 \left( \hat{\ell}[\text{rang}(\Pi) = r] - \hat{\ell}[\text{rang}(\Pi) = r + 1] \right) = -T \log(1 - \hat{\lambda}_{r+1}) \quad (31)$$

Comme dans le cas précédent, la distribution asymptotique de la statistique LRT n'est pas la loi de  $\chi^2$ . Des valeurs critiques pour ces tests ont été tabulées par Osterwald et Lenum (1992). Comme avec le test de Dickey-Fuller, les distributions asymptotiques et les valeurs critiques qui en découlent dépendent de la présence ou non de termes déterministes. Par ailleurs, elles vont également dépendre de la façon dont ces termes déterministes sont pris en compte dans le processus VECM (Conf. Infra). Fort heureusement, la plupart des logiciels d'économétrie affichent les valeurs critiques adaptées au modèle estimé.

En pratique le choix de  $r$  est souvent effectué via une succession de tests  $\lambda$ -max : on commence par opposer  $r = 0$  à  $r = 1$ . Si  $H_0$  n'est pas rejetée on conclut à la non-cointégration ; sinon, on oppose  $r = 1$  à  $r = 2$ . En l'absence de rejet on retient une seule relation de cointégration, sinon on continue (opposition de  $r = 2$  à  $r = 3$ , etc. ...).



Dans VARMAX, la mise en oeuvre de ces tests de Johansen est réclamée par la présence de l'option COINTEST dans la commande MODEL. Ils ne peuvent être réalisés que pour des systèmes de dimension  $k \leq 11$ . La ligne d'appel est de la forme :

MODEL ... / ... COINTEST=(JOHANSEN=(TYPE=mot-clef SIGLEVEL=value));

L'option TYPE= précise la nature du test à réaliser : TYPE=MAX pour le test  $\lambda$ -max, TYPE=TRACE pour le test de la trace. L'option SIGLEVEL gouverne le seuil de risque et donc les valeurs critiques employées. Les valeurs possibles sont 1%, 5% et 10%. Par défaut, TYPE=TRACE et SIGLEVEL= 0.05<sup>3</sup>.

## 4 L'identification des matrices $\alpha$ et $\beta$

On repart de l'écriture  $\Pi = \alpha\beta^T$  ou  $\alpha$  et  $\beta$  sont des matrices  $(k, r)$  de rang  $r$  telles que  $\beta^T y_t$  est  $I(0)$ . On sait donc que les colonnes de  $\beta$  sont les coefficients des  $r$  relations de cointégration et que la  $j^{\text{ème}}$  colonne de  $\alpha$  donne les coefficients de la  $j^{\text{ème}}$  relation de cointégration dans chacune des  $k$  relations définissant les accroissements  $(\Delta y_{1t}, \Delta y_{2t}, \dots, \Delta y_{kt})^T$ . Soit :

$$\Delta y_t = \alpha\beta^T y_{t-1} + \Phi_1^* \Delta y_{t-1} + \dots + \Phi_{p-1}^* \Delta y_{t-p+1} + u_t \quad (32)$$

A l'évidence, pour toute matrice  $M$  régulière de dimension  $(r, r)$  on peut constituer une autre matrice de vecteurs cointégrants  $\beta M^T$  ayant  $\alpha M^{-1}$  comme matrice de coefficients associés. Cette nouvelle écriture est indiscernable de la première puisque  $\alpha\beta^T = \alpha M^{-1} M \beta^T = \alpha^* \beta^{*T} = \Pi$ . Cela signifie que  $\alpha$  et  $\beta$  ne sont pas complètement identifiées. En fait,  $\beta$  est seulement une base de l'espace de cointégration. Il est donc facile d'imposer des pseudo-restrictions, c'est à dire des restrictions qui ne changent pas la valeur la fonction de vraisemblance. Pour pouvoir afficher un résultat parmi les infinités de solutions possibles, Johansen utilise une contrainte arbitraire de normalisation des  $\beta$  qui consiste à contraindre à l'unité le coefficient d'une des variables dans la relation de cointégration estimée. Attention, cette normalisation n'a aucun sens économique. Par exemple, supposons que l'on ait obtenu l'estimation suivante d'une relation de cointégration dans un système de 3 variables

$$2y_{1t} - .8y_{2t} + 4y_{3t} = w_t \sim I(0) \quad (33)$$

dont les coefficients seraient  $\alpha^T = (-0.6, +.3, -1)^T$ . On peut désirer forcer le coefficient de  $y_{1t}$  à être égal à l'unité en ayant en tête une interprétation plus structurelle de la forme  $y_{1t} = c_1 y_{2t} + c_2 y_{3t} + w_t^*$ , signifiant que  $y_{1t}$  est une variable qui s'ajuste à une cible de long terme, définie par  $y_{2t}$  et  $y_{3t}$  selon  $c_1 y_{2t} + c_2 y_{3t}$ , avec

3. Pour information, il existe une autre option associée à JOHANSEN : il s'agit de IORDER=value. Deux valeurs sont possibles : 1, qui signifie que l'on étudie la cointégration dans un système de variables  $I(1)$ , et 2, qui permet l'étude de la cointégration dans un système de variables  $I(2)$ . Ce dernier cas n'est pas abordé dans ce cours. Par défaut, IORDER=1.

un écart à cette position stationnaire égal à  $w_t^*$ . Clairement la prise en compte de cette contrainte dans la sortie précédente revient à diviser par 2 les coefficients  $\beta$  et à doubler les coefficients  $\alpha$  correspondants : cela donne donc comme "nouvelle" relation de cointégration<sup>4</sup>  $y_{1t} - .4y_{2t} + 2y_{3t}$  avec les poids associés  $(-1.2, +.6, -2)^\top$  et  $w^*t = w_t/2$ . Naturellement la valeur de la vraisemblance ne change pas : il s'agit bien d'une pseudo-contrainte. On peut tout aussi bien de normaliser à l'unité le coefficient de  $y_{2t}$  ou celui de  $y_{3t}$  : toutes ces écritures sont équivalentes<sup>5</sup>. Une leçon importante est qu'on ne doit pas faire d'interprétation structurelle avec de telles pseudo-contraintes. Il faut notamment prendre garde à ne pas raisonner en termes de variable expliquée et de variables explicatives en se référant à une écriture du type régression de cointégration. De ce point de vue, la procédure de Johansen est moins dangereuse que la procédure en deux étapes d'Engle-Granger qui, dans sa première étape oblige son utilisateur à faire un pari sur la nature de la régression de cointégration, à savoir expliquer  $y_{1t}$  par  $y_{2t}$  et  $y_{3t}$  ou expliquer  $y_{2t}$  par  $y_{1t}$  et  $y_{3t}$ , ou  $y_{3t}$  par  $y_{1t}$  et  $y_{2t}$  avec comme on le sait des possibles contradictions dans les conclusions des tests d'Engle-Granger réalisés à la seconde étape notamment lorsque le  $R^2$  de la première étape s'éloigne de l'unité.

En revanche il est naturellement possible d'imposer des restrictions significatives. Un exemple type est de forcer à zéro un des coefficients constituant  $\alpha$  : cela signifie que les déviations par rapport à l'équilibre de long terme que décrit la relation de cointégration correspondante n'expliquent pas les variations de la variable afférente à l'équation concernée. De même, lorsque l'on est en présence d'une seule régression de cointégration, on peut vouloir mettre à zéro le coefficient d'une variable dans le vecteur cointégrant : cela revient à imposer l'absence de cette variable dans la relation de long terme et ce n'est pas une pseudo-contrainte. Ce type de restrictions affecte alors la valeur de la vraisemblance et elles peuvent être étudiées au moyen de tests LRT. Dans ces cas, conditionnellement à la valeur de  $r$ , on retrouve des statistiques distribuées asymptotiquement comme des  $\chi^2$  et dont le nombre de degrés de liberté est le nombre de contraintes significatives imposées.

Les deux exemples précédents ne doivent pas masquer le fait que l'identification des relations d'équilibre de long terme est l'une des difficultés majeures de l'approche de Johansen. Pour illustrer ce point, on reprend l'exemple de mise à zéro du coefficient d'une variable dans une relation de cointégration en

4. attention à la petite gymnastique : si  $y_{1t} = c_1 y_{2t} + c_2 y_{3t} + w_t$  est la régression de cointégration, alors la relation de cointégration est  $y_{1t} - c_1 y_{2t} - c_2 y_{3t}$  et donc le vecteur cointégrant est  $(1, -c_1, -c_2)$ .

5. On peut imposer cette normalisation soit dans l'option COINTEST vue précédemment, soit dans l'option ECM qui réclame l'estimation du modèle à correction d'erreur. Par exemple,

```
MODEL y1 y2 y3 / ... COINTEG=(JOHANSEN=(normalize=y2));
```

va réaliser le test sur le rang de  $\Pi$  en forçant le coefficient de la variable  $y_2$  à être égal à l'unité dans le où les vecteurs de cointégration rangés dans  $\beta$ . Si vous avez compris les précédents développements, vous savez alors que le résultat du test de cointégration ne dépendra pas de la variable choisie dans cette normalisation.

supposant maintenant que l'on a trois variables dans le système étudié et deux vecteurs cointégrants. Soit donc les deux relations en question :

$$\begin{cases} \beta_{11}y_{1t} + \beta_{12}y_{2t} + \beta_{13}y_{3t} \\ \beta_{21}y_{1t} + \beta_{22}y_{2t} + \beta_{23}y_{3t} \end{cases} \quad (34)$$

Dans ce système, pour une raison quelconque, on pense que l'une des relations est définie uniquement à partir des deux variables  $y_{1t}$  et  $y_{3t}$ . Si on choisit la première relation pour représenter cette configuration, cela revient à imposer  $\beta_{12} = 0$ . Maintenant, si dans le système précédent on multiplie la première équation par  $-\frac{\beta_{22}}{\beta_{12}}$ , on obtient évidemment toujours une relation I(0) qui, si on l'ajoute à la seconde, nous donne une structure équivalente de deux relations stationnaires :

$$\begin{cases} -\frac{\beta_{22}\beta_{11}}{\beta_{12}}y_{1t} - \frac{\beta_{22}\beta_{13}}{\beta_{12}}y_{3t} \\ \beta_{21}y_{1t} + \beta_{22}y_{2t} + \beta_{23}y_{3t} \end{cases} \Leftrightarrow \begin{cases} \beta_{12}^*y_{1t} + \beta_{13}^*y_{3t} \\ \beta_{21}y_{1t} + \beta_{22}y_{2t} + \beta_{23}y_{3t} \end{cases} \quad (35)$$

En d'autres termes on peut sans imposer de contrainte particulière, passer de (34) à (35) et donc le fait d'imposer ici  $\beta_{12} = 0$  n'est absolument pas une contrainte testable.

Ce problème d'identification des matrices  $\alpha$  et  $\beta$ , qui contiennent chacune  $k \times r$  termes, est résolu dans un travail de Bauwens et Lubrano <sup>6</sup> montrant que sous la restriction  $\pi = \alpha\beta^\top$ , seuls  $2kr - r^2$  termes sont identifiés. Il faut donc introduire  $r^2$  restrictions supplémentaires pour avoir l'identification complète. Avec une seule relation de cointégration, il suffit par exemple de contraindre à l'unité le coefficient d'une des variables <sup>7</sup>. Le problème se complique lorsque le rang de cointégration est supérieur à 1 puisqu'il faut alors s'assurer que les restrictions supplémentaires sont indépendantes entre elles. Soit par exemple  $R_i\beta_{.i} = 0$  ou  $R_i$  est une matrice  $c \times k$  décrivant un ensemble de  $c$  contraintes portant sur la  $i^{me}$  colonne de  $\beta$ , alors l'indépendance des contraintes entre elles exige que  $\text{rang}(R\beta_c) = r - 1$ , où  $\beta_c$  est constitué des vecteurs cointégrants respectant les  $c$  contraintes. Cette condition de rang assure en effet qu'aucune restriction ne définit un vecteur de cointégration pouvant s'écrire comme combinaison linéaire des  $r - 1$  autres vecteurs.

Afin d'illustrer ce point, nous reprenons un exemple construit par Michel Lubrano. soit  $k=4$  et  $r=2$  :

$$\beta = \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \beta_{31} & \beta_{31} \\ \beta_{41} & \beta_{41} \end{pmatrix} \quad (36)$$

6. Bauwens, L., AND M. Lubrano (1996) : "Identifications Restrictions and Posterior Densities in Cointegrated Gaussian VAR Systems" in *Advances in Econometrics*, ed.by T. Fomby, Vol 11, part B, pp. 3-28. JAI Press, Greenwich, USA.

7. en évitant toutefois de faire porter cette restriction sur une variable dont le vrai coefficient serait nul

et soit les restrictions suivantes sur le premier vecteur cointégrant  $\beta_{21} = -\beta_{11}$ ,  $\beta_{41} = -\beta_{31}$ , et, sur le second :  $\beta_{12} = 0, \beta_{22} = 0, \beta_{42} = -\beta_{32}$ . Les matrices  $R_1$  et  $R_2$  correspondantes sont donc :

$$R_1 = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \quad (37)$$

$$R_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \quad (38)$$

avec les vecteurs cointégrants contraints suivants :

$$\beta_c = \begin{pmatrix} \beta_{11} & 0 \\ -\beta_{11} & 0 \\ \beta_{31} & \beta_{32} \\ -\beta_{31} & -\beta_{32} \end{pmatrix} \quad (39)$$

Dans ces conditions,

$$R_1\beta_c = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad (40)$$

et,

$$R_2\beta_c = \begin{pmatrix} \beta_{11} & 0 \\ -\beta_{11} & 0 \\ 0 & 0 \end{pmatrix} \quad (41)$$

Comme  $\text{rang}(R_1\beta_c) = 0 \neq r - 1 = 1$ , le vecteur cointégrant  $\beta_{.1}$  n'est pas identifié. En revanche, si  $\beta_{11} \neq 0$ , alors le second est identifié par les restrictions décrites par  $R_2$  puisque  $\text{rang}(R_2\beta_c) = 1$ <sup>8</sup>.

Un des enseignements pratique de cette section est ainsi de montrer que dès lors que l'on a plus d'une relation de cointégration il devient très difficile de les identifier : plus que l'économétrie, la théorie économique est alors essentielle à condition de fournir un cadre théorique de référence permettant de préciser les restrictions à tester.

## 5 La présence de termes déterministes

Jusqu'à présent nous avons omis la présence de composantes déterministes. Celle-ci est cependant souvent nécessaire, via par exemple l'inclusion d'une constante, ou d'un trend déterministe ou bien encore de variables indicatrices représentatives de mouvements saisonniers.

Depuis les tests DF et ADF de Dickey-Fuller, vous savez que la prise en compte ou non de termes déterministes ainsi que la nature de ceux-ci, constante ou trend linéaire, affecte la distribution asymptotique des statistiques de test. Ce

---

8. Et que par ailleurs le nombre de restrictions imposées, égal à 5, est supérieur à  $r^2 = 4$ .

résultat obtenu dans un cadre univarié continue malheureusement d'être valide dans le cadre multivarié.

En outre, le cadre multivarié amène des difficultés supplémentaires qui lui sont propres et qui vont un peu plus perturber ces distributions asymptotiques.

### 5.1 un exemple illustrant l'attention qui doit être portée à la position des termes déterministes dans le VECM

Pour illustrer les problèmes liés à l'intégration de termes déterministes dans l'écriture VECM, on considère simplement deux variables  $x_t$  et  $y_t$  intégrées d'ordre 1 et cointégrées. Si nous reprenons le théorème de représentation et que l'on insère une constante de la manière usuelle à savoir l'ajout d'une colonne de 1 dans la liste des explicatives, les équations auront une structure de la forme :

$$\Delta x_t = \alpha_1(\beta_1 x_{t-1} + \beta_2 y_{t-1}) + \sum_{i=1}^{p-1} a_{1i} \Delta x_{t-i} + \sum_{i=1}^{p-1} b_{1i} \Delta y_{t-i} + c_1 + u_{1t} \quad (42)$$

$$\Delta y_t = \alpha_2(\beta_1 x_{t-1} + \beta_2 y_{t-1}) + \sum_{i=1}^{p-1} a_{2i} \Delta x_{t-i} + \sum_{i=1}^{p-1} b_{2i} \Delta y_{t-i} + c_2 + u_{2t} \quad (43)$$

Dans cette structure, comme les expliquées sont des différences premières, on sait que la présence de constantes non nulles dans les explicatives revient à imposer l'existence d'un trend déterministe linéaire de pentes respectives  $c_1$  et  $c_2$  dans les niveaux de  $x$  et  $y$ .

On peut aisément imaginer des situations où l'incorporation de tels trends linéaires déterministes dans les niveaux de certaines variables est totalement déraisonnable. Un exemple type est celui de deux taux d'intérêt, l'un à court terme,  $S_t$ , et l'autre à long terme,  $L_t$ . Si on admet que le second est égal au premier augmenté d'une prime éventuellement variable mais d'espérance constante et positive pour couvrir le risque de transformation,  $p_t = p + u_t, p > 0$ , il vient  $L_t = S_t + p + u_t$ . Il existe bien une constante dans le système, à savoir  $p$ , mais celle-ci est intégrée dans l'équation de cointégration et ne provoque pas de dérive du niveau des taux d'intérêt.

Plus généralement, l'incorporation d'une constante dans les équations de cointégration peut être souvent légitime : c'est notamment le cas où précisément on désire interpréter l'équation de cointégration comme définissant une erreur et les coefficients  $\alpha$  comme mesurant la vitesse d'ajustement des variables à cette erreur. Dans ces conditions, si le terme  $\beta y_{t-1}$  définit une variable  $w_{t-1}$  stationnaire d'espérance  $c$  éventuellement non nulle, alors afin d'assurer la nullité en moyenne de l'erreur, il suffit de poser  $v_t = w_t - c$ . Il vient naturellement  $\beta y_{t-1} + c = v_t$ , et  $v_t$  est bien un terme d'erreur par rapport à une position d'équilibre de long terme qui est nul en moyenne. Cela revient dans l'exemple précédent à faire en sorte que l'écart du taux long à sa position d'équilibre de long terme est mesuré non pas simplement par  $L_t - S_t$  mais par  $L_t - S_t - p$ . Dans

ce deuxième cas de prise en compte d'une constante, les équations du VECM auront pour structure (toujours pour  $k=2$  et  $r=1$ ) :

$$\Delta x_t = \alpha_1(\beta_1 x_{t-1} + \beta_2 y_{t-1} + c) + \sum_{i=1}^{p-1} a_{1i} \Delta x_{t-i} + \sum_{i=1}^{p-1} b_{1i} \Delta y_{t-i} + u_{1t} \quad (44)$$

$$\Delta y_t = \alpha_2(\beta_1 x_{t-1} + \beta_2 y_{t-1} + c) + \sum_{i=1}^{p-1} a_{2i} \Delta x_{t-i} + \sum_{i=1}^{p-1} b_{2i} \Delta y_{t-i} + u_{2t} \quad (45)$$

La prise en compte de trends déterministes explicites dans les équations conduit avec un raisonnement similaire à celui qui vient d'être mené à distinguer deux cas : soit le trend est à l'extérieur de l'espace de cointégration et alors on impose des trends quadratiques dans le niveaux des variables modélisées. Soit le trend est à l'intérieur de cet espace, ce qui revient à poser que le terme  $\beta y_t$  définit une variable stationnaire autour d'un trend linéaire.

## 5.2 Les cinq cas types de prise en compte de variables déterministes

Au total on peut définir cinq configurations types qui peuvent maintenant être toutes estimées dans la plupart des logiciels d'économétrie avec disponibilité des valeurs critiques du test de trace ou du test  $\lambda$ -max afférentes<sup>9</sup>. Pour les illustrer, on considère un système bivarié ayant une écriture autorégressive d'ordre 2. Dans ces conditions, d'une part s'il existe une relation de cointégration elle est unique, d'autre part il n'y a qu'une augmentation dans l'écriture équivalente de type Dickey-Fuller. Les enseignements tirés de cette structure simple peuvent être aisément généralisés à des structures plus complexes mettant en jeu plus de deux variables et/ou dont l'ordre d'autorégression est plus élevé.

Pour cette présentation, nous reprenons la typologie cas 1-cas 5 que vous trouvez dans l'aide de la proc Varmax. Elle suppose un processus autorégressif d'ordre  $p = 2$  sur les niveaux des variables  $y = (y_1, \dots, y_k)^T$ . La logique qui prévaut dans les instructions SAS que nous allons rencontrer est simple : la structure du modèle autorégressif et la nature des variables déterministes sont précisées dans la commande `Model` . . . alors que le rang de cointégration dans le VECM associé et l'emplacement de ces variables déterministes sont gérés par la commande `Cointeg`<sup>10</sup> et la présence ou l'absence de son option `ectrend`.

9. En cas de besoin, vous trouverez un ensemble de tables statistiques utiles contenant notamment la table de McKinnon pour les tests de Dickey-Fuller ou de cointégration en deux étapes à la Engle-Granger, ainsi que la table d'Osterwald-Lenum pour les tests de trace et  $\lambda$ -max ici : <http://www2.warwick.ac.uk/fac/soc/economics/staff/academic/jeremysmith/manual/statisticaltables.pdf>.

10. L'option `ECM=` . . . de la commande `Model` est maintenant considérée comme obsolète.

- cas 1 : absence de termes déterministes. C'est le cas que nous avons considéré jusqu'ici :

$$\Delta y_t = \alpha \beta^\top y_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \Delta y_{t-i} + u_t \quad (46)$$

Par exemple, dans le cas d'un système bivarié  $(y_{1t}, y_{2t})$  avec un rang de cointégration  $r = 1$ <sup>11</sup> et  $p = 5$ , et le coefficient de  $y_2$  forcé à l'unité dans le vecteur cointégrant, soit :

$$\begin{aligned} \Delta y_{1t} &= \alpha_1(\beta_1 y_{1,t-1} + y_{2,t-1}) + \sum_{i=1}^4 a_{1i} \Delta y_{1,t-i} + \sum_{i=1}^4 b_{1i} \Delta y_{2,t-i} + u_{1t} \\ \Delta y_{2t} &= \alpha_2(\beta_1 y_{1,t-1} + y_{2,t-1}) + \sum_{i=1}^4 a_{2i} \Delta y_{1,t-i} + \sum_{i=1}^4 b_{2i} \Delta y_{2,t-i} + u_{2t} \end{aligned}$$

Son estimation sous SAS est réclamée par les commandes :

```
MODEL y1 y2 / P=5 NOINT;
COINTEG RANK=1 NORMALIZE=y2;
```

- cas 2 : no trend, constante dans l'espace de cointégration :

$$\Delta y_t = \alpha(\beta^\top, \beta_0)(y_{t-1}, 1) + \sum_{i=1}^{p-1} \Phi_i^* \Delta y_{t-i} + u_t \quad (47)$$

Toujours dans le cadre bivarié avec  $r = 1, p = 5$ , cela donnerait :

$$\begin{aligned} \Delta y_{1t} &= \alpha_1(\beta_1 y_{1,t-1} + y_{2,t-1} + \beta_0) + \sum_{i=1}^4 a_{1i} \Delta y_{1,t-i} + \sum_{i=1}^4 b_{1i} \Delta y_{2,t-i} + u_{1t} \\ \Delta y_{2t} &= \alpha_2(\beta_1 y_{1,t-1} + y_{2,t-1} + \beta_0) + \sum_{i=1}^4 a_{2i} \Delta y_{1,t-i} + \sum_{i=1}^4 b_{2i} \Delta y_{2,t-i} + u_{2t} \end{aligned}$$

Cette structure est estimée au moyen des commandes suivantes :

```
MODEL y1 y2 / P=5;
COINTEG RANK=1 NORMALIZE=y2 ECTREND;
```

- cas 3 : pas de trend dans le VECM, constante hors de l'espace de cointégration :

$$\Delta y_t = \alpha \beta^\top y_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \Delta y_{t-i} + c + u_t \quad (48)$$

---

11. Rappelez-vous que dans un cadre bivarié, si il y a cointégration, on a forcément  $r = 1$ .

soit par exemple :

$$\Delta y_{1t} = \alpha_1(\beta_1 y_{1,t-1} + y_{2,t-1}) + \sum_{i=1}^4 a_{1i} \Delta y_{1,t-i} + \sum_{i=1}^4 b_{1i} \Delta y_{2,t-i} + c_1 + u_{1t}$$

$$\Delta y_{2t} = \alpha_2(\beta_1 y_{1,t-1} + y_{2,t-1}) + \sum_{i=1}^4 a_{2i} \Delta y_{1,t-i} + \sum_{i=1}^4 b_{2i} \Delta y_{2,t-i} + c_2 + u_{2t}$$

Son estimation est obtenue avec les commandes :

```
MODEL y1 y2 / P=5;
COINTEG RANK=1 NORMALIZE=y2;
```

— cas 4 : trend linéaire dans l'espace de cointégration et trend linéaire dans le niveau des variables :

$$\Delta y_t = \alpha(\beta^\top, \beta_0)(y_{t-1}, t) + \sum_{i=1}^{p-1} \Phi_i^* \Delta y_{t-i} + c + u_t \quad (49)$$

ainsi :

$$\Delta y_{1t} = \alpha_1(\beta_1 y_{1,t-1} + y_{2,t-1} + \beta_0 t) + \sum_{i=1}^4 a_{1i} \Delta y_{1,t-i} + \sum_{i=1}^4 b_{1i} \Delta y_{2,t-i} + c_1 + u_{1t}$$

$$\Delta y_{2t} = \alpha_2(\beta_1 y_{1,t-1} + y_{2,t-1} + \beta_0 t) + \sum_{i=1}^4 a_{2i} \Delta y_{1,t-i} + \sum_{i=1}^4 b_{2i} \Delta y_{2,t-i} + c_2 + u_{2t}$$

On l'ajuste avec SAS au moyen des commandes :

```
MODEL y1 y2 / P=5 TREND=LINEAR;
COINTEG RANK=1 NORMALIZE=y2 ECTREND;
```

— cas 5 : trend linéaire hors de l'espace de cointégration :

$$\Delta y_t = \alpha \beta^\top y_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \Delta y_{t-i} + c + dt + u_t \quad (50)$$

et comme exemple :

$$\Delta y_{1t} = \alpha_1(\beta_1 y_{1,t-1} + y_{2,t-1}) + \sum_{i=1}^4 a_{1i} \Delta y_{1,t-i} + \sum_{i=1}^4 b_{1i} \Delta y_{2,t-i} + c_1 + d_1 t + u_{1t}$$

$$\Delta y_{2t} = \alpha_2(\beta_1 y_{1,t-1} + y_{2,t-1}) + \sum_{i=1}^4 a_{2i} \Delta y_{1,t-i} + \sum_{i=1}^4 b_{2i} \Delta y_{2,t-i} + c_2 + d_2 t + u_{2t}$$



L'ajustement de ce cinquième cas est réalisé au moyen de :

```
MODEL y1 y2 / P=5 TREND=LINEAR;  
COINTEG RANK=1 NORMALIZE=y2;
```

On rappelle que dans ce dernier cas un trend quadratique est imposé dans le niveau des variables. Cela peut donner un ajustement satisfaisant in-sample mais conduire à des prévisions complètement délirantes hors période d'estimation : en général il est déconseillé d'ajuster cette structure, sauf à avoir de bonnes justifications.

### 5.3 Test de rang dans VARMAX en présence d'une constante

Ce point ne sert qu'à aider à la compréhension des résultats affichés par la PROC VARMAX lorsqu'on demande un test de rang via l'option COINTEST et qu'une constante est autorisée dans le VECM, i.e. lorsque l'option NOINT n'est pas présente dans la commande MODEL ...

Dans ce cas, VARMAX effectue automatiquement le test de trace ou le test  $\lambda$ -max dans deux configurations : avec la constante hors de l'espace de cointégration d'une part, et avec la constante dans l'espace de cointégration d'autre part. Dans l'intitulé des tableaux en sortie, vous verrez apparaître la précision "Under Restriction" pour signaler que les résultats sont afférents au deuxième cas de figure.

Un autre moyen de repérer dans quelle configuration on se situe, est de lire le contenu des deux dernières colonne des tableaux de présentation des tests de rang. Celles-ci sont intitulées respectivement "Drift in ECM" et "Drift in Process" : la première est relative à l'écriture VECM, la seconde à celle du processus autorégressif initial. Ainsi, dans la colonne "Drift in ECM", l'indication "Constant" signifie qu'il y a présence d'une constante dans le VECM. Lorsque, dans la colonne "Drift in Process" on lit "Linear", cela signifie que cette constante dans le VECM entraîne l'introduction d'un trend linéaire dans les niveaux des variables. Si vous avez compris les précédents développements, on est alors dans le cas où la constante est hors de l'espace de cointégration : c'est le cas 3 des 5 cas types. En revanche si vous lisez "Constant" dans "Drift in Process", c'est que la constante présente dans le VECM n'introduit pas de trend linéaire dans les niveaux : on est dans le cas 2 des 5 cas types, la constante est dans l'espace de cointégration.

## 6 Les tests sur $\alpha$ et $\beta$

Il est possible de construire des tests LRT sur les coefficients des relations de cointégration  $\beta$ , ou sur les coefficients (*loading factors*)  $\alpha$  de ces relations

retardées. Avec SAS, c'est la commande COINTEG qui autorise la mise en oeuvre de ces tests via la spécification de deux matrices de transition notée respectivement  $H$  pour ce qui concerne les restrictions sur la matrice  $\beta$  et  $J$  pour celles afférentes à  $\alpha$ .

## 6.1 tests sur $\beta$

L'écriture de l'hypothèse nulle à tester est de la forme  $H_0 : \beta = H\Phi$ , où  $\Phi$  est une matrice  $(s, r)$ ,  $s$  étant le nombre de paramètres libres à estimer sous  $H_0$ , et  $H$  une matrice  $(k, s)$  de passage entre les paramètres libres et les coefficients  $\beta$  contraints. Il s'agit de la formulation duale de l'écriture matricielle de contraintes linéaires que vous utilisez plus habituellement :  $R\beta = 0$ . Un exemple simple est le suivant : on considère une relation de cointégration avec 2 variables :  $\beta_1 x_t + \beta_2 y_t$  et on soupçonne que la différence des deux variables est stationnaire, soit  $H_0 : \beta_1 = -\beta_2$ . Sous cette hypothèse il n'y a qu'un paramètre à estimer,  $\Phi$  est de dimension  $(1 \times 1)$  et on a donc  $H = (1, -1)^\top$ . En présence d'une constante dans l'équation de cointégration on aurait deux paramètres libres à estimer, à savoir un seul  $\beta$  (puisque  $\beta_1 = \beta$  et  $\beta_2 = -\beta$ ) et la constante :  $\Phi$  serait donc de dimensions  $(2, 1)$  et  $H$  serait donné par

$$H = \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \end{pmatrix}$$

Soit encore un VECM composé de trois variables avec une équation de cointégration d'écriture  $\beta_1 x_t + \beta_2 y_t + \beta_3 z_t + \beta_4$  (donc avec constante dans l'espace de cointégration) que l'on désire réduire à  $x_t - z_t$ . Le test de cette proposition revient à spécifier une matrice  $H = (1, 0, -1, 0)^\top$ .

*Attention* : avec SAS, une fois donnée la structure  $\beta = H\Phi$ , les restrictions qu'elle décrit s'imposent à chacune des  $r$  relations de cointégration présentes. En conséquence la statistique de test suit asymptotiquement une loi de  $\chi^2$  à  $r(k - s)$  degrés de liberté.

## 6.2 tests sur $\alpha$

L'écriture de l'hypothèse nulle à tester est de la forme  $H_0 : \alpha = J\Phi$ , où  $\Phi$  est une matrice  $(s, r)$ ,  $s$  étant le nombre de paramètres libres à estimer sous  $H_0$ , et  $J$  une matrice  $(k, s)$  de passage entre les paramètres libres et les coefficients  $\alpha$  contraints. Par exemple, dans un VECM à 3 variables et une relation de cointégration, on a  $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ . L'hypothèse de nullité jointe de  $\alpha_2$  et  $\alpha_3$  est testée en posant  $J = (1, 0, 0)^\top$ . L'hypothèse de nullité de  $\alpha_2$  est réalisé en spécifiant :

$$J = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}$$

Le test de nullité de  $\alpha$  est encore qualifié de test d'exogénéité faible. Avant de continuer sur ce thème, il est utile de rappeler quelques définitions.

### 6.2.1 Trois définitions de l'exogénéité

On considère un couple de variables  $\tilde{x}$  et  $\tilde{y}$  ayant les réalisations  $x_t$  et  $y_t$  au temps  $t$  et dont la fonction de densité jointe est  $f_{\tilde{x},\tilde{y}}(x_t, y_t|\lambda, \tilde{X}_{t-1}, \tilde{Y}_{t-1})$ . Dans cette expression,  $\lambda \in \Lambda$  est un vecteur de paramètres,  $\tilde{X}_{t-1} = \{x_{t-1}, x_{t-2}, \dots, x_0\}$  représente le passé de  $x$  et  $\tilde{Y}_{t-1} = \{y_{t-1}, y_{t-2}, \dots, y_0\}$  le passé de  $y$  jusqu'au temps  $t-1$ . On sait que l'on peut toujours écrire la loi jointe comme le produit d'une conditionnelle et d'une marginale :

$$f_{\tilde{x},\tilde{y}}(x_t, y_t|\lambda, \tilde{X}_{t-1}, \tilde{Y}_{t-1}) = f_{\tilde{y}|\tilde{x}}(y_t|\lambda, \tilde{X}_{t-1}, \tilde{Y}_{t-1}, x_t) \times f_{\tilde{x}}(x_t|\lambda, \tilde{X}_{t-1}, \tilde{Y}_{t-1}) \quad (51)$$

Supposons que l'on s'intéresse à l'estimation des paramètres  $\lambda$ . En règle générale il faudra considérer la loi jointe et, par exemple, une maximisation de la vraisemblance pour trouver  $\hat{\lambda}$ . Lorsque le nombre de variables est relativement important il devient vite fastidieux, voire impossible, d'écrire cette loi jointe qui oblige à modéliser la totalité des variables du système. On préfère donc souvent se limiter à l'étude de la loi conditionnelle ce qui peut simplifier considérablement les choses. Ainsi par exemple, si on se contente de poser  $E[y_t|x_t] = x_t\lambda$  et  $Var[y_t|x_t] = \sigma^2$  alors, comme vous devez le savoir, on peut aisément construire des estimateurs pour  $\lambda$ . Cependant dans l'écriture générale (51) une difficulté peut apparaître avec cette façon de procéder : les paramètres  $\lambda$  sont a-priori présents à la fois dans la loi conditionnelle et dans la loi marginale de  $\tilde{x}$ . En conséquence, considérer uniquement la partie conditionnelle revient à se priver d'une information utile pour l'estimation de  $\lambda$  ce qui provoque une perte d'efficacité. En clair il faudrait considérer simultanément la conditionnelle et la marginale (les marginales si on avait plusieurs variables explicatives dans la conditionnelle).

On peut donc se demander à quelles conditions le fait de considérer uniquement la loi conditionnelle ne provoquerait pas de perte d'efficacité. Le raisonnement précédent indique la solution : il faut que les paramètres qui sont dans  $f_{\tilde{y}|\tilde{x}(\cdot)}$  ne soient pas dans  $f_{\tilde{x}(\cdot)}$ . En d'autres termes, le vecteur  $\lambda$  peut être coupé en deux et faire apparaître deux sous-ensembles de paramètres libres  $\lambda_1 \in \Lambda_1$  dans la loi conditionnelle et  $\lambda_2 \in \Lambda_2$  dans la loi marginale avec donc  $\Lambda = \Lambda_1 \times \Lambda_2$ . Il faut enfin naturellement que les paramètres d'intérêt soient des éléments de  $\lambda_1$  ou des transformées  $\psi_1 = g(\lambda_1)$  de ces seuls paramètres. C'est précisément ce que précise la notion d'exogénéité faible :

1. Exogénéité faible (*Weak exogeneity*) : dans (51)  $x_t$  est faiblement exogène pour un ensemble de paramètres  $\psi_1$  s'il existe une partition  $(\lambda_1, \lambda_2)$  de  $\lambda$  telle que (i)  $\psi_1$  est fonction du seul  $\lambda_1$  et (ii)  $\lambda_1 \in \Lambda_1, \lambda_2 \in \Lambda_2$  et  $\lambda = (\lambda_1^\top, \lambda_2^\top)^\top \in \Lambda = (\Lambda_1 \times \Lambda_2)$ . Dans ce cas il vient :

$$f_{\bar{x},\bar{y}}(x_t, y_t | \lambda, \bar{X}_{t-1}, \bar{Y}_{t-1}) = f_{\bar{y}|\bar{x}}(y_t | \lambda_1, \bar{X}_{t-1}, \bar{Y}_{t-1}, x_t) \times f_{\bar{x}}(x_t | \lambda_2, \bar{X}_{t-1}, \bar{Y}_{t-1}) \quad (52)$$

et on peut, pour estimer  $\psi_1$ , considérer uniquement le modèle conditionnel sans perte d'information utile. En particulier on n'a pas à travailler sur les lois marginales des variables faiblement exogènes. Au passage, vous remarquerez que cette exogénéité est relative à un ensemble de paramètres d'intérêt donnés : si on s'intéresse à d'autres paramètres, une variable qui était auparavant exogène faible peut très bien perdre ce statut (et inversement).

2. Exogénéité forte (*Strong exogeneity*) :  $x_t$  est une variable exogène forte (i) si  $x_t$  est exogène faible et (ii) si  $y$  n'est pas un prédicteur avancé de  $x$  (ou encore, si  $y$  ne cause pas  $x$  au sens de Granger), soit

$$f_{\bar{x}}(x_t | \lambda_2, \bar{X}_{t-1}, \bar{Y}_{t-1}) = f_{\bar{x}}(x_t | \lambda_2, \bar{X}_{t-1})$$

L'exogénéité forte montre tout son intérêt lorsque l'on veut faire des prévisions. on peut en effet prévoir les valeurs futures de la variable  $x$  à partir de son écriture marginale, ou se donner une trajectoire, puis calculer des prévisions pour  $y$  conditionnellement à cette trajectoire des  $x$ . Vous devez vous persuader qu'en l'absence de cette condition, il devient impossible de spécifier a priori une trajectoire sur  $x$  et d'en étudier les effets sur l'évolution à venir de  $y$ .

3. Super exogénéité (*Super exogeneity*) :  $x_t$  est super exogène pour  $\psi_1 = g(\lambda_1)$  si (i)  $x_t$  est exogène faible et (ii) si  $\lambda_1$  est invariant lorsque des modifications affectent  $\lambda_2$ .

L'intérêt de la super-exogénéité est d'assurer que l'on peut continuer à travailler avec l'écriture conditionnelle même si la loi marginale de  $x$  change. Un exemple type est celui où  $x$  est un instrument de la politique économique et  $y$  une variable telle que le chômage, le PIB ou l'inflation,...Si on a réussi à construire et à ajuster une équation dans laquelle  $y$  est expliquée par  $x$  alors, si  $x$  est super-exogène, un changement de politique économique qui se traduirait par un nouveau mode de détermination de  $x$ , et donc une nouvelle loi marginale  $f_{\bar{x}}(\cdot)$ , n'empêche pas de continuer à utiliser l'équation conditionnelle précédente pour apprécier les conséquences de cette modification sur  $y$ .

## 6.2.2 Exogénéité faible dans le VECM

Considérons deux ensembles de variables intégrées d'ordre 1 :  $k_1$  variables  $x_t$  et  $k_2$  variables  $y_t$  avec  $k = k_1 + k_2$  et telles que le système  $(x'_t, y'_t)'$  soit caractérisé

par  $r$  relations de cointégration avec  $1 < r < k$  :

$$\begin{pmatrix} \Delta x_t \\ \Delta y_t \end{pmatrix} = \alpha \beta (x_{t-1}^\top, y_{t-1}^\top)^\top + \sum_{i=1}^{p-1} \Phi_i^* \begin{pmatrix} \Delta x_{t-i} \\ \Delta y_{t-i} \end{pmatrix} + u_t \quad (53)$$

Si les  $k_1$  premières lignes de  $\alpha$  sont nulles alors les  $r$  termes d'erreurs afférents aux équations de long terme n'interviennent pas dans l'explication des variations courantes des variables  $x$ . En particulier les coefficients de ces équations,  $\beta$ , n'apparaissent plus dans les  $k_1$  premières équations. En l'absence d'autres contraintes, les coefficients de ces  $k_1$  équations appartiennent à un ensemble  $\Lambda_1$  et ceux des  $k_2$  suivantes à un ensemble  $\Lambda_2$  et sont tels que la totalité des paramètres est dans  $\Lambda_1 \times \Lambda_2$ . Dans ces conditions  $x_t$  est exogène faible pour  $\beta$ . En d'autres termes, l'estimation de  $\beta$  peut s'effectuer uniquement en considérant le système partiel suivant :

$$\Delta y_t = \alpha_2 \beta (x_{t-1}^\top, y_{t-1}^\top)^\top + \sum_{i=1}^{p-1} \Phi_{2i}^* \begin{pmatrix} \Delta x_{t-i} \\ \Delta y_{t-i} \end{pmatrix} + u_{2t} \quad (54)$$

Cela peut être intéressant en pratique puisque les tailles d'échantillon sont souvent réduites et le passage du système complet au système partiel peut augmenter sensiblement le nombre de degrés de liberté. Cette réduction doit également accroître la précision des estimations (à condition naturellement que les contraintes de nullité imposées sur certaines lignes de  $\alpha$  soient vraies).

Sous l'hypothèse précédente, les variables  $x$  sont quand à elles gouvernées par le processus :

$$\Delta x_t = \sum_{i=1}^{p-1} \Phi_{1i}^* \begin{pmatrix} \Delta x_{t-i} \\ \Delta y_{t-i} \end{pmatrix} + u_{1t} \quad (55)$$

A la lecture de cette dernière écriture, vous devez parfaitement comprendre que parler de  $x$  comme étant des variables exogènes n'a pas de sens :  $x_t$  est exogène faible pour les paramètres  $\beta$  mais à l'évidence  $x_t$  n'est pas exogène faible pour les paramètres  $\Phi^*$  : l'exogénéité faible de long terme (définie par rapport aux paramètres des équations d'équilibre de long terme) n'implique pas l'exogénéité faible de court terme (définie par rapport aux coefficients des ajustements conjoncturels  $\Phi^*$ ). Elle ne permet notamment pas de conclure sur le fait que  $y$  est ou n'est pas un prédicteur avancé de  $x$ .

On terminera ce point en précisant que SAS permet d'effectuer un test d'exogénéité faible de long terme pour chacune des variables expliquée par le VECM grâce à l'option EXOGENEITY dans la commande COINTEG. Ainsi, dans un VECM sur 3 variables  $x_t, y_t$  et  $z_t$  la commande

```
model x y z / p=2; cointeg rank=1 exogeneity ;
```

conduit à la réalisation de trois tests : exogénéité faible (*de long terme*) de  $x_t$ , puis exogénéité faible de  $y_t$  et finalement exogénéité faible de  $z_t$ , chacune des statistiques ayant dans cet exemple sous  $H_0$  une distribution asymptotique  $\chi^2(1)$ .

## 7 La causalité selon Granger dans un système cointégré

Si on est intéressé à la conduite d'un test de causalité selon Granger sur des variables  $I(1)$ , il convient de prendre quelques précautions : Les statistiques des tests LRT ou de Wald usuels peuvent être calculées sur la représentation autorégressive afférente aux variables en niveau. La difficulté vient du fait, et cela ne doit pas vous surprendre, que travaillant sur des variables non stationnaires, ces statistiques ne possèdent pas leur distribution asymptotique standard. En clair, elles ne sont pas sous la nulle des réalisations d'une variable de  $\chi^2$ .

Apparemment, les choses semblent se simplifier s'il y a cointégration : plutôt que de travailler à partir du VAR en niveaux, on considère l'écriture VECM dans laquelle les variables sont stationnaires. On peut donc espérer retrouver cette distribution standard pour les tests de nullité des coefficients qui traduisent la non causalité. Cependant les choses sont plus complexes : la démarche précédente suppose que préalablement au test de non causalité on ait statué sur la cointégration, pour justifier l'emploi du VECM. En d'autres termes, le second test (de non causalité) est conditionnel à la décision qu'a faite prendre le premier (de cointégration). Or cette décision a été prise en prenant un risque de première espèce : il existe une probabilité non nulle de s'être trompé en concluant par exemple à la cointégration.

Au final, la statistique que l'on obtient à la seconde étape devrait être une moyenne pondérée de la statistique que l'on obtient sous hypothèse de cointégration et de celle qu'on obtiendrait sous hypothèse de non cointégration. La loi de ce mélange n'a aucune raison d'être un  $\chi^2$ , i.e. celle que l'on aurait si on savait a-priori qu'il y a cointégration dans le système initial. Contrairement aux apparences, le recours à un VECM n'est donc pas exempt de défaut, et le raisonnement précédent semble conseiller une solution qui permettrait de tester l'absence de causalité sans avoir à effectuer de test préalable sur la configuration à retenir pour réaliser le test final.

Dans ce cas, le test de non causalité devrait être effectué sur la représentation autorégressive en niveau à condition naturellement de pouvoir connaître la distribution de la statistique. C'est précisément ce que proposent Toda et Yamamoto<sup>12</sup>. Ils montrent que si  $y_t = (y_{1t}, y_{2t}, \dots, y_{kt})$  est un vecteur de  $k$  variables intégrées avec  $y_i \sim I(d_i), i = 1, \dots, k$ , obéissant à une écriture autorégressive d'ordre  $p$ , alors on peut construire une statistique de test d'absence de causalité

---

12. Toda, H. Y. & Yamamoto, T., *Statistical inference in vector autoregressions with possibly integrated processes*, Journal of Econometrics, 1995, 66, 225-250.

à l'horizon 1 qui possède asymptotiquement une distribution de  $\chi^2$  dans un VAR d'ordre  $p + d$  où  $d = \max(d_1, d_2, \dots, d_k)$ .

Formellement, soit l'écriture VAR(p) usuelle :

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \dots + \Phi_p y_{t-p} + c + u_t \quad (56)$$

où  $u_t$  est un processus en bruit blanc. L'absence de causalité à l'horizon 1 de  $y_i, 1 \leq i \leq k$  sur n'importe quelle autre variable du système  $y_j, 1 \leq j \leq k, j \neq i$  se traduit par la nullité des éléments  $\Phi_s(j, i), s = 1, \dots, p$  : le test à effectuer concerne donc :

$$H_0 : \Phi_1(j, i) = \Phi_2(j, i) = \dots = \Phi_p(j, i) = 0$$

versus

$H_1$  : l'un au moins de ces coefficients est non nul

Ils montrent alors que dans le VAR augmenté d'ordre  $p + d$

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \dots + \Phi_p y_{t-p} + \Phi_{p+1} y_{t-p-1} + \dots + \Phi_{p+d} y_{t-p-d} + c + u_t \quad (57)$$

la statistique de Wald afférente au  $H_0$  ci-dessus, et nommée alors MWald, possède asymptotiquement une distribution de  $\chi^2(2)$  à  $p$  degrés de liberté. Notez bien que le test porte sur la nullité des coefficients des  $p$  premières valeurs retardées de  $y_i$  dans l'équation de  $y_j$ . Les coefficients des valeurs de  $y_i$  retardées aux ordres  $p + 1, \dots, p + d$  ne sont pas pris en compte : leur présence dans les équations vise uniquement à permettre à la statistique de test de posséder sa distribution usuelle à savoir un  $\chi^2$  à  $p$  degrés de liberté.

## 8 Exemples

### 8.1 Un premier exemple sur séries stationnaires

Dans cet exemple nous reprenons les séries simulées à la fin de la première partie du cours consacrée à la modélisation VAR sur séries stationnaires. Nous avons trois séries,  $y_1, y_2, y_3$  et un VAR sans constante. En plus du test de cointégration de Johansen, nous avons demandé l'application de tests de racine unitaire et notamment du test de Dickey-Fuller. Aucune option n'étant spécifiée pour johansen, c'est le test de la trace qui sera effectué.

```
proc varmax data=simul;
model y1 y2 y3 / p=1 dftest cointtest=(johansen);
run;
```

Les résultats des tests DF sont présentés dans la table 1. Quelle que soit la version du test  $(\tau, \tau_\mu, \tau_i)$ , nous serions bien en présence de séries  $I(0)$ .

Dickey-Fuller Unit Root Tests					
Variable	Type	Rho	Pr < Rho	Tau	Pr < Tau
y1	Zero Mean	-29.19	< .0001	-3.64	0.0004
	Single Mean	-29.27	0.0009	-3.63	0.0067
	Trend	-29.42	0.0057	-3.66	0.0296
y2	Zero Mean	-29.01	< .0001	-3.52	0.0006
	Single Mean	-28.93	0.0009	-3.50	0.0100
	Trend	-29.44	0.0056	-3.56	0.0384
y3	Zero Mean	-54.40	< .0001	-4.86	< .0001
	Single Mean	-54.70	0.0009	-4.84	0.0002
	Trend	-53.65	0.0003	-4.84	0.0008

TABLE 1 – Résultats des tests de racine unitaire

Cointegration Rank Test Using Trace						
H0: Rank=r	H1: Rank>r	Eigenvalue	Trace	5% Critical Value	Drift in ECM	Drift in Process
0	0	0.9336	359.2010	24.08	NOINT	Constant
1	1	0.5531	90.7193	12.21		
2	2	0.1051	10.9921	4.14		

TABLE 2 – Résultats des tests de racine unitaire

Dans la table 2 on trouve le test de la trace. La première ligne oppose H0 :r=0 à H1 : r>0. La statistique vaut 359.2 et avec une valeur critique à 5% égale à 24.08, on rejette évidemment H0 : le rang de cointégration serait donc non nul. On passe alors à la ligne suivante qui oppose H0 :r=1 à H1 : r>1. Avec une statistique de 9.075 et une valeur critique de 12.21, on est à nouveau conduit à rejeter H0. On continue donc à descendre dans la table pour s’intéresser à H0 :r=2 versus H1 : r>2. La statistique de test égale à 10.99, comparée à la valeur critique de 4.14, conduit encore au rejet de H0. Nous serions donc conduit à admettre que le rang est égal à 3, ce qui dans ce système à trois variables est le rang maximal de la matrice  $\pi$ . On rappelle alors que si  $r$ , le rang de  $\pi$ , est égal à  $k$ , le nombre de variables constituant le VAR, alors les variables en question sont I(0). En d’autres termes le test de la trace de Johansen conforte ici les conclusions tirées sur chacune des séries au moyen du test DF. L’ajustement d’un VAR en niveau sur  $y_1, y_2, y_3$  serait donc justifié.

## 8.2 Un second exemple sur séries I(1)

Ici nous avons simulé 200 réalisations de trois aléatoires gouvernées par le système suivant :

$$x_t = x_{t-1} + u_{xt} \quad (58)$$

$$z_t = z_{t-1} + u_{zt} \quad (59)$$

$$y_t = 2.0 + 1.6x_t - 0.8z_t + u_{yt} \quad (60)$$

où  $u_{xt}$ ,  $u_{zt}$  et  $z_{xt}$  sont des pseudo-réalisations de gaussiennes *i.i.d.* centrées réduites indépendantes entre elles. Les deux marches au hasard  $x$  et  $z$  sont évidemment I(1) et  $y$ , somme de deux I(1) et d’une I(0) est également intégrée



Minimum Information Criterion Based on AICC	
Lag	MA 0
AR 0	6.0456494
AR 1	0.2281736
AR 2	0.2970287
AR 3	0.3710118
AR 4	0.4585681
AR 5	0.5045197

TABLE 3 – Critère AICC sur les séries I(1)

d'ordre (1). Sous-jacent à (60) on reconnaît également une relation de cointégration intégrant une constante avec un vecteur cointégrant  $(2.0, 1.6, -1.0, -0.8)^T$  associé aux variables  $(1., x_t, y_t, z_t)^T$ .

Un exemple de réalisation de ce processus est donné dans la Figure (1). Ce sont ces données qui sont utilisées par la suite.

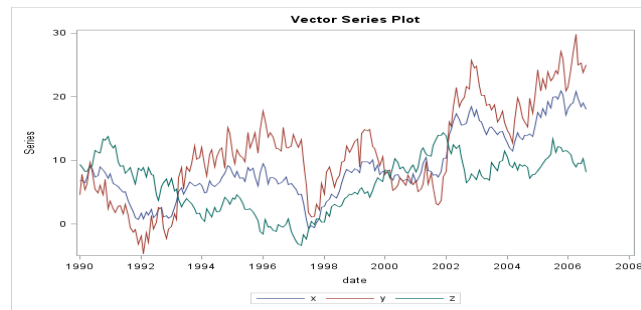


FIGURE 1 – Graphique de séries simulées selon les équations (58), (59), (60)

Comme vous l'aurez compris, la première étape consiste à déterminer l'ordre du processus autorégressif adapté à ces données. On emploie ici simplement l'option MINIC, via une commande de la forme :

```
model x y z / minic=(p=5 q=0) print=(roots parcoef);
```

Les valeurs du critère AICC présentées dans la table 3 nous conduisent à retenir un seul retard, ce que confirme l'examen de la table 4 des matrices de corrélations partielles.

Les autocorrélations résiduelles, non présentées ici sont également non significatives. Nous travaillerons donc par la suite avec un VAR(1). La recherche du rang de cointégration est réalisé au moyen du test de trace de Johansen avec :

```
model x y z / p=1 cointtest=(johansen);
```

L'ensemble des résultats associés à ce test est présenté dans la table 5. Notez que, toujours à des fins d'illustration, nous n'avons pas utilisé l'option NOINT

Schematic Representation of Partial Autoregression												
Variable/Lag	1	2	3	4	5	6	7	8	9	10	11	12
x	+. .	. . .	. . .	. . .	. . -	. . .	. . .	. . .	. . .	. . .	. . .	. . .
y	+. -	. . .	. . .	. . .	. . .	. + .	. . .	. . .	. . .	. . .	. . .	. . .
z	. . +	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .

+ is > 2\*std error, - is < -2\*std error, . is between

TABLE 4 – Matrices des Corrélations partielles

même si les constantes estimées s’avèrent finalement non significativement différentes de zéro. Selon les deux premiers tableaux, que la constante soit ou ne soit pas dans l’espace de cointégration<sup>13</sup>, l’hypothèse d’un rang nul est rejetée, ce qui n’est pas le cas de l’hypothèse d’un rang égal à l’unité. On retiendrait donc une seule relation de cointégration.

Les troisième et quatrième tableaux renseignent sur la place de la constante : le troisième rappelle simplement que l’hypothèse nulle imposant la constante dans l’espace de cointégration, elle correspond au second cas des cinq typologies vues précédemment. L’hypothèse H1 lève cette restriction et place la constante hors de l’espace de cointégration, c’est le troisième cas des cinq typologies qu’il est possible d’estimer avec la Proc VARMAX. Si on travaille sous l’hypothèse d’un rang égal à 1, la statistique de Chi2 égale à 0.58 ne permet pas de rejeter la nulle, i.e. on retiendrait la constante dans l’espace de cointégration ce qui correspond d’ailleurs au processus simulé.

On peut alors demander l’ajustement du modèle sélectionné par :

```
model x y z / p=1 print=(estimates) ecm=(rank=1 normalize=x
ectrend);
```

Ici nous avons arbitrairement imposé que le coefficient de la variable 'x' soit égal à l’unité. Les résultats suivants sont obtenus :

— Vecteur cointégrant, table 6, et loading factors estimés, table 7.

La relation de cointégration ajustée est donc  $x_t - 0.62y_t - 0.49z_t + 1.32$ . Cette combinaison linéaire serait stationnaire et définit l’écart à l’équilibre de long terme. La valeur en  $t-1$  de cet écart entre dans l’équation de  $\Delta x_t$  avec un coefficient égal à -0.13, dans celles de  $\Delta y_t$  et  $\Delta z_t$  avec les coefficients respectifs de 1.54 et -0.12.

L’estimation de la matrice  $\pi$ , donnée par  $\hat{\pi} = \hat{\alpha}\hat{\beta}^T$  est donnée dans la table 8. La procédure affiche également un schéma permettant de repérer aisément les coefficients significatifs et leur signe, table 9. Ici, les seuls coefficients non nuls concernent l’équation de  $\Delta y_t$ , avec une constante qui serait positive, et des coefficients de  $x$ ,  $y$ ,  $z$  respectivement positif, négatif et négatif.

13. Pour mémoire, lorsque la constante est en dehors de l’espace de cointégration, on a une constante dans le VECM qui impose un trend linéaire dans l’écriture autorégressive sur les niveaux des variables. Lorsque la constante est dans cet espace, on a toujours une constante dans le VECM et seulement une constante dans l’écriture autorégressive. Ceci explique l’intitulé des deux dernières colonnes des deux premiers tableaux de cette Figure (5).

Cointegration Rank Test Using Trace						
H0: Rank=r	H1: Rank>r	Eigenvalue	Trace	5% Critical Value	Drift in ECM	Drift in Process
0	0	0.5294	156.1788	29.38	Constant	Linear
1	1	0.0268	6.1739	15.34		
2	2	0.0039	0.7677	3.84		

Cointegration Rank Test Using Trace Under Restriction						
H0: Rank=r	H1: Rank>r	Eigenvalue	Trace	5% Critical Value	Drift in ECM	Drift in Process
0	0	0.5295	156.8053	34.80	Constant	Constant
1	1	0.0283	6.7563	19.99		
2	2	0.0052	1.0383	9.13		

Hypothesis of the Restriction		
Hypothesis	Drift in ECM	Drift in Process
H0(Case 2)	Constant	Constant
H1(Case 3)	Constant	Linear

Hypothesis Test of the Restriction					
Rank	Eigenvalue	Restricted Eigenvalue	DF	Chi-Square	Pr > ChiSq
0	0.5294	0.5295	3	0.63	0.8904
1	0.0268	0.0283	2	0.58	0.7473
2	0.0039	0.0052	1	0.27	0.6029

TABLE 5 – Résultats du test de la trace de Johansen dans VARMAX

- un tableau affichant pour chacune des équations du VECM, les coefficients estimés, leurs écarts-types, les students usuels pour le test de nullité ainsi que leurs niveaux de signification marginale, avec enfin en dernière colonne, le nom de l'explicative concernée, table 10. La seule surprise concerne peut être le fait qu'au sein d'une équation donnée de notre VECM, les statistiques de student soient toutes égales en valeur absolue, par exemple  $|t| = 6.67$  dans l'équation de  $\Delta y_t$ , ou encore,  $|t| = 1.03$  dans l'équation de  $\Delta z_t$  et cela quel que soit le coefficient considéré. Pour comprendre ce résultat, considérons arbitrairement la première équation du modèle<sup>14</sup> :

$$\Delta x_t = \alpha_1 (\beta_1 x_{t-1} + \beta_2 y_{t-1} + \beta_3 z_{t-1} + \beta_0) + u_{1t}, \text{ avec } \beta_1 = 1, \quad (61)$$

$$= \alpha_1 x_{t-1} + \alpha_1 \beta_2 y_{t-1} + \alpha_1 \beta_3 z_{t-1} + \alpha_1 \beta_0 + u_{1t} \quad (62)$$

$$= \pi_{11} x_{t-1} + \pi_{12} y_{t-1} + \pi_{13} z_{t-1} + \pi_{14} + u_{1t} \quad (63)$$

Les students de cette première équation de la table 10 ont donc comme nulle,  $H0 : \pi_{1i} = 0, i = 1, 2, 3, 4$ . Maintenant, le VECM ajusté suppose et impose une relation de cointégration dont les  $\beta$  sont les coefficients. dans ces conditions, les  $\hat{\beta}$  doivent être superconvergens avec une vitesse de convergence en  $T$ . En revanche,  $\alpha_1$  est le coefficient d'une stationnaire dans le VECM :  $\hat{\alpha}_1$  possède la vitesse de convergence usuelle, en  $\sqrt{T}$ . Cette différence de vitesse, permet à considérer que les  $\hat{\beta}$  sont des

14. Les trois équations ont la même structure : ce qui vaut pour l'une vaut pour les deux autres.

Long-Run Parameter Beta Estimates When RANK=1	
Variable	1
x	1.00000
y	-0.62404
z	-0.49183
1	1.32404

TABLE 6 – relation de cointégration estimée

Adjustment Coefficient Alpha Estimates When RANK=1	
Variable	1
x	-0.13463
y	1.54777
z	-0.11822

TABLE 7 – Coefficients d'ajustement à l'écart de long terme

constantes pour  $\hat{\alpha}_1$ . Au final :

$$\begin{aligned}
 t_{\pi_{1i}} &= \frac{\hat{\pi}_{1i}}{s_{\hat{\pi}_{1i}}} = \frac{\hat{\alpha}_1 \hat{\beta}_i}{s_{\hat{\alpha}_1 \hat{\beta}_i}} \text{ et, en considérant } \hat{\beta}_i \text{ comme étant une constante} \\
 &= \frac{\hat{\alpha}_1 \hat{\beta}_i}{\hat{\beta}_i s_{\hat{\alpha}_1}} = \frac{\hat{\alpha}_1}{s_{\hat{\alpha}_1}} = t_{\alpha_1}, i = 1, 2, 3, 4
 \end{aligned} \tag{64}$$

Ainsi, dans chacune des trois équations, les students construits sur les coefficients testent tous la même chose, à savoir la nullité du coefficient d'ajustement de l'écart à l'équilibre de long terme.

- Les tests d'exogénéité faible. Il s'agit de tester la nullité des coefficients  $\alpha$ . Cela s'effectue via la commande cointreg

Parameter Alpha * Beta' Estimates				
Variable	x	y	z	1
x	-0.13463	0.08401	0.06621	-0.17825
y	1.54777	-0.96588	-0.76123	2.04931
z	-0.11822	0.07377	0.05814	-0.15653

TABLE 8 – Estimation de la matrice  $\pi$

Schematic Representation of Parameter Estimates		
Variable/Lag	C	AR1
x	-	...
y	+	+...
z	-	...

+ is > 2\*std error, - is < -2\*std error, . is between, \* is N/A

TABLE 9 – Coefficients de  $\pi$  significatifs avec leur signe

Model Parameter Estimates						
Equation	Parameter	Estimate	Standard Error	t Value	Pr >  t	Variable
D_x	CONST1	-0.17825	0.15326	-1.16	0.2462	1, EC
	AR1_1_1	-0.13463	0.11575	-1.16	0.2462	x(t-1)
	AR1_1_2	0.08401	0.07223	1.16	0.2462	y(t-1)
	AR1_1_3	0.06621	0.05693	1.16	0.2462	z(t-1)
D_y	CONST2	2.04931	0.30704	6.67	<.0001	1, EC
	AR1_2_1	1.54777	0.23190	6.67	<.0001	x(t-1)
	AR1_2_2	-0.96588	0.14472	-6.67	<.0001	y(t-1)
	AR1_2_3	-0.76123	0.11405	-6.67	<.0001	z(t-1)
D_z	CONST3	-0.15853	0.15129	-1.03	0.3021	1, EC
	AR1_3_1	-0.11822	0.11427	-1.03	0.3021	x(t-1)
	AR1_3_2	0.07377	0.07131	1.03	0.3021	y(t-1)
	AR1_3_3	0.05814	0.05620	1.03	0.3021	z(t-1)

TABLE 10 – Le VECM estimé