

Chapitre I. Modèle VAR stationnaire

Stéphane Adjemian*

7 février 2010

1 Introduction

Dans ce chapitre nous discutons les propriétés du modèle VAR. On supposera que les paramètres du modèle sont connus et nous renvoyons la discussion de l'estimation de ce type de modèle au chapitre suivant.

2 Définitions

Soit $y_t \equiv (y_{1,t}, \dots, y_{K,t})'$ un vecteur colonne de K séries temporelles. Le modèle VAR(p) est défini par la récurrence suivante :

$$y_t = \nu + A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t \quad (2.1)$$

où les A_i ($i = 1, \dots, p$) sont des matrices $K \times K$ de coefficients, ν est un vecteur $K \times 1$ de coefficients et u_t est un bruit blanc multivarié ($K \times 1$) d'espérance nulle et de variance Σ_u (une matrice symétrique définie positive de dimension $K \times K$).

2.1 Le modèle VAR(1)

Concentrons nous sur le modèle avec un seul retard sur le vecteur de variables endogènes ($p = 1$). Nous verrons plus loin que nous pouvons toujours nous ramener à ce modèle.

Vous avez déjà rencontré, sans le savoir, des processus VAR(1). En effet, on peut montrer que tout processus AR(p) peut s'écrire sous la forme d'un processus VAR(1) de dimension p . Soit le processus scalaire $\{x_t\}$ AR(p) défini par :

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \dots + \varphi_p x_{t-p} + \epsilon_t$$

Posons :

$$y_t \equiv \begin{pmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-p+1} \end{pmatrix} \quad \text{et} \quad u_t \equiv \begin{pmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

*CEPREMAP, 142 rue Chevaleret, 75013 Paris. E-mail stephane [DOT] adjemian [AT] ens [DOT] fr.

des vecteurs $p \times 1$. Définissons la matrice $p \times p$:

$$A_1 = \begin{pmatrix} \varphi_1 & \varphi_2 & \dots & \dots & \varphi_p \\ 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 0 \end{pmatrix}$$

Nous avons alors :

$$y_t = A_1 y_{t-1} + u_t$$

en effet la première ligne de cette équation matricielle (il s'agit d'un système d'équations linéaires) nous dit que le premier élément de y_t est égal au produit de la première ligne de A_1 par y_{t-1} plus le premier élément de u_t , c'est-à-dire :

$$x_t = \sum_{i=1}^p A_{1,i} x_{t-i} + \epsilon_t$$

où $A_{1,i}$ est le i -ème élément de la première ligne de A_1 . Nous retrouvons donc notre processus AR(p) :

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \dots + \varphi_p x_{t-p} + \epsilon_t$$

Les équations suivantes du système de p équations définissent des identités (déterministes). Par exemple, l'équation j nous dit que x_{t-j} est égal au produit de la ligne j de A_1 par y_{t-1} plus zéro, c'est-à-dire : $x_{t-j} = x_{t-j}$. Si on peut toujours écrire un AR(p) sous la forme d'un VAR(1), la réciproque n'est pas vraie. Le premier élément de y_t n'est pas un AR(1) si la matrice A_1 n'est pas de la forme particulière postulée au-dessus. Plus généralement, par exemple pour une matrice A_1 pleine quelconque, le premier élément de y_t est un processus ARMA. Dans la suite on s'intéresse au modèle VAR(1) dans le cas général (c'est-à-dire pas dans le cas où il y a équivalence avec un modèle AR(p)).

Supposons que la condition initiale soit en $t = 0$. Par récurrence arrière nous pouvons exprimer y_t en fonction de la condition initiale y_0 , de ν , de A_1 , de l'innovation présente et des innovations passées :

$$y_t = (I_K + A_1 + \dots + A_1^{t-1})\nu + A_1^t y_0 + \sum_{i=0}^{t-1} A_1^i u_{t-i}$$

Plus généralement, si la condition initiale est $y_{t-\tau}$ avec éventuellement $\tau \rightarrow \infty$ lorsque nous remontons jusqu'à l'origine du monde, nous avons :

$$y_t = (I_K + A_1 + \dots + A_1^{\tau-1})\nu + A_1^\tau y_{t-\tau} + \sum_{i=0}^{\tau-1} A_1^i u_{t-i}$$

Dans le premier terme nous reconnaissons une série géométrique (matricielle). Si les valeurs propres de la matrices A_1 sont à l'intérieur du cercle unité alors la limite de ce terme est finie lorsque $\tau \rightarrow \infty$. Nous avons dans ce cas :

$$(I_K + A_1 + \dots + A_1^{\tau-1})\nu \xrightarrow{\tau \rightarrow \infty} (I_K - A_1)^{-1}\nu$$

Si les valeurs propres de A_1 sont inférieures à un en module alors A_1^τ tend vers 0 lorsque τ tend vers l'infini,

ainsi l'effet de la condition initiale sur y_t est d'autant plus faible que la condition initiale s'éloigne vers le passé. Enfin, nous pouvons montrer que sous la même condition (les valeurs propres de A_1 sont inférieures à 1 en module) $\sum_{i=0}^{\infty} A_1^i u_{t-i}$ existe en moyenne quadratique car la suite de matrices $\{A_1^i, i \in \mathbb{N}\}$ est absolument sommable¹. Au final, si toutes les valeurs propres de A_1 sont à l'intérieur du cercle unité, alors nous pouvons écrire y_t sous la forme suivante :

$$y_t = \mu + \sum_{i=0}^{\infty} A_1^i u_{t-i} \quad \text{pour tout } i \in \mathbb{Z} \quad (2.2)$$

avec

$$\mu \equiv (I_K - A_1)^{-1} \nu$$

L'équation (2.2) un modèle MA(∞) du modèle, il s'agit de la solution de l'équation récurrente stochastique définie par le modèle VAR(1).

Définition 2.1. *Le modèle VAR(1) est stable si toutes les valeurs propres de la matrice autorégressive A_1 sont inférieures à un en module.*

À partir de la forme MA(∞) on peut facilement calculer les moments d'ordre 1 et 2 du processus VAR(1). On montre directement que l'espérance est donnée par :

$$\mathbb{E}[y_t] = \mu \quad (2.3)$$

On obtient presque aussi facilement la fonction d'autocovariance. Par définition² nous avons :

$$\Gamma_y(h) \equiv \mathbb{E}[(y_t - \mu)(y_{t-h} - \mu)']$$

En substituant la forme MA(∞) il vient :

$$\Gamma_y(h) = \mathbb{E} \left[\lim_{n \rightarrow \infty} \sum_{i=0}^n \sum_{j=0}^n A_1^i \mathbb{E}[u_{t-i} u'_{t-h-j}] A_1^{j'} \right]$$

Puisque u_t est un bruit blanc, nous savons que $\mathbb{E}[u_{t-i} u'_{t-h-j}]$ est non nul si et seulement si $t-i = t-h-j$, c'est-à-dire si et seulement si $i = h+j$. Ainsi, nous avons :

$$\Gamma_y(h) = \lim_{n \rightarrow \infty} \sum_{j=0}^n A_1^{h+j} \Sigma_u A_1^{j'}$$

ou encore, puisque $\{A_1^i, i \in \mathbb{N}\}$ est absolument sommable,

$$\Gamma_y(h) = \sum_{i=0}^{\infty} A_1^{h+i} \Sigma_u A_1^{i'} \quad (2.4)$$

¹La limite $\lim_{n \rightarrow \infty} \sum_{i=-n}^n |A_1^i| < \infty$ existe.

²Notons que la matrice d'autocovariance d'ordre h est égale à la transposée de la matrice d'autocovariance d'ordre $-h$, $\Gamma_y(h) = \gamma_y(-h)'$.

En particulier la variance de y_t est définie par :

$$\mathbb{V}[y_t] = \Gamma_y(0) = \sum_{i=0}^{\infty} A_1^i \Sigma_u A_1^{i'} \quad (2.5)$$

Par construction il s'agit d'une matrice symétrique (ie $\mathbb{V}[y_t] = \mathbb{V}[y_t]'$) définie positive.

Remarque 2.1. Nous avons calculé les moments en supposant que la condition initiale est très éloignée dans le temps ($\tau \rightarrow \infty$), nous avons supposé que le processus VAR(1) tourne depuis l'origine du monde. Pour être plus précis, on parle parfois de moments asymptotiques. Les moments asymptotiques d'un modèle VAR stable ne dépendent pas du temps. Si la condition initiale n'est pas en $-\infty$ alors les moments dépendent généralement de t .

Exercice 2.1. Calculez l'espérance et la variance de $y_t \sim \text{VAR}(1)$ si la condition initiale est $y_0 \in \mathbb{R}^K$ et si les valeurs propres de la matrice autorégressive A_1 sont toutes inférieures à 1 en module. Montrez que ces moments tendent vers les moments asymptotiques lorsque t tend vers l'infini.

Exercice 2.2 (suite de l'exercice 2.1). Quelles conditions faut-il formuler sur la condition initiale y_0 pour que l'espérance et la variance de y_t soient invariantes ?

En pratique on ne calcule pas la variance à partir de la forme MA(∞). En appliquant l'opérateur variance à la définition du processus VAR(1), nous avons :

$$\begin{aligned} \mathbb{V}[y_t] &= \mathbb{V}[A_1 y_{t-1} + u_t] \\ &= \mathbb{V}[A_1 y_{t-1}] + \mathbb{V}[u_t] \\ &= A_1 \mathbb{V}[y_{t-1}] A_1' + \Sigma_u \end{aligned}$$

où la deuxième égalité vient de l'orthogonalité entre l'innovation et le passé de y_t et la troisième vient des propriétés de la variance et de la définition de u_t . Si le modèle VAR est stable alors la variance asymptotique ne dépend pas du temps, $\mathbb{V}[y_{t-1}] = \mathbb{V}[y_t]$, et nous avons donc :

$$\mathbb{V}[y_t] = A_1 \mathbb{V}[y_t] A_1' + \Sigma_u$$

Nous devons résoudre cette équation en $\mathbb{V}[y_t]$ pour obtenir la variance. En vectorisant cette équation, il vient :

$$\text{vec } \mathbb{V}[y_t] = (A_1 \otimes A_1) \text{vec } \mathbb{V}[y_t] + \text{vec } \Sigma_u$$

soit de façon équivalente :

$$\text{vec } \mathbb{V}[y_t] = (I_{K^2} - A_1 \otimes A_1)^{-1} \text{vec } \Sigma_u$$

2.2 Le modèle VAR(p)

Il est possible de transformer un modèle VAR(p) en modèle VAR(1) en augmentant le nombre de variables endogènes (de la même façon qu'il est toujours possible de se ramener d'une équation différentielle d'ordre n à une équation différentielle d'ordre 1). Nous pouvons donc généraliser assez facilement les résultats de la section 2.1.

Nous pouvons écrire le modèle VAR(p) défini par (2.1) de la façon suivante :

$$Y_t = \nu + \mathbf{A}Y_{t-1} + U_t \quad (2.6)$$

avec :

$$Y_t \equiv \begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{pmatrix}$$

$$\nu \equiv \begin{pmatrix} \nu \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$U_t \equiv \begin{pmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

et la matrice compagnon :

$$\mathbf{A} \equiv \begin{pmatrix} A_1 & A_2 & \dots & \dots & A_p \\ I_K & 0 & \dots & \dots & 0 \\ 0 & I_K & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & & & I_K & 0 \end{pmatrix}$$

Une condition de stabilité du modèle VAR(p) est donc simplement que les valeurs propres de la matrice compagnon \mathbf{A} sont inférieures à un en module. Nous proposons une condition alternative de la stabilité du modèle VAR(p) et nous montrerons plus loin en quoi elle est liée à la condition sur les valeurs propres de la matrice compagnon.

Définition 2.2. *Le modèle VAR(p) est stable si $\det(I_{Kp} - \mathbf{A}z) \neq 0$ pour tout $|z| \leq 1$.*

En utilisant les propriétés du déterminant, on établit facilement l'équivalence avec la définition suivante :

Proposition 2.1. *Le modèle VAR(p) est stable si $\det(I_K - A_1z - A_2z^2 - \dots - A_pz^p) \neq 0$ pour tout $|z| \leq 1$.*

Preuve 2.1. *Dans le cas d'un modèle VAR(p), nous avons :*

$$\det(I_{Kp} - \mathbf{A}z) = \begin{vmatrix} I_K - A_1z & -A_2z & \dots & \dots & A_pz \\ -zI_K & I_K & 0 & \dots & 0 \\ 0 & -zI_K & I_K & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & & & -zI_K & I_K \end{vmatrix}$$

RAPPEL : Si on multiplie une colonne (ou *a fortiori* un groupe de m colonnes) par une constante et additionne le résultat à une autre colonne (ou un autre groupe de m colonnes), le déterminant de la matrice n'est pas affecté.

En multipliant les K dernières colonnes (qui correspondent au retard p) par z et en ajoutant le résultat au K colonnes précédentes, nous obtenons :

$$\det(I_{Kp} - \mathbf{A}z) = \begin{vmatrix} I_K - A_1z & -A_2z & \dots & A_{p-1}z + A_pz^2 & A_pz \\ -zI_K & I_K & 0 & 0 & 0 \\ 0 & -zI_K & I_K & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & & & 0 & I_K \end{vmatrix}$$

Cette transformation permet d'éliminer le bloc $-zI_K$ sous la diagonale de la matrice. Nous répétons cette transformation en parcourant la matrice vers la gauche. Après $p - 2$ répétitions sur des bloc de K colonnes, nous obtenons :

$$\det(I_{Kp} - \mathbf{A}z) = \begin{vmatrix} I_K - A_1z - A_2z^2 - \dots - A_pz^p & \mathcal{X} \\ 0 & I_{K(p-1)} \end{vmatrix}$$

où \mathcal{X} est une matrice $K \times K(p-1)$ dont la connaissance précise n'est pas nécessaire. De façon équivalente, nous avons finalement :

$$\det(I_{Kp} - \mathbf{A}z) = |I_K - A_1z - A_2z^2 - \dots - A_pz^p|$$

□

Cette proposition nous dit que si les racines du polynôme retard³ sont supérieures à l'unité en module, alors le modèle VAR(p) est stable. Une condition alternative de stabilité est donnée par la proposition suivante :

Proposition 2.2. *Le modèle VAR(p) est stable si les valeurs propres de \mathbf{A} sont inférieures à un en module.*

Preuve 2.2. *En factorisant z , nous avons :*

$$\begin{aligned} \det(I_{Kp} - \mathbf{A}z) &= \det(z(I_{Kp}z^{-1} - \mathbf{A})) \\ &= z^{Kp} \det(I_{Kp}z^{-1} - \mathbf{A}) \end{aligned}$$

Les valeurs de $1/z$ qui annulent le déterminant sur le membre de droite de la dernière égalité sont les valeurs propres de la matrice compagnon \mathbf{A} . Si les valeurs propres sont toutes à l'intérieur du cercle unitaire, alors la condition (2.2) est nécessairement satisfaite, le modèle VAR(p) est donc stable. □

Exemple 2.1. *Soit le modèle VAR(1) dans \mathbb{R}^3 (ie $p = 1$ et $K = 3$) suivant :*

$$y_t = \begin{pmatrix} 0,1 \\ 0,2 \\ 0,3 \end{pmatrix} + \begin{pmatrix} 0,5 & 0 & 0 \\ 0,1 & 0,1 & 0,3 \\ 0 & 0,2 & 0,3 \end{pmatrix} y_{t-1} + u_t$$

Les racines du polynôme retard sont les valeurs de z qui satisfont :

$$\left| \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 0,5 & 0 & 0 \\ 0,1 & 0,1 & 0,3 \\ 0 & 0,2 & 0,3 \end{pmatrix} \times z \right| = 0$$

³Notons qu'il s'agit d'un polynôme retard *matriciel*, voir l'équation 2.9 pour une définition du polynôme retard associée à un VAR(p). Les racines d'un polynôme retard $\mathcal{A}(L)$ sont les valeurs de z telles que le déterminant $\det(\mathcal{A}(z))$ est nul.

c'est-à-dire les valeurs de z telles que :

$$\begin{vmatrix} 1 - 0,5z & 0 & 0 \\ -0,1z & 1 - 0,1z & -0,3z \\ 0 & -0,2z & 1 - 0,3z \end{vmatrix} = 0$$

ou encore :

$$(1 - 0,5z) \begin{vmatrix} 1 - 0,1z & -0,3z \\ -0,2z & 1 - 0,3z \end{vmatrix} = 0$$

$$\Leftrightarrow (1 - 0,5z)(1 - 0,4z - 0,03z^2) = 0$$

Les racines du polynôme retard sont donc :

$$z_1 = 2, \quad z_2 = 2,1525 \text{ et } z_3 = -15,4858$$

Elles sont toutes plus grandes que 1 en valeur absolue, le modèle VAR(1) est donc stable.

Notons que dans cet exemple les racines du polynôme retard sont réelles, ce n'est pas toujours le cas, comme le montre l'exemple suivant.

Exemple 2.2. Soit le processus VAR(2) dans \mathbb{R}^2 (ie $p = 2$ et $K = 2$) suivant :

$$y_t = \begin{pmatrix} 0,1 \\ 0,2 \end{pmatrix} + \begin{pmatrix} 0,5 & 0,1 \\ 0,4 & 0,5 \end{pmatrix} y_{t-1} + \begin{pmatrix} 0 & 0 \\ 0,25 & 0 \end{pmatrix} y_{t-2} + u_t$$

Les racines du polynôme retard sont données par :

$$\det \left(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} 0,5 & 0,1 \\ 0,4 & 0,5 \end{pmatrix} \times z - \begin{pmatrix} 0 & 0 \\ 0,25 & 0 \end{pmatrix} z^2 \right) = 1 - z + 0,21z^2 - 0,025z^3 = 0$$

c'est-à-dire :

$$z_1 = 1,3, \quad z_2 = 3,55 + 4,26i \text{ et } z_3 = 3,55 - 4,26i$$

Nous avons $|z_2| = |z_3| = \sqrt{3,55^2 + 4,26^2} = 5,545$, ainsi toutes les racines du polynôme retard sont supérieures à 1 en module, le modèle VAR(2) est stable.

Notons que, même si deux des racines du polynôme retard sont complexes, le VAR(2) décrit dans cet exemple caractérise la dynamique de variables réelles. La présence de racines non réelles (ie dans $\mathbb{C} - \mathbb{R}$) traduit la présence de composantes cycliques dans les variables étudiées.

2.3 La représentation moyenne mobile d'un modèle VAR(p)

Nous avons vu qu'il est possible d'écrire le modèle VAR(p) (2.1) sous la forme d'un modèle VAR(1) grâce à la forme compagnon (2.6) :

$$Y_t = \nu + \mathbf{A}Y_{t-1} + U_t$$

Si la condition de stabilité est satisfaite alors par récurrence arrière nous pouvons écrire ce modèle sous une forme moyenne mobile infinie :

$$Y_t = \mu + \sum_{i=0}^{\infty} \mathbf{A}^i U_{t-i}$$

À partir de la forme moyenne mobile de $\{Y_t\}$ on peut obtenir la forme moyenne mobile de $\{y_t\}$. Pour cela, on définit la matrice $K \times Kp$ de sélection :

$$J = (I_K, 0, \dots, 0)$$

On vérifie que cette matrice de sélection satisfait $JJ' = I_K$ et

$$J'J = \begin{pmatrix} I_K & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & \dots & 0 \end{pmatrix}$$

Cette matrice (disons $\mathcal{J} = J'J$) peut être considérée comme un élément neutre pour les vecteurs de la forme U_t (ou μ), dans le sens :

$$\mathcal{J}U_t = U_t$$

Notons que ceci tient à la forme très particulière du vecteur U_t qui est un vecteur dont toutes les entrées sont nulles à l'exception des K premières. En pré-multipliant la forme moyenne mobile de $\{Y_t\}$ par la matrice J , il vient :

$$y_t = J\mu + \sum_{i=0}^{\infty} J\mathbf{A}^i U_{t-i}$$

On a encore :

$$y_t = \mu + \sum_{i=0}^{\infty} J\mathbf{A}^i \mathcal{J}U_{t-i}$$

puisque l'application de \mathcal{J} à U_{t-h} laisse U_{t-h} inchangé. Ainsi, nous avons :

$$y_t = \mu + \sum_{i=0}^{\infty} J\mathbf{A}^i J' u_{t-i}$$

puisque J sélectionne les K premières lignes de U_t . En définissant $\Phi_i \equiv J\mathbf{A}^i J'$ une matrice $K \times K$, nous avons finalement :

$$y_t = \mu + \sum_{i=0}^{\infty} \Phi_i u_{t-i} \quad (2.7)$$

À partir de cette représentation, nous pouvons calculer les moments d'ordre 1 et 2 de $\{y_t\}$. En appliquant l'opérateur espérance nous avons directement :

$$\mathbb{E}[y_t] = \mu$$

Pour la fonction d'autocovariance, nous avons :

$$\begin{aligned}
\Gamma_y(h) &= \mathbb{E} [(y_t - \mu)(y_{t-h} - \mu)] \\
&= \mathbb{E} \left[\left(\sum_{i=0}^{\infty} \Phi_i u_{t-i} \right) \left(\sum_{i=0}^{\infty} \Phi_i u_{t-h-i} \right)' \right] \\
&= \mathbb{E} \left[\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \Phi_i u_{t-i} u'_{t-h-j} \Phi_j' \right] = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \Phi_i \mathbb{E} [u_{t-i} u'_{t-h-j}] \Phi_j'
\end{aligned}$$

Finalement, en remarquant que $\mathbb{E} [u_{t-i} u'_{t-h-j}]$ est non nul si et seulement si $t-i = t-h-j$, c'est-à-dire si et seulement si $j+h = i$, nous avons :

$$\Gamma_y(h) = \sum_{i=0}^{\infty} \Phi_{i+h} \Sigma_u \Phi_i'$$

Le problème est que nous ne connaissons pas explicitement les matrices Φ_i (nous savons seulement que $\Phi_0 = I_K$). Dans la suite nous décrivons un algorithme pour calculer les matrices de la représentation moyenne mobile.

On note L l'opérateur retard, c'est-à-dire l'opérateur tel que $Ly_t = y_{t-1}$ et plus généralement $L^h y_t = y_{t-h}$. En utilisant cet opérateur nous pouvons écrire le modèle VAR(p) de la façon suivante :

$$y_t = \nu + (A_1 L + A_2 L^2 + \dots + A_p L^p) y_t + u_t$$

ou encore :

$$A(L)y_t = \nu + u_t \tag{2.8}$$

ou le polynôme retard $A(L)$ est donné par :

$$A(L) = I_K - A_1 L - A_2 L^2 - \dots - A_p L^p \tag{2.9}$$

Posons :

$$\Phi(L) = \sum_{i=0}^{\infty} \Phi_i L^i$$

un autre polynôme retard tel que :

$$\Phi(L)A(L) = I_K$$

En pré-multipliant (2.8) par ce polynôme retard, il vient (par définition) :

$$y_t = \Phi(L)\nu + \Phi(L)u_t$$

comme le retard d'une constante est la constante elle même, on a $\Phi(L)\nu = \Phi(1)\nu$ et donc :

$$y_t = \nu \sum_{i=0}^{\infty} \Phi_i + \sum_{i=0}^{\infty} \Phi_i u_{t-i} \tag{2.10}$$

On dit que le polynôme $\Phi(L)$ est l'inverse du polynôme $A(L)$ et on notera souvent $\Phi(L) = A(L)^{-1}$. Le polynôme $A(L)$ est inversible si $\det(A(z)) \neq 0$ pour tout $|z| \leq 1$. Si cette condition est satisfaite les matrices de coefficients Φ_i sont absolument sommables et donc le processus moyenne mobile $A(L)^{-1}u_t$ est bien défini. Il

est possible d'identifier les matrices Φ_i en notant que par définition de $\Phi(L)$ nous avons :

$$I_K = (\Phi_0 + \Phi_1 L + \Phi_2 L^2 + \dots) (I_K - A_1 L - A_2 L^2 - \dots - A_p L^p)$$

soit en développant et regroupant les termes par puissance sur l'opérateur retard :

$$\begin{aligned} I_K &= \Phi_0 + (\Phi_1 - \Phi_0 A_1) L + (\Phi_2 - \Phi_1 A_1 - \Phi_0 A_2) L^2 + \dots \\ &\quad + \left(\Phi_h - \sum_{j=1}^h \Phi_{h-j} A_j \right) L^h \\ &\quad + \dots \end{aligned}$$

On voit immédiatement que Φ_0 doit être égal à la matrice identité I_K et que les termes associés aux puissances strictement positives de l'opérateur retard L doivent être nuls. Ainsi, nous pouvons identifier les matrices de la forme moyenne mobile infinie à partir de la récurrence suivante :

$$\begin{aligned} \Phi_0 &= I_K \\ \Phi_1 &= \Phi_0 A_1 \\ \Phi_2 &= \Phi_1 A_1 + \Phi_0 A_2 \\ &\vdots \\ \Phi_h &= \sum_{j=1}^h \Phi_{h-j} A_j \\ &\vdots \end{aligned} \tag{2.11}$$

où $A_j = 0$ pour tout $j > p$. L'espérance μ de $\{y_t\}$ est obtenue de la façon suivante :

$$\mu = \Phi(1)\nu = A(1)^{-1}\nu = (I_K - A_1 - \dots - A_p)^{-1}\nu$$

Exemple 2.3 (Suite de l'exemple 2.2). Dans le cas d'un processus VAR(2), les matrices de la représentation MA(∞) sont définies par la récurrence suivante :

$$\begin{aligned} \Phi_0 &= I_K \\ \Phi_1 &= A_1 \\ \Phi_2 &= \Phi_1 A_1 + A_2 = A_1^2 + A_2 \\ \Phi_3 &= \Phi_2 A_1 + \Phi_1 A_2 = A_1^3 + A_2 A_1 + A_1 A_2 \\ &\vdots \\ \Phi_h &= \Phi_{h-1} A_1 + \Phi_{h-2} A_2 \\ &\vdots \end{aligned}$$

Étant données les valeurs numériques dans l'exemple 2.2, nous avons :

$$\Phi_1 = \begin{pmatrix} 0,5 & 0,1 \\ 0,4 & 0,5 \end{pmatrix}, \quad \Phi_2 = \begin{pmatrix} 0,29 & 0,1 \\ 0,65 & 0,29 \end{pmatrix}, \quad \Phi_3 = \begin{pmatrix} 0,21 & 0,079 \\ 0,566 & 0,21 \end{pmatrix}, \dots$$

On peut vérifier que les matrices Φ_i tendent vers zéro (de façon non monotone ici) lorsque $i \rightarrow \infty$, autrement ces matrices ne seraient pas absolument sommables.

Notons que la représentation moyenne mobile d'un modèle VAR n'est pas toujours d'ordre infini.

Exercice 2.3. Soit le modèle VAR(1) dans \mathbb{R}^2 suivant :

$$y_t = \nu + \begin{pmatrix} 0 & \alpha \\ 0 & 0 \end{pmatrix} y_{t-1} + u_t$$

avec α un paramètre réel. Écrivez la représentation moyenne mobile de ce processus.

2.4 Processus stationnaire

Définition 2.3. Un processus est (asymptotiquement) stationnaire au second ordre si (pour une condition initiale particulière) ses moments d'ordre 1 et 2 sont invariants.

Autrement dit un processus $\{y_t\}$ est stationnaire si :

$$\mathbb{E}[y_t] = \mu \text{ pour tout } t$$

et

$$\mathbb{E}[(y_t - \mu)(y_{t-h} - \mu)'] = \Gamma_y(h) = \Gamma_y(-h)' \text{ pour tout } t \text{ et pour tout } h \in \mathbb{N}$$

La fonction d'autocovariance d'un processus stationnaire ne dépend que de h (et pas de t).

Définition 2.4. Un processus est stationnaire si la distribution jointe de n variables consécutives $(y_{t+1}, y_{t+2}, \dots, y_{t+n})$ ne dépend pas de t .

Notons que cette dernière définition est plus générale que la définition 2.3 de stationnarité au second ordre. Si un processus est stationnaire, alors il est stationnaire au second ordre (la réciproque n'est pas vraie). Par construction, un bruit blanc est stationnaire.

Proposition 2.3. Un processus VAR(p) stable est stationnaire au second ordre.

Preuve 2.3. Nous avons déjà vu qu'il est possible d'écrire la forme moyenne mobile infinie d'un processus VAR stable. Nous avons vu aussi comment calculer l'espérance ainsi que la fonction d'autocovariance du VAR(p) à partir de la forme moyenne mobile et vu que dans ce cas ces moments ne dépendent pas du temps. \square

Cette proposition explique pourquoi on associe souvent la condition de stabilité à une condition de stationnarité. Le théorème suivant donne un résultat général sur la représentation des processus stationnaires au second ordre.

Théorème 2.1 (Décomposition de Wold). Tout processus stationnaire au second ordre $\{x_t\}$ peut s'écrire comme la somme de deux processus non corrélés $\{z_t\}$ et $\{y_t\}$,

$$x_t = z_t + y_t$$

où z_t est un processus déterministe qui peut être prédit parfaitement à partir de son passé et y_t est un processus admettant une représentation moyenne mobile,

$$y_t = \sum_{i=0}^{\infty} \Phi_i u_{t-i}$$

où $\Phi_0 = I_K$, u_t est un bruit blanc et la somme $\sum_{i=0}^{\infty} \Phi_i u_{t-i}$ est définie en moyenne quadratique même si les $\{\Phi_i\}_{i=0}^{\infty}$ ne sont pas nécessairement absolument sommables.

Si nous supposons que la partie déterministe du processus qui nous intéresse est son espérance, ce théorème nous dit que tout processus stochastique stationnaire au second ordre peut s'écrire sous la forme d'un $MA(\infty)$. Si de plus nous supposons que les $\{\Phi_i\}_{i=0}^{\infty}$ sont absolument sommables, alors le polynôme retard $\Phi(L)$ est inversible, on notera $A(L) = \Phi(L)^{-1}$. Ainsi $\{y_t\}$ a une représentation VAR avec éventuellement un nombre infini de retards :

$$y_t = \sum_{i=1}^{\infty} A_i y_{t-i} + u_t$$

Puisque les $\{\Phi_i\}_{i=0}^{\infty}$ sont absolument sommables, nous avons $A_i \xrightarrow{i \rightarrow \infty} 0$ et donc nous pouvons approximer ce $VAR(\infty)$ par un modèle $VAR(p)$ pour un nombre de retard p assez grand. Ainsi le théorème de décomposition de Wold suggère que le modèle $VAR(p)$ peut nous permettre de caractériser tout processus stationnaire au second ordre. Ce résultat souligne l'intérêt de la modélisation VAR ; supposer que les données sont générées par un $VAR(p)$ n'est pas une hypothèse trop forte, dès lors que l'on admet que le processus d'intérêt est stationnaire au second ordre.

2.5 Moments d'ordre 2 d'un modèle VAR(p)

Nous pourrions obtenir la fonction d'autocovariance du modèle VAR (2.1) à partir de la forme compagnon (2.6) et de la représentation moyenne mobile. Comme dans la section 2.1 nous montrons ici comment calculer la fonction d'autocovariance sans passer par la forme $MA(\infty)$.

Nous avons par définition de l'espérance :

$$y_t - \mu = A_1(y_{t-1} - \mu) + \dots + A_p(y_{t-p} - \mu) + u_t$$

En post-multipliant $y_t - \mu$ par sa transposée puis en appliquant l'opérateur espérance, nous obtenons :

$$\Gamma_y(0) = A_1 \Gamma_y(-1) + \dots + A_p \Gamma_y(-p) + \Sigma_u$$

soit encore :

$$\Gamma_y(0) = A_1 \Gamma_y(1)' + \dots + A_p \Gamma_y(p)' + \Sigma_u \quad (2.12)$$

et $\forall h > 0$ nous avons :

$$\Gamma_y(h) = A_1 \Gamma_y(h-1) + \dots + A_p \Gamma_y(h-p) \quad (2.13)$$

Le système d'équations (2.12) et (2.13) sont les équations de Yule-Walker. Si on connaît $\Gamma_y(0), \Gamma_y(1), \dots, \Gamma_y(p-1)$, alors on peut décrire la fonction d'autocovariance pour tout $h \geq p$ à partir de la récurrence (2.13). Nous avons ici p matrices $K \times K$ inconnues et nous disposons de p équations matricielles. Nous pouvons donc obtenir la fonction d'autocovariance en résolvant un système d'équations linéaires.

Une approche alternative est de passer par la forme compagnon, c'est-à-dire l'écriture sous la forme d'un $VAR(1)$ du processus $VAR(p)$. En appliquant l'opérateur variance sur le $VAR(p)$, équation (2.6), on obtient directement :

$$\Gamma_Y(0) = \mathbf{A} \Gamma_Y(0) \mathbf{A}' + \Sigma_U$$

où

$$\Gamma_Y(0) = \begin{pmatrix} \Gamma_y(0) & \Gamma_y(1) & \Gamma_y(2) & \dots & \Gamma_y(p-1) \\ \Gamma_y(1)' & \Gamma_y(0) & \Gamma_y(1) & \dots & \Gamma_y(p-2) \\ \vdots & & \ddots & & \\ \Gamma_y(p-1)' & & & & \Gamma_y(0) \end{pmatrix}$$

En appliquant l'opérateur vec , on obtient directement :

$$\text{vec} \Gamma_Y(0) = (I_{(Kp)^2} - \mathbf{A} \otimes \mathbf{A})^{-1} \text{vec} \Sigma_U \quad (2.14)$$

Exemple 2.4 (Suite de l'exemple 2.2). *La matrice compagnon est donnée par :*

$$\begin{pmatrix} 0,5 & 0,1 & 0 & 0 \\ 0,4 & 0,5 & 0,25 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

en supposant que :

$$\Sigma_u = \begin{pmatrix} 0,09 & 0 \\ 0 & 0,04 \end{pmatrix}$$

on a :

$$\Sigma_U = \begin{pmatrix} 0,09 & 0 & 0 & 0 \\ 0 & 0,04 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Pour initialiser les équations de Yule-Walker, nous devons calculer :

$$\Gamma_Y(0) = \begin{pmatrix} \Gamma_y(0) & \Gamma_y(1) \\ \Gamma_y(1)' & \Gamma_y(0) \end{pmatrix}$$

En appliquant (2.14), nous avons :

$$\text{vec} \Gamma_Y(0) = (I_{16} - \mathbf{A} \otimes \mathbf{A})^{-1} \begin{pmatrix} 0,09 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0,04 \\ 0 \\ 0 \\ 0 \\ 0 \\ \mathbf{0}_{8 \times 1} \end{pmatrix}$$

On obtient :

$$\Gamma_y(0) = \begin{pmatrix} 0,131 & 0,066 \\ 0,066 & 0,181 \end{pmatrix} \quad \text{et} \quad \Gamma_y(1) = \begin{pmatrix} 0,072 & 0,051 \\ 0,104 & 0,143 \end{pmatrix}$$

puis en utilisant la récurrence (2.13) :

$$\Gamma_y(2) = \begin{pmatrix} 0,5 & 0,1 \\ 0,4 & 0,5 \end{pmatrix} \begin{pmatrix} 0,072 & 0,051 \\ 0,104 & 0,143 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0,25 & 0 \end{pmatrix} \begin{pmatrix} 0,131 & 0,066 \\ 0,066 & 0,181 \end{pmatrix} = \begin{pmatrix} 0,046 & 0,040 \\ 0,113 & 0,108 \end{pmatrix}$$

puis

$$\Gamma_y(3) = \begin{pmatrix} 0,5 & 0,1 \\ 0,4 & 0,5 \end{pmatrix} \begin{pmatrix} 0,046 & 0,040 \\ 0,113 & 0,108 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0,25 & 0 \end{pmatrix} \begin{pmatrix} 0,072 & 0,051 \\ 0,104 & 0,143 \end{pmatrix} = \begin{pmatrix} 0,035 & 0,031 \\ 0,093 & 0,083 \end{pmatrix}$$

et caetera.

Terminons en définissant la fonction d'auto-corrélation d'un processus VAR. La fonction d'autocovariance est souvent difficile à lire, pour en faciliter l'interprétation on la normalise. Soit D une matrice diagonale contenant les racines carrées des éléments sur la diagonale de $\Gamma_y(0)$:

$$D = \begin{pmatrix} \sqrt{\gamma_{y,11}} & 0 & \dots & 0 \\ 0 & \sqrt{\gamma_{y,22}} & 0 & 0 \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \sqrt{\gamma_{y,KK}} \end{pmatrix}$$

La fonction d'autocorrélation est définie par :

$$R_y(h) = D^{-1}\Gamma_y(h)D^{-1} \quad (2.15)$$

Par construction, les éléments sur la diagonale de $R_y(h)$ sont égaux à 1.

Exemple 2.5 (Suite de l'exemple 2.2). *On a directement*

$$R_h(0) = \begin{pmatrix} 0,3619 & 0 \\ 0 & 0,4254 \end{pmatrix}^{-1} \begin{pmatrix} 0,131 & 0,066 \\ 0,066 & 0,181 \end{pmatrix} \begin{pmatrix} 0,3619 & 0 \\ 0 & 0,4254 \end{pmatrix}^{-1} = \begin{pmatrix} 1,0000 & 0,4286 \\ 0,4286 & 1,0000 \end{pmatrix}$$

$$R_h(1) = \begin{pmatrix} 0,5496 & 0,3312 \\ 0,6754 & 0,7901 \end{pmatrix}$$

et caetera.

3 Les modèles VAR avec Octave

Dans cette section nous montrons comment utiliser Octave pour étudier des modèles VAR. Avant de montrer comment utiliser Octave pour générer des séries temporelles artificielles, nous utilisons ce langage matriciel pour calculer les matrices de la forme $MA(\infty)$ puis les moments asymptotiques d'ordre un et deux associés à un modèle VAR.

3.1 Représentation $MA(\infty)$ d'un VAR(2)

Nous reprenons l'exemple 2.2 et ses suites. Le plus simple est d'utiliser Octave en écrivant des scripts contenant une série de définitions et instructions. Pour définir les deux matrices autorégressives du VAR(2) :

```
A1 = [ 0.5 0.1 ; 0.4 0.5];
A2 = zeros(2,2);
A2(2,1) = 0.25;
```

En comparant ces trois lignes avec la définition du VAR(2) dans l'exemple 2.2 et, éventuellement, en vous documentant sur Octave/Matlab, vous devriez comprendre comment ces trois lignes permettent de créer les matrices A_1 et A_2 . Retenez les points suivants :

- Les crochets servent à définir une matrice. chaque point-virgule entre deux crochets termine une ligne de la matrice.
- La fonction `zeros(m,n)` permet de créer une matrice $m \times n$ pleine de zéros.
- Pour sélectionner l'élément à l'intersection de la ligne i et de la colonne j d'une matrice M il faut utiliser des parenthèses : `M(i,j)`.

Pour vérifier que nous avons bien défini les deux matrices nous pouvons demander à Octave de les afficher à l'écran en ajoutant les deux lignes suivantes dans le script :

```
A1
A2
```

Notez que ces deux lignes ne se terminent pas par un point-virgule, contrairement aux lignes précédentes. L'absence de cette ponctuation indique à Octave qu'il doit afficher ce que renvoi l'instruction sur une ligne.

Nous avons vu plus haut comment calculer récursivement les matrices correspondant à la représentation $MA(\infty)$. En pratique, il suffit de rajouter les lignes suivantes dans le script Octave :

```
% Compute Phi0, Phi1, Phi2 and Phi3 (first matrices
% of the MA(\infty) representation).
Phi0 = eye(2);
Phi1 = A1;
Phi2 = Phi1*A1 + Phi0*A2;
Phi3 = Phi2*A1 + Phi1*A2;
```

Les deux premières lignes sont des commentaires et ne sont pas « lues » par Octave. Les quatre lignes suivantes calculent les matrices $\Phi_0, \Phi_1, \Phi_2, \Phi_3$. Pour Φ_0 , il n'y a rien à calculer puisqu'il s'agit de la matrice identité. L'instruction `eye(m)` construit une matrice identité de dimension m . Le problème avec ce code est que si nous souhaitons calculer Φ_{50} nous devons écrire une cinquantaine de lignes de codes dans le script. Pour nous éviter cette peine nous pouvons utiliser une boucle :

```
tmp2 = eye(2); % Phi0
tmp1 = A1; % Phi1
for i = 2:50
    Phi = tmp1*A1 + tmp2*A2;
    tmp2 = tmp1;
    tmp1 = Phi;
end
```

À la sortie de la boucle, `Phi` contient la valeur de Φ_{50} . Cette boucle peut servir à calculer la variance à la volée. Pour une matrice de variance covariance $SIGMA_u$ donnée :

```

tmp2 = eye(2); % Phi0
tmp1 = A1; % Phi1
Gamma0 = SIGMAu + tmp1*SIGMAu*tmp1';
for i = 2:100
    Phi = tmp1*A1 + tmp2*A2;
    Gamma0 = Gamma0 + Phi*SIGMAu*Phi';
    tmp2 = tmp1;
    tmp1 = Phi;
end

```

Si 100 est une bonne approximation de ∞ pour le problème posé, alors à la sortie de la boucle Gamma0 contient (une bonne approximation de) la variance. Bien sûr nous ne savons pas à l'avance si 100 itérations sont suffisantes. Pour palier ce problème nous pouvons remplacer la boucle `for` par une boucle conditionnelle `while` de la façon suivante :

```

tmp2 = eye(2); % Phi0
tmp1 = A1; % Phi1
Gamma0 = SIGMAu + tmp1*SIGMAu*tmp1';
condition = 1;
while condition
    Phi = tmp1*A1 + tmp2*A2;
    Gamma0 = Gamma0 + Phi*SIGMAu*Phi';
    tmp2 = tmp1;
    tmp1 = Phi;
    condition = (max(max(Phi)) > 1e-6);
end

```

Cette boucle s'arrête dès que `condition` est égal à 0. L'instruction `(max(max(Phi)) > 1e-6)` renvoie 1 si le plus grand élément de Φ est supérieur à 10^{-6} , 0 sinon. Ainsi, la boucle se termine dès que Φ est arbitrairement proche de zéro.

Le problème (éventuel) avec ces boucles est que nous ne gardons pas en mémoire les matrices de la représentation $MA(\infty)$. Si nous souhaitons travailler avec ces matrices après la sortie de la boucle, nous pouvons les stocker dans un tableau en trois dimensions :

```

PHI = zeros(2,2,101);
PHI(:,:,1) = eye(2); % Phi0
PHI(:,:,2) = A1; % Phi1
for i = 2:100
    PHI(:,:,i+1) = PHI(:,:,i)*A1 + PHI(:,:,i-1)*A2;
end

```

L'instruction `PHI = zeros(2,2,101);` initialise un tableau 3D avec 2 lignes, 2 colonnes et 101 plans (dans la profondeur). On peut voir ce tableau comme une superposition de 101 matrices 2×2 . Les 101 matrices $\Phi_0, \Phi_1, \dots, \Phi_{100}$ sont stockés dans ce tableau. Le premier plan contient Φ_0 , le deuxième contient Φ_1 , et *caetera*.

3.2 Moments asymptotiques d'un VAR(1)

Nous reprenons le modèle VAR défini dans l'exemple 2.1. Le script suivant calcule l'espérance du vecteur de variables endogènes :

```
ConstantVector = [.1 ; .2 ; .3];  
AutoregressiveMatrix = [.5 0 0 ; .1 .1 .3 ; 0 .2 .3];  
Mu = inv(eye(3)-AutoregressiveMatrix)*ConstantVector;
```

L'instruction `inv(X)` inverse une matrice carrée X de plein rang. Pour calculer la variance de y_t , en supposant donnée la variance Σ_{Au} des innovations, on peut rajouter les lignes suivantes au script :

```
vSIGMAy = inv(eye(9)-kron(AutoregressiveMatrix,AutoregressiveMatrix))*SIGMAu(:);  
SIGMAy = reshape(vSIGMAy,3,3);
```

L'instruction `X(:)`, où X est une matrice, calcule $\text{vec } X$. La première ligne utilise le produit de Kronecker pour calculer $\text{vec } \mathbb{V}[y_t]$. Le résultat est un vecteur 9×1 . La dernière ligne transforme le vecteur $\text{vec } \mathbb{V}[y_t]$ en une matrice 3×3 , $\mathbb{V}[y_t]$.

3.3 Simulation d'un modèle VAR

On commence en écrivant une fonction qui permet de simuler un modèle VAR(1). Le code octave est enregistré dans un fichier nommé `simul_var_1.m`. Notons que cette fonction manque considérablement de généralité. Elle ne nous permet pas de changer la condition initiale, elle impose une matrice de variance-covariance diagonale des innovations et ne permet pas de considérer le cas où les variances des K innovations sont différentes. Pour simuler le VAR, il ne nous reste plus qu'à écrire un *script* qui appelle cette fonction et produit des graphiques. Voir les encadrés 1 et 2.

La fonction `simul_var.m` permet de simuler un modèle VAR(p) avec ou sans constante. On généralise le code en autorisant une matrice de variance covariance des innovations non diagonale. Vous pouvez essayer de généraliser le code dans d'autres directions (par exemple en considérant des conditions initiales quelconques). Voir les encadrés 3 et 4.

Pour simuler un modèle VAR avec une matrice de variance-covariance arbitraire pour les innovations, on la décompose de Cholesky. Pour toute matrice définie positive (les valeurs propres sont strictement positives) X il existe une matrice A triangulaire inférieure (ie les éléments de A au dessus de la diagonale sont tous nuls) telle que $X = AA'$. La matrice A est la matrice de Cholesky associée à la matrice X . Avec octave on obtient la matrice A à l'aide de la fonction `chol`. Si ε est un vecteur $K \times 1$ gaussien d'espérance 0 et de variance I_K , alors $u = A\varepsilon$, où A est la matrice de Cholesky associée à Σ_u , est un vecteur gaussien de dimension $K \times 1$ d'espérance nulle et de variance $AI_KA' = \Sigma_u$. Ainsi pour simuler des innovations multivariées de variance Σ_u il suffit de simuler des vecteurs gaussiens de variance unitaire et de pré-multiplier ces vecteurs par la matrice de Cholesky associée à Σ_u .

4 Prévisions avec un modèle VAR

4.1 Prédiction ponctuelle

À une période t le statisticien doit annoncer le futur d'un ensemble de variables y_1, \dots, y_K . Pour cela, il dispose d'un modèle générateur des données, on utilisera l'acronyme DGP (*Data Generating Process* en anglais) pour désigner ce modèle, et d'un ensemble d'information à la date t noté Ω_t .

On supposera ici que le DGP est un modèle VAR(p) et que l'ensemble d'information contient le présent et le passé des variables endogènes⁴ : $\Omega_t = \{y_s | s \leq t\}$. La période t où le statisticien produit sa prévision est l'*origine de la prévision*. Le nombre de périodes pour lesquelles le statisticien produit des prévisions est l'*horizon de prévision*.

Supposons un instant que le DGP soit un modèle VAR(1) et qu'à la date t nous souhaitons produire une prévision pour y_{t+1} . Nous savons que :

$$y_{t+1} = Ay_t + u_t$$

où $y_t \in \Omega_t$. Supposons que l'innovation est un bruit blanc gaussien d'espérance nulle et de variance Σ_u (une matrice $K \times K$ symétrique définie positive). Sous ces deux hypothèses nous savons donc que y_{t+1} conditionnellement à l'ensemble d'information Ω_t est une variable aléatoire gaussienne d'espérance Ay_t et de variance Σ_u . Le statisticien pourrait publier sa prévision en annonçant la densité conditionnelle de y_{t+1} . Mais il a aussi le soucis d'être compris par ses clients, qui ignorent les statistiques, et préfère donc communiquer un *point* dans la distribution conditionnelle. Par exemple, il préférerait dire « *le taux de croissance du PIB sera de 2% demain* », plutôt que « *la distribution du taux de croissance du PIB demain conditionnelle au taux de croissance du PIB demain est gaussienne d'espérance ay_t et de variance σ^2* ».

Choisir un point dans la distribution de $y_{t+1} | \Omega_t$ c'est produire une prédiction ponctuelle. Tout le problème du statisticien est de résumer une distribution par un point. Ce choix est évidemment arbitraire. Afin de réduire l'arbitraire à l'essentiel, et de rendre ce choix le plus transparent possible, on définit une *fonction de perte* qui associe à chaque prévision une perte lorsque la prévision est fautive. Le statisticien choisit alors la prévision ponctuelle qui minimise la *perte espérée*. On suppose souvent que la fonction de perte est quadratique. Par exemple si on note $y_t(1)$ la prévision à l'horizon 1 à la date t , on a :

$$y_t(1) = \arg \min_x \frac{1}{2} \int (x - y)' \Omega (x - y) f_Y(y) dy$$

où Ω est une matrice définie positive quelconque, $f_Y(y)$ la densité associée au vecteur y . Le terme $P(x, y) \equiv \frac{1}{2} (x - y)' \Omega (x - y)$ est la perte associée à une prévision x si la vraie valeur demain est y , il s'agit d'une fonction de perte quadratique. La matrice arbitraire Ω sert à pondérer les pertes sur les K variables du VAR. On note $\mathcal{P}(x)$ la perte espérée. Nous avons :

$$\frac{d}{dx} \mathcal{P}(x) = \int \Omega (x - y) f_Y(y) dy$$

⁴Il n'est pas nécessaire que l'ensemble d'information contienne tout le passé depuis l'origine du monde. Si le modèle considéré est un VAR(p), le passé est résumé par les p réalisations précédentes du processus stochastique multivarié.

En égalisant à zéro il vient :

$$\Omega x \int f_Y(y) dy = \Omega \int y f_Y(y) dy$$

Finalement, en pré-multipliant par Ω^{-1} et notant que $\int f_Y(y) dy = 1$ par définition d'une densité de probabilité, nous avons donc :

$$y_t(1) = \int y f_Y(y) dy$$

La prévision ponctuelle, qui minimise la perte espérée, est l'espérance de la distribution de y en $t + 1$ conditionnellement à Ω_t . Ce résultat vaut pour tout modèle dès lors que la fonction de perte est quadratique. Si nous avons choisi une autre fonction de perte nous aurions trouvé une autre prévision ponctuelle que l'espérance.

Proposition 4.1. *L'espérance conditionnelle, $y_t(h) \equiv \mathbb{E}[y_{t+h}|\Omega_t]$, est le prédicteur ponctuel qui minimise l'erreur quadratique moyenne (MSE).*

Soit $\tilde{y}_t(h) \neq y_t(h)$ un prédicteur ponctuel quelconque. L'erreur quadratique moyenne associée à ce prédicteur ponctuel est :

$$\text{MSE}(\tilde{y}_t(h)) = \mathbb{E}[(y_{t+h} - \tilde{y}_t(h))(y_{t+h} - \tilde{y}_t(h))']$$

La proposition nous dit simplement que :

$$\text{MSE}(\tilde{y}_t(h)) \geq \text{MSE}(\mathbb{E}[y_{t+h}|\Omega_t])$$

c'est-à-dire que la matrice $\text{MSE}(\tilde{y}_t(h)) - \text{MSE}(\mathbb{E}[y_{t+h}|\Omega_t])$ est semi-définie positive, ou, de façon équivalente, que pour tout vecteur c ($K \times 1$) nous avons :

$$\text{MSE}(c'\tilde{y}_t(h)) \geq \text{MSE}(c'\mathbb{E}[y_{t+h}|\Omega_t])$$

Preuve 4.1. Soit $\tilde{y}_t(h) \neq y_t(h)$ un prédicteur ponctuel quelconque. Par définition, nous avons :

$$\begin{aligned} \text{MSE}(\tilde{y}_t(h)) &= \mathbb{E}[(y_{t+h} - \tilde{y}_t(h))(y_{t+h} - \tilde{y}_t(h))'] \\ &= \mathbb{E}[(y_{t+h} - \mathbb{E}[y_{t+h}|\Omega_t] + \mathbb{E}[y_{t+h}|\Omega_t] - \tilde{y}_t(h))(y_{t+h} - \mathbb{E}[y_{t+h}|\Omega_t] + \mathbb{E}[y_{t+h}|\Omega_t] - \tilde{y}_t(h))'] \\ &= \text{MSE}(\mathbb{E}[y_{t+h}|\Omega_t]) + \mathbb{E}[(\mathbb{E}[y_{t+h}|\Omega_t] - \tilde{y}_t(h))(\mathbb{E}[y_{t+h}|\Omega_t] - \tilde{y}_t(h))'] \end{aligned}$$

où nous avons utilisé la nullité de $\mathbb{E}[(y_{t+h} - \mathbb{E}[y_{t+h}|\Omega_t])(\mathbb{E}[y_{t+h}|\Omega_t] - \tilde{y}_t(h))']$, car $y_{t+h} - \mathbb{E}[y_{t+h}|\Omega_t]$ dépend des innovations entre t et $t+h$ qui sont orthogonales aux termes dans $\mathbb{E}[y_{t+h}|\Omega_t] - \tilde{y}_t(h)$ (qui sont fonction de y_s avec $s < t$). Puisque le dernier terme sur le membre de droite est une matrice définie positive, on a nécessairement $\text{MSE}(\tilde{y}_t(h)) \geq \text{MSE}(\mathbb{E}[y_{t+h}|\Omega_t])$ \square

L'optimalité de l'espérance conditionnelle, dans le cas du modèle VAR (2.1), implique que :

$$y_t(h) = \nu + A_1 y_t(h-1) + \dots + A_p y_t(h-p) \quad (4.1)$$

est le prédicteur optimal à l'horizon h , si l'innovation u_t est un bruit blanc d'espérance nulle (u_t et u_s sont indépendants pour $s \neq t$) de sorte que $\mathbb{E}[u_{t+h}|\Omega_t] = 0$ pour tout $h > 0$. Nous pouvons donc obtenir les prévisions ponctuelles de façon récursive :

$$y_t(1) = \nu + A_1 y_t + \dots + A_p y_{t-p}$$

$$y_t(2) = \nu + A_1 y_t(1) + a_2 y_{t-2} + \dots + A_p y_{t-p+1}$$

Dans le cas d'un modèle VAR(1) c'est encore plus simple, nous avons :

$$y_t(h) = (I_K + A_1 + \dots + A_1^{h-1}) \nu + A_1^h y_t$$

Remarque 4.1. L'hypothèse importante pour établir la récurrence qui nous permet d'obtenir des prévisions ponctuelles est l'indépendance des innovations. Si les innovations ne sont pas indépendantes, nous n'avons pas $\mathbb{E}[u_{t+h}|\Omega_t] = 0$. Par exemple, dans le cas univarié, supposons que nous ayons :

$$u_t = \begin{cases} e_t & \text{si } t \text{ est pair,} \\ \frac{e_{t-1}^2 - 1}{\sqrt{2}} & \text{sinon.} \end{cases}$$

où $e_t \sim \mathcal{N}(0, 1)$. On vérifie facilement que l'espérance inconditionnelle de u_t est nulle, que la variance de u_t est finie et constante, que u_t et u_s sont non corrélés pour tout $s \neq t$ (ie $\mathbb{E}[u_t u_s] = 0 \quad \forall s \neq t$). Pour cela il suffit de se rappeler que les moments d'ordre impair d'une va gaussienne centrée réduite sont nulles et que les moments pairs sont donnés par la récurrence : $m_{2k} = (2k - 1)m_{2(k-1)}$, en particulier le moment d'ordre 4 est $m_4 = 3$. Pourtant les innovations u ne sont pas indépendantes. Ainsi $\mathbb{E}[u_{t+1}|\Omega_t] \neq 0$, en effet, si t est pair (et donc $t + 1$ impair) :

$$\begin{aligned} E[u_{t+1}|\Omega_t] &= E\left[\frac{e_t^2 - 1}{\sqrt{2}} \middle| \Omega_t\right] \\ &= \frac{e_t^2 - 1}{\sqrt{2}} \end{aligned}$$

car $e_t \in \Omega_t$.

4.2 Des fourchettes pour les prévisions

Le statisticien souhaitera sûrement donner une indication de la précision de sa prévision. L'idée est de construire des intervalles de prévision. Pour cela nous allons supposer que les innovations sont gaussiennes, de telle sorte, puisque le modèle VAR est linéaire, que les y_t sont eux aussi normalement distribués. Notons aussi que dans ce cas l'indépendance des innovations est équivalente à l'absence de corrélation des innovations. Sous cette condition les erreurs de prévisions sont elles aussi gaussiennes :

$$y_{t+h} - y_t(h) = \sum_{i=0}^{h-1} \Phi_i u_{t+h-i} \sim \mathcal{N}(0, \Sigma_y(h))$$

Notons que la variance de l'erreur de prévision à l'horizon h tend vers la variance asymptotique de y lorsque h tend vers l'infini. Nous savons marginaliser un vecteur aléatoire gaussien, et savons donc que l'erreur de prévision sur une des K variables est elle aussi gaussienne :

$$\frac{y_{k,t+h} - y_{k,t}(h)}{\sigma_k(h)} \sim \mathcal{N}(0, 1)$$

où $\sigma_k(h)$ est la racine carrée du k -ième élément de la diagonale de $\Sigma_y(h)$, la matrice de variance-covariance de l'erreur de prévision à l'horizon h .

Soit Z une variable aléatoire gaussienne centrée réduite, on note $f_Z(z)$ sa fonction de densité. On définit $t(\alpha)$

un réel tel que $\int_{t(\alpha)}^{\infty} f_Z(z)dz = \alpha$ (avec $\alpha \in [0, 1]$). On a :

$$\begin{aligned} 1 - \alpha &= \mathbb{P} \left(-t(\alpha/2) \leq \frac{y_{k,t+h} - y_{k,t}(h)}{\sigma_k(h)} \leq t(\alpha/2) \right) \\ &= \mathbb{P} (y_{k,t}(h) - t(\alpha/2)\sigma_k(h) \leq y_{k,t+h} \leq y_{k,t}(h) + t(\alpha/2)\sigma_k(h)) \end{aligned}$$

Ainsi un intervalle de prévision à $(1 - \alpha)100\%$ à l'horizon h pour la variable k est donné par :

$$y_{k,t}(h) \pm t(\alpha/2)\sigma_k(h)$$

Ici nous proposons un intervalle pour la prévision d'une unique variable. Nous pourrions construire un « intervalle » pour la prévision jointe d'un ensemble de variables...

5 Analyse structurelle avec un VAR

Nous nous intéressons ici à l'interprétation des modèles VAR.

5.1 Causalité

5.1.1 Définitions

Notons Ω_t l'ensemble contenant toute l'information pertinente de l'univers à la date t sur une variable ou un ensemble de variables. On note $z_t(h|\Omega_t)$ la meilleure prédiction – par exemple au sens de la minimisation de l'erreur quadratique moyenne – à la date t de z_{t+h} sachant l'ensemble d'information Ω_t . Cet ensemble contient en particulier le passé et le présent de z_t . Enfin, on note $\Sigma_z(h|\Omega_t)$ l'erreur quadratique moyenne de la prévision $z_t(h|\Omega_t)$.

Définition 5.1. Le processus $\{x_t\}$ cause le processus $\{z_t\}$ au sens de Granger si :

$$\Sigma_z(h|\Omega_t) < \Sigma_z(h|\Omega_t \setminus \{x_s | s \leq t\})$$

L'idée derrière cette définition est relativement simple : une cause ne peut apparaître avant son effet ! Ainsi si une variable x cause une variable z , la première doit aider à former des prévisions sur la seconde. Si z_t peut être prédit plus efficacement lorsque x_t fait partie de l'ensemble d'information, on dit que x cause z au sens de Granger.

Cette définition se généralise dans le cas multivarié où z et x sont éventuellement des vecteurs (les deux MSE doivent être différentes et la différence $\Sigma_z(h|\Omega_t) - \Sigma_z(h|\Omega_t \setminus \{x_s | s \leq t\})$ doit être semi définie positive). Enfin si x cause z au sens de Granger et si z cause x au sens de Granger, alors on dit que nous sommes en présence d'un *feedback system* (en anglais).

Définition 5.2. Il y a causalité instantanée entre $\{x_t\}$ et $\{z_t\}$ si :

$$\Sigma_z(1|\Omega_t \cup \{x_{t+1}\}) \neq \Sigma_z(1|\Omega_t)$$

Autrement dit, à la date t , rajouter x_{t+1} dans l'ensemble d'information aide à prévoir z_{t+1} . Nous allons voir dans la section suivante que ce concept de causalité est symétrique (dans le sens où si x cause instantanément z alors z cause instantanément x).

Notons que ces définitions de la causalité reposent sur l'erreur quadratique moyenne, ce qui est un choix arbitraire. Une autre mesure (de la qualité des prévisions) pourrait aboutir à une autre caractérisation de la causalité.

En pratique Ω_t ne peut couvrir toute l'information pertinente disponible dans l'univers (non accessible au statisticien). Ainsi on limitera l'ensemble Ω_t aux valeurs présentes et passées des processus étudiés : $\Omega_t \equiv \{z_s, x_s | s \leq t\}$

5.1.2 Causalité au sens de Granger dans un VAR

Soit un modèle VAR(p) y_t $K \times 1$ dont la représentation MA(∞) est donnée par :

$$y_t = \mu + \sum_{i=0}^{\infty} \phi_i u_{t-i} = \mu + \Phi(L)u_t \quad (5.1)$$

où u_t est un bruit blanc d'espérance nulle et de variance Σ_u , une matrice symétrique définie positive, et $\phi_0 = I_K$. On suppose que le vecteur y_t peut être partitionné sous la forme $y_t \equiv (z_t' | x_t')'$ où z_t est un vecteur $M \times 1$ et x_t un vecteur $(K - M) \times 1$. La représentation MA(∞) peut être partitionnée conformément à cette définition de y_t :

$$y_t = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \Phi_{11}(L) & \Phi_{12}(L) \\ \Phi_{21}(L) & \Phi_{22}(L) \end{pmatrix} \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix} \quad (5.2)$$

Ainsi la prévision optimale à l'horizon 1 pour z sachant $\{y_s | s \leq t\}$ est :

$$\begin{aligned} z_t(1 | s \leq t) &= (I_M \quad 0) y_t(1) \\ &= \mu_1 + \sum_{i=0}^{\infty} \phi_{11,i} u_{1,t-i} + \sum_{i=0}^{\infty} \phi_{12,i} u_{2,t-i} \end{aligned} \quad (5.3)$$

Et donc l'erreur de prévision est :

$$z_{t+1} - z_t(1 | s \leq t) = u_{1,t+1} \quad (5.4)$$

car la matrice ϕ_0 étant égale à la matrice identité de dimension K , la sous matrice $\phi_{12,0}$ est nulle et $u_{2,t+1}$ n'affecte pas z_{t+1} . Par ailleurs on peut montrer, en utilisant le théorème de décomposition de Wold 2.1, qu'un sous ensemble d'un processus stationnaire admet aussi une représentation MA(∞). Ainsi,

$$\begin{aligned} z_t &= \mu_1 + \sum_{i=0}^{\infty} \phi_{11,i} u_{1,t-i} + \sum_{i=0}^{\infty} \phi_{12,i} u_{2,t-i} \\ &= \mu_1 + \sum_{i=0}^{\infty} F_i v_{t-i} \end{aligned}$$

où $F_0 = I_M$. La prévision à l'horizon 1 conditionnelle à l'ensemble d'information réduit $\{z_s | s \leq t\}$ est donc :

$$z_t(1 | \{z_s | s \leq t\}) = \mu_1 + \sum_{i=1}^{\infty} F_i v_{t+1-i} \quad (5.5)$$

et l'erreur de prévision est :

$$z_{t+1} - z_t (1|\{z_s|s \leq t\}) = v_{t+1} \quad (5.6)$$

Les erreurs de prévisions (5.4) et (5.6), qui ne diffèrent que par l'ensemble d'information, sont identiques si et seulement si $v_t = u_{1,t}$ pour tout t . Dit autrement, les deux prédicteurs sont identiques si et seulement si la représentation MA de $\{z_t\}$ vérifie :

$$\begin{aligned} z_t &= \mu_1 + \sum_{i=0}^{\infty} F_i u_{1,t-i} = \sum_{i=0}^{\infty} (F_i \quad 0) u_{t-i} \\ &= \mu_1 + \sum_{i=0}^{\infty} (\phi_{11,i} \quad \phi_{12,i}) u_{t-i} \end{aligned}$$

L'unicité de la représentation MA implique que nous devons avoir $F_i = \phi_{11,i}$ et $\phi_{12,i} = 0$ pour tout i . D'où la proposition suivante :

Proposition 5.1 (Non causalité au sens de Granger et représentation MA). *Soit $\{y_t\}$ un modèle VAR stationnaire, défini comme dans les équations 5.1 et , dont la représentation MA est définie par le polynôme retard $\Phi(L)$. Alors,*

$$z_t (1|\{y_s|s \leq t\}) = z_t (1|\{z_s|s \leq t\})$$

est équivalent à

$$\Phi_{12}(L) = 0$$

Remarque 5.1. *Nous avons défini la non causalité au sens de Granger à partir de la représentation MA(∞) d'un processus. Celle-ci n'est pas forcément associée à un modèle VAR fini. La définition présentée ici est donc relativement générale.*

Pour tester la non causalité au sens de Granger il faut vérifier la nullité des matrices $\phi_{12,i}$. Ce qui n'est pas forcément très simple... D'autant moins que généralement nous travaillons plutôt sur un modèle VAR(p). Soit le modèle VAR(p) suivant :

$$y_t \equiv \begin{pmatrix} z_t \\ x_t \end{pmatrix} = \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix} + \sum_{i=1}^p \begin{pmatrix} A_{11,i} & A_{12,i} \\ A_{21,i} & A_{22,i} \end{pmatrix} \begin{pmatrix} z_{t-i} \\ x_{t-i} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix} \quad (5.7)$$

La condition donnée dans la proposition précédente est équivalente ici à $A_{12,i} = 0$ pour tout i . Pour voir cela, il suffit de remarquer, en utilisant une formule d'inversion matricielle par bloc :

$$\begin{pmatrix} \Phi_{11}(L) & \Phi_{12}(L) \\ \Phi_{21}(L) & \Phi_{22}(L) \end{pmatrix}^{-1} = \begin{pmatrix} \Phi_{11}(L)^{-1} & 0 \\ -\Phi_{22}(L)^{-1}\Phi_{21}(L)\Phi_{11}(L)^{-1} & \Phi_{22}(L)^{-1} \end{pmatrix}$$

Proposition 5.2 (Non causalité au sens de Granger dans un VAR). *Soit $\{y_t\}$ un modèle VAR(p) stationnaire défini par 5.7. le processus $\{x_t\}$ ne cause pas $\{z_t\}$ au sens de Granger si et seulement si $\phi_{12,i} = 0$ pour $i = 1, \dots, p$.*

Exemple 5.1. *Soit le processus VAR(1) suivant :*

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{pmatrix} = \nu + \begin{pmatrix} 0,5 & 0 & 0 \\ 0,1 & 0,1 & 0,3 \\ 0 & 0,2 & 0,3 \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{pmatrix} + u_t$$

Posons $x_t \equiv (y_{2,t}, y_{3,t})$ et $z_t \equiv y_{1,t}$. On voit immédiatement que x ne cause pas z au sens de Granger. Par contre, z cause x au sens de Granger. Pour rendre la discussion plus concrète, supposons que les trois variables soient dans l'ordre les variations de l'investissement, du PIB et de la conso. Ce VAR nous dit que la variation de l'investissement cause au sens de Granger les variations de la conso et du PIB, mais que la réciproque n'est pas vraie. On vérifie tout aussi facilement que la consommation cause au sens de Granger le système investissement/PIB et réciproquement. Il est important de noter que nous n'avons défini la causalité au sens de Granger que d'un sous ensemble de variables à un autre, et non pas d'une variable à une autre.

5.1.3 Causalité instantannée

Soit Σ_u la matrice de variance covariance de l'innovation d'un modèle MA(∞), cette matrice est symétrique et surtout définie positive, il est donc possible de la décomposer sous la forme $\Sigma_u = PP'$ où P est une matrice triangulaire inférieure dont la diagonale est composée de scalaires positifs. On peut alors écrire la représentation MA de la façon suivante :

$$y_t = \mu \sum_{i=0}^{\infty} \phi_i P P^{-1} u_{t-i} = \mu + \sum_{i=0}^{\infty} \Theta_i w_{t-i} \quad (5.8)$$

où $\Theta_i = \phi_i P$ et $w_t = P^{-1} u_t$ est un bruit blanc dont la matrice de variance covariance est :

$$\Sigma_w = P^{-1} \Sigma_u (P^{-1})' = I_K$$

Comme la matrice Σ_w est diagonale (les composantes de w sont non corrélées), on dit que w_t est une innovation orthogonale. En partitionnant y_t comme dans la section précédente, on obtient :

$$\begin{pmatrix} z_t \\ x_t \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \Theta_{11,0} & 0 \\ \Theta_{21,0} & \Theta_{22,0} \end{pmatrix} \begin{pmatrix} w_{1,t} \\ w_{2,t} \end{pmatrix} + \begin{pmatrix} \Theta_{11,1} & \Theta_{12,1} \\ \Theta_{21,1} & \Theta_{22,1} \end{pmatrix} \begin{pmatrix} w_{1,t-1} \\ w_{2,t-1} \end{pmatrix} + \dots$$

Nous avons donc :

$$z_{t+1} = \mu_1 + \Theta_{11,0} w_{1,t+1} + \Theta_{11,1} w_{1,t} + \Theta_{12,1} w_{2,t} + \dots$$

et

$$x_{t+1} = \mu_2 + \Theta_{21,0} w_{1,t+1} + \Theta_{22,0} w_{2,t+1} + \Theta_{21,1} w_{1,t} + \Theta_{22,1} w_{2,t} + \dots$$

Le prédicteur optimal à l'horizon 1 de x_t basé sur l'ensemble d'information $\{y_s | s \leq t\}$ et sur z_{t+1} est égal au prédicteur optimal de x_t basé sur $\{w_s | s \leq t\} \cup \{w_{t+1}\}$:

$$\begin{aligned} x_t (1 | \{y_s | s \leq t\} \cup \{z_{t+1}\}) &= x_t (1 | \{w_s | s \leq t\} \cup \{w_{1,t+1}\}) \\ &= \Theta_{21,0} w_{1,t+1} + x_t (1 | \{y_s | s \leq t\}) \end{aligned}$$

Ainsi nous avons :

$$x_t (1 | \{y_s | s \leq t\} \cup \{z_{t+1}\}) = x_t (1 | \{y_s | s \leq t\})$$

si et seulement si la matrice $\Theta_{21,i}$ est nulle. Cette condition est vérifiée si et seulement si la matrice de variance covariance Σ_u est bloc diagonal (un bloc pour $u_{1,t}$ et un autre pour $u_{2,t}$) :

$$\Sigma_u = \begin{pmatrix} \Sigma_{u_1} & 0 \\ 0 & \Sigma_{u_2} \end{pmatrix}$$

Cette structure particulière de la matrice de variance-covariance nous dit simplement que les innovations $u_{1,t}$ et $u_{2,t}$ sont non corrélées.

Proposition 5.3 (Absence de causalité instantanée). *Soit le processus stochastique multivarié stationnaire $\{y_t\}$ défini par 5.8. Alors il n'y a pas de causalité instantanée entre $\{x_t\}$ et $\{z_t\}$ si et seulement si $\mathbb{E}[u_{1,t}u'_{2,t}] = 0$.*

Comme cette proposition pose une condition sur la matrice de variance covariance de l'innovation, Σ_u , elle peut aussi bien être utilisée dans le cadre d'un modèle VAR(p) ou d'un modèle MA.

Exemple 5.2. *En poursuivant l'exemple précédant, si on a la matrice de variance-covariance :*

$$\Sigma_U = \begin{pmatrix} 2,25 & 0 & 0 \\ 0 & 1,0 & 0,5 \\ 0 & 0,5 & 0,74 \end{pmatrix}$$

alors il n'y a pas de causalité instantanée entre $(y_{2,t}, y_{3,t})$ et $y_{1,t}$.

5.1.4 Remarques

Il faut utiliser ces notions de causalité avec beaucoup de précaution... Ces définitions ne correspondent pas forcément à l'idée que nous pouvons avoir, par ailleurs de la causalité.

Finalement, la causalité instantanée n'est qu'une corrélation entre des variables. Les représentations VAR ou MA sont silencieuses sur le sens de la causalité. Pour dévoiler le sens de la causalité il faut utiliser des théories économiques.

La causalité au sens de Granger, si elle semble donner plus d'information sur le sens de la causalité, n'est pas exempte de problèmes. Un exemple simple montre que l'interprétation de l'absence de causalité au sens de Granger est difficile.

Exemple 5.3. *Soit le modèle VAR(1) bivarié suivant :*

$$\begin{pmatrix} z_t \\ x_t \end{pmatrix} = \begin{pmatrix} \alpha_{11} & 0 \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \begin{pmatrix} z_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix}$$

On voit immédiatement que x_t ne cause pas z_t au sens de Granger. Néanmoins, sans rien changer aux propriétés du processus stochastique, on peut toujours multiplier les deux membres de cette équations par une matrice non singulière de la forme :

$$B = \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix}$$

Ainsi il vient :

$$\begin{pmatrix} z_t \\ x_t \end{pmatrix} = \begin{pmatrix} 0 & -\beta \\ 0 & 0 \end{pmatrix} \begin{pmatrix} z_t \\ x_t \end{pmatrix} + \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \begin{pmatrix} z_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} v_{1,t} \\ v_{2,t} \end{pmatrix}$$

avec $\gamma_{11} = \alpha_{11} + \alpha_{21}\beta$, $\gamma_{12} = \alpha_{22}\beta$, $\gamma_{21} = \alpha_{21}$ et $\gamma_{22} = \alpha_{22}$. Il s'agit d'une représentation alternative du même processus. le lecteur est invité à vérifier que nous n'avons pas changer l'espérance et les moments d'ordre 2 du processus bivarié. Malheureusement ces deux représentations équivalentes ont des interprétations très différentes. Maintenant nous

voyons que x_t cause z_t au sens de Granger ! Ainsi, l'absence de causalité au sens de Granger, ne doit pas s'interpréter comme l'absence d'une relation de cause à effet.

La causalité au sens de Granger pose d'autres problèmes :

- (i) La causalité au sens de Granger peut dépendre de la fréquence des données. On peut n'observer aucune relation entre inflation et taux d'intérêt sur des données annuelles ou trimestrielles, mais percevoir une relation sur des données mensuelles.
- (ii) L'utilisation de données ajustée des variations saisonnières ou la présence d'erreurs de mesures peut affecter la causalité au sens de Granger.
- (iii) La causalité au sens de Granger dépend de la liste des variables dans le système. Nous avons supposé que toutes les variables pertinentes sont incluses... Cette hypothèse, peu crédible, est difficilement testable.
- (iv) Il est difficile de parler de la causalité entre deux variables plutôt qu'entre deux sous ensembles exclusifs de variables. Dans un VAR bivarié on identifie assez facilement la condition de non causalité au sens de Granger. Si $y_{2,t}$ ne permet pas d'améliorer la prévision à l'horizon 1 de $y_{1,t}$, alors cette variable n'aide pas plus à améliorer les prévisions à un horizon h quelconque de $y_{1,t}$. Ceci n'est plus vrai si le système comporte trois variables au lieu de deux. Dans ce cas il se peut que $y_{2,t}$ n'améliore pas les prévisions de $y_{1,t}$ à l'horizon 1 mais que $y_{2,t}$ améliore les prévisions à des horizons $h > 1$. Exhiber une condition de non causalité au sens de Granger dans ce cadre n'est pas une chose aussi triviale (restrictions non linéaires sur les matrices autorégressives du modèle VAR).

5.2 Fonctions de réponses

Nous venons de voir que la notion de causalité au sens de Granger ne peut suffire à caractériser les relations entre différentes variables. En pratique on veut connaître la réaction d'une variable à la suite d'une impulsion sur une autre variable, dans un système où il y a éventuellement plus de deux variables. L'idée est que si une variable réagit à une impulsion sur une autre variable, alors on a identifié une forme de relation de cause à effet.

Exemple 5.4. Dans un système bivarié décrivant la dynamique de l'inflation et du taux d'intérêt nominal, on aimerait identifier l'effet d'un choc exogène sur l'inflation (un choc pétrolier peut s'interpréter comme un choc exogène) ou l'effet d'un choc exogène sur le taux d'intérêt nominal (par exemple, une baisse surprise du taux décidée par le banquier central). Un choc est exogène dans le sens où la variation de la variable n'est pas liée aux autres variables présentes dans le système. Le dernier cas pose donc éventuellement problème, puisque le banquier central (en Europe) est censé maîtriser la volatilité des prix...

5.2.1 Réponses à des erreurs de prévisions

Supposons que nous intéressions à l'innovation sur la variable d'investissement dans un système de trois variables : l'investissement (y_1), le PIB (y_2) et la consommation (y_3). Supposons qu'avant l'instant initial ($t = 0$) les trois variables soient à leurs niveaux de long terme respectifs, ie $y_t = \mu$ pour tout $t < 0$. À l'instant initial, l'investissement subit un choc positif $u_{1,0} = 1$, mais pas les autres variables du système $u_{2,0} = u_{3,0} = 0$. On cherche la dynamique des variables du système suite à ce choc, si le système n'est plus perturbé par d'autres chocs par la suite ($u_t = 0$ pour tout $t > 0$). Pour simplifier nous allons supposer que $\mu = 0$, notre modèle est de la forme :

$$y_t = A_1 y_{t-1} + u_t$$

avec

$$A_1 = \begin{pmatrix} 0,5 & 0 & 0 \\ 0,1 & 0,1 & 0,3 \\ 0 & 0,2 & 0,3 \end{pmatrix}$$

$$y_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$u_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$u_t = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \text{ pour tout } t > 0$$

On a alors :

$$y_1 = A_1 y_0 = \begin{pmatrix} 0,5 \\ 0,1 \\ 0 \end{pmatrix}$$

$$y_2 = A_1 y_1 = A_1^2 y_0 = \begin{pmatrix} 0,25 \\ 0,06 \\ 0,02 \end{pmatrix}$$

et plus généralement $y_i = A_1^i y_0$ pour tout $i > 0$. Dans ce terme général nous reconnaissons la matrice de coefficient de la représentation $MA(\infty)$ associée au VAR : $\phi_i = A_1^i$. Ainsi, la réponse du vecteur y_t suite à un choc initial sur la première variable ($y_{1,t}$) est donnée par la première colonne des matrices de la représentation $MA(\infty)$. Plus généralement, la réponse du vecteur y_t suite à un choc initial sur la i -ième variable ($y_{i,t}$) est donnée par la i -ième colonne des matrices de la représentation $MA(\infty)$. On peut montrer que résultat général vaut aussi pour un modèle $VAR(p)$.

La réponse d'une variable j à un choc unitaire (erreur de prévision) sur une variable k est souvent représentée graphiquement.

<INSÉRER GRAPHIQUE ICI>

Toutes les réponses convergent (plus ou moins rapidement) vers zéro (l'espérance du processus considéré). Si les variables (l'innovation de chaque équation) ont des échelles différentes, il peut être pertinent de considérer des chocs initiaux d'un écart-type plutôt que des chocs initiaux unitaires. Cela revient à changer l'échelle des représentations graphiques mais pas la forme des réponses (quand un modèle est linéaire, la taille des chocs n'affecte pas la forme de la dynamique).

Remarque 5.2. Un choc sur une variable k n'a pas d'effet sur les autres variables si y_k ne cause pas les autres variables au sens de Granger. Dans ce cas les réponses sont nulles. Nous observons cette propriété dans le cas d'un choc sur le PIB ou la consommation qui n'affecte pas l'investissement. A contrario, si un choc sur une variable affecte une autre variable j , alors on ne peut pas dire que la variable k ne cause pas la variable j au sens de Granger. Ainsi pour vérifier qu'une

variable k ne cause pas une variable j , il suffit de vérifier que $\phi_{jk,i} = 0$ pour $i = 1, 2, \dots$. Il n'est pas nécessaire d'aller chercher toutes les matrices de la représentation MA (ie d'aller jusqu'à l'infini)... Comme l'établit la proposition suivante.

Proposition 5.4. Si y_t est un VAR(p) de dimension K alors

$$\phi_{jk,i} = 0 \quad \text{pour } i = 1, \dots, \infty$$

est équivalent à

$$\phi_{jk,i} = 0 \quad \text{pour } i = 1, \dots, p(K - 1)$$

Autrement dit, il suffit de montrer que $\phi_{jk,i} = 0$ pour $i = 1, \dots, p(K - 1)$ pour établir l'absence de causalité au sens de Granger entre la variable k et la variable j .

La construction de ces fonctions de réponses suppose que les innovations sont non corrélées. Si $y_{1,t}$ est corrélé avec les autres variables ($y_{2,t}$ et $y_{3,t}$) alors un choc sur la première devrait, en moyenne, avoir simultanément un effet sur les deux autres variables. Dans ce cas, en procédant comme nous le faisons ici, nous biaisons les résultats dans un sens a priori inconnu.

5.2.2 Réponses à des impulsions orthogonales

Pour palier le problème lié à une éventuelle corrélation des chocs, nous allons orthogonaliser ces chocs... En fait nous avons déjà rencontré cette transformation. Soit Σ_u la matrice de variance covariance des innovations u_t . Il s'agit d'une matrice symétrique et définie positive. Nous pouvons donc appliquer la décomposition de Cholesky ; il existe une matrice P triangulaire inférieure telle que $\Sigma_u = PP'$. En éliminant le terme lié à l'espérance du processus (qui n'est pas d'un grand intérêt si nous souhaitons seulement étudier la transmission des chocs dans le système), nous pouvons écrire sa représentation MA sous la forme :

$$y_t = \sum_{i=0}^{\infty} \phi_i PP^{-1} u_{t-i}$$

soit de façon équivalente :

$$y_t = \sum_{i=0}^{\infty} \Theta_i w_{t-i}$$

avec $\Theta_i = \phi_i P$ et $w_t = u_{t-i}$. On vérifie que la variance de w_t est la matrice identité de dimension K . La colonne k de la matrice Θ_i caractérise l'effet d'un choc sur la k -ième variable en t sur l'ensemble des variables du système en $t + i$.

Proposition 5.5. Si y_t est un VAR(p) de dimension K alors

$$\Theta_{jk,i} = 0 \quad \text{pour } i = 1, \dots, \infty$$

est équivalent à

$$\Theta_{jk,i} = 0 \quad \text{pour } i = 1, \dots, p(K - 1)$$

Calculer des IRFs (*Impulse Response Function*) pour des chocs orthogonaux est relativement simple. Il suffit de calculer les matrices de la représentation MA(∞) du modèle, puis de post multiplier ces matrices par P .

Tout le problème est que les résultats vont dépendre de l'ordre des variables dans le VAR. Changer l'ordre des variables dans le VAR (*ie* remplacer $y_t = (y_{1,t}, y_{2,t}, y_{3,t})'$ par $y_t = (y_{3,t}, y_{1,t}, y_{2,t})'$) peut considérablement affecter la forme des IRFs. Il faut donc toujours prendre soin de vérifier la robustesse des IRFs reportés à l'ordre des variables dans le VAR... Ou passer du temps pour justifier un ordre particulier... Notons qu'il existe d'autres façons d'orthogonaliser les innovations (on peut utiliser la racine carrée matricielle de matlab).

A Boîte à outils pour le calcul matriciel

A.1 La trace

Soit une matrice $m \times m$ carrée $A = (a_{ij})$. La trace de la matrice A est la somme des éléments sur sa diagonale.

Nous avons :

$$\text{tr } A = \sum_{i=1}^m a_{ii}$$

Soient A et B deux matrices $m \times m$ et $\lambda_1, \dots, \lambda_m$ les valeurs propres de A . C et D sont respectivement des matrices $m \times n$ et $n \times m$. On peut établir les propriétés suivantes :

1. $\text{tr } A + B = \text{tr } A + \text{tr } B$
2. $\text{tr } A' = \text{tr } A$
3. $\text{tr } CD = \text{tr } DC$
4. $\text{tr } A = \sum_{i=1}^m \lambda_i$

A.2 Le produit de Kronecker

Soient $A = (a_{ij})$ et $B = (b_{ij})$ deux matrices respectivement $m \times n$ et $p \times q$. La matrice $mp \times nq$:

$$A \otimes B = \begin{pmatrix} a_{11}B & \dots & a_{1m}B \\ \vdots & & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{pmatrix}$$

est le produit de Kronecker de la matrice A par la matrice B . Soient A, B, C et D des matrices avec les dimensions appropriées, on peut établir les propriétés suivantes :

1. $A \otimes B \neq B \otimes A$ en général
2. $(A \otimes B)' = A' \otimes B'$
3. $A \otimes (B + C) = A \otimes B + A \otimes C$
4. $(A \otimes B)(C \otimes D) = AC \otimes BD$
5. Si A et B sont des matrices inversibles, alors $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$
6. Si A et B sont des matrices carrées avec des valeurs propres respectives λ_A et λ_B et des vecteurs propres associés v_A et v_B , alors $\lambda_A \lambda_B$ est une valeur propre de $A \otimes B$ et $v_A \otimes v_B$ son vecteur propre associé.
7. Si A et B sont respectivement des matrices $m \times m$ et $n \times n$, alors $\det(A \otimes B) = \det(A)^n \det(B)^m$
8. Si A et B sont des matrices carrées, alors $\text{tr } A \otimes B = \text{tr } A \text{tr } B$

A.3 L'opérateur vec

Soit la matrice $m \times n$ $A \equiv (a_1, \dots, a_n)$ une matrice formée en concaténant horizontalement n vecteurs colonnes $m \times 1$. L'opérateur vec transforme la matrice A en un vecteur $mn \times 1$ en empilant (verticalement) les vecteurs a_i pour $i = 1, \dots, n$. Nous avons :

$$\text{vec } A = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

Soient A, B et C des matrices avec les dimensions appropriées, on peut établir les propriétés suivantes :

1. $\text{vec } A + B = \text{vec } A + \text{vec } B$
2. $\text{vec } ABC = (C' \otimes A) \text{vec } B$
3. $\text{vec } AB = (I \otimes A) \text{vec } B$
4. $\text{vec } ABC = (I \otimes AB) \text{vec } C = (C' B' \otimes I) \text{vec } A$
5. $(\text{vec } B')' \text{vec } A = \text{tr } BA = \text{tr } AB = (\text{vec } A')' \text{vec } B$
6. $\text{tr } ABC = (\text{vec } A')' (C' \otimes I) \text{vec } B = (\text{vec } A')' (I \otimes B) \text{vec } C$

A.4 Décomposition de Cholesky

Si A est une matrice carrée $m \times m$ définie positive alors il existe une matrice C $m \times m$ triangulaire inférieure telle que $A = C \times C$. Dans il faut utiliser la commande `chol` avec l'option `'lower'` . Par exemple :

A.5 Inversion par bloc

Soit la matrice carrée \mathcal{A} définie par :

$$\mathcal{A} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

où A et D sont des sous matrices carrées. En posant le système d'équation linéaire associé à $\mathcal{A}^{-1} \mathcal{A} = I$ et $\mathcal{A} \mathcal{A}^{-1} = I$ on peut alors facilement montrer que si \mathcal{A} est inversible, nous avons :

$$\mathcal{A}^{-1} = \begin{pmatrix} \mathcal{D} & -\mathcal{D} B D^{-1} \\ -D^{-1} C A^{-1} & D^{-1} + D^{-1} C \mathcal{D} B D^{-1} \end{pmatrix}$$

avec $\mathcal{D} = (A - B D^{-1} C)^{-1}$.

B Modèle VAR(1) et modèle ARMA

```

1 function DataSet = simul_var_1(AutoregressiveMatrix,ConstantVector,SampleSize)
2 % Simulation of a K-dimensional VAR(1) model:
3 %
4 %    $y_t = c + A * y_{t-1} + u_t$  for  $t=1,\dots,T$ 
5 %
6 % where  $u_t$  is a multivariate (K*1) zero mean white noise with unitary variance (the
7 % covariance matrix is a K*K identity matrix). The initial condition is zero ( $y_0 = 0$ ).
8 %
9 % INPUT
10 % [1] AutoregressiveMatrix    [double]    the (K*K) autoregressive matrix (A).
11 % [2] ConstantVector          [double]    a (K*1) vector (c)
12 % [3] SampleSize              [integer]   a scalar specifying the number of periods (T)
13 %
14 % OUTPUT
15 %
16 %
17 % [1] DataSet                  [double]   a (K*SampleSize) matrix holding the K simulated
18 %                               time series.
19 %
20 % stephane DOT adjemian AT ens DOT fr
21
22     K = length(ConstantVector);
23     DataSet = zeros(K,SampleSize+1);
24
25     size(DataSet)
26
27     for t=1:SampleSize+1
28         DataSet(:,t+1) = ConstantVector + AutoregressiveMatrix*DataSet(:,t) + randn(K,1) ;
29     end
30
31     DataSet = DataSet(:,2:SampleSize+1);
32     size(DataSet)

```

FIG. 1 – Simulation d'un VAR(1).

```

1 K = 2 ;           % Number of endogenous variables in the VAR(1) model.
2 A = zeros(K,K);  % Initialization of the autoregressive matrix.
3 C = zeros(K,1);  % Initialization of the constant vector.
4 T = 500;         % Size of the sample.
5
6 C(1) = .1;
7 C(2) = .2;
8
9 A(1,1) = 0.1;
10 A(1,2) = -0.2;
11 A(2,1) = 0;
12 A(2,2) = 0.4;
13
14 % Check if all the eigenvalues are less than one in modulus (stability condition).
15 EigenValues = eig(A);
16 test = (abs(EigenValues)<1)
17 if sum(test)==K
18     disp('The VAR(1) model is stable!')
19 else
20     disp('The stability condition is not satisfied!')
21 end
22
23 % Simulation of the K time series.
24 DataSet = simul_var_1(A,C,T);
25
26 size(DataSet)
27
28 % Plots.
29 close all
30
31 figure(1)
32 plot(1:500,DataSet(1,:), '-k');
33 box on
34 grid on
35 axis tight
36
37 figure(2)
38 plot(1:500,DataSet(2,:), '-k');
39 box on
40 grid on
41 axis tight

```

FIG. 2 – Simulation d'un VAR(1).

```

1 function DataSet = simul_var(ConstantVector,LowerCholSigma,SampleSize,varargin)
2 % Simulation of a K-dimensional VAR(p) model:
3 %
4 %  $y_t = c + A_1 * y_{t-1} + \dots + A_p * y_{t-p} + u_t$  for  $t=1,\dots,T$ 
5 %
6 % where  $u_t$  is a multivariate (K*1) zero mean white noise with variance
7 % LowerCholSigma*LowerCholSigma'.
8 % The initial condition is zero ( $y_0 = 0$ ).
9 %
10 % INPUT
11 % [1] ConstantVector [double] a (K*1) vector (c)
12 % [2] LowerCholSigma [double] a (K*K) matrix, the lower cholesky decomposition
13 % of the covariance matrix.
14 % [3] SampleSize [integer] a scalar specifying the number of periods (T)
15 % [+] Optional K*K autoregressive matrices
16 % OUTPUT
17 %
18 % [1] DataSet [double] a (K*SampleSize) matrix holding the K simulated
19 % time series.
20 %
21 % stephane DOT adjemian AT ens DOT fr
22
23 NumberOfLags = nargin-3;
24
25 if NumberOfLags<1
26     error(['simul_var:: Use another routine if you want to simulate a' ...
27         'multivariate white noise!'])
28 end
29
30 NumberOfVariables = length(ConstantVector);
31
32 if NumberOfVariables ~= length(LowerCholSigma)
33     error('simul_var:: Dimension mismatch!')
34 end
35 if NumberOfVariables ~= length(varargin{1})
36     error('simul_var:: Dimension mismatch!')
37 end
38 if size(varargin{1},1) ~= size(varargin{1},2)
39     error('simul_var:: Autoregressive matrices have to be squared matrices!')
40 end
41
42 % Get innovations from a multivariate gaussian distribution
43 % (with zero mean and unit variance).
44 Innovations = randn(NumberOfVariables,SampleSize) ;
45
46 % Rescale the innovations such that the covariance is Sigma.
47 Innovations = LowerCholSigma*Innovations ;
48
49 % Initialization of the returned array.
50 DataSet = zeros(NumberOfVariables,SampleSize);
51
52 % Set a column vector gathering NumberOfLags lagged data.
53 LaggedData = zeros(NumberOfVariables*NumberOfLags,1);
54
55 % Define a (NumberOfVariables,NumberOfVariables*NumberOfLags) matrix gathering the
56 % autoregressive matrices.
57 AutoregressiveMatrices = zeros(NumberOfVariables,NumberOfVariables*NumberOfLags);
58 for lag=1:NumberOfLags
59     AutoregressiveMatrices(:,NumberOfVariables*(lag-1)+1:NumberOfVariables*lag) = ...
60         varargin{lag};
61 end
62
63 % Simulation of the VAR(p) model.
64 for t=1:SampleSize
65     % Simulation for period t.
66     DataSet(:,t) = ConstantVector + ...
67         Innovations(:,t) + ...
68         AutoregressiveMatrices*LaggedData;
69     % Update LaggedData.
70     LaggedData(NumberOfVariables+1:NumberOfVariables*NumberOfLags) = ...
71         LaggedData(1:NumberOfVariables*(NumberOfLags-1));
72     LaggedData(1:NumberOfVariables) = DataSet(:,t);
73 end

```

FIG. 3 – Simulation d'un VAR(p).

```

1 K = 2 ;           % Number of endogenous variables in the VAR(2) model.
2 p = 2 ;           % Number of lags.
3 A1 = zeros(K,K); % Initialization of the first autoregressive matrix.
4 A2 = zeros(K,K); % Initialization of the first autoregressive matrix.
5 C = zeros(K,1);  % Initialization of the constant vector.
6 T = 500;         % Size of the sample.
7
8 % Define the vector of constants.
9 C(1) = .1;
10 C(2) = .2;
11
12 % Define two autoregressive matrices.
13 A1(1,1) = 0.2;
14 A1(1,2) = -0.1;
15 A1(2,1) = 0.1;
16 A1(2,2) = 0.4;
17 A2(1,1) = 0.1;
18 A2(1,2) = -0.2;
19 A2(2,2) = 0.3;
20
21 % Build the compagnon matrix.
22 CompagnonMatrix = zeros(K*p,K*p);
23 CompagnonMatrix(1:K,1:K) = A1;
24 CompagnonMatrix(1:K,K+1:2*K) = A2;
25 CompagnonMatrix(K+1:2*K,1:K) = eye(K);
26
27 % Check if all the eigenvalues are less than one in modulus (stability condition).
28 EigenValues = eig(CompagnonMatrix);
29 test = (abs(EigenValues)<1);
30 if sum(test)==K*p
31     disp(['The VAR(' int2str(p) ') model is stable!'])
32 else
33     disp('The stability condition is not satisfied!')
34     disp('The eigenvalues of the Compagnon matrix are:')
35     abs(EigenValues)
36     return
37 end
38
39 % Set the covariance matrix
40 Sigma = zeros(K,K);
41 Sigma(1,1) = 0.2;
42 Sigma(2,2) = 0.4;
43 Sigma(1,2) = 0.1;
44 Sigma(2,1) = Sigma(1,2);
45
46 % Simulation of the K time series.
47 %
48 % NOTE If Sigma isn't a symmetric definite positive matrix then matlab won't be able to compute
49 % the Cholesky factorization and will crash.
50 DataSet = simul_var(C,chol(Sigma,'lower'),T,A1,A2);
51
52 % Build Plots of the simulated times series.
53 for variable=1:K
54     figure(variable)
55     plot(1:T,DataSet(variable,:), '-k');
56     box on
57     grid on
58     axis tight
59 end

```

FIG. 4 – Simulation d'un VAR(p).