

La Modélisation VAR

June 20, 2001

1. Représentation VAR

1.1. Exemple introductif

On considère deux processus stationnaires $\{y_{1,t}, t \in \mathbb{Z}\}$ et $\{y_{2,t}, t \in \mathbb{Z}\}$ définies par les relations suivantes :

$$y_{1,t} = a_1 + \sum_{i=1}^p b_{1,i} y_{1,t-i} + \sum_{i=1}^p c_{1,i} y_{2,t-i} - d_1 y_{2,t} + \varepsilon_{1,t} \quad (1.1)$$

$$y_{2,t} = a_2 + \sum_{i=1}^p b_{2,i} y_{1,t-i} + \sum_{i=1}^p c_{2,i} y_{2,t-i} - d_2 y_{1,t} + \varepsilon_{2,t} \quad (1.2)$$

où les innovations $\{\varepsilon_{1,t}, t \in \mathbb{Z}\}$ et $\{\varepsilon_{2,t}, t \in \mathbb{Z}\}$ sont des bruits de variance respective σ_1^2 et σ_2^2 , et non corrélés : $E(\varepsilon_{1,t} \varepsilon_{2,t-j}) = 0, \forall j \in \mathbb{Z}$. On constate immédiatement que le processus vectoriel $Y_t = (y_{1,t} \ y_{2,t})'$ peut s'écrire sous la forme d'un processus $AR(p)$. En effet :

$$B = \begin{pmatrix} 1 & d_1 \\ d_2 & 1 \end{pmatrix} \quad A_0 = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad A_i = \begin{pmatrix} b_{1,i} & c_{1,i} \\ b_{2,i} & c_{2,i} \end{pmatrix} \quad \forall i \in [1, p]$$

On définit un processus vectoriel ε_t *i.i.d.* tel que :

$$\varepsilon_t = \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \quad E(\varepsilon_t \varepsilon_t') = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$$

Alors, on montre immédiatement que :

$$BY_t = A_0 + \sum_{i=1}^p A_i Y_{t-i} + \varepsilon_t \quad (1.3)$$

On qualifie cette représentation de processus *VAR* (*Vectorial Autoregressive*) d'ordre p , noté $VAR(p)$. Ce système initial donnée par les équations (1.1) et (1.2), ou par la définition matricielle (1.3) est qualifiée de *représentation structurelle*. On constate que dans cette

représentation le niveau de $y_{2,t}$ a un effet immédiat sur $y_{1,t}$ et vice et versa. L'estimation de ce modèle suppose donc d'estimer $4 * (p + 1) + 2 + 2$ paramètres.

C'est pourquoi on travaille généralement à partir de la *forme réduite* du modèle VAR. Ce modèle, obtenu en multipliant les deux membres de (1.3) par B^{-1} , s'écrit alors sous la forme :

$$Y_t = \tilde{A}_0 + \sum_{i=1}^p \tilde{A}_i Y_{t-i} + v_t \quad (1.4)$$

avec :

$$\begin{aligned} \tilde{A}_i &= B^{-1} A_i \quad \forall i \in [0, p] \\ v_t &= B^{-1} \varepsilon_t \quad \forall t \in \mathbb{Z} \end{aligned}$$

Ce qui peut s'écrire sous la forme :

$$y_{1,t} = \tilde{a}_1 + \sum_{i=1}^p \tilde{b}_{1,i} y_{1,t-i} + \sum_{i=1}^p \tilde{c}_{1,i} y_{2,t-i} + v_{1,t} \quad (1.5)$$

$$y_{2,t} = \tilde{a}_2 + \sum_{i=1}^p \tilde{b}_{2,i} y_{1,t-i} + \sum_{i=1}^p \tilde{c}_{2,i} y_{2,t-i} + v_{2,t} \quad (1.6)$$

On constate alors que le niveau de $y_{2,t}$ ne dépend plus directement de $y_{1,t}$, mais seulement des valeurs passées de $y_{2,t}$ et de $y_{1,t}$, et de l'innovation $v_{2,t}$. Les innovations $v_{1,t}$ et $v_{2,t}$ sont alors fonction des innovations de la forme structurelle ($\varepsilon_{1,t}$ et $\varepsilon_{2,t}$) et peuvent être corrélées même en l'absence de corrélation des innovations ε_t . En effet, on a $\forall t \in \mathbb{Z}$:

$$v_{1,t} = \frac{\varepsilon_{1,t} - d_1 \varepsilon_{2,t}}{1 - d_1 d_2} \quad (1.7)$$

$$v_{2,t} = \frac{\varepsilon_{2,t} - d_2 \varepsilon_{1,t}}{1 - d_1 d_2} \quad (1.8)$$

Dès lors, on vérifie que les processus $\{v_{1,t}, t \in \mathbb{Z}\}$ et $\{v_{2,t}, t \in \mathbb{Z}\}$ sont *i.i.d.* puisque :

$$E(v_{1,t}) = E(v_{2,t}) = 0$$

$$E(v_{1,t} v_{1,t-j}) = E(v_{2,t} v_{2,t-j}) = 0 \quad \forall j \in \mathbb{Z}^*$$

Les variances de ces innovations sont alors définies par $\forall t \in \mathbb{Z}$

$$E(v_{1,t}^2) = \frac{\sigma_1^2 + d_1^2 \sigma_2^2}{(1 - d_1 d_2)^2}$$

$$E(v_{2,t}^2) = \frac{\sigma_2^2 + d_2^2 \sigma_1^2}{(1 - d_1 d_2)^2}$$

Enfin, on constate que les innovations $\{v_{1,t}, t \in \mathbb{Z}\}$ et $\{v_{2,t}, t \in \mathbb{Z}\}$ peuvent être corrélées alors même que les innovations du modèle structurel $\{\varepsilon_{1,t}, t \in \mathbb{Z}\}$ et $\{\varepsilon_{2,t}, t \in \mathbb{Z}\}$ sont non corrélées.

$$E(v_{1,t}v_{2,t-h}) = \begin{cases} -\frac{d_2^2\sigma_1^2 + d_1^2\sigma_2^2}{(1-d_1d_2)^2} & h = 0 \\ 0 & h \neq 0 \end{cases} \quad (1.9)$$

On constate que cette covariance est nulle en particulier lorsque $d_1 = d_2 = 0$, puisque dans ce cas là le niveau de $y_{2,t}$ n'a pas d'influence sur celui de $y_{1,t}$ et vice et versa.

1.2. Représentation générale

La définition générale d'un processus $VAR(p)$ est la suivante.

Definition 1.1. Un processus vectoriel $\{X_t, t \in \mathbb{Z}\}$, de dimension $(n, 1)$, admet une représentation VAR d'ordre p , notée $VAR(p)$ si :

$$X_t = c - \Phi_1 X_{t-1} - \Phi_2 X_{t-2} - \dots - \Phi_p X_{t-p} + \varepsilon_t \quad (1.10)$$

ou de façon équivalente :

$$\Phi(L) X_t = c + \varepsilon_t \quad (1.11)$$

où c , dimension $(n, 1)$ désigne un vecteur de constantes, $\Phi(L) = \sum_{i=0}^{\infty} \Phi_i L^i$, où les matrices $\Phi_i, \forall i \in [0, p]$ de dimension (n, n) , satisfont $\Phi_0 = I_n$ et $\Phi_p \neq 0_n$. Le vecteur $(n, 1)$ des innovations ε_t est *i.i.d.* $(0_n, \Sigma)$ où Σ est une matrice (n, n) symétrique définie positive.

Le processus vectoriel des innovations $\{\varepsilon_t, t \in \mathbb{Z}\}$ est *i.i.d.* $(0_n, \Sigma)$, et satisfait par conséquent les propriétés suivantes :

$$E(\varepsilon_t) = 0$$

$$E(\varepsilon_t \varepsilon_{t-j}') = \begin{cases} \Sigma & j = 0 \\ 0 & j \neq 0 \end{cases}$$

De la même façon que pour un $AR(1)$, on peut donc exprimer le polynôme matriciel $\Phi(L)$, de dimension (n, n) , de la façon suivante :

$$\Phi(L) = \sum_{i=0}^{\infty} \Phi_i L^i = I_n + \Phi_1 L - \Phi_2 L^2 - \dots - \Phi_p L^p$$

où I_n désigne la matrice identité (n, n) . On pose les définitions suivantes :

$$\underset{(n,1)}{X_t} = \begin{pmatrix} x_{1,t} \\ x_{2,t} \\ \dots \\ x_{n,t} \end{pmatrix} \quad \underset{(n,1)}{\varepsilon_t} = \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \dots \\ \varepsilon_{n,t} \end{pmatrix} \quad \underset{(n,n)}{\Phi_i} = \begin{pmatrix} \Phi_{1,1}^i & \Phi_{1,2}^i & \dots & \Phi_{1,n}^i \\ \Phi_{2,1}^i & \Phi_{2,2}^i & \dots & \dots \\ \dots & \dots & \Phi_{j,k}^i & \dots \\ \Phi_{n,1}^i & \Phi_{n,2}^i & \dots & \Phi_{n,n}^i \end{pmatrix} \quad \forall i \in [1, p]$$

On retrouve alors la forme réduite évoquée dans l'exemple précédent puisque, les processus $\{x_{i,t}, t \in \mathbb{Z}\}$ sont respectivement définis en fonctions de leur passé et du passé des processus $\{x_{j,t}, t \in \mathbb{Z}\}$ pour $j \neq i$. Par exemple, pour $x_{1,t}$ on obtient, $\forall t \in \mathbb{Z}$:

$$\begin{aligned} x_{1,t} &= c_j + \Phi_{1,1}^1 x_{1,t-1} + \Phi_{1,2}^1 x_{2,t-1} + \dots + \Phi_{1,n}^1 x_{n,t-1} \\ &\quad + \Phi_{1,1}^2 x_{1,t-2} + \Phi_{1,2}^2 x_{2,t-2} + \dots + \Phi_{1,n}^2 x_{n,t-2} \\ &\quad + \dots \\ &\quad + \Phi_{1,1}^p x_{1,t-p} + \Phi_{1,2}^p x_{2,t-p} + \dots + \Phi_{1,n}^p x_{n,t-p} \end{aligned}$$

De façon plus générale, pour $x_{j,t}$, $\forall j \in [1, n]$, on a :

$$\begin{aligned} x_{j,t} &= c_1 + \Phi_{j,1}^1 x_{1,t-1} + \Phi_{j,2}^1 x_{2,t-1} + \dots + \Phi_{j,j}^1 x_{j,t-1} + \dots + \Phi_{j,n}^1 x_{n,t-1} \\ &\quad + \Phi_{j,1}^2 x_{1,t-2} + \Phi_{j,2}^2 x_{2,t-2} \dots + \Phi_{j,j}^2 x_{j,t-2} + \dots + \Phi_{j,n}^2 x_{n,t-2} \\ &\quad + \dots \\ &\quad + \Phi_{j,1}^p x_{1,t-p} + \Phi_{j,2}^p x_{2,t-p} \dots + \Phi_{j,j}^p x_{j,t-p} + \dots + \Phi_{j,n}^p x_{n,t-p} \end{aligned}$$

1.3. Conditions de stationnarité

La définition de la stationnarité d'ordre deux (ou stationnarité du second ordre) est identique à celle du cas des processus univariés.

Definition 1.2. *Un processus vectoriel $\{X_t, t \in \mathbb{Z}\}$, de dimension $(n, 1)$, est dit stationnaire au second ordre, ou stationnaire au sens faible, ou stationnaire d'ordre deux si*

- $\forall t \in \mathbb{Z}, E(X_t^2) < \infty$
- $\forall t \in \mathbb{Z}, E(X_t) = m_{(n,1)}$ (indépendant de t)
- $\forall (t, h) \in \mathbb{Z}^2, E[(X_{t+h} - m) (X_t - m)'] = \gamma(h)_{(n,n)}$ (indépendant de t)

Lorsque l'on considère un processus $VAR(p)$, on peut démontrer que ces conditions de stationnarité reviennent à imposer des conditions sur les racines du déterminant du polynôme matriciel $\Phi(L)$.

Proposition 1.3. *Un processus vectoriel $\{X_t, t \in \mathbb{Z}\}$, de dimension $(n, 1)$, satisfaisant une représentation $VAR(p)$ telle que $\forall t \in \mathbb{Z}$:*

$$\Phi(L) X_t = X_t - \Phi_1 X_{t-1} - \Phi_2 X_{t-2} - \dots - \Phi_p X_{t-p} = c + \varepsilon_t$$

est stationnaire si et seulement si les racines du déterminant du polynôme matriciel $\Phi(L)$, notée λ_i $i \in [1, n]$, sont toutes supérieures à l'unité en module.

$$\begin{aligned} \det[\Phi(\lambda_i)] &= |\Phi(\lambda_i)| = |I_n \lambda_i^p - \Phi_1 \lambda_i^{p-1} - \Phi_2 \lambda_i^{p-2} - \dots - \Phi_{p-1} \lambda_i - \Phi_p| = 0 \\ &|\lambda_i| > 1 \quad \forall i \in [1, n] \end{aligned} \tag{1.12}$$

Exemple : On considère un processus bi-varié $\{Y_t, t \in \mathbb{Z}\}$ admettant une représentation $VAR(1)$ telle que :

$$\begin{aligned} y_{1,t} &= 3 + 0.2y_{1,t-1} + 0.7y_{2,t-1} + \varepsilon_{1,t} \\ y_{2,t} &= 1 + 0.3y_{1,t-1} + 0.4y_{2,t-1} + \varepsilon_{2,t} \end{aligned}$$

On pose $Y_t = (y_{1,t} \ y_{2,t})'$ et $\varepsilon_t = (\varepsilon_{1,t} \ \varepsilon_{2,t})'$. Cette représentation peut s'exprimer sous la forme réduite suivante :

$$\Phi(L) Y_t = c + \varepsilon_t$$

avec $c = (3 \ 1)'$ et :

$$\Phi(L) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} -0.2 & -0.7 \\ -0.3 & -0.4 \end{pmatrix} L = \begin{pmatrix} 1 - 0.2L & -0.7L \\ -0.3L & 1 - 0.4L \end{pmatrix}$$

Donc la condition de stationnarité du processus $\{Y_t, t \in \mathbb{Z}\}$ se ramène à déterminer les racines du polynôme :

$$\det[\Phi(L)] = (1 - 0.2L)(1 - 0.4L) - 0.21L^2$$

Les deux racines $\lambda_1 = 1.30$ et $\lambda_2 = -5.91$ sont supérieures à 1 en module, le processus VAR est stationnaire. Donc les processus univariées $\{y_{1,t}, t \in \mathbb{Z}\}$ et $\{y_{2,t}, t \in \mathbb{Z}\}$ sont stationnaires.

On peut facilement démontrer que cette condition de stationnarité peut être exprimée en fonction des valeurs propres de la matrice $\Phi(L)$.

Proposition 1.4. *Un processus vectoriel $\{X_t, t \in \mathbb{Z}\}$, de dimension $(n, 1)$, satisfaisant une représentation $VAR(p)$ telle que $\forall t \in \mathbb{Z}$:*

$$\Phi(L) X_t = X_t - \Phi_1 X_{t-1} - \Phi_2 X_{t-2} - \dots - \Phi_p X_{t-p} = c + \varepsilon_t$$

est stationnaire si et seulement si les valeurs propres de l'application linéaire $\Phi(L)$, notée $\tilde{\lambda}_i$ $i \in [1, n]$, sont toutes inférieures à l'unité en module. Ces valeurs propres satisfont l'équation caractéristique associée :

$$\left| I_n \tilde{\lambda}_i^p - \Phi_1 \tilde{\lambda}_i^{p-1} - \Phi_2 \tilde{\lambda}_i^{p-2} - \dots - \Phi_{p-1} \tilde{\lambda}_i - \Phi_p \right| = 0 \quad (1.13)$$

$$\left| \tilde{\lambda}_i \right| < 1 \quad \forall i \in [1, n] \quad (1.14)$$

1.4. Ecriture $VAR(1)$ d'un $VAR(p)$

Les processus $VAR(p)$ possède une propriété particulièrement utile pour les démonstrations ultérieures.

Proposition 1.5. *Tout processus vectoriel $\{X_t, t \in \mathbb{Z}\}$, satisfaisant une représentation $VAR(p)$ peut être transformé en un processus $\{\tilde{X}_t, t \in \mathbb{Z}\}$ satisfaisant une représentation $VAR(1)$ d'espérance nulle.*

Preuve : On considère un processus $\{X_t, t \in \mathbb{Z}\}$, avec $X_t = (x_{1,t}, \dots, x_{n,t})'$ satisfaisant la représentation $VAR(p)$ suivante, $\forall t \in \mathbb{Z}$:

$$\underset{(n,n)}{\Phi}(L) X_t = X_t - \Phi_1 X_{t-1} - \Phi_2 X_{t-2} - \dots - \Phi_p X_{t-p} = c + \varepsilon_t \underset{(n,1)}{}$$

Déterminons tout d'abord l'espérance du processus X_t :

$$E(X_t) = \Phi(L)^{-1} [c + \varepsilon_t] = \Phi(1)^{-1} c = \mu \quad \forall t \in \mathbb{Z}$$

où μ est un vecteur de constante (hypothèse de stationnarité) de dimension $(n, 1)$. On peut alors réécrire le processus sous la forme suivante :

$$\Phi(L)(X_t - \mu) = \varepsilon_t$$

$$\iff (X_t - \mu) = \Phi_1(X_{t-1} - \mu) + \Phi_2(X_{t-2} - \mu) + \dots + \Phi_p(X_{t-p} - \mu) + \varepsilon_t$$

On pose $\forall t \in \mathbb{Z}$

$$\underset{(np,1)}{\tilde{X}}_t = \begin{pmatrix} X_t - \mu \\ X_{t-1} - \mu \\ X_{t-2} - \mu \\ \dots \\ X_{t-p+1} - \mu \end{pmatrix} \quad \underset{(np,1)}{v}_t = \begin{pmatrix} \varepsilon_t \\ 0_{(n,1)} \\ 0_{(n,1)} \\ \dots \\ 0_{(n,1)} \end{pmatrix}$$

et

$$\underset{(np,np)}{A} = \begin{pmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_{p-1} & \Phi_p \\ I_n & 0_{(n,n)} & \dots & 0_{(n,n)} & 0_{(n,n)} \\ 0_{(n,n)} & I_n & \dots & 0_{(n,n)} & 0_{(n,n)} \\ 0_{(n,n)} & 0_{(n,n)} & I_n & 0_{(n,n)} & 0_{(n,n)} \\ 0_{(n,n)} & 0_{(n,n)} & 0_{(n,n)} & I_n & 0_{(n,n)} \end{pmatrix}$$

Alors le processus $VAR(p)$ X_t peut se réécrire sous la forme d'un processus transformé \tilde{X}_t satisfaisant une représentation $VAR(1)$ tel que :

$$\tilde{X}_t = A\tilde{X}_{t-1} + v_t$$

Exemple : On considère un processus bi-varié $\{Y_t, t \in \mathbb{Z}\}$ admettant une représentation $VAR(2)$ telle que :

$$\begin{aligned} y_{1,t} &= 3 + 0.2y_{1,t-1} + 0.7y_{2,t-1} - 0.4y_{1,t-2} - 0.6y_{2,t-2} + \varepsilon_{1,t} \\ y_{2,t} &= 1 + 0.3y_{1,t-1} + 0.4y_{2,t-1} - 0.1y_{1,t-2} - 0.8y_{2,t-2} + \varepsilon_{2,t} \end{aligned}$$

On pose $Y_t = (y_{1,t} \ y_{2,t})'$ et $\varepsilon_t = (\varepsilon_{1,t} \ \varepsilon_{2,t})'$. Cette représentation peut s'exprimer sous la forme réduite suivante :

$$\Phi(L) Y_t = c + \varepsilon_t$$

avec $c = (3 \ 1)'$ et :

$$\begin{aligned} \Phi(L) &= \Phi_0 + \Phi_1 L + \Phi_2 L^2 \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} -0.2 & -0.7 \\ -0.3 & -0.4 \end{pmatrix} L + \begin{pmatrix} 0.4 & 0.6 \\ 0.1 & 0.8 \end{pmatrix} L^2 \\ &= \begin{pmatrix} 1 - 0.2L + 0.4L^2 & -0.7L + 0.6L^2 \\ -0.3L + 0.1L^2 & 1 - 0.4L + 0.8L^2 \end{pmatrix} \end{aligned}$$

Déterminons $E(Y)$:

$$E(Y_t) = \Phi(1)^{-1} c = \begin{pmatrix} 1.2 & -0.1 \\ -0.2 & 1.4 \end{pmatrix}^{-1} \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 2.59 \\ 1.8 \end{pmatrix} = \mu$$

On pose $\forall t \in \mathbb{Z}$

$$\tilde{Y}_t = \begin{pmatrix} Y_t - \mu \\ Y_{t-1} - \mu \end{pmatrix} = \begin{pmatrix} y_{1,t} - 2.59 \\ y_{2,t} - 1.8 \\ y_{1,t-1} - 2.59 \\ y_{2,t-1} - 1.8 \end{pmatrix} \quad v_t = \begin{pmatrix} \varepsilon_t \\ 0_{(2,1)} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ 0 \\ 0 \end{pmatrix}$$

et

$$A = \begin{pmatrix} \Phi_1 & \Phi_2 \\ I_2 & 0_{(2,2)} \end{pmatrix} = \begin{pmatrix} -0.2 & -0.7 & 0.4 & 0.6 \\ -0.3 & -0.4 & 0.1 & 0.8 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

Alors on montre que la représentation $VAR(1)$ du processus \tilde{Y}_t est identique à la représentation $VAR(p)$ du processus Y_t puisque :

$$\begin{aligned} \tilde{Y}_t &= A\tilde{Y}_{t-1} + v_t \\ \Leftrightarrow \begin{pmatrix} y_{1,t} - 2.59 \\ y_{2,t} - 1.8 \\ y_{1,t-1} - 2.59 \\ y_{2,t-1} - 1.8 \end{pmatrix} &= \begin{pmatrix} -0.2 & -0.7 & 0.4 & 0.6 \\ -0.3 & -0.4 & 0.1 & 0.8 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} y_{1,t-1} - 2.59 \\ y_{2,t-1} - 1.8 \\ y_{1,t-2} - 2.59 \\ y_{2,t-2} - 1.8 \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ 0 \\ 0 \end{pmatrix} \\ \Leftrightarrow \begin{cases} y_{1,t} = 3 + 0.2y_{1,t-1} + 0.7y_{2,t-1} - 0.4y_{1,t-2} - 0.6y_{2,t-2} + \varepsilon_{1,t} \\ y_{2,t} = 1 + 0.3y_{1,t-1} + 0.4y_{2,t-1} - 0.1y_{1,t-2} - 0.8y_{2,t-2} + \varepsilon_{2,t} \end{cases} \end{aligned}$$

1.5. Représentation VMA

Considérons un processus vectoriel $\{X_t, t \in \mathbb{Z}\}$, de dimension $(n, 1)$, stationnaire.

Remark 1. *Tout comme dans le cas univarié, sous la condition de stationnarité, il est possible d'appliquer le théorème de Wold et de représenter X_t sous la forme d'un processus vectoriel moyenne mobile infini VMA(∞).*

Nous allons à nouveau donner l'intuition du théorème de Wold appliqué aux processus vectoriels. On considère un processus stationnaire satisfaisant la représentation VAR(p) suivante, $\forall t \in \mathbb{Z}$:

$$\Phi(L)X_t = X_t - \Phi_1 X_{t-1} - \Phi_2 X_{t-2} - \dots - \Phi_p X_{t-p} = c + \varepsilon_t$$

On sait que $E(X_t) = \Phi(1)^{-1}c = \mu$, où μ est un vecteur de constante (hypothèse de stationnarité) de dimension $(n, 1)$.

On sait d'après la propriété précédente que ce processus VAR(p) peut être réexprimer sous la forme d'un processus VAR(1) tel que :

$$\tilde{\Phi}(L)\tilde{X}_t = v_t \iff \tilde{X}_t = A\tilde{X}_{t-1} + v_t$$

où les processus \tilde{X}_t et v_t , ainsi que la matrice A ont été définis précédemment. Dès lors, en itérant vers le passé, on montre que le processus \tilde{X}_t peut s'écrire sous la forme d'une moyenne mobile d'ordre fini à t donné.

$$\tilde{X}_t = v_t + Av_{t-1} + A^2v_{t-2} + \dots + A^k\tilde{X}_{t-k}$$

Si l'on suppose que les valeurs propres de la matrice A sont strictement inférieures à 1 en module (condition de stationnarité), alors $\lim_{k \rightarrow \infty} A^k = 0$. Dès lors, le processus \tilde{X}_t peut s'écrire sous la forme :

$$\tilde{X}_t = \lim_{k \rightarrow \infty} \sum_{i=0}^k A^i v_{t-i} = \sum_{i=0}^{\infty} A^i v_{t-i} = \tilde{\Psi}(L)v_t$$

En reprenant la définition du processus \tilde{X}_t , on peut alors déterminer la décomposition de Wold associée au processus VAR(p) X_t .

Proposition 1.6. *Tout processus $\{X_t, t \in \mathbb{Z}\}$, de dimension $(n, 1)$, stationnaire, satisfaisant une représentation VAR(p), telle que, $\forall t \in \mathbb{Z}$:*

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \dots + \Phi_p X_{t-p} + c + \varepsilon_t \quad (1.15)$$

admet une représentation moyenne mobile convergente définie par :

$$X_t = \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} = \mu + \Psi(L)\varepsilon_t$$

avec $\mu = E(X_t) = \Phi(1)^{-1}c$, où ε_t est un bruit blanc vectoriel et où la séquence des matrices de dimension (n, n) , $\{\psi_i\}_{i=0}^{\infty}$ satisfait $\psi_0 = I_n$ et est absolument sommable au sens où les éléments $\psi_{j,k}^i$ de la matrice ψ_i satisfont la condition :

$$\sum_{s=0}^{\infty} |(\psi_{j,k}^i)^s| < \infty \quad \forall i \geq 1, \forall (j, k) \in [1, n]^2$$

La condition de sommabilité, qui garantit la convergence des moments d'ordre deux du processus X_t , doit se comprendre dans ce contexte comme une condition de sommabilité sur une séquence de matrices $\{\psi_i\}_{i=0}^{\infty}$. Cette condition se ramène à démontrer la sommabilité de tous les éléments $\psi_{j,k}^i$ de ces matrices, au sens traditionnel de la sommabilité d'une séquence de scalaires (cf. chapitre 1).

Il est possible de déterminer de façon générale la forme des matrices de l'opérateur polynômial associée à la représentation $VMA(\infty)$.

Proposition 1.7. *Le polynôme matriciel $\Psi(L) = \sum_{i=0}^{\infty} \psi_i L^i$ associé à la représentation $VMA(\infty)$ d'un processus $VAR(p)$ stationnaire, $\{X_t, t \in \mathbb{Z}\}$, $\forall t \in \mathbb{Z}$:*

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \dots + \Phi_p X_{t-p} + c + \varepsilon_t \quad (1.16)$$

satisfait la relation de récurrence suivante :

$$\psi_0 = I_n$$

$$\psi_s = \Phi_1 \psi_{s-1} + \Phi_2 \psi_{s-2} + \dots + \Phi_p \psi_{s-p} \quad \forall s \geq 1$$

avec $\psi_s = 0$, $\forall s < 0$.

Preuve : Une façon simple d'obtenir la représentation $VMA(\infty)$ d'un processus $VAR(p)$ consiste à identifier les polynômes matriciels $\Phi(L)^{-1}$ et $\Psi(L)$. En effet, dans le cas d'un processus centré ($c = 0$), on a :

$$X_t = \Phi(L)^{-1} \varepsilon_t = \Psi(L) \varepsilon_t \iff \Phi(L) \Psi(L) = I_n$$

Ce égalité peut se réécrire sous la forme :

$$\lim_{k \rightarrow \infty} [(I_n - \Phi_1 L - \Phi_2 L^2 - \dots - \Phi_p L^p) (I_n - \Psi_1 L - \Psi_2 L^2 - \dots - \Psi_k L^k)] = I_n$$

Dès lors, on montre par identification des termes de même ordre que les matrices de l'opérateur polynômial associée à la représentation $VMA(\infty)$ satisfont une équation de récurrence qui correspond à celle de la proposition.

Preuve : On considère un processus bi-varié stationnaire $\{Y_t, t \in \mathbb{Z}\}$ admettant une représentation $VAR(1)$ telle que :

$$\Phi(L) Y_t = c + \varepsilon_t$$

avec ε_t *i.i.d.* $(0, \quad)$, $c = (3 \ 1)'$ et :

$$\Phi(L) = \Phi_0 + \Phi_1 L = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} -0.2 & -0.7 \\ -0.3 & -0.4 \end{pmatrix} L = \begin{pmatrix} 1 - 0.2L & -0.7L \\ -0.3L & 1 - 0.4L \end{pmatrix}$$

Par application du théorème de Wold, on sait que ce processus peut être représenté comme sous une forme $VMA(\infty)$ telle que :

$$X_t = \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} = \mu + \Psi(L) \varepsilon_t$$

Immédiatement, on montre que

$$\mu = E(Y_t) = \begin{pmatrix} 0.8 & -0.7 \\ -0.3 & 0.6 \end{pmatrix}^{-1} \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 9.25 \\ 6.29 \end{pmatrix}$$

Par définition, on a $\Phi(L) \Psi(L) = I_2$, ce qui peut se réécrire sous la forme :

$$\lim_{k \rightarrow \infty} [(I_2 - \Phi_1 L) (\Psi_0 - \Psi_1 L - \Psi_2 L^2 - \dots - \Psi_p L^p - \dots - \Psi_k L^k)] = I_2$$

Par identification des membres de même terme, on montre que :

$$\begin{aligned} \Psi_0 &= I_2 \\ \Psi_1 &= \Phi_1 = \begin{pmatrix} -0.2 & -0.7 \\ -0.3 & -0.4 \end{pmatrix} \\ \Psi_2 &= \Phi_1 \Psi_1 = \Phi_1^2 = \begin{pmatrix} -0.2 & -0.7 \\ -0.3 & -0.4 \end{pmatrix}^2 \end{aligned}$$

et de façon générale, on a :

$$\Psi_n = \Phi_1 \Psi_{n-1} = \Phi_1^n = \begin{pmatrix} -0.2 & -0.7 \\ -0.3 & -0.4 \end{pmatrix}^n \quad \forall n \geq 1$$

On retrouve ainsi la formule générale que l'on avait établi par itération vers le passé dans le cas d'un $VAR(1)$:

$$Y_t = \mu + \Psi(L) \varepsilon_t = \mu + \sum_{i=0}^{\infty} \Phi_1^i \varepsilon_{t-i}$$

On montre ainsi que :

$$Y_t = \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} = \begin{pmatrix} 9.25 \\ 6.29 \end{pmatrix} + \sum_{i=0}^{\infty} \begin{pmatrix} -0.2 & -0.7 \\ -0.3 & -0.4 \end{pmatrix}^i \begin{pmatrix} \varepsilon_{1,t-i} \\ \varepsilon_{2,t-i} \end{pmatrix}$$

2. Estimation des paramètres

Tous comme pour les processus *AR* univariés plusieurs méthodes d'estimation sont envisageables pour les processus *VAR*. La première consiste tout simplement à appliquer les *MCO*. La seconde principale méthode consiste en le maximum de vraisemblance.

2.1. Maximum de Vraisemblance

On considère un processus $\{X_t, t \in \mathbb{Z}\}$, avec $X_t = (x_{1,t}, \dots, x_{n,t})'$ satisfaisant la représentation *VAR*(p) suivante, $\forall t \in \mathbb{Z}$:

$$\Phi(L) X_t = X_t - \Phi_1 X_{t-1} - \Phi_2 X_{t-2} - \dots - \Phi_p X_{t-p} = c + \varepsilon_t$$

On suppose que les innovations ε_t sont *i.i.d.* $\mathcal{N}(0, \Sigma)$ et que l'on dispose de $T+p$ observations du processus X_t . On cherche à déterminer la vraisemblance conditionnelle de X_t en fonction des réalisations passées X_{t-i} , $i \in [1, p]$. Par définition, la distribution conditionnelle de X_t s'écrit :

$$D(X_t/X_{t-1}, X_{t-2}, \dots, X_{t-p}) \sim \mathcal{N}(c + \Phi_1 X_{t-1} + \Phi_2 X_{t-2} - \dots + \Phi_p X_{t-p}, \Sigma)$$

Afin de simplifier les calculs, on pose :

$$\tilde{X}_t = \begin{pmatrix} 1 \\ X_{t-1} \\ X_{t-2} \\ \dots \\ X_{t-p} \end{pmatrix}_{(p,1)} \quad \Pi' = \begin{pmatrix} c \\ \Phi_1 \\ \Phi_2 \\ \dots \\ \Phi_p \end{pmatrix}$$

On a alors :

$$D(X_t/X_{t-1}, X_{t-2}, \dots, X_{t-p}) \sim \mathcal{N}(\Pi' \tilde{X}_t, \Sigma)$$

Si l'on note Θ le vecteur des paramètres à estimer :

$$\Theta = \begin{pmatrix} \Pi' \\ \text{vect}(\Sigma) \end{pmatrix} = \begin{pmatrix} c \\ \Phi_1 \\ \Phi_2 \\ \dots \\ \Phi_p \\ \text{vect}(\Sigma) \end{pmatrix}$$

Dès lors la densité conditionnelle de X_t s'écrit :

$$f(X_t/X_{t-1}, X_{t-2}, \dots, X_{t-p}; \Theta) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp \left[-\frac{1}{2} (X_t - \Pi' \tilde{X}_t)' \Sigma^{-1} (X_t - \Pi' \tilde{X}_t) \right]$$

En partant de cette expression, il est possible de dériver la vraisemblance sur l'ensemble de l'échantillon $\{X_t\}_{t=1}^T$ conditionnellement aux valeurs initiales $(X_0, X_{-1}, X_{-2}, \dots, X_{-p})$ s'écrit

$$f(X_t, X_{t-1}, X_{t-2}, \dots, X_1/X_{t-1}, X_{t-2}, \dots, X_{t-p}; \Theta) = \prod_{t=1}^T f(X_t/X_{t-1}, X_{t-2}, \dots, X_{t-p}; \Theta)$$

La log-vraisemblance d'un processus $VAR(p)$ s'écrit donc :

$$\begin{aligned} \mathcal{L}(\Theta) &= \sum_{t=1}^T \log [f(X_t/X_{t-1}, X_{t-2}, \dots, X_{t-p}; \Theta)] \\ &= -\frac{Tn}{2} \log(2\pi) + \frac{1}{2} \log(|\Sigma^{-1}|) - \frac{1}{2} \sum_{t=1}^T \left[(X_t - \Pi' \tilde{X}_t)' \Sigma^{-1} (X_t - \Pi' \tilde{X}_t) \right] \end{aligned} \quad (2.1)$$

La maximisation de cette vraisemblance permet alors d'obtenir des estimateurs convergents des paramètres Π et de la matrice de variance covariance des innovations Σ .

2.2. Détermination du nombre de retards

Pour déterminer le nombre de retards optimal pour un $VAR(p)$, on peut utiliser plusieurs méthodes. En particulier toutes les méthodes de comparaison de modèles étudiées dans le chapitre précédent sont valides dès qu'elles ont été adaptées au cas vectoriel.

Une procédure type consiste à estimer tous les modèles VAR pour des ordres p allant de 0 à un certain ordre h fixé de façon arbitraire (nombre de retards maximum pour la taille d'échantillon considéré, ou nombre de retards maximum compatible avec une théorie ou une intuition économique). Pour chacun de ces modèles, on calcule les fonction $AIC(p)$ et $SC(p)$ de la façon suivante :

$$AIC(p) = \ln \left[\det \hat{\Sigma} \right] + 2 \frac{k^2 p}{T} \quad (2.2)$$

$$FC(p) = \ln \left[\det \hat{\Sigma} \right] + \frac{k^2 p \ln(T)}{T} \quad (2.3)$$

où T est le nombre d'observations, k le nombre de variable du système, $\hat{\Sigma}$ la matrice de variance covariance des résidus estimés du modèle.

2.3. Prévisions

2.3.1. Le cas d'un VAR(1)

Considérons le cas d'un modèle $VAR(1)$ centré tel que $\forall t \in \mathbb{Z}$:

$$X_t = \Phi_0 + \Phi_1 X_{t-1} + \varepsilon_t$$

Supposons que l'on dispose d'une réalisation sur un échantillon de T réalisations (X_1, X_2, \dots, X_T) d'un estimateur convergent $\hat{\Phi}_1$ de Φ_1 et d'un estimateur convergent $\hat{\Phi}_0$ de Φ_0 . La formule qui

permet d'obtenir une prévision de la réalisation à la date $T + 1$ du processus X_t est donc naturellement donnée par :

$$\widehat{X}_{T+1} = E(X_{t+1}/X_T, X_{T-1}, \dots, X_1) = \widehat{\Phi}_0 + \widehat{\Phi}_1 X_t \quad (2.4)$$

A l'horizon $T + 2$, on a :

$$\widehat{X}_{T+2} = E(X_{t+2}/X_T, X_{T-1}, \dots, X_1) = \widehat{\Phi}_0 + \widehat{\Phi}_1 \widehat{X}_{t+1} = (I + \widehat{\Phi}_1) \widehat{\Phi}_0 + \widehat{\Phi}_1^2 X_T \quad (2.5)$$

Proposition 2.1. *De la même façon à un horizon h , la prévision d'un VAR(1) est donnée par :*

$$\widehat{X}_{T+h} = E(X_{T+h}/X_T, X_{T-1}, \dots, X_1) = \left(I + \widehat{\Phi}_1 + \widehat{\Phi}_1^2 + \dots + \widehat{\Phi}_1^{h-1} \right) \widehat{\Phi}_0 + \widehat{\Phi}_1^h X_T$$

Dès lors, l'erreur de prévision s'écrit sous la forme :

$$\begin{aligned} X_{T+h} - \widehat{X}_{T+h} &= X_{T+h} - E(X_{T+h}/X_T, X_{T-1}, \dots, X_1) \\ &= X_{T+h} - E(X_{T+h}/\varepsilon_T, \varepsilon_{T-1}, \dots, \varepsilon_1) \\ &= \sum_{i=0}^{h-1} \Phi_1^i \varepsilon_{T+h-i} \end{aligned}$$

Par définition des bruits blancs, cette erreur de prévision a une espérance nulle. La matrice de variance covariance de l'erreur de prévision est donc :

$$\begin{aligned} &E \left[\left(X_{T+h} - \widehat{X}_{T+h} \right) \left(X_{T+h} - \widehat{X}_{T+h} \right)' / X_T, X_{T-1}, \dots, X_1 \right] \\ &= E \left[\left(\varepsilon_{T+h} + \Phi_1 \varepsilon_{T+h-1} + \dots + \Phi_1^{h-1} \varepsilon_{T+1} \right) \left(\varepsilon_{T+h} + \Phi_1 \varepsilon_{T+h-1} + \dots + \Phi_1^{h-1} \varepsilon_{T+1} \right)' \right] \\ &= \sum_{i=1}^{h-1} \Phi_1^i (\Phi_1^i)' \end{aligned}$$

Proposition 2.2. *Pour un processus VAR(1), la matrice de variance covariance de l'erreur de prévision à un horizon h est déterminée par la relation :*

$$E \left[\left(X_{T+h} - \widehat{X}_{T+h} \right)' \left(X_{T+h} - \widehat{X}_{T+h} \right) / X_T, X_{T-1}, \dots, X_1 \right] = \sum_{i=1}^{h-1} \Phi_1^i (\Phi_1^i)'$$

Les variances des erreurs de prévisions pour les processus univariés $(x_{1,t}, x_{2,t}, \dots, x_{3,t})$ sont déterminés par la diagonale principale de cette matrice.

2.3.2. Le cas d'un VAR(p)

La variance de l'erreur de prévision s'obtient très facilement à partir de la représentation $VMA(\infty)$ d'un VAR d'ordre p quelconque. En effet, si X_t est un processus stationnaire, alors on peut l'écrire sous la forme :

$$X_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} = \Psi(L) \varepsilon_t$$

Dès lors, l'erreur de prévision s'écrit sous la forme :

$$\begin{aligned} X_{T+h} - \widehat{X}_{T+h} &= X_{T+h} - E(X_{T+h}/X_T, X_{T-1}, \dots, X_1) \\ &= X_{T+h} - E(X_{T+h}/\varepsilon_T, \varepsilon_{T-1}, \dots, \varepsilon_1) \\ &= \sum_{i=0}^{h-1} \psi_i \varepsilon_{T+h-i} \end{aligned}$$

Par définition des bruits blancs, cette erreur de prévision a une espérance nulle. La matrice de variance covariance de l'erreur de prévision est donc :

$$\begin{aligned} &E \left[\left(X_{T+h} - \widehat{X}_{T+h} \right) \left(X_{T+h} - \widehat{X}_{T+h} \right)' / X_T, X_{T-1}, \dots, X_1 \right] \\ &= E \left[\left(\varepsilon_{T+h} + \psi_1 \varepsilon_{T+h-1} + \dots + \psi_{h-1} \varepsilon_{T+1} \right) \left(\varepsilon_{T+h} + \psi_1 \varepsilon_{T+h-1} + \dots + \psi_{h-1} \varepsilon_{T+1} \right)' \right] \\ &= \sum_{i=1}^{h-1} \psi_i (\psi_i)' \end{aligned}$$

Proposition 2.3. Pour un processus $VAR(p)$, la matrice de variance covariance de l'erreur de prévision à un horizon h est déterminée par la relation :

$$E \left[\left(X_{T+h} - \widehat{X}_{T+h} \right) \left(X_{T+h} - \widehat{X}_{T+h} \right)' / X_T, X_{T-1}, \dots, X_1 \right] = \sum_{i=1}^{h-1} \psi_i (\psi_i)'$$

Les variances des erreurs de prévisions pour les processus univariés $(x_{1,t}, x_{2,t}, \dots, x_{3,t})$ sont déterminés par la diagonale principale de cette matrice.

3. Dynamique d'un modèle VAR

Les modèles VAR sont souvent analysés au travers de leur dynamique et ce via la simulation de chocs aléatoires et l'analyse de la décomposition de leur variance.

3.1. Analyse des chocs

On considère un processus $\{X_t, t \in \mathbb{Z}\}$, avec $X_t = (x_{1,t}, \dots, x_{n,t})'$ satisfaisant la représentation $VAR(p)$ suivante, $\forall t \in \mathbb{Z}$:

$$\Phi(L)X_t = X_t - \Phi_1 X_{t-1} - \Phi_2 X_{t-2} - \dots - \Phi_p X_{t-p} = c + \varepsilon_t$$

On suppose que les innovations ε_t sont *i.i.d.* $(0, \Sigma)$ et que l'on dispose de $T + p$ réalisations de ce processus. On suppose en outre que les paramètres c, Φ_i sont connus, mais la même analyse peut être menée lorsque l'on ne dispose que d'estimateurs convergents de ces paramètres.

Quelle est l'idée générale de l'analyse des chocs ?

Idée Générale Une fonction de réponse aux innovations résume l'information concernant l'évolution d'une composante $x_{i,t}$ qui intervient suite à une impulsion sur $x_{j,t}$ à la date T , en supposant que toutes les autres variables sont constantes pour $t \leq T$.

3.1.1. Exemple

On considère un processus bi-varié $\{Y_t, t \in \mathbb{Z}\}$ admettant une représentation $VAR(1)$ telle que :

$$\begin{aligned} y_{1,t} &= 3 + 0.2y_{1,t-1} + 0.7y_{2,t-1} + \varepsilon_{1,t} \\ y_{2,t} &= 1 + 0.3y_{1,t-1} + 0.4y_{2,t-1} + \varepsilon_{2,t} \end{aligned}$$

On pose $Y_t = (y_{1,t} \ y_{2,t})'$ et $\varepsilon_t = (\varepsilon_{1,t} \ \varepsilon_{2,t})'$. On suppose que les chocs $\varepsilon_{1,t}$ et $\varepsilon_{2,t}$ sont corrélés. Cette hypothèse est particulièrement importante pour la suite de cet exemple.

$$E(\varepsilon_t \varepsilon_t') = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

On cherche à identifier l'impact d'un choc unitaire sur $y_{2,T}$ à la date T sur la dynamique de la variable $y_{1,t}$ aux périodes postérieures à T , en supposant les évolutions de ces deux variables pour $t \leq T$ connues et données. Cela revient à supposer :

$$\varepsilon_{2,T} = 1 \quad \varepsilon_{2,t} = 0 \quad \forall t > T$$

On cherche donc à déterminer la variation de $y_{1,t}$ engendrée par ce choc. Pour cela considérons la décomposition de Wold du processus Y_t déterminée *infra*:

$$y_{1,t} = \mu_1 + \sum_{i=0}^{\infty} \Psi_{1,i} \varepsilon_t$$

où $\Psi_{1,i}$ désigne la première ligne de la matrice Ψ_i issue de la représentation VMA :

$$Y_t = \Psi(L)\varepsilon_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$$

$$Y_t = \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} = \begin{pmatrix} 9.25 \\ 6.29 \end{pmatrix} + \sum_{i=0}^{\infty} \begin{pmatrix} -0.2 & -0.7 \\ -0.3 & -0.4 \end{pmatrix}^i \begin{pmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{pmatrix}$$

On pourrait penser que suite au choc $\varepsilon_{2,T} = 1$, dans ce cas, la suite des réalisations $y_{1,T+h}$ soit donnée directement par les coefficients correspondants du vecteur $\Psi_{1,i}$. On obtiendrait ainsi une fonction de réponse de la variable y_1 à une impulsion de la variable y_2 . C'est une première possibilité de fonctions de réponse.

Le problème ici c'est qu'en raison de la corrélation des deux chocs, l'impulsion initiale sur $\varepsilon_{2,T}$ n'est pas sans influence sur l'innovation $\varepsilon_{1,T}$ qui entre elle aussi dans la représentation moyenne mobile infinie de $y_{1,t}$. Conditionnellement à la réalisation de $\varepsilon_{2,T}$, du fait de la corrélation des deux chocs, la probabilité d'engendrer une innovations $\varepsilon_{1,T}$ non nulle est elle même non nulle.

Or généralement, ce qui intéressant sur le plan économique c'est d'envisager une impulsion sur la *composante orthogonale* de $\varepsilon_{2,t}$ à $\varepsilon_{1,t}$. C'est à dire, il convient d'isoler l'innovation "propre" au processus $y_{2,t}$ non "polluée" par la réaction de l'innovation $y_{1,t}$. C'est pourquoi, il convient dans le cas général où $E(\varepsilon_t \varepsilon_t') \neq I_n$, d'orthogonaliser les innovations. On considère donc la décomposition suivante de la matrice de covariance des innovations :

$$= ADA'$$

où A est une matrice (2, 2) triangulaire inférieure et où D est une matrice diagonale. Dans notre exemple, on a :

$$= \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \frac{\sigma_{12}}{\sigma_1^2} & 1 \end{pmatrix} \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2} \end{pmatrix} \begin{pmatrix} 1 & \frac{\sigma_{12}}{\sigma_1^2} \\ 0 & 1 \end{pmatrix}$$

D'où dans notre exemple :

$$A = \begin{pmatrix} 1 & 0 \\ 0.5 & 1 \end{pmatrix} \quad D = \begin{pmatrix} 1 & 0 \\ 0 & 0.75 \end{pmatrix}$$

On pose:

$$v_t = A^{-1}\varepsilon_t$$

On remarque alors que les innovations v_t sont des combinaisons linéaires des innovations du modèle initial ε_t qui possèdent une propriété d'indépendance puisque :

$$\begin{aligned} E(v_t v_t') &= A^{-1} E(\varepsilon_t \varepsilon_t') (A^{-1})' \\ &= A^{-1} (A^{-1})' \\ &= A^{-1} ADA' (A')^{-1} \\ &= D \end{aligned}$$

Donc la matrice de variance covariance des innovations v_t est diagonale, ce qui prouve que ces innovations ne sont pas corrélées. Dans notre exemple, il est très simple de constater que cette orthogonalisation correspond à la projection linéaire des $\varepsilon_{2,t}$ sur les $\varepsilon_{1,t}$. Le résidu $v_{2,t}$ correspond à la composante orthogonale des $\varepsilon_{2,t}$.

$$v_{2,t} = \varepsilon_{2,t} - \frac{\sigma_{12}}{\sigma_1^2} \varepsilon_{1,t} = \varepsilon_{2,t} - \frac{\varepsilon_{1,t}}{2}$$

$$E(v_2 \varepsilon_{1,t}') = E(v_2 v_{1,t}') = 0$$

Reprenons la décomposition de Wold associée à Y_t il vient :

$$Y_t = \mu + \Psi(L) \varepsilon_t = \mu + \sum_{i=0}^{\infty} \Phi_1^i \varepsilon_{t-i}$$

Or on pose $\varepsilon_t = Av_t$, dès lors cette représentation $VMA(\infty)$ peut se réécrire en fonction d'innovations v_t non corrélées.

$$Y_t = \tilde{\Psi}(L) \varepsilon_t = \mu + \sum_{i=0}^{\infty} \tilde{\psi}_i v_{t-i} = \mu + \sum_{i=0}^{\infty} (\Phi_1^i A) v_{t-i} \quad (3.1)$$

On obtient donc dans notre exemple :

$$Y_t = \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} = \begin{pmatrix} 9.25 \\ 6.29 \end{pmatrix} + \sum_{i=0}^{\infty} \begin{pmatrix} -0.2 & -0.7 \\ -0.3 & -0.4 \end{pmatrix}^i \begin{pmatrix} 1 & 0 \\ 0.5 & 1 \end{pmatrix} \begin{pmatrix} v_{1,t-1} \\ v_{2,t-1} \end{pmatrix}$$

De façon équivalente on peut réécrire le VAR en fonction des seules innovations orthogonales :

$$\varepsilon_t = Av_t \iff \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0.5 & 1 \end{pmatrix} \begin{pmatrix} v_{1,t} \\ v_{2,t} \end{pmatrix}$$

$$y_{1,t} = 3 + 0.2y_{1,t-1} + 0.7y_{2,t-1} + v_{1,t}$$

$$y_{2,t} = 1 + 0.3y_{1,t-1} + 0.4y_{2,t-1} + 0.5v_{1,t} + v_{2,t}$$

Dès lors, on construit de la même façon la séquence des $y_{1,T+h}$ obtenus conditionnellement à un choc unitaire sur la composante orthogonale $v_{2,T}$. Cela revient à supposer :

$$v_{2,T} = 1 \quad v_{2,t} = 0 \quad \forall t > T$$

Voici la représentation de ces IRF :

Figure 3.1: Fonction de Réponses de y_1

3.1.2. Cas général

On cherche ainsi de façon générale à se ramener à une représentation où les innovations sont orthogonales.

Proposition 3.1. *Dans le cas général où $E(\varepsilon_t \varepsilon_t') = \Sigma \neq I_n$, on orthogonalise les innovations de la façon suivante. On pose :*

$$v_t = A^{-1} \varepsilon_t \quad (3.2)$$

où la matrice A est issue de l'orthogonalisation de Σ :

$$\Sigma = ADA' \quad (3.3)$$

où A est une matrice $(2, 2)$ triangulaire inférieure et où D est une matrice diagonale. On cherche donc à récrire le système VAR non plus en fonction des innovations corrélés ε_t , mais en fonction des innovations orthogonales v_t qui satisfont :

$$E(v_t v_t') = D \text{ matrice diagonale} \quad (3.4)$$

Proposition 3.2. *Cette phase d'orthogonalisation implique toutefois que l'ordre dans lequel sont disposées les variables du VAR affecte l'analyse dynamique et en particulier l'analyse des fonctions de réponse.*

En effet, reprenons l'exemple précédent. On considère un processus bi-varié $\{Y_t, t \in \mathbb{Z}\}$ admettant une représentation $VAR(1)$ telle que :

$$\begin{aligned} y_{1,t} &= 3 + 0.2y_{1,t-1} + 0.7y_{2,t-1} + \varepsilon_{1,t} \\ y_{2,t} &= 1 + 0.3y_{1,t-1} + 0.4y_{2,t-1} + \varepsilon_{2,t} \end{aligned}$$

On suppose que les deux chocs $\varepsilon_{1,t}$ et $\varepsilon_{2,t}$ sont corrélés et ont des variances différentes.

$$cov(\varepsilon_{1,t}, \varepsilon_{2,t}) = \frac{1}{2} \quad \sigma_1^2 = 1 \quad \sigma_2^2 = 2$$

Il existe deux façons d'écrire le VAR , soit on pose $Y_t = (y_{1,t} \ y_{2,t})'$, soit l'on écrit $Y_t = (y_{2,t} \ y_{1,t})'$. Le choix de l'ordre des variables va dès lors conditionner notre schéma d'orthogonalisation :

1. **Cas 1** : on écrit $Y_t = (y_{1,t} \ y_{2,t})'$ et $\varepsilon_t = (\varepsilon_{1,t} \ \varepsilon_{2,t})'$. Dès lors, on pose :

$$E(\varepsilon_t \varepsilon_t') = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 2 \end{pmatrix}$$

Dans ce cas, on a :

$$= ADA' = \begin{pmatrix} 1 & 0 \\ 0.5 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1.75 \end{pmatrix} \begin{pmatrix} 1 & 0.5 \\ 0 & 1 \end{pmatrix}$$

Les innovations orthogonales sont donc définies par

$$\begin{aligned} v_t = A^{-1}\varepsilon_t &\iff \begin{pmatrix} v_{1,t} \\ v_{2,t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0.5 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -0.5 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \\ &\iff \begin{cases} v_{1,t} = \varepsilon_{1,t} \\ v_{2,t} = -\frac{1}{2}\varepsilon_{1,t} + \varepsilon_{2,t} \end{cases} \end{aligned}$$

Dès lors $v_{2,t}$ mesure la composante de $\varepsilon_{2,t}$ orthogonale à $\varepsilon_{1,t}$.

2. **Cas 2** : on écrit $Y_t = (y_{2,t} \ y_{1,t})'$ et $\varepsilon_t = (\varepsilon_{2,t} \ \varepsilon_{1,t})'$. Dès lors, on pose :

$$E(\varepsilon_t \varepsilon_t') = \begin{pmatrix} \sigma_2^2 & \sigma_{12} \\ \sigma_{12} & \sigma_1^2 \end{pmatrix} = \begin{pmatrix} 2 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

Dans ce cas, on a :

$$= ADA' = \begin{pmatrix} 1 & 0 \\ 0.25 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & \frac{7}{8} \end{pmatrix} \begin{pmatrix} 1 & 0.25 \\ 0 & 1 \end{pmatrix}$$

Les innovations orthogonales sont donc définies par :

$$\begin{aligned} v_t = A^{-1}\varepsilon_t &\iff \begin{pmatrix} v_{2,t} \\ v_{1,t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0.25 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \varepsilon_{2,t} \\ \varepsilon_{1,t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -0.25 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{2,t} \\ \varepsilon_{1,t} \end{pmatrix} \\ &\iff \begin{cases} v_{1,t} = -\frac{1}{4}\varepsilon_{2,t} + \varepsilon_{1,t} \\ v_{2,t} = \varepsilon_{2,t} \end{cases} \end{aligned}$$

Dans ce cas, $v_{2,t}$ n'est rien d'autre que $\varepsilon_{2,t}$, qui par nature est corrélé avec $\varepsilon_{1,t}$.

3. En conséquence, dans cet exemple on montre que (i) si l'on désire étudier l'impact d'une innovation "pure" du processus $y_{2,t}$ sur le niveau de $y_{1,t}$, et que (ii) on retient cette méthode d'orthogonalisation, il convient d'écrire le système VAR sous la forme $Y_t = (y_{1,t} \ y_{2,t})'$. Bien entendu, les fonctions de réponse obtenues dans les deux cas ne sont pas identiques.

Résultat *De façon générale dans l'écriture d'un VAR, la variable, que l'on suppose économiquement être la variable explicative, doit figurer après la variable dont on veut expliquer les évolutions.*

La démonstration de ce principe dans le cas général d'un VAR à n variables est donnée dans Hamilton (1994), pages 291 et suivantes.

3.2. Décomposition de la variance

Définition *Partant de la décomposition des résidus en innovations "pures" ou orthogonales, on peut calculer quelle est la contribution de chaque innovation à la variance totale de l'erreur de prévisions du processus $x_{i,t}$. C'est ce que l'on appelle la décomposition de la variance.*

On considère processus $\{X_t, t \in \mathbb{Z}\}$, avec $X_t = (x_{1,t}, \dots, x_{n,t})'$ satisfaisant la représentation VAR(p) suivante, $\forall t \in \mathbb{Z}$:

$$\Phi(L) X_t = X_t - \Phi_1 X_{t-1} - \Phi_2 X_{t-2} - \dots - \Phi_p X_{t-p} = c + \varepsilon_t$$

On suppose que les innovations ε_t sont *i.i.d.* (0, Σ) On suppose que ce processus est stationnaire et peut être représenté sous la forme d'un VMA(∞) :

$$X_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} = \Psi(L) \varepsilon_t$$

avec $\psi_i = I_n$ L'erreur de prévision à l'horizon h s'écrit est :

$$\begin{aligned} X_{t+h} - \widehat{X}_{T+h} &= X_{t+h} - E(X_{t+h}/X_T, X_{T-1}, \dots, X_1) \\ &= X_{t+h} - E(X_{t+h}/\varepsilon_T, \varepsilon_{T-1}, \dots, \varepsilon_1) \\ &= \sum_{i=0}^{h-1} \psi_i \varepsilon_{T+h-i} \end{aligned}$$

Par définition des bruits blancs, cette erreur de prévision a une espérance nulle. La matrice de variance covariance de l'erreur de prévision est donc :

$$E \left[\left(X_{t+h} - \widehat{X}_{T+h} \right) \left(X_{t+h} - \widehat{X}_{T+h} \right)' \right] = \sum_{i=0}^{h-1} \psi_i (\psi_i)'$$

Cette erreur de prévision est donc exprimée en fonction de la matrice de variance covariance non diagonale des résidus ε_t .

Pour obtenir une décomposition de la variance du vecteur $X_t = (x_{1,t}, \dots, x_{n,t})'$ il suffit de réexprimer cette matrice de variance covariance sous la forme d'une combinaison linéaire des variances des innovations orthogonales v_t .

$$v_t = A^{-1}\varepsilon_t \iff \varepsilon_t = Av_t \quad (3.5)$$

où la matrice A est issue de l'orthogonalisation de :

$$= ADA'$$

On suppose que $\forall t \in \mathbb{Z}$:

$$\varepsilon_t = \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \dots \\ \varepsilon_{n,t} \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ (n,1) & (n,1) & & (n,1) \end{pmatrix} \begin{pmatrix} v_{1,t} \\ v_{2,t} \\ \dots \\ v_{n,t} \end{pmatrix}$$

où a_i désigne la $i^{\text{ème}}$ colonne de la matrice A . Dès lors :

$$\begin{aligned} &= E(\varepsilon_t \varepsilon_t') = a_1 a_1' \text{var}(v_{1,t}) + a_2 a_2' \text{var}(v_{2,t}) + \\ &+ \dots + a_n a_n' \text{var}(v_{n,t}) \end{aligned} \quad (3.6)$$

En substituant cette expression dans la variance de la prévision pour un horizon h , cela donne permet de réexprimer cette variance en fonction de la variance des innovations "orthogonales" :

$$\begin{aligned} &E \left[\left(X_{t+h} - \widehat{X}_{T+h} \right) \left(X_{t+h} - \widehat{X}_{T+h} \right)' \right] \\ &= + \sum_{i=1}^{h-1} \psi_i (\psi_i)' \\ &= \sum_{j=1}^n \left\{ \text{var}(v_{j,t}) \sum_{i=0}^{h-1} \psi_i [a_j a_j'] (\psi_i)' \right\} \end{aligned}$$

Proposition 3.3. *A partir de cette formule, on est en mesure de calculer la contribution d'une innovation pure $v_{j,t}$ à la variance totale de la prévision à un horizon h :*

$$\text{var}(v_{j,t}) \left[a_j a_j + \psi_1 [a_j a_j'] (\psi_1)' + \dots + \psi_{h-1} [a_j a_j'] (\psi_{h-1})' \right] \quad (3.7)$$

4. La causalité

Une des questions que l'on peut se poser à partir d'un VAR est de savoir s'il existe une relation de causalité entre les différentes variables du système. Il existe ainsi plusieurs définitions de la causalité :

- causalité au sens de Granger
- causalité au sens de Sims

Nous nous limiterons à l'exposé de la causalité au sens de Granger qui est la plus fréquemment utilisée en économétrie. On se restreint au cas d'un processus *bi-varié* ($n = 2$) que l'on

$$Z_t = \begin{pmatrix} y_t \\ x_t \end{pmatrix}$$

4.1. Causalité au sens de Granger

La question est de savoir si la variable x "cause" ou non la variable y .

Definition 4.1. On dit que la variable x cause au sens de Granger la variable y si et seulement si la connaissance du passé de x améliore la prévision de y à tout horizon.

De cette définition découle un corollaire :

Corollary 4.2. On dit que la variable x **ne cause pas** la variable y au sens de Granger, si et seulement si :

$$E(y_{t+h}/y_t, y_{t-1}, \dots, y_1) = E(y_{t+h}/y_t, y_{t-1}, \dots, y_1, x_t, x_{t-1}, \dots, x_1)$$

De façon équivalente, on dit alors que la variable y est exogène au sens des séries temporelles.

4.1.1. Application au cas d'un $VAR(p)$ avec $n = 2$

Pour un $VAR(p)$ avec $n = 2$ la condition de la causalité de Granger est immédiate à obtenir.

Résultat Dans le système bi-varié suivant $VAR(p) : \forall t \in \mathbb{Z}$

$$\begin{aligned} Z_t = \begin{pmatrix} y_t \\ x_t \end{pmatrix} &= \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} \phi_{11}^1 & \phi_{12}^1 \\ \phi_{21}^1 & \phi_{22}^1 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} \phi_{11}^2 & \phi_{12}^2 \\ \phi_{21}^2 & \phi_{22}^2 \end{pmatrix} \begin{pmatrix} y_{t-2} \\ x_{t-2} \end{pmatrix} \\ &+ \dots + \begin{pmatrix} \phi_{11}^p & \phi_{12}^p \\ \phi_{21}^p & \phi_{22}^p \end{pmatrix} \begin{pmatrix} y_{t-p} \\ x_{t-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_{y,t} \\ \varepsilon_{x,t} \end{pmatrix} \end{aligned} \quad (4.1)$$

la variable x_t ne cause pas la variable y_t si et seulement si :

$$\phi_{12}^1 = \phi_{12}^2 = \phi_{12}^3 = \dots = \phi_{12}^p = 0 \quad (4.2)$$

Autrement dit, la variable x_t ne cause pas la variable y_t si et seulement si les matrices Φ_i sont toutes triangulaires inférieures pour $\forall i \in [1, p]$.

En effet, réécrivons le processus sous cette condition, on a :

$$\begin{aligned} Z_t = \begin{pmatrix} y_t \\ x_t \end{pmatrix} &= \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} \phi_{11}^1 & 0 \\ \phi_{21}^1 & \phi_{22}^1 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} \phi_{11}^2 & 0 \\ \phi_{21}^2 & \phi_{22}^2 \end{pmatrix} \begin{pmatrix} y_{t-2} \\ x_{t-2} \end{pmatrix} \\ &+ \dots + \begin{pmatrix} \phi_{11}^p & 0 \\ \phi_{21}^p & \phi_{22}^p \end{pmatrix} \begin{pmatrix} y_{t-p} \\ x_{t-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_{y,t} \\ \varepsilon_{x,t} \end{pmatrix} \end{aligned} \quad (4.3)$$

Dès lors,

$$E(y_{t+h}/y_t, y_{t-1}, \dots, y_1) = c_1 + \phi_{11}^1 y_t + \phi_{11}^2 y_{t-1} + \dots + \phi_{11}^p y_{t-p+1}$$

$$E(y_{t+h}/y_t, y_{t-1}, \dots, y_1, x_t, x_{t-1}, \dots, x_1) = c_1 + \phi_{11}^1 y_t + \phi_{11}^2 y_{t-1} + \dots + \phi_{11}^p y_{t-p+1}$$

On a bien alors :

$$E(y_{t+h}/y_t, y_{t-1}, \dots, y_1) = E(y_{t+h}/y_t, y_{t-1}, \dots, y_1, x_t, x_{t-1}, \dots, x_1)$$