

LES PROCESSUS VAR

notes du cours de séries temporelles multivariées

Gilbert COLLETAZ

24 janvier 2019

Résumé

Cours de séries temporelles multivariées - Semestre 2 - Master ESA.
Ces pages ont maintenant bénéficié de la relecture des étudiants, en conséquence, les erreurs présentes sont de leur seule responsabilité.

Table des matières

1	Introduction	2
1.1	L'écriture VMA	2
1.1.1	Le VMA(q)	3
1.1.2	Le VMA(∞)	4
2	Le processus VAR(p)	4
2.1	Écritures VMA ET VAR, stabilité, stationarité	5
2.2	Écriture d'un VAR(p) comme VAR(1), matrice companion	5
2.3	Les prévisions tirées d'un VAR	7
2.4	L'estimation d'un VAR	10
2.4.1	Estimation par le maximum de vraisemblance conditionnel	10
2.4.2	L'estimation d'un VAR(p) sous Proc VARMAX dans SAS 9.3	18
2.4.3	Estimation SURE de Zellner	20
2.4.4	La détermination de l'ordre du VAR	24
2.5	Les outils de validation	33
2.6	Les fonctions de réponse aux chocs	38
2.6.1	Les fonctions de réponse simples	38
2.6.2	Les fonctions de réponse orthogonalisées	39
2.6.3	L'interprétation des chocs orthogonalisés et les consé- quences de cette orthogonalisation	40
2.6.4	Intervalle de confiance sur les fonctions de réponses	43
2.6.5	L'obtention des fonctions de réponse et de leurs écart- types dans VARMAX	45

2.7	La décomposition de la variance des erreurs de prévisions . . .	47
2.8	La causalité au sens de Granger	49
2.8.1	Causalité dans un système bivarié	51
2.8.2	Causalité entre deux ensembles de variables : les mesures de GEWEKE	56
2.9	Un exemple	62
3	Une introduction aux VAR structurels	76
3.1	un exemple	77
3.2	Les difficultés d'estimation des VAR structurels et le recours aux VAR	77
3.3	du VAR au VAR structurel : les conditions d'identification . . .	78
3.3.1	Une relecture de la décomposition de cholesky	80

1 Introduction

Lorsque l'on veut étudier les interdépendances entre plusieurs variables il est évidemment nécessaire de se situer dans un cadre multivarié. Fort heureusement, une bonne connaissance des concepts développés dans le cadre univarié facilite grandement la compréhension des outils alors utilisés. Comme lors du premier semestre, nous supposerons dans un premier temps que les processus étudiés sont stationnaires. Cela permet d'introduire, dans un cadre relativement simple, les notions utiles pour traiter des dépendances dynamiques entre plusieurs variables telles que les prédicteurs avancés (causalité au sens de Granger et les mesures de dépendance linéaires de Geweke), la réaction du système aux chocs d'innovations (fonctions de réponse), la contribution des chocs à la variance des erreurs de prévision (décomposition de la variance). Dans un second temps nous considérerons le cas des processus intégrés avec l'introduction du concept de cointégration et les modèles à correction d'erreur.

1.1 L'écriture VMA

Soit donc $y_t = (y_{1t}, y_{2t}, \dots, y_{kt})^\top$ un ensemble de k aléatoires. Sous hypothèse de stationnarité au sens large, les deux premiers moments de y_t sont :

$$E(y_t) = \mu, \text{ un vecteur de } k \text{ constantes, et}$$

$$\Sigma_{y_j} = E[(y_t - \mu)(y_{t-j} - \mu)^\top], j = 0, 2, \dots \text{ des matrices } (k \times k) \text{ de constantes.}$$

Naturellement Σ_{y_0} est symétrique, mais ce n'est en général pas le cas de Σ_{y_j} pour $j \neq 0$. Par exemple, la dépendance de y_{1t} à $y_{2,t-j}$ n'a pas de raison d'être

égale à la dépendance de y_{2t} à $y_{1,t-j}$. Plus précisément :

$$\begin{aligned}\Sigma_{y_j} &= E[(y_t - \mu)(y_{t-j} - \mu)^\top] \\ &= E[(y_{t+j} - \mu)(y_t - \mu)^\top] \text{ en raison de la stationnarité} \\ \Rightarrow \Sigma_{y_j}^\top &= E[(y_t - \mu)(y_{t+j} - \mu)^\top] = \Sigma_{y_{-j}}\end{aligned}$$

Sous hypothèse de stationnarité, le théorème de Wold multivarié assure que y_t peut s'écrire comme la somme d'une composante déterministe et d'une somme pondérée éventuellement infinie sur un processus en bruit blanc vectoriel :

$$y_t = \mu + \sum_{i=0}^{\infty} \Psi_i u_{t-i} = (I + \Psi_1 L + \Psi_2 L^2 + \dots) u_t$$

avec

$$E(u_t) = 0,$$

et

$$E(u_t u_{t-s}^\top) = \begin{cases} \Sigma_u, & \text{si } s = 0, \\ 0 & \text{sinon} \end{cases}$$

Notez que Σ_u n'est pas nécessairement diagonale : les aléatoires centrées qui composent u_t peuvent être linéairement dépendantes entre elles en instantané mais ne dépendent pas des composantes de u passées ou futures.

1.1.1 Le VMA(q)

Il peut arriver que l'écriture en moyenne mobile de y_t soit telle que $\Psi_i = 0$ si $i > q$. Dans ce cas on est en présence d'un VMA(q) :

$$y_t = \mu + u_t + \Psi_1 u_{t-1} + \Psi_2 u_{t-2} + \dots + \Psi_q u_{t-q} \quad (1)$$

On montre qu'alors¹ :

$$\Sigma_{y_0} = \Sigma_u + \Psi_1 \Sigma_u \Psi_1^\top + \Psi_2 \Sigma_u \Psi_2^\top + \dots + \Psi_q \Sigma_u \Psi_q^\top$$

$$\Sigma_{y_j} = \begin{cases} \Psi_j \Sigma_u + \Psi_{j+1} \Sigma_u \Psi_1^\top + \Psi_{j+2} \Sigma_u \Psi_2^\top + \dots + \Psi_q \Sigma_u \Psi_{q-j}^\top & \text{pour } j = 1, 2, \dots, q \\ \Sigma_u \Psi_{-j}^\top + \Psi_1 \Sigma_u \Psi_{-j+1}^\top + \Psi_2 \Sigma_u \Psi_{-j+2}^\top + \dots + \Psi_{q+j} \Sigma_u \Psi_q^\top, & \text{pour } j = -1, -2, \dots, -q \\ 0 & \text{pour } |j| > q \end{cases}$$

Compte tenu des résultats précédents, on remarque que tout processus VMA(q) est stationnaire au sens large, ses deux premiers moments étant indépendants du temps.

1. Je vous invite, à titre d'exercice, à retrouver ces résultats

1.1.2 Le VMA(∞)

Lorsque l'on est en présence d'un VMA d'ordre infini, il est nécessaire d'imposer une condition de convergence similaire à celle déjà vue dans le cadre univarié. Une condition suffisante pour que les moments d'ordre deux de y_t existent est que la suite de matrice $\{\Psi_{y_j}\}_{j=0}^\infty$ soit absolument sommable²

En effet³ :

$$\Sigma_{y_j} = \sum_{s=0}^{\infty} \Psi_{j+s} \Sigma_u \Psi_s^T, \quad j = 0, 1, 2, \dots \quad (2)$$

Comme dans le cas univarié, on impose également une hypothèse d'ergodicité permettant d'appliquer les résultats de la loi des grands nombres de sorte que lorsque la taille de l'échantillon, T , tend vers l'infini, on ait la convergence en probabilité des moments empiriques vers les moments théoriques, ainsi :

$$\frac{1}{T} \sum_{t=1}^T y_{it} \xrightarrow{P} \mu \quad (3)$$

$$\forall s, \frac{1}{T} \sum_{t=1}^T y_{it} y_{jt-s} \xrightarrow{P} E(y_{it} y_{jt-s}) \quad (4)$$

2 Le processus VAR(p)

$y_t = (y_{1t}, \dots, y_{kt})^T$, obéit à un processus AutoRégressif Vectoriel d'ordre p , ou VAR(p), si on a :

$$y_t = c + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \dots + \Phi_p y_{t-p} + u_t \quad (5)$$

où u_t est un processus en bruit blanc vérifiant donc $E(u_t) = 0$ et $E(u_t u_{t-s}^T) = \Sigma_u$ si $s = 0$ et 0 sinon.

Ainsi, avec $k = 2$ et $p = 1$:

$$y_{1t} = c_1 + \phi_{11} y_{1t-1} + \phi_{12} y_{2t-1} + u_{1t} \quad (6)$$

$$y_{2t} = c_2 + \phi_{21} y_{1t-1} + \phi_{22} y_{2t-1} + u_{2t} \quad (7)$$

$$\Sigma_u = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} E(u_{1t}^2) & E(u_{1t} u_{2t}) \\ E(u_{1t} u_{2t}) & E(u_{2t}^2) \end{bmatrix} \quad (8)$$

Cette structure simple et d'estimation aisée, permet de construire un certain nombre d'outils qui se sont avérés utiles notamment pour la construction de

2. Une suite de matrices est absolument sommable si chacun de ses éléments définit une suite de scalaires absolument sommable, soit ici $\sum_{s=0}^{\infty} |\Psi_{s_{ij}}| < \infty$, où $\Psi_{s_{ij}}$ est l'élément en ligne i , colonne j de Ψ_s .

3. Je vous invite là encore à retrouver ce résultat

prévisions, l'étude des relations dynamiques entre variables et de la propagation des chocs au sein d'un système économique. Le travail fondateur qui a popularisé cette approche VAR est l'article de Sims, *Macroeconomics and reality*, paru en 1980 dans la revue *Econometrica*. Avant d'aborder ces différents points, nous allons préciser les liens entre les écritures VMA et VAR afin notamment d'explicitier les conditions de stationnarité ;

2.1 Écritures VMA ET VAR, stabilité, stationnarité

En recourant à l'opérateur de décalage L , un VAR s'écrit encore :

$$y_t = c + \Phi_1 L y_t + \Phi_2 L^2 y_t + \dots + \Phi_p L^p y_t + u_t, \quad (9)$$

ou :

$$\Phi(L)y_t = c + u_t, \text{ et} \quad (10)$$

$$\Phi(L) = I - \Phi_1 L - \Phi_2 L^2 - \dots - \Phi_p L^p \quad (11)$$

avec toujours $E(u_t) = 0, E(u_t u_t^T) = \Sigma_u$, non nécessairement diagonale, et $E(u_t u_{t-s}^T) = 0$ si $s \neq 0$. Si $\phi(L)$ est inversible alors le VAR est qualifié de stable et il vient :

$$y_t = |\phi(L)|^{-1} \Phi(L)^* [c + u_t] = \mu + |\phi(L)|^{-1} \Phi(L)^* u_t = \mu + \Psi(L) u_t \quad (12)$$

où $\mu = E[y_t] = \Phi(L)^{-1} c$, et $\Phi(L)^*$ est la matrice adjointe de $\Phi(L)$. Cette dernière étant constituée de polynômes en L de degré p , son déterminant, $|\phi(L)|$, est aussi un polynôme en L dont l'inverse si il existe, est un polynôme en L de degré ∞ . En conséquence $\Psi(L)$ est de degré ∞ : l'équation précédente est ainsi l'écriture de y_t sous la forme d'un $VMA(\infty)$ équivalente à son écriture $VAR(p)$ initiale. La seule condition autorisant le passage du $VAR(p)$ vers le $VMA(\infty)$ est donc une condition d'inversion du polynôme $|\phi(L)|$. On sait⁴ que cette condition est que les racines de $|\phi(L)|$ soient, en module, supérieures à l'unité. Si elle est vérifiée, le VAR est dit stable.

Il reste à montrer que la séquence de matrices $\Psi_j, j = 0, \dots, \infty$, tirées de $\Phi(L)^{-1} = \Psi(L) = (\Psi_0 + \Psi_1 L + \Psi_2 L^2 + \dots)$ est absolument sommable ce qui assurerait la stationnarité. Si tel est le cas alors la condition d'inversion de $|\phi(L)|$ est également une condition de stationnarité pour y_t . Ce résultat est plus simplement montré lorsque l'on considère une réécriture alternative d'un $VAR(p)$ sous forme de $VAR(1)$.

2.2 Écriture d'un VAR(p) comme VAR(1), matrice companion

Partant de l'écriture (5), il vient :

$$\begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{pmatrix} = \begin{pmatrix} c \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_p \\ I & 0 & \dots & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & I & 0 \end{bmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{pmatrix} + \begin{pmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (13)$$

4. Cf. cours du premier semestre.

où I est la matrice identité de dimensions $(k \times k)$. Soit encore :

$$\begin{matrix} z_t & = & \tilde{c} & + & F & z_{t-1} & + & \epsilon_t \\ (kp \times 1) & & (kp \times 1) & & (kp \times kp) & (kp \times 1) & & (kp \times 1) \end{matrix} \quad (14)$$

Le système (13) est constitué de l'écriture initiale du VAR(p) complété par $p - 1$ identités : il s'agit donc d'écritures équivalentes. En particulier, si le VAR(p) décrit par (5) est stable, alors le VAR(1) de (13) est également stable et réciproquement.

La matrice F est appelée matrice companion du VAR(p). Un des intérêts de (13) est de faciliter l'obtention de certains résultats, et notamment celui qui nous intéresse maintenant.

On opère des substitutions répétées dans la dernière égalité :

$$\begin{aligned} z_t &= \tilde{c} + Fz_{t-1} + \epsilon_t \\ &= \tilde{c} + F(\tilde{c} + Fz_{t-2} + \epsilon_{t-1}) + \epsilon_t \\ &= (I + F)\tilde{c} + F^2z_{t-2} + F\epsilon_{t-1} + \epsilon_t \\ &\vdots \\ &= (I + F + F^2 + \dots + F^{i-1})\tilde{c} + F^i z_{t-i} + \sum_{j=0}^{i-1} F^j \epsilon_{t-j} \end{aligned}$$

Soit Λ la matrice diagonale constituée des valeurs propres de F et V celle formée par les vecteurs propres associés⁵. Si $|\lambda_j| < 1$, $j = 1, \dots, kp$ alors :

— $F^i = V\Lambda^i V^{-1}$: les éléments des matrices F^i sont constitués de combinaisons linéaires de termes du type λ_j^i , $j = 1, \dots, kp$. En conséquence,

$$\lim_{i \rightarrow \infty} F^i = \lim_{i \rightarrow \infty} V\Lambda^i V^{-1} = 0, \text{ et}$$

— La suite de matrices F^i est absolument sommable,
 — $\sum_{j=0}^{i-1} F^j \epsilon_{t-j}$ définit bien une variable aléatoire lorsque $i \rightarrow \infty$

En d'autres termes, pour qu'un processus gouverné par un VAR soit stable, il faut que les racines de $|I - \Phi_1 L - \Phi_2 L^2 - \dots - \Phi_p L^p|$ soient situées en dehors du cercle complexe unitaire, ou, de façon équivalente, que les valeurs propres de sa matrice companion soient de module inférieur à l'unité⁶. Par ailleurs, la stabilité implique la stationnarité⁷.

5. Pour mémoire, ceux-ci sont orthogonaux entre eux et de norme unitaire.

6. Notez que les estimateurs de ces valeurs propres peuvent être affichés par la Proc VARMAX en utilisant l'option ROOTS de la commande MODEL qui gouverne l'estimation d'un VAR(p) et qui sera détaillée plus loin

7. L'inverse n'est pas vrai : stabilité \Rightarrow stationnarité, mais stationnarité \nRightarrow stabilité.

Notez encore que l'on peut aisément passer des coefficients du VAR contenus dans les matrices Φ_j à ceux du VMA correspondant. En effet, partant de l'identité $\phi(L)\phi(L)^{-1} = I$, il vient :

$$(I - \Phi_1 L - \Phi_2 L^2 - \dots - \Phi_p L^p)(\Psi_0 + \Psi_1 L + \Psi_2 L^2 + \Psi_3 L^3 + \dots) = I \quad (15)$$

En distribuant le produit à gauche puis en égalisant les coefficients afférents à la même puissance de L à gauche et à droite, on a :

$$\Psi_0 = I \quad (16)$$

$$\Psi_1 = \Phi_1 \quad (17)$$

$$\Psi_2 = \Phi_1 \Psi_1 + \Phi_2, \text{ et plus généralement, pour tout } j, \quad (18)$$

$$\Psi_j = \Phi_1 \Psi_{j-1} + \Phi_2 \Psi_{j-2} + \dots + \Phi_p \Psi_{j-p} \quad (19)$$

2.3 Les prévisions tirées d'un VAR

Soit ${}_t y_{t+h}$ la prévision faite en t de la valeur de y en $t+h$, y_{t+h} . On sait que l'espérance conditionnelle en t de y_{t+h} , soit $E_t[y_{t+h}]$ minimise l'erreur quadratique moyenne : $E[y_{t+h} - {}_t y_{t+h}]^2$ est minimum lorsque ${}_t y_{t+h} = E_t[y_{t+h}]$. En ce sens, des prévisions définies par les espérances conditionnelles sont optimales⁸. Ainsi, si on part de l'écriture usuelle d'un VAR(p) :

$$y_{t+1} = c + \Phi_1 y_t + \Phi_2 y_{t-1} + \dots + \Phi_p y_{t-p+1} + u_{t+1}$$

et comme $E_t[u_{t+1}] = 0$, une prévision optimale en t à un horizon de une période est :

$${}_t y_{t+1} = E_t[y_{t+1}] = c + \Phi_1 y_t + \Phi_2 y_{t-1} + \dots + \Phi_p y_{t-p+1} \quad (20)$$

A l'horizon $h = 2$, on a :

$$\begin{aligned} {}_t y_{t+2} &= E_t[y_{t+2}] = c + \Phi_1 E_t[y_{t+1}] + \Phi_2 y_t + \dots + \Phi_p y_{t-p+2} \\ &= c + \Phi_1 {}_t y_{t+1} + \Phi_2 y_t + \dots + \Phi_p y_{t-p+2} \end{aligned} \quad (21)$$

En itérant il est donc aisé de construire le prédicteur pour un horizon h quelconque à partir des prévisions construites pour les horizons inférieurs, $h-1, h-2, \dots, 1$:

$${}_t y_{t+h} = c + \sum_{i=1}^p \Phi_i {}_t y_{t+h-i} \quad (22)$$

où ${}_t y_{t+h-i} = y_{t+h-i}$ si $h \leq i$.

8. En fait, pour être optimales, il faut aussi que la fonction de perte associée aux erreurs de prévisions soit quadratique.

La variance des erreurs de prévisions, est aisément obtenue à partir de l'écriture VMA associée au VAR stable :

$$\Phi(L)y_t = c + u_t \Rightarrow y_t = \mu + \Phi(L)^{-1}u_t = c + (\Psi_0 + \Psi_1L + \Psi_2L^2 + \dots)u_t \quad (23)$$

Le vecteur des erreurs de prévision commises à l'horizon h est ainsi donné par :

$$e_{t+h} = y_{t+h} - E_t[y_{t+h}] = \sum_{i=0}^{h-1} \Psi_i u_{t+h-i} \quad (24)$$

En conséquence, ces erreurs vérifient $E_t[e_{t+h}] = 0$ pour tout horizon h et toute date t , et leur matrice de variance-covariance $\Sigma_e(h)$ est :

$$\Sigma_e(h) = \text{MSE}[e_{t+h}] = \sum_{i=0}^{h-1} \Psi_i \Sigma_u \Psi_i^\top \quad (25)$$

On peut également montrer que $\lim_{h \rightarrow \infty} E_t[y_{t+h}] = E[y_t]$ et $\lim_{h \rightarrow \infty} \Sigma_e(h) = \Sigma_{y_0}$: les prévisions convergent vers l'espérance non conditionnelle de y et la variance des erreurs tend vers la variance non conditionnelle de y lorsque l'horizon de prévision tend vers l'infini.

En pratique, les coefficients $\Phi_i, i = 1, \dots, p$ doivent être estimés, et les prévisions sont calculées non pas au moyen de (22), mais de :

$${}_t\hat{y}_{t+h} = \hat{c} + \sum_{i=1}^p \hat{\Phi}_i {}_t\hat{y}_{t+h-i} \quad (26)$$

avec toujours ${}_t\hat{y}_{t+h-i} = y_{t+h-i}$ si $h \leq i$, et où $\hat{\Phi}_i$ est l'estimation de $\Phi_i, i = 1, \dots, p$. Dans ces conditions, les erreurs de prévisions empiriques sont données par :

$$\begin{aligned} \hat{e}_{t+h} &= y_{t+h} - {}_t\hat{y}_{t+h} \\ &= (y_{t+h} - {}_t y_{t+h}) + ({}_t y_{t+h} - {}_t\hat{y}_{t+h}) \\ &= \sum_{i=0}^{h-1} \Psi_i u_{t+h-i} + ({}_t y_{t+h} - {}_t\hat{y}_{t+h}) \end{aligned} \quad (27)$$

Elles ont donc deux sources : d'une part les innovations qui se produiront sur l'horizon de prévision et d'autre part l'erreur d'estimation des coefficients du VAR, représentée par le terme $({}_t y_{t+h} - {}_t\hat{y}_{t+h})$. En conséquence,

$$\widehat{\Sigma}_e(h) = \text{MSE}[\hat{e}_{t+h}] \neq \Sigma_e(h) = \text{MSE}[e_{t+h}] \quad (28)$$

Généralement la seconde source d'erreur est ignorée : on fait comme si les coefficients estimés étaient les vrais et les variances covariances des erreurs empiriques sont calculées selon :

$$\widehat{\Sigma}_e(h) = \text{MSE}[\hat{e}_{t+h}] = \sum_{i=0}^{h-1} \widehat{\Psi}_i \widehat{\Sigma}_u \widehat{\Psi}_i^\top \quad (29)$$

A partir de (25), à condition de connaître la loi des erreurs⁹ et en remplaçant les matrices Σ_u et Ψ_j par leurs estimations, il est possible d'afficher des intervalles de confiance pour les prévisions ponctuelles issues de (22) si on connaît la loi des résidus. Par défaut, la plupart des logiciels font alors une hypothèse de normalité sur ces derniers.

Avec VARMAX, la commande `output lead=h` où h est un entier positif demande le calcul de prévisions sur un horizon de h périodes.

L'affichage des matrices de variance covariance estimée, $\hat{\Sigma}_e(h)$ est gérée par l'option `PRINT=COVPE(entier)` de la commande `MODEL` servant à réclamer l'estimation d'un VAR¹⁰ où `entier` est un entier positif qui précise l'horizon jusqu'auquel ces informations sont demandées. Par exemple,

```
PROC VARMAX DATA=...;
MODEL y1 y2 y3 / P=2 print=COVPE(8);
OUTPUT LEAD=24;
run;
```

demande l'impression des 24 valeurs prévues par le modèle estimé, la première étant celle construite pour la période suivant immédiatement la dernière observation de l'échantillon d'estimation. Si T est le nombre d'observations, on obtiendra donc : $(T\hat{y}_{1,T+1,T} \hat{y}_{2,T+1,T} \hat{y}_{3,T+1,T})^\top, \dots, (T\hat{y}_{1,T+24,T} \hat{y}_{2,T+24,T} \hat{y}_{3,T+24,T})^\top$. Dans cet exemple, on demande également l'affichage des matrices de variance covariance estimées pour les erreurs de prévision commises sur les trois variables y_1, y_2, y_3 pour les horizons 1,2...8. Cette commande `output` est relativement flexible grâce ensemble d'options qui lui sont attachées :

- `ALPHA= α` , avec $0 < \alpha < 1$, qui demande l'affichage des bornes d'un intervalle de confiance à $100(1 - \alpha)\%$ autour des valeurs prévues,
- `BACK=d`, où d est un entier positif qui demande à ce que le calcul des prévisions se fasse à partir de $T - d$ périodes, c'est à dire d périodes avant celle afférente à la dernière observation de l'échantillon. Par défaut, `d=0` : les prévisions sont calculées à partir de T . C'est ce qui est obtenu dans l'exemple précédent.
- `OUT= nom d'une table SAS` : entraîne la création d'une table SAS dont le nom est celui spécifié et dans laquelle vont être stockées les prévisions réclamées.

Il est aussi possible d'obtenir des graphiques présentant à la fois les valeurs in-sample et les valeurs prévues. Ils sont générés par l'option `PLOTS` qui apparaît dans la ligne d'appel de la Proc VARMAX. Ainsi :

- `PROC VARMAX ... PLOTS=(FORECASTS)` ;
graphique les observations des variables constituant y_t , leurs valeurs calculées pendant la période d'estimation, qui correspondent aux prévisions

9. La chose est simple si on suppose notamment que u_t est un bruit blanc gaussien : selon (24), les erreurs sont des combinaisons linéaires de gaussiennes et sont donc elles-même gaussiennes.

10. Commande qui sera examinée dans la section suivante.

in-sample à une période¹¹ ${}_{t-1}\hat{y}_{i,t+1}$, avec $t = 1, 2, \dots, T - 1$ et $i = 1, 2, \dots, k$, ainsi que les prévisions réalisées selon les instructions précisées par LEAD= et BACK=, avec une représentation de l'intervalle de confiance correspondant à ALPHA=.

- PROC VARMAX . . . PLOTS=(FORECASTSONLY);
 graphe seulement les prévisions réclamées par LEAD=h et BACK=d, soit :

$$\begin{pmatrix} (T-d)\hat{y}_{1,T-d+1}, (T-d)\hat{y}_{2,T-d+1}, \dots, (T-d)\hat{y}_{k,T-d+1} \\ \vdots \\ (T-d)\hat{y}_{1,T-d+h}, (T-d)\hat{y}_{2,T-d+h}, \dots, (T-d)\hat{y}_{k,T-d+h} \end{pmatrix}^T$$

ainsi que leurs intervalles de confiance.

- PROC VARMAX . . . PLOTS=(ALL);
 produit l'ensemble des graphiques générés par les deux options précédentes.

On aura par exemple :

```
PROC VARMAX DATA=... PLOTS=(FORECASTS);
MODEL y1 y2 y3 / P=2;
OUTPUT LEAD=4 ALPHA=0.10;
RUN;
```

Ce programme réclame le calcul de prévisions pour les quatre premières dates suivant immédiatement la dernière période de l'échantillon d'estimation avec leurs intervalles de confiance à 90%, ainsi que des graphes présentant les valeurs observées et calculées in-sample et les valeurs prévues avec indication des intervalles pour chacune des trois variables y_1 , y_2 et y_3 .

2.4 L'estimation d'un VAR

Nous allons voir deux méthodes d'estimation : par le maximum de vraisemblance d'une part, par une estimation SURE d'autre part.

2.4.1 Estimation par le maximum de vraisemblance conditionnel

On reprend ici intégralement la démonstration donnée par Hamilton¹².
 On part de l'écriture usuelle du VAR :

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

Soit encore, en notant \tilde{Y}_t les valeurs des explicatives de y_t et β leur coefficients :

11. Ces prévisions in-sample à une période vérifient naturellement $\hat{u}_t = y_t - {}_{t-1}\hat{y}_{i,t}$
 12. *Time Series Analysis*, 1994, Princeton University Press, pp. 291-293.

$$\begin{matrix} \bar{Y}_t \\ (1+kp, 1) \end{matrix} = \begin{bmatrix} 1 \\ y_{t-1} \\ \vdots \\ y_{t-p} \end{bmatrix} \quad (30)$$

$$\begin{matrix} \beta^\top \\ (k, 1+kp) \end{matrix} = [c, \Phi_1, \dots, \Phi_p] \quad (31)$$

$$y_t = \beta^\top \bar{Y}_t + u_t$$

où u_t est supposé être un processus en bruit blanc gaussien de dimension $(k \times 1)$ possédant une matrice de variance-covariance Σ_u . On rappelle que la densité jointe de u_t est alors donnée par :

$$f(u_t) = (2\pi)^{-k/2} |\Sigma_u^{-1}|^{1/2} \exp\left(-\frac{1}{2} u_t^\top \Sigma_u^{-1} u_t\right) \quad (32)$$

Dans ces conditions, y_t défini par le processus d'innovation u_t est lui-même gaussien. On note ainsi immédiatement que :

$$y_t | \bar{Y}_t \sim N(\beta^\top \bar{Y}_t, \Sigma_u) \quad (33)$$

et donc :

$$f(y_t | \bar{Y}_t) = (2\pi)^{-k/2} |\Sigma_u^{-1}|^{1/2} \exp\left[-\frac{1}{2} (y_t - \beta^\top \bar{Y}_t)^\top \Sigma_u^{-1} (y_t - \beta^\top \bar{Y}_t)\right] \quad (34)$$

On retrouve ici une difficulté déjà rencontrée lors de l'estimation des processus ARMA univariés : l'espérance conditionnelle de y_t fait intervenir les p réalisations précédentes y_{t-1}, \dots, y_{t-p} , qui ne sont pas observées pour les p premières observations. Comme on sait, la solution la plus simple est de traiter ces p premières valeurs comme des vecteurs de constantes définissant les conditions initiales du processus à estimer. On maximisera donc la vraisemblance conditionnelle¹³ :

$$f(y_T, y_{T-1}, \dots, y_2, y_1 | y_0, y_{-1}, \dots, y_{-p+1}; \beta, \Sigma_u) \quad (35)$$

écriture que l'on va ramener à des produits de termes plus simples, utilisant le fait qu'une densité jointe peut toujours s'écrire comme produit d'une marginale par une conditionnelle :

13. Pour mémoire, cela conduit à construire la vraisemblance sur les observations qui suivent les p premières. Dans les expressions qui suivent, les indices sont choisis de sorte que $T+p$ est la taille totale de l'échantillon, y_0, \dots, y_{-p+1} les conditions initiales, et $y_t, t = 1, \dots, T$ les observations sur lesquelles on écrit la vraisemblance conditionnelle à maximiser.

$$\begin{aligned}
& f(y_T, y_{T-1}, \dots, y_2, y_1 | y_0, y_{-1}, \dots, y_{-p+1}; \beta, \Sigma_u) \\
&= f(y_T | y_{T-1}, \dots, y_2, y_1, y_0, y_{-1}, \dots, y_{-p+1}; \beta, \Sigma_u) \\
&\quad \times f(y_{T-1} | y_{T-2}, \dots, y_2, y_1, y_0, y_{-1}, \dots, y_{-p+1}; \beta, \Sigma_u) \\
&= f(y_T | y_{T-1}, \dots, y_2, y_1, y_0, y_{-1}, \dots, y_{-p+1}; \beta, \Sigma_u) \\
&\quad \times f(y_{T-1} | y_{T-2}, \dots, y_2, y_1, y_0, y_{-1}, \dots, y_{-p+1}; \beta, \Sigma_u) \\
&\quad \times f(y_{T-2} | y_{T-3}, \dots, y_2, y_1, y_0, y_{-1}, \dots, y_{-p+1}; \beta, \Sigma_u) \\
&= \\
&\quad \vdots \\
&= \prod_{t=1}^T f(y_t | y_{t-1}, y_{t-2}, \dots, y_{-p+1}; \beta, \Sigma_u) \\
&= \prod_{t=1}^T (2\pi)^{-k/2} |\Sigma_u^{-1}|^{1/2} \exp \left[-\frac{1}{2} (y_t - \beta^\top \bar{Y}_t)^\top \Sigma_u^{-1} (y_t - \beta^\top \bar{Y}_t) \right]
\end{aligned}$$

La dernière égalité étant obtenue au moyen de l'équation (34).

La log-vraisemblance conditionnelle s'écrit donc :

$$l(\beta, \Sigma_u) = -\frac{Tk}{2} \log(2\pi) + \frac{T}{2} \log(|\Sigma_u^{-1}|) - \frac{1}{2} \sum_{t=1}^T (y_t - \beta^\top \bar{Y}_t)^\top \Sigma_u^{-1} (y_t - \beta^\top \bar{Y}_t) \quad (36)$$

Les estimateurs du maximum de vraisemblance conditionnelle des paramètres du VAR seront donc les solutions des systèmes d'équation :

$$\frac{\delta l(\beta, \Sigma_u)}{\delta \beta} = 0 \quad \text{et}, \quad (37)$$

$$\frac{\delta l(\beta, \Sigma_u)}{\delta \Sigma_u} = 0 \quad (38)$$

— **L'estimation des coefficients β :**

Plutôt que de se lancer dans la dérivation, Hamilton va introduire les estimateurs des OLS dans les écritures précédentes. Le VAR est en effet constitué de k équations linéaires ayant toutes $pk + 1$ variables explicatives : chacune des variables expliquée $y_{it}, i = 1 \dots, k$ est écrite sur une constante et p valeurs retardées d'elle-même et des $k-1$ autres variables. Il est donc possible d'estimer chacune de ces équations par les moindres carrés ordinaires. Soit $\hat{\beta}$ la matrice correspondante à β remplie par ces estimations OLS et $\hat{\epsilon}_t \equiv y_t - \hat{\beta}^\top \bar{Y}_t$, le vecteur collectant les k résidus empiriques de ces régressions OLS au temps t . On sait que ces résidus sont par construction orthogonaux aux explicatives : $\sum_{t=1}^T \bar{Y}_t \hat{\epsilon}_t^\top = 0$.

On considère ensuite le troisième terme de (43), le seul faisant intervenir β . C'est une somme de formes quadratiques construites sur une matrice définie-positive. Étant affecté d'un coefficient négatif, il sera donc négatif ou nul¹⁴. En conséquence si on veut maximiser la vraisemblance il faut minimiser ce terme.

$$\begin{aligned}
& \sum_{t=1}^T (y_t - \beta^\top \bar{Y}_t)^\top \Sigma_u^{-1} (y_t - \beta^\top \bar{Y}_t) \\
&= \sum_{t=1}^T (y_t - \hat{\beta}^\top \bar{Y}_t + \hat{\beta}^\top \bar{Y}_t - \beta^\top \bar{Y}_t)^\top \Sigma_u^{-1} (y_t - \hat{\beta}^\top \bar{Y}_t + \hat{\beta}^\top \bar{Y}_t - \beta^\top \bar{Y}_t) \\
&= \sum_{t=1}^T (\hat{\epsilon}_t + [\hat{\beta} - \beta]^\top \bar{Y}_t)^\top \Sigma_u^{-1} (\hat{\epsilon}_t + [\hat{\beta} - \beta]^\top \bar{Y}_t) \\
&= \sum_{t=1}^T \hat{\epsilon}_t^\top \Sigma_u^{-1} \hat{\epsilon}_t + \underbrace{\sum_{t=1}^T \hat{\epsilon}_t^\top \Sigma_u^{-1} [\hat{\beta} - \beta]^\top \bar{Y}_t}_{(1,k) \quad \underbrace{\quad}_{\{(k,1+kp)\}} \quad \underbrace{\quad}_{(1+kp,1)} \quad (1,1)} \\
&\quad + \sum_{t=1}^T \underbrace{\bar{Y}_t^\top [\hat{\beta} - \beta] \Sigma_u^{-1} \hat{\epsilon}_t}_{\text{transposée du terme précédent}} \\
&\quad + \sum_{t=1}^T \bar{Y}_t^\top [\hat{\beta} - \beta] \Sigma_u^{-1} [\hat{\beta} - \beta]^\top \bar{Y}_t \\
&= \sum_{t=1}^T \hat{\epsilon}_t^\top \Sigma_u^{-1} \hat{\epsilon}_t + 2 \sum_{t=1}^T \hat{\epsilon}_t^\top \Sigma_u^{-1} [\hat{\beta} - \beta]^\top \bar{Y}_t \\
&\quad + \sum_{t=1}^T \bar{Y}_t^\top [\hat{\beta} - \beta] \Sigma_u^{-1} [\hat{\beta} - \beta]^\top \bar{Y}_t
\end{aligned}$$

Le second terme de la dernière égalité est donc un scalaire. On sait que la trace d'une matrice (1×1) est égale à l'élément qu'elle contient. On vérifie donc, $\forall x \in \mathbb{R}$, $\text{trace}(x) = x$. Par ailleurs si A et B sont deux matrices ayant des dimensions telles que les produits AB et BA sont possibles, alors $\text{trace}(AB) = \text{trace}(BA)$. Ici, on voit qu'en posant $A = \hat{\epsilon}_t^\top$ et $B = \Sigma_u^{-1} [\hat{\beta} - \beta]^\top \bar{Y}_t$, on est en présence de matrices de dimensions respectives $(1 \times k)$ et $(k \times 1)$: la commutation du produit est possible. Dans ces conditions :

14. Nul si $y_t - \beta^\top \bar{Y}_t = 0$.

$$\begin{aligned}
\sum_{t=1}^T \hat{\epsilon}_t^\top \Sigma_u^{-1} [\hat{\beta} - \beta]^\top \bar{Y}_t &= \text{trace} \left(\sum_{t=1}^T \hat{\epsilon}_t^\top \Sigma_u^{-1} [\hat{\beta} - \beta]^\top \bar{Y}_t \right) \\
&= \text{trace} \left(\sum_{t=1}^T \Sigma_u^{-1} [\hat{\beta} - \beta]^\top \bar{Y}_t \hat{\epsilon}_t^\top \right) \\
&= \text{trace} \left(\Sigma_u^{-1} [\hat{\beta} - \beta]^\top \sum_{t=1}^T \bar{Y}_t \hat{\epsilon}_t^\top \right) \\
&= 0 \text{ grâce à l'orthogonalité de } \hat{\epsilon}_t \text{ et } \bar{Y}_t
\end{aligned} \tag{39}$$

Au total, si on reprend le troisième terme de la log-vraisemblance on a :

$$\sum_{t=1}^T (y_t - \beta^\top \bar{Y}_t)^\top \Sigma_u^{-1} (y_t - \beta^\top \bar{Y}_t) = \sum_{t=1}^T \hat{\epsilon}_t^\top \Sigma_u^{-1} \hat{\epsilon}_t + \sum_{t=1}^T \bar{Y}_t^\top [\hat{\beta} - \beta] \Sigma_u^{-1} [\hat{\beta} - \beta]^\top \bar{Y}_t$$

Ce terme que l'on veut minimiser, est donc la somme de deux éléments. Le premier est une forme quadratique de résidus non nuls sur une matrice définie-positive. Il est donc strictement positif. Le minimum est donc obtenu avec l'annulation du second élément, annulation qui est obtenue en posant $\beta = \hat{\beta}$.

Ce résultat est important : il montre que les estimateurs du maximum de vraisemblance conditionnels des coefficients d'un VAR s'obtiennent simplement en appliquant les OLS équation par équation. Cette facilité d'estimation explique en partie le succès de la modélisation VAR.

— L'estimation de la matrice Σ_u :

La log-vraisemblance donnée par (43) évaluée en $\beta = \hat{\beta}$ est :

$$l(\hat{\beta}, \Sigma_u) = -\frac{Tk}{2} \log(2\pi) + \frac{T}{2} \log(|\Sigma_u^{-1}|) - \frac{1}{2} \sum_{t=1}^T (\hat{\epsilon}_t^\top \Sigma_u^{-1} \hat{\epsilon}_t) \tag{40}$$

Cette expression est écrite en Σ_u^{-1} , *i.e.* sur l'inverse de la matrice qui nous intéresse. Il est en conséquence plus simple de considérer la condition d'ordre 1 du problème de maximisation en dérivant la log-vraisemblance par rapport à Σ_u^{-1} . Normalement des contraintes supplémentaires devraient être prises en considération puisque l'on sait que Σ_u est définie positive et symétrique. Nous verrons que ceci n'est pas nécessaire : la solution du problème tel qu'il est posé appartenant déjà au sous-ensemble des matrices définies positives : les solutions du problème général et du problème contraint sont identiques.

Sous réserve de satisfaire cette dernière contrainte, l'estimateur cherché est donc l'inverse de la solution du système :

$$\frac{\delta l(\hat{\beta}, \Sigma_u)}{\delta \Sigma_u^{-1}} = \frac{T}{2} \frac{\delta \log |\Sigma_u^{-1}|}{\delta \Sigma_u^{-1}} - \frac{1}{2} \sum_{t=1}^T \frac{\delta \hat{\epsilon}_t^\top \Sigma_u^{-1} \hat{\epsilon}_t}{\delta \Sigma_u^{-1}} = 0 \quad (41)$$

Nous allons considérer successivement chacune des dérivées du membre de droite.

- $\frac{T}{2} \frac{\delta \log |\Sigma_u^{-1}|}{\delta \Sigma_u^{-1}}$

Soit une matrice $M = \{m_{ij}\}_{n \times n}$, on montre que $\frac{\delta \log |M|}{\delta m_{ij}} = m^{ji}$, où m^{ji} est l'élément situé au croisement de la ligne j et de la colonne i dans M^{-1} , soit encore $\frac{\delta \log |M|}{\delta M} = (M^\top)^{-1}$. En effet :

$$|M| \equiv \sum_{i=1}^n (-1)^{i+j} m_{ij} |M_{i,j}|$$

où $|M_{i,j}|$ est le déterminant de M privée de la ligne i et de la colonne j . En conséquence,

$$\frac{\delta |M|}{\delta m_{ij}} = (-1)^{i+j} |M_{i,j}|,$$

et, si on se rappelle que l'inverse fait appel à la transposée de la matrice des cofacteurs,

$$\frac{\delta \log |M|}{\delta m_{ij}} = |M|^{-1} (-1)^{i+j} |M_{i,j}| = m^{ji}$$

- $\frac{1}{2} \sum_{t=1}^T \frac{\delta \hat{\epsilon}_t^\top \Sigma_u^{-1} \hat{\epsilon}_t}{\delta \Sigma_u^{-1}}$

Soit M et la forme quadratique $x^\top M x = \sum_{i=1}^n \sum_{j=1}^n x_i m_{ij} x_j$, on a :

$$\frac{\delta x^\top M x}{\delta m_{ij}} = x_i x_j,$$

soit, sous forme matricielle :

$$\frac{\delta x^\top M x}{\delta M} = x x^\top.$$

En appliquant ces deux résultats à (41), il vient :

$$\frac{\delta l(\hat{\beta}, \Sigma_u)}{\delta \Sigma_u^{-1}} = \frac{T}{2} \Sigma_u^\top - \frac{1}{2} \sum_{t=1}^T \hat{\epsilon}_t \hat{\epsilon}_t^\top = 0$$

Ce qui conduit à l'estimateur :

$$\hat{\Sigma}_u = \frac{1}{T} \sum_{t=1}^T \hat{\epsilon}_t \hat{\epsilon}_t^\top = \{s_{i,j}\}_{i,j=1,\dots,k} = \left\{ \frac{1}{T} \sum_{t=1}^T \hat{\epsilon}_{it} \hat{\epsilon}_{jt} \right\}_{i,j=1,\dots,k} \quad (42)$$

On remarque que cet estimateur, recherché dans la classe des matrices carrées, est de plus symétrique et défini positif : la prise en compte de ces deux contraintes dans la maximisation aurait donc conduit au même résultat.

— **La valeur de la log-vraisemblance à l'optimum et test LRT sur l'ordre du VAR**

Au final, la valeur de la fonction objectif évaluée en $\hat{\beta}$ et $\hat{\Sigma}_u$ est donnée par :

$$l(\hat{\beta}, \hat{\Sigma}_u) = -\frac{Tk}{2} \log(2\pi) + \frac{T}{2} \log(|\hat{\Sigma}_u|^{-1}) - \frac{1}{2} \underbrace{\sum_{t=1}^T \underbrace{\hat{\epsilon}_t^\top}_{(1 \times k)} \underbrace{\hat{\Sigma}_u^{-1} \hat{\epsilon}_t}_{(k \times 1)}}_{(1,1)} \quad (43)$$

Cette expression peut être simplifiée en remarquant que le dernier terme est un scalaire défini par des matrices dont le produit peut être commuté. En utilisant une démarche précédemment employée¹⁵ :

$$\begin{aligned} l(\hat{\beta}, \hat{\Sigma}_u) &= -\frac{Tk}{2} \log(2\pi) + \frac{T}{2} \log(|\hat{\Sigma}_u|^{-1}) - \frac{1}{2} \text{trace} \left[\sum_{t=1}^T (\hat{\epsilon}_t^\top \hat{\Sigma}_u^{-1} \hat{\epsilon}_t) \right] \\ &= -\frac{Tk}{2} \log(2\pi) + \frac{T}{2} \log(|\hat{\Sigma}_u|^{-1}) - \frac{1}{2} \text{trace} \left[\sum_{t=1}^T (\hat{\Sigma}_u^{-1} \hat{\epsilon}_t \hat{\epsilon}_t^\top) \right] \\ &= -\frac{Tk}{2} \log(2\pi) + \frac{T}{2} \log(|\hat{\Sigma}_u|^{-1}) - \frac{1}{2} \text{trace} \left[\hat{\Sigma}_u^{-1} \sum_{t=1}^T \hat{\epsilon}_t \hat{\epsilon}_t^\top \right] \\ &= -\frac{Tk}{2} \log(2\pi) + \frac{T}{2} \log(|\hat{\Sigma}_u|^{-1}) - \frac{1}{2} \text{trace} \left[\hat{\Sigma}_u^{-1} (T \hat{\Sigma}_u) \right] \\ &= -\frac{Tk}{2} \log(2\pi) + \frac{T}{2} \log(|\hat{\Sigma}_u|^{-1}) - \frac{T}{2} \text{trace} \left[\begin{matrix} I \\ (k \times k) \end{matrix} \right] \\ &= -\frac{Tk}{2} \log(2\pi) + \frac{T}{2} \log(|\hat{\Sigma}_u|^{-1}) - \frac{Tk}{2} \end{aligned} \quad (44)$$

Avec cette écriture il est aisé de calculer des tests de type LRT notamment sur l'ordre du VAR.

Pour mémoire, soient deux modèles M_{nc} et M_c , le second (modèle contraint) se déduisant du premier (modèle non contraint) par l'imposition d'un

15. Revoir ce qui est fait autour de l'équation (39).

nombre n_c de contraintes. Soient \hat{l}_{nc} et \hat{l}_c les valeurs à l'optimum de leur log-vraisemblances respectives. On sait déjà que $\hat{l}_{nc} \geq \hat{l}_c$, par ailleurs, si les contraintes imposées sont vraies alors on peut montrer que :

$$LRT = -2(\hat{l}_c - \hat{l}_{nc}) \quad (45)$$

est la réalisation d'un χ^2 à n_c degrés de liberté.

On peut illustrer un cas type d'application de ce test : supposez un échantillon constitué d'observations trimestrielles relatives à k variables. Un premier VAR a été ajusté avec $p = p_0$ retards. Soit donc :

$$y_t = c + \sum_{i=1}^{p_0} \Phi_i y_{t-i} + u_t \quad (46)$$

$(k \times 1)$ $(k \times k)$ $(k \times 1)$

dont on tire notamment l'estimation de Σ_u , notée $\hat{\Sigma}_{u|p=p_0}$ afin de montrer sa dépendance au nombre de retards retenus dans le modèle dont elle est issue. On se demande si un modèle avec seulement p_1 retards, avec $p_1 < p_0$ ne serait pas équivalent empiriquement au précédent, ce qui conduirait à la structure

$$y_t = c + \sum_{i=1}^{p_1} \Phi_i y_{t-i} + u_t \quad (47)$$

$(k \times 1)$ $(k \times k)$ $(k \times 1)$

et à une estimation de Σ_u , notée $\hat{\Sigma}_{u|p=p_1}$.

Dans cette situation, le modèle non contraint correspond à (46) et le modèle contraint à (47), les contraintes faisant le passage du premier au deuxième étant la nullité des coefficients des explicatives retardées de plus de p_1 périodes dans (46). Les hypothèses à considérer sont ainsi :

$$H0 : \Phi_{p_1+1} = \Phi_{p_1+2} = \dots = \Phi_{p_0} \text{ versus}$$

$$H1 : \exists \text{ au moins un } i, i \in \{p_1 + 1, p_1 + 2, \dots, p_0\} \text{ tel que } \Phi_i \neq 0$$

Le test LRT associé à ce jeu d'hypothèse est ¹⁶ :

$$\begin{aligned}
 LRT &= -2[\hat{l}_c - \hat{l}_{nc}] \\
 &= -2[\hat{l}_{p=p_1} - \hat{l}_{p=p_0}] \\
 &= -T \left[\log |\hat{\Sigma}_{u|p=p_1}|^{-1} - \log |\hat{\Sigma}_{u|p=p_0}|^{-1} \right], \text{ grâce à (44)} \\
 &= T \left[\log |\hat{\Sigma}_{u|p=p_1}| - \log |\hat{\Sigma}_{u|p=p_0}| \right]
 \end{aligned}$$

Sous H_0 , la quantité LRT est la réalisation d'un χ^2 à $(p_0 - p_1)k^2$ degrés de liberté ¹⁷.

2.4.2 L'estimation d'un VAR(p) sous Proc VARMAX dans SAS 9.3

Le calcul des estimateurs précédents est mis en oeuvre par la commande MODEL. Afin d'estimer un VAR d'ordre 2 sur trois variables : $y_t = (y_{1t}, y_{2t}, y_{3t})^T$, on pourra faire

```
PROC VARMAX DATA=...;
MODEL y1 y2 y3 / P=2;
run;
```

Par défaut, chaque équation incorpore une constante. La suppression de celle-ci demande à ce que l'on spécifie l'option NOINT. Par exemple :

```
PROC VARMAX DATA=...;
MODEL y1 y2 y3 / p=2 NOINT;
run;
```

L'incorporation de $n_s - 1$ indicatrices saisonnières peut aussi être gérée via l'option NSEASON= n_s . Sur données trimestrielles, on pourrait faire :

```
PROC VARMAX DATA=...;
MODEL y1 y2 y3 / P=2 NOINT NSEASON=4;
run;
```

16. Dans les égalités qui suivent, on suppose que l'utilisateur a pris soin d'ajuster les deux modèles sur le même nombre d'observations T , ce qui entraîne la disparition des termes constants de (44) dans la troisième égalité. Par exemple, sur un échantillon de taille 50, avec $p_0 = 8$ et $p_1 = 4$, la vraisemblance du non contraint peut, a priori, être évaluée sur 42 observations, les observations 1 à 8 constituant les conditions initiales. Celle du modèle contraint peut être évaluée sur 46 observations, les conditions initiales étant constituée par les 4 premières observations. Dans la suite d'égalités du texte, on suppose que les deux vraisemblances sont calculées sur 42 points (observations des rang 9 à 50)

17. En effet, H_0 impose la nullité de $(p_2 - p_1)$ matrices, chacune de dimensions $(k \times k)$. En pratique la question de la réduction de l'ordre du VAR peut être importante. Pour montrer son intérêt, reprenons la configuration $p_0 = 8$, $p_1 = 4$ et soit un VAR sur 6 variables ($k = 6$). Le passage de 8 à 4 retards économise l'estimation de $4 \times 6^2 = 144$ coefficients, où de manière équivalente, fait regagner 144 degrés de liberté, ce qui peut être non négligeable au regard de la taille de l'échantillon disponible pour les estimations.

Trois variables de type (0, 1) seront alors ajoutées à chacune des équations.

Au moyen de l'option `TREND=` on peut gérer l'incorporation à la liste des explicatives d'un trend déterministe linéaire, `TREND=LINEAR`, ou quadratique, `TREND=QUAD`.

Il est également possible de réclamer l'estimation du VAR sur des différences, identiques ou non, des variables y_1, y_2, y_3 via l'emploi de l'option `DIFY` :

```
PROC VARMAX DATA=...;
MODEL y1 y2 y3 / P=2 DIFY(1);
run;
```

réclame l'estimation de $\Delta y_t = \Phi_1 \Delta y_{t-1} + \Phi_2 \Delta y_{t-2} + u_t$. Il est possible de spécifier une liste de différences : `DIF(1, 4)` ajuste un VAR sur les observations créées selon $(1-L)(1-L^4)y_t$. Ces opérations de différenciation peuvent également être différentes selon les variables. Ainsi :

```
PROC VARMAX DATA=...;
MODEL y1 y2 y3 / P=2 DIFY=(y2(1) y3(1,4));
run;
```

réalisera l'estimation d'un VAR(2) sur les observations $(y_{1t}, \Delta y_{2t}, \Delta \Delta^4 y_{3t})^T$.

Notez encore que `P=` peut préciser une liste de retards. Dans ce cas, seules les variables retardées des ordres indiqués apparaîtront comme variables explicatives. Par exemple sur données trimestrielles, un VAR saisonnier d'ordre 2, où `SVAR(2)`, d'écriture :

$$y_t = c + \Phi_{1,s} y_{t-4} + \Phi_{2,s} y_{t-8} + u_t \quad (48)$$

sera ajusté par la commande :

```
PROC VARMAX DATA=...;
MODEL y1 y2 y3 / P=(4, 8);
run;
```

Ainsi, l'option `P=2` est équivalente à l'option `P=(1, 2)`. Enfin l'option `P=0` entraîne automatiquement la sélection d'un ordre optimal via le calcul d'un critère de sélection. La valeur par défaut de `P` étant 0, il en va de même si l'option `P=` n'est pas renseignée dans la ligne de commande¹⁸.

La méthode d'estimation est gouvernée par l'option `METHOD=`, maximum de vraisemblance conditionnelle vu précédemment via `METHOD=LS`, ou maximum de vraisemblance exact par `METHOD=ML`.

¹⁸. Nous n'avons détaillé que des spécifications de type VAR(p). L'extension de la commande `MODEL` à des VARMA(p,q) est immédiate. On fera par exemple `MODEL y1 y2 y3 / P=2 Q=1` ; pour ajuster un VARMA(2,1).

Les modalités d'impression des résultats sont gouvernées par les options PRINTALL et/ou PRINTFORM=. Cette dernière peut prendre trois valeurs :

- PRINTFORM=UNIVARIATE : les résultats sont affichés variable par variable,
- PRINTFORM=MATRIX : les résultats sont présentés sous une forme matricielle,
- PRINTFORM=BOTH : affichage selon chacun des deux formats précédents.

L'option PRINTALL active notamment PRINTFORM=BOTH ainsi que l'impression de la plupart des statistiques calculées par VARMAX¹⁹.

2.4.3 Estimation SURE de Zellner

Cette section n'apporte aucun résultat nouveau par rapport aux précédentes : elle constitue pour l'essentiel une révision rapide de l'estimation GLS et une présentation de son application à l'estimation d'un ensemble de coefficients d'un modèle à équations simultanées linéaires apparemment indépendantes (*Seemingly Unrelated Regressions*), ainsi qualifiées parce qu'elles ne sont liées que par des covariances non nulles entre leurs résidus.. A ce titre, les estimateurs SURE proposés par Zellner(1962)²⁰ sont justifiés pour estimer les coefficients d'un VAR. On montrera que l'on est alors dans un cas particulier pour lequel l'estimation SURE est équivalente à une succession d'estimations OLS réalisées sur chacune des équations du VAR.

rappel : les GLS et FGLS Dans le cas des Moindres Carrés Généralisés, on a une variable expliquée y , un ensemble de k explicatives X et l'équation usuelle

$$y = X\beta + u \quad (49)$$

A la différence du cadre OLS, on admet toutefois que le résidu u bien que toujours centré possède une matrice de variance-covariance $\Sigma_u \neq \sigma^2 I$. Dans cette configuration l'estimateur des OLS $\hat{\beta}_{OLS} = (X^T X)^{-1} X^T Y$ reste non biaisé mais n'est plus à variance minimale. L'idée est alors de transformer les variables de l'équation précédente pour arriver à une équation ayant toujours les β comme coefficients et un résidu possédant une matrice de variance-covariance $\sigma^2 I$, de sorte que les estimateurs OLS appliqués à ces transformés retrouvent l'ensemble de leurs propriétés. L'objectif est donc de diagonaliser la matrice définie-positve Σ_u . Pour cela, on peut par exemple utiliser une décomposition de Cholesky. Soit donc une matrice P triangulaire inférieure inversible, à éléments positifs sur la diagonale telle que $\Sigma_u = PP^T$. En prémultipliant cette dernière égalité par P^{-1} et en la post-multipliant par P^{T-1} il vient

19. On peut également mobiliser l'option PRINT(liste d'objets), cette liste étant constituée de mnémoniques référençant diverses quantités, par exemple CORBB est la matrice de corrélation des $\hat{\beta}$, CORRY celle des variables constituant y_t , etc...(voir l'aide de VARMAX pour la liste exhaustive.)

20. L'estimateur SURE a été proposée par Zellner : Zellner, A., *An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias*, J. Am. Stat. Assoc, 57, 348-368 (1962).

$P^{-1}\Sigma_u P^{\top-1} = P^{-1}PP^{\top}P^{\top-1} = I^{21}$. Dans ces conditions la matrice de var-cov de $\tilde{u} = P^{-1}u$ est la matrice identité : les \tilde{u} vérifient l'hypothèse d'orthogonalité et d'homoscédasticité. Pour retrouver des estimateurs sans biais et à variance minimale pour β , il suffit d'arriver à une régression dans laquelle les coefficients d'intérêt β sont toujours les coefficients des explicatives mais ayant \tilde{u} comme résidus et non pas u .

Ceci est obtenu aisément en multipliant la régression d'origine par P^{-1}

$$P^{-1}y = P^{-1}X\beta + P^{-1}u \Leftrightarrow \tilde{y} = \tilde{X}\beta + \tilde{u} \quad (50)$$

en conséquence on obtient l'expression de l'estimateur $\hat{\beta}_{GLS}$ qui est l'estimateur OLS appliqué aux variables transformées :

$$\hat{\beta}_{GLS} = (\tilde{X}^{\top}\tilde{X})^{-1}\tilde{X}^{\top}\tilde{y} = (X^{\top}\Sigma_u^{-1}X)^{-1}X^{\top}\Sigma_u^{-1}y \quad (51)$$

est un estimateur sans biais et à variance minimale de β .

Le résultat précédent suppose que Σ_u est connue. Lorsque ce n'est pas le cas, il faut trouver une matrice $\hat{\Sigma}_u$ qui soit un estimateur convergent de Σ_u . Les estimateurs des moindres carrés quasi généralisés (Feasible Generalized Least Squares) sont alors donnés par :

$$\hat{\beta}_{FGLS} = (X^{\top}\hat{\Sigma}_u^{-1}X)^{-1}X^{\top}\hat{\Sigma}_u^{-1}y \quad (52)$$

Rappelez-vous qu'à distance finie on ne connaît en général pas la distribution de $\hat{\beta}_{FGLS}$ et que si on ne dévie "pas trop" des hypothèses de base des OLS, alors $\hat{\beta}_{OLS}$ peut fort bien être plus efficace que $\hat{\beta}_{FGLS}$.

Les estimateurs SURE On se trouve maintenant face à un système de k variables expliquées et chaque expliquée y_i possède son propre modèle de régression, soit

$$y_i = X_i \beta_i + u_i, \quad i = 1, \dots, k \quad (53)$$

$(T,1)$ $(T,k_i)(k_i,1)$ $(T,1)$

et sur chacune de ces k régressions on suppose que les hypothèses usuelles des OLS sont vérifiées. On peut naturellement empiler ces k régressions pour n'en former plus qu'une :

$$y = X \beta + u, \quad (54)$$

$(kT,1)$ $(kT,\sum_{i=1}^k k_i)(\sum_{i=1}^k k_i,1)$ $(kT,1)$

selon,

21. On utilise ici la décomposition de Cholesky, mais on pourrait tout aussi bien utiliser la décomposition en valeurs et vecteurs propres, *i.e.* écrire $\Sigma_u = C\Lambda C^{\top}$, avec C la matrice des vecteurs propres vérifiant $CC^{\top} = I$ et Λ la matrice diagonale formée des valeurs propres de Σ_u . On aurait alors dans les écritures du texte une matrice P définie par $P = C\Lambda^{1/2}C^{\top}$.

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{pmatrix} = \begin{pmatrix} X_1 & 0 & 0 & \dots & 0 \\ 0 & X_2 & 0 & \dots & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \dots & & X_k \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_k \end{pmatrix} \quad (55)$$

Ces régressions ne sont cependant pas indépendantes : on admet l'existence de covariances instantanées non nulles entre les séries de résidus $u_i, i = 1, \dots, k$. Ainsi $cov(u_{it}, u_{jt}) = \sigma_{ij}$ et donc,

$$\Sigma_u = E(uu^\top) = \begin{pmatrix} \sigma_{11}I_T & \sigma_{12}I_T & \dots & \sigma_{1k}I_T \\ \sigma_{12}I_T & \sigma_{22}I_T & \dots & \sigma_{2k}I_T \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1k}I_T & \sigma_{2k}I_T & \dots & \sigma_{kk}I_T \end{pmatrix} \quad (56)$$

où I_T est la matrice identité ($T \times T$).

Dans ces conditions, l'équation (54) est caractéristique d'un modèle de régression pour lequel, à Σ_u connu, les GLS donnent des estimateurs de β sans biais et à variance minimale. On obtient donc :

$$\hat{\beta}_{SURE} = (X^\top \Sigma_u^{-1} X)^{-1} X^\top \Sigma_u^{-1} y \quad (57)$$

En notant $\Sigma = \{\sigma_{ij}\}_{i,j=1,\dots,k}$ on peut encore écrire Σ_u comme :

$$\Sigma_u = \begin{matrix} \Sigma \\ (kT, kT) \end{matrix} \otimes I_T \quad (58)$$

où \otimes désigne le produit de Kronecker²². Il est alors possible de simplifier l'expression de l'estimateur SURE en utilisant le fait que si A et B sont deux matrices inversibles alors $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ et :

$$\hat{\beta}_{SURE} = (X^\top \Sigma_u^{-1} X)^{-1} X^\top \Sigma_u^{-1} y \quad (59)$$

$$= (X^\top (\Sigma^{-1} \otimes I_T) X)^{-1} X^\top (\Sigma^{-1} \otimes I_T) y \quad (60)$$

$$= \begin{pmatrix} \sigma^{11}(X_1^\top X_1) & \sigma^{12}(X_1^\top X_2) & \dots & \sigma^{1k}(X_1^\top X_k) \\ \sigma^{21}(X_2^\top X_1) & \sigma^{22}(X_2^\top X_2) & \dots & \sigma^{2k}(X_2^\top X_k) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^{k1}(X_k^\top X_1) & \sigma^{k2}(X_k^\top X_2) & \dots & \sigma^{kk}(X_k^\top X_k) \end{pmatrix}^{-1} \begin{pmatrix} X_1^\top \left[\sum_{i=1}^k \sigma^{1i} y_i \right] \\ X_2^\top \left[\sum_{i=1}^k \sigma^{2i} y_i \right] \\ \vdots \\ X_k^\top \left[\sum_{i=1}^k \sigma^{ki} y_i \right] \end{pmatrix} \quad (61)$$

22. Pour mémoire, soit deux matrices A et B, avec $A = \{a_{ij}\}, i = 1, \dots, L, j = 1, \dots, M$, alors

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1M}B \\ a_{21}B & a_{22}B & \dots & a_{2M}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{L1}B & a_{L2}B & \dots & a_{LM}B \end{pmatrix}$$

où $\Sigma^{-1} = \sigma^{ij}$.

Sous cette forme, on peut mettre aisément en évidence deux cas dans lesquels l'estimateur SURE de β est identique à l'estimateur de β que l'on obtiendrait en appliquant les OLS équation par équation :

1. Lorsque les covariances entre les résidus des k équations sont orthogonales entre eux, *i.e.* lorsque $\sigma_{ij} = 0$, si $i \neq j$. Dans ce cas :

$$\hat{\beta}_{SURE} = \begin{pmatrix} \sigma^{11}(X_1^T X_1) & 0 & \dots & 0 \\ 0 & \sigma^{22}(X_2^T X_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^{kk}(X_k^T X_k) \end{pmatrix}^{-1} \begin{pmatrix} \sigma^{11} X_1^T y_1 \\ \sigma^{22} X_2^T y_2 \\ \vdots \\ \sigma^{kk} X_k^T y_k \end{pmatrix} = \begin{pmatrix} \hat{\beta}_{1,OLS} \\ \hat{\beta}_{2,OLS} \\ \vdots \\ \hat{\beta}_{k,OLS} \end{pmatrix}$$

L'intuition de ce résultat est évidemment que si les covariances entre les résidus sont nulles, alors les k équations ne sont pas seulement apparemment indépendantes : elles ne partagent linéairement aucune information en commun et il n'y a aucun avantage à considérer le système d'équations plutôt que chacune prise isolément.

2. Lorsque les explicatives des k régressions sont identiques, *i.e.* pour $X_i = X_0, i = 1, \dots, k$, soit dans (54) lorsque

$$\underset{(kT, kk_0)}{X} = \begin{pmatrix} X_0 & 0 & 0 & \dots & 0 \\ 0 & X_0 & 0 & \dots & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \dots & & X_0 \end{pmatrix} = \underset{k,k}{I_k} \otimes \underset{T,k_0}{X_0}$$

L'écriture VAR correspond précisément à ce cas particulier, X_0 étant alors constitué des valeurs retardées de toutes les variables constituant y .

Il vient, en reprenant (60) et la distributivité de l'opérateur \otimes ²³ :

$$\begin{aligned}
\hat{\beta}_{SURE} &= (X^\top(\Sigma^{-1} \otimes I_T)X)^{-1}X^\top(\Sigma^{-1} \otimes I_T)y \\
&= \left((I_k \otimes X_0)^\top(\Sigma^{-1} \otimes I_T)(I_k \otimes X_0) \right)^{-1} (I_k \otimes X_0)^\top(\Sigma^{-1} \otimes I_T)y \\
&= \left(\Sigma^{-1} \otimes X_0^\top X_0 \right)^{-1} \left(\Sigma^{-1} \otimes X_0^\top \right) y \text{ (distributivité de } \otimes \text{)} \\
&= \left(I_k \otimes (X_0^\top X_0)^{-1} X_0^\top \right) y \text{ (règle d'inversion et distributivité de } \otimes \text{)} \\
&= \begin{pmatrix} (X_0^\top X_0)^{-1} X_0^\top y_1 \\ (X_0^\top X_0)^{-1} X_0^\top y_2 \\ \vdots \\ (X_0^\top X_0)^{-1} X_0^\top y_k \end{pmatrix} \\
&= \begin{pmatrix} \hat{\beta}_{1OLS} \\ \hat{\beta}_{2OLS} \\ \vdots \\ \hat{\beta}_{kOLS} \end{pmatrix}
\end{aligned}$$

C.Q.F.D.

Notez encore que les estimateurs SURE donnés par (57) sont également utiles pour estimer des systèmes d'équations qualifiés de *near-VAR*, i.e. des VAR dans lesquels sont imposés des contraintes sur les coefficients, très souvent des contraintes de nullité, dont les conséquences sont que le nombre de retards de chaque explicative et/ou les listes des explicatives peuvent différer entre les équations²⁴.

2.4.4 La détermination de l'ordre du VAR

Dans les développements précédents nous avons vu, à p donné, comment estimer les coefficients du VAR(p) ainsi que Σ_u la matrice de variance covariance des résidus. La question qui reste en suspend est celui de la détermination de l'ordre vrai du processus p .

Ce point est important. Par exemple, si on surestime l'ordre vrai, alors l'inclusion d'explicatives non pertinentes dans les équations accroît l'erreur quadratique moyenne des prévisions tirées du VAR. D'un autre côté, sa sous-estimation génère des autocorrélations dans les résidus du VAR.

Afin de fixer une valeur raisonnable pour p , une première solution qui utilise certains de nos précédents résultats mais n'est cependant pas recommandée est de mettre en oeuvre des tests de type LRT. On peut en effet imaginer d'estimer

23. Rappel : pour des matrices de tailles adaptées, $(A \otimes B)(C \otimes D) = AC \otimes BD$.

24. Avec SAS, ces estimateurs sont obtenus via la Proc SYSLIN et son option SUR.

un VAR d'ordre p_0 avec p_0 relativement élevé puis de mener un test de

$$H0 : p = p_0 - 1 \text{ versus}$$

$$H1 : p = p_0$$

avec k variables, sous $H0$, le test serait la réalisation d'un χ^2 à k^2 degrés de liberté²⁵. En cas de rejet, on reste sur $p = p_0$, où éventuellement, on teste l'opportunité d'augmenter p_0 d'une unité. En cas de non rejet, on passe à $p = p_0 - 1$ et on reprend la logique précédente en se demandant s'il n'est pas déraisonnable de baisser ce nombre de retards d'une unité. L'objectif étant qu'à la fin de ces tests séquentiels, on soit conduit à retenir un p^* tel que en opposant $H0 : p = p^*$ versus $H1 : p = p^* + 1$ on ne rejette pas $H0$, et, en opposant $H0 : p = p^* - 1$ versus $H1 : p = p^*$ on rejette $H0$. Le problème avec cette démarche²⁶ est que l'on ne maîtrise pas le seuil de risque qui permet de sélectionner p^* à l'issue d'une succession de tests qui sont individuellement réalisés à un seuil α , même si ce dernier est clairement identifié²⁷.

1. **L'emploi des critères de sélection** : une démarche couramment employée pour éviter la difficulté précédente est de recourir aux critères de sélection vus au premier semestre. Si on considère un modèle M , on sait que l'expression générale des critères associés à ce modèle est de la forme :

$$\text{Critère}_M = -\frac{2l_M}{T} + \frac{k_M g(T)}{T} \quad (62)$$

où l_M est la log-vraisemblance du modèle M considéré, k_M , le nombre de coefficients que l'on doit estimer avec ce modèle, et $g(T)$ une fonction de pénalité spécifique à chaque critère.

Dans le cas des modèles VAR(p), et compte tenu de l'expression de la log-vraisemblance qui leur est associée, donnée par (44), il vient, pour les principaux critères utilisés que sont Akaike (AIC), Schwarz (BIC) et Hannan et Quinn (HQ) :

25. Car test de la nullité de k coefficients correspondant à chacune des variables retardées de p_0 et cela dans les k équations.

26. Démarche qualifiée de "general-to-specific" : on part d'une valeur initiale de p relativement élevée et on cherche à la réduire par une succession de tests, la démarche inverse "specific-to-general" partirait d'un p faible, par exemple 0 et chercherait le p optimal en l'augmentant. Lorsque ce type de procédure itérative est utilisée, on recommande de suivre la démarche "general-to-specific".

27. Dit autrement, avec cette démarche, on effectue un test conditionnellement à la conclusion tirée d'un test réalisé précédemment. Or ce dernier est toujours soumis à un risque d'erreur : on ne peut pas considérer que la conclusion est certaine : par exemple, il existe toujours un risque de montant α d'avoir rejeté à tort une hypothèse vraie. En conséquence, même s'il est aussi mené au seuil nominal α , le second test n'est plus réalisé à un seuil réel α , ce qui serait le cas uniquement si on était certain de la conclusion du premier.

Akaike :

$$g(T) = 2,$$

$$AIC = k [1 + \log(2\pi)] + \log(|\hat{\Sigma}_{u(p)}|) + \frac{2}{T}pk^2 \quad (63)$$

Schwarz :

$$g(T) = \log T,$$

$$BIC = k [1 + \log(2\pi)] + \log(|\hat{\Sigma}_{u(p)}|) + \frac{\log T}{T}pk^2 \quad (64)$$

Hannan-Quinn :

$$g(T) = 2 \log \log T,$$

$$HQ = k [1 + \log(2\pi)] + \log(|\hat{\Sigma}_{u(p)}|) + \frac{2 \log \log T}{T}pk^2 \quad (65)$$

On se souvient que, quel que soit le critère choisit, la règle est de sélectionner le modèle pour lequel ce critère est minimum²⁸. En pratique, on se fixe un ordre de retard maximal p_{max} , la recherche du retard optimal se faisant sur la famille de modèles créée par $p = 0, 1, 2, \dots, p_{max}$.

— **Le modèle vrai est d'ordre fini et est présent dans la famille de modèles concurrents :**

On retrouve alors les caractéristiques vues dans le cadre univarié : lorsque le modèle vrai est dans la famille des modèles concurrents alors AIC tend à surestimer l'ordre vrai alors que BIC et HQ sont asymptotiquement consistants²⁹. Comme d'habitude, il faut toutefois se poser la question de la validité des propriétés asymptotiques lorsque l'on applique ces outils sur des échantillons finis. Les études de Monte Carlo sont de ce point de vue particulièrement utiles. Dans un travail de 1985, Lütkepohl compare les performances de différentes procédures de sélection de l'ordre du VAR en utilisant comme critère les performances en prévision du modèle sélectionné. Sur cette base les critères BIC et HQ seraient préférables à AIC³⁰. La généralité

28. Ce qui explique que vous trouverez souvent des formulations de AIC, BIC et HQ qui ne font pas apparaître le terme constant $k [1 + \log(2\pi)]$: la valeur de p qui minimise un critère est évidemment la même que l'on incorpore ou pas cette constante dans son expression

29. Si on note p^* l'ordre vrai du VAR, \hat{p} l'ordre sélectionné avec un de ces critères, on a d'une part avec les 3 critères $\lim_{T \rightarrow \infty} Pr[\hat{p} < p^*] = 0$, d'autre part $\lim_{T \rightarrow \infty} Pr[\hat{p} = p^*] = 1$ avec BIC et HQ, et enfin $\lim_{T \rightarrow \infty} Pr[\hat{p} > p^*] > 0$ avec AIC. On peut cependant rappeler que Shibata montre que cette probabilité

de surestimation décroît avec la dimension k du VAR et devient pratiquement négligeable lorsque $k \geq 5$ (Shibata, *Some Properties of the Order of an Autoregressive model selected by the Generalization of Akaike's EPF Criterion*, *Biometrika*, 1977 64(3),547-551.)

30. Lütkepohl, H. , *Comparison of criteria for estimating the order of a vector autoregressive process*. *Journal of TimeSeries Analysis* 14, 47-69 (1985).

de cette conclusion doit toutefois être relativisée : l'estimation d'un VAR peut obéir à d'autres objectifs que la construction de prévisions. On peut par exemple citer les résultats obtenus par Ivanov et Kilian³¹ où Kilian³² qui jugent de la performance d'un critère par sa capacité à sélectionner le modèle fournissant les fonctions de réponse les plus précises³³. Dans l'ensemble, le critère AIC est alors préférable aux deux autres. On peut avoir l'intuition de ce résultat. En effet, dans l'exercice considéré, il s'agit de retrouver l'écriture VMA, c'est à dire une structure de retards infinie. Dans ce cas, partir d'une structure autorégressive finie trop parcimonieuse risque d'être dommageable. Or, sur des échantillons de tailles finis, les deux critères consistants BIC et HQ ont tendance à retenir des ordres inférieurs à celui retenu par AIC. Fondamentalement, ce que ces travaux semblent enseigner est que, dans le choix de tel ou tel critère, ce n'est pas tant la sélection de l'ordre vrai du VAR qui est importante que l'utilisation que l'on veut faire du VAR estimé.

On notera aussi que dans leur travail de 2005, Ivanov & Kilian recommandent de ne pas employer une procédure séquentielle telle que la succession de tests LRT discutée auparavant. Sur ce point, en revanche, les simulations réalisées par Hatemi et Hacker³⁴ sont favorables à l'emploi du test LRT lorsque les ordres sélectionnés par BIC et HQ diffèrent. Notez qu'il ne s'agit toutefois pas d'une application séquentielle. Par exemple, si l'ordre sélectionné par BIC est supérieur à celui retenu par HQ, soit $p_{BIC} > p_{HQ}$, le test portera sur $H_0 : p = p_{HQ}$ versus $H_1 : p = p_{BIC}$.

On peut également s'interroger sur les performances des critères lorsque les séries possèdent de la saisonnalité. Cette question est abordée dans la classe des VAR saisonniers³⁵ par Kadilar et Erdemir³⁶. Ils observent alors une forte dégradation de la capacité de AIC à retrouver l'ordre vrai d'un modèle SVAR simulé. Dans un autre travail³⁷, ces mêmes auteurs considèrent à la fois des VAR et des SVAR. Ils

31. Ivanov, V. & Kilian, L., *A Practitioner's Guide to Lag Order Selection For VAR Impulse Response Analysis*, Studies in Nonlinear Dynamics & Econometrics. Volume 9, Issue 1, Pages -, ISSN (Online) 1558-3708, DOI : 10.2202/1558-3708.1219, March 2005.

32. Kilian, L., *Impulse response analysis in vector autoregressions with unknown lag order*, Journal of Forecasting, 20, 161-179 (2001).

33. Ces fonctions de réponses servent à étudier la propagation des chocs à l'intérieur d'un système décrit par un VAR. Elles sont étroitement liées à la réécriture VMA infinie d'un VAR stable. Nous les étudions dans une des sections qui suivent.

34. Hatemi-J, A. ; Hacker, R. S. *Can LR Test Be Helpful in Choosing the Optimal Lag order in the VAR Model When Information Criteria Suggest Different Lag Orders ?*, Applied Economics 41, 09 (2009) 1121-1125" DOI : 10.1080/00036840601019273

35. Conformément à l'écriture des processus univariés déjà vu, le SVAR possède la structure suivante : $\Phi_p(L^s)y_t = u_t$, avec $\Phi_p(L^s) = (I - \Phi_1 L^s - \Phi_2 L^{2s} - \dots - \Phi_1 L^{ps})$.

36. Kadilar, C. & Erdemir, C. , *Modification of the Akaike information criterion to account for seasonal effects*, Journal of Statistical Computation and Simulation, 2003, Vol. 73(2), pp. 135-143.

37. Kadilar, C. & Erdemir, C. , *Comparison of performance among information criteria in VAR and Seasonal VAR Models*, Journal of Mathematics and Statistics, 2002, Volume 31 , 127-137

concluent que pour retrouver l'ordre vrai des modèles simulés, le critère AIC est moins performant que HQ, ce dernier étant lui-même inférieur à BIC. Dans la classe SVAR, ils déconseillent l'emploi de AIC et HQ.

— **Le modèle vrai est d'ordre infini :**

Ce cas mérite d'être considéré en raison de sa fréquence d'apparition sur données réelles. Par exemple, tout processus VARMA(p, q) inversible a une écriture VAR(∞) équivalente. On sait également que la somme de VAR d'ordres finis a toute les chances d'être un VARMA(p, q), $q > 0$. Or de nombreuses séries économiques sont des séries agrégées et/ou intègrent des sommes de chocs : la probabilité que la majeure partie de ces variables, même stationnaires, soient des VARMA et des VAR(∞) est certainement plus grande qu'elles soient des VAR($p, 0$). La question est alors de trouver un VAR(p), $p < \infty$, qui donne une bonne approximation du vrai modèle. Ce dernier s'écrit alors :

$$y_t = c + \sum_{i=1}^{\infty} \Phi_i y_{t-i} + u_t \quad (66)$$

avec u_t processus en bruit blanc. Son approximation d'ordre p est donnée par :

$$y_t = c_p + \sum_{i=1}^p \Phi_{p,i} y_{t-i} + u_{p,t} \quad (67)$$

Dans ce cadre, on conçoit aisément que le paramètre p , qui est alors un paramètre de troncature puisqu'on néglige tous les retards qui lui sont supérieurs, doit évoluer avec la taille de l'échantillon T . Plus précisément, p doit augmenter avec T car d'une part la qualité de l'approximation doit s'améliorer lorsque p augmente et, d'autre part, on peut d'autant plus accroître p que le nombre d'observations est important. On trouve alors dans la littérature des résultats sur cette dépendance de p à T notamment si l'objectif est d'obtenir via l'estimation de (67), des estimateurs $\hat{\Phi}_{p,i}$ convergents et asymptotiquement gaussiens des matrices $\Phi_i, i = 1, \dots, p$. Ainsi, les travaux de Berk³⁸ et Lewis et Reinsel³⁹ montrent que c'est le cas si p vérifie simultanément

38. Berk, K.N., 1974, *Consistent autoregressive spectral estimates*, Annals of Statistics 2, 489-502.

39. Lewis, R. & G. Reinsel, 1985, *Prediction of multivariate time series by autoregressive model fitting*, Journal of Multivariate Analysis 64, 393-411.

ment :

$$\lim_{T \rightarrow \infty} p^3/T = 0 : p \text{ ne doit pas augmenter trop vite, et}$$

$$\lim_{T, p \rightarrow \infty} \sqrt{T} \sum_{i=p+1}^{\infty} \Phi_i = 0 : p \text{ ne doit pas augmenter trop lentement}$$

Ng et Perron⁴⁰ ont d'ailleurs montré que les critères de type AIC, BIC et HQ ne respectaient pas la dernière condition et conduisaient à sélectionner un modèle tronqué dont les coefficients estimés sont des estimateurs biaisés des matrices correspondantes dans le vrai modèle.

En revanche, Kuersteiner⁴¹ justifie l'emploi d'une procédure séquentielle de test du type *general-to-specific* : soit $J(p)$ la statistique de test utilisée pour tester la nullité des coefficients afférents au dernier retard p . La démarche consiste à partir d'un p_{max} respectant les deux conditions précédentes,

- à retenir $\hat{p} = p$ si, pour un seuil de risque fixé α , $J(p)$ est la première statistique de la séquence $J(p), p = p_{max}, p_{max-1}, \dots, 1$ autorisant le rejet de H_0 .
- à retenir $\hat{p} = 0$ si H_0 n'est jamais rejetée pour tous les $p = p_{max}, p_{max-1}, \dots, 1$

Cependant, indépendamment du mode de détermination de p_{max} , on retrouve la difficulté soulignée auparavant : les statistiques $J(p)$ ne sont pas indépendantes entre elles et donc la loi marginale de $J(p)$ n'est pas celle de $J(p | J(p+1), \dots, J(p_{max}))$. En d'autres termes, les valeurs critiques utilisées, par exemple tirées d'une loi de χ^2 dans le cas de l'application d'un test de type LRT ou Wald, pour statuer sur la significativité de $J(p)$ ne sont sans doute pas les bonnes.

Fort heureusement, dans le même article, Kuersteiner justifie également le recours à une procédure de sélection faisant appel au critère AIC qui permettrait l'obtention d'estimateurs de $\hat{\Phi}_{p,i}$ via l'ajustement du VAR(p) tronqué (67) qui seraient asymptotiquement gaussiens et convergeraient vers les Φ_i de (66) notamment lorsque le VAR(∞) est la réécriture équivalente d'un VARMA(p, q). La démarche à suivre consiste alors

- à se donner une valeur p_{max} qui dépende de T de façon à respecter les conditions de Berk et Lewis & Reinsel, par exemple de la forme $p_{max} = (\log T)^a$, avec $1 < a < \infty$,

40. Ng, S. & P. Perron, 1995, *Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag*, Journal of the American Statistical Association 90, 268-281

41. Kuersteiner, G., 2005, *Automatic inference for infinite order vector autoregressions*, Econometric Theory, Volume 21, pp 85-115.

— et, si on note \hat{p}_{AIC} l'ordre sélectionné par AIC parmi l'ensemble $p = 0, 1, 2, \dots, p_{max}$, à retenir comme ordre pour le VAR tronqué l'entier $\hat{p} = M \hat{p}_{AIC}$, où M est une constante prise arbitrairement et telle que $M > 2$.

Dans VARMAX, l'option MINIC permet de réaliser cette recherche d'un ordre optimal au moyen de plusieurs critères de sélection. Outre les trois discutés auparavant, il est possible de calculer un critère AIC corrigé, qui est le critère utilisé par défaut, et un critère FPE, pour "final prediction error", fondé sur la matrice de variance covariance des erreurs de prévisions à une période ou MSE(1). La syntaxe est la suivante MINIC=(TYPE=critere P=entier Q=0 PERRORS=entier)⁴². Les mnémoniques à employer pour spécifier le critere choisi sont précisées ci-après avec un rappel de leur expression. Dans celles-ci, r est le nombre de paramètres estimés dans le modèle considéré⁴³ :

— TYPE=AIC pour Akaike,

$$AIC = \log(|\widehat{\Sigma}_u|) + \frac{2r}{T} \quad (68)$$

— TYPE=AICC .

SUGIURA⁴⁴ a suggéré une correction du biais de surestimation afférent à AIC visant à augmenter la probabilité de sélection de l'ordre vrai notamment sur petits échantillons. Cela se traduit par une augmentation de la pénalité liée au nombre de paramètres à estimer. Le critère AIC corrigé, AIC_C se définit alors comme :

$$AIC_C = \log(|\widehat{\Sigma}_u|) + \frac{2r}{T - r/k} \quad (69)$$

Hurvich et Tsai⁴⁵ soulignent une amélioration notable de la capacité de sélectionner le vrai modèle en utilisant AIC_C par rapport à AIC. Par ailleurs ces mêmes auteurs⁴⁶ concluent, au moyen de processus AR(∞) simulés, que les MSE des erreurs de prévisions tirées des AR d'ordre fini sélectionné par AIC_C sont toujours inférieures au MSE des erreurs issus des modèles retenus par AIC ou BIC.

42. Pour rester dans la classe VAR, penser à spécifier Q=0 : si vous n'imposez pas la nullité de Q, la procédure va prendre la valeur par défaut de ce paramètre qui est 5. Par exemple, MINIC=(type=AIC p=8) va construire la famille de processus concurrents VARMA(p,q) avec $p = 1, \dots, 8$ et $q = 0, \dots, 5$.

43. r peut différer de $k(pk + 1)$ si, outre la constante et l'ensemble des y_i décalés, les équations contiennent par exemple un trend, des indicatrices saisonnières, où d'autres explicatives comme dans le cas d'un VARX.

44. SUGIURA, N., *Further analysis of the data by Akaike's information criterion and the finite corrections*, Comm. Statist. A 7, 1978, 13-26.

45. Hurvich, C. M. & C.-L. Tsai, *Regression and time series model selection in small samples*, Biometrika 76, 1989, 297-307.

46. Hurvich, C. M. & C.-L. Tsai, *Bias of the corrected AIC criterion for underfitted regression and time series models*, Biometrika 78, 1991 pp. 499-509.

— TYPE=FPE pour le critère FPE.

Premier critère proposé par Akaike ⁴⁷, il est fondé sur une estimation de la MSE des erreurs de prévisions à une période. L'idée est de sélectionner le processus qui minimise cette MSE. Son expression est :

$$FPE = |\widehat{\Sigma}_u| \left(\frac{T + r/k}{T - r/k} \right)^k \quad (70)$$

où r/k est donc le nombre de paramètres à estimer dans chacune des équations du VAR.

Shibata ⁴⁸ a souligné l'équivalence asymptotique des critères AIC, FPE, et AIC_C. Une conséquence est qu'ils sont tous trois non consistants au sens où la probabilité de surestimation de l'ordre vrai est non nulle même sur très grands échantillons.

— TYPE=HQC pour Hannan et Quinn,

$$HQC = \log(|\widehat{\Sigma}_u|) + \frac{2r \log \log T}{T} \quad (71)$$

— TYPE=BIC ou TYPE=SBC pour Schwarz,

$$BIC = SBC = \log(|\widehat{\Sigma}_u|) + \frac{r \log T}{T} \quad (72)$$

Lorsque seul MINIC est spécifié, la procédure évalue uniquement le critère AIC_C : c'est le critère employé par défaut.

P= va préciser la valeur de p_{max} , l'ordre du processus VAR le plus élevé dans la famille des processus concurrents. Par exemple, P=8 fera une recherche de \hat{p} dans $0, 1, \dots, 8$. Il est aussi possible de préciser l'ordre inférieur avec la syntaxe P=($p_{min}; p_{max}$). Si on ne donne aucune valeur à P, la procédure utilise par défaut P=(0:5).

Dans PERRORS=entier, la valeur de entier impose l'ordre du VAR qui sera estimé afin de récupérer $|\widehat{\Sigma}_u|$. En effet, VARMAX n'estime pas tous les modèles concurrents : elle utilise les résidus de cette première approximation pour construire l'estimateur de $|\Sigma_u|$ qui entre dans les expressions des divers critères. En conséquence, la valeur imposée à PERRORS doit être relativement large. Par défaut, elle est égale à la valeur p_{max} précisée dans P=.

On aura par exemple :

47. Akaike, H. Statistical predictor identification. Ann. Ins. Statist. Math., 22, 1970, 203-217.

48. Shibata, R., An optimal selection of regression variables, Biometrika, 68, 1981, pp. 45-54.

```

PROC VARMAX DATA=...;
MODEL y1 y2 y3 / MINIC=(TYPE=BIC P=8 Q=0);
run;

```

En plus de l’affichage des valeurs du critère pour les modèles concurrents spécifiés, VARMAX réalisera automatiquement l’estimation du processus optimal, *i.e.* celui pour lequel le critère en question est minimal.

Une dernière remarque sur l’utilisation de ces critères : lorsque l’on fait une recherche sur $p = 1, \dots, p_{max}$ sur des séries observées de $t = 1$ à $t = T$, et que l’on utilise l’estimation par le maximum de vraisemblance conditionnel, le nombre d’observations sur lequel est calculée la vraisemblance du VAR(p) varie avec p puisque par défaut, le logiciel va construire cette vraisemblance sur $T(p) = T - p$ points où T est ici la taille de l’échantillon disponible⁴⁹. Si on regarde l’expression générale des critères, on voit que plusieurs ajustements sont possibles : Partant de (62) :

$$\text{Critère}_M = -\frac{2l_M}{T(p)} + \frac{k_M g(T(p))}{T(p)}$$

on peut imaginer de faire un ajustement de type OLS sur la partie vraisemblance, *i.e.* prendre pour Σ_u l’estimateur $\hat{u}\hat{u}^T / (T - p - k_M)$ plutôt que $\hat{u}\hat{u}^T / T$. On peut aussi vouloir faire cette correction sur la composante associée à la pénalité, *i.e.* $g(T - p) / (T - p)$ de préférence à $g(T) / T$. Sur des échantillons de grandes tailles ces corrections ne changent évidemment rien, mais il n’en va pas de même lorsque T est relativement faible, ce qui arrive assez souvent en pratique. Ces aspects ont été étudiés par Ng et Perron⁵⁰ via un ensemble de simulations. Leur travail montre d’une part que la conclusion, c’est à dire la sélection de l’ordre du VAR optimal, est surtout dépendante du nombre d’observations sur lequel la vraisemblance est construite et, d’autre part, que ce nombre doit être identique pour tous les modèles concurrents. En d’autres termes, quel que soit p , $0 \leq p \leq p_{max}$, tous les modèles concurrents, VAR(p), devraient être estimés sur $T(p) = T - p_{max}$ observations.

2. **Les matrices de corrélations partielles** : un autre outil d’aide à la décision peut également être utile : nous savons que pour un AR(p), la fonction d’autocorrélation partielle s’annule au-delà de p . On peut généraliser ce résultat : pour un VAR(p), les matrices d’autocorrélation partielles s’annulent également au-delà de p retards. Des équations de Yule-Walker relient les matrices de corrélation Γ aux matrices de corrélations partielles

49. Rappelez-vous la note de bas de page 16.

50. Ng, S. & Perron, P., *PRACTITIONERS’ CORNER A Note on the Selection of Time Series Models*, Oxford Bulletin of Economics and Statistics, 2005, 67, 1, pp. 115-134.

Φ , selon :

$$\Gamma_k = \sum_{i=1}^m \Gamma_{k-i} \Phi_{im}, \quad k = 1, 2, \dots, m \quad (73)$$

On retrouve les mêmes caractéristiques que celles énoncées dans le cadre univarié : si le système est un VAR(p) alors ces matrices de corrélation partielles vérifient $\Phi_{pp} = \Phi_p$ et $\Phi_{mm} = 0$, pour $m > p$. En pratique, à condition de préciser dans la commande `model...` l'option `print=(parcoef)`, la procédure VARMAX donne un visuel permettant de repérer les coefficients significatifs et non significatifs dans les matrices de corrélations partielles, visuel devant permettre de mettre en évidence assez aisément l'ordre p au-delà duquel ces matrices seraient nulles.

2.5 Les outils de validation

Dans un VAR(p), $p < \infty$, le processus u_t est un processus en bruit blanc et donc les autocorrélations de u_t avec u_{t+s} sont nulles pour $s \neq 0$. Dans un VAR d'ordre infini, l'approximation par un VAR(p) fini est satisfaisante si ces autocorrélations peuvent être considérées comme négligeables. En conséquence, comme pour les processus univariés, il faut s'assurer que cette hypothèse d'absence de corrélation est raisonnable pour le processus sélectionné. Lorsque de l'autocorrélation résiduelle est détectée, la procédure usuelle est d'augmenter l'ordre du VAR en espérant que cela finisse par la gommer.

Dans la proc VARMAX, on dispose de plusieurs outils pour réaliser cet objectif :

- Un ensemble de tests univariés qui vont porter successivement sur chacune des séries résiduelles des k équations.

Pour obtenir ces statistiques, il suffit de mettre le mot clef `DIAGNOSE` dans l'option `PRINT` de la commande `MODEL`. Par exemple :

```
PROC VARMAX DATA=...;
MODEL y1 y2 y3 / P=2 LAGMAX=8 PRINT=(DIAGNOSE);
run;
```

- un ensemble de graphiques portant également sur les résidus de chacune des équations : fonctions d'autocorrélation, d'autocorrélation partielle, d'autocorrélation inverse, et de statistiques de Ljung-Box. Ces graphes vous sont normalement familiers depuis que vous connaissez la procédure ARIMA. Ils ne sont donc pas présentés par la suite.

Ils sont obtenus via l'option `PLOTS` qui apparaît dans la ligne d'appel de la Proc VARMAX, avec le mot-clef `residual(diagnostics)` :

```
PROC VARMAX DATA=... PLOTS=(RESIDUAL(DIAGNOSTICS));
```

```
MODEL y1 y2 y3 / P=2 LAGMAX=8 PRINT=(DIAGNOSE);
run;
```

- Un test multivarié qui est une généralisation de la statistique de Box-Pierce.

1. Lest tests univariés : Ceux-ci sont au nombre de quatre :

- (a) Un test de Durbin-Watson. Comme vous le savez, celui-ci teste la nullité de l'autocorrélation d'ordre 1 dans la série résiduelle en supposant que l'alternative au bruit blanc est un processus AR(1). Dans le cas présent, cette statistique n'est pas particulièrement intéressante dans le cas de non rejet de l'hypothèse nulle notamment parce que sa puissance est faible : on sait que dans une régression OLS, si l'expliquée décalée apparaît dans la liste des explicatives alors DW est biaisé vers 2, ce qui est favorable au non rejet⁵¹. Or, précisément, dans un VAR, chaque expliquée est projetée notamment sur ses propres valeurs décalées. En d'autres termes, lorsque DW fait conclure à l'absence de corrélation, il ne faut pas lui faire particulièrement confiance. En revanche, lorsqu'il aboutit au rejet de l'hypothèse nulle, on devrait être alors d'autant plus enclin à douter de l'ordre sélectionné que sa puissance est faible.
- (b) Un test de Fisher : Sur les séries de résidus $\hat{u}_{it}, i = 1, \dots, k$. Quatre AR(j), $j = 1, \dots, 4$ sont ajustés et la processus test la nullité de leurs coefficients autorégressifs. Le rejet de la nulle signifiant un rejet de l'hypothèse de bruit blanc sur ces résidus, et donc une remise en cause de l'ordre sélectionné.
- (c) Un test de Fisher d'effet ARCH. Il s'agit ici de tester l'hypothèse de constance de la variance conditionnelle des résidus. La modélisation de la variance conditionnelle faisant l'objet d'un cours de M2, nous ne détaillerons pas ce point ici⁵². Très rapidement, le rejet de l'hypothèse nulle ici ne conduit pas à une remise en cause de l'ordre sélectionné. Elle plaide plutôt pour la mise en oeuvre de tests "robustifiés" via notamment le calcul d'estimateurs robustes des matrices de variance-covariances utilisées par exemple dans les tests de Wald que l'on peut

51. Usuellement dans ce cas de figure, on substitue la statistique h de Durbin à la statistique DW.

52. Juste pour une rapide illustration : soit un bruit blanc faible ω_t vérifiant donc $E[\omega_t] = 0$, $E[\omega_t^2] = \sigma^2$ et $E[\omega_t \omega_{t+s}] = 0$ si $s \neq 0$. La variance non conditionnelle de ω_t est donc constante, mais on va autoriser sa variance conditionnelle à se modifier dans le temps en posant $E_{t-1}[\omega_t^2] = h_t, h_t > 0$.

Dans le processus ARCH(1) d'Engle, on va ainsi poser $h_t = a_0 + a_1 \frac{\omega_{t-1}^2}{h_{t-1}}$, avec les contraintes $a_0 > 0$ et $a_1 \geq 0$ pour assurer la positivité de h_t . Ainsi, la variance conditionnelle en t de ω dépend positivement du carré de l'innovation standardisée qui s'est réalisée en $t - 1$. Si ω_t est par exemple le rendement d'un actif financier, alors ce modèle simple est compatible avec les clusters de volatilité que l'on peut observer sur les marchés, qui correspondent au fait qu'une journée calme, i.e. amplitude de l'innovation faible, soit plutôt suivie par une autre journée calme, alors qu'une journée agitée, i.e. innovation de forte amplitude, soit plutôt suivie par une autre journée agitée.

être amené à utiliser⁵³.

- (d) un test de Jarque-Bera de normalité des résidus. Ce test se fonde sur les mesures de skewness (asymétrie), et de kurtosis (aplatissement) définis comme :

$$sk = \frac{m_{(3)}^2}{m_{(2)}^3} \quad (74)$$

$$K = \frac{m_{(4)}}{m_{(2)}^2} \quad (75)$$

où $m_{(i)}$ est le moment centré d'ordre i de l'aléatoire considérée.

Ce test exploite le fait que pour une gaussienne, le coefficient sk est nul, car la densité de la normale est symétrique, et que son K est égal à 3. Jarque et Bera⁵⁴ ont montré que

$$JB = T \left[\frac{\hat{sk}}{6} + \frac{(\hat{K} - 3)^2}{24} \right] \xrightarrow[T \rightarrow \infty]{d} \chi^2(2) \quad (76)$$

où \hat{sk} et \hat{K} sont obtenus en prenant les estimateurs empiriques des moments dans les définitions précédentes.

Comme pour le test d'effet ARCH, le rejet de la nulle ici ne remet pas forcément en cause le choix de l'ordre p . D'ailleurs, si la taille de l'échantillon T est élevée, l'hypothèse de normalité n'est pas cruciale. En revanche, le rejet peut être dû à la présence d'outliers dans les résidus des régressions, qui eux, peuvent affecter la qualité des estimations : on rappelle qu'indépendamment de la valeur de JB, et de tous les tests précédents, il faut toujours faire une représentation graphique des séries de travail d'une part, et des résidus estimés de chaque équation d'autre part. L'objectif étant d'ajouter éventuellement des variables indicatrices d'évènements atypiques si des outliers sont repérés, de corriger les observations en cas d'erreur sur les données, voire d'introduire des trends linéaires déterministes si des tendances sont apparentes⁵⁵.

2. le test portmanteau multivarié de nullité des corrélations.

53. Par exemple lors des tests de causalité selon Granger que nous abordons plus loin.

54. Jarque, C.M. & Bera, A. K., *A test for normality of observations and regressions residuals*, International Statistical Review, 1987, 55, 163-172.

55. Sur ce dernier aspect, la prudence s'impose cependant : on sait que la non stationarité peut résulter de la présence de trends stochastiques et non pas déterministes. Déjà dans le cadre univarié vous savez que confondre l'un et l'autre type de trend peut conduire à des résultats fallacieux. Les choses se compliquent encore dans le cadre multivarié comme nous le verrons dans la seconde partie de ce cours.

En préambule, on dira que les plus téméraires doivent utiliser ce test avec beaucoup de prudence, et que les plus prudents ne l'utiliseront pas.

Ce test a été initialement avancé par Chitturi ⁵⁶. Il est implémenté dans VARMAX dans la version proposé par Hosking ⁵⁷.

Soit $\widehat{\Sigma}_{u,s}$ la matrice de variance-covariance empirique de u_t avec u_{t+s} :

$$\widehat{\Sigma}_{u,s} = \frac{1}{T} \sum_{t=1}^{T-s} \hat{u}_t \hat{u}_{t+s}^\top$$

et $\widehat{\rho}_{u,s}$ leur matrice de corrélations estimées :

$$\widehat{\rho}_{u,s} = \text{Diag}(s_{u_1}, \dots, s_{u_k}) \widehat{\Sigma}_{u,s} \text{Diag}(s_{u_1}, \dots, s_{u_k})$$

où $\text{Diag}(s_{u_1}, \dots, s_{u_k})$ est une matrice (k, k) dont la diagonale est constituée des écarts-types estimés des k innovations.

La statistique de Hosking porte sur l'hypothèse de nullité de l'ensemble des corrélations théoriques autres que les corrélations instantanées, et cela jusqu'à un ordre S .

$$H_0 : \rho_{u,s} = 0 \text{ pour } s = 1, 2, \dots, S.$$

Elle est définie comme :

$$Q_S = T^2 \sum_{s=1}^S (T-s)^{-1} \text{trace} \left[\widehat{\rho}_{u,s} \widehat{\rho}_{u,0}^{-1} \widehat{\rho}_{u,s}^\top \widehat{\rho}_{u,0}^{-1} \right]$$

et est calculée dès lors que l'option DIAGNOSE est utilisée. Le nombre de matrice de corrélations considérées, S , est calé sur la valeur attribué au paramètre LAGMAX.

On connaît la loi de cette statistique dans un cas très particulier : si u_t est un ensemble de k variables indépendantes et observées alors sous la nulle, Q_S possède asymptotiquement une distribution de χ^2 dont le nombre de degrés de liberté est le nombre d'autocorrélations dont on teste la nullité, *i.e.* $k^2 S$ degrés de liberté. Appliquée aux résidus d'un VAR(p), elle est censée avoir une distribution de χ^2 à $k^2(S-p)$ degrés de liberté.

56. Chitturi R.V., Distribution of residual autocorrelations in multiple autoregressive schemes, JASA, 1974, 69, 928-934.

57. Hosking, J.R.M., the multivariate portmanteaustatistic, JASA, 1980, 75, 602-608.

Maintenant, dans le cas d'un VAR, les résidus u_t sont inobservés et on va appliquer la statistique sur les résidus empiriques \hat{u}_t : un risque d'estimation apparaît, puisque $\Phi(L) \neq \widehat{\Phi}(L)$, qui n'est pas sans conséquence sur la convergence vers la loi limite précédente. Afin d'éliminer cet impact, Box et Pierce⁵⁸ conseillent, dans le cas univarié, de prendre une valeur élevée pour S . Dans le cas multivarié, si on applique cette règle de S élevé, on peut aisément rencontrer des matrices de variance-covariance presque singulières générant notamment un comportement erratique du seuil de risque effectif associé à la statistique. Ceci semble bien apparent dans les simulations réalisées par Dufour, Khalaf et Beaulieu⁵⁹ : le seuil de risque effectif semble dépendre de la dimension du VAR, k , et du nombre de corrélations considérées, via S : il peut monter à 50% pour k et S élevés pour un nominal de 5%. Avec douze variables et $S = 12$, ils estiment un risque effectif déjà plus de deux fois supérieur au risque nominal.

Par ailleurs, la convergence vers la loi asymptotique est fortement dépendante des paramètres autorégressif du VAR. En particulier, lorsque le VAR(p) est proche d'être non stable, *i.e.* qu'au moins une des racines de $|\Phi(L)|$, bien qu'extérieure au cercle complexe unitaire, possède un module proche de l'unité alors le seuil de risque effectif peut s'éloigner fortement du seuil de risque nominal choisit a priori, l'approximation par le χ^2 n'étant alors pas satisfaisante même pour des échantillons de grande taille. Ce phénomène apparaît ainsi nettement dans certaines des simulations réalisées par Francq et Raïssi⁶⁰.

En outre, pour obtenir la convergence asymptotique en distribution de Q_S vers un χ^2 on a besoin d'une hypothèse de bruit blanc fort sur les résidus : les u_t doivent être identiquement distribués, indépendants, avec $E[u_t] = 0$ et $Var[u_t] = \Sigma_u$. Frank et Raïssi montrent qu'avec une hypothèse moins contraignante de bruit blanc faible où l'indépendance est remplacée par une hypothèse d'orthogonalité, la distribution limite χ^2 n'est plus valide. Dans certaines de leurs simulations construites avec un bruit blanc faible, le niveau de risque associé à Q_S peut dépasser les 40% alors que l'utilisateur croit travailler au seuil de 5%.

Pour résumer,

- Il faudrait choisir un nombre de matrices de corrélation, S élevé pour limiter l'impact du risque d'estimation, mais alors on peut facilement se retrouver avec un comportement erratique du risque effectif associé à Q_S .

58. Box, G. & Pierce, D., Distribution of residual autocorrelations in autoregressive integrated moving average time series models, 1970, JASA, 65, 1509-1527.

59. Dufour, J.-M., Khalaf, L. & M.C.- Beaulieu, Multivariate residual-based finite-sample tests for serial dependence and ARCH effects with applications to asset pricing models, Journal of Applied Econometrics, 2010, 25, 263-285.

60. Francq, C. & Raïssi, H., Multivariate portmanteau test for autoregressive models with uncorrelated but nonindependent errors, Journal of Time Series Analysis, 2007, 28, 454-470.

- Si les tests univariés concluent à la présence d’effet ARCH dans vos résidus, le risque effectif peut décaler de façon importante de votre risque nominal.
 - si les valeurs propres de la matrice companion sont inférieures à l’unité mais proches de celle-ci, alors là encore le risque effectif peut dévier fortement du risque nominal
- Pour ces raisons, le risque nominal ne peut pratiquement pas être contrôlé avec cette statistique et je vous invite donc à relire le préambule à cette section.

2.6 Les fonctions de réponse aux chocs

Il s’agit ici de préciser l’impact d’un choc exogène sur les valeurs contemporaines et futures des variables constituant le VAR. Il s’agit par exemple d’apprécier l’impact sur les valeurs futures de l’ensemble des composantes de y d’une augmentation de 1 point dans le niveau de l’une de ces variables. Étant donné la structure d’un VAR, ces chocs d’innovation ne peuvent provenir que d’une modification imposée dans une des composantes du vecteur u .

2.6.1 Les fonctions de réponse simples

Ces réponses présentes et futures du système y à des chocs de type u sont aisément identifiées à partir de l’écriture $VMA(\infty)$:

$$y_t = \mu + u_t + \Psi_1 u_{t-1} + \Psi_2 u_{t-2} + \dots \quad (77)$$

Et donc :

$$\frac{\delta y_t}{\delta u_{t-h}} = \Psi_h \quad (78)$$

Ainsi, un choc d’amplitude unitaire en t se produisant sur la $j^{\text{ème}}$ composante de u_t impacte la $i^{\text{ème}}$ composante de y_{t+h} d’un montant égal à $\Psi_{h(i,j)}$ qui est la dérivée partielle de $y_{i,t+h}$ par rapport à u_{jt} ⁶¹ Habituellement on représente ces fonctions de réponse simples⁶² en portant en abscisse les valeurs de h et en ordonnée les valeurs estimées des coefficients $\Psi_{h(i,j)}$ ⁶³. Il existe cependant une difficulté : la mesure de l’impact d’une innovation au moyen de la dérivée partielle qui permet de retrouver l’interprétation usuelle que l’on donne à un multiplicateur⁶⁴ suppose que les autres innovations soient maintenues à zéro. En effet, si en mettant une innovation non nulle dans la $j^{\text{ème}}$ composante de u_t , cela génère également une déviation dans la $l^{\text{ème}}$ composante,

61. Pour mémoire, en raison de la stationnarité, $\Psi_{h(i,j)}$ est également la dérivée partielle de $y_{i,t}$ par rapport à u_{jt-h} .

62. Le calcul de ces fonctions est activé par l’option (IMPULSE=SIMPLE) dans VARMAX.

63. Vous noterez que la masse d’informations à exploiter devient vite problématique : dans un VAR de dimension k , il existe donc k innovations dont on peut étudier les réponses et k variables qui répondent, soit k^2 fonctions de réponse.

64. A savoir la mesure de l’impact de la variation d’une variable toutes autres variables inchangées.

alors la réponse de y va cumuler les deux impacts et ne pourra donc pas être attribuée en totalité à la seule $j^{\text{ème}}$ composante. Or les composantes de u_t n'ayant aucune raison d'être de covariance nulle, on ne peut pas écarter ce risque. Ce n'est que lorsque Σ_u est diagonale que l'effet multiplicateur d'une innovation peut être mesuré par la dérivée partielle.

La solution passe par la substitution d'un ensemble d'aléatoires orthogonales aux u_t initiaux. Il sera alors possible d'étudier la réponse du système à ces nouveaux chocs via les fonctions de réponse orthogonalisées.

2.6.2 Les fonctions de réponse orthogonalisées

Il existe plusieurs possibilités de réaliser une telle orthogonalisation des chocs u_t . Sims utilise une décomposition de Cholesky de la matrice Σ_u . Cette méthode de décomposition via une matrice triangulaire inférieure unique est depuis la plus populaire. Elle repose sur le fait que toute matrice définie-positive peut s'écrire comme un produit PP^T où P est une matrice inversible à éléments nuls au-dessus de la diagonale et à éléments positifs sur la diagonale⁶⁵. Comme Σ_u est une matrice de variance covariance, elle est définie positive. On est donc assuré de l'existence d'une telle matrice $P_{(k \times k)}$ unique telle que $\Sigma_u = PP^T$.

Soit $v_t = P^{-1}u_t$ un nouvel ensemble de k aléatoires construits à partir de combinaisons linéaires des u_t initiaux, combinaisons dont les poids sont donnés par les éléments de P^{-1} . Il vient :

$$E[v_t] = P^{-1}E[u_t] = 0, \quad (79)$$

$$\Sigma_v = E[v_t v_t^T] = P^{-1}E[u_t u_t^T]P^{-1T} = P^{-1}\Sigma_u P^{-1T} \quad (80)$$

$$\text{comme } \Sigma_u = PP^T \Rightarrow P^{-1}\Sigma_u P^{-1T} = P^{-1}PP^T P^{-1T} = I \quad (81)$$

$$\text{on a donc } \Sigma_v = I \quad (82)$$

En d'autres termes, les v_t ainsi construites sont des aléatoires centrées, de variance unitaire et orthogonales entre elles⁶⁶.

Pour obtenir les fonctions de réponse orthogonalisées afférentes aux chocs v_t , il suffit de reprendre leur expression de définition $v_t = P^{-1}u_t$ pour obtenir $u_t = Pv_t$ et de remplacer u_t par cette expression dans la formulation du $VMA(\infty)$. Il vient :

$$y_t = \phi(L)^{-1}[c + u_t] = \mu + \Psi(L)u_t = \mu + \Psi(L)Pv_t = \mu + \Xi(L)v_t \quad (83)$$

Grâce à l'orthogonalisation, on peut modifier une des composantes de v_t tout en maintenant les autres composantes à une valeur constante, notamment zéro : le raisonnement toutes autres variables inchangées est ici valide, et la dérivée partielle mesure bien l'effet multiplicateur du seul choc que l'on provoque. En

⁶⁵. cette contrainte de positivité assure l'unicité de P

⁶⁶. On montre aussi aisément que $E[v_t v_{t-s}] = 0$ si $s \neq 0$: v_t est, de même que u_t , un processus en bruit blanc

pratique, les coefficients de ces fonctions de réponse sont naturellement donnés par

$$\frac{\delta y_{it}}{\delta v_{j,t-h}} = \hat{\Xi}_{h(i,j)} \quad (84)$$

Ici, $\hat{\Xi}_{h(i,j)}$ est l'estimation⁶⁷ de l'impact d'un choc unitaire dans la $j^{\text{ième}}$ composante de v_t sur la valeur de la $i^{\text{ième}}$ composante de y_{t+h} . Comme pour les fonctions simples, les réponses fonctions de réponse orthogonalisées sont souvent représentées graphiquement avec h en abscisse et $\hat{\Xi}_{h(i,j)}$ en ordonnée et permettent d'apprécier la propagation dans le temps des effets d'un choc sur chacune des variables du système étudié.

Enfin, Rappelez-vous que dans un système stationnaire les effets des chocs sont transitoires : lorsque h augmente on doit tendre vers un multiplicateur nul. Ceci doit être observé aussi bien sur les fonctions de réponse simples que sur les réponses orthogonalisées.

Maintenant que les chocs dans ces fonctions de réponse orthogonalisées sont représentés par des valeurs attribuées aux v , il importe de connaître leur signification.

2.6.3 L'interprétation des chocs orthogonalisés et les conséquences de cette orthogonalisation

On considère la $j^{\text{ième}}$ équation d'un VAR(p) afférent à un y de dimension k :

$$y_{jt} = c_j + \sum_{i=1}^k \Phi_{1,ji} y_{it-1} + \sum_{i=1}^k \Phi_{2,ji} y_{it-2} + \cdots + \sum_{i=1}^k \Phi_{p,ji} y_{it-p} + u_{j,t}, \quad (85)$$

soit encore dans sa version centrée :

$$y_{jt} - \mu_j = \sum_{i=1}^k \Phi_{1,ji} (y_{it-1} - \mu_i) + \cdots + \sum_{i=1}^k \Phi_{p,ji} (y_{it-p} - \mu_i) + u_{j,t}, \quad (86)$$

on peut concevoir que les fonctions de réponse précédentes apprécient les conséquences d'un choc en comparant la dynamique générée par une innovation à une situation de référence telle que, en l'absence d'innovation, le système se reproduirait à l'identique période après période. Clairement cette position d'équilibre est donnée par $y_{t-1} = y_{t-2} = \cdots = y_{t-p} = \mu$ puisqu'alors, en l'absence de chocs, c'est à dire pour $u_t = 0$, nous aurions $y_t - \mu = y_{t+1} - \mu = y_{t+2} - \mu = \cdots = 0$. Dès lors, si à une date t_0 nous fixons $u_{jt_0} = 1$ et $u_{jt_0+k} = 0$, $k = 1, 2, \dots$, alors on va impliquer sur y_j et aussi sur les autres variables des trajectoires

67. Ces valeurs estimées sont obtenues en plusieurs étapes : l'estimation du VAR(p) fournit les matrices Φ_i , $i = 1, \dots, p$. Celles-ci sont injectées dans les équations de passage de l'écriture VAR vers la VMA donnant ainsi des Ψ_h , $h = 1, 2, \dots$. La décomposition de $\hat{\Sigma}_u$, également estimée à la première étape, aboutit à \hat{P} , et finalement, $\hat{\Xi}_h = \Psi_h \hat{P}$, $h = 1, 2, \dots$. L'emploi de l'option (IMPULSE=ORTH) demande dans VARMAX le calcul de ces fonctions de réponse orthogonalisées. Avec IMPULSE=ALL, on affiche les fonctions simples et orthogonalisées, ainsi que leur cumulés respectifs.

qui vont dévier de zéro et qui donc représenterons l'impact du choc unitaire effectué en t_0 , mais aussi les impacts des modifications que cela aura entraîné dans d'autres composantes de u dès lors que Σ_u n'est pas diagonale : les covariances non nulles interdiront de maintenir à zéro ces autres composantes lorsque u_j prendra la valeur 1. C'est précisément la raison qui pousse à l'orthogonalisation. Le seul intérêt du raisonnement précédent est de faire comprendre que ce qu'on appelle innovation à une date t_0 pour une équation quelconque d'un VAR consiste obligatoirement en une modification de son résidu en t_0 . En conséquence, si nous construisons un choc tel qu'il n'affecte pas u_{jt_0} , alors y_j ne peut pas répondre en t_0 . Naturellement les valeurs postérieures à t_0 de y_j pourront être affectées⁶⁸, mais la réaction contemporaine sera nulle.

Précisément l'orthogonalisation de Choleski crée des chocs v qui génèrent de telles configurations. En effet, vous savez que $u = Pv$ avec P matrice triangulaire inférieure. On a donc :

$$\begin{pmatrix} u_{1t} \\ u_{2t} \\ \vdots \\ u_{kt} \end{pmatrix} = \begin{bmatrix} p_{11} & 0 & \dots & \dots & 0 \\ p_{21} & p_{22} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 & \vdots \\ p_{k1} & p_{k2} & \dots & \dots & p_{kk} \end{bmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \\ \vdots \\ v_{kt} \end{pmatrix} \quad (87)$$

soit encore :

$$\begin{aligned} u_{1t} &= p_{11}v_{1t} \\ u_{2t} &= p_{21}v_{1t} + p_{22}v_{2t}, \dots \\ u_{kt} &= p_{k1}v_{1t} + p_{k2}v_{2t} + \dots + p_{kk}v_{kt} \end{aligned}$$

Ainsi, une innovation dans v_{1t} pourra affecter l'ensemble des u en t et donc une réaction contemporaine dans l'ensemble des variables. En revanche, une innovation dans v_{2t} n'affecte pas u_{1t} et donc ne provoquera pas de réaction contemporaine dans y_{1t} . Par contre elle peut affecter les u_{jt} tels que $j \geq 2$ et donc il sera possible d'observer une réponse dans $y_{2t}, y_{3t}, \dots, y_{kt}$. Enfin, pour les mêmes raisons, une modification de v_{kt} ne peut pas affecter immédiatement $y_{1t}, y_{2t}, \dots, y_{k-1t}$, seules y_{kt} et les explicatives de rangs supérieurs dans l'ordre d'entrée pourront répondre à l'instant même du choc. Plus généralement, l'aléatoire v_{jt} de rang j ne pourra pas provoquer de réponse contemporaine dans les variables y_{it} de rang inférieur à j , i.e. telles que $i < j$ mais qu'elle est susceptible d'avoir un impact immédiat sur les y_{it} de rang supérieur ou égal à j . Vous devez bien comprendre que cette absence de réponse immédiate de certaines variables n'est absolument une propriétés réelle du système étudié : c'est une conséquence de la méthode de décomposition de qui est seulement liée à l'ordre d'entrée des variables dans le VAR⁶⁹. En d'autres termes, les valeurs des fonctions de réponses dépendent de l'ordre d'entrée des variables dans le vecteur y .

68. puisque une où plusieurs autres variables présentes dans la liste des explicatives de y_j seront affectées en t_0 .

69. Évitez par exemple d'interpréter cette absence de réponse immédiate comme la preuve de l'absence de dépendance instantanée entre les variables concernées.

La question qui se pose alors naturellement est de savoir s'il existe un ordre d'entrée optimal, et, si oui, comment le déterminer ? Si on se limite aux précédents développements⁷⁰, nous avons un élément de réponse qui malheureusement en pratique se révèle souvent difficile à appliquer. Il faut entrer en premier la variable dont les modifications sont les plus susceptibles d'affecter de façon contemporaine l'ensemble des autres, et en dernier celle qui est le moins à même d'avoir un impact immédiat sur toutes celles qui la précèdent. Par exemple, supposez que vous soyez intéressé par la modélisation des ventes d'une entreprise et que vous pensiez que celles-ci dépendent de la conjoncture nationale ou/et régionale, représentée chacune par un indicateur conjoncturel. Vous pourriez alors construire y comme :

$$y_t = \begin{pmatrix} \text{indicateur national}_t \\ \text{indicateur régional}_t \\ \text{ventes de l'entreprise}_t \end{pmatrix}$$

Cependant, dans beaucoup de cas il sera difficile de justifier l'ordre choisit. En pratique, la prudence recommande d'essayer quelques permutations sur l'ordre d'entrée initial et de comparer les fonctions de réponse ainsi obtenues afin de vérifier la robustesse des conclusions que vous pourriez être amené à énoncer sur la seule base du classement initial⁷¹.

Outre la difficulté précédemment soulevée, une autre interrogation se pose avec l'orthogonalisation : si les chocs dont on étudie les réponses sont les aléatoires v , il reste à en comprendre la signification : étudier l'impact de chocs dont on ne sait pas ce qu'ils signifient n'est pas particulièrement intéressant.

Pour clarifier les choses, on peut repartir des équations de reconstruction des u à partir des v en nous limitant à un système de dimension 3, la généralisation à des systèmes supérieurs étant immédiate :

$$u_{1t} = p_{11}v_{1t} \tag{88}$$

$$u_{2t} = p_{21}v_{1t} + p_{22}v_{2t} \tag{89}$$

$$u_{3t} = p_{31}v_{1t} + p_{32}v_{2t} + p_{33}v_{3t} \tag{90}$$

D'après la première, v_{1t} n'est rien d'autre que u_{1t} réduite. Comme cette dernière est de variance σ_1^2 et que v_{1t} est de variance unitaire, on a immédiatement la valeur du coefficient de normalisation $p_{11} = \sigma_1$: un choc unitaire sur v_{1t} correspond à un choc d'un écart-type sur u_{1t} . Pour la deuxième, on peut encore l'écrire comme $u_{2t} = (p_{21}/p_{11})u_{1t} + p_{22}v_{2t}$. Comme v_{2t} est orthogonal à v_{1t} et donc à u_{1t} , $p_{22}v_{2t}$ est donc le résidu de la projection de u_{2t} sur u_{1t} : c'est la partie de u_{2t} non expliquée linéairement par u_{1t} . Ainsi, v_{1t} et v_{2t} engendrent le même espace

70. Les VAR structurels que nous verrons ultérieurement constituent un autre moyen de répondre à cette interrogation

71. Souvenez-vous qu'en présence de k variables dans le vecteur y , $k!$ permutations sont possibles, ce qui peut rendre l'exercice rapidement impraticable, à moins de ne retenir que quelques-unes d'entre elles en se fondant sur des considérations théoriques telles que, par exemple, la plus ou moins grande capacité d'une variable à affecter instantanément les autres variables du système.

que u_{1t} et u_{2t} . Dans ces conditions, comme v_{3t} est orthogonal à v_{1t} et v_{2t} , $p_{33}v_{3t}$ est le résidu de la projection de u_{3t} sur cet espace, et c'est donc la partie de u_{3t} non expliquée linéairement par u_{1t} et u_{2t} .

En généralisant ce raisonnement, v_{jt} apparaît donc comme étant la partie de u_{jt} non expliquée linéairement par les u_{it} de rangs inférieurs à j .

2.6.4 Intervalles de confiance sur les fonctions de réponses

Comme à chaque fois que l'on dispose d'une estimation, il est nécessaire d'avoir une idée de sa précision, via sa variance afin, par exemple, de pouvoir mener des tests d'hypothèses où construire des intervalles de confiance. Dans le cas qui nous intéresse ici, avant de commenter la forme des fonctions de réponse estimées, il est utile de statuer sur l'hypothèse de leur nullité : il serait dangereux de raconter une histoire sur l'impact d'un choc (signe de l'impact, propagation dans le temps), si dans le même temps on ne peut pas rejeter une hypothèse de nullité de ces coefficients de réponse. Il est donc nécessaire de construire des intervalles de confiance autour des coefficients estimés pour en particulier savoir si 0 est ou n'est pas dans l'intervalle en question.

Pour construire ceux-ci on a recourt soit à la méthode delta, soit à des intervalles bootstrappés. La première est la seule qui est actuellement implémentée dans la proc VARMAX (SAS 9.3). C'est donc naturellement la voie la plus simple à suivre sous ce logiciel pour l'obtention des IC. Il est assez cependant généralement admis qu'elle conduit à des intervalles trop larges et beaucoup de publications préfère utiliser les IC obtenus au moyen d'une méthode bootstrap.

Intervalles obtenus via la distribution asymptotique et la méthode delta

Avec les précédents calculs on dispose des estimateurs ponctuels de la réponse de la variable y_i à un choc u_j ou v_j après h périodes via $\hat{\Psi}_{h(i,j)}$ ou $\hat{\Xi}_{h(i,j)}$. On rappelle que $\Psi(L) = \Phi(L)^{-1}$ et que les estimateurs des matrices autorégressives $\hat{\Phi}(L)_i, i = 1 \dots, p$, sont des estimateurs du maximum de vraisemblance. D'après la propriété d'invariance de ceux-ci, les estimateurs du maximum de vraisemblance de $\Psi(L)$ sont donc $\hat{\Phi}(L)^{-1}$. Ils sont convergents, asymptotiquement gaussiens. Soit encore : $\hat{\Psi}_h(L) = g_h(\hat{\Phi})$. Connaissant la matrice de variance covariance des coefficients du VAR, $V(\hat{\Phi})$, la méthode delta⁷² permet de trouver celle des coefficients $\hat{\Psi}(L)$, soit $V(\hat{\Psi}_h(L)) = G_h V(\hat{\Phi}) G_h^\top$ ou $G_h = \frac{\delta g_h(\hat{\Phi})}{\delta \Phi|_{\hat{\Phi}=\hat{\Phi}}}$ ⁷³. Les expressions analytiques de ces variances asymptotiques sont cependant

72. Rappel, soit X un vecteur de k aléatoires d'espérance μ_X , de variance V_X , et $Y = g(X)$, $g()$ continue dérivable alors un développement de Taylor d'ordre 1 au voisinage de μ_X donne : $Y = g(X) \approx g(\mu_X) + G(X - \mu_X)$, $G()$ étant la dérivée partielle de $g()$ évaluée en μ_X . Ainsi $V_Y = Var(GX) = GV_X G^\top$.

73. Dans le cadre d'un VAR, les fonctions de passage $g_h()$ des coefficients autorégressifs Φ vers les coefficients Ψ du VMA sont relativement complexes et en tout cas non linéaires. On peut illustrer les manipulations dans le cadre simple d'un AR(1) univarié stationnaire : $x_t = \phi x_{t-1} + u_t = \sum_{i=0}^{\infty} \phi^i u_{t-i} = \sum_{i=0}^{\infty} \psi^i u_{t-i}$ avec x_t une variable aléatoire réelle, $|\phi| < 1$, $\psi_0 = 1$. La réponse de x_{t+h} à un choc u_t est donc donnée par $\psi_{i+h} = \phi^h$. Dans ce cadre, $G() = h\hat{\phi}^{h-1}$ est un scalaire, et finalement $Var(\hat{\psi}_{i+h}) = (h\hat{\phi}^{h-1})^2 Var(\hat{\phi})$, où $Var(\hat{\phi})$ est donnée par l'estimation de l'AR(1).

compliquées et cela explique en partie la popularité de la méthode bootstrap de construction des intervalles : celle-ci se passe en effet de la connaissance des expressions en question. L'autre argument est que sur petits échantillons les propriétés asymptotiques ne sont pas vérifiées et on espère alors que le taux de couverture des IC obtenus par bootstrap améliore celui des IC associés à la méthode delta.

Intervalles obtenus via la méthode bootstrap

On sait que si le VAR(p) est bien spécifié, alors u_t est un processus en bruit blanc : ses composants, $u_{1t}, u_{2t}, \dots, u_{kt}$ ne sont pas nécessairement orthogonaux mais, en revanche, $\forall s \neq 0, u_t$ et u_{t+s} , le sont. Soit $\hat{U} = (\hat{u}_1^\top, \hat{u}_2^\top, \dots, \hat{u}_T^\top)$ la matrice $(T \times k)$ dont la $i^{\text{ème}}$ ligne est constitué par le vecteurs des résidus estimés au temps i . Chacun de ces vecteurs peut être vu comme la réalisation d'un tirage aléatoire équiprobable de u_i . Le fait qu'il se soit réalisé au temps i n'est pas essentiel pour les propriétés des estimateurs du VAR puisqu'il avait autant de chance de se réaliser à un autre temps j . La méthode bootstrap exploite cette idée comme suit :

1. Estimation du VAR, récupération des coefficients $\hat{\Phi}_k, k = 1, 2, \dots, p$ et des fonctions de réponse $\hat{\Psi}_{h(i,j)}$ et/ou $\hat{\Xi}_{h(i,j)}$, construction de \hat{U} .
2. construction d'une matrice \tilde{U} de dimension $(T \times k)$ par tirage aléatoire avec remise dans les lignes de \hat{U}
3. les p premières observations de y_t étant considérées comme données⁷⁴, construction de k séries \tilde{y}_t selon :

$$\tilde{y}_t = \begin{cases} y_t & \text{pour } t = -p+1, -p+2, \dots, 0 \text{ (conditions initiales)} \\ \hat{c} + \hat{\Phi}_1 \tilde{y}_{t-1} + \hat{\Phi}_2 \tilde{y}_{t-2} + \dots + \hat{\Phi}_p \tilde{y}_{t-p} + \tilde{u}_t^\top & \text{pour } t = 1, 2, \dots, T \end{cases} \quad (91)$$

ou \tilde{u}_t est la $t^{\text{ème}}$ ligne de \tilde{U}

4. estimation du VAR(p) au moyen des nouvelles séries \tilde{y}_t et des fonctions de réponse associées $\tilde{\Psi}_{h(i,j)}$ et/ou $\tilde{\Xi}_{h(i,j)}$
5. les étapes 2 à 4 sont répétées n fois. Au final on dispose, pour tout h , de n valeurs $\tilde{\Psi}_{h(i,j)}$ ou $\tilde{\Xi}_{h(i,j)}$. Un encadrement à, par exemple, 90% de confiance des $\hat{\Psi}_{h(i,j)}$ ou $\hat{\Xi}_{h(i,j)}$ obtenus à l'étape 1 est construit en prenant les $5i^{\text{ème}}$ et $95i^{\text{ème}}$ centiles obtenus sur ces n valeurs $\tilde{\Psi}_{h(i,j)}$ ou $\tilde{\Xi}_{h(i,j)}$ selon que l'on veut un IC sur les fonctions simples ou orthogonalisées.

Avec cette méthode, aucune hypothèse de distribution n'est imposée sur les innovations, si ce n'est leur indépendance dans le temps. C'est

74. Notez que dans certaines implémentations, le vecteur des conditions initiales est lui-même constitué d'une séquence de $p+1$ observations tirée au hasard parmi les $T+p+1$ valeurs observées. Cette pratique est dangereuse si le système est non stationnaire, ou proche de la non stationarité, ces tirages successifs ne décrivant alors pas bien la distribution non conditionnelle des conditions initiales

donc une procédure non paramétrique de construction d'estimateurs de bornes d'IC. Naturellement, si l'hypothèse d'indépendance n'est pas vérifiée, le tirage aléatoire avec remise dans les lignes de \hat{U} fait que la méthode de bootstrapping est déficiente puisque \tilde{U} ne peut plus être considérée comme une réalisation d'aléatoires de même loi que les innovations initiales.

On notera cependant que cette procédure simple à mettre en oeuvre donne des IC ayant un taux de couverture de la vraie fonction de réponse pouvant être très éloigné du taux nominal $(1 - \alpha)^{75}$. Pour certains, l'origine du problème est que le biais d'estimation sur les $\Phi_i, i = 1, \dots, p$ sont amplifiés sur les $\Psi_{h(i)}$. Afin de corriger ces biais, une proposition relativement populaire a été avancée par Killian⁷⁶ : cette technique, dite de *bootstrap after bootstrap*, consiste à effectuer une première fois les étapes 1 à 5 précédentes pour récupérer un ensemble de n matrices de coefficients autorégressifs sur les échantillons bootstrappés, $\hat{\Phi}_k^* = \hat{\Phi}_{1k}^*, \hat{\Phi}_{2k}^*, \dots, \hat{\Phi}_{pk}^*, k = 1, \dots, n$. Si $\hat{\Phi}$ est la matrice des coefficients obtenus sur l'échantillon initial, on estime le biais par $\hat{\Phi} - \overline{\Phi^*}$, ou $\overline{\Phi^*}$ est la moyenne des $\hat{\Phi}_k^*$, pour construire un estimateur des coefficients autorégressifs corrigés du biais :

$$\hat{\hat{\Phi}} = \hat{\Phi} + (\hat{\Phi} - \overline{\Phi^*}) = 2\hat{\Phi} - \overline{\Phi^*}$$

Les 5 étapes précédentes sont alors reprises une deuxième fois en remplaçant $\hat{\Phi}$ par $\hat{\hat{\Phi}}$ dans la troisième.

2.6.5 L'obtention des fonctions de réponse et de leurs écart-types dans VAR-MAX

Les fonctions de réponses estimées s'affichent en fonction des mots-clefs spécifiés dans l'option PRINT de la commande MODEL. Il est possible d'obtenir les valeurs estimées :

- des fonctions de réponse simples, i.e. des $\widehat{\Psi}_h$ dans

$$y_t = \mu + u_t + \widehat{\Psi}_1 u_{t-1} + \widehat{\Psi}_2 u_{t-2} + \dots$$

où $\widehat{\Psi}_{ijh}$ est la réponse après h périodes de la $i^{\text{ème}}$ variable du système à un choc affectant la $j^{\text{ème}}$ variable,

- des fonctions de réponse simples cumulées : pour un horizon h , elles sont définies comme

$$\sum_{k=0}^h \widehat{\Psi}_{ij,k}$$

75. Voir par exemple Sims, C. A. & ZHA, T., Error bands for impulse responses, *Econometrica*, Vol. 67, No. 5, September 1999, ou Benkwitz, A., Neumann, M. H. & Lütekpohl, H., Problems related to confidence intervals for impulse responses of autoregressive processes, *Econometric Review*, Volume 19, Issue 1, 2000, pp. 69-103.

76. Killian, L., Small-sample confidence intervals for impulse response functions, *Review of Economics and Statistics*, 1998, 80, 218-230.

— des fonctions de réponse orthogonalisées, i.e. des $\widehat{\Xi}_h$ dans

$$y_t = \mu + u_t + \widehat{\Xi}_1 u_{t-1} + \widehat{\Xi}_2 u_{t-2} + \dots$$

où $\widehat{\Xi}_{ij,h}$ est la réponse après h périodes de la $i^{\text{ème}}$ variable du système au choc orthogonalisé selon Cholesky afférant à la $j^{\text{ème}}$ variable,

— des fonctions de réponse orthogonalisées cumulées : pour un horizon h , elles sont définies comme

$$\sum_{k=0}^h \widehat{\Xi}_{ij,k}$$

— des écarts-types des fonctions de réponse simples et orthogonalisées, cumulées et non cumulées.

En ce qui concerne le choc dont on étudie l'impact sur la dynamique du système, son amplitude est généralement adaptée à l'innovation concernée : les u_i ayant des variances différentes, on va imposer une amplitude qui soit comparable lorsque l'on passe d'un choc sur une variable à un choc sur une autre variable⁷⁷. En pratique, on va imposer une amplitude égale à l'écart-type estimé de la partie inexpliquée de u_i , ainsi, dans le système trivarié suivant :

$$u_{1t} = p_{11}v_{1t} \quad (92)$$

$$u_{2t} = p_{21}v_{1t} + p_{22}v_{2t} \quad (93)$$

$$u_{3t} = p_{31}v_{1t} + p_{32}v_{2t} + p_{33}v_{3t} \quad (94)$$

les amplitudes des trois chocs orthogonaux seront respectivement égales à l'écart-type de $p_{11}v_{1t}$, soit p_{11} pour le premier, à l'écart-type de $p_{22}v_{2t}$, soit p_{22} pour le deuxième et à p_{33} pour le dernier⁷⁸.

L'affichage des fonctions précédentes est obtenu en spécifiant dans la commande `PRINT=(IMPULSE=(. . .))` le où les mots clefs correspondant à l'objet que

77. supposez par exemple que $\sigma_{u_1}^2 = 0.5$ et $\sigma_{u_2}^2 = 16$: un choc d'une amplitude unitaire équivaut à une amplitude de 4 écart-type sur u_1 , ce qui correspond à une événement extrêmement rare de type outlier, et à une amplitude d'un quart d'écart-type pour u_2 , événement que l'on ne peut pas qualifier d'anormal dans cette distribution.

78. Par exemple, dans un VAR bivarié, si la matrice de variance covariance estimée est

$$\widehat{\Sigma}_u = \begin{pmatrix} 1.32267 & 0.38206 \\ 0.38206 & 1.40146 \end{pmatrix}$$

et que l'on s'intéresse aux chocs orthogonalisés, comme u_1 n'est expliquée que par lui-même, le choc afférant à la première variable aura une ampleur égale à $\sqrt{1.32267} = 1.15007$. Pour la seconde, il faut évaluer la variance inexpliquée de u_2 . Si \hat{a} est le coefficient de la régression de u_2 sur u_1 , on sait d'une part que $\hat{a} = \text{cov}(u_1, u_2) / \text{var}(u_1) = 0.38206 / 1.32267$, et d'autre part que la variance estimée du résidu est égale à $\text{var}(u_2) - \hat{a}^2 \text{var}(u_1)$, soit finalement ici 1.29110. L'amplitude du choc afférant à la seconde variable sera donc de $\sqrt{1.29110} = 1.13627$. Ces résultats sont évidemment directement obtenus en calculant la décomposition de Cholesky de $\widehat{\Sigma}_u$, soit $\hat{P} = \begin{pmatrix} 1.15007 & 0.00000 \\ 0.33220 & 1.13627 \end{pmatrix}$.

l'on veut : SIMPLE, ORTH, ACCUM, STDERR, ALL. La syntaxe est la suivante :

PRINT=(IMPULSE=SIMPLE) ou PRINT=(IMPULSE=(SIMPLE ACCUM ORTH)).

L'horizon sur lequel on veut examiner ces réponses, *i.e.* le nombre de valeurs que l'on veut afficher pour chacune d'elle, est gouverné soit par l'option LAGMAX= de cette même commande MODEL⁷⁹, soit par le nombre mis entre parenthèses immédiatement après le mot-clef IMPULSE

Les graphes de ces fonctions, avec indication d'un intervalle de confiance de ± 2 écart-types sont obtenus en activant l'option PLOTS= dans la ligne d'appel de la proc VARMAX. On aura par exemple :

```
PROC VARMAX DATA=... PLOTS=(IMPULSE(SIMPLE ORTH));
MODEL y1 y2 y3 / P=2 LAGMAX=12 PRINT=(IMPULSE=((orth stderrs)));
run;
```

Noter qu'il est possible d'activer l'option PLOTS= sans réclamer l'affichage des valeurs dans PRINT : dans le programme précédent, nous aurons le rendu graphique des fonctions de réponse simples sans pour autant les avoir appelées dans PRINT=(IMPULSE=...).

2.7 La décomposition de la variance des erreurs de prévisions

La prévision à un horizon h de y est donnée par $y_{t+h|t} = E_t[y_{t+h}]$. En conséquence, l'expression de l'erreur associée est, si on utilise l'écriture du VMA sur les chocs orthogonaux (équation 83),

$$e_{t+h|t} = y_{t+h} - y_{t+h|t} = \Xi_0 v_{t+h} + \Xi_1 v_{t+h-1} + \dots + \Xi_{h-1} v_{t+1} \quad (95)$$

Fort logiquement, les erreurs de prévision à un horizon h sont déterminées par les innovations qui se produiront au cours des h périodes à venir. On a évidemment :

$$E_t[e_{t+h|t}] = 0 \quad (96)$$

L'expression de l'erreur afférente à une des variables y_i , $1 \leq i \leq k$ quelconque est ainsi :

$$e_{i,t+h|t} = y_{i,t+h} - y_{i,t+h|t} \quad (97)$$

$$= \sum_{j=1}^k \Xi_{0,i,j} v_{j,t+h} + \sum_{j=1}^k \Xi_{1,i,j} v_{j,t+h-1} + \dots + \sum_{j=1}^k \Xi_{h-1,i,j} v_{j,t+1} \quad (98)$$

Sous cette forme, on pourrait mesurer la contribution de l'ensemble des innovations se produisant à un temps $t+l$ à la variance de l'erreur à l'horizon h , avec $1 \leq l \leq h$.

Si on désire faire apparaître la contribution à cette variance de chacune des

79. Par défaut, LAGMAX=min(12, T-2).

innovations v_j , $1 \leq j \leq k$, il suffit de réécrire l'expression précédente en les isolants, soit :

$$e_{i,t+h|t} = \sum_{l=1}^h \Xi_{h-l(i,1)} v_{1,t+l} + \sum_{l=1}^h \Xi_{h-l(i,2)} v_{2,t+l} + \cdots + \sum_{l=1}^h \Xi_{h-l(i,k)} v_{k,t+l} \quad (99)$$

Compte-tenu du fait que v_t est un processus en bruit blanc constitué de variables aléatoires de variance unitaire et orthogonales entre elles, il vient alors :

$$\text{var}(e_{i,t+h|t}) = E[e_{i,t+h|t}^2] = \sum_{l=1}^h \Xi_{h-l(i,1)}^2 + \sum_{l=1}^h \Xi_{h-l(i,2)}^2 + \cdots + \sum_{l=1}^h \Xi_{h-l(i,k)}^2 \quad (100)$$

et donc $\sum_{l=1}^h \Xi_{h-l(i,j)}^2$, avec $1 \leq j \leq k$ est la contribution du choc v_j à la variance de $e_{i,t+h|t}$.

En termes de proportions, ce qui a l'avantage de révéler rapidement quels chocs contribuent le plus à cette variance, la proportion de $\text{var}(e_{i,t+h|t})$ attribuable à v_j , $j = 1, 2, \dots, k$ est donnée par :

$$\frac{\sum_{l=1}^h \Xi_{h-l(i,j)}^2}{\sum_{l=1}^h \Xi_{h-l(i,1)}^2 + \sum_{l=1}^h \Xi_{h-l(i,2)}^2 + \cdots + \sum_{l=1}^h \Xi_{h-l(i,k)}^2} \quad (101)$$

Les valeurs estimées sont obtenues en remplaçant les termes $\Xi_{h-l(i,j)}^2$ par leurs estimations $\hat{\Xi}_{h-l(i,j)}^2$ dans (101). Enfin, la procédure de bootstrapping décrite lors de la construction d'IC sur les fonctions de réponse peut également être employée pour le calcul d'intervalles de confiance sur ces proportions.

L'affichage de la décomposition de la variance des erreurs de prévisions est demandé en spécifiant `DECOMPOSE(hmax)` dans l'option `PRINT` de la commande `MODEL`, le nombre `hmax` précisant l'horizon maximal de prévision sur lequel on désire cette décomposition⁸⁰. Outre les contributions des différents chocs à la variance, la procédure affiche également les parts attribuables à chacune d'eux conformément à l'équation (101).

Ainsi,

```
PROC VARMAX DATA=... PLOTS=(IMPULSE(SIMPLE ORTH));
MODEL y1 y2 y3 / P=2 PRINT=(IMPULSE(12)=(orth) decompose(36));
run;
```

demande l'affichage des 12 premières valeurs des fonctions de réponses orthogonalisées et leurs représentations graphiques ainsi que le graphe des fonctions de réponse simples. Les tableaux relatifs aux contributions absolues et relatives

80. Si `hmax` n'est pas précisé, VARMAX cale l'horizon maximal sur la valeur de `LAGMAX`.

des chocs dans la variance des erreurs de prévisions jusqu'à une horizon de 36 périodes sont aussi réclamés.

2.8 La causalité au sens de Granger

Soit un ensemble de k variables $y_t = (y_{1t}, \dots, y_{kt})$. Soit \tilde{Y}_{it} l'ensemble des réalisations de y_i jusqu'à la date t exclue,

$$\tilde{Y}_{it} = \{y_{i,\tau}\}, \tau < t$$

et $\tilde{Y}_t = \{\tilde{Y}_{it}\}, i = 1, \dots, k$, l'historique des réalisations des k variables jusqu'en $t-1$. On note $\tilde{Y}_t^{(i)}$ l'ensemble constitué par \tilde{Y}_t privé des réalisations passées de la variable y_i . Soit enfin ${}_t y_{j,t+1}$ le prédicteur linéaire optimal au sens du critère MSE formulé en t pour y_j à la date $t+1$.

A la suite de Granger⁸¹, on dira que y_i ne cause pas y_j au sens de Granger si

$$\forall t, {}_t y_{j,t+1} | \tilde{Y}_t = {}_t y_{j,t+1} | \tilde{Y}_t^{(i)} \quad (102)$$

Dans ce cas, l'information contenue dans l'historique de y_i est non pertinente pour prévoir y_j à un horizon d'une période une fois pris en compte les historiques des autres variables., ce que l'on notera $y_i \xrightarrow{(1)} y_j$. Dans le cas contraire, on dira que y_i cause y_j selon Granger, ce qui sera noté $y_i \xrightarrow{(1)} y_j$. Le concept de causalité d'une variable y_i vers une autre variable y_j présenté par Granger est donc directement lié à la capacité prédictive de y_i : lorsque, conditionnellement à un ensemble d'informations, l'anticipation en t de $y_{j,t+1}$ n'est pas modifiée par le retrait du passé de y_i de cet ensemble d'informations alors $y_i \xrightarrow{(1)} y_j$.

Quelques remarques s'imposent déjà :

1. On ne traite que des dépendances linéaires,
2. La définition précédente est donné pour un horizon de prévision d'une seule période. Nous aurons à discuter de sa généralisation à un horizon de h périodes, $h > 1$, et des rapports qui existent, où n'existent pas, entre $y_i \xrightarrow{(1)} y_j$ et $y_i \xrightarrow{(h)} y_j$ et entre $y_i \xrightarrow{(1)} y_j$ et $y_i \xrightarrow{(h)} y_j$.
3. Soit e le vecteur des erreurs de prévision à une période : $e_{t+1} = y_{t+1} - {}_t y_{t+1}$. La définition donnée en (102) est équivalente à

$$V(e_{t+1} | \tilde{Y}_t) = V(e_{t+1} | \tilde{Y}_t^{(i)})$$

81. Granger, C. W. J., *Investigating causal relations by econometric models and cross-spectral methods*, *Econometrica*, 1969, 37, 424-459.

4. On ne doit pas confondre le concept de causalité selon Granger avec le sens plus profond du terme causalité qui renvoie à une notion de responsabilité. Chez Granger, ce qui est essentiel est l'antériorité temporelle : si dans les mouvements passés de y_i il y a une information corrélée avec les mouvements futurs de y_j et que cette information n'est pas en totalité déjà présente dans les évolutions passées des autres variables alors $y_i \xrightarrow{(1)} y_j$.
Il se peut que rien ne soit vraiment causal dans cette causalité selon Granger⁸². On pourra ainsi remplacer les affirmations " y_i cause y_j selon Granger" ou " y_i ne cause pas y_j selon Granger" par " y_i est (ou n'est pas) un prédicteur avancé de y_j ". Il faudra en tout cas éviter d'employer les seuls termes " y_i cause y_j ".
5. On ne peut conclure que y_i est ou n'est pas un prédicteur avancé de y_j que conditionnellement à un ensemble d'informations \tilde{Y}_t . Si \tilde{Z}_t est un autre ensemble de variables en partie différent de \tilde{Y}_t ⁸³, alors chacune des propositions $y_i \xrightarrow{(1)} y_j | \tilde{Y}_t$ ou $y_i \not\xrightarrow{(1)} y_j | \tilde{Y}_t$ est compatible avec l'une et l'autre des propositions $y_i \xrightarrow{(1)} y_j | \tilde{Z}_t$ ou $y_i \not\xrightarrow{(1)} y_j | \tilde{Z}_t$.

Un exemple bien connu est donné par les travaux de Sims : dans un des premiers articles utilisant le concept de causalité à la Granger⁸⁴, il conclut que dans un ensemble d'informations limité à deux variables, la quantité de monnaie et l'activité réelle, la première est un prédicteur avancé de la seconde alors que l'activité ne cause pas selon Granger la quantité de monnaie. En revanche, dans un travail ultérieur⁸⁵ où l'ensemble d'information est constitué de trois variables (la quantité de monnaie, l'activité réelle, et un taux d'intérêt), la causalité au sens de Granger entre monnaie et activité n'est plus mise en évidence : c'est au taux d'intérêt qu'est attribué l'essentiel de l'impact sur l'activité réelle.

6. lorsque $y_i \xrightarrow{(1)} y_j$ et $y_j \xrightarrow{(1)} y_i$ on dira qu'il existe des retro-actions (feedback) entre les deux variables : chacune est un prédicteur avancé de l'autre.
7. L'absence de causalité à la Granger entre y_i et y_j ne signifie pas l'indépendance des deux variables. D'une part car seules les dépendances linéaires éventuelles sont saisies, et on sait qu'hormis pour des variables gaussiennes, l'orthogonalité n'équivaut pas à l'indépendance. D'autre part, car même si seules des dépendances linéaires sont autorisées, le

82. Par exemple vous savez que certaines personnes ont des rhumatismes qui deviennent douloureux à l'approche d'un changement de temps. En conséquence les douleurs rhumatismales causent au sens de Granger les variations des conditions climatiques. Clairement, si ces douleurs sont effectivement des prédicteurs avancés des changements météorologiques, la vraie causalité est inverse.

83. \tilde{Y}_t et \tilde{Z}_t doivent au moins avoir en commun les historiques de y_i et de y_j .

84. Sims, C. A., *Money, income, and causality*. American Economic Review, 1972, 62, 540-52.

85. Sims, C. A., *Comparison of interwar and postwar business cycles : Monetarism reconsidered*, American Economic Review, 1980, 70, 250-57.

fait que $y_i \xrightarrow{(1)} y_j$ ou que $y_j \xrightarrow{(1)} y_i$ n'interdit pas $cov(y_{it}, y_{jt}) \neq 0$: les valeurs contemporaines des variables peuvent être corrélées entre elles.

2.8.1 Causalité dans un système bivarié

Le cadre le plus simple pour l'étude de la causalité au sens de Granger est celui d'un système constitué uniquement de deux variables x et y . Il permet normalement de se familiariser aisément avec les outils et représentations qui seront mobilisés dans des cadres plus généraux.

Nous verrons tout d'abord ce qu'implique l'absence de causalité selon Granger d'une variable x vers une variable y sur les coefficients du VAR afférent au couple $(x_t, y_t)^\top$. Dans un second temps, nous montrerons l'équivalence des approches VAR ou VMA pour traiter de la causalité selon Granger dans ce cadre bivarié. Dans un troisième temps, nous exposerons la mise en oeuvre des tests de causalité dans la procédure VARMAX.

Causalité au sens de Granger et coefficients du VAR.

Soit donc $z_t = (x_t, y_t)^\top$. Conformément au cadre retenu jusqu'alors, z_t est supposé être stationnaire. On a donc une écriture de Wold :

$$z_t = \mu_z + \Psi(L)u_t$$

où u_t est un processus en bruit blanc. Si ce processus est inversible, on a également une écriture VAR, éventuellement d'ordre infini :

$$\Phi(L)z_t = c + u_t,$$

avec $\phi(L) = I - \phi_1 L - \phi_2 L^2 + \dots = \Psi(L)^{-1}$ et $c = \Psi(L)^{-1} \mu_z$.

Si x n'est pas un prédicteur avancé de y , alors, d'après la définition, les coefficients des x retardés dans l'équation de y doivent être nuls. En effet, si tel est le cas, la présence ou l'absence de l'historique de x dans cette équation ne modifie évidemment pas la prévision que l'on peut faire de y , qui est alors tirée d'un processus univarié de type AR. L'absence de causalité à la Granger revient donc à imposer des contraintes de nullité sur les coefficients hors-diagonaux des matrices $\Phi_i, i = 1, 2, \dots$. Ainsi, pour un VAR(1), on aura :

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{bmatrix} \phi_{xx} & \phi_{xy} \\ \phi_{yx} & \phi_{yy} \end{bmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} c_x \\ c_y \end{pmatrix} + \begin{pmatrix} u_{x_t} \\ u_{y_t} \end{pmatrix}$$

et :

$$\begin{aligned}
 x \rightarrow y &\Leftrightarrow \phi_{yx} = 0 & x \rightarrow y &\Leftrightarrow \phi_{yx} \neq 0 \\
 &\Leftrightarrow \Phi(L) = \begin{bmatrix} 1 - \phi_{xx}L & -\phi_{xy}L \\ 0 & 1 - \phi_{yy}L \end{bmatrix} \\
 y \rightarrow x &\Leftrightarrow \phi_{xy} = 0 & y \rightarrow x &\Leftrightarrow \phi_{xy} \neq 0 \\
 &\Leftrightarrow \Phi(L) = \begin{bmatrix} 1 - \phi_{xx}L & 0 \\ -\phi_{yx}L & 1 - \phi_{yy}L \end{bmatrix}
 \end{aligned}$$

Pour un VAR(2) :

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{bmatrix} \phi_{xx,1} & \phi_{xy,1} \\ \phi_{yx,1} & \phi_{yy,1} \end{bmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{bmatrix} \phi_{xx,2} & \phi_{xy,2} \\ \phi_{yx,2} & \phi_{yy,2} \end{bmatrix} \begin{pmatrix} x_{t-2} \\ y_{t-2} \end{pmatrix} + \begin{pmatrix} c_x \\ c_y \end{pmatrix} + \begin{pmatrix} u_{x_t} \\ u_{y_t} \end{pmatrix}$$

avec,

$$\begin{aligned}
 x \rightarrow y &\Leftrightarrow \phi_{yx,1} = \phi_{yx,2} = 0 & x \rightarrow y &\Leftrightarrow \phi_{yx,1} \neq 0 \text{ ou } \phi_{yx,2} \neq 0 \\
 &\Leftrightarrow \Phi(L) = \begin{bmatrix} 1 - \phi_{xx,1}L - \phi_{xx,2}L^2 & -\phi_{xy,1}L - \phi_{xy,2}L^2 \\ 0 & 1 - \phi_{yy,1}L - \phi_{yy,2}L^2 \end{bmatrix} \\
 y \rightarrow x &\Leftrightarrow \phi_{xy,1} = \phi_{xy,2} = 0 & y \rightarrow x &\Leftrightarrow \phi_{xy,1} \neq 0 \text{ ou } \phi_{xy,2} \neq 0 \\
 &\Leftrightarrow \Phi(L) = \begin{bmatrix} 1 - \phi_{xx,1}L - \phi_{xx,2}L^2 & 0 \\ -\phi_{yx,1}L - \phi_{yx,2}L^2 & 1 - \phi_{yy,1}L - \phi_{yy,2}L^2 \end{bmatrix}
 \end{aligned}$$

Plus généralement, pour un VAR d'ordre p quelconque, si on partitionne $\phi(L)$ conformément au découpage pris sur z_t , il vient :

$$\begin{aligned}
 x \rightarrow y &\Leftrightarrow \Phi(L) = \begin{bmatrix} 1 - \sum_{i=1}^p \phi_{xx,i}L^i & -\sum_{i=1}^p \phi_{yx,i}L^i \\ 0 & 1 - \sum_{i=1}^p \phi_{yy,i}L^i \end{bmatrix} \\
 y \rightarrow x &\Leftrightarrow \Phi(L) = \begin{bmatrix} 1 - \sum_{i=1}^p \phi_{xx,i}L^i & 0 \\ -\sum_{i=1}^p \phi_{yx,i}L^i & 1 - \sum_{i=1}^p \phi_{yy,i}L^i \end{bmatrix}
 \end{aligned}$$

Dans l'un ou l'autre cas, $\phi(L)$ est une matrice triangulaire, triangulaire supérieure lorsque $x \rightarrow y$, triangulaire inférieure lorsque $y \rightarrow x$.

Conséquence de la non causalité pour la représentation VMA.

Si le VAR est stable, on peut basculer vers sa représentation VMA infinie : $\Phi(L)z_t = c + u_t \Rightarrow z_t = \Psi(L)u_t + \mu_z$, avec $\Psi(L) = \Phi(L)^{-1}$. Or l'inverse d'une triangulaire est une triangulaire de même nature. Dans ces conditions :

$$z_t = \begin{pmatrix} x_t \\ y_t \end{pmatrix} = \Psi(L) \begin{pmatrix} u_{x_t} \\ u_{y_t} \end{pmatrix} + \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}$$

et,

$$\begin{aligned}
 x \rightarrow y & \Leftrightarrow \Psi(L) = \begin{bmatrix} 1 + \sum_{i=1}^{\infty} \psi_{xx,i}L^i & \sum_{i=1}^{\infty} \psi_{yx,i}L^i \\ 0 & 1 - \sum_{i=1}^{\infty} \psi_{yy,i}L^i \end{bmatrix} \\
 y \rightarrow x & \Leftrightarrow \Psi(L) = \begin{bmatrix} 1 + \sum_{i=1}^{\infty} \psi_{xx,i}L^i & 0 \\ \sum_{i=1}^{\infty} \psi_{yx,i}L^i & 1 + \sum_{i=1}^{\infty} \psi_{yy,i}L^i \end{bmatrix}
 \end{aligned}$$

Ainsi, dans un VAR bivarié, on a l'ensemble des équivalence suivantes :

- $x \rightarrow y \Leftrightarrow \phi_{yx,i} = 0, i = 1 \dots, p$: les coefficients des x retardés dans l'équation de y sont nuls,
- $\Leftrightarrow \psi_{yx,i} = 0, i = 1 \dots \infty$: la fonction de réponse de y au choc u_x est nulle.
- \Leftrightarrow la totalité de la variance des erreurs de prévision commises sur y pour n'importe quel horizon est attribuable à u_y

Naturellement, le même type de conclusions est obtenu dans le cas symétrique où $y \rightarrow x$.

En d'autres termes, dans le cadre d'un VAR bivarié, on peut indifféremment traiter de la causalité selon Granger à partir de l'écriture VAR et de l'écriture VMA. Malheureusement, ainsi que nous le verrons ultérieurement, cette équivalence ne tient plus si on traite de la causalité de x vers y dans un VAR de dimension supérieure, i.e. lorsque $z_t = (x_t, y_t, w_t^\top)^\top$ avec $w_t = (w_{1t}, \dots, w_{n_w,t})^\top$ et $n_w \geq 1$

Tests de non causalité avec la Proc VARMAX.

La plupart des travaux réalisés dans un cadre multivarié partent d'une représentation VAR(p), que le modèle vrai soit un VAR d'ordre fini, ou infini⁸⁶. Dans cette représentation, les développements qui précèdent ont montré que :

$$\begin{aligned}
 H0 : x \not\rightarrow y & \text{ versus } H1 : x \xrightarrow{1} y \\
 \Leftrightarrow H0 : \phi_{yx,1} = \dots = \phi_{yx,p} = 0 & \text{ versus } H1 : \exists i, 1 \leq i \leq p, \phi_{yx,i} \neq 0 \\
 \Leftrightarrow H0 : \psi_{yx,1} = \psi_{yx,2} = \dots = 0 & \text{ versus } H1 : \exists i, 1 \leq i \leq \infty, \psi_{yx,i} \neq 0
 \end{aligned}$$

Disposant de l'estimation d'un VAR, il est plus simple de tester les restrictions imposées par $H0$ sur les coefficients du VAR, qui ne sont qu'un nombre fini de contraintes linéaires sur ces coefficients, plutôt que sur ceux de l'écriture VMA, les coefficients de celle-ci étant des combinaisons non linéaires des précédents, ce qui complique le calcul de la statistique.

Classiquement, trois tests peuvent être réalisés :

⁸⁶. Dans ce dernier cas l'utilisateur pense disposer d'une approximation satisfaisante du vrai processus

- Un test LRT : estimation du modèle non contraint correspondant à l'estimation par OLS de l'équation afférente à y avec présence des x et y retardés à l'ordre p , récupération de l_{nc} . Estimation du modèle contraint correspondant à la régression par OLS de y_t sur ses propres p premières valeurs retardées, récupération de l_c . Construction de la statistique $LRT = -2(l_c - l_{nc})$ qui, sous H_0 est la réalisation d'un χ^2 à p degrés de liberté⁸⁷.
- Un test de Lagrange, ou du score, qui passe par l'estimation du seul modèle contraint.
- Un test de Wald qui nécessite l'estimation du modèle non contraint, celui-ci n'étant rien d'autre que le VAR(p) estimé initialement.

Les deux derniers peuvent être aisément obtenus dans VARMAX en faisant appel respectivement aux commandes RESTRICT et TEST. Pour le problème qui nous intéresse ici, RESTRICT est toutefois non adaptée : elle renvoie un test d'hypothèse simple pour chaque restriction et non pas un test joint sur l'ensemble des restrictions indiquées dans la commande⁸⁸. Aussi, à moins d'avoir un VAR(1), on mobilisera le plus souvent la commande TEST qui lui affiche le test joint. Pour les mettre en oeuvre il faut naturellement être capable d'indiquer à la procédure quels sont les coefficients impliqués dans les tests. Pour cela, la syntaxe suivante est utilisée :

$AR(l, i, j)$: coefficient de la $j^{\text{ème}}$ variable retardée de l périodes dans l'équation de la $i^{\text{ème}}$ variable

Supposons par exemple un VAR(2) toujours sur $z_t = (x_t, y_t)^\top$. L'absence de causalité au sens de Granger de x vers y renvoie au test :

$$H_0 : \phi_{yx,1} = \psi_{yx,2} = 0 \quad \text{versus } H_1 : \phi_{yx,1} \neq 0 \text{ ou } \psi_{yx,2} \neq 0$$

On aura un programme du type :

```
PROC VARMAX DATA=...;
MODEL x y / P=2;
TEST AR(1,2,1)=0,AR(2,2,1)=0;
run;
```

En effet l'ordre d'entrée des variables dans le système VAR est imposé par la commande MODEL, la variable x est identifiée comme étant l'expliquée de la première équation, et y celle de la seconde. En conséquence :

$AR(1,2,1)$ = coefficient de x_{t-1} dans l'équation de y_t

$AR(2,2,1)$ = coefficient de x_{t-2} dans l'équation de y_t

⁸⁷. Naturellement, chacune des régressions précédentes peut éventuellement incorporer une constante.

⁸⁸. Un multiplicateur de Lagrange est estimé pour chacune des conditions listées dans la commande RESTRICT, et un test de significativité de chacun d'eux est réalisé.

La statistique de Wald sera alors la réalisation d'un χ^2 à 2 degrés de liberté sous l'hypothèse nulle d'absence de causalité selon Granger de x vers y ⁸⁹.

Pour réaliser ce test de Wald de non causalité selon Granger, VARMAX dispose également de la commande CAUSAL. La syntaxe est la suivante :

CAUSAL GROUP1=(y_1, \dots, y_{k_y}) GROUP2=(x_1, \dots, x_{k_x});

où (y_1, \dots, y_{k_y}) et (x_1, \dots, x_{k_x}) sont des ensembles disjoints de respectivement k_x et k_y variables qui doivent être obligatoirement référencées dans la commande MODEL. Elle est donc utilisable dans des VAR de dimension $\geq k_x + k_y$. Son objet est de tester l'absence de causalité selon Granger des variables constituant le groupe 2 vers les variables du groupe 1. Clairement, un VAR bivarié est un cas particulier où GROUP1 et GROUP2 ne sont composés chacun que d'une seule variable. Elle réalise les actions suivantes :

- extraction parmi l'ensemble des variables apparaissant dans la commande MODEL des variables référencées par GROUP1 et GROUP2 ,
- construction d'un VAR dont la structure (présence d'une constante ou pas, nombre de retards,...) est guidée par les options données dans la commande MODEL
- test de Wald de la nullité des coefficients des variables retardées référencées par GROUP2 dans les équations dont les expliquées sont celles précisées par GROUP1.

par ailleurs, on peut faire apparaître plusieurs appels à la commande CAUSAL. Dans ces conditions, le programme précédent peut être avantageusement remplacé par :

```
PROC VARMAX DATA=...;
MODEL x y / P=2;
CAUSAL GROUP1=(y) GROUP2=(x) ;
CAUSAL GROUP1=(x) GROUP2=(y) ;
run;
```

Le premier CAUSAL porte sur $H_0 : x \xrightarrow{1} y$ et fournit évidemment la même statistique de Wald que notre commande TEST AR(1,2,1)=0,AR(2,2,1)=0; , le deuxième effectue le test de $H_0 : y \xrightarrow{1} x$ et est donc équivalent à la commande TEST AR(1,1,2)=0,AR(2,1,2)=0.

Enfin, notez que sur petits échantillons, on conseille à la suite de Sims, de remplacer le test de Wald par un test de Fisher, la statistique de Wald n'atteignant alors pas sa distribution asymptotique. Pour mémoire, on a : $F=W/c$, où W est la statistique de Wald, c le nombre de contraintes testées et F la statistique

⁸⁹. Si on avait fait RESTRICT AR(1,2,1)=0,AR(2,2,1)=0; on aurait reçu en retour deux tests d'hypothèse simple de nullité des deux multiplicateurs de Lagrange associés respectivement à chacune des deux contraintes. Par ailleurs VARMAX aurait aussi affiché l'estimation du VAR contraint, *i.e.* avec les deux coefficients référencés mis à 0.

de Fisher construite pour le test de ces mêmes contraintes. Dans nos exemples, elle possède une distribution de Fisher à p et $(T - 2p - 1)$ degrés de liberté.

2.8.2 Causalité entre deux ensembles de variables : les mesures de GEWEKE

GEWEKE⁹⁰ propose d'appréhender la causalité selon Granger en termes de mesures de dépendance. Cela va lui permettre d'offrir une description des relations dynamiques à la fois cohérente et plus complète que la seule causalité au sens de Granger présentée ci-dessus au moins dans les systèmes linéaires.

La cohérence tient à ce que la mesure de dépendance complète de X et Y sera la somme des mesures de dépendance de X au passé de Y , de Y au passé de X et de la mesure de dépendance instantanée. L'approche est également plus complète puisque seule les deux premières formes de dépendance sont appréhendées dans la causalité au sens de Granger.

Par ailleurs, l'approche de GEWEKE peut aisément s'appliquer à l'étude des causalités entre deux blocs de variables disjoints mais exhaustifs : supposons que le vecteur d'intérêt soit un ensemble de k variable que l'on accepte de décomposer en deux sous-ensembles :

$$\begin{aligned} z_t &= (x_{1t}, \dots, x_{n_X,t}, y_{1t}, \dots, y_{n_Y,t})^\top \\ &= (X_t^\top, Y_t^\top)^\top \end{aligned}$$

où

$$X_t = \begin{pmatrix} x_{1t} \\ \vdots \\ x_{n_X,t} \end{pmatrix}, \quad Y_t = \begin{pmatrix} y_{1t} \\ \vdots \\ y_{n_Y,t} \end{pmatrix}$$

l'exhaustivité signifiant que Z est de dimension $(n_X + n_Y, 1)$. Il s'agit alors d'étudier les dépendances entre le bloc X et le bloc Y . On voit aussi que le cadre bivarié traité dans le paragraphe précédent est un cas particulier du système étudié maintenant, cas particulier dans lequel $n_X = n_Y = 1$.

Pour cela, Geweke considère

1. les autorégressions marginales, où l'état en t de chacun des blocs est projeté sur son seul passé :

$$X_t = \Phi_X(L)X_t + u_{X,t} = \sum_{i=1}^p \Phi_{X,1} X_{t-1} + \dots + \Phi_{X,p} X_{t-p} + u_{X,t} \quad (103)$$

$$Y_t = \Phi_Y(L)Y_t + u_{Y,t} = \sum_{i=1}^p \Phi_{Y,1} Y_{t-1} + \dots + \Phi_{Y,p} Y_{t-p} + u_{Y,t} \quad (104)$$

⁹⁰. Geweke, J., *Measurement of Linear Dependence and feedback between multiple time series*, JASA, 1982, 77, 304-313.

où $u_{X,t}$ et $u_{Y,t}$ sont supposés être des processus en bruit blanc⁹¹, de matrices de variance covariance respectives Σ_{u_X} et Σ_{u_Y} .

2. les autorégressions jointes, où l'état en t de chacun des blocs est projeté sur son propre passé et le passé de l'autre bloc⁹² :

$$X_t = \Phi_{XX}(L)X_t + \Phi_{XY}(L)Y_t + u_{XY,t} \quad (105)$$

$$= \sum_{i=1}^p \Phi_{XX,i}X_{t-i} + \dots + \Phi_{XX,p}X_{t-p} + \sum_{i=1}^p \Phi_{XY,i}X_{t-i} + \dots + \Phi_{XY,p}X_{t-p} + u_{XY,t}$$

$$Y_t = \Phi_{YX}(L)X_t + \Phi_{YY}(L)Y_t + u_{YX,t} \quad (106)$$

$$= \sum_{i=1}^p \Phi_{YX,i}X_{t-i} + \dots + \Phi_{YX,p}X_{t-p} + \sum_{i=1}^p \Phi_{YY,i}Y_{t-i} + \dots + \Phi_{YY,p}Y_{t-p} + u_{YX,t}$$

où $u_{XY,t}$ et $u_{YX,t}$ sont supposés être des processus en bruits blanc, de matrices de variance covariance respectives $\Sigma_{u_{XY}}$ et $\Sigma_{u_{YX}}$.

3. Le VAR(p) complet :

$$Z_t = \Phi_Z(L)Z_t + u_{Z_t} \quad (107)$$

$$\Leftrightarrow \begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{bmatrix} \Phi_{XX}(L) & \Phi_{XY}(L) \\ \Phi_{YX}(L) & \Phi_{YY}(L) \end{bmatrix} \begin{pmatrix} X_t \\ Y_t \end{pmatrix} + u_{Z_t}$$

avec $E(u_{Z_t} u_{Z_t}^\top) = \Sigma_{u_Z}$.

Comme l'ordre de troncature p est identique sur le VAR et sur les autorégressions jointes, on note que les n_X premières équations du VAR(p) ne sont rien d'autre que les autorégressions jointes afférentes au bloc X et que les n_Y dernières équations du VAR(p) sont les jointes afférentes au bloc Y. En conséquence :

$$u_{Z_t} = \begin{pmatrix} u_{XY,t} \\ u_{YX,t} \end{pmatrix} \quad (108)$$

et,

$$\Sigma_{u_Z} = \begin{bmatrix} \Sigma_{u_{XY}} & Cov(u_{XY,t}, u_{YX,t}) \\ Cov(u_{XY,t}, u_{YX,t})^\top & \Sigma_{u_{YX}} \end{bmatrix} \quad (109)$$

91. Ainsi, l'autorégression marginale de X (resp. de Y) est un VAR(p) sur X (resp. sur Y)

92. La règle suivante est adoptée en ce qui concerne les indices des différents objets : le premier indice précise l'expliquée du bloc, le second, l'explicative. Par exemple $\Phi_{YX,p}$ donne les coefficients des variables constituant le bloc X retardées de p périodes dans l'explication de l'état en t du vecteur Y.

Enfin, du fait de l'exhaustivité de X et Y par rapport à Z , le VAR(p) sur Z_t peut être vu comme un VAR(p) bivarié sur les blocs X_t et Y_t . Ainsi,

$$\begin{aligned}
Y \rightarrow X &\Leftrightarrow \Phi_{XY}(L) = 0 \text{ dans (105)} \\
&\Leftrightarrow \text{autorégressions marginale et jointe sur } X \text{ sont identiques} \\
&\Leftrightarrow u_{Xt} = u_{XY,t} \\
&\Rightarrow \Sigma_{u_X} = \Sigma_{u_{XY}} \\
&\Leftrightarrow |\Sigma_{u_X}| = |\Sigma_{u_{XY}}| \\
X \rightarrow Y &\Leftrightarrow \Phi_{YX}(L) = 0 \text{ dans (106)} \\
&\Leftrightarrow \text{autorégressions marginale et jointe sur } Y \text{ sont identiques} \\
&\Leftrightarrow u_{Yt} = u_{YX,t} \\
&\Rightarrow \Sigma_{u_Y} = \Sigma_{u_{YX}} \\
&\Rightarrow |\Sigma_{u_Y}| = |\Sigma_{u_{YX}}|
\end{aligned}$$

En revanche :

$$\begin{aligned}
Y \rightarrow X &\Leftrightarrow \Phi_{XY}(L) \neq 0 \text{ dans (105)} \\
&\Leftrightarrow \text{autorégressions marginale et jointe sur } X \text{ sont différentes} \\
&\Leftrightarrow \text{la composante inexpliquée de } X \text{ dans la jointe est plus "petite"} \\
&\quad \text{que la composante inexpliquée de } X \text{ dans la marginale} \\
&\Leftrightarrow |\Sigma_{u_X}| > |\Sigma_{u_{XY}}|, \text{ l'inégalité étant d'autant plus forte que le contenu en} \\
&\quad \text{information du passé de } Y \text{ est élevé pour expliquer le présent de } X. \\
X \rightarrow Y &\Leftrightarrow \Phi_{YX}(L) \neq 0 \text{ dans (106)} \\
&\Leftrightarrow \text{autorégressions marginale et jointe sur } Y \text{ sont différentes} \\
&\Leftrightarrow \text{la composante inexpliquée de } Y \text{ dans la jointe est plus "petite"} \\
&\quad \text{que la composante inexpliquée de } Y \text{ dans la marginale} \\
&\Leftrightarrow |\Sigma_{u_Y}| > |\Sigma_{u_{YX}}|, \text{ l'inégalité étant d'autant plus forte que le contenu en} \\
&\quad \text{information du passé de } X \text{ est élevé pour expliquer le présent de } Y.
\end{aligned}$$

On remarque donc que, pour statuer sur la présence ou l'absence de causalité selon Granger entre les deux blocs de variables X et Y , il faut comparer les déterminants des matrices de variance-covariance des innovations des autorégressions marginales d'une part et jointes d'autre part. Maintenant, en considérant le VAR complet sur Z , il est également possible de statuer sur l'existence ou non de dépendance linéaires instantanées entre les deux blocs.

Une dépendance instantanée signifie ici qu'une innovation affectant l'une des composantes de Z en t affecte à cette même date t une ou plusieurs autres composante de Z . On sait déjà que la matrice Σ_{u_Z} du VAR complet n'est pas nécessairement diagonale : un choc sur la $i^{\text{ème}}$ composante de u_Z en t peut affecter sa la $j^{\text{ème}}$ composante si la covariance des deux innovations n'est pas nulle. Ce sont précisément ces dépendances instantanées qui nous ont amené

à passer des innovations u aux innovations orthogonales v lors de l'étude des fonctions de réponse et des décompositions de la variance des erreurs de prévision. En conséquence, les dépendances instantanées entre les variables du bloc X et celles du bloc Y proviennent des covariances qui peuvent exister entre les innovations afférentes à ces deux blocs. Si ces covariances sont nulles, il n'y a pas de dépendance instantanée entre les deux ensembles de variables : un choc d'innovation en t sur la $i^{\text{ème}}$ composante de X affectera évidemment x_{it} et éventuellement d'autres variables de X si Σ_{u_X} n'est pas diagonale, mais elle ne touchera pas les innovations afférentes au bloc Y et donc il n'y aura pas de réponse des composantes y_1, \dots, y_{n_Y} à cette date t .

Comme le montre (109), les covariances des deux ensembles d'innovation sont les blocs hors diagonale dans Σ_{u_Z} . Si on note $X \leftrightarrow Y$ la présence de dépendances instantanées et $X \nleftrightarrow Y$ son absence, il vient :

$$X \leftrightarrow Y \Leftrightarrow \Sigma_{u_Z} = \begin{bmatrix} \Sigma_{u_{XY}} & 0 \\ 0 & \Sigma_{u_{YX}} \end{bmatrix}$$

$$\Leftrightarrow |\Sigma_{u_Z}| = |\Sigma_{u_{XY}}| |\Sigma_{u_{YX}}|.$$

$X \leftrightarrow Y \Leftrightarrow$ la composante inexpliquée de Z dans le VAR complet, qui tient compte des covariances instantanées entre les innovations des deux blocs, est plus "petite" que la composante inexpliquée de Z associée aux autorégressions jointes qui ne les prend pas en compte.
 $\Leftrightarrow |\Sigma_{u_Z}| \leq |\Sigma_{u_{XY}}| |\Sigma_{u_{YX}}|$, l'inégalité étant d'autant plus forte que les covariances inter-blocs des innovations sont élevées.

Il s'agit maintenant de construire des mesures de dépendance pour les trois cas possibles, $X \rightarrow Y$, $Y \rightarrow X$ et $X \leftrightarrow Y$, sachant que chacune de ces mesures devrait être nulle si le cas de dépendance qu'elle représente n'existe pas, et devrait être d'autant plus élevée que la dépendance en question est forte. Par ailleurs, on peut aussi être intéressé à la construction d'une mesure globale des dépendances linéaires entre les deux blocs qui soit cohérente avec les trois précédentes : la somme de celles-ci devrait être égale à celle-là. Les mesures proposées par GEWEKE, qui respectent ces exigences, sont les suivantes :

— Mesure de dépendance de Y au passé de X :

$$C_{X \rightarrow Y} = \log \frac{|\Sigma_{u_Y}|}{|\Sigma_{u_{YX}}|} \quad (110)$$

— Mesure de dépendance de X au passé de Y :

$$C_{Y \rightarrow X} = \log \frac{|\Sigma_{u_X}|}{|\Sigma_{u_{XY}}|} \quad (111)$$

— Mesure de dépendance instantanée⁹³ entre X et Y :

$$C_{Y \leftrightarrow X} = \log \frac{|\Sigma_{u_{XY}}| |\Sigma_{u_{YX}}|}{|\Sigma_{u_Z}|} \quad (112)$$

— Mesure de dépendance globale entre X et Y :

$$C_{Y,X} = C_{X,Y} \quad (113)$$

$$= C_{X \rightarrow Y} + C_{Y \rightarrow X} + C_{X \leftrightarrow Y} \quad (114)$$

$$= \log \frac{|\Sigma_{u_X}| |\Sigma_{u_Y}|}{|\Sigma_{u_Z}|} \quad (115)$$

En ce qui concerne cette dernière mesure, on peut voir aisément que l'absence de dépendance signifie l'absence de causalité selon Granger de X vers Y , $|\Sigma_{u_Y}| = |\Sigma_{u_{YX}}|$, l'absence de causalité de Y vers X , $|\Sigma_{u_X}| = |\Sigma_{u_{XY}}|$, et l'absence de dépendance instantanée, $|\Sigma_{u_{XY}}| |\Sigma_{u_{YX}}| = |\Sigma_{u_Z}|$. Au total, dans ce cas, on vérifie bien $C_{Y,X} = 0$. Dès lors qu'il y a des causalités, la partie inexpliquée de Z qui tient compte des 3 cas possibles de dépendance⁹⁴ diminue par rapport à ce qu'expliquent les autorégressions marginales qui n'en retiennent aucun : $C_{X,Y}$ sera donc croissant avec l'intensité des dépendances.

Un point mérite encore d'être souligné : dans les développements qui précèdent une hypothèse centrale est que $Z = (X^T, Y^T)^T$ est gouverné par un VAR(p) stable. En conséquence, s'il est évident que les régressions jointes doivent être tronquées à l'ordre p , il n'en est pas de même pour les autorégressions marginales : si les jointes sont d'ordre p , il n'est absolument pas certain que le fait de retirer le passé d'un des deux blocs ne modifie pas l'ordre des retards sur le bloc restant : la marginale n'a aucune raison d'être d'ordre p . Ce point a été éclairci par Gourieroux, Monfort & Renault⁹⁵ qui justifient que toutes les autorégressions doivent effectivement être tronquées à l'ordre du VAR pour construire ces mesures de causalité.

En pratique, on construira des estimateurs des mesures à partir de ceux obtenus sur les diverses matrices de variance-covariance :

— Mesure de dépendance estimée de Y au passé de X :

$$\widehat{C}_{X \rightarrow Y} = \log \frac{|\widehat{\Sigma}_{u_Y}|}{|\widehat{\Sigma}_{u_{YX}}|} \quad (116)$$

93. Notez qu'une dépendance instantanée reflète des co-mouvements contemporains systématiques entre les diverses variables concernées : il est impossible d'identifier statistiquement la où les variables à l'origine de ces mouvements.

94. En effet, le VAR(p) sur Z prend en compte $C_{X \rightarrow Y}$ et $C_{Y \rightarrow X}$ via les équations du VAR, ainsi que $C_{X \leftrightarrow Y}$ via les blocs non diagonaux de Σ_Z .

95. Gourieroux, C., Monfort, A., Renault, E., *Kullback Causality Measures*, Annales d'Économie et de Statistique, 6/7, 1987, 369-410.

— Mesure de dépendance estimée de X au passé de Y :

$$\widehat{C}_{Y \rightarrow X} = \log \frac{|\widehat{\Sigma}_{u_X}|}{|\widehat{\Sigma}_{u_{XY}}|} \quad (117)$$

— Mesure de dépendance instantanée estimée entre X et Y :

$$\widehat{C}_{Y \leftrightarrow X} = \log \frac{|\widehat{\Sigma}_{u_{XY}}| |\widehat{\Sigma}_{u_{YX}}|}{|\widehat{\Sigma}_{u_Z}|} \quad (118)$$

— Mesure de dépendance globale estimée entre X et Y :

$$\widehat{C}_{Y,X} = \widehat{C}_{X,Y} \quad (119)$$

$$= \widehat{C}_{X \rightarrow Y} + \widehat{C}_{Y \rightarrow X} + \widehat{C}_{X \leftrightarrow Y} \quad (120)$$

$$= \log \frac{|\widehat{\Sigma}_{u_X}| |\widehat{\Sigma}_{u_Y}|}{|\widehat{\Sigma}_{u_Z}|} \quad (121)$$

Enfin, en se souvenant de (44), on sait que les log-vraisemblances des divers modèles font apparaître un terme de la forme $\frac{T}{2} \times$ [déterminant de la matrice de var-cov des résidus]. En conséquence, un test $LRT = -2(\hat{l}_c - \hat{l}_{nc})$ de l'hypothèse de nullité de chaque mesure est aisément construit :

— $H_0 : C_{X \rightarrow Y} = 0$ versus $H_1 : C_{X \rightarrow Y} > 0$
 modèle contraint = autorégression marginale
 modèle non contraint = autorégression jointe
 nombre de coefficients contraints à zéro = $p n_X n_Y$

$$\Rightarrow \text{ sous } H_0 : T \widehat{C}_{X \rightarrow Y} \xrightarrow[T \rightarrow \infty]{d} \chi^2(p n_X n_Y).$$

— $H_0 : C_{Y \rightarrow X} = 0$ versus $H_1 : C_{Y \rightarrow X} > 0$
 modèle contraint = autorégression marginale
 modèle non contraint = autorégression jointe
 nombre de coefficients contraints à zéro = $p n_X n_Y$

$$\Rightarrow \text{ sous } H_0 : T \widehat{C}_{Y \rightarrow X} \xrightarrow[T \rightarrow \infty]{d} \chi^2(p n_X n_Y).$$

— $H_0 : C_{X \leftrightarrow Y} = 0$ versus $H_1 : C_{X \leftrightarrow Y} > 0$
 modèle contraint = autorégressions jointes
 modèle non contraint = VAR complet
 nombre de coefficients contraints à zéro ⁹⁶ = $n_X n_Y$

96. Sous H_0 , il faut contraindre à zéro toutes les covariances entre les n_X innovations afférentes à X et les n_Y afférentes à Y . Si vous raisonnez sur Σ_{u_Z} il faut, pour traduire H_0 , y mettre à zéro tous les éléments des deux blocs hors diagonale de dimensions respectives (n_X, n_Y) et (n_Y, n_X) . En raison de la symétrie de la matrice de var-cov il suffit de contraindre à zéro un seul de ces blocs, ce qui donne bien le nombre indiqué.

$$\Rightarrow \text{ sous } H_0 : T\widehat{C}_{X \leftrightarrow Y} \xrightarrow[T \rightarrow \infty]{d} \chi^2(n_X n_Y).$$

- $H_0 : C_{X,Y} = 0$ versus $H_1 : C_{X,Y} > 0$
 modèle contraint = autorégressions marginales
 modèle non contraint = VAR complet
 nombre de coefficients contraints à zéro = $n_X n_Y(2p + 1)$

$$\Rightarrow \text{ sous } H_0 : T\widehat{C}_{Y,X} \xrightarrow[T \rightarrow \infty]{d} \chi^2(n_X n_Y[2p + 1]).$$

2.9 Un exemple

Afin d'illustrer les divers points abordés précédemment, nous allons travailler sur un vecteur $y = (y_1, y_2, y_3)^\top$ de longueur 100 dont les observations sont obtenues par simulation du VAR(1) suivant :

$$y_{1t} = 1.2y_{1t-1} - 0.5y_{2t-1} + u_{1t} \quad (122)$$

$$y_{2t} = 0.6y_{1t-1} + 0.3y_{2t-1} + u_{2t} \quad (123)$$

$$y_{3t} = .6y_{1t-1} + 0.4y_{2t-1} + 0.7y_{3t-1} + u_{3t} \quad (124)$$

soit

$$y_t = \Phi y_{t-1} + u_t \quad (125)$$

et

$$\Sigma_u = \begin{pmatrix} 1.0 & 0.5 & 0.3 \\ 0.5 & 1.25 & 0.4 \\ 0.3 & 0.4 & 0.5 \end{pmatrix} \quad (126)$$

On notera que dans ce VAR, y_3 n'est pas un prédictor avancé pour les deux autres variables et qu'il existe des covariances instantanées non nulles entre les trois résidus.

Les observations sont facilement créées par appel de la commande `varmasim` disponible dans la Proc IML dans laquelle on spécifie le contenu des matrices Φ et Σ_u via respectivement `phi=...` et `sig=...`. Le nombre d'observations simulées est précisé par `n=`. L'option `seed=34567` utilisée ici est utile si vous voulez reproduire les résultats de cet exemple. La commande `create=...` construit la table simul contenant les trois variables y_1 , y_2 et y_3 .

```
proc iml;
sig = 1.0 0.5 0.3 , 0.5 1.25 0.4, 0.3 0.4 .5;
phi = 1.2 -0.5 0., 0.6 0.3 0., .6 .4 .7;
call varmasim(y,phi) sigma = sig n = 100 seed = 34657;
cn = 'y1' 'y2' 'y3';
create simul from y[colname=cn];
append from y;
quit;
```

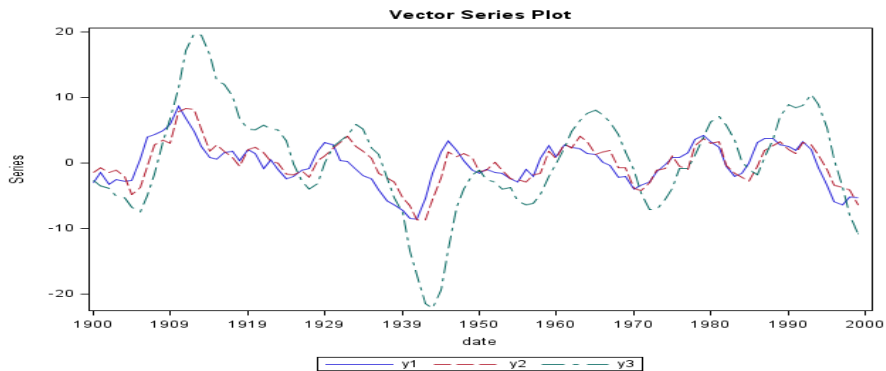


FIGURE 1 – Graphes des séries simulées

On ajoute ensuite une date dans le fichier de données en précisant que les données simulées correspondent à des observations mensuelles avec une fenêtre d'observation commençant en janvier 2007 :

```
data simul;
set simul;
date = intnx( 'month', '01jan2007'd, _n_-1 );
format date monyy.;
run;
```

Après exécution de toutes ces instructions, on peut obtenir dans la Figure 1 les représentations graphiques des trois séries de travail via par exemple :

```
proc timeseries data=simul vectorplot=series;
id date interval=month;
var y1 y2 y3;
run;
```

1. Détermination de l'ordre du VAR

Pour la recherche de l'ordre du VAR, nous allons utiliser d'une part l'option `minic`, en nous limitant à la classe $\text{VAR}(p)$, *i.e.* $\text{VARMA}(p,0)$, et le critère par défaut AICC, et d'autre part examiner les matrices des corrélations partielles. Soit donc les commandes :

```
proc varmax data=simul;
model y1 y2 y3 / minic=(p=5 q=0) print=(parcoef);
run;
```

L'option `lagmax=` n'étant pas utilisée, ce paramètre est mis à sa valeur par défaut $\min(12, T - 2)$ avec T = nombre d'observations non manquantes

Minimum Information Criterion Based on AICC	
Lag	MA 0
AR 0	6.533396
AR 1	-0.432164
AR 2	-0.251984
AR 3	-0.071189
AR 4	0.1513793
AR 5	0.3166587

TABLE 1 – Critères de sélection des modèles concurrents générés par `minic=(p=5 q=0)`

Schematic Representation of Partial Autoregression												
Variable/Lag	1	2	3	4	5	6	7	8	9	10	11	12
y1	+..	+..
y2	+..
y3	+.+
+ is > 2*std error, - is < -2*std error, . is between												

TABLE 2 – tests de significativité des autocorrélations partielles associés à `print=(parcoef)`

et donc ici `lagmax=12`. Selon les résultats présentés dans la table 1, le processus optimal selon AICC parmi les six modèles concurrents serait donc le VAR(1). Dans la table 2 on peut remarquer que les matrices (3×3) des corrélations partielles sont constituées d'éléments pratiquement toujours non significativement différents de zéro au-delà de l'ordre un, ce qui conforte le choix de $p = 1$. Notez que si dans cet exemple les outils de sélection de l'ordre du VAR permettent de retrouver l'ordre vrai du processus simulé, ce n'est évidemment pas toujours le cas.

2. Estimation du VAR(1)

L'étape suivante consiste donc à estimer les coefficients du VAR(1) ainsi que la matrice de var-cov des résidus. Comme dans la commande `model...` utilisée précédemment aucune valeur de p n'est explicitement fixée, `Varmx` va par défaut estimer le VAR optimal au sens du critère utilisé dans `minic=(...)`, soit donc ici un VAR(1). En d'autres termes, cette commande `model...` donne de ce point de vue les mêmes résultats que `model y1 y2 y3 / p=1;`

Ici, nous avons simplement ajouté l'option `roots` qui réclame l'affichage des valeurs propres de la matrice companion du VAR estimé. Remarquez que le présent VAR étant d'ordre 1, le VAR companion n'est rien

Model Parameter Estimates						
Equation	Parameter	Estimate	Standard Error	t Value	Pr > t	Variable
y1	CONST1	0.05963	0.11923	0.50	0.6181	1
	AR1_1_1	1.34032	0.06999	19.15	0.0001	y1(t-1)
	AR1_1_2	-0.57307	0.10114	-5.67	0.0001	y2(t-1)
y2	AR1_1_3	0.00342	0.02639	0.13	0.8971	y3(t-1)
	CONST2	0.08955	0.11810	0.76	0.4502	1
	AR1_2_1	0.66814	0.06933	9.64	0.0001	y1(t-1)
y3	AR1_2_2	0.28381	0.10018	2.83	0.0056	y2(t-1)
	AR1_2_3	0.00218	0.02614	0.08	0.9336	y3(t-1)
	CONST3	0.16268	0.08013	2.03	0.0451	1
	AR1_3_1	0.63929	0.04704	13.59	0.0001	y1(t-1)
	AR1_3_2	0.40463	0.06797	5.95	0.0001	y2(t-1)
	AR1_3_3	0.69324	0.01773	39.09	0.0001	y3(t-1)

TABLE 3 – Estimation de Φ avec la commande `model y1 y2 y3 / p=1 ;`

Covariances of Innovations			
Variable	y1	y2	y3
y1	1.38586	0.69798	0.50899
y2	0.69798	1.35990	0.43236
y3	0.50899	0.43236	0.62598

TABLE 4 – Estimation de Σ_u avec la commande `model y1 y2 y3 / p=1 ;`

d'autre que lui-même et la matrice companion est évidemment Φ : nous récupérerons donc 3 valeurs propres.

`model y1 y2 y3 / p=1 print=(roots) ;`

Les estimations obtenues pour Φ sont présentées dans la table 3. Selon les statistiques de student, tous les coefficients sont significatifs aux seuils de risque usuels à l'exception des constantes dans les trois équations et du coefficient de y_{3t-1} dans les équations de y_{1t} et y_{2t} , tout ceci étant parfaitement raisonnable compte-tenu des valeurs des paramètres du modèle simulé. L'estimation de Σ_u est présentée dans la table 4. Enfin, les valeurs propres estimées de la matrice companion apparaissent dans la table 5 : elles sont toutes de module inférieur à l'unité ce qui, sans être un test formel, plaide en faveur de la stabilité du VAR estimé.

3. Tests de validation du VAR estimé

Nous nous intéressons maintenant aux tests de validation offerts par la proc `varmax`. Ceux-ci portent essentiellement sur les propriétés des résidus empiriques \hat{u}_{1t} , \hat{u}_{2t} et \hat{u}_{3t} qui, si le processus estimé est satisfaisant,

Roots of AR Characteristic Polynomial					
Index	Real	Imaginary	Modulus	Radian	Degree
1	0.81010	0.31672	0.8698	0.3727	21.3537
2	0.81010	-0.31672	0.8698	-0.3727	-21.3537
3	0.69718	0.00000	0.6972	0.0000	0.0000

TABLE 5 – Valeurs propres estimées de la matrice companion, option `print=(roots)`

devraient être des estimateurs des processus en bruit blanc u_{1t} , u_{2t} et u_{3t} . Ces tests sont obtenus avec l'option `diagnose` dans la commande `model`. Dans ce qui suit, nous avons également supprimé les constantes non significatives dans les trois équations. Par ailleurs les graphiques des autocorrélations, des autocorrélations partielles et inverses et de la statistique de Ljung-Box, tous bien connus depuis la proc `ARIMA`, ont aussi été réclamés via l'option `plots=(residual)` dans l'appel de la procédure. Les commandes sont donc :

```
proc varmax data=simul plots=(residual);
model y1 y2 y3 / noint p=1 print=(diagnose);
run;
```

En sortie, nous récupérons les estimations des matrices de covariances et de corrélations entre \hat{u}_t et \hat{u}_{t-j} , $j = 0 \dots, 12$ conformément à la valeur de `lagmax`. Des tests de nullité des coefficients de corrélation sont réalisés⁹⁷. Les résultats sont donnés dans la table 6 selon une représentation graphique similaire à celle déjà vue pour les matrices des autocorrélations partielles. Dans le présent exercice, à l'exception de la corrélation positive entre \hat{u}_{1t} et \hat{u}_{1t-6} et de celle, négative entre \hat{u}_{2t} et \hat{u}_{1t-8} , on ne peut rejeter leur nullité pour tout retard non nul. En d'autres termes, les seules corrélations significatives intéressantes entre les résidus correspondent à des dépendances instantanées. On trouve également dans les sorties la statistique portemanteau de Hogsking de test de nullité des corrélations jusqu'à l'ordre j , $j = 1, \dots, \text{lagmax}$. Comme on peut le voir dans la table 7, cette statistique ne permet pas de rejeter la nullité des corrélations quel que soit l'ordre j retenu. Pour mémoire, on a ajusté le vrai processus générateur des données, les résidus sont indépendants, homoscédastiques on est donc dans les conditions idéales d'application du test. Le fait qu'il donne une conclusion raisonnable ici n'enlève rien à nos remarques le concernant.

On trouve ensuite un ensemble de statistiques relatives à chacune des équations du VAR estimé. Tout d'abord des mesures usuelles dans le cadre OLS de la qualité d'ajustement : R^2 , écart-type résiduel estimé, test de Fisher de nullité des coefficients de la régression (voir la table 8). Dans

⁹⁷. au moyen d'un écart-type estimé égal à $T^{-1/2}$ et donc un intervalle de confiance autour de zéro égal à $\pm 2/\sqrt{T}$.

Schematic Representation of Cross Correlations of Residuals													
Variable/Lag	0	1	2	3	4	5	6	7	8	9	10	11	12
y1	+++	+..
y2	+++	-..
y3	+++
+ is > 2*std error, - is < -2*std error, . is between													

TABLE 6 – tests de significativité des autocorrélations associés à print=(diagnose)

Portmanteau Test for Cross Correlations of Residuals			
Up To Lag	DF	Chi-Square	Pr > ChiSq
2	9	7.49	0.5858
3	18	10.16	0.9267
4	27	18.17	0.8984
5	36	25.81	0.8955
6	45	38.14	0.7555
7	54	54.73	0.4467
8	63	61.85	0.5175
9	72	72.48	0.4619
10	81	78.53	0.5571
11	90	85.85	0.6043
12	99	94.40	0.6121

TABLE 7 – Statistique de Hogsking associée à l'option print=(diagnose)

Univariate Model ANOVA Diagnostics				
Variable	R-Square	Standard Deviation	F Value	Pr > F
y1	0.8745	1.17262	334.45	<.0001
y2	0.8675	1.16356	314.33	<.0001
y3	0.9898	0.80395	4679.92	<.0001

TABLE 8 – Mesures de qualité de l’ajustement de chaque équation du VAR estimé, associées à l’option `print=(diagnose)`

cet exemple, on rejette la nullité simultanée des coefficients des explicatives $y_{1,t-1}$, $y_{1,t-1}$ et $y_{3,t-1}$ dans chacune des trois régressions. Puis des informations tirées des résidus empiriques de chaque régression : test de Durbin-Watson, de Jarque-Bera et test d’effet ARCH (voir table 9) et des tests de Fisher de nullité des coefficients de processus $AR(j)$, $j = 1 \dots, 4$ ajustés sur chacune des séries résiduelles (table 10). Ici, quelle que soit l’équation considérée, on conclut que le résidu courant ne dépend pas de ses valeurs passées, *i.e.* à l’absence d’autocorrélation significative jusqu’à l’ordre quatre.

Pour finir, un ensemble de sorties graphiques afférentes aux séries de résidus sont générées⁹⁸ : graphes de ces séries elles-mêmes (Figure 2), utiles par exemple pour repérer d’éventuels outliers pouvant conduire à introduire des indicatrices d’événements dans les équations du VAR, puis les trois graphes des fonctions d’aide à la sélection des filtres univariés complétés par celui indiquant la valeur de la statistique de Ljung-Box pour un ensemble de retards allant jusqu’à lagmax (Figure 3). Ces éléments vous sont évidemment connus depuis que vous savez utiliser la proc ARIMA. L’idée est ici de vérifier que chaque série résiduelle se comporte comme un bruit blanc. Enfin est également produit un graphe utile pour juger de la normalité des séries de résidus (histogramme avec superposition de la densité d’une gaussienne centrée et ayant un écart-type égal à celui estimé sur la série concernée) associé à un Q-Q plot mettant en regard les fractiles théoriques de cette gaussienne avec ceux observés dans la série résiduelle : en cas d’adéquation parfaite, le nuage de points est aligné le long de la première bissectrice (Figure 4).

4. Les fonctions de réponses

98. Dans ce document ne seront reproduits que les graphiques relatifs à la série \hat{u}_{2t} .

Univariate Model White Noise Diagnostics					
Variable	Durbin	Normality		ARCH	
	Watson	Chi-Square	Pr > ChiSq	F Value	Pr > F
y1	1.99158	2.37	0.3056	3.43	0.0671
y2	2.02490	1.68	0.4310	0.48	0.4914
y3	1.91428	3.97	0.1377	0.88	0.3509

TABLE 9 – Statistiques DW et Jarque-Bera, test d'effet ARCH sur les résidus, sortie associée à l'option print=(diagnose)

Univariate Model AR Diagnostics								
Variable	AR1		AR2		AR3		AR4	
	F Value	Pr > F	F Value	Pr > F	F Value	Pr > F	F Value	Pr > F
y1	0.01	0.9108	0.18	0.8382	0.32	0.8128	1.22	0.3080
y2	0.18	0.6752	0.11	0.8931	0.16	0.9244	0.19	0.9426
y3	0.00	0.9617	0.05	0.9489	0.31	0.8202	0.26	0.9057

TABLE 10 – Tests de nullité des coefficients autorégressifs sur les séries résiduelles, sortie associée à l'option print=(diagnose)

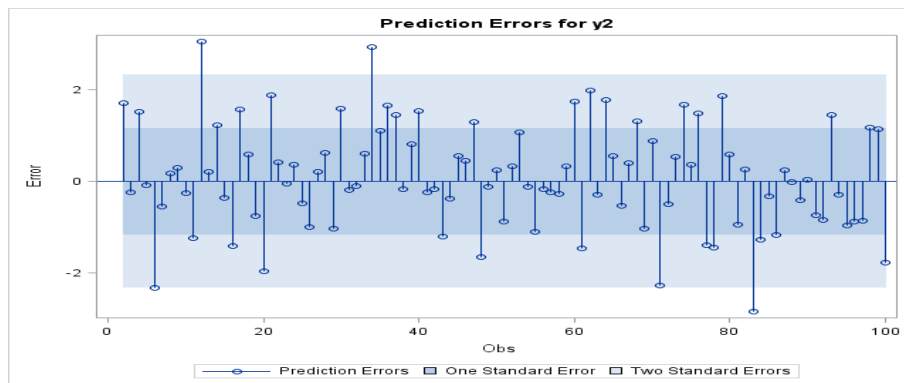


FIGURE 2 – Graphe de \hat{u}_{2t} généré par l'option plots=(residual)

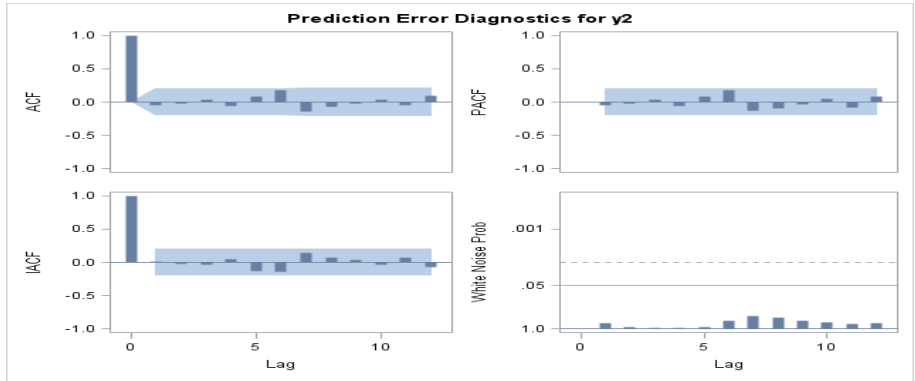


FIGURE 3 – Série \hat{u}_{2t} , graphe des autocorrélations, des autocorrélations partielles et inverses, de la statistique de Ljung-Box généré par l'option plots=(residual)

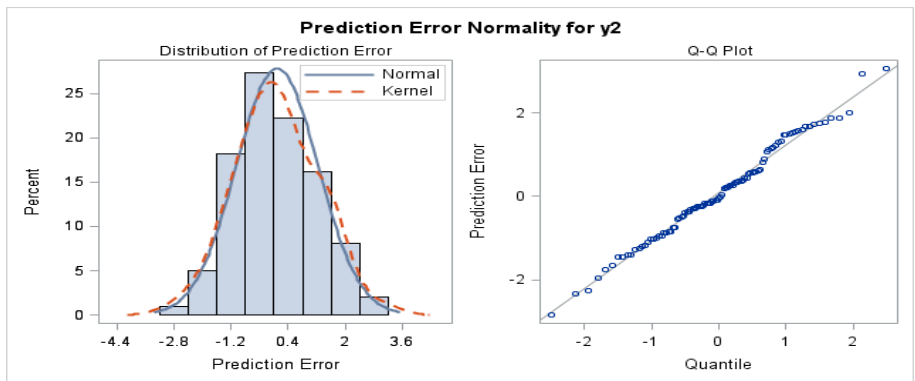


FIGURE 4 – Indications sur la normalité de \hat{u}_{2t} généré par l'option plots=(residual)

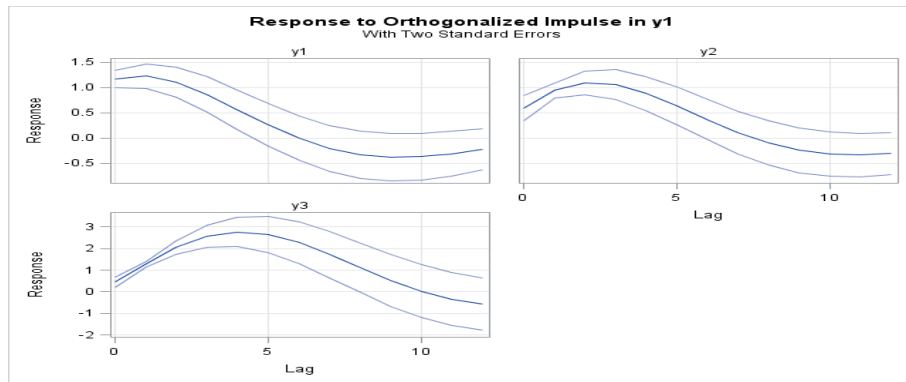


FIGURE 5 – Réponses de y_1 , y_2 et y_3 à un choc affectant v_1 généré par l'option `plots=(impulse(orth))`

On se limitera ici aux seules fonctions de réponse orthogonalisées et à leur représentation graphique⁹⁹ Avec

```
proc varmax data=simul plots=(impulse(orth));
model y1 y2 y3 / noint p=1;
run;
```

On obtient les graphes 5, 6 et 7 qui traduisent les trajectoires estimées des trois variables suite à trois chocs affectant successivement chacune des composantes des résidus orthogonalisés v_1 , v_2 et v_3 . Par défaut, un intervalle de confiance ponctuel à 95% est également représenté autour de ces fonctions de réponse. On voit par exemple qu'un choc positif sur v_2 fait passer y_1 sous sa position d'équilibre, la déviation maximale étant obtenue environ quatre mois après l'événement déclencheur suivie d'un retour à l'équilibre après 8 mois. Par ailleurs, ce choc positif élève naturellement instantanément la variable y_2 au-dessus de sa position d'équilibre, puis la fait rapidement basculer significativement sous sa valeur d'équilibre, l'impact négatif maximal étant atteint 5 mois après le choc, le retour à l'équilibre prenant environ une dizaine de mois.

Vous noterez aussi qu'une modification de la composante v_3 n'affecte pas les variables y_1 et y_2 , ce que vous devriez être en mesure de comprendre en examinant les équations (122), (123) et (124) du système simulé pour construire cet exemple.

99. Pour mémoire, l'affichage de leurs valeurs, des écart-types associés,... sont obtenus via l'option `print=(impulse=(orth stderr))`.

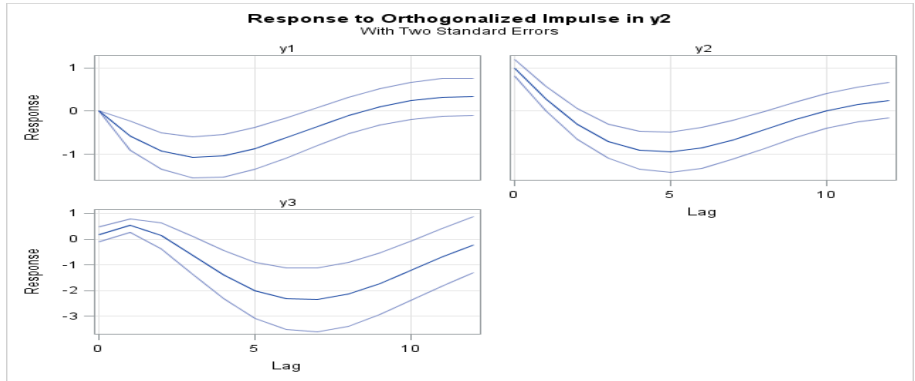


FIGURE 6 – Réponses de y_1 , y_2 et y_3 à un choc affectant v_2 généré par l'option `plots=(impulse(orth))`

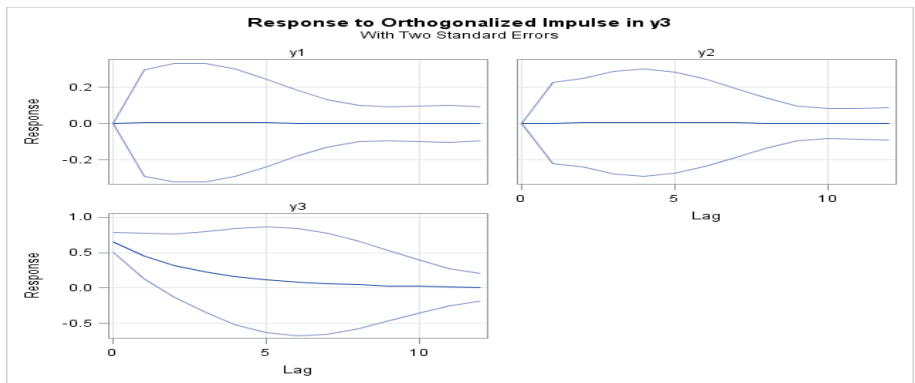


FIGURE 7 – Réponses de y_1 , y_2 et y_3 à un choc affectant v_3 généré par l'option `plots=(impulse(orth))`

5. la décomposition de la variance des erreurs de prévisions

Nous avons demandé le calcul de la contribution de chacun des chocs orthogonaux à la variance de erreurs de prévision jusqu'à un horizon de 10 mois via :

```
proc varmax data=simul;  
model y1 y2 y3 / noint p=1 print=decompose(10));  
run;
```

La procédure affiche les valeurs des variances et celles des contributions dans un premier tableau (Table 11), puis les pourcentages correspondants dans un second (Table 12). La variable sur laquelle porte la prévision est en en-tête de ligne, les contributions des chocs orthogonalisés, v_1 , v_2 et v_3 sont en colonne et repérés par le nom des variables à laquelle chacun revoie soit ici, $v_1 \rightarrow y_1$, $v_2 \rightarrow y_2$ et $v_3 \rightarrow y_3$.

Nous indiquons seulement les résultats afférents aux trois premiers horizons et ceux des deux derniers. Pour trouver la variance des erreurs elles-mêmes, il suffit d'additionner les contributions. Par exemple, à l'horizon 2, la variance des erreurs commises sur y_2 est de 2.33792 (=1.25439+1.08353). A l'horizon 10 elle passe à 9.76 (=4.96775 + 4.79221 + 0.00013). Sans surprise, on remarque que plus l'horizon de prévision s'élève et plus la variance des erreurs augmente, *i.e.* la précision des prévisions décroît.

En ce qui concerne y_1 , en regardant les pourcentages des contributions, on doit noter sans surprise qu'à l'horizon 1, y_1 (*i.e.* v_1) contribue à 100% à la variance de l'erreur commise sur elle-même et que la part de la variance des innovations afférentes à y_2 , augmente régulièrement pour expliquer jusqu'à la 45% de la variance des erreurs à dix mois. Si à un mois, la variance des erreurs sur y_3 est à 65% expliquée par celle de v_3 , cette part devient rapidement négligeable, la variance totale étant expliquée à parts égales par v_1 et v_2 , soulignant ainsi l'importance des variables y_1 et y_2 pour comprendre les évolutions de y_3 .

Comme on le sait, les fonctions de réponses aux chocs orthogonalisés et la décomposition de la variance des erreurs de prévision dépendent de l'ordre d'entrée des variables. Afin d'apprécier cet impact dans le présent exemple, nous avons estimé à nouveau la décomposition de la variance dans l'ordre d'entrée y_3 , y_2 , y_1 via :

```
proc varmax data=simul;  
model y3 y2 y1 / noint p=1 print=decompose(10));  
run;
```

Dans cet ordre on autorise y_3 à avoir un impact maximal sur les deux autres variables, et on minimise l'influence de y_1 . On est donc, pour ces deux variables, à l'opposé des contraintes de l'ordre précédent. Pour autant, on peut voir dans la Table 13 que l'influence de v_3 décroît rapidement avec l'horizon pour finir, à l'horizon 10 à expliquer moins de

Decomposition of Prediction Error Covariances				
Lead	Variable	y1	y2	y3
1	y1	1.37503	0.00000	0.00000
	y2	0.35243	1.00145	0.00000
	y3	0.19180	0.03332	0.42122
2	y1	2.89002	0.32658	0.00001
	y2	1.25439	1.08353	0.00000
	y3	1.84981	0.31982	0.62485
3	y1	4.11920	1.18404	0.00003
	y2	2.44887	1.17106	0.00002
	y3	6.09550	0.33620	0.72493
⋮	⋮			
9	y1	5.40835	4.66703	0.00009
	y2	4.90953	4.75218	0.00013
	y3	36.78353	22.07360	0.82820
10	y1	5.55064	4.67693	0.00009
	y2	4.96775	4.79221	0.00013
	y3	37.06370	25.04986	0.82927

TABLE 11 – Décomposition de la variance des erreurs de prévisions, associées à l’option print=(decompose) et à l’ordre d’entrée y1, y2, y3.

20% de la variance des erreurs quelle que soit la variable considérée. En revanche, la part attribuable à v_1 , toujours supérieure à 60% pour y_1 augmente fortement également sur les deux autres variables pour dépasser les 50% à l’horizon de 10 mois au détriment de v_2 et surtout de v_3 . Ainsi, cet exercice souligne le rôle central de la variable y_1 dans l’évolution du vecteur y_t .

6. La causalité selon Granger

Il semble, compte-tenu des résultats précédents, que les évolutions de y_3 ne contiennent que peu d’information pour comprendre celles des deux autres variables y_1 et y_2 , alors qu’à l’inverse, ces dernières seraient une source d’informations utiles pour prévoir y_3 . Le test des propositions

$$H0 : \{y_3\} \nrightarrow \{y_1, y_2\} \text{ versus } H1 : \{y_3\} \rightarrow \{y_1, y_2\} \text{ et,} \quad (127)$$

$$H0 : \{y_1, y_2\} \nrightarrow \{y_3\} \text{ versus } H1 : \{y_1, y_2\} \rightarrow \{y_3\} \quad (128)$$

peuvent être réalisés soit par la commande Test, soit par Causal Group1=() group2=(). Ainsi les deux ensembles de commandes suivants sont équivalents :

```
proc varmax data=simul;
model y1 y2 y3 / noint p=1;
```

Proportions of Prediction Error Covariances				
Lead	Variable	y1	y2	y3
1	y1	1.00000	0.00000	0.00000
	y2	0.26031	0.73969	0.00000
	y3	0.29675	0.05155	0.65171
2	y1	0.89847	0.10153	0.00000
	y2	0.53654	0.46346	0.00000
	y3	0.66195	0.11445	0.22360
3	y1	0.77673	0.22327	0.00003
	y2	0.77673	0.22327	0.00000
	y3	0.85173	0.04698	0.10129
⋮	⋮			
9	y1	0.53678	0.46321	0.00001
	y2	0.50814	0.49185	0.00001
	y3	0.61629	0.36983	0.01388
10	y1	0.54271	0.45728	0.00001
	y2	0.50899	0.49100	0.00001
	y3	0.58885	0.39798	0.01317

TABLE 12 – Proportions de la variance des erreurs de prévisions attribuable à chaque composante associée à l’option print=(decompose) et à l’ordre d’entrée y1, y2, y3.

Proportions of Prediction Error Covariances				
Lead	Variable	y3	y2	y1
1	y3	1.00000	0.00000	0.00000
	y2	0.22392	0.77608	0.00000
	y1	0.29675	0.08211	0.62115
2	y3	0.73570	0.14229	0.12201
	y2	0.27562	0.56330	0.16109
	y1	0.21850	0.04091	0.74060
3	y3	0.56384	0.09979	0.33637
	y2	0.25600	0.36475	0.37925
	y1	0.16227	0.06740	0.77033
⋮	⋮			
9	y3	0.16574	0.14244	0.69182
	y2	0.12734	0.30279	0.56988
	y1	0.10125	0.21992	0.67883
10	y3	0.15725	0.16344	0.67931
	y2	0.12924	0.30566	0.56510
	y1	0.10303	0.21669	0.68027

TABLE 13 – Proportion de la variance des erreurs de prévisions attribuable à chaque composante, associé à l’ordre y3, y2, y1.

Granger-Causality Wald Test			
Test	DF	Chi-Square	Pr > ChiSq
1	2	0.03	0.9860
2	2	904.82	<.0001
Test 1 :	Group 1 Variables :	y1 y2	
	Group 2 Variables :	y3	
Test 2 :	Group 1 Variables :	y3	
	Group 2 Variables :	y1 y2	

TABLE 14 – Test de causalité selon Granger , associées aux options causal group1=(y1 y2) group2=(y3);causal group1=(y3) group2=(y1 y2);

```
test ar(1,1,3)=0, ar(1,2,3)=0;
test ar(1,3,1)=0, ar(1,3,2)=0;
run;
```

et :

```
proc varmax data=simul;
model y1 y2 y3 / noint p=1;
causal group1=(y1 y2) group2=(y3);
causal group1=(y3) group2=(y1 y2);
run;
```

La première commande test porte sur la nullité des coefficients de $y_{3,t-1}$ dans les équations de y_1 et y_2 , la seconde réclame un test de Wald de nullité des coefficients de $y_{1,t-1}$ et $y_{2,t-1}$ dans l'équation de y_3 . Pour mémoire, avec causal, on réclame le test de nullité des coefficients des variables référencées dans group2= dans les équations des variables précisées dans group1=, ce dernier test étant réalisé dans le VAR constitué par l'union des variables référencées dans les deux options group1 et group2. Dans cet exemple, cette union est l'ensemble des variables du système initial $(y_{-1}, y_{-2}, y_{-3})^\top$, et donc le test de WALD réalisé via causal est mené dans le même VAR que celui dans lequel est appelée la commande test. Les résultats seront donc identiques. La structure de la table 14 est celle associée à la commande causal. A sa lecture, on constate bien que y_1 et y_2 sont, à l'horizon un, des prédicteurs avancés de y_3 alors que y_3 ne cause pas selon Granger le couple (y_1, y_2) .

3 Une introduction aux VAR structurels

Une critique faite à l'encontre de la modélisation VAR présentée auparavant est qu'elle est complètement a-théorique : aucune analyse économique ne vient étayer la construction du modèle à l'exception de la liste des variables constituant le système : on laisse les données parler. La difficulté est naturellement de comprendre ce qu'elles disent, et pour cela il est nécessaire de disposer d'un

cadre d'analyse qui traduit en propositions compréhensibles d'un point de vue théorique les messages envoyés. Les VAR structurels ont été conçus comme une réponse à cette critique. Dans un VAR structurel, la théorie économique peut spécifier les liens de dépendance, ou au contraire d'indépendance, instantanés entre telle et telle variables. Elle peut aussi préciser les relations de causalité qui les relient.

Si on omet la possible présence de variables exogènes et de termes déterministes, l'écriture générale d'un VAR structurel est la suivante :

$$Ay_t = B(L)y_t + \epsilon_t \quad (129)$$

où

- les ϵ_t sont qualifiés de chocs structurels et constituent un bruit blanc vectoriel de matrice de variance-covariance Σ_ϵ diagonale
- la matrice A possède généralement des 1 sur sa diagonale afin d'identifier la variable expliquée de chaque équation, et contient les opposés des coefficients des relations instantanées entre les y .

3.1 un exemple

Considérons les équations d'un VAR structurel, ou SVAR, sur deux variables y_{1t} et y_{2t} :

$$y_{1t} = a_1 y_{2t} + b_{11} y_{1t-1} + b_{12} y_{2t-1} + \epsilon_{1t} \quad (130)$$

$$y_{2t} = a_2 y_{1t} + b_{21} y_{1t-1} + b_{22} y_{2t-1} + \epsilon_{2t} \quad (131)$$

On notera la présence de l'explicative y_{2t} (resp. y_{1t}) dans l'équation de y_{1t} (resp. y_{2t}). Cette présence des autres y contemporaines de l'expliquée dans chaque équation explique que les dépendances instantanées entre ces variables n'aient plus besoin d'être saisie par des covariances non nulles des ϵ_t entre eux. On peut ici imposer l'orthogonalité des chocs structurels. Dans ces écritures, ϵ_{jt} représente bien le choc afférent à la $j^{\text{ième}}$ variable du système. Sous forme matricielle, les deux équations précédentes possèdent bien l'écriture (129) avec

$$A = \begin{pmatrix} 1 & -a_1 \\ -a_2 & 1 \end{pmatrix}, \quad \Phi(L) = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \quad \text{et} \quad \Sigma_\epsilon = \begin{pmatrix} \sigma_{\epsilon_1}^2 & 0 \\ 0 & \sigma_{\epsilon_2}^2 \end{pmatrix}$$

3.2 Les difficultés d'estimation des VAR structurels et le recours aux VAR

Si on désire confronter la théorie qui fonde le VAR structurel aux données observées, où plus simplement utiliser cette théorie pour raconter une histoire sur l'évolution du système économique qu'elle est censée décrire, il est nécessaire d'estimer les objets A , $\Phi(L)$ et Σ_ϵ . Cependant, à la lecture des équations (130) et (131) on peut deviner que cette estimation est compliquée. En particulier on remarque aisément que les OLS sont inadaptés en raison de la non

orthogonalité entre le résidu de l'équation et certaines explicatives¹⁰⁰. Une solution est de recourir à la forme réduite du VAR structurel, *i.e.* d'écrire y_t sur des variables orthogonales aux innovations. Lorsque A est inversible, cela se réalise aisément puisqu'alors :

$$Ay_t = B(L)y_t + \epsilon_t \Rightarrow y_t = A^{-1}B(L)y_t + A^{-1}\epsilon_t \quad (132)$$

soit encore

$$y_t = \Phi(L)y_t + u_t \quad (133)$$

et on reconnaît là, l'équation usuelle d'un VAR¹⁰¹, avec $\Phi(L) = A^{-1}B(L)$, $u_t = A^{-1}\epsilon_t$, et $\Sigma_u = A^{-1}\Sigma_\epsilon A^{-1\top}$.

Dans ces conditions, sachant que l'on sait trouver aisément les estimateurs du maximum de vraisemblance conditionnels de Φ et de Σ_u on peut chercher, en partant de ces derniers, à remonter à ceux des paramètres du VAR structurel.

3.3 du VAR au VAR structurel : les conditions d'identification

Malheureusement, cette remontée vers les seuls paramètres ayant une signification théorique n'est pas possible sans la prise en compte de contraintes supplémentaires. Il est en effet impossible d'identifier les coefficients de la forme structurelle à partir des coefficients de la forme réduite, *i.e.* du VAR. Il suffit pour s'en apercevoir de compter simplement les coefficients des deux structures. Supposons ainsi que le vecteur y soit de dimensions $k \times 1$ et que $B(L)$ soit composé de polynômes en L de degré p .

— dans le VAR structurel :

- $A = \{a_{i,j}\}_{(k,k)}$ avec $a_{i,i} = 1, i = 1, \dots, k$ soit $k(k-1)$ inconnues,
- $B(L) = (B_1L + B_2L^2 + \dots + B_pL^p)$ avec $B_i, i = 1, \dots, p$, soit pk^2 inconnues
- $\Sigma_\epsilon = \{\sigma_{i,j}\}_{(k,k)} = \begin{cases} \sigma_{\epsilon_i}^2 & \text{si } i = j \\ 0 & \text{sinon} \end{cases}$ soit k inconnues

Ce qui fait un total de $(p+1)k^2$ inconnues

— dans le VAR :

- $p * k^2$ coefficients sont estimés avec $\hat{\Phi}(L)$,
- $\frac{k^2+k}{2}$ coefficients sont estimés via la matrice symétrique $\hat{\Sigma}_u$

Ce qui donne un total de $\frac{(2p+1)k^2+k}{2}$ coefficients estimés.

Au final, il y a un déficit de coefficients estimés : le passage du jeu de coefficients estimés dans la forme réduite au jeu de coefficients de la forme structurelle nécessite que l'on impose $k(k-1)/2$ contraintes d'identification. C'est précisément

100. En effet, selon (130), ϵ_{1t} est dans y_{1t} . Si a_2 n'est pas nul, il est alors également présent d'après (131) dans y_{2t} et donc $\epsilon_{1t} \not\perp y_{2t}$. Un raisonnement symétrique montre que généralement, $\epsilon_{2t} \not\perp y_{1t}$.

101. Dans cette littérature, les ϵ_t sont qualifiés de chocs structurels et les u_t de chocs canoniques.

à ce niveau que les arguments théoriques peuvent intervenir. Lorsque l'on regarde l'équation (132), on voit que ces contraintes peuvent porter sur deux objets :

- sur la matrice A constituée des coefficients des variables contemporaines à l'expliquée. Le plus souvent certains coefficients sont mis à zéro, ce qui revient à exclure les variables concernées de la liste des explicatives d'une ou de plusieurs équations.

Pour illustration, dans l'exemple précédent constitué des équations (130) et (131), il y a 8 inconnues dans le VAR structurel et 7 coefficients estimés dans le VAR¹⁰². Il faut donc imposer au moins une contrainte d'identification. Une théorie économique pourrait alors justifier que $a_1 = 0$, *i.e.* que la valeur courante de y_1 ne dépend pas du y_2 contemporain. Dans ce cadre, il devient alors possible de retrouver les valeurs estimées de l'ensemble des coefficients du VAR structurel à partir de celui des coefficients estimés du VAR.

- sur la matrice A^{-1} . Sachant que $u_t = A^{-1}\epsilon_t$, on peut par exemple, pour des raisons théoriques supposer que l'un des chocs structurel affecte une variable mais pas l'autre.

Ainsi, toujours dans l'exemple bivarié, en posant $a^{22} = 0$, on force les deux chocs structurels ϵ_1 et ϵ_2 à définir le choc canonique u_1 de la première équation alors que celui de la seconde, u_2 , ne dépendrait pas de ϵ_2 et donc y_{2t} ne dépendrait pas de ϵ_{2t} ^{103, 104}.

En théorie, on rien n'interdit d'imposer ces contraintes sur la dynamique du système, *i.e.* sur les coefficients $b_{11}, b_{12}, b_{21}, b_{22}$, mais les justifications théoriques paraissent beaucoup plus ténues que celles que l'on peut avancer pour contraindre A ou A^{-1} .

Retenez encore, qu'à une même forme réduite, *i.e.* à un VAR donné, peuvent correspondre différents VAR structurels selon les contraintes d'identification que l'on impose. Ceci souligne bien qu'une forme réduite ne renseigne en rien sur la structure du système économique¹⁰⁵.

102. soit respectivement $\{a_1, a_2, b_{11}, b_{12}, b_{21}, b_{22}, \sigma_{\epsilon_1}^2, \sigma_{\epsilon_2}^2\}$ et $\{\phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}, \sigma_{u_1}^2, \sigma_{u_2}^2, \sigma_{u_1, u_2}\}$.

103. Remarquez que dans cet exemple bivarié, la distinction entre imposer une contrainte sur A ou sur A^{-1} est sans objet dans la mesure où l'inverse d'une matrice triangulaire est aussi une matrice triangulaire et donc contraindre par exemple $a_2 = 0$ équivaut à imposer $a^{21} = 0$ et réciproquement.

104. Cet exemple permet de donner l'intuition de l'apport important d'un travail de Blanchard et Quah (Blanchard, O. J. & D. QUAH, The dynamic effects of aggregate demand and supply disturbances, The American Economic Review, 1989, 655-73). Il s'agit pour ces auteurs d'identifier des chocs d'offre et des chocs de demande et d'étudier la réponse d'un système économique constitué d'une variable d'activité (le logarithme du PIB) supposé intégrer un trend stochastique, et du taux de chômage supposé stationnaire. Leur contrainte d'identification porte sur la dynamique de long terme : ils imposent que les chocs de demande n'ont aucun effet à long terme sur le PIB contrairement aux chocs d'offre. Notez bien que si l'idée est proche de celle du texte ci-dessus, les aspects techniques sont bien différents dans la mesure où pour l'instant nos variables sont stationnaires et qu'il ne peut donc y avoir ici de l'hystérèse dans les effets des innovations.

105. Ce qui ne l'empêche d'ailleurs pas d'être éventuellement toujours utile pour construire des prévisions.

3.3.1 Une relecture de la décomposition de cholesky

Nous savons que lorsque la matrice de variance-covariance des innovations canoniques n'est pas diagonale, les fonctions de réponse associées à l'une de ces innovations ne peuvent être recouvertes par les dérivées partielles de l'écriture VMA puisque qu'une modification de l'une d'entre elles génère alors des réponses instantanées sur une ou plusieurs autres composantes du vecteur u . Nous savons aussi que la solution, préconisée par Sims, met en oeuvre la décomposition de Cholesky qui consiste à substituer aux u_t des innovations orthogonales définies par $v_t = P^{-1}u_t$, où P est une matrice inversible triangulaire inférieure, à éléments positifs sur la diagonales et telle que $PP^T = \Sigma_u$.

Ainsi, le VAR

$$y_t = \Phi(L)y_t + u_t, \quad (134)$$

à partir duquel nous trouvons les fonctions de réponse orthogonalisées aux innovations v_t définies ci-dessus, peut encore s'écrire de manière équivalente comme :

$$P^{-1}y_t = P^{-1}\Phi(L)y_t + P^{-1}u_t$$

Soit encore

$$P^{-1}y_t = \tilde{\Phi}(L)y_t + v_t$$

avec $\Sigma_v = I$.

On remarque immédiatement que cette dernière équation correspond à un VAR structural dans lequel la matrice des coefficients des explicatives contemporaines à l'expliquée est P^{-1} , matrice triangulaire inférieure. Ainsi, la $i^{\text{ème}}$ équation de ce système spécifie que l'expliquée est déterminée par les valeurs retardées de toutes les variables, par les valeurs contemporaines des variables dont le rang d'entrée dans le VAR est supérieur à i et par un résidu orthogonal à ceux des autres équations. Pour illustration, avec $k = 3$ et $p = 1$, nous avons :

$$\begin{bmatrix} p^{11} & 0 & 0 \\ p^{21} & p^{22} & 0 \\ p^{31} & p^{31} & p^{31} \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} \tilde{\phi}_{11} & \tilde{\phi}_{1,2} & \tilde{\phi}_{1,3} \\ \tilde{\phi}_{21} & \tilde{\phi}_{2,2} & \tilde{\phi}_{2,3} \\ \tilde{\phi}_{31} & \tilde{\phi}_{3,2} & \tilde{\phi}_{3,3} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-1} \end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \end{bmatrix} \quad (135)$$

Les résidus orthogonalisés du VAR initial donné en (134) sont les innovations structurelles de (135) et les réponses aux chocs orthogonalisés par Cholesky de (134) sont donc aussi les réponses simples aux chocs structurels de (135).

On observe par ailleurs bien que la première équation détermine la première variable du système, ici y_{1t} , indépendamment des valeurs contemporaines des deux autres variables, que dans la seconde équation, y_{2t} est expliquée par y_{1t} mais non par y_{3t} , et qu'enfin y_{3t} est expliquée par y_{1t} et y_{2t} *i.e.* par les deux variables qui la précède. Ce type de système est qualifié de récursif.

Ainsi, la décomposition de Cholesky impose implicitement un VAR structural récursif calé uniquement sur l'ordre d'introduction des variables. Il faut donc se méfier de son utilisation automatique afin d'éviter de considérer des modélisations qui pourraient n'avoir aucun sens.