



Econométrie des Séries Temporelles Multivariées

Gilles de Truchis

Master 1 ESA

January 19, 2025

Plan du cours

VMA

VAR

Prévision

Causalité de Granger

IRF et décomposition de variance

Estimation et inférence

Tests et diagnostics

VAR structurels

Théorie de la cointégration

Engle et Granger

Endogénéité

Analyse canonique

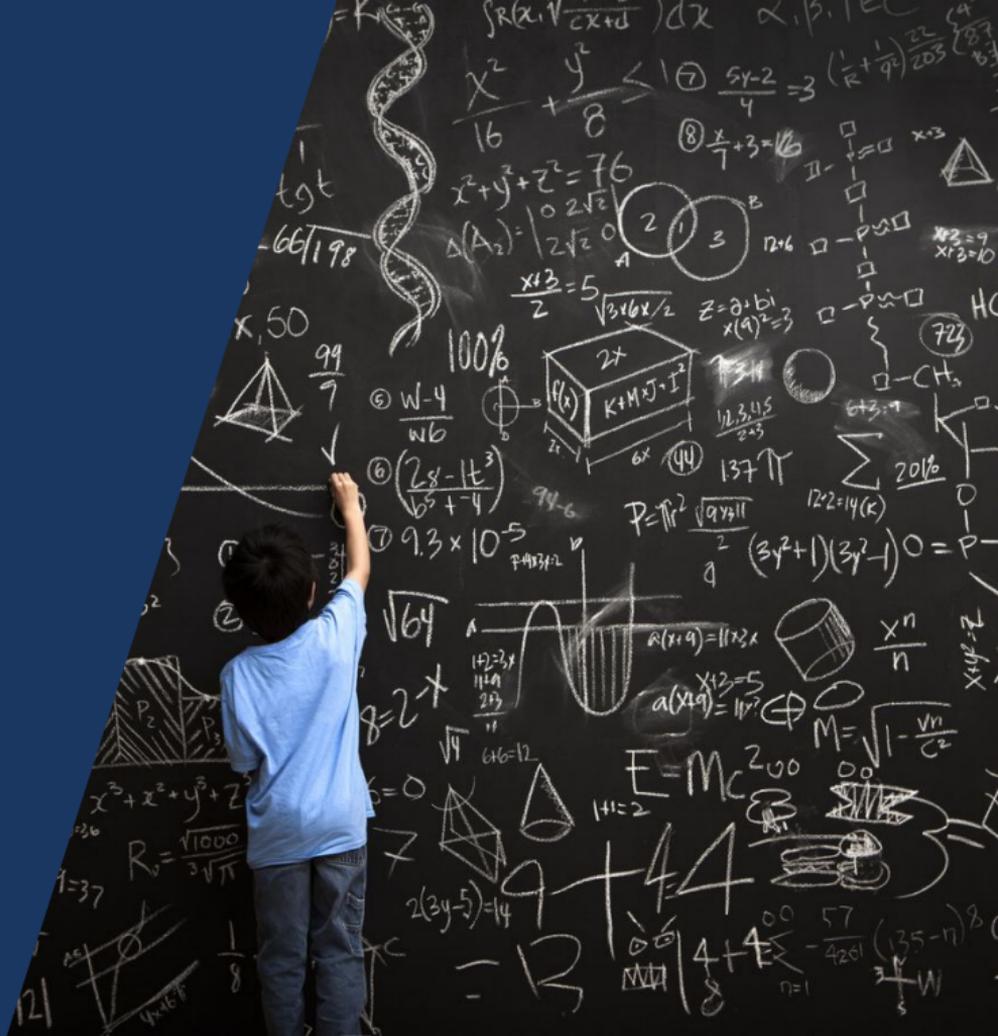
Johansen

FCVAR

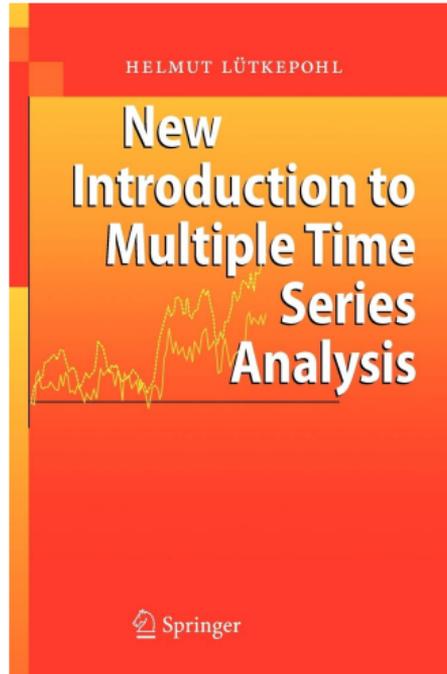


Chapitre 1

Les VAR Stationnaires



Ouvrage de référence



Les Séries Temporelles Multivariées

- L'analyse de interdépendance entre plusieurs variables aléatoires impose un cadre multivarié
- L'objet d'analyse sera donc un vecteur de k séries temporelles

$$Y_t = (y_{1,t}, y_{2,t}, \dots, y_{kt})'$$

avec $t = 1, 2, \dots, n$

La stationnarité en covariance

- Dans ce Chapitre on suppose que Y_t est stationnaire en covariance
- Dans ce cadre multivarié cela impose que

$$\mathbb{E}(Y_t) = (\mu_1, \mu_2, \dots, \mu_k)' = \mu, \quad \forall t$$

et

$$\mathbb{E}((Y_t - \mu)(Y_t - \mu)') = \Gamma_Y(0) = \begin{pmatrix} \sigma_{11}^2 & \dots & \dots & \sigma_{1k} \\ \vdots & \sigma_{22}^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ \sigma_{1k} & \dots & \dots & \sigma_{kk}^2 \end{pmatrix} = \Sigma_Y, \quad \forall t$$

qui découle de l'indépendance au temps de l'autocovariance

$$\begin{aligned} \Gamma_Y(h) &= \mathbb{E}((Y_t - \mu)(Y_{t-h} - \mu)') \\ &= \mathbb{E}((Y_{t+h} - \mu)(Y_t - \mu)') = \Gamma_Y(-h), \quad \forall h \end{aligned}$$

Note Pour $h \neq 0$, $\Gamma_Y(h)$ n'est pas forcément symétrique car, e.g.,

$$\text{Cov}(y_{1,t}, y_{2,t-h})$$

n'a pas de raison d'être identique à $\text{Cov}(y_{2,t}, y_{1,t-h})$

Théorème de Wold multivarié

- Sous l'hypothèse de stationnarité de Y_t et si ce dernier est causale

... le Th. de Wold assure que Y_t peut s'écrire comme

$$Y_t = \mu + \sum_{i=0}^{\infty} \Psi_i \varepsilon_{t-i} = (I_k + \Psi_1 L + \Psi_2 L^2 + \dots) \varepsilon_t \quad (1)$$

où $\varepsilon_t = (\varepsilon_{1,t}, \varepsilon_{2,t}, \dots, \varepsilon_{k,t})' \sim \text{WN}(\mathbf{0}_k, \Sigma_\varepsilon)$, ce qui implique

$$\mathbb{E}(\varepsilon_t \varepsilon_{t-h}') = \Gamma_\varepsilon(h) = 0, \quad \forall h \neq 0$$

⇒ Il s'agit donc d'une représentation Vectorielle MA(∞)

Note Σ_ε n'est pas nécessairement diagonale ce qui implique que, e.g.,

$$\text{Cov}(\varepsilon_{1,t}, \varepsilon_{2,t}) = \text{Cov}(\varepsilon_{2,t}, \varepsilon_{1,t}) \neq 0$$

est possible en instantané

VMA(∞), stationnarité et ergodicité

- Tous les processus stationnaires causaux admettent une représentation VMA(∞)
- ... mais tous les VMA(∞) ne sont pas stationnaires
- Comme pour les MA(∞), une condition suffisante pour la stationnarité sera que

$$\{\Psi_j\}_{j=0}^{\infty}$$

soient absolument sommable

Note 1 Une matrice est absolument sommable si chacun de ses éléments définit une suite de scalaires absolument sommables

$$\sum_{s=0}^{\infty} |\Psi_{s_{i,j}}| < \infty$$

avec $s_{i,j}$ l'élément en ligne i et colonne j de Ψ_s

Note 2 Cette condition suffit pour l'ergodicité qui implique la stationnarité

Note 3 L'ergodicité assure que les moments empiriques convergent vers les moments théoriques

VMA(q)

- Tout comme en univarié, au-delà d'un certain retard

$$q < \infty$$

les coefficients de la représentation VMA(∞) peuvent être nuls

- Dans ce cas on parle de modèle VMA(q)

$$Y_t = \mu + \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \Psi_2 \varepsilon_{t-2} + \dots + \Psi_q \varepsilon_{t-q}$$

- Comme pour les MA(q), si $q < \infty$, Y_t est stationnaire et

$$\Gamma_Y(0) = \Sigma_\varepsilon + \Psi_1 \Sigma_\varepsilon \Psi_1' + \Psi_2 \Sigma_\varepsilon \Psi_2' + \dots + \Psi_q \Sigma_\varepsilon \Psi_q' = \Sigma_Y$$

et

$$\Gamma_Y(h) = \Psi_h \Sigma_\varepsilon + \Psi_{h+1} \Sigma_\varepsilon \Psi_1' + \Psi_{h+2} \Sigma_\varepsilon \Psi_2' + \dots + \Psi_q \Sigma_\varepsilon \Psi_{q-h}'$$

$$= \sum_{i=0}^{q-h} \Psi_{i+h} \Sigma_\varepsilon \Psi_i', \quad h = 1, 2, \dots, q$$

$$\Gamma_Y(h) = 0, \quad \forall |h| > q$$

VMA(1) et inversibilité

- Si le VMA(1) est inversible il aura une représentation Vectorielle AR
- Considérons $Y_t = \varepsilon_t + \Theta_1 \varepsilon_{t-1}$, avec $k = 2$ et Θ_1 une matrice 2×2 , ayant pour écriture en équations

$$y_{1,t} = \varepsilon_{1,t} + \theta_{11}\varepsilon_{1,t-1} + \theta_{12}\varepsilon_{2,t-1}$$

$$y_{2,t} = \varepsilon_{2,t} + \theta_{22}\varepsilon_{2,t-1} + \theta_{21}\varepsilon_{1,t-1}$$

- En écriture matricielle on a alors que

$$\begin{aligned}\varepsilon_t &= Y_t - \Theta_1 \varepsilon_{t-1} = Y_t - \Theta_1(Y_{t-1} - \Theta_1 \varepsilon_{t-2}) = Y_t - \Theta_1 Y_{t-1} + \Theta_1^2 \varepsilon_{t-2} \\ &= Y_t - \Theta_1 Y_{t-1} + \dots + (-\Theta_1)^j Y_{t-j} + (-\Theta_1)^{j+1} \varepsilon_{t-j-1} \\ &= Y_t + \sum_{i=1}^{\infty} (-\Theta_1)^i Y_{t-i} = \sum_{i=0}^{\infty} (-\Theta_1)^i Y_{t-i}\end{aligned}\tag{2}$$

à condition que $(-\Theta_1)^i \rightarrow 0$ quand $i \rightarrow \infty$ de telle sorte que

$$(-\Theta_1)^{j+1} \varepsilon_{t-j-1} \rightarrow 0$$

et donc la représentation VAR(∞) apparaît

$$Y_t = - \sum_{i=1}^{\infty} (-\Theta_1)^i Y_{t-i} + \varepsilon_t$$

VMA(1) et conditions d'inversibilité

- Comment garantir que $(-\Theta_1)^i \rightarrow 0$ dans ce contexte matriciel ?
- Comme dans le cas univarié, on constate que

$$\begin{aligned} Y_t &= \varepsilon_t + \Theta_1 \varepsilon_{t-1} \\ &= (I_k - \Theta_1 L) \varepsilon_t \\ \varepsilon_t &= (I_k + \Theta_1 L)^{-1} Y_t \\ \Rightarrow (I_k + \Theta_1 L)^{-1} &= \Theta(L)^{-1} = \sum_{i=0}^{\infty} (-\Theta_1)^i L^i \end{aligned}$$

⇒ les conditions de convergence sont donc liées à l'inversibilité de $\Theta(L)$

- Par l'algèbre matriciel il vient donc que $\Theta(L)^{-1}$ existe si

$$\det(I_k + \Theta_1 z) \neq 0, \text{ for } z \in \mathbb{C} \text{ and } |z| \leq 1$$

ou de manière équivalente,

$$\det(I_k + \Theta_1 z) = 0, \text{ for } z \in \mathbb{C} \text{ and } |z| > 1$$

ou encore, **les valeurs propres de Θ_1 sont toutes inférieures à 1 en module**

VMA(q) et conditions d'inversibilité

- Si le VMA(1) est inversible on pourra donc l'écrire comme

$$\varepsilon_t = \sum_{i=0}^{\infty} (-\Theta_1)^i Y_{t-i} = \Theta(L)^{-1} Y_t$$

ou comme

$$\varepsilon_t = \Pi(L) Y_t = Y_t - \sum_{i=1}^{\infty} \Pi_i Y_{t-i}$$

avec $\Pi(L) = \Theta(L)^{-1}$, ce qui implique $Y_t = \sum_{i=1}^{\infty} \Pi_i Y_{t-i} + \varepsilon_t$

- Plus généralement, pour un VMA(q), la représentation VAR(∞)

$$Y_t = \sum_{i=1}^{\infty} \Pi_i Y_{t-i} + \varepsilon_t$$

existera si

$$\det(I_k + \Theta_1 z + \Theta_2 z^2 + \dots + \Theta_q z^q) \neq 0, \text{ for } z \in \mathcal{C} \text{ and } |z| \leq 1$$

avec

$$\Pi(L) = I_k - \sum_{i=1}^{\infty} \Pi_i L^i = \Theta(L)^{-1}$$

VAR(p)

- On a vu que les VMA(q) inversibles admettent des représentations VAR(∞)
- On peut bien entendu définir un VAR(p) qui pour $p = 1$ et $k = 2$ donne

$$\begin{aligned}Y_{1,t} &= c_1 + \phi_{11}Y_{1,t-1} + \phi_{12}Y_{2,t-1} + \varepsilon_{1,t} \\Y_{2,t} &= c_2 + \phi_{22}Y_{2,t-1} + \phi_{21}Y_{1,t-1} + \varepsilon_{2,t}\end{aligned}$$

dont l'écriture matricielle sera $Y_t = C + \Phi_1 Y_{t-1} + \varepsilon_t$ avec

$$\Sigma_\varepsilon = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{21} & \sigma_{22}^2 \end{pmatrix} \text{ et } \Gamma_\varepsilon(h \neq 0) = 0$$

- Plus généralement, pour tout $p < \infty$ on aura

$$Y_t = C + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \dots + \Phi_p Y_{t-p} + \varepsilon_t$$

Note Comme pour les AR(p) et à la différence des VMA(q), $\mathbb{E}(Y) \neq C$ mais

$$\mathbb{E}(Y) = (I_k - \Phi_1)^{-1} C$$

$p = 1$ par exemple (nous verrons plus tard le cas $p > 1$)

VAR(1) et stabilité

Rappel Les séries d'autocovariance absolument sommables sont stationnaires

- Pour les VAR(p) cette condition est assurée par la **stabilité**
- Considérons un VAR(1) et itérons depuis le début des temps

$$Y_1 = C + \Phi_1 Y_0 + \varepsilon_1$$

$$\begin{aligned} Y_2 &= C + \Phi_1 Y_1 + \varepsilon_2 = C + \Phi_1(C + \Phi_1 Y_0 + \varepsilon_1) + \varepsilon_2 \\ &= (I_k + \Phi_1)C + \Phi_1^2 Y_0 + \Phi_1 \varepsilon_1 + \varepsilon_2 \end{aligned}$$

$$Y_t = \dots = (I_k + \Phi_1 + \dots + \Phi_1^{t-1})C + \Phi_1^t Y_0 + \sum_{i=0}^{t-1} \Phi_1^i \varepsilon_{t-i}$$

- En considérant un passé infini on s'affranchit de Y_0 et

$$Y_t = (I_k + \Phi_1 + \dots + \Phi_1^j)C + \Phi_1^{j+1} Y_{t-j-1} + \sum_{i=0}^j \Phi_1^i \varepsilon_{t-i}$$

⇒ Une première étape vers la stabilité est que

$$\Phi_1^{j+1} \rightarrow 0, \quad j \rightarrow \infty$$

ce qui sera le cas si toutes les valeurs propres de Φ_1 sont inférieures à 1

VAR(1) et conditions de stabilité

- Les conditions sur les valeurs propres vont également assurer que

$\{\Phi_1^i\}_{i=0}^\infty$ est absolument sommable

et que $(I_k + \Phi_1 + \dots + \Phi_1^j)C \rightarrow (I_k - \Phi_1)^{-1}C = \mathbb{E}(Y)$

- Mais d'où viennent ces conditions sur les valeurs propres ?
- Comme pour l'AR(1), on voit $Y_t - \Phi_1 Y_{t-1} = C + \varepsilon_t$ s'écrit aussi

$$Y_t = (I_k - \Phi_1 L)^{-1}C + (I_k - \Phi_1 L)^{-1}\varepsilon_t$$

$$\Rightarrow (I_k - \Phi_1 L)^{-1} = \Phi(L)^{-1} = \sum_{i=0}^{\infty} \Phi_1^i L^i$$

⇒ Les conditions de sommabilité sont donc liées à l'inversibilité de $\Phi(L)$

- Comme pour le VMA, par l'algèbre matricielle, $\Phi(L)^{-1}$ existe si

$$\det(I_k - \Phi_1 z) \neq 0, \text{ for } z \in \mathcal{C} \text{ and } |z| \leq 1$$

ou de manière équivalente, $\det(I_k - \Phi_1 z) = 0$, for $z \in \mathcal{C}$ and $|z| > 1$ ou si **les valeurs propres de Φ_1 sont toutes inférieures à 1 en module**

VAR(1) stationnaire

- La stabilité implique la stationnarité car $\mathbb{E}(Y) = \mu$ pour tout t

... et car on a également que

$$\begin{aligned}\Gamma_Y(h) &= \mathbb{E}((Y_t - \mu)(Y_{t-h} - \mu)') \\ &= \lim_{n \rightarrow \infty} \sum_{i=0}^n \sum_{j=0}^n \Phi_1^i \mathbb{E}(\varepsilon_{t-i} \varepsilon'_{t-h-j}) (\Phi_1^j)' \\ &= \lim_{n \rightarrow \infty} \sum_{i=0}^n \Phi_1^{h+i} \Sigma_\varepsilon (\Phi_1^i)' = \sum_{i=0}^{\infty} \Phi_1^{h+i} \Sigma_\varepsilon (\Phi_1^i)'\end{aligned}$$

⇒ La convergence des autocovariances est assurée par la stabilité

- Comme $\Gamma_Y(h)$ est indépendante de t , Y_t est stationnaire

Note La stationnarité n'implique pas la stabilité

Note La stationnarité implique l'existence de la représentation de Wold

$$Y_t = \mu + \Psi(L)\varepsilon_t = \mu + \sum_{i=0}^{\infty} \Psi_i \varepsilon_{t-i} \text{ avec } \Psi(L) = \Phi(L)^{-1}$$

Exemple de VAR(1) trivarié

- Soit le VAR(1) suivant avec $\varepsilon_t \sim \text{i. i. d. } (0, \Sigma_\varepsilon)$

$$Y_t = C + \begin{pmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{pmatrix} Y_{t-1} + \varepsilon_t$$

- Écrivez les 3 équations du VAR et démontrez sa stabilité

Exemple de VAR(1) trivarié

- Soit le VAR(1) suivant avec $\varepsilon_t \sim \text{i. i. d. } (0, \Sigma_\varepsilon)$

$$Y_t = C + \begin{pmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{pmatrix} Y_{t-1} + \varepsilon_t$$

- Écrivez les 3 équations du VAR et démontrez sa stabilité
- Le polynôme caractéristique est ici donné par

$$\begin{aligned} \det \left(I_3 - \begin{pmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{pmatrix} z \right) &= \det \begin{pmatrix} 1 - 0.5z & 0 & 0 \\ -0.1z & 1 - 0.1z & -0.3z \\ 0 & -0.2z & 1 - 0.3z \end{pmatrix} \\ &= (1 - 0.5z)(1 - 0.4z - 0.03z^2) \end{aligned}$$

Exemple de VAR(1) trivarié

- Soit le VAR(1) suivant avec $\varepsilon_t \sim \text{i. i. d. } (0, \Sigma_\varepsilon)$

$$Y_t = C + \begin{pmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{pmatrix} Y_{t-1} + \varepsilon_t$$

- Écrivez les 3 équations du VAR et démontrez sa stabilité
- Le polynôme caractéristique est ici donné par

$$\begin{aligned} \det \left(I_3 - \begin{pmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{pmatrix} z \right) &= \det \begin{pmatrix} 1 - 0.5z & 0 & 0 \\ -0.1z & 1 - 0.1z & -0.3z \\ 0 & -0.2z & 1 - 0.3z \end{pmatrix} \\ &= (1 - 0.5z)(1 - 0.4z - 0.03z^2) \end{aligned}$$

- On voit alors que les 3 racines qui annulent ce polynôme sont

$$z_1 = 2, \quad z_2 = 2.1525, \quad z_3 = -15.4858$$

- Toutes les racines satisfont $|z| > 1$, ce VAR(1) est donc stable

VAR(p) et réécriture VAR(1)

- La généralisation aux VAR(p) est facilitée par une réécriture en VAR(1)
- Considérons un VAR(p)

$$Y_t = C + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \dots + \Phi_p Y_{t-p} + \varepsilon_t$$

- On peut alors encapsuler Y_t dans un VAR(1) en définissant

$$Y_t = \mu + \Phi Y_{t-1} + \varepsilon_t$$

dont la forme développée donne

$$\begin{pmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-p+1} \end{pmatrix} = \begin{pmatrix} c \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_p \\ I_k & 0 & \cdots & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & I_k & 0 \end{pmatrix} \begin{pmatrix} Y_{t-1} \\ Y_{t-2} \\ \vdots \\ Y_{t-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

La structure de Φ est atypique et porte le nom de matrice compagnon

Décomposition de la matrice compagnon

- Bien que Φ soit de dimension $kp \times kp$ la logique d'itération appliquée pour la stabilité du VAR(1) reste la même
- On a donc après j itérations

$$Y_t = (I_k + \Phi + \dots + \Phi^{j-1})C + \Phi^j Y_{t-j} + \sum_{i=0}^{j-1} \Phi^i \varepsilon_{t-i}$$

- On sait à présent que l'absolue sommabilité dépend des valeurs propres

$$\lambda_1, \lambda_2, \dots, \lambda_{kp}$$

regroupées sur la diagonal de Λ , qui apparaît dans la décomposition

$$\Phi = V\Lambda V^{-1}$$

et implique que $\Phi^j = V\Lambda^j V^{-1}$

⇒ En conséquence, si $|\lambda_j| < 1$, quand $j \rightarrow \infty$

$$\lim_{j \rightarrow \infty} \Phi^j = \lim_{j \rightarrow \infty} V\Lambda^j V^{-1} = 0$$

et on voit que la séquence $\{\Phi^j\}_{j=0}^{\infty}$ est absolument sommable

VAR(p) et conditions de stationnarité

- On peut également réécrire \mathbf{Y}_t comme

$$\mathbf{Y}_t = (\mathbf{I}_{kp} - \Phi \mathbf{L})^{-1} \mathbf{C} + (\mathbf{I}_{kp} - \Phi \mathbf{L})^{-1} \varepsilon$$

- Comme $(\mathbf{I}_{kp} - \Phi \mathbf{z}) = (\mathbf{I}_k - \Phi_1 \mathbf{z} - \Phi_2 \mathbf{z}^2 - \dots - \Phi_p \mathbf{z}^p)$ il vient que

$$\det(\mathbf{I}_k - \Phi_1 \mathbf{z} - \Phi_2 \mathbf{z}^2 - \dots - \Phi_p \mathbf{z}^p) \neq 0, \text{ for } z \in \mathbb{C} \text{ and } |z| \leq 1$$

est une condition équivalente de stabilité d'un VAR(p)

⇒ La stationnarité en découle puisque $\mathbb{E}(\mathbf{Y}_t) = (\mathbf{I}_{kp} - \Phi \mathbf{L})^{-1} \mathbf{C}$ et

$$\Gamma_Y(h) = \mathbb{E}((\mathbf{Y}_t - \boldsymbol{\mu})(\mathbf{Y}_{t-h} - \boldsymbol{\mu})') = \sum_{i=0}^{\infty} \Phi^{h+i} \Sigma_{\varepsilon} (\Phi^i)'$$

- Retrouver le VAR(p) originel, il suffit de considérer $Y_t = \mathbf{J} \mathbf{Y}_t$ où

$$\mathbf{J} = (\mathbf{I}_k, \mathbf{0}_{k \times (kp-k)})$$

est une matrice de dimension $k \times kp$

⇒ On a alors $\mathbb{E}(Y_t) = \mathbf{J} \boldsymbol{\mu}$ et $\Gamma_Y(h) = \mathbf{J} \Gamma_Y(h) \mathbf{J}'$ et enfin

$$Y_t = \mathbf{J} \mathbf{Y} = \mathbf{J} \boldsymbol{\mu} + \sum_{i=0}^{\infty} \mathbf{J} \Phi^i \mathbf{J}' \varepsilon_{t-i} = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \Psi_i \varepsilon_{t-i}$$

Exemple de VAR(2) bivarié

- Soit le VAR(2) suivant avec $\varepsilon_t \sim \text{i. i. d. } (0, \Sigma_\varepsilon)$

$$Y_t = C + \begin{pmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{pmatrix} Y_{t-1} + \begin{pmatrix} 0 & 0 \\ 0.25 & 0 \end{pmatrix} Y_{t-2} + \varepsilon_t$$

- Écrivez les 2 équations du VAR et étudiez sa stabilité

Exemple de VAR(2) bivarié

- Soit le VAR(2) suivant avec $\varepsilon_t \sim \text{i. i. d. } (0, \Sigma_\varepsilon)$

$$Y_t = C + \begin{pmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{pmatrix} Y_{t-1} + \begin{pmatrix} 0 & 0 \\ 0.25 & 0 \end{pmatrix} Y_{t-2} + \varepsilon_t$$

- Écrivez les 2 équations du VAR et étudiez sa stabilité
- Le polynôme caractéristique est ici donné par

$$\begin{aligned} & \det \left(I_2 - \begin{pmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{pmatrix} z - \begin{pmatrix} 0 & 0 \\ 0.25 & 0 \end{pmatrix} z^2 \right) \\ &= \det \begin{pmatrix} 1 - 0.5z & -0.1z \\ -0.4z - 0.25z^2 & 1 - 0.5z \end{pmatrix} \\ &= 1 - z + 0.21z^2 - 0.025z^3 = 0.025(1.3 - z)(30.8 - 7.1z + z^2) \end{aligned}$$

Exemple de VAR(2) bivarié

- Soit le VAR(2) suivant avec $\varepsilon_t \sim \text{i. i. d. } (0, \Sigma_\varepsilon)$

$$Y_t = C + \begin{pmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{pmatrix} Y_{t-1} + \begin{pmatrix} 0 & 0 \\ 0.25 & 0 \end{pmatrix} Y_{t-2} + \varepsilon_t$$

- Écrivez les 2 équations du VAR et étudiez sa stabilité
- Le polynôme caractéristique est ici donné par

$$\begin{aligned} & \det \left(I_2 - \begin{pmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{pmatrix} z - \begin{pmatrix} 0 & 0 \\ 0.25 & 0 \end{pmatrix} z^2 \right) \\ &= \det \begin{pmatrix} 1 - 0.5z & -0.1z \\ -0.4z - 0.25z^2 & 1 - 0.5z \end{pmatrix} \\ &= 1 - z + 0.21z^2 - 0.025z^3 = 0.025(1.3 - z)(30.8 - 7.1z + z^2) \end{aligned}$$

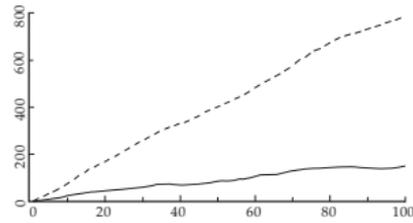
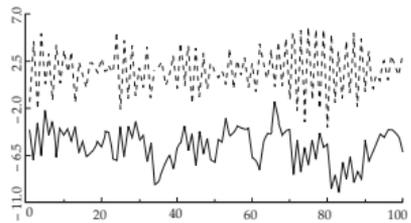
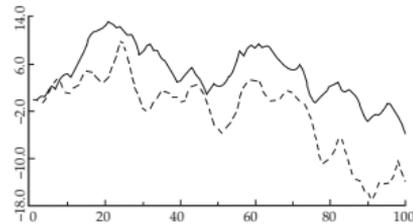
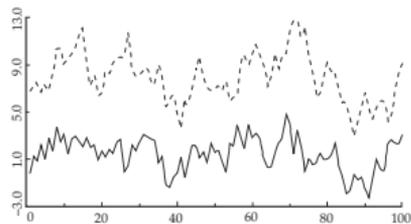
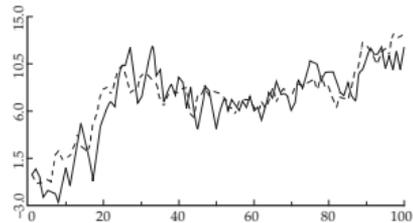
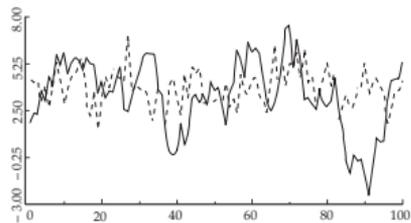
- On voit alors que les 3 racines qui annulent ce polynôme sont

$$z_1 = 1.3, \quad z_2 = 3.55 + 4.26i, \quad z_3 = 3.55 - 4.26i$$

et que les modules sont $|z_2| = |z_3| = \sqrt{3.55^2 + 4.26^2} = 5.545$

- Toutes les racines satisfont $|z| > 1$, ce VAR(2) est donc stable

Exemples de trajectoires stables et instables



Bivariate time series generated by stable processes

Bivariate time series generated by unstable VAR processes

Dérivation de la fonction d'autocovariance

- Considérons un VAR(1) centré

$$Y_t - \mu = \Phi_1(Y_{t-1} - \mu) + \varepsilon_t$$

- Pour obtenir $\Gamma(h)$, on multiplie par $(Y_{t-h} - \mu)'$ et on prend l'espérance

$$\mathbb{E}((Y_t - \mu)(Y_{t-h} - \mu)') = \Phi_1 \mathbb{E}((Y_{t-1} - \mu)(Y_{t-h} - \mu)') + \mathbb{E}(\varepsilon_t(Y_{t-h} - \mu)')$$

- On voit alors que pour $h = 0$ on obtient

$$\Gamma_Y(0) = \Phi_1 \Gamma_Y(-1) + \Sigma_\varepsilon = \Phi_1 \Gamma_Y(1)' + \Sigma_\varepsilon$$

et pour $h > 0$ on a

$$\Gamma_Y(h) = \Phi_1 \Gamma_Y(h-1)$$

- L'analogie est donc immédiate avec les équations de Yule-Walker en univarié

Calcul de la fonction d'autocovariance

- En supposant Φ_1 et Σ_ε donné, on peut déterminer $\Gamma_Y(0)$
- En effet, on sait que $\Gamma_Y(1) = \Phi_1 \Gamma_Y(0)$ ce qui donne

$$\Gamma_Y(0) = \Phi_1 \Gamma_Y(0) \Phi_1' + \Sigma_\varepsilon$$

- L'opérateur $\text{vec}(A)$ permet d'empiler les colonnes de

$$A_{a \times b}$$

pour convertir la matrice en vecteur $B = \text{vec}(A)$ de dimension $ab \times 1$

- En faisant appel au produit de Kronecker on a alors

$$\text{vec } \Gamma_Y(0) = \text{vec}(\Phi_1 \Gamma_Y(0) \Phi_1') + \text{vec } \Sigma_\varepsilon \quad (3)$$

$$= (\Phi_1 \otimes \Phi_1) \text{vec } \Gamma_Y(0) + \text{vec } \Sigma_\varepsilon \quad (4)$$

$$= (I_{k^2} - \Phi_1 \otimes \Phi_1)^{-1} \text{vec } \Sigma_\varepsilon \quad (5)$$

- Pour un VAR(p), en utilisant la représentation VAR(1) on obtient

$$\text{vec } \Gamma_Y(0) = (I_{(kp)^2} - \Phi \otimes \Phi)^{-1} \text{vec } \Sigma_\varepsilon$$

Exercice : autocovariance d'un VAR(1)

- Soit le VAR(1) vu précédemment

$$Y_t = C + \begin{pmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{pmatrix} Y_{t-1} + \varepsilon_t \text{ avec } \Sigma_\varepsilon = \begin{pmatrix} 2.25 & 0 & 0 \\ 0 & 1 & 0.5 \\ 0 & 0.5 & 0.74 \end{pmatrix}$$

- Exprimez sous forme matricielle $\text{vec } \Gamma_Y(0) = (I_9 - \Phi_1 \otimes \Phi_1)^{-1} \text{vec } \Sigma_\varepsilon = V$

Exercice : autocovariance d'un VAR(1)

- Soit le VAR(1) vu précédemment

$$Y_t = C + \begin{pmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{pmatrix} Y_{t-1} + \varepsilon_t \text{ avec } \Sigma_\varepsilon = \begin{pmatrix} 2.25 & 0 & 0 \\ 0 & 1 & 0.5 \\ 0 & 0.5 & 0.74 \end{pmatrix}$$

- Exprimez sous forme matricielle $\text{vec } \Gamma_Y(0) = (I_9 - \Phi_1 \otimes \Phi_1)^{-1} \text{vec } \Sigma_\varepsilon = V$

$$V = \begin{pmatrix} .75 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -.05 & .95 & -.15 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -.1 & .85 & 0 & 0 & 0 & 0 & 0 & 0 \\ -.05 & 0 & 0 & .95 & 0 & 0 & -.15 & 0 & 0 \\ -.01 & -.01 & -.03 & -.01 & .99 & -.03 & -.03 & -.03 & -.09 \\ 0 & -.02 & -.03 & 0 & -.02 & .97 & 0 & -.06 & -.09 \\ 0 & 0 & 0 & -.01 & 0 & 0 & .85 & 0 & 0 \\ 0 & 0 & 0 & -.02 & -.02 & -.06 & -.03 & .97 & -.09 \\ 0 & 0 & 0 & 0 & -.04 & -.06 & 0 & -.06 & .91 \end{pmatrix}^{-1} \begin{pmatrix} 2.25 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0.5 \\ 0 \\ 0.5 \\ 0.74 \end{pmatrix}$$

$$= \begin{pmatrix} 3.000 \\ 0.161 \\ 0.019 \\ 0.161 \\ 1.172 \\ 0.674 \\ 0.019 \\ 0.674 \\ 0.954 \end{pmatrix} \Leftrightarrow \Gamma_Y(0) = \begin{pmatrix} 3 & 0.161 & 0.019 \\ 0.161 & 1.172 & 0.674 \\ 0.019 & 0.674 & 0.954 \end{pmatrix}, \Gamma_Y(1) = \begin{pmatrix} 1.500 & 0.080 & 0.009 \\ 0.322 & 0.335 & 0.355 \\ 0.038 & 0.437 & 0.421 \end{pmatrix}$$

Autocorrélations d'un VAR(p) stable

- La fonction d'autocorrélation découle de l'ACovF et s'écrit

$$R_Y(h) = D^{-1}\Gamma_Y(h)D^{-1}$$

où

$$D^{-1} = \begin{pmatrix} 1/\sqrt{\gamma_{11}(0)} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1/\sqrt{\gamma_{kk}(0)} \end{pmatrix}$$

avec $\gamma_{ij}(h)$ l'élément ligne i et colonne j de la matrice $\Gamma_Y(h)$

Exercice Calculez l'ACF du VAR(1) précédent pour $h = 0, 1, 2$ sachant que

$$D = \begin{pmatrix} \sqrt{3} & 0 & 0 \\ 0 & \sqrt{1.172} & 0 \\ 0 & 0 & \sqrt{0.954} \end{pmatrix}$$

Autocorrélations d'un VAR(p) stable

- La fonction d'autocorrélation découle de l'ACovF et s'écrit

$$R_Y(h) = D^{-1}\Gamma_Y(h)D^{-1}$$

où

$$D^{-1} = \begin{pmatrix} 1/\sqrt{\gamma_{11}(0)} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1/\sqrt{\gamma_{kk}(0)} \end{pmatrix}$$

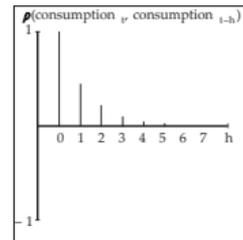
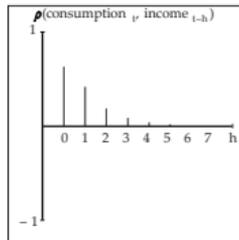
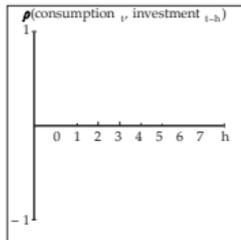
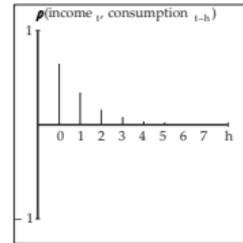
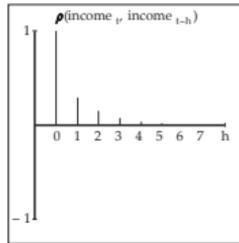
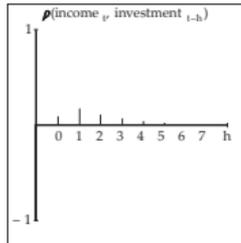
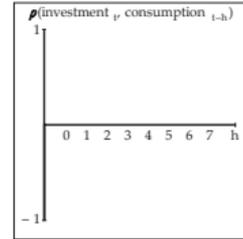
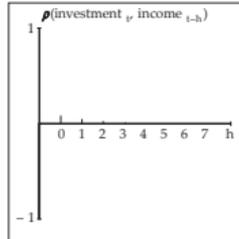
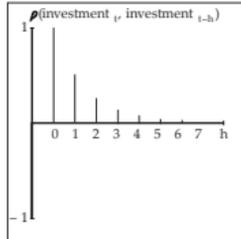
avec $\gamma_{ij}(h)$ l'élément ligne i et colonne j de la matrice $\Gamma_Y(h)$

Exercice Calculez l'ACF du VAR(1) précédent pour $h = 0, 1, 2$ sachant que

$$D = \begin{pmatrix} \sqrt{3} & 0 & 0 \\ 0 & \sqrt{1.172} & 0 \\ 0 & 0 & \sqrt{0.954} \end{pmatrix}$$

$$R(0) = \begin{pmatrix} 1 & 0.86 & 0.011 \\ & 1 & 0.637 \\ & & 1 \end{pmatrix}, R(1) = \begin{pmatrix} 0.5 & 0.043 & 0.005 \\ 0.172 & 0.286 & 0.336 \\ 0.022 & 0.413 & 0.441 \end{pmatrix}, R(2) = \begin{pmatrix} 0.25 & 0.021 & 0.003 \\ 0.103 & 0.148 & 0.154 \\ 0.045 & 0.187 & 0.206 \end{pmatrix}$$

Exemples d'autocorrélation d'un VAR(1) trivarié



MSE et prévisions optimales

- Tout comme en univarié, on peut montrer qu'au sens de la MSE

$$MSE = \mathbb{E}({}_t Y_{t+h} - Y_{t+h})^2 = \mathbb{E}(\epsilon)^2$$

i.e. l'erreur quadratique moyenne, le prédicteur optimal sera

$${}_t Y_{t+h} = \mathbb{E}(Y_{t+h} | \mathcal{F}_t)$$

l'espérance conditionnelle, avec $\mathcal{F}_t = (Y_t, Y_{t-1}, \dots)'$

- Dans le cadre d'un VAR(p) il vient que

$$\mathbb{E}(Y_{t+h} | \mathcal{F}_t) = C + \sum_{i=1}^p \Phi_i \mathbb{E}(Y_{t+h-i} | \mathcal{F}_t)$$

si $\epsilon_t \sim$ i. i. d. $(0, \Sigma_\epsilon)$ car $\mathbb{E}(\epsilon_{t+h} | \mathcal{F}_t)$ pour $h > 0$ et où

$$\mathbb{E}(Y_{t+h-i} | \mathcal{F}_t) = Y_{t+h-i}, \text{ si } h \leq i$$

- Bien entendu on peut montrer que

$$\mathbb{E}(Y_{t+h} - \mathbb{E}(Y_{t+h} | \mathcal{F}_t)) = 0$$

et si $\epsilon_t \sim$ i. i. d. que $MSE(\mathbb{E}(Y_{t+h} | \mathcal{F}_t)) = MSE(\mathbb{E}(Y_{t+h} | \mathcal{F}_t) | \mathcal{F}_t)$

... et donc plus aucune information pertinente pour prédire reste dans \mathcal{F}_t

Prévisions d'un VAR(1)

- Pour $h = 1$ la formule de $\mathbb{E}(Y_{t+h}|\mathcal{F}_t)$ donne par récursion

$$\mathbb{E}(Y_{t+1}|\mathcal{F}_t) = C + \Phi_1 Y_t + \dots + \Phi_p Y_{t+1-p} \quad (6)$$

$$\mathbb{E}(Y_{t+2}|\mathcal{F}_t) = C + \Phi_1 \mathbb{E}(Y_{t+1}|\mathcal{F}_t) + \Phi_2 Y_t + \dots + \Phi_p Y_{t+2-p} \quad \vdots \quad (7)$$

- Pour un VAR(1) et $h \geq 1$ cela conduit à

$$\mathbb{E}(Y_{t+h}|\mathcal{F}_t) = (I_k + \Phi_1 + \dots + \Phi_1^{h-1})C + \Phi_1^h Y_t$$

Exemple Avec le VAR(1) précédent, si $Y_t = (-6, 3, 5)'$ et $C = (0, 2, 1)'$ on obtient pour $h = 1$?

Prévisions d'un VAR(1)

- Pour $h = 1$ la formule de $\mathbb{E}(Y_{t+h}|\mathcal{F}_t)$ donne par récursion

$$\mathbb{E}(Y_{t+1}|\mathcal{F}_t) = C + \Phi_1 Y_t + \dots + \Phi_p Y_{t+1-p} \quad (6)$$

$$\mathbb{E}(Y_{t+2}|\mathcal{F}_t) = C + \Phi_1 \mathbb{E}(Y_{t+1}|\mathcal{F}_t) + \Phi_2 Y_t + \dots + \Phi_p Y_{t+2-p} \quad \vdots \quad (7)$$

- Pour un VAR(1) et $h \geq 1$ cela conduit à

$$\mathbb{E}(Y_{t+h}|\mathcal{F}_t) = (I_k + \Phi_1 + \dots + \Phi_1^{h-1})C + \Phi_1^h Y_t$$

Exemple Avec le VAR(1) précédent, si $Y_t = (-6, 3, 5)'$ et $C = (0, 2, 1)'$ on obtient pour $h = 1$?

$$\mathbb{E}(Y_{t+1}|\mathcal{F}_t) = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix} + \begin{pmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0. & 0.2 & 0.3 \end{pmatrix} \begin{pmatrix} -6 \\ 3 \\ 5 \end{pmatrix} = \begin{pmatrix} -3 \\ 3.2 \\ 3.1 \end{pmatrix}$$

Prévisions d'un VAR(1)

- Pour $h = 1$ la formule de $\mathbb{E}(Y_{t+h}|\mathcal{F}_t)$ donne par récursion

$$\mathbb{E}(Y_{t+1}|\mathcal{F}_t) = C + \Phi_1 Y_t + \dots + \Phi_p Y_{t+1-p} \quad (6)$$

$$\mathbb{E}(Y_{t+2}|\mathcal{F}_t) = C + \Phi_1 \mathbb{E}(Y_{t+1}|\mathcal{F}_t) + \Phi_2 Y_t + \dots + \Phi_p Y_{t+2-p} \quad \vdots \quad (7)$$

- Pour un VAR(1) et $h \geq 1$ cela conduit à

$$\mathbb{E}(Y_{t+h}|\mathcal{F}_t) = (I_k + \Phi_1 + \dots + \Phi_1^{h-1})C + \Phi_1^h Y_t$$

Exemple Avec le VAR(1) précédent, si $Y_t = (-6, 3, 5)'$ et $C = (0, 2, 1)'$ on obtient pour $h = 1$?

$$\mathbb{E}(Y_{t+1}|\mathcal{F}_t) = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix} + \begin{pmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0. & 0.2 & 0.3 \end{pmatrix} \begin{pmatrix} -6 \\ 3 \\ 5 \end{pmatrix} = \begin{pmatrix} -3 \\ 3.2 \\ 3.1 \end{pmatrix}$$

- Pour $h = 2$?

$$\mathbb{E}(Y_{t+2}|\mathcal{F}_t) = (I_3 + \Phi_1)C + \Phi_1^2 Y_t = \begin{pmatrix} -1.5 \\ 2.95 \\ 2.57 \end{pmatrix}$$

Exercice : prévisions d'un VAR(2)

Exemple Avec le VAR(2) précédent, si $Y_t = (0.06, 0.03)'$, $Y_{t-1} = (0.055, 0.03)'$ et $C = (0.02, 0.03)'$ on obtient pour $h = 1$?

Exercice : prévisions d'un VAR(2)

Exemple Avec le VAR(2) précédent, si $Y_t = (0.06, 0.03)'$, $Y_{t-1} = (0.055, 0.03)'$ et $C = (0.02, 0.03)'$ on obtient pour $h = 1$?

$$\mathbb{E}(Y_{t+1}|\mathcal{F}_t) = \begin{pmatrix} 0.02 \\ 0.03 \end{pmatrix} + \begin{pmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{pmatrix} \begin{pmatrix} 0.06 \\ 0.03 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0.25 & 0 \end{pmatrix} \begin{pmatrix} 0.055 \\ 0.03 \end{pmatrix} = \begin{pmatrix} 0.053 \\ 0.08275 \end{pmatrix}$$

Exercice : prévisions d'un VAR(2)

Exemple Avec le VAR(2) précédent, si $Y_t = (0.06, 0.03)'$, $Y_{t-1} = (0.055, 0.03)'$ et $C = (0.02, 0.03)'$ on obtient pour $h = 1$?

$$\mathbb{E}(Y_{t+1}|\mathcal{F}_t) = \begin{pmatrix} 0.02 \\ 0.03 \end{pmatrix} + \begin{pmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{pmatrix} \begin{pmatrix} 0.06 \\ 0.03 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0.25 & 0 \end{pmatrix} \begin{pmatrix} 0.055 \\ 0.03 \end{pmatrix} = \begin{pmatrix} 0.053 \\ 0.08275 \end{pmatrix}$$

- Pour $h = 2$?

$$\mathbb{E}(Y_{t+1}|\mathcal{F}_t) = \begin{pmatrix} 0.02 \\ 0.03 \end{pmatrix} + \begin{pmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{pmatrix} \begin{pmatrix} 0.053 \\ 0.08275 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0.25 & 0 \end{pmatrix} \begin{pmatrix} 0.06 \\ 0.03 \end{pmatrix} = \begin{pmatrix} 0.0548 \\ 0.1076 \end{pmatrix}$$

Prédicteur linéaire optimal

- Si l'on relâche l'hypothèse que i. i. d. , trouver un prédicteur optimal est plus complexe
- Plus modestement on peut rechercher le meilleur prédicteur linéaire
- Au sens de la MSE,

$$MSE = \mathbb{E}(\epsilon_{t+h})^2 = \mathbb{E}(\epsilon_{t+h})^2$$

on retrouvera alors le résultat que le prédicteur linéaire optimal sera

$${}_t Y_{t+h} = \mathbb{E}(Y_{t+h} | \mathcal{F}_t)$$

mais cette fois sous l'hypothèse plus générale que $\epsilon_t \sim \text{WN}(0, \Sigma_\epsilon)$

- On obtient également que

$$\Sigma_\epsilon(h) = MSE(\epsilon_{t+h}) = \sum_{i=0}^{h-1} \Psi_i \Sigma_\epsilon \Psi_i' = \Sigma_\epsilon(h-1) + \Psi_{h-1} \Sigma_\epsilon \Psi_{h-1}$$

puisque

$$\epsilon_{t+h} = \sum_{i=0}^{h-1} \Psi_i \epsilon_{t+h-i}$$

MSE et inférence

- En pratique les Ψ_i sont inobservés car les Φ_i le sont aussi
- En considérant $\widehat{\Psi}_i$, un estimateur de Φ_i , on aura

$${}_t\widehat{Y}_{t+h} = \widehat{C} + \sum_{i=1}^p \widehat{\Psi}_i \widehat{Y}_{t+h-i}$$

- Les erreurs de prévisions sont alors estimées par

$$\begin{aligned}\widehat{\epsilon}_{t+h} &= Y_{t+h} - \widehat{Y}_{t+h} \\ &= (Y_{t+h} - {}_t Y_{t+h})(Y_{t+h} - \widehat{Y}_{t+h}) \\ &= \sum_{i=0}^{h-1} \Psi_i \epsilon_{t+h-i} + ({}_t Y_{t+h} - \widehat{Y}_{t+h})\end{aligned}$$

et reflètent l'erreur de prévision et celle d'estimation (souvent ignorée)

- Ces deux sources d'erreur contribueront à la variance de la MSE

$$\widehat{\Sigma}_{\widehat{\epsilon}}(h) = \text{MSE}(\widehat{\epsilon}_{t+h}) = \sum_{i=0}^{h-1} \widehat{\Psi}_i \widehat{\Sigma}_{\epsilon} \widehat{\Psi}_i'$$

⇒ Sous une hypothèse de loi pour ϵ_t , on peut calculer des CI pour $\widehat{\epsilon}_{t+h}$

Exemple : prévisions et intervalles de confiance

- Calculez $\Sigma_\epsilon(h = 1, 2)$ pour le VAR(1) précédent étant donné que

$$\Sigma_\epsilon(h) = \Sigma_\epsilon(h - 1) + \Psi_{h-1} \Sigma_\epsilon \Psi'_{h-1}$$

et sachant que

$$\Sigma_\epsilon = \begin{pmatrix} 2.25 & 0 & 0 \\ 0 & 1 & 0.5 \\ 0 & 0.5 & 0.74 \end{pmatrix}$$

Exemple : prévisions et intervalles de confiance

- Calculez $\Sigma_\epsilon(h = 1, 2)$ pour le VAR(1) précédent étant donné que

$$\Sigma_\epsilon(h) = \Sigma_\epsilon(h-1) + \Psi_{h-1} \Sigma_\epsilon \Psi_{h-1}'$$

et sachant que

$$\Sigma_\epsilon = \begin{pmatrix} 2.25 & 0 & 0 \\ 0 & 1 & 0.5 \\ 0 & 0.5 & 0.74 \end{pmatrix}$$

- On voit immédiatement que $\Sigma_\epsilon(1) = \Sigma_\epsilon$ et vient

$$\Sigma_\epsilon(2) = \Sigma_\epsilon + \Psi_1 \Sigma_\epsilon \Psi_1 = \begin{pmatrix} 2.813 & 0.113 & 0 \\ 0.113 & 1.129 & 0.632 \\ 0 & 0.632 & 0.907 \end{pmatrix}$$

- En supposant que $\epsilon_t \sim \mathcal{N}(0, \Sigma_\epsilon)$, calculez les CI à 95% de

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

où $h = \{1, 2\}$ et $k = 1, 2, 3$, les variables de notre VAR(1) trivarié

Exemple : prévisions et intervalles de confiance

- Calculez $\Sigma_\epsilon(h = 1, 2)$ pour le VAR(1) précédent étant donné que

$$\Sigma_\epsilon(h) = \Sigma_\epsilon(h-1) + \Psi_{h-1} \Sigma_\epsilon \Psi_{h-1}'$$

et sachant que

$$\Sigma_\epsilon = \begin{pmatrix} 2.25 & 0 & 0 \\ 0 & 1 & 0.5 \\ 0 & 0.5 & 0.74 \end{pmatrix}$$

- On voit immédiatement que $\Sigma_\epsilon(1) = \Sigma_\epsilon$ et vient

$$\Sigma_\epsilon(2) = \Sigma_\epsilon + \Psi_1 \Sigma_\epsilon \Psi_1 = \begin{pmatrix} 2.813 & 0.113 & 0 \\ 0.113 & 1.129 & 0.632 \\ 0 & 0.632 & 0.907 \end{pmatrix}$$

- En supposant que $\epsilon_t \sim \mathcal{N}(0, \Sigma_\epsilon)$, calculez les CI à 95% de

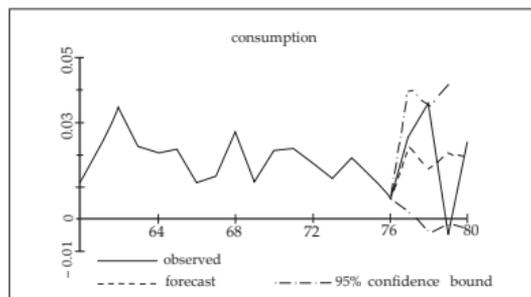
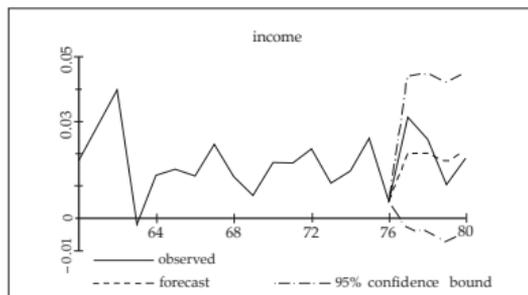
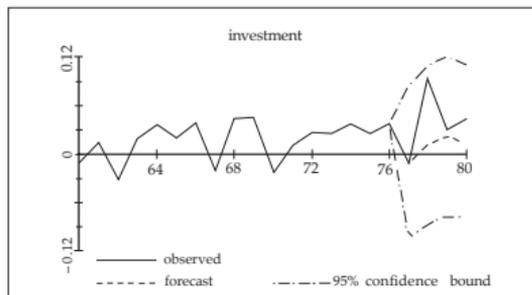
$$\frac{\epsilon_{k,t+h}}{\sigma_k(h)} \sim \mathcal{N}(0, 1)$$

où $h = \{1, 2\}$ et $k = 1, 2, 3$, les variables de notre VAR(1) trivarié

- De cette formule on déduit qu'au seuil de risque α , les CI sont

$$y_{k,t+h} \pm z_{\alpha/2} \sigma_k(h) \Rightarrow y_{1,t+1} \pm 1.96\sqrt{2.25}, \dots, y_{3,t+2} \pm 1.96\sqrt{0.907}$$

Exemples : prévisions et IC



- Dans cet exemple en petit échantillon, on voit que l'accumulation d'incertitude autour de $\hat{\Sigma}_{\epsilon}(h)$ a des conséquences

⇒ parfois les réalisations futures sortent de l'intervalle de confiance

Définition de la Causalité

- Granger a proposé un concept de causalité adapté au contexte des VAR
- L'idée de causalité est ici que l'effet ne peut précéder la cause

⇒ Si x cause z , x devrait aider à prédire z

- Soit $z_t(h|\mathcal{F}_t)$ un prédicteur optimal à l'horizon h et

$$\sigma_z(h|\mathcal{F}_t)$$

la MSE correspondante

⇒ On dira que x_t cause z_t au sens de Granger ($x_t \xrightarrow{G} z_t$) si

$$\sigma_z(h|\mathcal{F}_t) < \sigma_z(h|\mathcal{F}_t \setminus \{x_s | s \leq t\}), \quad h = 1, 2, \dots$$

- Cette inégalité dit que la MSE est plus faible (prédiction meilleure) si

$$\{x_s | s \leq t\},$$

i.e. l'information du passé et présent de x , est dans le set d'information

Définition de la Causalité en multivarié

- On peut généraliser l'inégalité précédente au cas matriciel
- Soit $Z_t(h|\mathcal{F}_t)$ un prédicteur optimal à l'horizon h et

$$\Sigma_Z(h|\mathcal{F}_t)$$

la MSE correspondante

⇒ On dira que X_t cause Z_t au sens de Granger si

$$\Sigma_Z(h|\mathcal{F}_t) < \Sigma_Z(h|\mathcal{F}_t \setminus \{X_s | s \leq t\}) \Leftrightarrow \Sigma_Z(h|\mathcal{F}_t) - \Sigma_Z(h|\mathcal{F}_t \setminus \{X_s | s \leq t\}) \leq \mathbf{0}$$

Note La matrice nulle $\mathbf{0}$ étant semi-définie positive, on doit avoir

$$\Sigma_Z(h|\mathcal{F}_t) \neq \Sigma_Z(h|\mathcal{F}_t \setminus \{X_s | s \leq t\})$$

Intuition de la causalité instantanée

- On parle de causalité instantanée si

$$X_t \xrightarrow{G} Z_t \text{ et } Z_t \xrightarrow{G} X_t$$

- Formellement, on aura causalité instantanée si

$$\Sigma_Z(1|\mathcal{F}_t \cup \{X_{t+1}\}) \neq \Sigma_Z(h|\mathcal{F}_t)$$

⇒ l'information de X_{t+1} aide à améliorer la prévision de Z_{t+1}

- Cette causalité est purement symétrique : effet équivalent pour $Z_t \xrightarrow{G} X_t$

Note La définition de la causalité sera liée à la fonction de perte, ici la MSE, puisque pour une autre fonction de perte on pourrait aboutir à un résultat différent

Limite de la causalité de Granger

- Dans la pratique, le choix de \mathcal{F}_t n'est pas simple

⇒ L'ensemble informationnel optimal ne peut être déterminé

- En général on se contentera de définir \mathcal{F}_t ainsi

$$\{X_s, Z_s | s \leq t\}$$

- Par ailleurs, on se limitera aux prédicteurs linéaires

⇒ les relations non-linéaires échapperont à ce type de causalité

Causalité de Granger dans la représentation de Wold

- Soit un VAR(p), k -varié, sous sa forme MA(∞)

$$Y_t = \mu + \sum_{i=0}^{\infty} \Psi_i \varepsilon_{t-i} = \mu + \Psi(L)\varepsilon_t, \quad \Psi_0 = I_k$$

avec $\varepsilon_t \sim \text{WN}(0, \Sigma_\varepsilon)$

- Supposons que Y_t se partage en Z_t (m -varié) et X_t ($k - m$ -varié)
- On peut alors représenter ce VAR comme

$$Y_t = \begin{pmatrix} Z_t \\ X_t \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \Psi_{11}(L) & \Psi_{12}(L) \\ \Psi_{21}(L) & \Psi_{22}(L) \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}$$

et donc le prédicteur de Z_t pour $h = 1$ sera

$$Z_t(1|\{Y_s|s \leq t\}) = (I_M, \mathbf{0})' Y_{1,t} = \mu_1 + \sum_{i=1}^{\infty} \Psi_{11,i} \varepsilon_{1,t+1-i} + \sum_{i=1}^{\infty} \Psi_{12,i} \varepsilon_{2,t+1-i}$$

- On peut alors montrer que l'absence de causalité s'exprime comme

$$Z_t(1|\{Y_s|s \leq t\}) = Z_t(1|\{Z_s|s \leq t\}) \Leftrightarrow \Psi_{12,i} = 0 \quad i = 1, 2, \dots$$

i.e. les $k - m$ variables de X_t ne causent pas Z_t

Causalité de Granger dans les VAR(p)

- Soit le VAR(p) précédent exprimé sous sa forme VAR

$$Y_t = \begin{pmatrix} Z_t \\ X_t \end{pmatrix} = \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} + \begin{pmatrix} \Phi_{11,1} & \Phi_{12,1} \\ \Phi_{21,1} & \Phi_{22,1} \end{pmatrix} \begin{pmatrix} Z_{t-1} \\ X_{t-1} \end{pmatrix} + \dots + \begin{pmatrix} \Phi_{11,p} & \Phi_{12,p} \\ \Phi_{21,p} & \Phi_{22,p} \end{pmatrix} \begin{pmatrix} Z_{t-p} \\ X_{t-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}$$

- La condition d'absence de causalité de Granger étant

$$\begin{pmatrix} \Psi_{11}(L) & 0 \\ \Psi_{21}(L) & \Psi_{22}(L) \end{pmatrix}$$

on passera sur les matrices AR par l'inverse qui donne

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

et implique donc, sans besoin de développement supplémentaire, que

$$\Phi_{12,i} = 0 \quad i = 1, 2, \dots, p$$

- Ce résultat a pour corollaire que Z_t ne causera pas X_t si

$$X_t(h|\{Y_s|s \leq t\}) = X_t(h|\{X_s|s \leq t\}) \Leftrightarrow \Phi_{21,i} = 0 \quad h = 1, 2, \dots \quad i = 1, 2, \dots, p$$

Exercice : causalité de Granger dans un VAR(1)

- Soit le VAR(1) trivarié des exemples précédents

$$\begin{pmatrix} \gamma_{1,t} \\ \gamma_{2,t} \\ \gamma_{3,t} \end{pmatrix} = C + \begin{pmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{pmatrix} \begin{pmatrix} \gamma_{1,t-1} \\ \gamma_{2,t-1} \\ \gamma_{3,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{pmatrix}$$

- Analysez les causalités au sens de Granger

Exercice : causalité de Granger dans un VAR(1)

- Soit le VAR(1) trivarié des exemples précédents

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{pmatrix} = C + \begin{pmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{pmatrix}$$

- Analysez les causalités au sens de Granger

⇒ On voit que $X_t := (y_{2,t}, y_{3,t})'$ ne cause pas $Z_t = y_{1,t}$

⇒ En revanche on voit que Z_t cause au sens de Granger X_t

- Supposons que le VAR(2) des exemples précédents porte en réalité sur

$$\begin{pmatrix} y_{2,t} \\ y_{3,t} \end{pmatrix} = A + \begin{pmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{pmatrix} \begin{pmatrix} y_{2,t-1} \\ y_{3,t-1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0.25 & 0 \end{pmatrix} \begin{pmatrix} y_{2,t-2} \\ y_{3,t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{pmatrix}$$

- Analysez les causalités au sens de Granger

Orthogonalisation des innovations

- Comment caractériser la causalité instantanée ? Partons de

$$\Sigma_\varepsilon$$

dont la semi-définie positivité permet la décomposition en

$$\Sigma_\varepsilon = PP'$$

où P est une matrice triangulaire inférieure avec $\text{diag}(P) > 0$

- Avec $\Omega_i = \Psi_i P$ et $\omega_t = P^{-1}\varepsilon_t$, il vient alors que

$$Y_t = \mu + \sum_{i=0}^{\infty} \Psi_i P P^{-1} \varepsilon_{t-i} = \mu + \sum_{i=0}^{\infty} \Omega_i \omega_{t-i}$$

et donc à la manière d'une standardisation

$$\Sigma_\omega = P^{-1} \Sigma_\varepsilon (P^{-1})' = I_k$$

⇒ On qualifie ω_t d'innovations orthogonalisées car $\Sigma_\omega = I_k$

Formalisation de la causalité instantanée

- Comme précédemment considérons Z_t et X_t avec ω_t

$$\begin{pmatrix} Z_t \\ X_t \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \Omega_{11,0} & 0 \\ \Omega_{21,0} & \Omega_{22,0} \end{pmatrix} \begin{pmatrix} \omega_{1,t} \\ \omega_{2,t} \end{pmatrix} + \dots + \begin{pmatrix} \Omega_{11,1} & \Omega_{12,1} \\ \Omega_{21,1} & \Omega_{22,1} \end{pmatrix} \begin{pmatrix} \omega_{1,t-1} \\ \omega_{2,t-1} \end{pmatrix} + \dots$$

- Les équations marginales prédictives de Z_{t+1} et X_{t+1} sont donc

$$Z_{t+1} = \mu_1 + \Omega_{11,0}\omega_{1,t+1} + \Omega_{11,1}\omega_{1,t} + \Omega_{12,1}\omega_{2,t} + \dots$$

$$\text{et } X_{t+1} = \mu_2 + \Omega_{21,0}\omega_{1,t+1} + \Omega_{22,0}\omega_{2,t+1} + \Omega_{21,1}\omega_{1,t} + \Omega_{22,1}\omega_{2,t} + \dots$$

- Pour $h = 1$ le prédicteur de X_t basé sur $\{Y_s | s \leq t\}$ et sur Z_{t+1} sera alors

$$\begin{aligned} X_t(\mathbf{1}_{\{Y_s | s \leq t\}} \cup \{Z_{t+1}\}) &= X_t(\mathbf{1}_{\{\omega_s | s \leq t\}} \cup \{\omega_{1,t+1}\}) \\ &= \Omega_{21,0}\omega_{1,t+1} + X_t(\mathbf{1}_{\{\omega_s | s \leq t\}}) \end{aligned}$$

$$\Rightarrow X_t(\mathbf{1}_{\{Y_s | s \leq t\}} \cup \{Z_{t+1}\}) = X_t(\mathbf{1}_{\{\omega_s | s \leq t\}}) \text{ uniquement si } \Omega_{21,0} = 0$$

- Comme $\Omega_i = \Psi_i P$, $\Omega_{21,0} = 0$ si Σ_ε est bloc-diagonal avec un bloc de zéros $\mathbf{0}_{(k-m) \times m}$ dans le coin gauche inférieur et $\mathbf{0}_{m \times (k-m)}$ dans le coin droit supérieur qui indique que

$$\mathbb{E}(\varepsilon_{1,t} \varepsilon'_{2,t}) = 0$$

Définition de la causalité instantanée

- Du résultat précédent on déduit qu'il n'y a pas causalité instantanée si

$$\mathbb{E}(\varepsilon_{1,t}\varepsilon'_{2,t}) = 0$$

- En d'autres termes, la causalité instantanée ne peut venir que de chocs non-orthogonaux

⇒ Comme par construction Σ_ε est non-singulière et symétrique

$$\sigma_{12} = \sigma_{21}$$

et la causalité instantanée existe forcément dans les deux sens

Note Dans ce cadre d'analyse, la causalité instantanée ne nous apprend rien sur les causes et les effets de ces relations instantanées

Exercice : causalité instantanée

- Pour rappel, dans le VAR(1) trivarié on a

$$\Sigma_{\varepsilon} = \begin{pmatrix} 2.25 & 0 & 0 \\ 0 & 1 & 0.5 \\ 0 & 0.5 & 0.74 \end{pmatrix}$$

- Que peut-on dire de la causalité instantanée dans le VAR orthogonalisé ?

Exercice : causalité instantanée

- Pour rappel, dans le VAR(1) trivarié on a

$$\Sigma_{\varepsilon} = \begin{pmatrix} 2.25 & 0 & 0 \\ 0 & 1 & 0.5 \\ 0 & 0.5 & 0.74 \end{pmatrix}$$

- Que peut-on dire de la causalité instantanée dans le VAR orthogonalisé ?
- On constate que Σ_{ε} est bloc-diagonal est donc l'absence de causalité instantanée entre $\gamma_{1,t}$ et $(\gamma_{2,t}, \gamma_{3,t})'$ (i.e. entre Z_t et X_t)
- Dans la représentation orthogonalisée cela implique que

$$\begin{aligned} \begin{pmatrix} Z_t \\ X_t \end{pmatrix} &= \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \Omega_{11,0} & \mathbf{0} \\ \mathbf{0} & \Omega_{22,0} \end{pmatrix} \begin{pmatrix} \omega_{1,t} \\ \omega_{2,t} \end{pmatrix} + \dots + \begin{pmatrix} \Omega_{11,1} & \mathbf{0} \\ \Omega_{21,1} & \Omega_{22,1} \end{pmatrix} \begin{pmatrix} \omega_{1,t-1} \\ \omega_{2,t-1} \end{pmatrix} + \dots \\ &= \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \Omega_{11}(L) & \mathbf{0} \\ \Omega_{21}(L) & \Omega_{22}(L) \end{pmatrix} \begin{pmatrix} \omega_{1,t} \\ \omega_{2,t} \end{pmatrix} \end{aligned}$$

$\Rightarrow \Omega_{12,0} = \Omega_{21,0} = \mathbf{0}$ reflète l'absence de causalité instantanée et

$$\Omega_{12}(L) = \mathbf{0}$$

reflète l'absence de causalité de Granger de X_t vers Z_t

Absence de causalité et interprétation

- La causalité de Granger est directionnelle et s'interprète donc aisément
- Mais que peut-on dire de l'absence de causalité de Granger ? Par exemple,

$$\begin{pmatrix} Z_t \\ X_t \end{pmatrix} = \begin{pmatrix} \phi_{11} & 0 \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} Z_{t-1} \\ X_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}$$

représente un VAR(1) dans lequel $X_t \xrightarrow{G} Z_t$ mais

$$\begin{pmatrix} Z_t \\ X_t \end{pmatrix} = \begin{pmatrix} 0 & -\beta \\ 0 & 0 \end{pmatrix} \begin{pmatrix} Z_t \\ X_t \end{pmatrix} + \begin{pmatrix} \varphi_{11} & \varphi_{12} \\ \varphi_{21} & \varphi_{22} \end{pmatrix} \begin{pmatrix} Z_{t-1} \\ X_{t-1} \end{pmatrix} + \begin{pmatrix} \nu_{1,t} \\ \nu_{2,t} \end{pmatrix}$$

est une représentation équivalente obtenue en multipliant le système par

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

et donc $\varphi_{11} = \phi_{11} + \phi_{21}\beta$, $\varphi_{12} = \phi_{22}\beta$, $\varphi_{21} = \phi_{21}$, $\varphi_{22} = \phi_{22}$ et $\nu_t = B\varepsilon_t$

- Via $\beta \neq 0$ on voit alors que X_t peut impacter Z_t ce qui implique que

$$X_t \xrightarrow{G} Z_t \not\Rightarrow X_t \xrightarrow{C} Z_t$$

où $X_t \xrightarrow{C} Z_t$ reflète une forme de causalité plus générale

Au delà de la causalité de Granger

- Quand l'existence de $X_t \xrightarrow{c} Z_t$ échappe à la causalité de Granger

... c'est souvent que le set d'information est incomplet

e.g. 1 une variable hors du système cause X_t et Z_t

- Dans un système bivarié (monnaie et PIB) on peut trouver que la monnaie cause le PIB
- Quand on intègre le taux d'intérêt dans le système cette causalité disparaît entre monnaie et PIB et il apparaît que le taux d'intérêt cause le PIB

e.g. 2 X_t cause Z_{t+h} pour un h non considéré dans le modèle

- Il est plus rigoureux de préciser l'horizon dans les relations étudiées

$$X_t \xrightarrow{G(h)} Z_t \text{ et } X_t \not\xrightarrow{G(h)} Z_t$$

e.g. 3 X cause non-linéairement Z

L'approche de la causalité de Geweke

- Geweke propose une mesure de dépendance plus complète que celle de Granger qui regroupe
 - la dépendance de X au passé de Z
 - la dépendance de Z au passé de X
 - la dépendance instantanée entre Z et X
- Comme précédemment on considère un VAR(p)

$$Y_t = \mu + \sum_{i=0}^{\infty} \Psi_i \varepsilon_{t-i} = \mu + \Psi(L) \varepsilon_t, \quad \Psi_0 = I_k$$

avec $\varepsilon_t \sim \text{WN}(0, \Sigma_\varepsilon)$ et on scinde Y_t en 2 groupes

$$Y_t = \begin{pmatrix} Z_t \\ X_t \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \Psi_{11}(L) & \Psi_{12}(L) \\ \Psi_{21}(L) & \Psi_{22}(L) \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}$$

- Mais en parallèle on va considérer les équations marginales

$$X_t = \Psi_{11}(L) \eta_{1,t} = \sum_{i=1}^p \phi_{i,1} X_{t-i} + \eta_{1,t}, \quad \eta_{1,t} \sim \text{WN}(0, \Sigma_{\eta_1})$$

$$Z_t = \Psi_{22}(L) \eta_{2,t} = \sum_{i=1}^p \phi_{i,2} Z_{t-i} + \eta_{2,t}, \quad \eta_{2,t} \sim \text{WN}(0, \Sigma_{\eta_2})$$

L'approche de Geweke pour la causalité de Granger

- Si $Z \xrightarrow{G} X$

- $\Psi_{21}(L) = 0$
- les équations marginales sur Z_t sont équivalentes à Y_t
- $\eta_{2,t} = \varepsilon_{2,t}$, $\Sigma_{\eta_2} = \Sigma_{\varepsilon_2}$ et $\det \Sigma_{\eta_2} = \det \Sigma_{\varepsilon_2}$

- Si $X \xrightarrow{G} Z$

- $\Psi_{12}(L) = 0$
- les équations marginales sur X_t sont équivalentes à Y_t
- $\eta_{1,t} = \varepsilon_{1,t}$, $\Sigma_{\eta_1} = \Sigma_{\varepsilon_1}$ et $\det \Sigma_{\eta_1} = \det \Sigma_{\varepsilon_1}$

- Si $Z \xrightarrow{G} X$

- $\Psi_{21}(L) \neq 0$
- les équations marginales sur Z_t sont différentes de Y_t
- $\eta_{2,t} = \varepsilon_{2,t}$ et $\det \Sigma_{\eta_2} > \det \Sigma_{\varepsilon_2}$

- Si $X \xrightarrow{G} Z$

- $\Psi_{12}(L) \neq 0$
- les équations marginales sur X_t sont équivalentes à Y_t
- $\eta_{1,t} = \varepsilon_{1,t}$ et $\det \Sigma_{\eta_1} > \det \Sigma_{\varepsilon_1}$

L'approche de Geweke pour la causalité instantanée

- On a vu précédemment que de manière générale, pour Y_t ,

$$\Sigma_Y = \begin{pmatrix} \Sigma_{\varepsilon_1} & \text{Cov}(\varepsilon_{1,t}, \varepsilon_{2,t}) \\ \text{Cov}(\varepsilon_{1,t}, \varepsilon_{2,t}) & \Sigma_{\varepsilon_2} \end{pmatrix}$$

et qu'en l'absence de causalité instantanée, i.e. $X \overset{C}{\leftrightarrow} Z$

$$\Sigma_Y = \begin{pmatrix} \Sigma_{\varepsilon_1} & \mathbf{0} \\ \mathbf{0} & \Sigma_{\varepsilon_2} \end{pmatrix}$$

i.e. Σ_Y est bloc-diagonal

⇒ On voit alors l'implication pour le déterminant de Σ_Y

$$\det \Sigma_Y = \det(\Sigma_{\varepsilon_1}) \det(\Sigma_{\varepsilon_2})$$

- A l'inverse, en présence de causalité instantanée, i.e. $X \overset{C}{\leftarrow} Z$

$$\det \Sigma_Y \leq \det(\Sigma_{\varepsilon_1}) \det(\Sigma_{\varepsilon_2})$$

Les mesures de dépendance de Geweke

- Geweke identifie donc des inégalités entre les déterminants pour les deux causalités de Granger et la causalité instantanée
- ⇒ plus l'inégalité est forte, plus la causalité est prononcée
- Geweke propose donc les 4 mesures suivantes :

- Mesure de dépendance de Z au passé de X

$$X \xrightarrow{c} Z := \log \frac{\det \Sigma_{\eta_1}}{\det \Sigma_{\varepsilon_1}}$$

- Mesure de dépendance de Z au passé de X

$$Z \xrightarrow{c} X := \log \frac{\det \Sigma_{\eta_2}}{\det \Sigma_{\varepsilon_2}}$$

- Mesure de dépendance instantanée entre Z et X

$$X \longleftrightarrow Z := \log \frac{\det \Sigma_{\varepsilon_1} \det \Sigma_{\varepsilon_2}}{\det \Sigma_Y}$$

- Mesure de dépendance globale entre Z et X

$$\begin{aligned} C_{Z,X} &= C_{X,Z} = X \xrightarrow{c} Z + Z \xrightarrow{c} X + X \longleftrightarrow Z \\ &:= \log \frac{\det \Sigma_{\eta_1} \det \Sigma_{\eta_2}}{\det \Sigma_Y} \end{aligned}$$

Interprétation des mesures de dépendance de Geweke

- La mesure de dépendance globale revient vers les marginales
- En effet, si aucune causalité n'est présente on a

$$\det \Sigma_{\eta_1} \det \Sigma_{\eta_2} = \det \Sigma_Y$$

car modéliser de manière indépendante ou jointe X et Z revient au même

- A l'inverse, si au moins un type de causalité survient, $C_{z,x} > 0$

Note Si Y_t suit un VAR(p), qu'en est-il des équations marginales ?

⇒ on pourrait penser qu'elles suivent un sous-VAR($p' \neq p$)

⇒ heureusement, Gouriéroux, Monfort et Renault montrent qu'un VAR(p) convient aussi pour les marginales

Fonctions de réponses impulsionnelles

- Dans les travaux appliqués il peut être intéressant de connaître
... la réponse d'une variable à une impulsion d'une autre variable
⇒ c'est le rôle des fonctions de réponses impulsionnelles
- Soit un système : investissement ($y_{t,1}$), PIB ($y_{t,2}$), consommation ($y_{t,3}$)

⇒ on s'intéresse à l'effet d'un choc d'investissement

- Pour isoler cet effet on suppose tout d'abord que

$$(y_{1,t}, y_{2,t}, y_{3,t})' = \mu, \quad t \leq 0$$

sont à leur valeur moyenne avant le début des temps

- Puis on augmente l'investissement d'une unité en $t = 0$: $\varepsilon_{0,1} = 1$
- Enfin on trace la dynamique du système en $t = 1, 2, \dots$ ceteris paribus

$$(\varepsilon_{2,0}, \varepsilon_{3,0})' = 0, \varepsilon_1 = 0, \varepsilon_2 = 0, \dots$$

Calcul des d'IRF

- Considérons encore le VAR(1) trivarié avec $C = 0$

$$Y_t = C + \begin{pmatrix} 0.5 & 0 & 0 \\ 0.1 & 0.1 & 0.3 \\ 0 & 0.2 & 0.3 \end{pmatrix} Y_{t-1} + \varepsilon_t$$

- On veut à présent suivre un choc unitaire en $t = 0$

$$Y_0 = \begin{pmatrix} \gamma_{1,0} \\ \gamma_{2,0} \\ \gamma_{3,0} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1,0} \\ \varepsilon_{2,0} \\ \varepsilon_{3,0} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$Y_1 = \begin{pmatrix} \gamma_{1,1} \\ \gamma_{2,1} \\ \gamma_{3,1} \end{pmatrix} = \Phi_1 Y_0 = \begin{pmatrix} 0.5 \\ 0.1 \\ 0 \end{pmatrix}$$

$$Y_2 = \begin{pmatrix} \gamma_{1,2} \\ \gamma_{2,2} \\ \gamma_{3,2} \end{pmatrix} = \Phi_1 Y_1 = \Phi_1^2 Y_0 = \begin{pmatrix} 0.25 \\ 0.01 \\ 0 \end{pmatrix}$$

- En itérant jusqu'en i on peut montrer que $Y_i = (\gamma_{1,i}, \gamma_{2,i}, \gamma_{3,i})'$ sera juste $(\Phi_1)_{.,1}^i$

Note $(\Phi_1)_{.,1}$ désigne toutes les lignes de la colonne 1 de Φ_1

Calcul des IRF

- Plus généralement, pour un choc sur la variable j , après i périodes on aura

$$(\Phi_1)^i_{.j}$$

où $(\Phi_1)^i_{.j}$ désigne toutes les lignes de la colonne j de Φ_1

Note Les IRF sont aussi appelées multiplicateurs dynamiques

- Pour un VAR, on peut utiliser la forme VMA(∞) pour le calcul d'IRF

⇒ Pour $Y_t = \mu + \varepsilon_t + \Psi_1\varepsilon_{t-1} + \Psi_2\varepsilon_{t-2} + \dots$ on aura

$$\frac{\partial Y_t}{\partial \varepsilon_{t-h}} = \Psi_h$$

- Cette formulation plus générale est basée sur les dérivées partielles
- En l'état elle pose problème car, e.g. pour $\varepsilon_{2,t}$, elle ne neutralise pas

$$(\varepsilon_{1,t+h}, \varepsilon_{3,t+h})', \quad h > 0$$

qui pourrait être non-nul si Σ_ε n'est pas diagonal

IRF orthogonalisés

- Souvenons-nous de la méthode d'orthogonalisation du S42
- Elle repose sur le fait que si Σ_ε est non-singulière et définie positive

$$\exists P \text{ tel que } \Sigma_\varepsilon = PP'$$

- Cette décomposition (de Choleski), des innovations permet de définir

$$\omega_t = P^{-1}\varepsilon_t$$

les innovations orthogonalisées de Y_t puisque $\Sigma_\omega = P^{-1}\omega(P^{-1})' = I_k$

⇒ En les utilisant dans la forme VMA(∞) on a alors

$$Y_t = \mu + \Psi(L)\varepsilon_t = \mu + \Psi(L)P\omega_t = \mu + \Xi(L)\omega_t$$

et à partir de $\Xi(L)$ on obtient les coefficients des IFR orthogonalisées

$$\frac{\partial Y_{i,t}}{\partial \omega_{j,t-h}} = (\Xi_h)_{.j}$$

également obtenus par $\Xi_i = \Psi_i P$

Impacte de la décomposition de Choleski

- La méthode d'orthogonalisation n'est pas neutre sur le système
- Étudions les conséquences de la décomposition de Choleski

$$\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \vdots \\ \varepsilon_{k,t} \end{pmatrix} = \begin{pmatrix} p_{11} & 0 & \dots & \dots & 0 \\ p_{21} & p_{22} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 & \vdots \\ p_{k1} & p_{k2} & \dots & & p_{kk} \end{pmatrix} = \begin{pmatrix} \omega_{1,t} \\ \omega_{2,t} \\ \vdots \\ \omega_{k,t} \end{pmatrix}$$

ce qui donne

$$\varepsilon_{1,t} = p_{11}\omega_{1,t}$$

$$\varepsilon_{2,t} = p_{21}\omega_{1,t} + p_{22}\omega_{2,t}$$

$$\varepsilon_{k,t} = p_{k1}\omega_{1,t} + p_{k2}\omega_{2,t} + \dots + p_{kk}\omega_{k,t}$$

- On voit qu'un choc dans $\omega_{1,t}$ peut impacter tous les $\varepsilon_{j,t}, j = 1, 2, \dots, k$
 - Mais $\omega_{i,t}$ pourra seulement affecter les $\varepsilon_{j,t}, j = i + 1, \dots, k$
- ⇒ $\omega_{i,t}$ n'entraînera aucune réponse contemporaine dans $y_{l,t}, l < i$
- ⇒ Avec Choleski, l'ordre des variables dans le système compte

Ordonnancement des variables du VAR

- Comment choisir l'ordre des variables dans le VAR ?
- Au regard de la décomposition de Choleski, on doit placer en premier la variable la plus susceptible d'impacter les autres contemporainement
etc ...

dernier la variable la moins susceptible d'avoir un impact immédiat sur le système

e.g. on cherche à modéliser les ventes d'une firme à l'aide de la conjoncture

— si on dispose de ces 3 indicateurs : international, national et régional

... l'ordre pourrait être le suivant

$$\begin{pmatrix} \textit{international}_t \\ \textit{national}_t \\ \textit{rgional}_t \\ \textit{ventes}_t \end{pmatrix}$$

Interprétation des chocs orthogonalisés

- Comment interpréter ces innovations transformées ?
- Considérons le cas d'un VAR trivarié

$$\varepsilon_{1,t} = p_{11}\omega_{1,t}$$

$$\varepsilon_{2,t} = p_{21}\omega_{1,t} + p_{22}\omega_{2,t}$$

$$\varepsilon_{3,t} = p_{31}\omega_{1,t} + p_{32}\omega_{2,t} + p_{33}\omega_{3,t}$$

- Comme $\Sigma_\omega = I_k$, $\omega_{1,t}$ est de variance unitaire, ce qui implique

$$p_{11} = \sigma_{11}$$

pour préserver la variance de $\varepsilon_{1,t}$

⇒ un choc unitaire sur $\omega_{1,t}$ équivaut à un choc d'un écart-type sur $\varepsilon_{1,t}$

- Pour $\varepsilon_{2,t}$ on voit que $\varepsilon_{2,t} = (p_{21}/p_{11})\varepsilon_{1,t} + p_{22}\omega_{2,t}$ et comme $\omega_{1,t} \perp \omega_{2,t}$,

$p_{22}\omega_{2,t}$ est le résidu de la projection de $\varepsilon_{2,t}$ sur $\varepsilon_{1,t}$

i.e. la partie non expliquée linéairement par $\varepsilon_{1,t}$

⇒ $\omega_{j,t}$ est la partie de $\varepsilon_{j,t}$ non expliquée linéairement par les $\varepsilon_{i,t}$ où $i < j$

⇒ Pour $\varepsilon_{j,t}$, l'amplitude du choc sera de l'écart-type de $p_{jj}\omega_{j,t}$, donc p_{jj}

Exercice : IRF et chocs orthogonalisés

- Pour le VAR(1) trivarié, la décomposition de Choleski de Σ_ε donne

$$P = \begin{pmatrix} 1.5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0.5 & 0.7 \end{pmatrix}$$

- Calculez Ξ_0, Ξ_1, Ξ_2 à partir de $\Psi_i = \Phi^i$, car $Y_t \sim \text{VAR}(1)$

Exercice : IRF et chocs orthogonalisés

- Pour le VAR(1) trivarié, la décomposition de Choleski de Σ_ε donne

$$P = \begin{pmatrix} 1.5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0.5 & 0.7 \end{pmatrix}$$

- Calculez Ξ_0, Ξ_1, Ξ_2 à partir de $\Psi_i = \Phi^i$, car $Y_t \sim \text{VAR}(1)$
- On obtient les résultats suivants

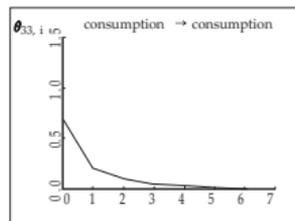
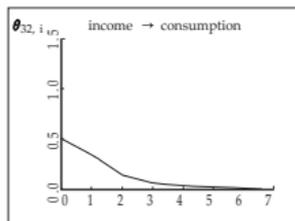
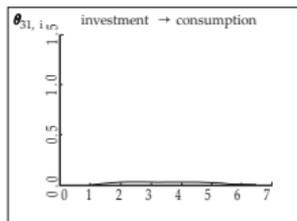
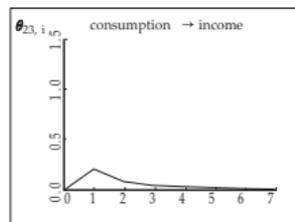
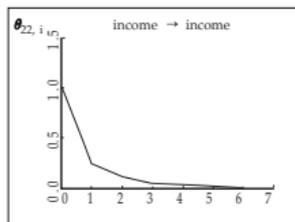
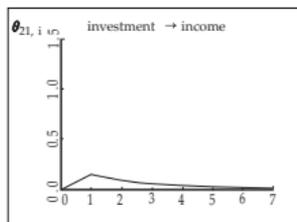
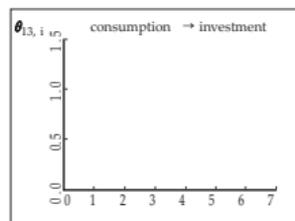
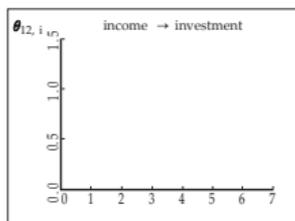
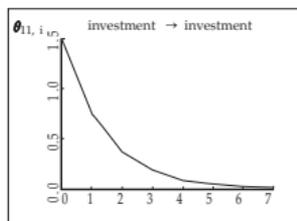
$$\Xi_0 = \begin{pmatrix} 1.5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0.5 & 0.7 \end{pmatrix}$$

$$\Xi_1 = \begin{pmatrix} 0.75 & 0 & 0 \\ 0.15 & 0.25 & 0.21 \\ 0 & 0.35 & 0.21 \end{pmatrix}$$

$$\Xi_2 = \begin{pmatrix} 0.375 & 0 & 0 \\ 0.09 & 0.130 & 0.084 \\ 0.030 & 0.055 & 0.105 \end{pmatrix}$$

Exemple : IRF et chocs orthogonalisés

- En dénotant : investissement ($y_{t,1}$), PIB ($y_{t,2}$), consommation ($y_{t,3}$)



Les multiplicateurs de long terme

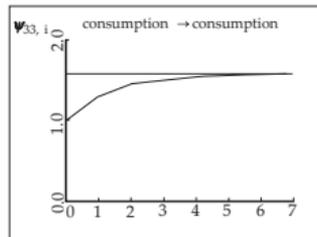
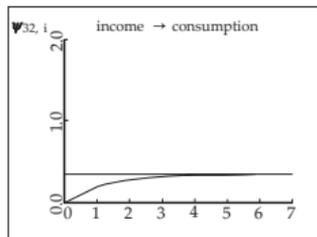
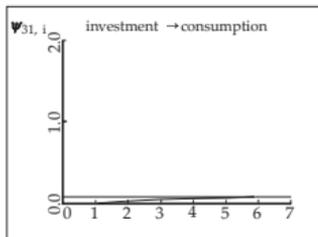
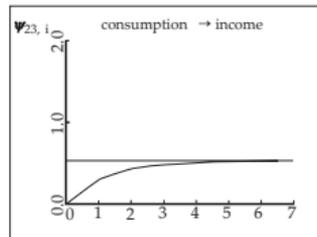
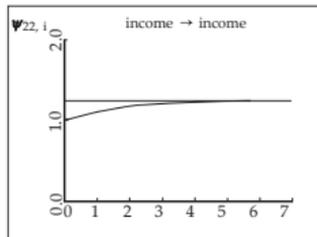
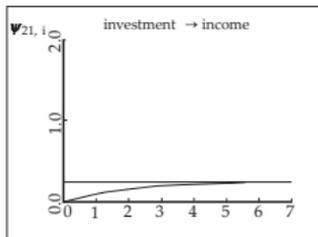
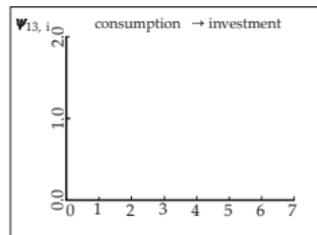
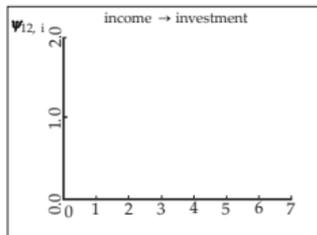
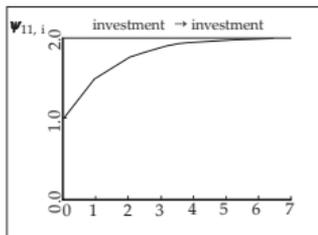
- L'effet total accumulé pour toutes périodes futures est donné par

$$\Psi_{\infty} := \sum_{i=0}^{\infty} \Psi_i$$

- Comme on sait que $\Psi(L) = \Phi(L)^{-1}$ on obtient aisément

$$\Psi_{\infty} = \Psi(1) = (I_k - \Phi_1 - \dots - \Phi_p)^{-1}$$

Exemple : multiplicateurs de long terme et chocs orthogonalisés



Décomposition de variance des erreurs de prévisions

- De la décomposition Σ_ε on peut construire un autre outil d'analyse
- ⇒ la décomposition variance des erreurs de prévisions
- Bien que proche de la causalité de Granger, cet outil est différent
- ⇒ il n'est pas déterminé uniquement car dépend de la décomposition
- Repartons de $Y_t = \mu + \sum_{i=0}^{\infty} \Xi_i \omega_{t-i}$ et considérons

$$\epsilon_{t+h} = Y_{t+h} - {}_t Y_{t+h} = \sum_{i=0}^{h-1} \Xi_i \omega_{t+h-i}$$

- Pour une variable $y_{j,t}$ parmi les k variables du système, on aura

$$\begin{aligned} \epsilon_{j,t+h} &= \sum_{i=0}^{h-1} (\xi_{j1,i} \omega_{1,t+h-i} + \dots + \xi_{jk,i} \omega_{k,t+h-i}) \\ &\equiv \sum_{l=0}^k (\xi_{jl,0} \omega_{l,t+h} + \dots + \xi_{jk,h-1} \omega_{k,t+1}) \end{aligned}$$

Décomposition de variance des erreurs de prévisions

- Comme les $\omega_{l,t}$ sont non-corrélés et de variance unitaire, on a

$$MSE({}_t Y_{t+h}) = \mathbb{E}(\epsilon_{j,t+h}^2) = \sum_{l=1}^k (\xi_{jl,0}^2 + \dots + \xi_{jk,h-1}^2) = \sum_{i=0}^{h-1} \sum_{l=1}^k \xi_{jl,i}^2$$

- En notant ι_j la j -ième colonne de I_k on peut réécrire

$$\xi_{jl,0}^2 + \dots + \xi_{jk,h-1}^2 = \sum_{i=0}^{h-1} (\iota_j' \Xi_i \iota_l)^2$$

qui sera la contribution de l'innovation de la variable l à la variance de l'erreur de prévision à la période h de la variable j

- De cette formulation on peut construire

$$\zeta_{jl,h} = \frac{\sum_{i=0}^{h-1} (\iota_j' \Xi_i \iota_l)^2}{\sum_{i=0}^{h-1} \sum_{l=1}^k \xi_{jl,i}^2}$$

la proportion de la variance de l'erreur de prévision à la période h de la variable j attribuable au choc orthogonal $\omega_{l,t}$

Note Le dénominateur s'obtient aussi via $\mathbb{E}(\epsilon_{j,t+h}^2) = \sum_{i=0}^{h-1} \Xi_i \Xi_i' = \sum_{i=0}^{h-1} \Psi_i \Sigma_\epsilon \Psi_i'$

Exemple : décomposition de variance des erreurs de prévisions

forecast error in	forecast horizon h	proportions of forecast error variance h periods ahead accounted for by innovations in		
		investment	income	consumption
investment	1	1	0	0
	2	1	0	0
	3	1	0	0
	4	1	0	0
	5	1	0	0
	10	1	0	0
	∞	1	0	0
income	1	0	1	0
	2	.020	.941	.039
	3	.026	.930	.044
	4	.029	.926	.045
	5	.030	.925	.045
	10	.030	.925	.045
	∞	.030	.925	.045
consumption	1	0	.338	.662
	2	0	.411	.589
	3	.001	.421	.578
	4	.002	.423	.576
	5	.002	.423	.575
	10	.002	.423	.575
	∞	.002	.423	.575

Interprétation de la variance des erreurs de prévisions

e.g. 1 On voit que 66% de la variance des erreurs de prévisions à 1 période de la consommation est imputable à ses propres innovations

... et 34% de la variance des erreurs de prévisions à 1 période est attribuable aux innovations du PIB

e.g. 2 Toujours pour la consommation, pour les prévisions de longues, $h = \infty$

... les valeurs se rapprochent avec 42% de la variance des erreurs de prévisions attribuables aux innovations du PIB versus 57%

Note Un parallèle peut être fait avec la causalité de Granger

- Considérons de nouveau 2 sous-groupes $Y_t = (Z_t, X_t)'$
- En l'absence de causalité de Granger et si Σ_ε est bloc-diagonale

$$\zeta_{jl,h} = 0$$

sera observée pour toutes les interactions entre Z_t et X_t

- Pour autant, $\zeta_{jl,h} \neq 0$ dépend de la forme de la décomposition et n'est donc pas unique

⇒ $\zeta_{jl,h} \neq 0$ ne saurait représenter une forme de causalité

Notations

- Jusqu'à présent on a considéré les modèles comme connus
- Bien entendu en pratique les $\Psi(L)$ doivent être estimés
- Avant de décrire les diverses méthodes d'estimation clarifions les notations

— $\mathbf{Y} = (y_{1,.}, \dots, y_{k,.})$ est de dimension $k \times n$ où

$$y_{j,.} = (y_{j,1}, \dots, y_{j,n})'$$

représente les n réalisations de la variable aléatoire

— $\mathbf{X}_t = (1, Y_t, \dots, Y_{t-p+1})'$ est de dimension $(kp + 1) \times 1$

— $\mathbf{X} = (\mathbf{X}_0, \dots, \mathbf{X}_{n-1})$ est de dimension $(kp + 1) \times n$

— $\mathbf{B} = (\mathbf{C}, \Phi_1, \dots, \Phi_p)'$ est de dimension $k \times (kp + 1)$

— $\boldsymbol{\varepsilon} = (\varepsilon_t, \dots, \varepsilon_n)'$ est de dimension $k \times n$

— $\mathbf{y} = \text{vec}(Y)$ est $kn \times 1$

— $\boldsymbol{\beta} = \text{vec}(\mathbf{B})$ est $(k^2p + k) \times 1$

— $\mathbf{e} = \text{vec}(\boldsymbol{\varepsilon})$ est $(kn \times 1)$

Moindres carrés multivariés (LS)

- A l'aide des notations précédentes on voit que le VAR se réécrit

$$\mathbf{Y} = \mathbf{B}\mathbf{X} + \boldsymbol{\varepsilon}$$

ou de manières équivalentes

$$\begin{aligned}\text{vec}(\mathbf{Y}) &= \text{vec}(\mathbf{B}\mathbf{X}) + \text{vec}(\boldsymbol{\varepsilon}) \\ &= (\mathbf{X}' \otimes \mathbf{I}_k) \text{vec}(\mathbf{B}) + \text{vec}(\boldsymbol{\varepsilon}) \\ \boldsymbol{\psi} &= (\mathbf{X}' \otimes \mathbf{I}_k)\boldsymbol{\beta} + \mathbf{e}\end{aligned}$$

avec $\Sigma_{\boldsymbol{\varepsilon}} = \mathbf{I}_n \otimes \Sigma_{\boldsymbol{\varepsilon}}$

- Comme pour les LS en univarié, on cherche à minimiser

$$\begin{aligned}\mathbf{e}'\Sigma_{\boldsymbol{\varepsilon}}^{-1}\mathbf{e} &= \mathbf{e}'(\mathbf{I}_n \otimes \Sigma_{\boldsymbol{\varepsilon}}^{-1})\mathbf{e} \\ &= (\boldsymbol{\psi} - (\mathbf{X}' \otimes \mathbf{I}_k)\boldsymbol{\beta})'(\mathbf{I}_n \otimes \Sigma_{\boldsymbol{\varepsilon}}^{-1})(\boldsymbol{\psi} - (\mathbf{X}' \otimes \mathbf{I}_k)\boldsymbol{\beta}) \\ &= \text{vec}(\mathbf{Y} - \mathbf{B}\mathbf{X})'(\mathbf{I}_n \otimes \Sigma_{\boldsymbol{\varepsilon}}^{-1}) \text{vec}(\mathbf{Y} - \mathbf{B}\mathbf{X}) \\ &= \text{tr} \left((\mathbf{Y} - \mathbf{B}\mathbf{X})'\Sigma_{\boldsymbol{\varepsilon}}^{-1}(\mathbf{Y} - \mathbf{B}\mathbf{X}) \right)\end{aligned}$$

⇒ Un peu d'algèbre matricielle (non demandée) nous donne

$$\hat{\boldsymbol{\beta}}_{LS} = ((\mathbf{X}\mathbf{X}')^{-1}\mathbf{X} \otimes \mathbf{I}_k)\boldsymbol{\psi}$$

Moindres carrés multivariés (LS) et OLS

- Par la suite on souhaitera parfois réécrire l'estimateur $\hat{\beta}_{LS}$ ainsi

$$\hat{\beta}_{LS} = \beta + ((\mathbf{XX}')^{-1}\mathbf{X} \otimes I_k)\mathbf{e}$$

ou encore ainsi

$$\text{vec}(\hat{\mathbf{B}}) = \text{vec}(\mathbf{YX}'(\mathbf{XX}')^{-1}) = \hat{\beta}_{LS}$$

ce qui montre que $\hat{\mathbf{B}} = \mathbf{YX}'(\mathbf{XX}')^{-1} = \mathbf{B} + \epsilon\mathbf{X}'(\mathbf{XX}')^{-1}$

Note Cette autre écriture montre le lien avec les OLS sur chaque $y_{i,t}$ du VAR

$$\hat{\beta}'_{LS} = \text{vec}(\hat{\mathbf{B}}') = (I_k \otimes (\mathbf{XX}')^{-1}\mathbf{X}) \text{vec}(\mathbf{Y}')$$

puisque $\text{vec}(\mathbf{Y}') = (y_{1,\cdot}, \dots, y_{k,\cdot})'$

- Notons que nous sommes pourtant partis d'une équation incluant

$$\Sigma_{\epsilon}^{-1}$$

et donc de type GLS, pour aboutir à un estimateur de type OLS

Estimation SURE à la Zellner

- Cette équivalence entre GLS et OLS est liée au cadre multivarié
- Plus précisément les k équations possèdent les mêmes régresseurs

⇒ Ce résultat, mis en évidence par Zellner (1962) n'est pas unique

- L'analyse de Zellner porte sur les Seemingly Unrelated Regressions

SURE Équations simultanées linéaires apparemment indépendantes

- En générale l'estimation SURE est de type GLS et sera de la forme

$$\begin{aligned}\hat{\beta}_{SURE} &= (\mathbf{X}'\Sigma_{\varepsilon}^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma_{\varepsilon}^{-1}\boldsymbol{\gamma} \\ &= (\mathbf{X}'(\Sigma_{\varepsilon}^{-1} \otimes I_n)\mathbf{X})^{-1}\mathbf{X}'(\Sigma_{\varepsilon}^{-1} \otimes I_n)\boldsymbol{\gamma}\end{aligned}$$

... sauf dans deux situations où il se confond avec les OLS

- Lorsque les équations sont réellement indépendantes : Σ_{ε} est diagonale
- Lorsque les équations possèdent toutes les mêmes régresseurs

Note Pour divers raisons, des contraintes peuvent être imposées dans les VAR (near-VAR) et l'estimation SURE peut en tenir compte (pas les OLS)

Propriétés limites des LS

Univarié En univarié, les propriétés en petit échantillon sont connues

Multivarié Même si la solution est une forme fermée, c'est plus complexe

⇒ On retiendra donc les propriétés asymptotiques

- Sous certaines hypothèses de régularité, si Y_t suit un VAR(p) stable,

$$\widehat{B} \xrightarrow{p} B$$

$$\sqrt{n}(\widehat{\beta}_{LS} - \beta) = \sqrt{n} \text{vec}(\widehat{B} - B) \xrightarrow{d} \mathcal{N}(0, \Gamma^{-1} \otimes \Sigma_\varepsilon)$$

avec $\mathbf{X}\mathbf{X}'/n \xrightarrow{p} \Gamma$ quand $n \rightarrow \infty$

- La construction de tests inférentiels requiert d'estimer $\Gamma^{-1} \otimes \Sigma_\varepsilon$

⇒ Les candidats naturels sont $\widehat{\Gamma} = \mathbf{X}\mathbf{X}'/n$ et

$$\widetilde{\Sigma}_\varepsilon = n^{-1} \sum_{t=1}^n \widehat{\varepsilon}_t \widehat{\varepsilon}_t' = n^{-1} \mathbf{Y}'(I_n - \mathbf{X}'(\mathbf{X}\mathbf{X}')^{-1}\mathbf{X})\mathbf{Y}$$

- En présence de régresseurs non stochastique on préférera l'ajustement

$$\widehat{\Sigma}_\varepsilon = n/(n - kp - 1) \widetilde{\Sigma}_\varepsilon$$

Estimateur LS et tests inférentiels

- La théorie limite nous autorise à implémenter les tests usuels

e.g. Pour un certain paramètre estimé, le t -test associé sera de la forme

$$t_{\beta_i} = \frac{(\hat{\beta}_i - \beta_i)}{\hat{s}_i}$$

où $\hat{\beta}_i$ et \hat{s}_i sont les i -ième éléments de $\hat{\beta}_{LS}$ et

$$(\mathbf{X}\mathbf{X}')^{-1} \otimes \hat{\Sigma}_\varepsilon$$

respectivement

- Pour n petit, $\mathcal{L}(t_{\beta_i})$ a tendance à présenter des queues plus épaisses qu'une normale

⇒ La distribution de Student est plus indiquée mais avec quel $d.f.$?

- On peut opter pour 2 choix : $d.f. = kn - k^2p - k$ ou $n - kp - 1$

Exemple : estimation d'un VAR(2) trivarié

- Considérons des données trimestrielles allemandes entre 1960-1978

⇒ investissement ($\tilde{y}_{t,1}$), PIB ($\tilde{y}_{t,2}$), consommation ($\tilde{y}_{t,3}$) et $n = 73$

- Pour s'assurer de la stationnarité on travaillera avec $Y_t = \Delta(\log \tilde{Y}_t)$
- On estime un VAR(p) pour $p = 2$ sélectionné arbitrairement

$$\hat{B} = (\hat{C}, \hat{\Phi}_1, \hat{\Phi}_2) = \begin{pmatrix} -0.017 & -0.320 & 0.146 & 0.961 & -0.161 & 0.115 & 0.934 \\ 0.016 & 0.044 & -0.153 & 0.289 & 0.050 & 0.019 & -0.010 \\ 0.013 & -0.002 & 0.225 & -0.264 & 0.034 & 0.355 & -0.022 \end{pmatrix}$$

- L'analyse des racines du polynôme ($I_3 - \hat{\Phi}_1 z - \hat{\Phi}_2 z^2$) conduit à

$$z_1 = 1.753, \quad z_2 = -2.694, \quad z_{3/4} = -0.320 \pm 2.008i, \quad z_{5/6} = -1.285 \pm 1.280i$$

ce qui permet de conclure que le modèle est stable

- Par ailleurs, on obtient également

$$\hat{\Sigma}_\varepsilon = \begin{pmatrix} 21.30 & 0.72 & 1.23 \\ 0.72 & 1.37 & 0.61 \\ 1.23 & 0.61 & 0.89 \end{pmatrix} \times 10^{-4}$$

Exemple : inférence pour d'un VAR(2) trivarié

- A partir de $(\mathbf{XX}')^{-1} \otimes \widehat{\Sigma}_\varepsilon$ on obtient les t -stats suivantes

$$\begin{pmatrix} -0.97 & -2.55 & 0.27 & 1.45 & -1.29 & 0.21 & 1.41 \\ 3.60 & 1.38 & -1.10 & 1.71 & 1.58 & 0.14 & -0.06 \\ 3.67 & -0.09 & 2.01 & -1.94 & 1.33 & 3.24 & -0.16 \end{pmatrix}$$

- Considérons les 2 choix de $d.f.$: $kn - k^2p - k = 198$ et $n - kp - 1 = 66$

⇒ Au seuil de risque de 5% les valeurs critiques sont 1.972 et 1.997 ⇒ ± 2

⇒ C'est donc également proche d'une normale : 1.96 ⇒ ± 2

- On constate que bien peu de coefficients sont significatifs

... cela peut suggérer une sur-paramétrisation du VAR

⇒ Nous verrons à la section suivante les tests de sélection

Densité Gaussienne et VAR(p)

- Sous une hypothèse distributionnelle l'estimation par MLE est possible
- Nous considérons ici le cas Gaussien mais sans perte de généralité

$$\mathbf{e} \sim \mathcal{N}(\mathbf{0}, I_n \otimes \Sigma_\varepsilon)$$

- La fonction de densité multivariée est donc donnée par

$$f_{\mathbf{e}}(\mathbf{e}) = \frac{1}{(2\pi)^{kn/2}} \det(I_n \otimes \Sigma_\varepsilon)^{-1/2} \exp\left(-\frac{1}{2}\mathbf{e}'(I_n \otimes \Sigma_\varepsilon^{-1})\mathbf{e}\right)$$

avec sous l'hypothèse que \mathbf{y} suit un VAR(p), $\mathbf{e} =$

$$\begin{pmatrix} I_k & 0 & \cdots & 0 & \cdots & \cdots & 0 \\ -\Phi_1 & I_k & \cdots & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \cdots & \cdots & \vdots \\ -\Phi_p & -\Phi_{p-1} & \cdots & I_k & \cdots & \cdots & 0 \\ 0 & -\Phi_p & \cdots & \cdots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \cdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & -\Phi_p & \cdots & \cdots & I_k \end{pmatrix} (\mathbf{y} - \tilde{\boldsymbol{\mu}}) + \begin{pmatrix} -\Phi_1 & -\Phi_2 & \cdots & -\Phi_p \\ -\Phi_2 & -\Phi_3 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\Phi_p & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} (\mathbf{X}_0 - \boldsymbol{\mu})$$

où $\mathbf{X}_0 = (Y'_0, \dots, Y'_{-p+1})'$, $\tilde{\boldsymbol{\mu}} = (\mu'_1, \dots, \mu'_{kn})'$ et $\boldsymbol{\mu} = (\mu'_1, \dots, \mu'_{kp})'$

Vraisemblance Gaussienne d'un VAR(p)

- En introduisant la notation $\boldsymbol{\varphi} = \text{vec}(\Phi_1, \dots, \Phi_p)$ on peut montrer que

$$\mathbf{e} = \mathbf{y} - \tilde{\boldsymbol{\mu}} - (\tilde{\mathbf{X}}' \otimes I_k)\boldsymbol{\varphi}$$

où $\tilde{\mathbf{X}}$ est équivalent à \mathbf{X} mais centré sur $\boldsymbol{\mu}$

- En découle la densité Gaussienne de \mathbf{y}

$$f_{\mathbf{y}}(\mathbf{y}) = \frac{1}{(2\pi)^{kn/2}} \det(I_n \otimes \Sigma_{\varepsilon})^{-1/2} \\ \times \exp\left(-\frac{1}{2}(\mathbf{y} - \tilde{\boldsymbol{\mu}} - (\tilde{\mathbf{X}}' \otimes I_k)\boldsymbol{\varphi})'(I_n \otimes \Sigma_{\varepsilon}^{-1})(\mathbf{y} - \tilde{\boldsymbol{\mu}} - (\tilde{\mathbf{X}}' \otimes I_k)\boldsymbol{\varphi})\right)$$

- On obtient alors la fonction de log-vraisemblance $\ell(\boldsymbol{\mu}, \boldsymbol{\varphi}, \Sigma_{\varepsilon})$ comme

$$\ln L(\boldsymbol{\mu}, \boldsymbol{\varphi}, \Sigma_{\varepsilon}) = -\frac{kn}{2} \ln(2\pi) - \frac{n}{2} \ln \det(\Sigma_{\varepsilon}^{-1}) - \frac{1}{2} \text{tr} \left((\tilde{\mathbf{Y}} - \Phi \tilde{\mathbf{X}})' \Sigma_{\varepsilon}^{-1} (\tilde{\mathbf{Y}} - \Phi \tilde{\mathbf{X}}) \right)$$

où $\tilde{\mathbf{Y}}$ est équivalent à \mathbf{Y} mais centré sur $\boldsymbol{\mu}$

Estimateur du MLE Gaussien

- Il existe des solutions en forme fermée à

$$\hat{\Upsilon} = \arg \max_{\Upsilon} \ell(\mu, \varphi, \Sigma_{\varepsilon})$$

où $\Upsilon = (\mu, \varphi, \Sigma_{\varepsilon})'$

- En annulant les dérivées partielles de $\ell(\mu, \varphi, \Sigma_{\varepsilon})$ on peut montrer que

$$\hat{\mu} = n^{-1} \left(I_k - \sum_{i=1}^p \hat{\Phi}_i \right)^{-1} \sum_{t=1}^n \left(Y_t - \sum_{i=1}^p \hat{\Phi}_i \right)$$

$$\hat{\varphi} = ((\tilde{\mathbf{X}}\tilde{\mathbf{X}}')^{-1} \tilde{\mathbf{X}} \otimes I_k) (\boldsymbol{\psi} - \tilde{\boldsymbol{\mu}})$$

$$\hat{\Sigma}_{\varepsilon} = n^{-1} (\tilde{\mathbf{Y}} - \hat{\boldsymbol{\Phi}}) (\tilde{\mathbf{Y}} - \hat{\boldsymbol{\Phi}})'$$

- On voit alors que le MLE $\hat{\varphi}$ correspond à l'estimateur LS du S69

$$\hat{\boldsymbol{\beta}}_{LS} = ((\mathbf{X}\mathbf{X}')^{-1} \mathbf{X} \otimes I_k) \boldsymbol{\psi}$$

Propriétés limites du MLE Gaussien

- L'étude du comportement asymptotique de $\hat{\Upsilon}$ est complexe
- Il est néanmoins possible de montrer que si Y_t suit un VAR stable

$$\hat{\Upsilon} \xrightarrow{p} \Upsilon$$

- Il est également possible de montrer que

$$\sqrt{n} \begin{pmatrix} \hat{\mu} - \mu \\ \hat{\varphi} - \varphi \\ \hat{\sigma} - \sigma \end{pmatrix} \xrightarrow{d} \mathcal{N} \left(\mathbf{0}, \begin{pmatrix} \Sigma_{\hat{\mu}} & 0 & 0 \\ 0 & \Sigma_{\hat{\varphi}} & 0 \\ 0 & 0 & \Sigma_{\hat{\sigma}} \end{pmatrix} \right)$$

où $\sigma = \text{vech}(\Sigma_\varepsilon)$

Note 1 La matrice de VCV est bloc-diagonale \Rightarrow indépendance asymptotique

\Rightarrow cela est lié au fait qu'on a considéré un VAR centré

Note 2 Le MLE de $(C, \varphi, \Sigma_\varepsilon)'$ a des propriétés similaires mais l'indépendance asymptotique ne tient plus pour C et φ

\Rightarrow on retrouve les résultats des LS qui portaient sur $B = (C, \Phi_1, \dots, \Phi_p)'$

Estimateur MLE et tests inférentiels

- Comme pour les LS, la distribution limite permet le calcul de t -tests
- Pour cela détaillons les VCV asymptotiques

$$\Sigma_{\hat{\mu}} = \left(I_k - \sum_i \Phi_i \right)^{-1} \Sigma_{\varepsilon} \left(I_k - \sum_i \Phi_i' \right)^{-1}$$

$$\Sigma_{\hat{\varphi}} = \Gamma_Y^{-1} \otimes \Sigma_{\varepsilon}$$

$$\Sigma_{\hat{\sigma}} = 2\mathbf{D}_k^+ (\Sigma_{\varepsilon} \otimes \Sigma_{\varepsilon}) \mathbf{D}_k^{+'}$$

où $\Gamma_Y = \tilde{\mathbf{X}}\tilde{\mathbf{X}}'/n$ et $\mathbf{D}_k^+ = (\mathbf{D}_k' \mathbf{D}_k)^{-1} \mathbf{D}_k'$, l'inverse de Moore-Penrose

Note 1 \mathbf{D}_k est une matrice de duplication qui lie les opérateurs vec et vech

$$\text{vec}(\Phi_1) = \mathbf{D}_k \text{vech}(\Phi_1)$$

vech étant un vec appliqué uniquement sur et sous la diagonale

Note 2 Comme \mathbf{D}_k est sparse (creuse) et composée de 0 et de 1, elle approche la singularité et requiert une inverse généralisée

Note 3 Les quantités inconnues sont à remplacer par leur estimateur ML

Décomposition de variance des erreurs de prévisions et inférence

- Précédemment on a vu que

$$\zeta_{jl,h} = \frac{\sum_{i=0}^{h-1} (l'_j \Xi_i l_i)^2}{\sum_{i=0}^{h-1} \Xi_i \Xi_i' = \sum_{i=0}^{h-1} \Psi_i \Sigma_\varepsilon \Psi_i'} = \frac{\sum_{i=0}^{h-1} (l'_j \Xi_i l_i)^2}{MSE({}_t Y_{t+h})}$$

représentait la contribution de l'innovation d'une variable l à la variance de l'erreur de prévisions de la variable j

- On peut montrer grâce à la méthode δ que

$$\sqrt{n}(\hat{\zeta}_{jl,h} - \zeta_{jl,h}) \xrightarrow{d} \mathcal{N}(0, \Sigma_\zeta)$$

où l'expression de Σ_ζ ne sera pas détaillée en raison de sa complexité

⇒ Il est donc théoriquement possible de construire des IC

Rappel Si $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\hat{\beta}})$ la méthode δ permet de dériver la distribution de $g(\beta)$ pour $g(\cdot)$ une fonction continûment différentiable

⇒ On obtient alors que

$$\sqrt{n}(g(\hat{\beta}) - g(\beta)) \xrightarrow{d} \mathcal{N}\left(0, \frac{\partial g}{\partial \beta'} \Sigma_{\hat{\beta}} \frac{\partial g'}{\partial \beta}\right)$$

IFR et inférence

- On s'intéresse ici à la construction d'IC pour les IRF
- Pour rappel, les IRF reposent sur les matrices $\Psi(L) = \Phi(L)^{-1}$
- L'étude du MLE de φ nous a donné la VCV asymptotique de $\hat{\varphi}$

$$\Sigma_{\hat{\varphi}} = \Gamma_{\mathbf{Y}}^{-1} \otimes \Sigma_{\varepsilon}$$

- En invoquant de nouveau la méthode δ on cherche à obtenir la VCV de

$$g(\hat{\Phi}(L)) = \hat{\Phi}(L)^{-1} = \Psi(L)$$

qui sera de la forme

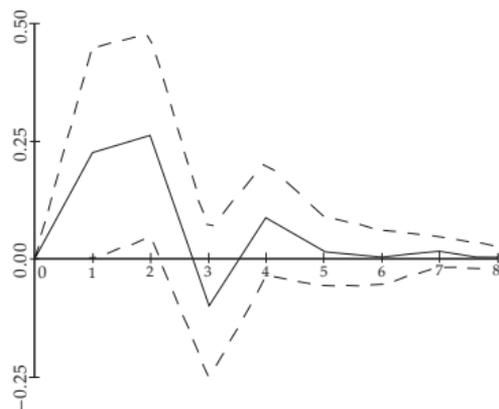
$$\sqrt{n} \text{vec}(\hat{\Psi}_i - \Psi_i) \xrightarrow{d} \mathcal{N}(0, G_i \Sigma_{\hat{\varphi}} G_i'), \quad i = 1, 2, \dots$$

où G_i est une matrice de 1 et de 0 dont la forme générale est complexe

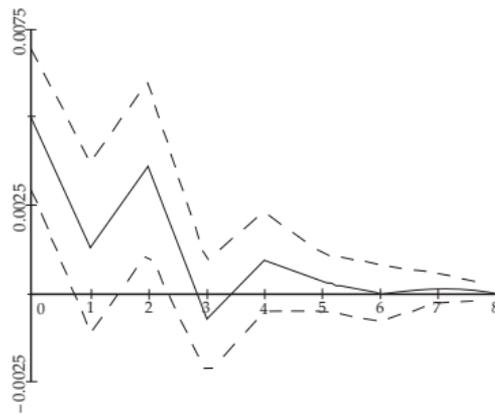
⇒ De nouveau, il est donc théoriquement possible de construire des IC

Note Des résultats similaires tiennent pour les IRF de long terme et orthogonalisées

Exemple : IFR et inférence



Estimated responses of consumption to a forecast error impulse in income with estimated asymptotic standard error bounds.



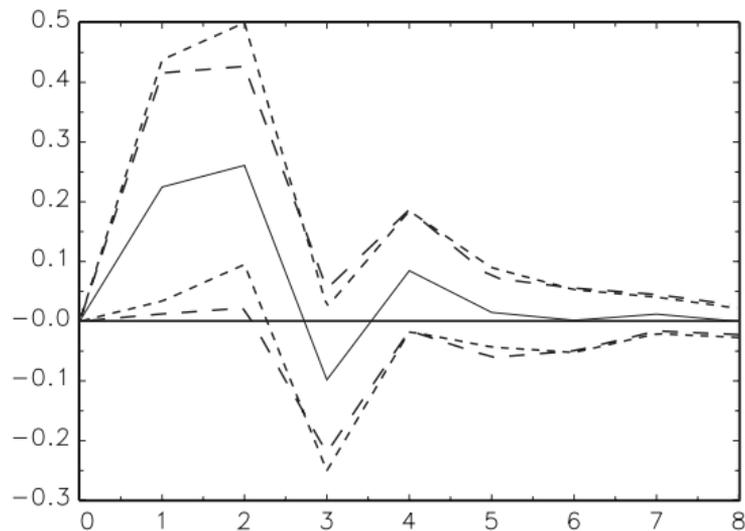
Estimated responses of consumption to an orthogonalized impulse in income with estimated asymptotic standard error bounds.

Intervalles de confiance et bootstrap

- La forme complexe des formules asymptotiques des IRF et décomposition de variance appelle à des alternatives
- ⇒ la puissance des ordinateurs n'étant plus une barrière, les méthodes de bootstrap ont gagné en popularité
- On va présenter ici une méthode basée sur l'hypothèse $\varepsilon_t \sim \text{i. i. d.}$
- ⇒ si cette hypothèse ne tient pas d'autres méthodes devront être envisagées
- 1 Estimation du VAR sur \mathbf{Y} , et stockage des $\widehat{\Phi}_i$, $\widehat{\Psi}_i$, $\widehat{\Xi}_i$, $\widehat{\Sigma}_\varepsilon$ et $\widehat{\varepsilon}_{n \times k}$
 - 2 Mise en place d'un tirage aléatoire avec remise de $\widehat{\varepsilon}_{n \times k}$ qui donne $\tilde{\varepsilon}_{n \times k}$
 - 3 Utilisation des $\widehat{\Phi}_i$ pour construire $\tilde{\mathbf{Y}}$ une série artificielle analogue à \mathbf{Y}
 - 4 Estimation du VAR sur $\tilde{\mathbf{Y}}$ et stockage des $\widehat{\Psi}_i^{(1)}$, $\widehat{\Xi}_i^{(1)}$ et $\widehat{\Sigma}_\varepsilon^{(1)}$
 - 5 Répétition de [2] à [4] \mathbf{b} fois afin d'obtenir $\widehat{\Psi}_i^{(b)}$, $\widehat{\Xi}_i^{(b)}$ et $\widehat{\Sigma}_\varepsilon^{(b)}$, $\mathbf{b} = 1, 2, \dots, \mathbf{b}$
 - 6 Les séries $\widehat{\Psi}_i^{(b)}$, $\widehat{\Xi}_i^{(b)}$ et $\widehat{\Sigma}_\varepsilon^{(b)}$, permettent de construire les IFR et les $\zeta_{j,l,h}$ bootstrappées ainsi que les IC associés

Note Vous verrez en cours de bootstrap les méthodes pour construire les IC

Exemple : IFR et bootstrap



Estimated responses of consumption to a forecast error impulse in income with 95% bootstrap confidence bounds based on 2000 bootstrap replications (— standard intervals, - - - Hall's percentile intervals).

Test LR pour l'ordre d'un VAR

- Jusque là nous avons considéré comme connue la spécification du VAR
- Dans la pratique l'ordre p des retards doit être déterminé
- Un VAR($p' < p$) étant un VAR contraint, un test LR est envisageable
- Repartons de la log-vraisemblance à l'optimum pour le construire

$$\begin{aligned}\ell(\hat{\mu}, \hat{\varphi}, \Sigma_{\hat{\varepsilon}}) &= -\frac{kn}{2} \ln(2\pi) - \frac{n}{2} \ln \det(\hat{\Sigma}_{\varepsilon}^{-1}) - \frac{1}{2} \text{tr} \left(\hat{\mathbf{e}}' \hat{\Sigma}_{\varepsilon}^{-1} \hat{\mathbf{e}} \right) \\ &= -\frac{kn}{2} \ln(2\pi) - \frac{n}{2} \ln \det(\hat{\Sigma}_{\varepsilon}^{-1}) - \frac{1}{2} \text{tr} \left(\sum_{t=1}^n \hat{\varepsilon}_t' \hat{\Sigma}_{\varepsilon}^{-1} \hat{\varepsilon}_t \right) \\ &= -\frac{kn}{2} \ln(2\pi) - \frac{n}{2} \ln \det(\hat{\Sigma}_{\varepsilon}^{-1}) - \frac{1}{2} \text{tr} \left(\hat{\Sigma}_{\varepsilon}^{-1} \sum_{t=1}^n \hat{\varepsilon}_t' \hat{\varepsilon}_t \right) \\ &= -\frac{kn}{2} \ln(2\pi) - \frac{n}{2} \ln \det(\hat{\Sigma}_{\varepsilon}^{-1}) - \frac{1}{2} \text{tr} \left(\hat{\Sigma}_{\varepsilon}^{-1} n \hat{\Sigma}_{\varepsilon} \right) \\ &= -\frac{kn}{2} \ln(2\pi) - \frac{n}{2} \ln \det(\hat{\Sigma}_{\varepsilon}^{-1}) - \frac{n}{2} \text{tr} (I_k) \\ &= -\frac{kn}{2} \ln(2\pi) - \frac{n}{2} \ln \det(\hat{\Sigma}_{\varepsilon}^{-1}) - \frac{n}{2}\end{aligned}$$

Test LR pour l'ordre d'un VAR

- La réécriture facilite grandement la construction du test \mathcal{LR}

$$\mathcal{LR} = -2(\widehat{\ell}_c - \widehat{\ell}_{nc})$$

- En effet, on voit que pour $p' = p - 1$ on aura juste

$$\begin{aligned}\mathcal{LR} &= -n(\ln |(\widehat{\Sigma}_\varepsilon^{-1})_{p-1}| - \ln |(\widehat{\Sigma}_\varepsilon^{-1})_p|) \\ &= n(\ln |(\widehat{\Sigma}_\varepsilon)_{p-1}| - \ln |(\widehat{\Sigma}_\varepsilon)_p|)\end{aligned}$$

où sous $H_0 : \Phi_p = 0$ versus $H_1 : \Phi_p \neq 0$, $\mathcal{LR} \sim \chi_{(p-p')k^2}^2 = \chi_{k^2}^2$

- Conditionnellement au non-rejet de $H_0 : \Phi_p = 0$ on peut ensuite tester

— $H_0^2 : \Phi_{p-1} = 0$ versus $H_1^2 : \Phi_{p-1} \neq 0$ avec toujours $\mathcal{LR} \sim \chi_{k^2}^2$

... puis $H_0^i : \Phi_{p-i+1} = 0$ versus $H_1^i : \Phi_{p-i+1} \neq 0 | \Phi_p = \dots = \Phi_{p-i+2} = 0$

... jusqu'à $H_0^p : \Phi_1 = 0$ versus $H_1^p : \Phi_1 \neq 0 | \Phi_p = \dots = \Phi_2 = 0$

- Un schéma de test alternatif consisterait à tester pour $i > 1$

$$H_0^i : \Phi_p = \dots = \Phi_{p-i+1} = 0 \text{ versus } H_1^i : \Phi_p \neq 0 \text{ ou } \dots \text{ ou } \Phi_{p-i+1} \neq 0$$

mais dans ce cas les tests ne sont plus indépendants et le seuil de risque globale est identifiable (correction à la Bonferroni)

Les critères de sélection

- Les critères vus en univarié s'appliquent et restent de la forme

$$C_k = -\frac{2\hat{\ell}}{n} + \frac{kg(n)}{n}$$

i.e. une combinaison de la qualité d'ajustement et d'une pénalisation

- Compte tenu du cadre multivarié, pour les principaux critères on a

Akaike $AIC = k(1 + \log(2\pi)) + \log(|\hat{\Sigma}_\varepsilon|) + 2pk^2n^{-1}$

Schwarz $BIC = k(1 + \log(2\pi)) + \log(|\hat{\Sigma}_\varepsilon|) + \log(n)pk^2n^{-1}$

Hannan $HQ = k(1 + \log(2\pi)) + \log(|\hat{\Sigma}_\varepsilon|) + 2(\log \log n)pk^2n^{-1}$

- Comme en univarié on cherche à minimiser ces critères et les mêmes remarques s'appliquent quant à leurs particularités

Mises en garde sur les critères de sélection

- Un modèle peu évoqué jusqu'à présent c'est le VARMA(p, q)
 - Comme en univarié, si la composante VMA est inversible
- ... il existe une représentation VAR(∞) du modèle
- Il existe également des VARFIMA(p, D, q)
- ... si $D = (\delta_1, \dots, \delta_k)'$ est négligé, l'approximation devient un VAR(∞)
- ⇒ Dans ces situations la littérature montre que si l'on souhaite trouver l'ordre de troncature optimal les critères doivent respecter :
- $\lim_{n \rightarrow \infty} p^3/n = 0$, i.e. p ne doit pas augmenter trop vite
 - $\lim_{n, p \rightarrow \infty} \sqrt{(n)} \sum_{i=p+1}^{\infty} \Phi_i = 0$, i.e. p ne doit pas augmenter trop lentement
- Malheureusement ni AIC, ni BIC, ni HQ, ne satisfont ces propriétés

Note Des procédures reposant sur des hypothèses difficiles à vérifier existent pour AIC (voir Kuersteiner 2005)

Test de diagnostic

- Beaucoup de tests provenant de l'univarié restent valides

... leur interprétation devient cependant plus délicate

- L'analyse des résidus pour détecter des problèmes de spécification est souvent l'objectif principal des tests de diagnostic

⇒ On s'intéressera alors principalement aux tests suivants :

- Test d'autocorrélation sous l'hypothèse $\varepsilon_t \sim WN(0, \Sigma_\varepsilon)$
- Test d'autocorrélation sur $\hat{\varepsilon}_t \sim WN(0, \Sigma_{\hat{\varepsilon}})$
- Test de nullité jointe des autocorrélations de $\hat{\varepsilon}_t$

⇒ D'autres tests usuels sont d'intérêt mais ne seront pas détaillés

- Test ARCH multivarié
- Test de non-normalité multivarié

Test sur les autocorrélations

- L'inférence autour des h premières autocorrélations résiduelles

$$\mathbf{r}_{h,\hat{\varepsilon}} = \text{vec}(\mathbf{R}_{h,\hat{\varepsilon}}) = \text{vec}(R_{1,\hat{\varepsilon}}, \dots, R_{h,\hat{\varepsilon}})$$

est possible si l'on connaît la distribution limite de $\hat{\mathbf{r}}_h$

- Cette distribution est connue si $\hat{\varepsilon}_t \sim WN(0, \Sigma_{\hat{\varepsilon}})$, i.e. $\mathbf{r}_{h,\hat{\varepsilon}} = 0$

$$\sqrt{n}\hat{\mathbf{r}}_{h,\hat{\varepsilon}} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma_{\hat{\mathbf{r}}}(h))$$

avec $\Sigma_{\hat{\mathbf{r}}}(h)$ la VCV associée dont la forme est complexe mais connue

- Notons que ces résultats tiennent compte du fait que $\hat{\varepsilon}_t$ est estimé

⇒ Si nous observions ε_t on retrouverait un résultat de l'univarié

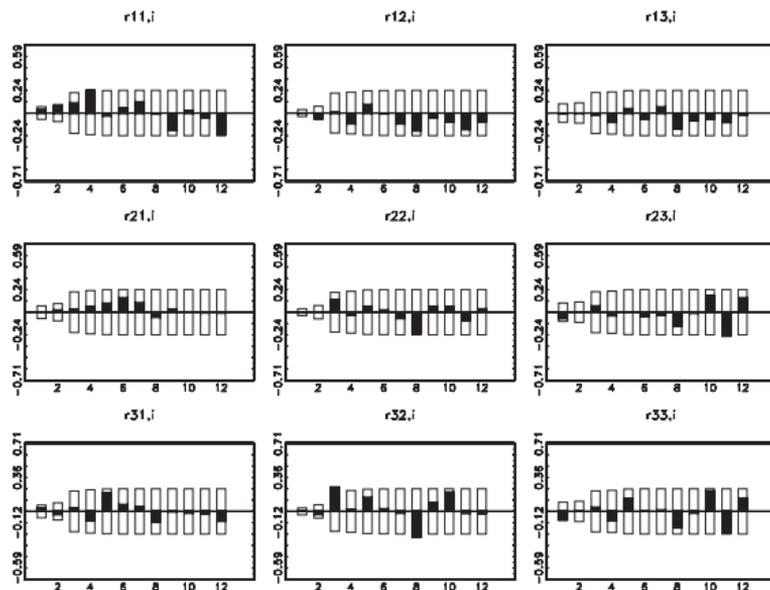
$$\sqrt{n}\mathbf{r}_{h,\varepsilon} \xrightarrow{d} \mathcal{N}(\mathbf{0}, I_h)$$

et donc les intervalles de confiance seraient simplement $\pm 1.96/\sqrt{n} \approx \pm 2/\sqrt{n}$

- Pour (n, h) grands, $\hat{\mathbf{r}}_{h,\hat{\varepsilon}}$ aura tendance à se rapprocher de $\mathbf{r}_{h,\varepsilon}$ sous H_0

Exemple : autocorrélations et inférences sur les résidus estimés

- Appliquons cela aux résidus de notre VAR(2) trivarié où $n = 73$



Test de type portemanteau

- Pour rappel, il s'agit de tester la nullité jointe des autocorrélations

$$H_0 : \mathbf{R}_h = (R_{1,\varepsilon}, \dots, R_{h,\varepsilon}) = 0 \text{ versus } H_1 : \mathbf{R}_h \neq 0$$

- Bien qu'on puisse construire une statistique de Ljung-Box multivariée
... cette dernière aura une puissance faible pour h et n petits
- La statistique ajustée suivante aura de meilleures propriétés

$$\bar{Q}_h := n^2 \sum_{i=1}^h (n-i) \text{tr} \left(\hat{\Gamma}_\varepsilon(i)' \hat{\Gamma}_\varepsilon^{-1} \hat{\Gamma}_\varepsilon(i) \hat{\Gamma}_\varepsilon^{-1} \right)$$

où $\hat{\Gamma}_\varepsilon = n^{-1} \varepsilon \mathbf{I}_n \varepsilon'$ et

$$\hat{\Gamma}_\varepsilon(i) = n^{-1} \sum_{t=i+1}^n \hat{\varepsilon}_t \hat{\varepsilon}_{t-1}'$$

- On peut montrer que pour $n \rightarrow \infty$

$$\bar{Q}_h \approx \chi^2(k^2(h-p))$$

- Dans l'exemple du VAR(2) trivarié, on obtient pour $h = 12$

$$\bar{Q}_{12} = 81.9$$

que l'on compare à $\chi^2(90) \approx 113$ au seuil de risque de 5%

Test de Wald d'un VAR contraint linéairement

- Nous avons vu que l'absence de causalité de Granger équivaut à

$$\Phi_{12,i} = 0, i = 1, 2, \dots, p$$

et cela revient à tester des contraintes sur les coefficients

$$H_0 : K\beta = \kappa \text{ versus } H_0 : K\beta \neq \kappa$$

- A partir des résultats limites des estimateurs LS ou ML

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, \Gamma^{-1} \otimes \Sigma_\varepsilon)$$

on obtient $\sqrt{n}(K\hat{\beta} - K\beta) \xrightarrow{d} \mathcal{N}(0, K(\Gamma^{-1} \otimes \Sigma_\varepsilon)K')$ et donc

$$n(K\hat{\beta} - \kappa)'(K(\Gamma^{-1} \otimes \Sigma_\varepsilon)K')^{-1}(K\hat{\beta} - \kappa) \xrightarrow{d} \chi_c^2$$

où c correspond au nombre de contraintes et au rang de $K_{c \times (k^2 p + k)}$

- En remplaçant Γ^{-1} et Σ_ε par $\hat{\Gamma}^{-1}$ et $\hat{\Sigma}_\varepsilon$ on obtient la statistique Wald

$$\mathcal{W} = n(K\hat{\beta} - \kappa)'(K(\hat{\Gamma}^{-1} \otimes \hat{\Sigma}_\varepsilon)K')^{-1}(K\hat{\beta} - \kappa)$$

Contraintes linéaires

- Testons l'absence de causalité de Granger de $(y_{2,t}, y_{3,t})$ vers $y_{1,t}$ dans le VAR(2) trivarié

$$H_0 : \Phi_{12,i} = \Phi_{13,i} = 0, i = 1, 2$$

ce qui fait $c = 4$ contraintes et $\kappa = (0, 0, 0, 0)'$

- A partir des estimations S74 on a donc β et K comme suit

$$\beta = \begin{pmatrix} -0.017 \\ -0.32 \\ 0.146 \\ 0.961 \\ -0.161 \\ 0.115 \\ 0.934 \\ 0.016 \\ 0.044 \\ -0.153 \\ 0.289 \\ 0.05 \\ 0.019 \\ -0.01 \\ 0.013 \\ -0.002 \\ 0.225 \\ -0.264 \\ 0.034 \\ 0.355 \\ -0.022 \end{pmatrix}, \quad K = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Test de causalité de Granger

- En pratique, pour n petit, $\mathcal{L}(\mathcal{W})$ a des queues plus épaisses que χ_c^2

⇒ Cela provient de la présence de $\hat{\Gamma}^{-1}$ et $\hat{\Sigma}_\varepsilon$

- On peut alors considérer $\mathcal{F} = \mathcal{W}/c$ dont la distribution de Fisher sera

$$\mathcal{F} \approx F(c, n - kp - 1) \equiv F(c, kn - k^2p - k)$$

- Testons l'absence de causalité de Granger de $(y_{2,t}, y_{3,t})$ vers $y_{1,t}$ dans le VAR(2) trivarié

$$H_0 : \Phi_{12,i} = \Phi_{13,i} = 0, i = 1, 2$$

ce qui fait $c = 4$ contraintes et $\kappa = (0, 0, 0, 0)'$

- $K_{3 \times (18+3)}$ (le 3 venant des 3 constantes) sera pleine de 0 avec quatre 1

⇒ Sur la base des estimations du S74 on obtient $\mathcal{F} = 1.59$

- La valeur critique à 95% est $F(4, 198) \approx F(4, 66) \approx 2.5$

⇒ Au seuil de risque de 5% on ne peut rejeter l'absence de causalité de Granger allant du groupe PIB/conso vers l'investissement

Test de causalité de instantanée

- Tester la causalité instantanée repose sur le même principe

⇒ les contraintes reposent cette fois sur les éléments de $\sigma = \text{vech}(\Sigma_\varepsilon)$

$$H_0 : K\sigma = 0 \text{ versus } H_0 : K\sigma \neq 0$$

- Comme la distribution de $\hat{\sigma}$ est compliquée, on va considérer $\Sigma_\varepsilon = PP'$

... la décomposition de Choleski de Σ_ε et reformuler les hypothèses

$$H_0 : K \text{vech}(P) = 0 \text{ versus } H_0 : K \text{vech}(P) \neq 0$$

car $\sqrt{n} \text{vech}(\hat{P} - P) \xrightarrow{d} \mathcal{N}(0, H\Sigma_{\hat{\sigma}}H')$ mène à une VCV plus simple

- On ne reportera pas H car basé sur des concepts matriciels avancés mais

$$\mathcal{W} = n \text{vech}(\hat{P})' K' (KH\Sigma_{\hat{\sigma}}H'K')^{-1} K \text{vech}(\hat{P}) \xrightarrow{d} \chi_c^2$$

avec c le nombre de contraintes

- Testons l'absence de causalité de instantanée de $(y_{2,t}, y_{3,t})$ vers $y_{1,t}$

⇒ $\mathcal{W} = 5.70$ pour une valeur critique à 95% de 5.99, ce qui nous amène à ne pas rejeter H_0

Test de causalité de Granger à horizon h

- Au niveau M1, nous ne traiterons pas des tests de causalité à horizon h

⇒ en effet, ces derniers relèvent de la modélisation non-linéaire par les contraintes non-linéaires qu'ils imposent dans le VAR

Note Retenons cependant qu'une implémentation naïve du test de Wald mènera malheureusement à une inférence non valide

Cadre général

- L'analyse des VAR, notamment au travers des IRF, questionne
 - Si les IRF ne sont pas uniques, que reflètent-elles des interactions économiques ?
 - Les VAR sont-ils des formes réduites de modèles théoriques ?
 - Comment s'assurer que ces modèles racontent une histoire fondée économiquement ?

⇒ L'analyse structurelle des VAR cherche à répondre à ces questions

- Le point de départ est la mise en place de restrictions structurelles
 - i.e. des restrictions qui, fondées sur la théorie économique, contraignent la dynamique du VAR et les liens causaux entre les variables qui le composent

Cadre formel

- Considérons un VAR(p) stable, donc stationnaire, k -varié

$$Y_t = \sum_{i=1}^p \Phi_i Y_{t-i} + \varepsilon_t$$

où la présence d'un constante C est omise sans perte de généralité et

$$\varepsilon_t \sim WN(\mathbf{0}, \Sigma_\varepsilon)$$

- On sait donc que ce VAR admet une représentation de Wold

$$Y_t = \sum_{i=0}^{\infty} \Psi_i \varepsilon_{t-i}$$

dont les matrices Ψ_i reflètent de manière impropre les IRF

⇒ En effet, les possibles corrélations instantanées des éléments de ε_t ne permettent pas d'isoler l'impact d'une variable sur une autre

- On a vu qu'une solution reposant sur la décomposition de Cholesky de

$$\Sigma_\varepsilon = PP'$$

existe mais elle est arbitraire car il existe d'autres possibilités

The A-Model

- Comment s'extraire de ce cadre pour définir des IRF uniques ?
- C'est l'objectif du modèle dit **A-Model** dont la forme est

$$AY_t = \Phi_1^* Y_{t-1} + \dots + \Phi_p^* Y_{t-p} + \varepsilon_t^*$$

où la matrice A va imposer des relations instantanées directement sur les observables

- On voit immédiatement que $\Phi_i^* := A\Phi_i$, $i = 1, \dots, p$ et que

$$\varepsilon_t^* \sim WN(0, A\Sigma_\varepsilon A')$$

ce qui permettra via un choix approprié de A d'obtenir Σ_ε^* diagonale

- ε_t^* représentera ce que l'on appelle des chocs structurels
- Bien entendu la forme VMA(∞) de ce VAR devient

$$Y_t = \sum_{i=0}^{\infty} \Psi_i^* \varepsilon_{t-i}^*$$

avec $\Psi_i^* := \Psi_i A^{-1}$, $i = 1, \dots$

La structure de A

- Comment structurer A ? Tout d'abord on voit dans le SVAR que

$$\Sigma_{\varepsilon}^*$$

contient k coefficients à estimer car diagonale

$$\Phi(L)^*$$

contient $k^2 p$ coefficients à estimer et sans restrictions

$$A_{k \times k}$$

contient $k(k - 1)$ coefficients à estimer, i.e. un total de $(p + 1)k^2$ inconnus

- Du côté du VAR en forme réduite on voit que

$$\Sigma_{\varepsilon}$$

contient $(k^2 + k)/2$ coefficients à estimer car symétrique et

$$\Phi(L)$$

contient $k^2 p$ coefficients à estimer, i.e. un total de $((2p + 1)k^2 + k)/2$

Contraintes d'identification

- Pour identifier les paramètres du modèle

$$k(k - 1)/2$$

contraintes doivent être imposées pour équilibrer le nombre d'inconnues

- Ces contraintes prendront souvent la forme de 0 ou de 1 dans A , e.g.

$$A = \begin{pmatrix} 1 & 0 & \dots & 0 \\ \mathbf{a}_{21} & 1 & 0 & 0 \\ \vdots & & \ddots & \vdots \\ \mathbf{a}_{k1} & \mathbf{a}_{k2} & \dots & 1 \end{pmatrix}$$

e.g. Ici on voit qu'assez naturellement on normalise a_{ii} à 1, $i = 1, \dots, k$

... et que l'on impose la nullité aux éléments au dessus de la diagonale

⇒ On a exactement $k(k - 1)/2$ contraintes : le modèle est juste-identifié

Note On aurait pu en imposer plus (sur-identification) mais pas moins

Les chocs structurels à la Cholesky

- Sauf si la théorie économique nous guide vers un ordonnancement très particulier des variables, la forme de A précédente est ad hoc
- Cette forme rappelle d'ailleurs celle issue de la décomposition de Choleski

$$P = \begin{pmatrix} p_{11} & 0 & \cdots & 0 \\ p_{21} & p_{22} & 0 & 0 \\ \vdots & & \ddots & \vdots \\ p_{k1} & p_{k2} & \cdots & p_{kk} \end{pmatrix}$$

⇒ On voit ainsi que la décomposition de Choleski reflète un SVAR où

$$A = P^{-1}$$

mais ce SVAR n'est pertinent que si l'ordonnancement des variables a un sens économique

Note L'argument selon lequel l'analyse des VAR non-structurels est agnostique de toute théorie économique sous-jacente est donc trompeur

⇒ En pensant être agnostique, on impose en fait via Choleski une structure théorique qui n'a peut-être aucune pertinence économique

Conditions théoriques d'identification du A-Model

- A ne peut être arbitraire, même avec le bon nombre de contraintes
- En d'autres termes, A doit être tel que le système d'équations

$$A\Sigma_\varepsilon A' = \Sigma_\varepsilon^*$$

a une unique solution, au moins localement

- Il existe des conditions complexes d'identification locale

$$\text{rank } \mathcal{C}_A = \text{rank} \begin{pmatrix} -2\mathbf{D}_k^+(\Sigma_\varepsilon \otimes A^{-1}) & \mathbf{D}_k^+(A^{-1} \otimes A^{-1})\mathbf{D}_k \\ S_A & 0 \\ 0 & S_\sigma \end{pmatrix} = k^2 + \frac{k(k+1)}{2}$$

... qui peuvent devenir des conditions d'identification globale si $\text{diag}(A) = \mathbf{1}$

- Les matrices type S_X sont des matrices de sélection qui sélectionnent les éléments de $\text{vech}(X)$ sous la diagonale et

$$\mathbf{D}_k^+ = (\mathbf{D}_k' \mathbf{D}_k)^{-1} \mathbf{D}_k'$$

est l'inverse de Moore-Penrose de \mathbf{D}_k , une matrice de duplication qui lie les opérateurs vec et vech

Conditions d'identification en pratique du A-Model

- Les conditions d'identification sont invérifiables avec certitude
- ... car elles impliquent des paramètres inconnus
- Au mieux on remplace les quantités inconnues par les estimées

Note Heureusement, si le modèle est identifié, on peut montrer que

$$\text{rank } C_A \leq k^2 + \frac{k(k+1)}{2}$$

dans tout l'espace des paramètres

⇒ Si le modèle est au moins localement identifié, il devrait remplir la condition de rang avec une probabilité unitaire

The B-Model

- Le **A-Model** permet d'orienter les interactions entre les variables
... de manière fondée économiquement
 - Si A est soigneusement structuré, les IRF sont uniquement définies
... mais l'interprétation des chocs ε_t^* qui les composent reste à étudier
- ⇒ On peut alors chercher à identifier ε_t^* directement à partir de

$$\varepsilon_t = B\varepsilon_t^*$$

i.e. les erreurs de prévision (résidus) de la forme réduite

Note Ici A est absent du modèle et l'on travaille directement sur ε_t dans

$$Y_t = \Phi_1 Y_{t-1} + \dots + \Phi_p Y_{t-p} + \varepsilon_t$$

- On voit alors que

$$\Sigma_\varepsilon = B\Sigma_\varepsilon^*B' = BB'$$

où sans perte de généralité on a supposé que $\Sigma_\varepsilon^* = I_k$

⇒ Pour identifier les éléments de B il faut $k(k-1)/2$ restrictions

Conditions théoriques d'identification du B-Model

- Comme pour A , des conditions d'identification existent pour B

... et elles sont satisfaites par Choleski mais cette approche est ad hoc

- Les conditions d'identification locale sont

$$\text{rank } C_B = \text{rank} \begin{pmatrix} 2D_k^+(B \otimes I_k) \\ S_B \end{pmatrix} = k^2$$

- Notons à ce stade que les restrictions dans A et B peuvent s'écrire

$$\text{vec}(A) = K_A \varrho_A + \iota_A \text{ et } \text{vec}(B) = K_B \varrho_B + \iota_B$$

où ϱ_A et ϱ_B contiennent les coefficients contraints de A et B

- Comme pour le test de Wald, K_A et K_B sont des matrices de 0 et de 1

The AB-Model

- On peut bien entendu considérer une combinaison du modèle A et B

$$AY_t = \Phi_1^* Y_{t-1} + \dots + \Phi_p^* Y_{t-p} + A\varepsilon_t$$

où $A\varepsilon_t = B\varepsilon_t^*$

- En général, on mènera néanmoins l'analyse sur

$$Y_t = \Phi_1 Y_{t-1} + \dots + \Phi_p Y_{t-p} + \varepsilon_t$$

en considérant $\varepsilon_t = A^{-1}B\varepsilon_t^*$ qui implique

$$\Sigma_\varepsilon = A^{-1}BB'A^{-1}$$

- On voit donc que pour le modèle réduit on dispose de toujours de

$$k(k+1)/2$$

équations quand A et B sont composées $2k^2$ éléments

⇒ Il nous faut donc à présent $2k^2 - k - k(k+1)/2$ restrictions, si $\text{diag}(A) = 1$

- Le **AB-Model** est donc difficile à mettre en œuvre dans la pratique

Conditions théoriques d'identification du AB-Model

- Des conditions d'identification existent aussi pour ce modèle général
- Les conditions d'identification locale sont

$$\text{rank } \mathcal{C}_{AB} = \text{rank} \begin{pmatrix} -2\mathbf{D}_k^+(\Sigma_\varepsilon \otimes \mathbf{A}^{-1}) & 2\mathbf{D}_k^+(\mathbf{A}^{-1}\mathbf{B} \otimes \mathbf{A}^{-1}) \\ \mathbf{S}_A & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_B \end{pmatrix} = 2k^2$$

Exemple d'application du AB-Model

- Pour illustrer l'utilisation du **AB-Model** considérons

$$Y_t = \Phi_1 Y_{t-1} + \dots + \Phi_p Y_{t-p} + \varepsilon_t$$

un VAR(p) trivarié en forme réduite composé de

- $y_{1,t} :=$ le PIB
- $y_{2,t} :=$ le taux d'intérêt
- $y_{3,t} :=$ l'offre réelle de monnaie

- Dans l'esprit d'un modèle Keynésien on se dit alors que les équations

$$\varepsilon_{1,t} = -a_{12}\varepsilon_{2,t} + b_{11}\varepsilon_{1,t}^*, \text{ (Courbe IS)}$$

$$\varepsilon_{2,t} = -a_{21}\varepsilon_{1,t} - a_{23}\varepsilon_{3,t} + b_{22}\varepsilon_{2,t}^*, \text{ (Courbe LM)}$$

$$\varepsilon_{3,t} = b_{33}\varepsilon_{3,t}^*, \text{ (Offre de Monnaie Exogène)}$$

représentent les équilibres économiques d'une petite économie fermée

- $\varepsilon_{1,t}^*$ et $\varepsilon_{2,t}^*$ représentent les chocs structurels de demande (IS et LM)
- $\varepsilon_{3,t}^*$ représentent les chocs d'offre structurels

Exemple d'application du AB-Model

- De cette formulation économiquement fondée viennent les contraintes

$$\begin{pmatrix} 1 & a_{12} & 0 \\ a_{21} & 1 & a_{23} \\ 0 & 0 & 1 \end{pmatrix} \varepsilon_t = \begin{pmatrix} b_{11} & 0 & 0 \\ 0 & b_{22} & 0 \\ 0 & 0 & b_{33} \end{pmatrix} \varepsilon_{1,t}^*$$

- Sur la base de $\text{vec}(A) = K_A \varrho_A + \iota_A$ et $\text{vec}(B) = K_B \varrho_B + \iota_B$ on a donc

$$\begin{pmatrix} 1 \\ a_{21} \\ 0 \\ a_{12} \\ 1 \\ 0 \\ 0 \\ a_{23} \\ 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} a_{21} \\ a_{12} \\ a_{23} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \text{ et } \begin{pmatrix} b_{11} \\ 0 \\ 0 \\ 0 \\ b_{22} \\ 0 \\ 0 \\ 0 \\ b_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} b_{11} \\ b_{22} \\ b_{33} \end{pmatrix}$$

- Ici $k = 3$ ce qui implique $2k^2 - k(k + 1)/2 = 12$ restrictions que nous avons avec les 3 uns et 3 zéros dans A et les 6 zéros dans B

Note Avec $\text{vec}(A)$ et $\text{vec}(B)$ on peut calculer S_A et S_B : la condition d'identification locale est satisfaite

Les restrictions de long terme à la Blanchard-Quah

- Construire un modèle comme le précédent n'est pas toujours aisé
- ... car supprimer certaines interactions est parfois dur à justifier
- Blanchard-Quah proposent une autre approche basée sur

$$Y_t = \sum_{i=0}^{\infty} \Psi_i \varepsilon_{t-i}, \quad \varepsilon_{t-i} = A^{-1} B \varepsilon_{t-i}^*$$

i.e. la forme VMA(∞), dont on sait que les

$$\Psi_{\infty} = (I_k - \Phi_1 - \dots - \Phi_p)^{-1}$$

seront les multiplicateurs de long terme des innovations et donc les

$$\Xi_{\infty} = (I_k - \Phi_1 - \dots - \Phi_p)^{-1} A^{-1} B$$

seront les multiplicateurs de long terme des innovations structurelles

⇒ Les restrictions seront alors vues comme des contraintes sur l'impact de long terme des chocs structurels

Exemple : Blanchard-Quah (1989)

- Soit un VAR(p) bivarié ($k = 2$) composé de
 - $y_{1,t} :=$ le PIB
 - $y_{2,t} :=$ le taux de chômage
- On souhaite que les chocs de demande (PIB) n'aient qu'un effet transitoire sur $y_{2,t}$

⇒ le multiplicateur de long terme de $\varepsilon_{1,t}^*$ doit être nul

$$\bar{\Xi}_{12,\infty} = 0$$

- Dans ce petit modèle, aucune restriction instantanée n'est requise par la théorie et donc $A = I_k$ est acceptable ⇒ **B-Model**

⇒ Il reste donc $k(k - 1)/2 = 1$ restriction à imposer: donc $\bar{\Xi}_{12,\infty} = 0$ suffit

Chapitre 2

Les VAR Cointégrés



Surmonter le risque de régression factice

- Nous avons vu que le risque de régression factice existait pour

$$y_t \sim I(\delta_y) \text{ et } x_t \sim I(\delta_x) \text{ avec } \delta_y, \delta_x \in (0, \infty)$$

si $\delta_y + \delta_x > 1/2$

⇒ comment s'assurer que le résultat de la régression n'est pas fallacieux ?

- Grâce à la théorie de la **cointégration** de Granger (1981)
 - pour comprendre la cointégration repartons d'un cadre simple
 - Soit $x_t = x_{t-1} + \eta_t$ une marche aléatoire et

$$y_t = \beta x_t + \varepsilon_t \tag{8}$$

avec $\varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2 < \infty)$ et $\beta = 1$

Régression non-stationnaire

- Dans le modèle (8) il apparaît que

- $x_t \sim I(\delta_x = 1)$

- $\varepsilon_t \sim I(\delta_\varepsilon = 0)$

- ⇒ $y_t \sim I(\max(\delta_x, \delta_\varepsilon) = 1)$

- Le modèle (8) implique donc deux variables non-stationnaires

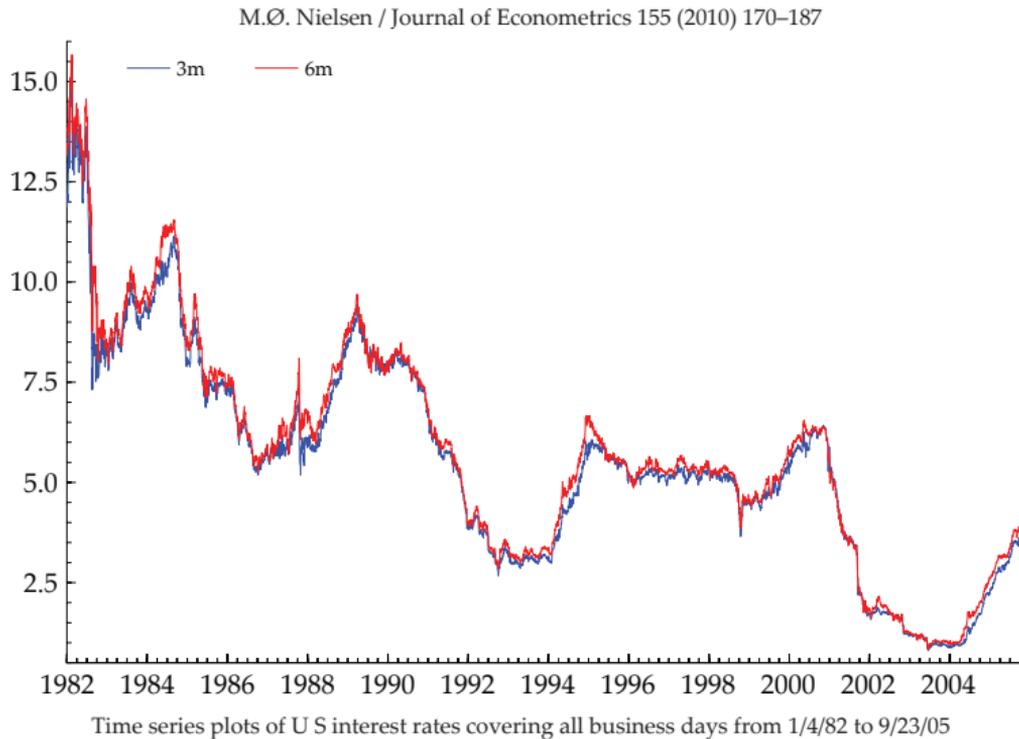
- ⇒ y_t et x_t sont animées par des **tendances stochastiques** et possèdent des **dynamiques de long terme**

- Pourtant, comme $\beta = 1$, y_t et x_t **ne s'éloignent jamais indéfiniment l'un de l'autre**

- On constate en effet que $y_t - x_t = \varepsilon_t \sim I(0)$ est stationnaire

- ⇒ les déviations entre y_t et x_t sont temporaires

Intuition graphique



Intuition de la cointégration

- Dans le modèle (8) une combinaison de variables $I(1)$ est $I(0)$
- En renversant le raisonnement on obtient l'idée de la cointégration

Definition 1 (Cointégration bivariée et paradigme $I(1)/I(0)$)

Soient y_t et x_t deux variables $I(1)$. S'il existe un coefficient $\beta \neq 0$ produisant une combinaison linéaire

$$y_t - \beta x_t = \varepsilon_t$$

tel que $\varepsilon_t \sim I(0)$ alors y_t et x_t possèdent un équilibre de long terme commun et sont cointégrées

- Par son existence, cet équilibre de long terme commun assure que la régression n'est pas factice

Cointégration et vecteur cointégrant

- Dans un papier fondateur, Engle et Granger (1987) définissent ainsi la cointégration

Definition 2 (Cointégration selon Engle et Granger, 1987)

Les éléments du vecteur de variables aléatoires $X_t = (x_{1,t}, \dots, x_{k,t})$ sont dits cointégrés d'ordres (δ, b) , noté $CI \sim (\delta, b)$ si

tous les éléments du vecteur X_t sont $I(\delta)$

il existe un vecteur $\beta \neq 0$ tel que $\varepsilon_t = \beta' X_t \sim I(\delta - b)$ avec $b > 0$

Le vecteur β est de dimension k et se nomme vecteur cointégrant.

- Dans ce chapitre, on supposera $\delta = 1$ et $b = 1$

Composition du vecteur cointégrant

- Pour mieux comprendre le modèle $\varepsilon_t = \beta' X_t$, détaillons β
- Le plus souvent on normalisera β_1 à 1 et on s'attend à avoir

$$\beta' = \begin{pmatrix} 1 \\ -\beta_2 \\ -\beta_3 \\ \vdots \\ -\beta_k \end{pmatrix} \text{ et } X_t = (x_{1,t}, \dots, x_{k,t})$$

de sorte que $\varepsilon_t = \beta' X_t$ s'écrit

$$\varepsilon_t = x_{1,t} - \beta_2 x_{2,t} - \dots - \beta_k x_{k,t} \text{ ou } x_{1,t} = \beta_2 x_{2,t} + \dots + \beta_k x_{k,t} + \varepsilon_t$$

- Dans le cas le plus simple, on a $x_{1,t} = \beta_2 x_{2,t} + \varepsilon_t$

Représentation VAR

- Considérons un modèle VAR(1) bivarié

$$X_t = \Phi X_{t-1} + \varepsilon_t \quad (9)$$

avec $(x_{1,t}, x_{2,t}), \varepsilon_t \sim \text{i. i. d. } (0, \sigma_\varepsilon^2 < \infty)$ et

$$\Phi = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}$$

- A partir du polynôme caractéristique $\det(\Phi - zI) = 0$ on sait que
 - La relation est factice si les deux racines sont sur le cercle unité
 - X_t est stationnaire si les racines ne sont pas sur le cercle unité
- ⇒ Étudions le cas intermédiaire où $z_1^{(\Phi)} = 1$ et $z_2^{(\Phi)} = \lambda^{-1}$ avec $|\lambda| < 1$

VAR et diagonalisation

- Puisque $z_1^{(\Phi)} = 1$, $x_{1,t} \sim I(1)$ et $x_{2,t} \sim I(1)$
- Puisque $\mathbf{z} \neq 0$, Φ est de rang plein (valeurs propres non-nulles)

⇒ Φ admet une représentation de la forme

$$\Phi = P\Lambda P^{-1} \text{ avec } \Lambda = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix} \text{ et } P = \begin{pmatrix} \beta & \alpha \\ d & g \end{pmatrix}, \text{ des vecteurs propres}$$

- Sans perte de généralité on suppose $\det P = \beta d - g\alpha = 1$ et donc

$$P^{-1} = \begin{pmatrix} d & -\alpha \\ -g & \beta \end{pmatrix}$$

⇒ On en déduit aisément une expression pour Φ

$$\Phi = \begin{pmatrix} \beta & \alpha \\ g & d \end{pmatrix} \begin{pmatrix} d & -\alpha \\ -g\lambda & \beta\lambda \end{pmatrix} = \begin{pmatrix} \beta d - \lambda\alpha g & -\beta\alpha(1 - \lambda) \\ g d(1 - \lambda) & -g\alpha + \lambda\beta d \end{pmatrix} \quad (10)$$

VAR cointégré

- En multipliant tous les termes de X_t par P^{-1} on obtient

$$P^{-1}X_t = P^{-1}\Phi X_{t-1} + P^{-1}\varepsilon_t$$

- On observe alors que $P^{-1}\Phi = P^{-1}P\Lambda P^{-1} = \Lambda P^{-1}$

- En notant $W_t = (w_{1t}, -w_{2t})' = P^{-1}X_t$ on obtient

$$W_t = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix} W_{t-1} + \eta_t, \text{ avec } \eta_t = P^{-1}\varepsilon_t$$

- A partir de Λ , on voit que pour ce nouveau VAR,

$$\underbrace{w_{1t} = dx_{1t} - \alpha x_{2t}}_{\text{non-stationnaire}} \text{ et } \underbrace{-w_{2t} = gx_{1t} - \beta x_{2t}}_{\text{stationnaire}}$$

- On sait que $(x_{1t}, x_{2t})' \sim I(1)$ mais $w_{2t} \sim I(0)$ car $|\lambda| < 1$

$\Rightarrow X_t \sim CI(1, 1)$ de vecteur cointégrant $(g, -\beta)'$

Représentation en tendance stochastique commune

- A partir de $W_t = P^{-1}X_t$ on peut exprimer X_t comme

$$X_t = PW_t$$

- On a alors,

$$x_{1t} = \beta \underbrace{w_{1t}}_{I(1)} - \alpha \underbrace{w_{2t}}_{I(0)} \text{ et } x_{2t} = g \underbrace{w_{1t}}_{I(1)} - d \underbrace{w_{2t}}_{I(0)}$$

⇒ $X_t = (x_{1t}, x_{2t})'$ contient donc deux composantes

- une composante stationnaire $I(0)$: w_{2t}
- une composante non-stationnaire $I(1)$: w_{1t}

⇒ cointégration = existence d'une tendance stochastique commune

Représentation ECM

- Pour rappel, $\det P = \beta d - g\alpha = 1$ qu'on peut réécrire comme

$$\beta d - \gamma\alpha + g\alpha - \lambda g\alpha = 1 + g\alpha - \lambda g\alpha \Rightarrow \beta d - \lambda g\alpha = 1 + g\alpha(1 - \lambda)$$

ou encore

$$\beta d - g\alpha - \beta d + \lambda\beta d = 1 - \beta d + \lambda\beta d \Rightarrow -g\alpha + \lambda\beta d = 1 - \beta d(1 - \lambda)$$

- Dans l'expression de Φ (équation 10), cette réécriture donne

$$\Phi = \begin{pmatrix} \beta d - \lambda g\alpha & -\beta\alpha(1 - \lambda) \\ g d(1 - \lambda) & -g\alpha + \lambda\beta d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + (1 - \lambda) \begin{pmatrix} \alpha \\ d \end{pmatrix} \begin{pmatrix} g & -\beta \end{pmatrix}$$

$\Rightarrow X_t = \Phi X_{t-1} + \varepsilon_t$ possède une forme à correction d'erreur (ECM)

$$X_t - X_{t-1} = (1 - \lambda) \begin{pmatrix} \alpha \\ d \end{pmatrix} \underbrace{\begin{pmatrix} g x_{1,t-1} & -\beta x_{2,t-1} \end{pmatrix}}_{\text{EC: correction d'erreur}} + \varepsilon_t \quad (11)$$

\Rightarrow EC = correction du déséquilibre de long terme passé

Interprétation de l'ECM

- Afin d'interpréter les coefficients de l'ECM, réécrivons ΔX_t

$$\Delta X_t = (1 - \lambda) \begin{pmatrix} \alpha \\ d \end{pmatrix} (g \quad -\beta) X_{t-1} + \varepsilon_t = \alpha \beta' X_{t-1} + \varepsilon_t \quad (12)$$

- avec α la force de rappel vers l'équilibre de long terme

$$\alpha = \begin{pmatrix} (1 - \lambda)\alpha \\ (1 - \lambda)d \end{pmatrix} \quad (13)$$

- et β le vecteur cointégrant décrivant l'équilibre de long terme

$$\beta = \begin{pmatrix} g \\ -\beta \end{pmatrix} \quad (14)$$

- Par convention, on note $\Pi = \alpha \beta'$ et on parle de VECM(0)

$$\Delta X_t = \Pi X_{t-1} + \varepsilon_t \quad (15)$$

Cointégration et rang réduit

- Remarquons tout d'abord que

$$\Pi = \alpha\beta' = \Phi - I \quad (16)$$

- D'après les racines de $|\Phi - \mathbf{z}I| = 0$, on déduit que

$$z_1^{(\Pi)} = 0 \text{ et } z_2^{(\Pi)} = (\beta d - \alpha g)(\lambda - 1) = \lambda - 1$$

- La valeur propre $z_1^{(\Pi)}$ étant nulle, Π est de rang réduit non nul

⇒ le rang de Π est révélateur de la présence de cointégration ou non

⇒ Si $z_1^{(\Phi)} = 1$ et $|z_2^{(\Phi)}| > 1$, $\text{rang}(\Pi) = 1$ et il y a cointégration

⇒ Si $z_1^{(\Phi)} = z_2^{(\Phi)} = 1$, $\text{rang}(\Pi) = 0$ et il n'y a pas de cointégration

⇒ Si $|z_1^{(\Phi)}| > 1$ et $|z_2^{(\Phi)}| > 1$, $\text{rang}(\Pi) = 2$ et le VAR est stationnaire

Engle et Granger (1987), Johansen (1991), Hansen (2005)

- Soit un VAR k -varié

$$X_t = \Phi X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d.} (0, \sigma_\varepsilon^2 < \infty) \quad (17)$$

Theorem 1 (Théorème de représentation de Granger)

Soit un VAR de dimension k décrit par l'équation (17) et admettant une forme à correction d'erreur décrite par l'équation (15) avec $\Pi = \alpha\beta'$, le polynôme caractéristique de X_t ayant $k - r$ racines égales à l'unité et toutes les autres en dehors du cercle unité. Alors,

ΔX_t est $I(0)$ de représentation de Wold $\Delta X_t = C(L)\varepsilon_t$

$\beta'X_t$ est $I(0)$ et le rang de cointégration de X_t est r , $0 \leq r < k$

X_t est $I(1)$ de forme $X_t = X_0 + C(1) \sum_{j=1}^t \varepsilon_j$

$\alpha_\perp \sum_{j=1}^t \varepsilon_j$ représente les tendances stochastiques communes (α_\perp étant le complément orthogonal de α : $\alpha'_\perp \alpha = 0$)

VECM et dynamique de court terme

- Le modèle (15) est simple et sans dynamique de court terme

⇒ Supposons un système cointégré bivarié de la forme

$$\begin{aligned}\Delta x_{1t} &= \alpha_1(x_{1,t-1} - \beta_2 x_{2,t-1}) + \gamma_{11} \Delta x_{1,t-1} + \gamma_{12} \Delta x_{2,t-1} + \varepsilon_{1t} \\ \Delta x_{2t} &= \alpha_2(x_{1,t-1} - \beta_2 x_{2,t-1}) + \gamma_{21} \Delta x_{1,t-1} + \gamma_{22} \Delta x_{2,t-1} + \varepsilon_{2t}\end{aligned}\tag{18}$$

- Sous forme vectorielle, il s'agit d'un VECM($p = 1$)

$$\Delta X_t = \alpha \beta' X_{t-1} + \Gamma_1 \Delta X_{t-1} + \varepsilon_t$$

avec les coefficients de la dynamique de court terme

$$\Gamma_1 = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix}$$

$$\text{et } \beta = (\gamma, -\beta)' = (1, -\beta_2)'$$

Coefficients attendus de l'ECM

$$\Delta x_{1t} = \alpha_1(x_{1,t-1} - \beta_2 x_{2,t-1}) + \gamma_{11} \Delta x_{1,t-1} + \gamma_{12} \Delta x_{2,t-1} + \varepsilon_{1t}$$

$$\Delta x_{2t} = \alpha_2(x_{1,t-1} - \beta_2 x_{2,t-1}) + \gamma_{21} \Delta x_{1,t-1} + \gamma_{22} \Delta x_{2,t-1} + \varepsilon_{2t}$$

- On s'attend à trouver $\alpha_1 < 0$ car
 - on suppose qu'on était à l'équilibre de long terme et que $x_{2t} \nearrow$
 - ⇒ $z_{t-1} = x_{1,t-1} - \beta_2 x_{2,t-1} < 0 \Rightarrow$ déséquilibre de long terme négatif
 - si $\alpha_1 < 0$, $\alpha_1 z_{t-1} > 0$ on observera en t une augmentation de x_{1t}
 - ⇒ cela viendra compenser le déséquilibre
- On s'attend à trouver $\alpha_2 > 0$ car
 - on suppose qu'on était à l'équilibre de long terme et que $x_{1t} \nearrow$
 - ⇒ $z_{t-1} = x_{1,t-1} - \beta_2 x_{2,t-1} > 0 \Rightarrow$ déséquilibre de long terme positif
 - si $\alpha_2 > 0$, $\alpha_2 z_{t-1} > 0$ on observera en t une augmentation de x_{2t}
 - ⇒ cela viendra compenser le déséquilibre

Représentation VAR d'un VECM

- Précédemment, en partant d'un VAR(1) cointégré (modèle 9)

... nous avons obtenu un VECM(0) (modèle 15)

⇒ Il existe une représentation VAR(2) d'un VECM(1)

$$X_t = (I + \Gamma_1 + \alpha\beta')X_{t-1} - \Gamma_1X_{t-2} + \varepsilon_t = A_1X_{t-1} + A_2X_{t-2} + \varepsilon_t$$

- Comme précédemment la remarque de l'équation 16 s'applique

$$\Pi = -(I - A_1 - A_2) = -(I - I - \Gamma_1 - \alpha\beta' + \Gamma_1) = \alpha\beta'$$

- Cela tient pour le passage d'un VECM(p) vers un VAR($p + 1$)

$$\Delta X_t = \Pi X_{t-1} + \sum_{j=1}^p \Gamma_j \Delta X_{t-j} + \varepsilon_t \Rightarrow X_t = \sum_{j=1}^{p+1} A_j X_{t-j} + \varepsilon_t \quad (19)$$

avec $A_1 = I + \Pi + \Gamma_1$, $A_j = \Gamma_j - \Gamma_{j-1}$ et $A_{p+1} = -\Gamma_{p+1}$

Exogénéité faible

- La notion **exogénéité faible** est définie par Engle et al. (1983)

⇒ Une variable X_t sera faiblement exogène

— pour θ un ensemble de paramètre d'intérêt

— si elle peut être prise comme **prédéterminée** dans le modèle

⇒ ne pas décrire le modèle marginal de X_t n'engendre aucune perte d'information

- A titre d'illustration considérons l'équation ECM suivante

$$\Delta y_t = \alpha_1(y_{t-1} - \beta_1 x_{t-1}) + \gamma_1 \Delta x_{t-1} + \varepsilon_t$$

- Si le fait de ne pas modéliser x_t conjointement à y_t n'engendre aucune perte d'information, x_t est **faiblement exogène**

Causalité dans les VECM

Definition 3 (Causalité de court terme)

En considérant le VECM bivarié (x_{1t}, x_{2t}) , x_{1t} ne cause pas x_{2t} à court terme au sens de Granger si et seulement si $\gamma_{21} = 0$ ($\gamma_{12} = 0$)

Definition 4 (Causalité de long terme)

En considérant le VECM bivarié (x_{1t}, x_{2t}) , x_{1t} ne cause pas x_{2t} à long terme au sens de Granger si et seulement si $\alpha_2\beta_{x_1} = 0$ ($\alpha_1\beta_{x_2} = 0$)

- La causalité à court terme implique une contrainte linéaire, elle est donc simple à tester
- La causalité à long terme implique une contrainte non-linéaire, elle est donc difficile à tester

Système triangulaire à la Phillips (1991)

- Phillips (1991) propose une représentation triangulaire
- Dans un cadre à 1 dimension le système est bivarié de la forme

$$\begin{aligned}x_{1t} &= \gamma' \mathbf{x}_{2,t} + \varepsilon_{1t} \\ \Delta \mathbf{x}_{2t} &= \varepsilon_{2t}\end{aligned}$$

avec $\gamma = (\beta_2, \dots, \beta_k)$ et $\mathbf{x}_{2t} = (x_{2t}, \dots, x_{kt})$

- Sous l'hypothèse que ε_{1t} et ε_{2t} sont faiblement dépendants
 - cette représentation est **semi-paramétrique**
 - cette représentation permet une représentation ECM simple
 - ⇒ plus générale que Engle et Granger (1987)

Note Un à priori est fait sur la variable **faiblement exogène**

Système triangulaire de dimension k

- La représentation triangulaire existe pour toute dimension

$$\begin{aligned}X_{1t} &= \Upsilon X_{2t} + \varepsilon_{1t} \\ \Delta X_{2t} &= \varepsilon_{2t}\end{aligned}$$

- Υ une matrice de coefficient $k \times (k - r)$
- X_{1t} est un ensemble de r régresseurs cointégrés avec X_{2t}
- X_{2t} est un ensemble de $k - r$ régresseurs non cointégrés entre eux
- $X_t = (X_{1t}, X_{2t})'$ est un ensemble de k variables aléatoire $I(1)$
- $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ des erreurs faiblement dépendantes

Note Un à priori est fait sur les $k - r$ variables **faiblement exogènes**

Système triangulaire et forme ECM

- La forme ECM des systèmes à la Phillips (1991) s'obtient par différenciation de X_t

$$\Delta X_t = -\Xi \aleph X_{t-1} + v_t$$

avec

$$\Xi = \begin{pmatrix} I_r \\ 0 \end{pmatrix}, \quad \aleph = (I \quad -\Upsilon) \text{ et } v_t = \begin{pmatrix} I & \Upsilon \\ 0 & I \end{pmatrix} \varepsilon_t$$

Approche de Engle et Granger (1987)

- Comment estimer les modèles de cointégration ?
 - commençons par les systèmes cointégrés de dimension 1

- Soit $x_t \sim I(1)$ un processus stochastique d'innovations v_t et

$$y_t = \zeta + \beta x_t + \varepsilon_t \text{ avec } \varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2 < \infty)$$

et $\beta \neq 0$ de sorte que $(y_t, x_t)' \sim CI(1, 1)$

- Il s'agit d'un équilibre de long terme de vecteur cointégrant

$$\beta = \begin{pmatrix} 1 \\ -\beta \end{pmatrix}$$

- L'estimation de β par OLS paraît alors envisageable car
 - $\hat{\beta}$ devrait être consistant si $\beta \neq 0$
 - on sait que $\varepsilon_t \sim I(0)$ si $\beta \neq 0$

Normalisation

- L'exemple précédent repose sur une hypothèse importante

$$- y_t = \beta x_t + \varepsilon_t \text{ et } \beta = (1, -\beta)'$$

⇒ il s'agit d'une normalisation

- On aurait pu considérer $x_t = \theta y_t + v_t$ et $\beta = (1, -\theta)'$

⇒ mais attention, $\theta \neq 1/\beta$

- La normalisation a un impact chez Engle et Granger (1987)

⇒ car les valeurs de $\beta = (\beta_1, \beta_2)'$ tel que $\beta'X_t$ est stationnaire ne sont pas uniques

- L'approche d'Engle et Granger (1987) n'est interprétable facilement que dans un système de dimension 1

Procédure en deux étapes

- Engle et Granger (1987) proposent une estimation en 2 étapes

1 on estime β par OLS

... on récupère les résidus estimés $\widehat{\varepsilon}_t = y_t - \hat{\zeta} - \widehat{\beta}x_t$

2 on teste si $\widehat{\varepsilon}_t \sim I(0)$

- On a vu que si $(y_t, x_t) \sim I(1)$ et $\beta \neq 0$ les OLS sont super-consistants

$$- T(\widehat{\beta} - \beta) \xrightarrow{p} 0$$

- On montre par la suite que

— $\widehat{\beta}$ est biaisé si ε_t et v_t ne sont pas indépendants

— la distribution de $\widehat{\beta}$ étant non-standard, tester $\widehat{\varepsilon}_t \sim I(0)$ implique des valeurs critiques non-standards

Étape 1 de Engle et Granger (1987): intuition

- Si $\beta \neq 0$ on sait qu'il y a cointégration
⇒ au sein du paradigme $I(1)/I(0)$, $\varepsilon_t \sim I(0)$

- Les OLS sont consistants mais quelle est la distribution de $\hat{\beta}$?

$$\mathbf{x}_{1t} = \zeta + \beta \mathbf{x}_{2t} + \varepsilon_t, \quad (20)$$

$$\Delta \mathbf{x}_{2t} = \mathbf{v}_t \quad (21)$$

- Avant de répondre à cette question, précisons que

$$\begin{pmatrix} \varepsilon_t \\ \mathbf{v}_t \end{pmatrix} = \Psi(L)\varepsilon_t, \text{ avec } \mathbb{E}(\varepsilon_t \varepsilon_t') = \Sigma = PP' < \infty$$

et $\varepsilon_t \sim$ i. i. d. , $\Psi(L)$ absolument sommable et $\Psi(1)$ de rang plein

- On permet ici que les éléments non-diagonaux de Σ soient $\neq 0$
⇒ cela introduit 2 problèmes : endogénéité + autocorrélation via $\Psi(L)$

Étape 1 de Engle et Granger (1987): théorie

- Supposons qu'on peut partitionner $\Psi(1) \cdot P$ ainsi

$$\Psi(1) \cdot P = \Lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}$$

avec $\dim \lambda_1 = 1 \times k$ et $\dim \lambda_2 = k - 1 \times k$ pour $k = 2$ ici

- Stock (1987) montre alors que

$$\begin{pmatrix} \sqrt{n}(\hat{\zeta} - \zeta) \\ n(\hat{\beta} - \beta) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} 1 & \int_0^1 \mathbf{W}(r)' dr \cdot \lambda_2' \\ \lambda_2 \int_0^1 \mathbf{W}(r)' dr & \lambda_2 \int_0^1 \mathbf{W}(r) \mathbf{W}(r)' dr \lambda_2' \end{pmatrix}^{-1} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$$

avec $\mathbf{W}(r)'$ un Wiener k -varié, $h_1 := \lambda_1' \mathbf{W}(1)$ et

$$h_2 := \lambda_2 \left(\int_0^1 \mathbf{W}(r) \mathbf{W}(r)' dr \right) \lambda_1 + \sum_{i=0}^{\infty} \mathbb{E}(v_t \varepsilon_{t+i}) \quad (22)$$

- La distribution n'est pas centrée sur 0: **biais de non-centralité**

⇒ donc sauf si $\varepsilon_t \perp v_t$, $\hat{\beta}$ fait face à un biais $O_p(n^{-1})$

Étape 1 de Engle et Granger (1987): intuition si $\varepsilon_t \perp v_t$

- Les résultats de Stock (1987) décrivent le cas général
- Pour voir le lien avec le **Théorème 12** du Chapitre 1 repartons de

$$\begin{aligned}x_{1t} &= \zeta + \beta x_{2t} + \varepsilon_t, \\ \Delta x_{2t} &= v_t\end{aligned}$$

1 Supposons ε_t et v_t i.i.d. avec

$$\begin{pmatrix} \varepsilon_t \\ v_t \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\varepsilon^2 & 0 \\ 0 & \sigma_v^2 \end{pmatrix} \right)$$

2 Supposons également que $\Psi(L) = I$ et

$$\Sigma = PP' \text{ avec } P = \begin{pmatrix} \sigma_\varepsilon & 0 \\ 0 & \sigma_v \end{pmatrix} \Rightarrow \Psi(1) \cdot P = P$$

- On en déduit que $\lambda_1 = (\sigma_\varepsilon \quad 0)$ et $\lambda_2 = (0 \quad \sigma_v)$

Étape 1 de Engle et Granger (1987): théorie si $\varepsilon_t \perp v_t$

- Ces 2 hypothèses éliminent le problème d'endogénéité
- La distribution limite devient alors

$$\begin{pmatrix} \sqrt{n}(\hat{\zeta} - \zeta) \\ n(\hat{\beta} - \beta) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} 1 & \sigma_v \int_0^1 W_2(r) dr \\ \sigma_v \int_0^1 W_2(r) dr & \sigma_v^2 \int_0^1 W_2(r)^2 dr \end{pmatrix}^{-1} \\ \times \begin{pmatrix} \sigma_\varepsilon W_1(1) \\ \sigma_v \sigma_\varepsilon \int_0^1 W_2(r) dr \int_0^1 W_1(r) dr \end{pmatrix}$$

- En se focalisant sur $\hat{\beta}$ on retrouve le résultat du **Théorème 12**

$$n(\hat{\beta} - \beta) \xrightarrow{d} \left(\sigma_v \int_0^1 W_2(r)^2 dr \right)^{-1} \left(\sigma_\varepsilon \int_0^1 W_2(r) W_1(r) dr \right) \quad (23)$$

mais **sans le problème lié à la normalisation par n**

- Dans la pratique, on rencontre rarement une situation si simple
 \Rightarrow nous verrons par la suite comment traiter l'endogénéité

Étape 2 de Engle et Granger (1987): intuition

- Si $\beta = 0$ comment se comporte ε_t dans le système

$$\begin{aligned}x_{1t} &= \beta x_{2t} + \varepsilon_t, \\ \Delta x_{2t} &= v_t\end{aligned}$$

- Afin de répondre à cette question, supposons la régression OLS

$$x_{1t} = \hat{\beta} x_{2t} + \hat{\varepsilon}_t \text{ ou } \hat{\beta}' X_t = \hat{\varepsilon}_t$$

- $\hat{\varepsilon}_t$ devrait être $I(1)$ en l'absence de cointégration donc

$$\hat{\varepsilon}_t = \rho \hat{\varepsilon}_{t-1} + \epsilon_{1t} = \hat{\varepsilon}_{t-1} + \epsilon_{1t}$$

- Dans la pratique on peut estimer ρ par OLS et tester si $\hat{\rho} = 1$

⇒ les résidus étant estimés, quelle sera la distribution de $\hat{\rho}$?

$$\hat{\rho} = \left(\sum_{t=1}^n \hat{\varepsilon}_{t-1}^2 \right)^{-1} \left(\sum_{t=1}^n \hat{\varepsilon}_t \hat{\varepsilon}_{t-1} \right)$$

Étape 2 de Engle et Granger (1987): théorie

- Phillips et Ouliaris (1990) montrent alors que

$$(n-1)(\hat{\rho}-1) \xrightarrow{d} \left[\frac{1}{2} \begin{pmatrix} 1 & -h'_2 \end{pmatrix} \cdot \mathbf{W}(1)\mathbf{W}(1)' \cdot \begin{pmatrix} 1 \\ -h_2 \end{pmatrix} \right. \\ \left. - h_1\mathbf{W}(1)' \begin{pmatrix} 1 \\ -h_2 \end{pmatrix} \right. \\ \left. - \frac{1}{2} \begin{pmatrix} 1 & -h'_2 \end{pmatrix} \mathbf{L}' \left(\mathbb{E}(\Delta X_t \Delta X_t') \right) \mathbf{L} \begin{pmatrix} 1 \\ -h_2 \end{pmatrix} \right] H_n^{-1} \quad (24)$$

avec $(\Lambda\Lambda')^{-1} = \mathbf{L}\mathbf{L}'$, $\mathbf{W}(r) = (W_1(r) \quad W_2(r))'$ et

$$\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} := \begin{pmatrix} 1 & \int_0^1 W_2(r)dr \\ \int_0^1 W_2(r)dr & \int_0^1 W_2(r)^2 dr \end{pmatrix}^{-1} \begin{pmatrix} \int_0^1 W_1(r)dr \\ \int_0^1 W_2(r)W_1(r)dr \end{pmatrix}$$

et finalement

$$H_n = \int_0^1 W_1(r)^2 dr - \begin{pmatrix} \int_0^1 W_1(r)dr & \int_0^1 W_1(r)W_2(r)dr \end{pmatrix} \begin{pmatrix} 1 \\ h_2 \end{pmatrix}$$

Étape 2 de Engle et Granger (1987): théorie si $\varepsilon_t \perp v_t$

- Si $\varepsilon_t \perp v_t$, $\Psi(L) = I$ et $\Lambda\Lambda' = \mathbb{E}(\Delta X_t \Delta X_t')$ ce qui implique

$$L' \left(\mathbb{E}(\Delta X_t \Delta X_t') \right) L = L' \Lambda \Lambda' L = I \text{ car } (\Lambda \Lambda')^{-1} = L L'$$

- La distribution asymptotique de $(n-1)(\hat{\rho} - 1)$ devient alors

$$(n-1)(\hat{\rho} - 1) \xrightarrow{d} \left[\frac{1}{2} \left((1 \quad -h_2') \cdot \mathbf{W}(1) \mathbf{W}(1)' \cdot \begin{pmatrix} 1 \\ -h_2 \end{pmatrix} \right) - h_1 \mathbf{W}(1)' \begin{pmatrix} 1 \\ -h_2 \end{pmatrix} - \frac{1}{2} (1 + h_2' h_2) \right] H_n^{-1}$$

avec $\mathbf{W}(r) = (W_1(r) \quad W_2(r))'$

Correction de Saikkonen (1991) pour l'endogénéité

- L'équation (22) montre qu'en présence d'endogénéité $\hat{\beta}$ est biaisé
- Saikkonen (1991) propose d'ajouter des lags et des leads de Δx_{2t}

$$x_{1t} = \zeta + \beta x_{2t} + \sum_{j=-l}^l \gamma_j \Delta x_{2,t-j} + \tilde{\varepsilon}_t$$

⇒ cela va rendre $\tilde{\varepsilon}_t$ et v_{t-j} orthogonaux pour $j \in \{-l, l\}$ car

$$\varepsilon_t = \sum_{j=-l}^l \varphi_j v_{t-j} + \tilde{\varepsilon}_t$$

- La sélection de l peut se faire à l'aide d'un critère d'information
- Grâce à cette approche **paramétrique**, à présent on a

$$\begin{pmatrix} \tilde{\varepsilon}_t \\ v_t \end{pmatrix} = \tilde{\Psi}(L) \tilde{\varepsilon}_t, \text{ avec } \mathbb{E}(\tilde{\varepsilon}_t \tilde{\varepsilon}_t') = \tilde{\Sigma} = \tilde{P}\tilde{P}' < \infty$$

et $\tilde{\varepsilon}_t \sim \text{i. i. d.}$, $\tilde{\Psi}(L)$ absolument sommable et $\tilde{\Psi}(1)$ de rang plein

Théorie limite des dynamic OLS (DOLS)

- L'approche dite **dynamic OLS** de Saikkonen (1991) assure que

$$\tilde{P} = \begin{pmatrix} \tilde{\sigma}_{11} & 0 \\ 0 & \tilde{\sigma}_{22} \end{pmatrix} \text{ et } \tilde{\Psi}(L) = \begin{pmatrix} \tilde{\psi}_{11}(L) & 0 \\ 0 & \tilde{\psi}_{22}(L) \end{pmatrix}$$

- Pour $\varphi = (\varphi_{-1}, \dots, \varphi_1)'$, Saikkonen (1991) montre que

$$\begin{pmatrix} \sqrt{n}(\hat{\varphi} - \varphi) \\ \sqrt{n}(\hat{\zeta} - \zeta) \\ n(\hat{\beta} - \beta) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \mathbf{Q}^{-1}h_1 \\ \tilde{\lambda}_{11}v_1 \\ \tilde{\lambda}_{11}v_2 \end{pmatrix}, \text{ avec } \mathbf{Q}^{-1} = \mathbb{E}(w_t w_t')$$

et $n^{-1/2} \sum w_t \tilde{\epsilon}_t \rightarrow h_1$ pour $w_t = (v_{t-1}, \dots, v_{t+1})'$

- $\tilde{\lambda}_{11} = \tilde{\psi}_{11}(1)\tilde{\sigma}_{11}$, $\tilde{\lambda}_{22} = \tilde{\psi}_{22}(1)\tilde{\sigma}_{22}$ et $(v_1 \quad v_2)'$ donné par

$$\begin{pmatrix} 1 & \tilde{\lambda}_{22} \int_0^1 W_2(r) dr \\ \tilde{\lambda}_{22} \int_0^1 W_2(r) dr & \tilde{\lambda}_{22}^2 \int_0^1 W_2(r)^2 dr \end{pmatrix}^{-1} \begin{pmatrix} W_1(1) \\ \tilde{\lambda}_{22} \int_0^1 W_2(r) W_1(r) dr \end{pmatrix}$$

Théorie limite des DOLS en absence d'autocorrélation

- Supposons une corrélation contemporaine entre ε_t et v_t

$$\Sigma = PP' \text{ avec } P = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

mais qu'à présent $\tilde{\Psi}(1) = I$ et donc $\tilde{\lambda}_{11} = \tilde{\sigma}_{11}$ et $\tilde{\lambda}_{22} = \tilde{\sigma}_{22}$

- Les **DOLS** nous ramènent ici aux résultats de l'équation (23)
- En effet, en se focalisant sur β on obtient pour $\tilde{\lambda}_{11}v_2$

$$\begin{aligned} \tilde{\lambda}_{11}v_2 &= \left(\tilde{\lambda}_{22}^2 \int_0^1 W_2(r)^2 dr \right)^{-1} \left(\tilde{\lambda}_{11}\tilde{\lambda}_{22} \int_0^1 W_2(r)W_1(r)dr \right) \\ &= \left(\tilde{\sigma}_{22}^2 \int_0^1 W_2(r)^2 dr \right)^{-1} \left(\tilde{\sigma}_{11}\tilde{\sigma}_{22} \int_0^1 W_2(r)W_1(r)dr \right) \\ &= \left(\tilde{\sigma}_{22} \int_0^1 W_2(r)^2 dr \right)^{-1} \left(\tilde{\sigma}_{11} \int_0^1 W_2(r)W_1(r)dr \right) \end{aligned}$$

Correction de Phillips et Hansen (1990) pour l'endogénéité

- L'équation (22) montre qu'en présence d'endogénéité, $\hat{\beta}$ fait face
 - à un biais de **second-ordre** (endogénéité): Ξ_0
 - et un biais de **non-centralité** si autocorrélation sérielle: Ξ_1
- Phillips et Hansen (1990) proposent une double correction car
 - le biais de second-ordre vient des éléments non-diagonaux non-nuls dans $\Sigma = \mathbb{E}(\epsilon_t \epsilon_t')$ venant de $(\epsilon_t, v_t)' = \Psi(L)\epsilon_t$
 - la non-centralité vient de $\Xi_1 = \sum_{i=0}^{\infty} \mathbb{E}(v_t \epsilon_{t+i}) \neq 0$ dans l'Eq. (22)
- Concernant Σ , la correction est simple
 - Dans l'Eq. (20) on soustrait $\Xi_0 = \sigma_{21}^2 (\sigma_{22}^2)^{-1} \Delta x_{2t} = \sigma_{21}^2 (\sigma_{22}^2)^{-1} v_t$
 - On obtient alors l'équation **modifiée**

$$x_{1t}^* = \zeta + \beta^* x_{2t} + \varepsilon_t^* \text{ où } x_{1t}^* = x_{1t} - \Xi_0 \text{ et } \varepsilon_t^* = \varepsilon_t - \Xi_0 \quad (25)$$

Correction de l'endogénéité via Ξ_0

- Dans le système triangulaire (20,21), la correction Ξ_0 donne

$$\begin{pmatrix} \varepsilon_t^* \\ v_t \end{pmatrix} = \boldsymbol{\xi} \begin{pmatrix} \varepsilon_t \\ v_t \end{pmatrix} \text{ avec } \boldsymbol{\xi} = \begin{pmatrix} 1 & -\sigma_{21}^2(\sigma_{22}^2)^{-1} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{pmatrix}$$

- Phillips et Hansen (1990) montrent alors que

$$\begin{pmatrix} n^{1/2}(\hat{\zeta}^* - \zeta) \\ n(\hat{\beta}^* - \beta) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} 1 & \left(\int_0^1 \mathbf{W}(r)' dr \right) \boldsymbol{\Lambda}' \boldsymbol{\xi}_2 \\ \boldsymbol{\xi}_2' \boldsymbol{\Lambda} \int_0^1 \mathbf{W}(r) dr & \boldsymbol{\xi}_2' \boldsymbol{\Lambda} \left(\int_0^1 \mathbf{W}(r) \mathbf{W}(r)' dr \right) \boldsymbol{\Lambda}' \boldsymbol{\xi}_2 \end{pmatrix}^{-1} \\ \times \begin{pmatrix} \boldsymbol{\xi}_1 \boldsymbol{\Lambda} \mathbf{W}(1) \\ \boldsymbol{\xi}_2' \boldsymbol{\Lambda} \left(\int_0^1 \mathbf{W}(r) \mathbf{W}(r)' dr \right) \boldsymbol{\Lambda}' \boldsymbol{\xi}_1 + \boldsymbol{\Xi}_1^* \end{pmatrix} \quad (26)$$

avec $\Psi(1) \cdot P = \boldsymbol{\Lambda}$ sachant $\boldsymbol{\Sigma} = PP'$ et $\boldsymbol{\Xi}_1^*$ donné par

$$\boldsymbol{\Xi}_1^* = \sum_{i=0}^{\infty} \mathbb{E}(v_t \varepsilon_{t+i}^*) = \sum_{i=0}^{\infty} \mathbb{E} \left(v_t (\varepsilon_t - \sigma_{21}^2 (\sigma_{22}^2)^{-1} v_{t+i}) \right)$$

Correction de l'endogénéité via Ξ_1

- L'Eq. (26) fait toujours apparaître le biais de non-centralité, Ξ_1^*
- Phillips et Hansen (1990) proposent un estimateur de Ξ_1^*

1 estimer le système triangulaire (20,21)

2 estimer $\widehat{\Sigma}$ via $\widehat{\Sigma} = \widehat{\Omega}_0 + \sum_{i=1}^q (1 - i(q+1)^{-1})(\widehat{\Omega}_i + \widehat{\Omega}_i')$ avec

$$\widehat{\Omega}_i = n^{-1} \sum_{t=i+1}^n \begin{pmatrix} \widehat{\varepsilon}_t \widehat{\varepsilon}_{t-i} & \widehat{\varepsilon}_t \widehat{v}_{t-i} \\ \widehat{v}_{t-i} \widehat{\varepsilon}_t & \widehat{v}_{t-i} \widehat{v}_{t-i} \end{pmatrix} = \begin{pmatrix} \widehat{\omega}_{11}^{(i)} & \widehat{\omega}_{12}^{(i)} \\ \widehat{\omega}_{21}^{(i)} & \widehat{\omega}_{22}^{(i)} \end{pmatrix} \quad (27)$$

3 on note alors que $\Xi_1^* = \sum_{i=0}^{\infty} \begin{pmatrix} \omega_{12}^{(i)} & \omega_{22}^{(i)} \end{pmatrix} (1 \quad -\sigma_{21}^2 (\sigma_{22}^2)^{-1})'$ car

$$\Xi_1^* = \sum_{i=0}^{\infty} \mathbb{E} \left(v_t \begin{pmatrix} \varepsilon_t & v_{t+i} \end{pmatrix} \right) (1 \quad -\sigma_{21}^2 (\sigma_{22}^2)^{-1})'$$

ce qui donne un estimateur pour Ξ_1^*

$$\widehat{\Xi}_1^* = \sum_{i=0}^q (1 - i(q+1)^{-1}) \begin{pmatrix} \widehat{\omega}_{12}^{(i)} & \widehat{\omega}_{22}^{(i)} \end{pmatrix} (1 \quad -\widehat{\sigma}_{21}^2 (\widehat{\sigma}_{22}^2)^{-1})' \quad (28)$$

Fully modified OLS (FMOLS)

- Via $\hat{\Xi}_0$ et $\hat{\Xi}_1^*$ obtenu par l'Eq. (28) on corrige les 2 biais
 - ⇒ contrairement aux **DOLS** cette correction est non-paramétrique
 - ⇒ on parle de **fully modified OLS** de $\theta^* = (\zeta, \beta)'$

$$\hat{\theta}^* = \left(n^{-1} \sum_{t=1}^n \mathbf{x}_{2t} \mathbf{x}_{2t}' \right)^{-1} \left(n^{-1} \sum_{t=1}^n \mathbf{x}_{2t} \underbrace{(x_{1t} - \hat{\Xi}_0)}_{x_{1t}^*} - n \hat{\Xi}_1^* \right), \quad \mathbf{x}_{2t} = (1, x_{2t})'$$

- Phillips et Hansen (1990) montrent alors que

$$\begin{pmatrix} n^{1/2}(\hat{\zeta}^* - \zeta) \\ n(\hat{\beta}^* - \beta) \end{pmatrix} \xrightarrow{d} \sigma_{11}^* \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \quad (29)$$

avec $(v_1 \quad v_2)'$ donné par

$$\begin{pmatrix} 1 & \sigma_{22} \int_0^1 W_2^*(r) dr \\ \sigma_{22} \int_0^1 W_2^*(r) dr & \sigma_{22}^2 \int_0^1 W_2^*(r)^2 dr \end{pmatrix}^{-1} \begin{pmatrix} W_1^*(1) \\ \sigma_{22} \int_0^1 W_2^*(r) W_1^*(r) dr \end{pmatrix}$$

- Les **FMOLS** nous ramènent également aux résultats de l'Eq. (23)

Correction de Park (1992) pour l'endogénéité

- Park (1992) propose une approche très similaire aux FMOLS
- Le biais de **second-ordre** est corrigé similairement
- Le biais de **non-centralité** est corrigé différemment

— A partir de l'équation (27) de $\widehat{\Omega}$ et $\widehat{\Sigma}$ on définit

$$\widehat{\Lambda} = \widehat{\Omega}_0 + \sum_{i=1}^q (1 - i(q+1)^{-1}) \widehat{\Omega}_i = \begin{pmatrix} \widehat{\lambda}_{11} & \widehat{\lambda}_{12} \\ \widehat{\lambda}_{21} & \widehat{\lambda}_{22} \end{pmatrix} = \begin{pmatrix} \widehat{\lambda}_1 \\ \widehat{\lambda}_2 \end{pmatrix}$$

— On construit les variables **modifiées**

$$\circ \mathbf{x}_{1t}^+ = \mathbf{x}_{1t} - \left(\widehat{\beta}_* \widehat{\lambda}_2 \widehat{\Omega}_0^{-1} + \begin{pmatrix} 0 & \widehat{\sigma}_{12}^2 (\widehat{\sigma}_{22}^2)^{-1} \end{pmatrix} \right) \begin{pmatrix} \widehat{\varepsilon}_{*t} \\ \mathbf{v}_t \end{pmatrix}$$

$$\circ \mathbf{x}_{2t}^+ = \mathbf{x}_{2t} - \widehat{\lambda}_2 \widehat{\Omega}_0^{-1} \begin{pmatrix} \widehat{\varepsilon}_{*t} \\ \mathbf{v}_t \end{pmatrix} \text{ avec } \mathbf{v}_t = \Delta \mathbf{x}_{2t}$$

$$\circ \mathbf{x}_{2t}^+ = (1 \quad \mathbf{x}_{2t}^+)'$$

— $\widehat{\beta}_*$ et $\widehat{\varepsilon}_{*t}$ proviennent d'une estimation OLS naive de l'Eq. (20)

Régression canonique cointégrante (CCR)

- Chez Park (1992) on voit que $\hat{\Xi}_0^+ = \hat{\Xi}_0$ mais $\hat{\Xi}_1^+ = \hat{\lambda}_2 \hat{\Omega}_0^{-1}$

⇒ on parle de **Canonical Cointegration Regression (CCR)**

$$\hat{\theta}^+ = \left(n^{-1} \sum_{t=1}^n \mathbf{x}_{2t}^+ \mathbf{x}_{2t}^{+'} \right)^{-1} \left(n^{-1} \sum_{t=1}^n \mathbf{x}_{2t}^+ \mathbf{x}_{1t}^+ \right) \text{ avec } \theta^+ = (\zeta, \beta)'$$

- Park (1992) montre alors que

$$\begin{pmatrix} n^{1/2}(\hat{\zeta}^+ - \zeta) \\ n(\hat{\beta}^+ - \beta) \end{pmatrix} \xrightarrow{d} \sigma_{11}^+ \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \quad (30)$$

avec $(v_1 \ v_2)'$ donné par

$$\begin{pmatrix} 1 & \sigma_{22}^+ \int_0^1 W_2^+(r) dr \\ \sigma_{22}^+ \int_0^1 W_2^+(r) dr & (\sigma_{22}^+)^2 \int_0^1 W_2^+(r)^2 dr \end{pmatrix}^{-1} \begin{pmatrix} W_1^+(1) \\ \sigma_{22}^+ \int_0^1 W_2^+(r) W_1^+(r) dr \end{pmatrix}$$

- Les **CCR** nous ramènent également aux résultats de l'Eq. (23)

Bilan des estimateurs dans l'approche d'Engle et Granger

- Chez Engle et Granger (1987), l'estimation de β est cruciale
 - Nous avons analysé 4 estimateurs
 - **OLS**: biaisé en présence d'endogénéité
 - **DOLS**: corrige paramétriquement le biais d'endogénéité
 - **FMOLS**: corrige non-paramétriquement le biais d'endogénéité
 - **CCR**: corrige non-paramétriquement le biais d'endogénéité
 - Kurozumi et Hayakawa (2009) rappellent et démontrent que
 - **DOLS**, **FMOLS** et **CCR** sont asymptotiquement équivalents
 - **DOLS**, **FMOLS** et **CCR** subissent un biais
 - si l'autocorrélation sérielle des erreurs est très forte
 - pouvant ramener leurs distributions limite à celle des **OLS**
- ⇒ ce biais apparaît donc en échantillon fini

Approche de Johansen (1991)

- Soit un VECM(p) pour X_t un vecteur de k variables $I(1)$

$$\Delta X_t = \Xi + \Pi X_{t-1} + \sum_{j=1}^p \Gamma_j \Delta X_{t-j} + \varepsilon_t \quad (31)$$

- Johansen (1991) propose une procédure d'estimation de

$$\theta = (\Xi, \Gamma_1, \dots, \Gamma_p, \Pi, \Sigma)$$

- le paramètre Ξ (constante)
 - les paramètres de court terme $\Gamma_1, \dots, \Gamma_p$
 - les paramètres de long terme α et β pour $\Pi = \alpha\beta$
 - les paramètres de Σ sous l'hypothèse que $\varepsilon_t \sim \mathcal{N}(\mathbf{0}_{k \times 1}, \Sigma)$
- Cette procédure est basée sur une **analyse canonique**
 - Revient à implémenter un **maximum de vraisemblance**

L'analyse canonique: introduction

- Soit \mathbf{y} et \mathbf{x} des variables dont on ignore le degré de corrélation
- L'**analyse canonique** permet de rechercher les combinaisons linéaires entre \mathbf{y} et \mathbf{x} les plus corrélées possibles
- Ces combinaisons forment des **variables canoniques**
- Formellement, soient \mathbf{y} et \mathbf{x} deux vecteurs de variables tels que

$$\dim(\mathbf{y}) = k_y \times 1 \text{ et } \dim(\mathbf{x}) = k_x \times 1$$

⇒ La matrice de covariance jointe résume l'ensemble des corrélations

$$\Sigma = \begin{pmatrix} \mathbb{E}(\mathbf{y}\mathbf{y}') & \mathbb{E}(\mathbf{y}\mathbf{x}') \\ \mathbb{E}(\mathbf{x}\mathbf{y}') & \mathbb{E}(\mathbf{x}\mathbf{x}') \end{pmatrix} = \begin{pmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix}$$

- Σ est bien sur complexe à étudier: solution ?

L'analyse canonique: intuition

- L'**analyse canonique** permet une approche parcimonieuse
- Invoquons deux nouveaux vecteurs η_t et ζ_t tels que

$$\eta_t = \mathcal{V}'\gamma_t \text{ et } \zeta_t = \mathcal{A}'x_t$$

avec $\dim(\mathcal{V}) = k_\wedge \times k_y$ et $\dim(\mathcal{A}) = k_\wedge \times k_x$ et $k_\wedge = k_y \wedge k_x$

- η_t et ζ_t constituent les **variables canoniques**

e.g. Le i -ième élément de η_t est un combinaison linéaire de γ_t pondérée par la i -ième ligne de \mathcal{V}'

- \mathcal{V} et \mathcal{A} sont choisis telles que η_t et ζ_t sont de variances unitaires et leurs éléments respectifs sont non-corrélés entre eux:

$$\mathbb{E}(\eta_t \eta_t') = \mathcal{V}' \Sigma_{\gamma\gamma} \mathcal{V} = I \text{ et } \mathbb{E}(\zeta_t \zeta_t') = \mathcal{A}' \Sigma_{xx} \mathcal{A} = I$$

$\Rightarrow \mathcal{R}_{k_\wedge \times k_\wedge} = \mathcal{A}' \Sigma_{xy} \mathcal{V} = \mathbb{E}(\zeta_t \eta_t')$ résume les $\text{corr}(\eta_i, \zeta_j)$ pour $i = j$

- On peut ordonner η_t et ζ_t tel que $(1 \geq r_1 \geq \dots \geq r_{k_\wedge} \geq 0)$

\Rightarrow l'élément r_i de $\text{diag}(\mathcal{R})$ est la i ème **corrélation canonique**

Rappel sur le problème des valeurs propres

- Soit une matrice carrée \mathcal{M} avec $\dim(\mathcal{M}) = k \times k$
- Les solutions de la représentation

$$\mathcal{M}\mathbf{v} = \lambda\mathbf{v} \text{ ou } (\mathcal{M} - \lambda I)\mathbf{v} = 0$$

donne λ un scalaire nommé valeur propre et \mathbf{v} un vecteur propre

- Rappelons que cette représentation admet une solution ssi

$$\det(\mathcal{M} - \lambda I) = 0$$

- Il s'agit d'un polynôme caractéristique qui se factorise comme

$$\det(\mathcal{M} - \lambda I) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_k - \lambda)$$

avec $\lambda_1, \dots, \lambda_k$ les k valeurs propres associées à \mathcal{M}

- Dans le cadre de la régression canonique, déterminer \mathcal{V} , \mathcal{A} et donc \mathcal{R} , peut se faire à partir des valeurs propres de $\Sigma_{\gamma\gamma}$, Σ_{xx} et Σ_{xy}

L'analyse canonique: fonctionnement

- On note $\lambda_1, \dots, \lambda_{k_y}$ et μ_1, \dots, μ_{k_x} les valeurs propres et v_1, \dots, v_{k_y} et $\alpha_1, \dots, \alpha_{k_x}$ les vecteurs propres de

$$\left(\Sigma_{yy}^{-1}\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}\right)v_i = \lambda_i v_i \text{ et } \left(\Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\right)\alpha_i = \mu_i \alpha_i$$

- Notons que si v_i est un vecteur propre, $c v_i$ l'est aussi $\forall c$
- Dans l'analyse canonique on choisit c tel que

$$\tilde{v}_i' \Sigma_{yy} \tilde{v}_i = 1 \text{ avec } \tilde{v}_i = v_i / \left(\sqrt{v_i' \Sigma_{yy} v_i} \right), \quad \forall i$$

- \mathcal{V} et \mathcal{A} sont alors obtenus grâce aux vecteurs propres normalisés

$$\mathcal{V}' = (\tilde{v}_1, \dots, \tilde{v}_{k_y}) \text{ et } \mathcal{A}' = (\tilde{\alpha}_1, \dots, \tilde{\alpha}_{k_x})$$

- Les corrélations canoniques sont simplement déduites de

$$\mathcal{R}^2 = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_{k \wedge} \end{pmatrix} \Rightarrow r_i = \sqrt{\lambda_i}$$

Approche de Johansen (1991)

- On part d'un VECM(p) pour X_t un vecteur de k variables $I(1)$

$$\Delta X_t = \Xi + \Pi X_{t-1} + \sum_{j=1}^p \Gamma_j \Delta X_{t-j} + \varepsilon_t \text{ avec } \varepsilon_t \sim \mathcal{N}(\mathbf{0}_{k \times 1}, \Sigma)$$

et $\theta = (\Xi, \Gamma_1, \dots, \Gamma_p, \Pi, \Sigma)$

- L'hypothèse de normalité mène à la **log-vraisemblance** suivante

$$\mathcal{L}(\theta; X_{2-p}, X_{3-p}, \dots, X_0) = -\frac{nk}{2} \log(2\pi) - \frac{n}{2} \log |\Sigma| - \frac{1}{2} \sum_{t=1}^n \left(\varepsilon_t' \Sigma^{-1} \varepsilon_t \right)$$

avec $\varepsilon_t = \Delta X_t - \Xi - \Pi X_{t-1} - \sum_{j=1}^p \Gamma_j \Delta X_{t-j}$

- A l'aide de l'analyse canonique, Johansen (1991) développe une procédure en 3 étapes pour obtenir

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \mathcal{L}(\theta; X_{2-p}, X_{3-p}, \dots, X_0)$$

Johansen (1991): étape 1 (régression auxiliaire)

- On construit une première régression auxiliaire

$$\Delta X_t = \Omega_0 + \sum_{j=1}^p \Omega_j \Delta X_{t-j} + \omega_t \quad (32)$$

où $\hat{\theta}_\omega = (\hat{\Omega}_0, \hat{\Omega}_1, \dots, \hat{\Omega}_p)$ et $\hat{\omega}_t$ sont obtenus par OLS

- On construit une seconde régression auxiliaire

$$X_{t-1} = \Psi_0 + \sum_{j=1}^p \Psi_j \Delta X_{t-j} + \psi_t \quad (33)$$

où $\hat{\theta}_\psi = (\hat{\Psi}_0, \hat{\Psi}_1, \dots, \hat{\Psi}_p)$ et $\hat{\psi}_t$ sont obtenus par OLS

Johansen (1991): étape 2 (corrélation canonique)

- A partir de $\hat{\omega}_t$ et $\hat{\psi}_t$, on peut construire

$$\hat{\Sigma}_{\omega\omega} = \frac{1}{n} \sum_{t=1}^n \hat{\omega}_t \hat{\omega}_t', \quad \hat{\Sigma}_{\psi\psi} = \frac{1}{n} \sum_{t=1}^n \hat{\psi}_t \hat{\psi}_t', \quad \hat{\Sigma}_{\psi\omega} = \hat{\Sigma}_{\omega\psi} = \frac{1}{n} \sum_{t=1}^n \hat{\omega}_t \hat{\psi}_t'$$

- A l'aide de la régression canonique, déterminer les valeurs propres ordonnées $\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_k$ de

$$\hat{\Sigma}_{\psi\psi}^{-1} \hat{\Sigma}_{\psi\omega} \hat{\Sigma}_{\omega\omega}^{-1} \hat{\Sigma}_{\omega\psi}$$

- Johansen (1991) montre que $\mathcal{L}(\theta; \cdot)$ atteint un maximum pour

$$\mathcal{L}^* = -\frac{nk}{2} \log(2\pi) - \frac{n}{2} \log |\hat{\Sigma}_{\omega\omega}| - \frac{1}{2} \sum_{i=1}^r \log(1 - \hat{\lambda}_i)$$

avec r le **nombre de relations de cointégration**

⇒ Possibilité d'un test \mathcal{LR} sur le nombre de relations de cointégration

Johansen (1991): étape 3 (MLE)

- L'analyse canonique a révélé les r premiers vecteurs propres normalisés et ordonnés

$\widehat{\mathcal{V}}_r' = (\widehat{v}_1, \widehat{v}_2, \dots, \widehat{v}_r)$ de

$$(\widehat{\Sigma}_{\psi\psi}^{-1} \widehat{\Sigma}_{\psi\omega} \widehat{\Sigma}_{\omega\omega}^{-1} \widehat{\Sigma}_{\omega\psi}) \widehat{v}_i = \widehat{\lambda}_i \widehat{v}_i$$

- Johansen (1991) montre alors que

- 1 Le MLE de Π est donné par

$$\widehat{\Pi} = \widehat{\Sigma}_{\omega\omega} \widehat{\mathcal{V}} \widehat{\mathcal{V}}' = \widehat{\alpha} \widehat{\beta}' \text{ avec } \widehat{\alpha} = \widehat{\Sigma}_{\psi\omega} \widehat{\mathcal{V}}_r \text{ et } \widehat{\beta} = \widehat{\mathcal{V}}_r$$

- 2 Le MLE de Γ_i pour $i = 1, 2, \dots, p$ est donné par

$$\widehat{\Gamma}_i = \widehat{\Omega}_i - \widehat{\Pi} \widehat{\Psi}_i$$

- 3 Le MLE de Ξ est donné par

$$\widehat{\Xi} = \widehat{\Omega}_0 - \widehat{\Pi} \widehat{\Psi}_0$$

- 4 Le MLE de Σ est donné par

$$\widehat{\Sigma} = n^{-1} \sum_{t=1}^n (\widehat{\omega}_t - \widehat{\Pi} \widehat{\psi}_t)(\widehat{\omega}_t - \widehat{\Pi} \widehat{\psi}_t)'$$

Test de trace de Johansen (1988)

- On a vu que $\mathcal{L}(\theta; \cdot)$ atteint un maximum pour

$$\mathcal{L}^* = -\frac{nk}{2} \log(2\pi) - \frac{n}{2} \log |\hat{\Sigma}_{\omega\omega}| - \frac{1}{2} \sum_{i=1}^r \log(1 - \hat{\lambda}_i)$$

avec r le nombre de relations de cointégration

- Définissons que sous H_0 il y a r relations de cointégration
- Définissons que sous H_1 il y a $r = k$ relations de cointégration

$$\mathcal{L}_{H_1}^* = -\frac{nk}{2} \log(2\pi) - \frac{n}{2} \log |\hat{\Sigma}_{\omega\omega}| - \frac{1}{2} \sum_{i=1}^k \log(1 - \hat{\lambda}_i)$$

- Un test du ratio de vraisemblance est alors donné par

$$\mathcal{LR}_g^{(\tau)} = 2(\mathcal{L}_{H_1}^* - \mathcal{L}_{H_0}^*) = -n \sum_{i=r+1}^k \log(1 - \hat{\lambda}_i)$$

avec $g = k - r$, le nombre de racines unitaires dans le VECM

Distribution limite du test de trace

- Considérons le modèle VECM(p) avec $\Xi = 0$

$$\Delta X_t = \Pi X_{t-1} + \sum_{j=1}^p \Gamma_j \Delta X_{t-j} + \varepsilon_t \text{ avec } \varepsilon_t \sim \mathcal{N}(\mathbf{0}_{k \times 1}, \Sigma)$$

- La distribution limite de $\mathcal{LR}_g^{(\tau)}$ est donnée par la **trace** de

$$\mathbf{Q} = \left(\int_0^1 \mathbf{W}(r) d\mathbf{W}(r)' \right)' \left(\int_0^1 \mathbf{W}(r) \mathbf{W}(r)' dr \right)^{-1} \left(\int_0^1 \mathbf{W}(r) d\mathbf{W}(r)' \right)$$

avec $\mathbf{W}(r)$ un Wiener de dimension g

- Dans le cas $g = 1$, la statistique \mathbf{Q} devient un scalaire

$$Q = \frac{\left(\int_0^1 W(r) dW(r) \right)^2}{\int_0^1 W(r)^2} = \frac{(1/2)^2 \left((W(1))^2 - 1 \right)^2}{\left(\int_0^1 W(r)^2 dr \right)}$$

⇒ Il s'agit de $(\hat{t}_\rho)^2$ dans la théorie limite des OLS non-standard !!

Test $\lambda - \max$ de Johansen (1988)

- Dans le même esprit que le test de trace on peut considérer

$$\mathcal{L}_{H_1}^* = -\frac{nk}{2} \log(2\pi) - \frac{n}{2} \log |\widehat{\Sigma}_{\omega\omega}| - \frac{1}{2} \sum_{i=1}^{r+1} (1 - \widehat{\lambda}_i)$$

- Un test du ratio de vraisemblance est alors donné par

$$\mathcal{LR}_g^{(\lambda)} = 2(\mathcal{L}_{H_1}^* - \mathcal{L}_{H_0}^*) = -n \log(1 - \widehat{\lambda}_{r+1})$$

- La distribution limite de $\mathcal{LR}_g^{(\lambda)}$ est donnée par la **valeur propre la plus large** de la matrice \mathbf{Q}
- Pour $g = 1$, les deux tests sont identiques
- Notons que le test de trace a plus de **puissance** que le test $\lambda - \max$

Méthodologie d'application des tests de Johansen (1988)

- Les deux tests s'appliquent séquentiellement pour $r = 0, 1, 2, \dots$

⇒ soit le test de trace pour un VECM trivarié ($k = 3$)

$r = 0$ calculer $\mathcal{LR}_g^{(\tau)}$ sous $H_0 : \text{rank}(\Pi) = 0$ contre $H_1 : \text{rank}(\Pi) = 3$

⇒ si on rejette H_0 , on rejette l'absence de cointégration

$r = 1$ calculer $\mathcal{LR}_g^{(\tau)}$ sous $H_0 : \text{rank}(\Pi) = 1$ contre $H_1 : \text{rank}(\Pi) = 3$

⇒ si on rejette H_0 , le nombre de relation de cointégration > 1

$r = 2$ calculer $\mathcal{LR}_g^{(\tau)}$ sous $H_0 : \text{rank}(\Pi) = 2$ contre $H_1 : \text{rank}(\Pi) = 3$

⇒ si on rejette H_0 , le VECM est un VAR stationnaire

⇒ si on ne rejette pas H_0 , il y a 2 relations de cointégration

Le cas d'un CVAR(1) trivarié

- Considérons le modèle VECM(0) associé à un CVAR(1) trivarié
- Supposons que le test de trace nous donne $r = 2$

$$\Delta X_t = \Pi X_{t-1} + \varepsilon_t = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{pmatrix} \begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \end{pmatrix} X_{t-1} + \varepsilon_t$$

- Π n'étant pas unique, une normalisation naturelle vient : $\beta_{11} = \beta_{22} = 1$

$$\begin{aligned} \Delta x_{1t} &= \alpha_{11}^* (y_{1,t-1} + \beta_{12}^* y_{2,t-1} + \beta_{13}^* y_{3,t-1}) \\ &\quad + \alpha_{12}^* (\beta_{21}^* y_{1,t-1} + y_{2,t-1} + \beta_{31}^* y_{3,t-1}) + \varepsilon_{1t} \end{aligned}$$

$$\Delta x_{2t} = \alpha_{21}^* z_{1t-1} + \alpha_{22}^* z_{2t-1} + \varepsilon_{2t}$$

$$\Delta x_{3t} = \alpha_{31}^* z_{1t-1} + \alpha_{32}^* z_{2t-1} + \varepsilon_{3t}$$

où les * rappellent que la normalisation impacte les coefficients de Π

$$\alpha_{11}^* = \beta_{11} \alpha_{11}, \quad \beta_{12} = \beta_{12} / \beta_{11}, \quad \beta_{13} = \beta_{13} / \beta_{11}$$

Note Normaliser ne suffit pas à identifier, il manquera une restriction dans chaque vecteur de cointégration : exclusion ou restriction de signe

VECM et termes déterministes

- Jusqu'à présent nous avons omis la présence de termes déterministes
- Mais on a pu voir qu'ils sont impactant dans le cadre non-stationnaire

⇒ la théorie limite des tests est affectée (e.g. ADF)

- Dans le cadre multivarié non-stationnaire on distingue 3 situations
 - Les termes déterministes sont dans l'espace de cointégration
 - Les termes déterministes sont hors de l'espace de cointégration
 - Les termes déterministes sont partout

⇒ Cela va mener à 5 types de modèle

Cas 1 : aucun termes déterministes

- Sous forme vectorielle, considérons un VECM(p) bivarié

$$\Delta X_t = \alpha \beta' X_{t-1} + \sum_{i=1}^p \Gamma_i \Delta X_{t-i} + \varepsilon_t$$

dont l'écriture par équations donne

$$\Delta x_{1t} = \alpha_1(x_{1,t-1} - \beta_2 x_{2,t-1}) + \sum_{i=1}^p \gamma_{11,i} \Delta x_{1,t-i} + \sum_{i=1}^p \gamma_{12,i} \Delta x_{2,t-i} + \varepsilon_{1t}$$

$$\Delta x_{2t} = \alpha_2(x_{1,t-1} - \beta_2 x_{2,t-1}) + \sum_{i=1}^p \gamma_{21,i} \Delta x_{1,t-i} + \sum_{i=1}^p \gamma_{22,i} \Delta x_{2,t-i} + \varepsilon_{2t}$$

⇒ Ce modèle ne pose donc aucune complication

Note Ce modèle étant bivarié, $k = 2$ et $r = 1$ (on suppose la cointégration)

Cas 2 : constante dans l'espace $CI(1, 1)$

- Soit $Y_t \sim CI(1, 1)$ un processus d'espérance $\mu_0 \neq 0$ et définissons

$$X_t = Y_t - \mu_0$$

avec bien entendu $\Delta X_t = \Delta Y_t$

- La forme VECM(p) bivarié sera

$$\begin{aligned}\Delta X_t &= \alpha \beta' (Y_{t-1} - \mu_0) + \sum_{i=1}^p \Gamma_i \Delta X_{t-i} + \varepsilon_t \\ &= \alpha \beta'_0 (Y_{t-1}, 1)' + \sum_{i=1}^p \Gamma_i \Delta X_{t-i} + \varepsilon_t\end{aligned}$$

où $\beta'_0 = (\beta' : \tau')$ avec $\tau = -\beta \mu_0$ un vecteur $r \times 1$ (ici un scalaire)

- On voit alors que $\mu_0 \neq 0$ implique une constante dans l'espace cointégré
- ... ou une constante hors de l'espace cointégré mais dont la forme est

$$\Delta X_t = \nu + \alpha \beta' Y_{t-1} + \sum_{i=1}^p \Gamma_i \Delta X_{t-i} + \varepsilon_t$$

avec $\nu = \alpha \tau'$ uniquement

Cas 3 : constante et tendance dans l'espace $CI(1, 1)$

- Soit, de $\mu_1 t$, $Y_t \sim CI(1, 1)$ un processus d'espérance $\mu_0 \neq 0$ tel que

$$X_t = Y_t - \mu - \mu_1 t$$

avec bien entendu $\Delta X_t = \Delta Y_t - \mu_1$

- La forme VECM(p) bivarié sera

$$\Delta X_t = \alpha \beta' (Y_{t-1} - \mu_0 - \mu_1(t-1)) + \sum_{i=1}^p \Gamma_i \Delta X_{t-i} + \varepsilon_t$$

- Cette fois, $\mu_1 t \neq 0$ implique une tendance dans l'espace cointégré

... et une constante hors de l'espace cointégré puisque

$$\Delta Y_t = \nu + \alpha (\beta' : \eta') (Y_{t-1}, t-1)' + \sum_{i=1}^p \Gamma_i \Delta Y_{t-i} + \varepsilon_t$$

où $\nu = -\Pi \mu_0 + (I_k - \Gamma_1 - \dots - \Gamma_{p-1}) \mu_1$ et $\eta' = -\beta' \mu_1$ (ici un scalaire)

Note Avec $\beta = (\beta_1, \beta_2)$ et η un scalaire, $(\beta' : \eta) = (\beta_1, \beta_2, \eta)'$

Cas 4 : constante hors de l'espace $CI(1, 1)$

- Sur la base du cas 2, on voit que

$$\Delta X_t = \nu_0 + \alpha\beta'Y_{t-1} + \sum_{i=1}^p \Gamma_i \Delta X_{t-i} + \varepsilon_t$$

ne correspond à une constante dans l'espace cointégré que si $\nu_0 = \alpha\tau'$

- Si en revanche on ne restreint pas ν_0 , cela engendrera un trend dans X_t

... Mais cette tendance existera en dehors de l'espace de cointégration

Note Cela implique aussi que $\text{rank}(\Pi) < k$ car si $r = k$, le VECM est un VAR stable ce qui est incompatible avec une tendance

Cas 5 : tendance hors de l'espace $CI(1, 1)$

- Sur la base du cas 3, on voit que

$$\Delta X_t = \nu_0 + \nu_1(\mathbf{0}, t-1)' + \alpha\beta'Y_{t-1} + \sum_{i=1}^p \Gamma_i \Delta X_{t-i} + \varepsilon_t$$

ne correspond à une tendance dans l'espace cointégré que si

$$\nu_0 = -\Pi\mu_0 + (I_k - \Gamma_1 - \dots - \Gamma_{p-1})\mu_1$$

et

$$\nu_1 = \alpha(\beta' : \eta')$$

- Si en revanche on ne restreint pas

$$\nu_0 \text{ et } \nu_1$$

cela engendrera une tendance quadratique dans X_t

... Mais cette tendance existera en dehors de l'espace de cointégration

Note Cela implique aussi que $\text{rank}(\Pi) < k$ car si $r = k$, le VECM est un VAR stable ce qui est incompatible avec une tendance

Définition de la cointégration fractionnaire

- Repartons de la définition d'Engle et Granger (1987)

Definition 5 (Cointégration selon Engle et Granger, 1987)

Les éléments du vecteur de variables aléatoires $X_t = (x_{1,t}, \dots, x_{k,t})$ sont dits cointégrés d'ordres (δ, b) , noté $CI \sim (\delta, b)$ si

tous les éléments du vecteur X_t sont $I(\delta)$

il existe un vecteur $\beta \neq 0$ tel que $\varepsilon_t = \beta' X_t \sim I(\delta - b)$ avec $b > 0$

Le vecteur β est de dimension k et se nomme vecteur cointégrant.

- Pour $\delta, b \in \mathbb{R}$ on parle de **cointégration fractionnaire**

Cointégration, forte, faible et stationnaire

- Selon les sous-espaces de δ et $\gamma = \delta - b$ on distingue
 - la **cointégration forte**
 - si $\delta > 1/2$ et $\delta - \gamma > 1/2$ (i.e. $b > 1/2$)
 - ⇒ ε_t est stationnaire et les déviations à l'équilibre de long terme sont peu persistantes relativement à X_t
 - la **cointégration faible**
 - si $\delta \begin{matrix} \leq \\ > \end{matrix} 1/2$ et $\delta - \gamma < 1/2$ (i.e. $b < 1/2$)
 - ⇒ ε_t est très persistant relativement à X_t impliquant de larges et durables déviations à l'équilibre de long terme
 - la **cointégration stationnaire**
 - si $\delta < 1/2$ et $\delta - \gamma < 1/2$ (i.e. $b < 1/2$)
 - ⇒ ε_t et X_t sont **stationnaires** donc les déviations à l'équilibre de long terme sont persistantes relativement à X_t
 - ⇒ il s'agit d'un sous-cas très particulier de **cointégration faible**

VAR fractionnairement cointégré

- En résonance avec les ARFIMA, on peut définir des VARFIMA
- Leur adaptation au contexte de la cointégration mène au FCVAR
- Ce modèle fut proposé par Johansen (2008)

$$\Delta^d X_t = \alpha \Delta^{d-b} L_b(\beta' X_t + \rho') + \sum_{i=1}^k \Phi_i \Delta^d L_b^i X_t + \varepsilon_t$$

où $L_b = 1 - \Delta^b$

- Afin de faciliter l'estimation, Johansen et Nielsen (2012) s'intéressent à

$$\Delta^d (X_t - \mu) = \alpha \beta' \Delta^{d-b} L_b (X_t - \mu) + \sum_{i=1}^k \Phi_i \Delta^d L_b^i (X_t - \mu) + \varepsilon_t, \quad \rho = \beta' \mu$$

- Ces auteurs montrent que le MLE Gaussien est consistant et de la forme

$$\log \mathcal{L}(\theta) = -k \frac{n}{2} \log(2\pi) - k \frac{n}{2} - n \frac{1}{2} \log \det \left(n^{-1} \sum_{t=1}^n \varepsilon_t(\theta) \varepsilon_t'(\theta) \right)$$

mais $\hat{\beta}$ et $\hat{\rho}$ sont asymptotiquement normaux uniquement quand $b_0 < 1/2$ et localement asymptotiquement mix-normaux quand $b_0 > 1/2$

Ce qu'il faut retenir

- La cointégration permet d'éviter les régressions factices
 - ⇒ si une **combinaison linéaire** de séries $I(1)$ est $I(0)$, elles partagent un équilibre de long terme commun
- Il existe 2 principales représentations des systèmes cointégrés
 - la représentation triangulaire à la Phillips (1991)
 - la représentation vectorielle à correction d'erreur (VECM)
- Il existe 2 principales méthodologies d'estimation
 - l'approche en 2 étapes à la Engle et Granger (1987)
 - estimation de la relation de long terme: $\hat{\beta}$
 - test sur les résidus de la relation: $I(1)$ v.s. $I(0)$
 - l'approche en système à la Johansen (1991)
 - estimation du VECM par maximum de vraisemblance
 - test sur la matrice Π pour détecter les relations de long terme

References

-  Engle, R. F., Granger, C. W. J. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica*, 55(2), 251–276.
-  Engle, R. F., Hendry, D., Richard, J. (1983). Exogeneity. *Econometrica*
-  Granger, C. W. J. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, 16, 121–130.
-  Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 1551-1580.
-  Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of economic dynamics and control* 12, 231-254.
-  Hansen, P. R. (2005). Granger's representation theorem: A closed-form expression for I (1) processes. *The Econometrics Journal*, 8(1), 23-38.
-  Hamilton, J. D. (1994). *Time series analysis (Vol. 2)*. Princeton: Princeton university press.
-  Kurozumi, E., Hayakawa, K., 2009. Asymptotic properties of the efficient estimators for cointegrating regression models with serially dependent errors. *Journal of Econometrics* 149, 118-135.
-  Park, J. (1992). Canonical cointegrating regressions. *Econometrica*, 60(1), 119–143.
-  Phillips, P. C. (1991). Optimal inference in cointegrated systems. *Econometrica: Journal of the Econometric Society*, 283-306.
-  Phillips, P., Hansen, B. (1990). Statistical inference in instrumental variables regression with I (1) processes. *The Review of Economic Studies*, 57, 99–125.
-  Phillips, P., Ouliaris, S., 1990. Asymptotic properties of residual based tests for cointegration. *Econometrica: Journal of the Econometric Society* 58, 165-193.
-  Saikkonen, P. (1991). Asymptotically efficient estimation of cointegration regressions. *Econometric theory*, 7(1), 1-21.
-  Stock, J. H. (1987). Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica: Journal of the*

Thank you!