

# Processus vectoriels stationnaires

## Processus VAR stationnaires

Inspiré du cours de C. Doz

Mars-Avril 2003

### Introduction

On peut s'intéresser aux généralisations vectorielles des MA et des ARMA. Mais en pratique, on s'est souvent ramené dans le cas univarié à des processus AR. C'est d'autant plus le cas dans les processus vectoriels, où les VARMA donnent lieu à des calculs de polynômes matriciels. On va donc considérer principalement les processus VAR, généralisation vectorielle des processus AR. Cependant, on évoquera des processus VMA( $\infty$ ) dans le cadre d'une généralisation du théorème de Wold.

Les modèles VAR servent à la prévision, au test de relations entre variables, sans modélisation économique *a priori*, car on ne sépare pas les endogènes et les exogènes.

**Exemple 0.1.** On peut réaliser des tests de causalité (mal nommés), des tests d'exogénéité, de modèles structurels.

### Table des matières

<b>1</b>	<b>Processus vectoriels stationnaires</b>	<b>2</b>
1.1	Définition et propositions générales . . . . .	2
1.2	Densité spectrale . . . . .	4
1.3	Innovation et théorème de Wold . . . . .	5
1.4	Convergence des moments empiriques . . . . .	7
<b>2</b>	<b>Processus VAR stationnaires</b>	<b>7</b>
2.1	Définition et propriétés générales . . . . .	7
2.2	Prévision d'un processus VAR stationnaire . . . . .	9
<b>3</b>	<b>Estimation d'un modèle VAR</b>	<b>11</b>
3.1	Écriture du modèle . . . . .	11
3.2	Estimation du modèle empilé . . . . .	12
3.3	EMV sous l'hypothèse de normalité . . . . .	14
3.4	Propriétés de l'EMV sous l'hypothèse de normalité . . . . .	15
3.5	Tests linéaires sur les paramètres du modèle . . . . .	15
3.5.1	Test de Wald . . . . .	15
3.5.2	Test du rapport de vraisemblance . . . . .	16

# 1 Processus vectoriels stationnaires

## 1.1 Définition et propositions générales

**Définition 1 (Processus vectoriel).**  $(X_t)_{t \in \mathbb{Z}}$  est un processus vectoriel stationnaire à valeurs dans  $\mathbb{R}^n$  signifie qu'il s'écrit :

$$X_t = \begin{pmatrix} x_{1,t} \\ \vdots \\ x_{n,t} \end{pmatrix}$$

où les  $(x_{i,t})_{t \in \mathbb{Z}}$  sont des processus à valeurs dans  $\mathbb{R}$ .

**Définition 2 (Processus vectoriel stationnaire).**  $(X_t)_{t \in \mathbb{Z}}$  est un processus vectoriel stationnaire s'il vérifie :

- (i)  $\mathbb{E}(X_t) = m, \quad \forall t \in \mathbb{Z}$
- (ii)  $\mathbb{V}(x_t) = \Gamma_0 \quad \forall t \in \mathbb{Z}$ , où :

$$\begin{aligned} \mathbb{V}(x_t) &= \mathbb{E}[(X_t - m)(X_t - m)'] \\ &= (\text{Cov}(x_{i,t}, x_{j,t}))_{i,j} \\ &= (\sigma_{i,j,t})_{i,j} \end{aligned}$$

- (iii)  $\mathbb{E}[(X_t - m)(x_{t+h} - m)'] = \gamma(h) \quad \forall t \in \mathbb{Z}$

*Remarque.*

$$\begin{aligned} &\mathbb{E}[(X_t - m)(X_{t+h} - m)'] \\ &= \mathbb{E} \left[ \begin{pmatrix} x_{1,t} - m_1 \\ \vdots \\ x_{n,t} - m_n \end{pmatrix} (x_{1,t+h} - m_1, \dots, x_{n,t+h} - m_n) \right] \\ &= [\mathbb{E}((x_{i,t} - m_i)(x_{j,t+h} - m_j)')]_{i,j} \\ &= (\text{Cov}(x_{i,t}, x_{j,t+h}))_{i,j} \end{aligned}$$

On a donc :

$$\forall t \in \mathbb{Z}, \forall (i, j), \text{Cov}(x_{i,t}, x_{j,t-h}) = \gamma_{i,j}(h)$$

**Proposition 1.1.**  $(X_t)_{t \in \mathbb{Z}}$  stationnaire  $\Rightarrow \forall i \in \llbracket 1, n \rrbracket, (x_{i,t})$  stationnaire.  
La réciproque est fausse.

**Démonstration.**

(i)

$$\forall t, \mathbb{E}(X_t) = m \Leftrightarrow \forall i, \begin{pmatrix} \mathbb{E}(x_{1,t}) \\ \vdots \\ \mathbb{E}(x_{n,t}) \end{pmatrix} = \begin{pmatrix} m_1 \\ \vdots \\ m_n \end{pmatrix}$$

- (ii)  $(X_t)$  stationnaire  $\Leftrightarrow \forall t, \forall (i, j), \text{Cov}(x_{i,t}, x_{j,t-h}) = \gamma_{i,j}(h)$ . Si  $i = j$ , on a bien :  $\forall t, \text{Cov}(x_{i,t}, x_{i,t-h}) = \gamma_i(h)$ , donc  $(x_{i,t})$  est stationnaire.

■

**Exemple 1.1 (Bruit blanc vectoriel).**  $X_t \rightsquigarrow \mathcal{BB}(0, \Omega)$  ssi :

- (i)  $\forall t, \mathbb{E}(X_t) = 0$

$$(ii) \quad \forall t, \mathbb{V}(X_t) = \Omega$$

$$(iii) \quad \forall t, \forall h \neq 0, \mathbb{E}(X_t, X_{t-h}) = 0$$

**Exemple 1.2 (VMA(1)).** Soit  $\varepsilon_t \rightsquigarrow \mathcal{BB}(0, \Omega) : X_t = \varepsilon_t - A\varepsilon_{t-1}$ , où  $A \in \mathcal{M}_n(\mathbb{R})$ . Vérifions que  $(X_t)$  est bien stationnaire :

$$(i) \quad \mathbb{E}(x_t) = \mathbb{E}(\varepsilon_t - A\varepsilon_{t-1}) = 0$$

$$(ii) \quad \begin{aligned} \mathbb{V}(X_t) &= \mathbb{E}(X_t X_t') = \mathbb{E}[(\varepsilon_t - A\varepsilon_{t-1})(\varepsilon_t' - \varepsilon_{t-1}'A')] \\ &= \mathbb{E}[\varepsilon_t \varepsilon_t' - A\varepsilon_{t-1} \varepsilon_t' - \varepsilon_t \varepsilon_{t-1}' A' + A' \varepsilon_{t-1} \varepsilon_{t-1}' A'] \\ &= \Omega - A \times 0 - 0 \times A' + A\Omega A' \end{aligned}$$

(iii) Soit  $h \neq 0$  :

$$h = 1 : \mathbb{E}((\varepsilon_t - A\varepsilon_{t-1})(\varepsilon_{t-1} - \varepsilon_{t-2}A')) = -A\Omega$$

$$h > 1 : \mathbb{E}((\varepsilon_t - A\varepsilon_{t-1})(\varepsilon_{t-h} - \varepsilon_{t-h-1}A')) = 0$$

$$h = -1 : \mathbb{E}((\varepsilon_t - A\varepsilon_{t-1})(\varepsilon_{t+1} - \varepsilon_t A')) = -\Omega A'$$

$$h < -1 : \mathbb{E}((\varepsilon_t - A\varepsilon_{t-1})(\varepsilon_{t+h} - \varepsilon_{t+h-1}A')) = 0$$

On note que :  $\Gamma(1) = -A\Omega = [\Gamma(-1)]'$

**Proposition 1.2.** Si  $(X_t)_{t \in \mathbb{Z}}$  est stationnaire, alors,  $\forall h \in \mathbb{Z}, \Gamma(-h) = [\Gamma(h)]'$

**Démonstration.**

$$\begin{aligned} \Gamma(h) &= \mathbb{E}[(X_t - m)(X_{t-h} - m)'] \\ \Gamma(-h) &= \mathbb{E}[(X_t - m)(X_{t+h} - m)'] \\ &= \mathbb{E}[(X_{t-h} - m)'(X_t - m)] \text{ par stationnarité} \\ &= \mathbb{E}([(X_t - m)(X_{t-h} - m)']') \\ &= (\mathbb{E}[(X_t - m)(X_{t+h} - m)'])' = \Gamma(h)' \end{aligned}$$

■

*Remarque.* 1.  $\Gamma(h) = (\gamma_{i,j}(h))_{1 \leq i,j \leq n}$ , où  $\gamma_{i,j}(h) = \mathbb{E}((x_{i,t} - m_i)(x_{j,t-h} - m_j)) = \text{Cov}(x_{i,t}, x_{j,t-h})$

$$2. \Gamma(h) = \mathbb{E}(X_t X_{t-h}' - m X_{t-h}' - X_t m' + m m') = \mathbb{E}(X_t X_{t-h}') + m m'$$

**Proposition 1.3.** Soit  $(X_t)_{t \in \mathbb{Z}}$  stationnaire,  $(A_i)_{i \in \mathbb{Z}}$  une suite de matrices  $n \times n$  non aléatoires, telles que  $\sum_{i \in \mathbb{Z}} |A_i| < +\infty$ . Alors,

- $Y_t = \sum_{i \in \mathbb{Z}} A_i X_i$  est défini presque sûrement pour tout  $t$ , et appartient à  $\mathcal{L}_{\mathbb{R}^n}^2(\Omega, \mathcal{A}, P)$  ;
- $(Y_t)$  est stationnaire ;
- $\mathbb{E}(Y_t) = \left( \sum_{i \in \mathbb{Z}} A_i \right) m_X$  ;
- $\Gamma_Y(h) = \sum_{i,j} A_i \Gamma_X(h+j-i) A_j'$

**Démonstration.** Ce sont des applications du théorème de Fubini. On ne donne donc ici que l'idée de la démonstration, qui est laissée en exercice :

$$\begin{aligned} - \int y_{i,t}^2 dP &= \int \left( \sum_j a_{i,j} x_{j,t-k} \right)^2 dP < +\infty \\ - \mathbb{E}(Y_t) &= \mathbb{E}(\sum_k A_k X_{t-k}) \stackrel{\text{Fubini}}{=} \sum_k \mathbb{E}(A_k X_{t-k}) = \sum_k A_k \mathbb{E}(X_{t-k}) \end{aligned}$$

$$\begin{aligned}
\Gamma_Y(h) &= \mathbb{E}((Y_t - m_Y)(Y_{t-h} - m_Y)') \\
&= \mathbb{E} \left[ \left( \sum_k A_k (X_{t-k} - m_X) \right) \left( \sum_j A_j (X_{t-h-j} - m_X) \right)' \right] \\
&\stackrel{\text{Fubini}}{=} \sum_{k,j} \mathbb{E} [A_k (X_{t-k} - m_X) (A_j (X_{t-h-j} - m_X))'] \\
&= \sum_{j,k} A_k \mathbb{E} [A_k (X_{t-k} - m_X) (X_{t-h-j-k} - m_X)'] A_j' \\
&= \sum_{k,j} A_k \Gamma_X(j+h-k) A_j'
\end{aligned}$$

■

*Remarque.* En général,  $\|A\| = \max_{j \in \llbracket 1, n \rrbracket} (\lambda_j(A'A))^{1/2}$ , où  $\lambda_j(A'A)$  est une valeur propre de  $A'A$ , supérieure à 0. On a alors :  $\forall x \in \mathbb{R}^n$ ,  $\|Ax\| \leq \|A\| \|x\|$  pour  $\|x\|^2 = \sum x_i^2$

**Définition 3 (Processus VMA( $\infty$ )).** Soit  $\varepsilon_t \rightsquigarrow \mathcal{BB}(0, \Omega)$ ,  $(A_k)_{k \in \mathbb{Z}}$ ,  $\sum_k \|A_k\| < +\infty$ . Alors,  $Y_t = \sum_k A_k \varepsilon_{t-k}$  suit un processus VMA( $\infty$ ).

**Exemple 1.3.**  $Y_t = \sum_{k=0}^{+\infty} A_k \varepsilon_{t-k}$ ,  $A_0 = I$ ,  $\sum_k \|A_k\| < +\infty$

## 1.2 Densité spectrale d'un processus vectoriel stationnaire

**Proposition 1.4.** Soit  $\varepsilon_t \rightsquigarrow \mathcal{BB}(0, \Omega)$ ,  $X_t = (m) + \sum_{k \in \mathbb{Z}} A_k \varepsilon_{t-k}$ ,  $\sum_k \|A_k\| < +\infty$ .

Alors,  $\sum_h \|\Gamma_X(h)\| < +\infty$

**Définition 4 (Densité spectrale).** La densité spectrale de  $(X_t)$  est l'application :

$$\begin{aligned}
[-\pi, \pi] &\rightarrow \mathcal{M}_n(\mathbb{C}) \\
\omega &\mapsto S_X(\omega)
\end{aligned}$$

Avec :

$$S_X(\omega) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \Gamma_X(h) e^{i\omega h}$$

**Proposition 1.5.** Ainsi définie, la densité spectrale vérifie les propriétés suivantes :

- (i) Densité spectrale d'un bruit blanc :  $\varepsilon_t \rightsquigarrow \mathcal{BB}(0, \Omega) \Rightarrow S_\varepsilon(\omega) = \frac{1}{2\pi} \Omega$
- (ii) Si  $X_t = \sum_k A_k \varepsilon_{t-k}$ ,  $Y_t = \sum_j B_j X_{t-j}$ ,  $\sum \|A_k\| < +\infty$ ,  $\sum_j \|B_j\| < +\infty$ , alors,  $\exists (C_i)_i / Y_t = \sum_i C_i \varepsilon_{t-i}$  et  $S_Y(\omega) = B_i(e^{i\omega}) S_X(\omega) \overline{B(e^{i\omega})}'$

**Démonstration.** On remarque que :  $Y_t = B(L)X_t$ , avec  $B(L) = \sum_j B_j L^j$ ,  $B(e^{i\omega}) = \sum_j B_j e^{i\omega j}$  et  $\overline{B(e^{i\omega})} = \sum_j B_j e^{-i\omega j}$  puisque  $B_k \in \mathcal{M}_n(\mathbb{R})$

Le reste est similaire à la démonstration dans le cas univarié. ■

**Théorème 1.6 (Théorème d'injectivité).** Soit  $X_t = \sum_j A_j \varepsilon_{t-j}$ , où  $\varepsilon_t \rightsquigarrow \mathcal{BB}$  et  $\sum_j \|A_j\| < +\infty$ . Alors,

$$\Gamma_X(h) = \int_{-\pi}^{\pi} S_X(\omega) e^{-i\omega h} d\omega$$

**Démonstration.** C'est une application du théorème de Fubini, comme en univarié. ■

### 1.3 Innovation et théorème de Wold

*Rappel.* Soit  $(x_t)_{t \in \mathbb{Z}}$  un processus réel stationnaire,  $\varepsilon_t = x_t - x_t^* = x_t - \mathbb{E}\mathbb{L}(x_t | \underline{x}_{t-1}) = x_t - P_{H_{t-1}}(x_t)$ , où  $H_{t-1} = \overline{\mathcal{L}}(1, \underline{x}_{t-1}, \dots)$ .

Ici,  $X_t = (x_{1,t}, \dots, x_{n,t})'$ . À la date  $t-1$ , les  $\underline{x}_{1,t-1}, \dots, \underline{x}_{n,t-1}$  sont supposés observés.

Dans le cas des processus vectoriels, on va considérer  $H_t = \overline{\mathcal{L}}(1, \underline{x}_{1,t}, \dots, \underline{x}_{n,t})$ , sous-espace vectoriel fermé de  $L^2_{\mathbb{R}}(\Omega, \mathcal{A}, P)$ .

*Remarque.*  $X_t \in L^2_{\mathbb{R}^n}(\Omega, \mathcal{A}, P)$

**Définition 5 (Prévision optimale de  $X_t$  à la date  $t-1$ ).**

$$X_t^* = \begin{pmatrix} x_{1,t}^* \\ \vdots \\ x_{n,t}^* \end{pmatrix} = \begin{pmatrix} \mathbb{E}\mathbb{L}(x_{1,t} | 1, \underline{x}_{1,t-1}, \dots, \underline{x}_{n,t-1}) \\ \vdots \\ \mathbb{E}\mathbb{L}(x_{n,t} | 1, \underline{x}_{1,t-1}, \dots, \underline{x}_{n,t-1}) \end{pmatrix} = \begin{pmatrix} P_{H_{t-1}}(x_{1,t}) \\ \vdots \\ P_{H_{t-1}}(x_{n,t}) \end{pmatrix}$$

**Définition 6 (Processus des innovations de  $(X_t)$ ).** On définit le processus des innovations  $(\varepsilon_t)$  de  $(X_t)$  comme :

$$\varepsilon_t = X_t - X_t^*$$

ce qui revient à :

$$\varepsilon_t = \begin{pmatrix} \varepsilon_{1,t} \\ \vdots \\ \varepsilon_{n,t} \end{pmatrix} \text{ où } \forall i \in \llbracket 1, n \rrbracket, \varepsilon_{i,t} = x_{i,t} - \mathbb{E}\mathbb{L}(x_{i,t} | 1, \underline{x}_{1,t-1}, \dots, \underline{x}_{n,t-1})$$

**Notation 1.1.** Dans ce qui suit, on note  $\overline{\mathcal{L}}_t(X_t) = H_t = \overline{\mathcal{L}}(1, \underline{x}_{1,t-1}, \dots, \underline{x}_{n,t-1})$ ,  $X_t^* = \mathbb{E}\mathbb{L}(X_t | \underline{X}_{t-1})$  et  $\varepsilon_t = X_t - \mathbb{E}\mathbb{L}(X_t | \underline{X}_{t-1})$

Comme on se sert du passé de toutes les variables pour faire la prévision de chaque composante, il est clair que  $\mathbb{V}(P_{H_{t-1}}(x_{1,t})) \leq \mathbb{V}(\mathbb{E}\mathbb{L}(x_t | \underline{x}_{t-1}))$

On applique le théorème de Pythagore :

$$\begin{aligned} \mathbb{V}(x_{1,t} - \mathbb{E}\mathbb{L}(x_{1,t} | \underline{X}_{t-1})) &\leq \mathbb{V}(\mathbb{E}\mathbb{L}(x_{1,t} | \underline{x}_{1,t-1})) \\ x_{1,t} - \mathbb{E}\mathbb{L}(x_{1,t} | \underline{X}_{t-1}) &= \varepsilon_{1,t} \neq \eta_{1,t} = \mathbb{E}\mathbb{L}(x_{1,t} | \underline{x}_{1,t-1}) \end{aligned}$$

**Proposition 1.7.**

$$\forall i, j \in \llbracket 1, n \rrbracket \forall k > 0, \varepsilon_{i,t} \perp x_{j,t-k}$$

ce qui équivaut à :

$$\forall i, \mathbb{E}(\varepsilon_{i,t} x_{j,t-k}) = 0$$

**Démonstration.**  $\varepsilon_{i,t} = x_{i,t} - \mathbb{E}\mathbb{L}(x_{i,t} | \underline{X}_{t-1}) \stackrel{L^2}{\perp} H_{t-1}$  ■

*Remarque.*  $\perp^{L^2}$  signifie orthogonal au sens de  $L^2$ , et équivaut à la non-corrélation.

**Proposition 1.8 (Optimalité de la prévision optimale).**  $X_t^*$  est la prévision optimale de  $X_t$  parmi toutes les prévisions linéaires en fonction du passé des  $x_{j,t-1}$ , c'est-à-dire que  $\mathbb{V}(X_t - X_t^*)$  est minimale au sens des matrices symétriques.

En d'autres termes, si  $\widehat{X}_t$  est une prévision linéaire de  $X_t$ ,  $\mathbb{V}(X_t - X_t^*) < \mathbb{V}(X_t - \widehat{X}_t)$  au sens des matrices symétriques.

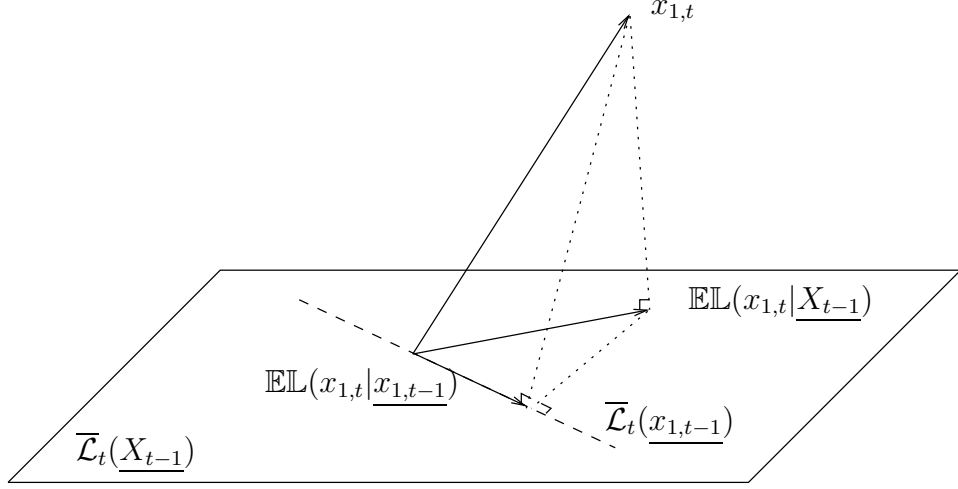


FIG. 1 – Plans de projection en dimension 2

*Rappel.* Soient  $A$  et  $B \in \mathcal{M}_n(\mathbb{R})$  symétriques :

$$\begin{aligned} B \geq A &\Leftrightarrow B - A \geq 0 \Leftrightarrow \forall x \in \mathbb{R}^n, x'(B - A)x \geq 0 \\ B > A &\Leftrightarrow B - A > 0 \Leftrightarrow \forall x \in \mathbb{R}^n, x'(B - A)x > 0 \end{aligned}$$

**Démonstration de l'optimalité.** Soit  $\hat{X}_t$  prévision linéaire de  $X_t$  s'exprimant en fonction des  $x_{j,t-k}$  :

$$\begin{aligned} \mathbb{V}(X_t - \hat{X}_t) &= \mathbb{V}((X_t - X_t^*) + (X_t^* - \hat{X}_t)) = \\ &= \mathbb{V}(X_t - X_t^*) + \mathbb{V}(X_t^* - \hat{X}_t) + \mathbb{E} \left[ (X_t - X_t^*)(X_t^*)' + (X_t - \hat{X}_t)(X_t - X_t^*)' \right] \\ \mathbb{E} \left[ (X_t - X_t^*)(X_t^* - \hat{X}_t)' \right] &= \mathbb{E} \left( \underbrace{(x_{i,t} - x_{i,t}^*)}_{\perp H_{t-1}} \left( \underbrace{x_{j,t}^*}_{\in \mathcal{L}(\underline{X_{t-1}})} - \underbrace{\hat{x}_{j,t}}_{\in \mathcal{L}(\underline{X_{t-1}})} \right) \right)_{i,j} = 0 \end{aligned}$$

$$\text{Symétriquement, } \mathbb{E} \left[ (X_t - X_t^*)(X_t^*)' + (X_t - \hat{X}_t)(X_t - X_t^*)' \right] = 0$$

Ainsi,

$$\mathbb{V}(X_t - \hat{X}_t) = \mathbb{V}((X_t - X_t^*) + (X_t^* - \hat{X}_t)) = \mathbb{V}(X_t - X_t^*) + \mathbb{V}(X_t^* - \hat{X}_t)$$

Or,  $\mathbb{V}(X_t^* - \hat{X}_t)$  matrice symétrique positive, donc  $\mathbb{V}(X_t - \hat{X}_t) > \mathbb{V}(X_t - X_t^*)$  au sens des matrices symétriques. Or,  $\mathbb{V}(X_t^* - \hat{X}_t)$  est symétrique définie positive sauf si  $X_t^* \stackrel{p.s.}{=} \hat{X}_t$ .

Donc :

$$\hat{X}_t \neq X_t^* \Leftrightarrow \mathbb{V}(X_t - \hat{X}_t) > \mathbb{V}(X_t - X_t^*)$$

■

**Corollaire.**  $\forall u \in \mathbb{R}^n$ , la prévision optimale parmi les prévisions linéaires en fonction de  $\underline{X_{t-1}}$  de  $u'X_t = \sum_{i=1}^n u_i x_{i,t}$

(i) est  $u'X_t^*$  ;

(ii) en particulier,  $x_{i,t}^*$  est la prévision optimale de  $x_{i,t}$  parmi les prévisions linéaires en fonction de  $X_{t-1}$ .

**Démonstration.**  $\mathbb{V}(X_t - X_t^*) < (X_t - \hat{X}_t)$ , donc  $\forall u \in \mathbb{R}^n, U \neq 0$ ,

$$\begin{aligned} & u'[\mathbb{V}(X_t - \hat{X}_t) - \mathbb{V}(X_t - X_t^*)]u > 0 \\ \Leftrightarrow & u'\mathbb{V}(X_t - \hat{X}_t)u - u'\mathbb{V}(X_t - X_t^*)u > 0 \\ \Leftrightarrow & \mathbb{V}(u'X_t - u'\hat{X}_t) - \mathbb{V}(u'X_t - u'X_t^*) > 0 \end{aligned}$$

$X_t^*$  est donc optimal pour toute combinaison linéaire des  $x_{i,t}$ . Pour  $u = e_i$ , on a le second résultat. ■

**Théorème 1.9 (Théorème de Wold en multivarié).** Soit  $(X_t)_{t \in \mathbb{Z}}$  processus vectoriel stationnaire,  $(\varepsilon_t)_{t \in \mathbb{Z}}$  le processus des innovations de  $(X_t)$ .

Alors,

$$\begin{aligned} \exists (C_k)_{k \in \mathbb{N}}, \quad C_0 = I, \quad \sum_k \|C_k\| < +\infty \quad \text{tel que} \\ X_t = m_X + \sum_{k=0}^{+\infty} C_k \varepsilon_{t-k} \end{aligned}$$

## 1.4 Convergence des moments empiriques

**Définition 7 (moments empiriques).** On définit la moyenne empirique comme :

$$\hat{m} = \bar{X}_T = \frac{1}{T} \sum_{t=1}^T X_t$$

et l'auto-covariance empirique comme :

$$\hat{\Gamma}_X(h) = \frac{1}{T-h} \sum_{t=h+1}^T (X_t - \bar{X}_T)(X_{t-h} - \bar{X}_T)'$$

**Proposition 1.10.** De fait du théorème central-limite, ces estimateurs ont les propriétés suivantes :

1.  $\hat{m}$  et  $\hat{\Gamma}^X(h)$  sont des estimateurs convergents en probabilité de  $m_X$  et de  $\Gamma_X(h)$  ;
2. Si de plus  $(\varepsilon_t)$  est stationnaire à l'ordre 4,

$$\sqrt{T}(\bar{X}_T - m_X) \xrightarrow{L} \mathcal{N}\left(0, \sum_{k \in \mathbb{Z}} \Gamma_X(h)\right) = \mathcal{N}(0, 2\pi S_X(0))$$

## 2 Processus VAR stationnaires

### 2.1 Définition et propriétés générales

**Définition 8.**  $(X_t)_{t \in \mathbb{Z}}$  est un processus VAR d'ordre  $p$  si et seulement si :

1.  $(X_t)$  est stationnaire ;

2. Il existe  $\varepsilon_t \rightsquigarrow \mathcal{BB}$ ,  $\Phi_j \in \mathcal{M}_n(\mathbb{R})$ ,  $j \in \llbracket 1, p \rrbracket$ ,  $\Phi_p \neq 0$  tels que :

$$X_t = \mu + \Phi_1 X_{t-1} + \dots + \Phi_p X_{t-p} + \varepsilon_t$$

On note aussi :  $\Phi(L)X_t = \mu + \varepsilon_t$ , où  $\Phi(L) = I - \Phi_1 L - \dots - \Phi_p L^p$ .

*Remarque.* Soit  $(X_t)$  VAR tel que  $\Phi(L)X_t = \mu + \varepsilon_t$ . Il existe des solutions non stationnaires en espérance vérifiant la même équation. Il suffit pour le montrer considérer  $(Y_t)$  déterministe tel que  $\Phi(L)Y_t = 0$ , et de considérer  $Z_t = X_t + Y_t$ .

**Proposition 2.1.** *Si  $\det(\Phi(z))$  a toutes ses racines de module strictement supérieur à 1, l'équation  $\Phi(L)X_t = \mu + \varepsilon_t$  admet effectivement une solution stationnaire.*

*Remarque.*  $\Phi(L) = I - \Phi_1 L - \dots - \Phi_p L^p = (\varphi_{i,j}(z))_{i,j}$

Si  $\Phi_k = \varphi_{i,j}^k$ , alors

$$\begin{aligned} \varphi_{i,i}(z) &= 1 - \varphi_{i,i}^1 z - \dots - \varphi_{i,i}^p z^p \\ &= 1 - \sum_{k=1}^p \varphi_{i,i}^k z^k \\ \varphi_{i,j}(z) &= - \sum_{k=1}^p \varphi_{i,j}^k z^k, \quad i \neq j \end{aligned}$$

$\det(\Phi(z))$  est donc un polynôme en  $z$  de degré très élevé :  $n \times p$ .

**Démonstration de la proposition.** Soit  $\Phi(L) = 1 - \Phi_1 L - \dots - \Phi_p L^p = (\varphi_{i,j})_{i,j}$ . Soit  $\tilde{\Phi}(L)$  la transposée de la matrice des cofacteurs de  $\Phi(L)$  :  $\Phi(L)\tilde{\Phi}(L) = \tilde{\Phi}(L)\Phi(L) = (\det(\Phi(L)))I_n$ .  $\Phi(L)$  est donc inversible ssi  $(\det \Phi(L))$  est inversible. Dans ce cas,  $\Phi(L)^{-1} = (\det \Phi(L))^{-1} \tilde{\Phi}(L)$ .

Si  $\det(\Phi(z))$  a toutes ses racines de modules strictement supérieur à 1, on sait que :  $(\det(\Phi(z))) = \prod_{i=1}^{np} (1 - \lambda_i z)$  avec  $|\lambda_i| < 1$ . Dans ce cas,  $\exists (a_k)$  tq  $(\det \Phi(z))^{-1} = \sum_{k=0}^{+\infty} a_k z^k$ ,  $a_0 = 1$ ,  $\sum |a_k| < +\infty$ .

Alors,  $\Phi(z)^{-1} = \left( \sum_{k=0}^{+\infty} a_k z^k \right) \tilde{\Phi}(z)$ . Pour les mêmes raisons, il existe  $A_k$  telle que  $\Phi(L)^{-1} = \sum_{k=0}^{+\infty} A_k L^k$ .

On a alors :

$$\begin{aligned} \Phi(L)X_t = \mu + \varepsilon_t &\Leftrightarrow X_t = \Phi(L)^{-1}(\mu + \varepsilon_t) \\ &\Leftrightarrow X_t = (\Phi(1))^{-1}\mu + \Phi(L)^{-1}\varepsilon_t \\ &\Leftrightarrow X_t = \left( \sum_{k=0}^{+\infty} A_k \right) \mu + \sum_{k=0}^{+\infty} A_k \varepsilon_{t-k} \end{aligned}$$

Ainsi défini,  $(X_t)$  est bien une solution stationnaire du modèle. ■

*Remarque.* Si  $\det(\Phi(1)) = 0$ , on peut montrer que  $(X_t)$  est non stationnaire, et  $\Phi(1)$  est lié à la dimension du sous-espace de co-intégration.

*Remarque.* On ne s'intéresse pas au cas où  $\det(\Phi(z))$  a des racines de module strictement inférieur à 1.

**Proposition 2.2.** *Soit  $(X_t) \rightsquigarrow VAR(p)$  tel que  $\Phi(L)X_t = \mu + \varepsilon_t$ . On suppose que  $|z| < 1 \Leftrightarrow \det \Phi(z) \neq 0$ .*

*Alors,  $(\varepsilon_t)$  est le processus des innovations de  $(X_t)$ ,  $\bar{\mathcal{L}}(\varepsilon_t) = \bar{\mathcal{L}}(X_t)$ .*

*La représentation  $\Phi(L)X_t = \mu + \varepsilon_t$  est la représentation canonique de  $(X_t)$ .*

**Démonstration.** On sait que  $X_t = m + A(L)\varepsilon_t$  avec  $m = (\sum_k A_k)\mu = \Phi(1)^{-1}\mu$ ,  $A(L) = \sum_k A_k L^k$ ,  $A_0 = 1$ ,  $\sum \|A_k\| < +\infty$ .



On a alors :  $\varepsilon_t = \Phi(L)X_t - \mu$ , donc

$$\begin{aligned} \forall i \in \llbracket 1, n \rrbracket, \varepsilon_{i,t} &= x_{i,t} - \sum_{k=0}^p \sum_{j=1}^n \varphi_{i,j}^k x_{j,t-k} - \mu_i \\ \Rightarrow \varepsilon_{i,t} &\in \mathcal{L}(1, x_{i,t}, x_{j,t-k}, k \in \llbracket 1, p \rrbracket) \subset \overline{\mathcal{L}}(\underline{X}_t) \\ \Rightarrow \mathcal{L}(1, \varepsilon_{1,t}, \dots, \varepsilon_{n,t}) &\subset \overline{\mathcal{L}}(\underline{X}_t) \\ \Rightarrow \mathcal{L}(1, \varepsilon_{1,t-k}, \dots, \varepsilon_{n,t-k}) &\subset \overline{\mathcal{L}}(\underline{X}_{t-k}) \subset \overline{\mathcal{L}}(\underline{X}_t) \\ \Rightarrow \mathcal{L}(\underline{\varepsilon}_t) &\subset \overline{\mathcal{L}}(\underline{X}_t) \end{aligned}$$

$$\begin{aligned} X_t &= m + \sum_{k=0}^{+\infty} A_k \varepsilon_{t-k} = m + \varepsilon_t + \sum_{K \geq 1} A_K \varepsilon_{t-K} \\ x_{i,t} &= m_i + \varepsilon_{i,t} + \sum_{k \geq 1} A_k \varepsilon_{t-k} \in \mathcal{L}(\underline{\varepsilon}_t) \end{aligned}$$

On montre de m me que  $\overline{\mathcal{L}}(\underline{X}_t) \subset \overline{\mathcal{L}}(\underline{\varepsilon}_t)$ , et donc que

$$\overline{\mathcal{L}}(\underline{X}_t) = \overline{\mathcal{L}}(\underline{\varepsilon}_t)$$

Or,

$$x_{i,t} = \mu_i + \varepsilon_{i,t} + \sum_{k=1}^p \sum_{j=1}^n \varphi_{i,j}^k x_{j,t-k}$$

$$x_{i,j}^* = \mathbb{E}\mathbb{L}(x_{i,t} | \underline{X}_{t-1}) = \mu_i + \mathbb{E}\mathbb{L}(\varepsilon_{i,t} | \underline{X}_{t-1}) + \sum_{k=1}^p \sum_{j=1}^n \varphi_{i,j}^k x_{j,t-k}$$

Comme  $\varepsilon_t \rightsquigarrow \mathcal{BB}$ ,  $\mathbb{E}\mathbb{L}(\varepsilon_{i,t} | \underline{X}_{t-1}) = 0$ , donc :

$$\begin{aligned} x_{i,t}^* &= x_{i,t} - \varepsilon_{i,t} \\ \Rightarrow \varepsilon_{i,t} &= x_{i,t} - x_{i,t}^* \\ \Rightarrow \varepsilon_t &= X_t - X_t^* \end{aligned}$$

■

## 2.2 Pr vision d'un processus VAR stationnaire

**Pr vision optimale** Soit  $X_t$  tel que  $\Phi(L)X_t = \mu + \varepsilon_t$ , avec  $|z| \leq 1 \Rightarrow \det \Phi(z) \neq 0$ .

$$\begin{aligned} X_t &= \mu + \Phi_1 X_{t-1} + \dots + \Phi_p X_{t-p} + \varepsilon_t \quad \text{avec } \Phi_j \in \mathcal{M}_n(\mathbb{R}) \\ X_{t+1}^* &= \mathbb{E}\mathbb{L}(X_{t+1} | \underline{X}_t) \\ X_{t+1} &= \mu + \Phi_1 X_t + \dots + \Phi_p X_{t+1-p} + \varepsilon_{t+1} \\ \Rightarrow \mathbb{E}\mathbb{L}(X_{t+1} | \underline{X}_t) &= \mu + \Phi_1 X_t + \dots + \Phi_p X_{t+1-p} + 0 \quad \text{car } \varepsilon_t \text{ est l'innovation} \\ \text{Donc } X_{t+1}^* &= \mu + \sum_{k=1}^p \Phi_k X_{t+1-k} \end{aligned}$$

Comme dans un AR, on a alors :

$$\begin{aligned} X_{t+2} &= \mu + \Phi_1 X_{t+1} + \Phi_2 X_t + \dots + \Phi_p X_{t+2-p} + \varepsilon_{t+2} \\ {}_t X_{t+2}^* &= \mu + \Phi_1 X_{t+1}^* + \Phi_2 X_t + \dots + \Phi_p X_{t+2-p} \\ {}_t e_{t+2} &= X_{t+2} - {}_t X_{t+2}^* = \Phi_1 (X_{t+1} - {}_t X_{t+1}^*) + \varepsilon_{t+2} = \Phi_1 \varepsilon_{t+1} - \varepsilon_{t+2} \\ \text{Donc } \mathbb{V}({}_t e_{t+2}) &= \mathbb{V}(\Phi_1 \varepsilon_{t+1} - \varepsilon_{t+2}) = \Phi_1 \Omega \Phi_1' + \Omega \end{aligned}$$

De manière générale,

$$\begin{aligned} X_{t+h} &= \mu + \sum_{k=1} \Phi_k X_{t+h-k} + \varepsilon_{t+h} \\ {}_tX^*_{t+h} &= \mu + \sum_{k=1} \Phi_k {}_tX^*_{t+h-k} \end{aligned}$$

avec la convention  ${}_tX^*_{t+h-k} = X_{t+h-k}$  si  $h \leq k$ . La prévision optimale se calcule donc par itérations successives sur  $h$ .

**Régions de prévision** Supposons  $\varepsilon_t \rightsquigarrow \mathcal{N}(0, \Omega)$ .

Alors,  ${}_te_{t+h} \rightsquigarrow \mathcal{N}(0, V_h)$

$${}_te_{t+h} = \begin{pmatrix} e_{1,t+h} \\ \vdots \\ e_{n,t+h} \end{pmatrix}$$

Il y a donc deux manières de déterminer la région de prévision :

1. Composante par composante :

$$\forall i \in \llbracket 1, n \rrbracket, e_{i,t+h} = x_{i,t+h} - {}_tx^*_{i,t+h} \rightsquigarrow \mathcal{N}(0, \sigma_{i,i}^h)$$

On a alors un intervalle de prévision au niveau  $1 - \alpha$  pour  $x_{i,t+h}$  de la forme :

$$\left[ {}_tx^*_{i,t+h} - u_{1-\alpha/2} \sqrt{\sigma_{i,i}^h}, {}_tx^*_{i,t+h} + u_{1-\alpha/2} \sqrt{\sigma_{i,i}^h} \right]$$

2. De manière globale : la détermination d'une région de confiance pose en effet problème, car :

$$\begin{aligned} X_{t+h} - X_y X^*_{t+h} &\rightsquigarrow \mathcal{N}(O, V_h) \\ \Leftrightarrow (X_{t+h} - X_y X^*_{t+h}) V_h^{-1} (X_{t+h} - X_y X^*_{t+h})' &\rightsquigarrow \chi^2(n) \end{aligned}$$

La région de confiance globale est donc un ellipsoïde dans  $\mathbb{R}^n$  :

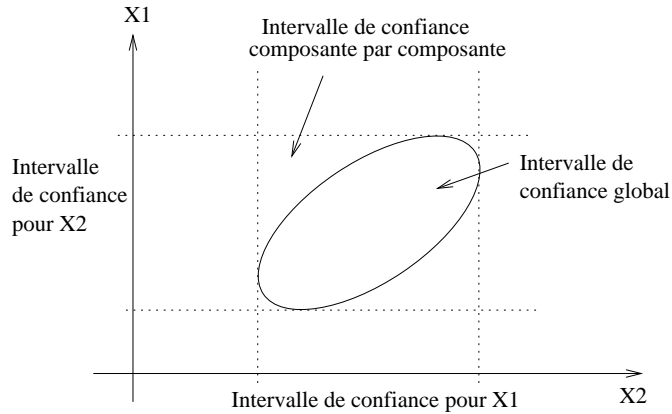


FIG. 2 – Intervalles de confiance

### 3 Estimation d'un modèle VAR sous l'hypothèse de normalité

#### 3.1 Écriture du modèle

**Exemple :**  $n = 2, p = 2$

$$X_t = \begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} \begin{pmatrix} x_{1,t-1} \\ x_{2,t-1} \end{pmatrix} + \begin{pmatrix} c_1 & d_1 \\ c_2 & d_2 \end{pmatrix} \begin{pmatrix} x_{1,t-2} \\ x_{2,t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}$$

$$\mathbb{V}(\varepsilon_t) = \Sigma = \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} \\ \sigma_{2,1} & \sigma_{2,2} \end{pmatrix}$$

On a en fait deux modèles linéaires :

$$(M_1) : x_{1,t} = \mu_1 + a_1 x_{1,t-1} + b_1 x_{2,t-1} + c_1 x_{1,t-2} + d_1 x_{2,t-2} + \varepsilon_{1,t}$$

$$(M_2) : x_{2,t} = \mu_2 + a_2 x_{1,t-1} + b_2 x_{2,t-1} + c_2 x_{1,t-2} + d_2 x_{2,t-2} + \varepsilon_{2,t}$$

Il s'agit d'un modèle de régressions empilées (SURE).

Supposons que les observations commencent en  $t = 1$  :

$$\begin{pmatrix} x_{1,3} \\ \vdots \\ x_{1,T} \\ x_{2,3} \\ \vdots \\ x_{2,T} \end{pmatrix} = \left( \begin{array}{ccccc|ccccc} 1 & x_{1,2} & x_{2,2} & x_{1,1} & x_{2,1} & & & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & & & & \\ 1 & x_{1,T-1} & x_{2,T-1} & x_{1,T-2} & x_{2,T-2} & & & & \\ \hline & & 0 & & & 1 & x_{1,2} & x_{2,2} & x_{1,1} & x_{2,1} \\ & & & & & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & & & 1 & x_{1,T-1} & x_{2,T-1} & x_{1,T-2} & x_{2,T-2} \end{array} \right) \begin{pmatrix} \mu_1 \\ a_1 \\ b_1 \\ c_1 \\ d_1 \\ \mu_2 \\ a_2 \\ b_2 \\ c_2 \\ d_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,3} \\ \vdots \\ \varepsilon_{1,T} \\ \varepsilon_{2,3} \\ \vdots \\ \varepsilon_{2,T} \end{pmatrix}$$

Cette expression s'écrit de manière condensée :

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} Z & 0 \\ 0 & Z \end{pmatrix} + \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

et

$$\mathbb{V} \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} = \begin{pmatrix} \sigma_{1,1} & & 0 & \sigma_{2,1} & & 0 \\ & \ddots & & & \ddots & \\ 0 & & \sigma_{1,1} & 0 & & \sigma_{2,1} \\ \sigma_{1,2} & & 0 & \sigma_{2,2} & & 0 \\ & \ddots & & & \ddots & \\ 0 & & \sigma_{1,2} & 0 & & \sigma_{2,2} \end{pmatrix} = \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} \\ \sigma_{2,1} & \sigma_{2,2} \end{pmatrix} \otimes I_{T-2}$$

**Cas général**

$$X_t = \begin{pmatrix} x_{1,t} \\ \vdots \\ x_{n,t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} + \sum_{k=1}^p \begin{pmatrix} \varphi_{1,1}^k & \cdots & \varphi_{1,n}^k \\ \vdots & & \vdots \\ \varphi_{n,1}^k & \cdots & \varphi_{n,n}^k \end{pmatrix} \begin{pmatrix} x_{1,t-k} \\ \vdots \\ x_{n,t-k} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \vdots \\ \varepsilon_{n,t} \end{pmatrix}$$

On a donc les modèles :

$$(M_i) : x_{i,t} = \mu_t + \sum_{k=1}^p \sum_{j=1}^n \varphi_{i,j}^k x_{j,t-k} + \varepsilon_{i,t}$$

Si les observations commencent en  $t = 1$ , on a :

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} Z & 0 \\ & \ddots \\ 0 & Z \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

avec

$$Z = \begin{pmatrix} 1 & x_{1,p} & \dots & n_{n,p} & \dots & x_{1,1} & \dots & x_{n,1} \\ \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ 1 & x_{1,T-1} & \dots & n_{n,T-1} & \dots & x_{1,T-p} & \dots & x_{n,T-p} \end{pmatrix}$$

et  $\beta_i = (\mu_i, \varphi_{i,1}^k, \dots, \varphi_{i,n}^k, \dots, \varphi_{i,1}^p, \dots, \varphi_{i,n}^p)$

$$\mathbb{V} \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} = \begin{pmatrix} \sigma_{1,1} I_{T-p} & \dots & \sigma_{1,n} I_{T-p} \\ \vdots & \ddots & \vdots \\ \sigma_{n,1} I_{T-p} & \dots & \sigma_{n,n} I_{T-p} \end{pmatrix}$$

On peut ainsi écrire le modèle VAR

$$x = X\beta + \varepsilon$$

avec  $X = I_n \otimes Z$  et  $\mathbb{V}(\varepsilon) = \Sigma \otimes I_{T-p}$ .

Le théorème de Zellner dit :

$$\begin{aligned} x &= (I_n \otimes Z)\beta + \varepsilon, \mathbb{V}(\varepsilon) = \Sigma \otimes I_{T-p} \\ \Rightarrow \hat{\beta}_{MCQG} &= \hat{\beta}_{MCG} = \hat{\beta}_{MCO} = \begin{pmatrix} \hat{\beta}_{1,MCO} \\ \vdots \\ \hat{\beta}_{n,MCO} \end{pmatrix} = \beta \end{aligned}$$

Il suffit finalement d'estimer les MCO dans chaque modèle  $(M_i)$ .

### 3.2 Estimation du modèle empilé

**Proposition 3.1.**  $\hat{\beta}$  vérifie les propriétés suivantes :

- (i)  $\mathbb{E}(\hat{\beta}) = \beta$
- (ii)  $\mathbb{V}(\hat{\beta}) = \Sigma \otimes (Z'Z)^{-1}$
- (iii)  $\hat{\beta}$  est efficace

**Démonstration.**

(i)

$$\begin{aligned} \hat{\beta} &= \begin{pmatrix} \hat{\beta}_1 \\ \vdots \\ \widehat{\text{beta}}_n \end{pmatrix} = \begin{pmatrix} (Z'Z)^{-1} Z' x_1 \\ \vdots \\ (Z'Z)^{-1} Z' x_n \end{pmatrix} = (I_n \otimes (Z'Z)^{-1} Z') x \\ \hat{\beta} &= (I_n \otimes (Z'Z)^{-1} Z') ((I_n \otimes Z)\beta + \varepsilon) \\ &= (I_n \otimes (Z'Z)^{-1} Z' Z)\beta + (I_n \otimes (Z'Z)^{-1} Z') \varepsilon \\ &= \beta + (I_n \otimes (Z'Z)^{-1} Z') \varepsilon \\ \Rightarrow \mathbb{E}(\hat{\beta}) &= \beta \end{aligned}$$

$$\begin{aligned}
(ii) \quad \mathbb{V}(\hat{\beta}) &= (I_n \otimes (Z'Z)^{-1}Z')(\Sigma \otimes I_{T-p})(I_n \otimes Z(Z'Z)^{-1}) \\
&= \Sigma \otimes (Z'Z)^{-1} \underbrace{Z' I_{T-p} Z (Z'Z)^{-1}}_I = \Sigma \otimes (Z'Z)^{-1}
\end{aligned}$$

■

**Définition 9 (Estimateur de  $\Sigma$ ).**

$$\begin{aligned}
\hat{\Sigma} &= \frac{1}{T-p} \sum_{t=p+1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t' \\
\Rightarrow \hat{\sigma}_{i,j} &= \frac{1}{T-p} \sum_{t=p+1}^T \hat{\varepsilon}_{i,t} \hat{\varepsilon}_{j,t}' \quad \hat{\sigma}_{i,i} = \frac{1}{T-p} \sum_{t=p+1}^T \hat{\varepsilon}_{i,t}^2
\end{aligned}$$

*Remarque.* Commencer les observations en 1 fait apparaître des facteurs  $T-p$ . On va donc les renuméroter en les faisant commencer à  $t = -p + 1$  :

$$\Rightarrow x_i = \begin{pmatrix} x_{i,1} \\ \vdots \\ x_{i,T} \end{pmatrix} \quad \hat{\sigma}_{i,j} = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_{i,t} \hat{\varepsilon}_{j,t}'$$

On utilisera cette notation par la suite.

*Remarque.*

$$\sigma_{i,i} = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_{i,t}^2$$

cf l'estimateur usuel des MCO des  $\sigma_{i,i}$  dans  $(M_i)$ , mais sans correction par le nombre de variables explicatives, ce qui revient en fait à considérer l'estimateur du maximum de vraisemblance.

*Remarque.*

$$\hat{\sigma}_{i,i} = \text{Cov}_{\text{emp}}(\hat{\varepsilon}_{i,t}, \hat{\varepsilon}_{j,t})$$

*Remarque.*

$$\hat{\mathbb{V}}(\hat{\beta}) = \hat{\Sigma} \otimes (Z'Z)^{-1}$$

est la matrice de variance-covariance estimée de  $\hat{\beta}$ .

*Aparté (Nombre de paramètres dans un VAR).* Si on a  $n$  variables, on a  $n + n^2p + n(n+1)$  paramètres (pour  $\mu, \Phi$  et  $\Sigma$ ) avec  $nt$  observations. On ne peut faire d'estimations correctes que pour  $n$  petit, de l'ordre de 4 ou 5, et  $T$  assez grand.

*Aparté (À quoi sert le théorème de Wold ?).* En univarié comme en multivarié, il sert à :

- définir la densité spectrale ;
- prouver que tout processus stationnaire admet une représentation  $\text{MA}(\infty)$  (respectivement  $\text{VMA}(\infty)$ ) avec  $(\varepsilon_t)$  l'innovation du processus, ce qui permet d'étudier les effets de propagation des chocs (étude des fonctions impulsion-réponse, IRF) ;
- pour la même raison, montrer l'équivalence entre processus  $\text{MA}(q)$ ,  $\text{AR}(\infty)$ ,  $\text{MA}(p)$  et  $\text{ARMA}(p,q)$ .

En univarié, il permet en outre de définir la représentation canonique par le biais de la densité spectrale.

### 3.3 EMV sous l'hypothèse de normalité

$$X_t = \mu + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + \varepsilon_t$$

$$\varepsilon_t \text{ iid } \mathcal{N}(O, \Sigma)$$

$$\theta = \text{Vect}(\mu_1, \dots, \mu_n, \varphi_{i,j}^k, 1 \leq i, j \leq n, j \in \llbracket 1, n \rrbracket, \sigma_{i,j}, 1 \leq i, j \leq n)$$

On suppose que les observations commencent à  $t = -p + 1$ .

Quel est l'EMV de  $\theta$  ?

$$x = X\beta + \varepsilon$$

$$EMV(\beta) = \hat{\beta}_{MCG} = \hat{\beta}$$

$$EMV(\Sigma) = \hat{\Sigma} = (\hat{\sigma}_{i,j})_{i,j} \hat{\sigma}_{i,j} = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_{i,t} \hat{\varepsilon}_{j,t}$$

Écriture de la log-vraisemblance conditionnelle aux conditions initiales :

On note :

$$l(X_1, \dots, X_T | \theta) = l(X_1, \dots, X_T | X_{-p+1}, \dots, X_0, \theta)$$

$$l(X_1, \dots, X_T | \theta) = \prod_{i=1}^T l(X_i | X_{i-1}, \dots, X_1, \theta) \text{ par itération}$$

Or,  $l(X_i | X_{i-1}, \dots, X_1, \theta) \rightsquigarrow \mathcal{N}(\mu + \varphi_1 X_{i-1} + \dots + \varphi_p X_{i-p}, \Sigma)$ , donc :

$$l(X_1, \dots, X_T | \theta) = \prod_{i=1}^T \frac{1}{\sqrt{2\pi^n}} \frac{1}{\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(X_i - (\mu + \varphi_1 X_{i-1} + \dots + \varphi_p X_{i-p}))' \Sigma^{-1} (X_i - (\mu + \varphi_1 X_{i-1} + \dots + \varphi_p X_{i-p}))}$$

donc

$$\ln(l) = -\frac{nT}{2} \ln(2\pi) - \frac{T}{2} \ln(\det(\Sigma)) - \frac{1}{2} \sum_{i=1}^T (X_i - (\mu + \varphi_1 X_{i-1} + \dots + \varphi_p X_{i-p}))' \Sigma^{-1} (X_i - (\mu + \varphi_1 X_{i-1} + \dots + \varphi_p X_{i-p}))$$

On pose :  $\mu_t = \mu + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p}$ .

On admet que  $\hat{\theta}$  est obtenu par :

$$\hat{\beta}_{EMV} = \hat{\beta}$$

$$\hat{\Sigma} = \frac{1}{T} \sum_{i=1}^T \hat{\varepsilon}_i \hat{\varepsilon}_i'$$

avec  $\hat{\varepsilon}_t = X_t - (\hat{\mu} + \hat{\varphi}_1 X_{t-1} + \dots + \hat{\varphi}_p X_{t-p})$

Calcul de la vraisemblance à son maximum :

$$\ln(l(x, \hat{\theta}_T)) = -\frac{nT}{2} \ln(2\pi) - \frac{T}{2} \ln(\det(\Sigma)) - \frac{1}{2} \sum_{t=1}^T \hat{\varepsilon}_t' \Sigma^{-1} \hat{\varepsilon}_t$$

avec l'astuce habituelle :

$$\begin{aligned} \sum_{t=1}^T \hat{\varepsilon}_t' \Sigma^{-1} \hat{\varepsilon}_t &= \sum_{i=1}^T \text{Tr}(\hat{\varepsilon}_i' \Sigma^{-1} \hat{\varepsilon}_i) \\ &= \sum_{i=1}^T \text{Tr}(\hat{\Sigma}^{-1} \hat{\varepsilon}_i \hat{\varepsilon}_i') \\ &= \text{Tr}((\sum_{i=1}^T \hat{\varepsilon}_i \hat{\varepsilon}_i') \Sigma^{-1}) \\ &= T \text{Tr}(\hat{\Sigma} \hat{\Sigma}^{-1}) \\ \sum_{t=1}^T \hat{\varepsilon}_t' \Sigma^{-1} \hat{\varepsilon}_t &= nT \end{aligned}$$

D'où une forme très simple de la log-vraisemblance maximale :

$$\ln(l(X_1, \dots, X_T, \hat{\theta})) = -\frac{nT}{2} \ln(2\pi) - \frac{nT}{2} - \frac{T}{2} \ln(\det(\hat{\Sigma}))$$

### 3.4 Propriétés de l'EMV sous l'hypothèse de normalité

**Proposition 3.2 (Normalité).**

$$\hat{\beta} \rightsquigarrow \mathcal{N}(\beta, \Sigma \otimes (Z'Z)^{-1})$$

**Proposition 3.3 (Convergence des estimateurs empiriques).**

$$\hat{\beta} \xrightarrow{P} \beta$$

$$\hat{\sigma} \xrightarrow{P} \Sigma$$

**Proposition 3.4 (Normalité asymptotique).**

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \text{plim } \Sigma \otimes \left(\frac{Z'Z}{T}\right)^{-1}\right)$$

Avec :

$$\text{plim } \Sigma \otimes \left(\frac{Z'Z}{T}\right)^{-1} = \Sigma \otimes \text{plim } \left(\frac{Z'Z}{T}\right)^{-1}$$

le dernier membre existant en vertu de la stationnarité des variables.

*Remarque.* Ces résultats sont valides asymptotiquement, même si  $(\varepsilon_t)$  n'est pas gaussien.

### 3.5 Tests linéaires sur les paramètres du modèle

#### 3.5.1 Test de Wald

$$\begin{aligned} H_0 &= R\beta = r \\ H_1 &= R\beta \neq r \end{aligned}$$

$$\begin{aligned} \sqrt{T}(\hat{\beta} - \beta) &\xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \Sigma \otimes \text{plim } \left(\frac{Z'Z}{T}\right)^{-1}\right) \\ \Rightarrow \sqrt{T}(R\hat{\beta} - r) &\xrightarrow{H_0, \mathcal{L}} \mathcal{N}\left(0, R\left(\Sigma \otimes \text{plim } \left(\frac{Z'Z}{T}\right)^{-1}\right)R'\right) \\ \Rightarrow T(R\hat{\beta} - r)' &\left[R\left(\Sigma \otimes \left(\frac{Z'Z}{T}\right)^{-1}\right)R'\right]^{-1}(R\hat{\beta} - r) \xrightarrow{H_0, \mathcal{L}} \chi^2(\text{rg}(R)) \end{aligned}$$

D'où la statistique de test :

$$\xi_W = (R\hat{\beta} - r)' \left[R\left(\Sigma \otimes (Z'Z)^{-1}\right)R'\right]^{-1}(R\hat{\beta} - r) \xrightarrow{H_0, \mathcal{L}} \chi^2(\text{rg}(R))$$

On rejette  $H_0$  au seuil  $1 - \alpha$  si  $\xi_W > \chi_{1-\alpha}^2(\text{rg}(R))$ .

*Remarque.*

$$(M_i) : x_{i,t} = \mu_i + \sum_{k=1}^p \sum_{j=1}^n \varphi_{i,j}^k x_{j,t-k} + \varepsilon_t$$

$$\beta_i = (\mu_i, \varphi_{i,1}^1, \dots, \varphi_{i,n}^p)'$$

$$\beta = \begin{pmatrix} \beta_1 \\ \dots \\ \beta_n \end{pmatrix} \quad \hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \dots \\ \hat{\beta}_n \end{pmatrix}$$

$$\mathbb{V}(\varepsilon_i) = \sigma_{i,i} = I_T \quad \hat{\sigma}_{i,i} = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_{i,t}^2$$

Si  $R\beta$  ne met en jeu que des coefficients « appartenant » au même  $\beta_i$ , on se ramène aux tests usuels sur le modèle  $(M_i)$ .

*Exemple 3.1.* Significativité d'un coefficient.

Si  $R\beta$  met en jeu des paramètres de plusieurs sous-modèles, il faut prendre en compte les covariances des  $\varepsilon_i$  entre eux.

*Exemple 3.2* ( $n = 2, p = 2$ ).

$$\begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} \begin{pmatrix} x_{1,t-1} \\ x_{2,t-1} \end{pmatrix} + \begin{pmatrix} c_1 & d_1 \\ c_2 & d_2 \end{pmatrix} \begin{pmatrix} x_{1,t-2} \\ x_{2,t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}$$

- Si  $H_0 : b_1 = d_1$ , on va faire un test de Fisher usuel dans  $(M_1)$ ;
- Si  $H_0 : b_1 = b_2$ , la loi de  $(\hat{b}_1, \hat{b}_2)$  dépend de  $\Sigma$ , et il faut faire le test de Fisher dans le modèle vectoriel.

Supposons qu'on obtienne  $b_1$  non significativement différent de 0. Il peut être utile de n'en pas tenir compte, car si l'on contraint  $b_1 = 0$ , on sort des hypothèses du théorème de Zellner. On ne va donc imposer des contraintes qu'en cas de non-significativité symétrique, du type  $b_1 = b_2$ .

### 3.5.2 Test du rapport de vraisemblance

$$\ln(l(x, \hat{\theta})) = -\frac{nT}{2}(1 + \ln(2\pi)) - \frac{T}{2} \ln(\det(\hat{\Sigma}))$$

Si Zellner s'applique au modèle contraint,

$$\ln(l(x, \hat{\theta}_0)) = -\frac{nT}{2}(1 + \ln(2\pi)) - \frac{T}{2} \ln(\det(\hat{\Sigma}_0))$$

où

$$\hat{\Sigma}_0 = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t^0 \hat{\varepsilon}_t^{0'}$$

Statistique de test :

$$\xi_{LR} = 2 \ln \left( \frac{l(x, \hat{\theta})}{l(x, \hat{\theta}_0)} \right) = -T \ln(\det(\hat{\Sigma})) + \ln(\det(\hat{\Sigma}_0))$$

$$\xi_{LR} = T \ln \left( \frac{\ln(\det(\hat{\Sigma}_0))}{\ln(\det(\hat{\Sigma}))} \right) \xrightarrow{H_0, \mathcal{L}} \chi^2(k)$$

où  $k$  est le nombre de contraintes testées.

**Exemple 3.3 (Applications).** Ce test peut s'appliquer dans les cas suivants :

- Nullité du  $p$ -ième retard :  $H_0 : \Phi_p = 0$  contre  $H_1 : \Phi_p \neq 0$ . On est censé en voir une application en TD.
- Tests de causalité. Attention, il ne s'agit pas de relation causale au sens physique. La causalité au sens de Granger se demande qu'une variable aide à en prévoir une autre. Voir TD.