



Deuxième année
2004-2005

SERIES TEMPORELLES

LINEAIRES

Polycopié librement inspiré du cours de **Madame Doz**¹

¹La rédaction a été commencée par la "cuisine expérimentale" pour les chapitres 1, 2 et 3 puis complétée et achevée par Joachim CONNAULT pour les chapitres 4 et 5.

Table des matières

Introduction	1
1 Processus réels stationnaires du second ordre	3
1.1 Processus stationnaire du second ordre	3
1.1.1 Définitions	3
1.1.2 Rappels sur $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$	6
1.2 Outils pour l'étude des processus stationnaires	6
1.2.1 Transformée d'un processus stationnaire par une moyenne mobile infinie	6
1.2.2 Régression linéaire ou affine théorique sur un nombre fini de retards	8
1.2.3 Régression linéaire théorique sur un nombre infini de retards	10
1.2.4 Densité spectrale et auto-corrélations inverses	11
1.2.5 Estimateurs associés et lois limites	14
1.3 Polynômes retard et avance	15
1.3.1 Définitions et propositions	15
1.3.2 Inversibilité des polynômes en L	17
2 Processus ARMA et ARIMA	21
2.1 Processus auto-régressifs d'ordre p ($AR(p)$)	21
2.1.1 Définition et représentation canonique	21
2.1.2 Propriétés des processus $AR(p)$	25
2.1.3 Auto-corrélations partielle et inverse d'un processus $AR(p)$	26
2.2 Processus moyenne mobile d'ordre q ($MA(q)$)	28
2.2.1 Définition et représentation canonique	28
2.2.2 Propriétés des processus $MA(q)$	30
2.3 Processus $ARMA(p, q)$	31
2.3.1 Définition et représentation canonique minimale	31
2.3.2 Propriétés des processus $ARMA(p, q)$	33
2.4 Processus $ARIMA(p, d, q)$	34
2.4.1 Approximation auto-régressive d'un $ARIMA(p, d, q)$	36
2.4.2 Approximation moyenne mobile d'un $ARIMA(p, d, q)$	36
3 Identification et estimation d'un modèle ARMA ou ARIMA	39
3.1 Première phase de l'identification : choix de d	39
3.1.1 Approche empirique : l'auto-corrélogramme	39
3.1.2 Approche par les tests de racine unité	41
3.2 Deuxième phase de l'identification : choix de p et q	47
3.2.1 Résultats préliminaires	47

3.2.2	Choix de p pour un $AR(p)$	48
3.2.3	Choix de q pour un $MA(q)$	48
3.2.4	Choix de (p, q) pour un $ARMA(p, q)$	49
3.3	Estimation	49
3.3.1	Cas d'un $AR(p)$	50
3.3.2	Cas d'un $MA(q)$	51
3.3.3	Cas d'un $ARMA(p, q)$	51
3.4	Vérifications <i>a posteriori</i>	51
3.4.1	Tests sur les paramètres	51
3.4.2	Tests sur les résidus	52
3.5	Choix du modèle	53
4	Prévision dans les $ARMA$ et les $ARIMA$	55
4.1	Prévisions dans un $AR(p)$	55
4.2	Prévision dans un $MA(q)$	56
4.3	Cas d'un $ARMA(p, q)$	58
4.3.1	Forme $AR(\infty)$	58
4.3.2	Utilisation d'une équation de récurrence	58
4.4	Cas d'un $ARIMA(p, d, q)$	59
4.5	Intervalles de précision	60
5	Processus vectoriels stationnaires - Processus VAR stationnaires	63
5.1	Processus vectoriels stationnaires du second ordre	63
5.1.1	Définition et proposition	63
5.1.2	Densité spectrale d'un processus vectoriel stationnaire	65
5.1.3	Innovation d'un processus vectoriel	67
5.1.4	Convergence des moments empiriques	69
5.2	Processus VAR stationnaires	70
5.2.1	Définition et proposition générale	70
5.2.2	Prévision dans un VAR stationnaire	73
5.3	Estimation d'un modèle VAR sous hypothèse de normalité	74
5.3.1	Ecriture empilée du modèle	74
5.3.2	Estimation par les MCQG	76
5.3.3	EMV sous l'hypothèse de normalité	77
5.3.4	Propriétés de l'EMV sous l'hypothèse de normalité	79
5.3.5	Tests de restrictions linéaires sur les paramètres du modèle sous hypothèse de normalité	79
	Bibliographie	83
	Index	84

Introduction

Les séries temporelles sont des données mesurées à des intervalles de temps régulier. Les données macroéconomiques sont relevées par année, trimestres, mois, . . . Les données financières sont mensuelles, hebdomadaires, quotidiennes, infra-journalières (on peut généraliser à temps continu, $t \in \mathbb{R}$).

On fera des études en temps discret donc on indicera de façon dénombrable, $t \in \mathbb{Z}$.

On étudiera des séries univariées : elles résultent de l'observation d'une seule série. On modélise la valeur en t en fonction des valeurs passées.

On peut aussi étudier des séries multivariées, c'est-à-dire vectorielles. Par exemple on a un contenu économique qui repose sur un *a priori* économique mais on n'a pas d'*a priori* sur le poids des variables (rôle symétrique?). On parle de modèles *VAR* évoqués dès 1981 par SIMS.

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

Chapitre 1

Processus réels stationnaires du second ordre

Formalisme : On observe une grandeur donnée sur des dates de 1 à T . On considère des observations x_1, \dots, x_T , réalisations des variables aléatoires $X_1, \dots, X_T : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathbb{R}$, $\omega \in \Omega$ est un état de la nature tel que $x_t = X_t(\omega)$.

On dit que $(X_t)_{t \in \mathbb{Z}}$ est un **processus stochastique** et que $(x_t)_{t \in \mathbb{Z}}$ une **trajectoire du processus** $(X_t)_{t \in \mathbb{Z}}$.

Hypothèses supplémentaires : Si $\mathbb{E}(X_t) = m_t$, on a une seule observation (x_t en l'occurrence) pour estimer m_t . En revanche si pour tout $t \in \mathbb{Z}$, $\mathbb{E}(X_t) = m$, on peut estimer m par $\hat{m} = \frac{1}{T} \sum_{t=1}^T X_t$.

Il paraît donc nécessaire de supposer que la suite X_t a certaines propriétés de régularité.

1.1 Processus stationnaire du second ordre

1.1.1 Définitions

Dans toute la suite on considérera $(X_t)_{t \in \mathbb{Z}}$ et on supposera $X_t \in \mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$, $\forall t \in \mathbb{Z}$.

Définition 1.1.1 (Stationnarité stricte ou forte) $(X_t)_{t \in \mathbb{Z}}$ est un **processus stationnaire au sens strict** si :

$\forall n \in \mathbb{N}$, $\forall (t_1, \dots, t_n)$, $\forall h \in \mathbb{Z}$, la loi de $(X_{t_1}, \dots, X_{t_n})$ est identique à la loi de $(X_{t_1+h}, \dots, X_{t_n+h})$

Théorème 1.1.1 (Théorème de Kolmogorov) $(X_t)_{t \in \mathbb{Z}}$ est un processus stationnaire au sens strict si et seulement si la loi de $(X_t)_{t \in \mathbb{Z}}$ est identique à la loi de $(Y_t)_{t \in \mathbb{Z}}$ où $Y_t = X_{t+h}$.

Définition 1.1.2 (Stationnarité faible) $(X_t)_{t \in \mathbb{Z}}$ est un **processus stationnaire du second ordre** (ou un processus faiblement stationnaire) s'il vérifie :

- (i) $\forall t \in \mathbb{Z}$, $\mathbb{E}(X_t) = m$
- (ii) $\forall t \in \mathbb{Z}$, $\mathbb{V}(X_t) = \sigma^2 = \gamma(0)$
- (iii) $\forall t \in \mathbb{Z}$, $\forall h \in \mathbb{Z}$, $\text{Cov}(X_t, X_{t+h}) = \gamma(h)$ (ne dépend que de h)
 $\gamma(h)$ est l'**auto-covariance** d'ordre h de X_t .

Remarque 1.1.1 (1) Dans la suite, les processus stationnaires désignent les processus de la définition 1.1.2;

(2) (iii) \Rightarrow (ii) : $h = 0$ et $\gamma(0) = \sigma^2$;

(3) Si un processus est stationnaire au sens strict alors il est faiblement stationnaire ;

(4) Si $(X_t)_{t \in \mathbb{Z}}$ est un processus gaussien alors il y a équivalence entre stationnarité faible et forte ;

$$(5) \mathbb{E} \begin{pmatrix} X_{t_1} \\ \vdots \\ X_{t_n} \end{pmatrix} = \begin{pmatrix} m \\ \vdots \\ m \end{pmatrix} \quad \mathbb{V} \begin{pmatrix} X_{t_1} \\ \vdots \\ X_{t_n} \end{pmatrix} = \begin{pmatrix} \gamma(0) & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \gamma(0) \end{pmatrix}$$

Exemple 1.1.1 (Processus stationnaire) (1) **Bruit blanc faible** (*white noise*), $(\varepsilon_t)_{t \in \mathbb{Z}}$, si et seulement si :

$$\begin{aligned} \mathbb{E}(\varepsilon_t) &= 0, \quad \forall t \in \mathbb{Z} \\ \mathbb{V}(\varepsilon_t) &= \sigma^2, \quad \forall t \in \mathbb{Z} \\ \text{Cov}(\varepsilon_t, \varepsilon_\tau) &= 0, \text{ si } t \neq \tau \end{aligned}$$

On notera $\varepsilon_t \rightsquigarrow \mathcal{BB}(0, \sigma^2)$.

(2) ε_t est un **bruit blanc fort** si et seulement si les ε_t sont i.i.d., $\mathbb{E}(\varepsilon_t) = 0$ et $\mathbb{V}(\varepsilon_t) = \sigma^2$.

(3) Processus **moyenne mobile** d'ordre 1, noté $MA(1)$ (*moving average of order 1*)

Soit $\theta \in \mathbb{R}^*$.

Soit $\varepsilon_t \rightsquigarrow \mathcal{BB}(0, \sigma^2)$.

Soit $(X_t)_{t \in \mathbb{Z}}$ défini par : $\forall t \in \mathbb{Z}, X_t = \varepsilon_t - \theta \varepsilon_{t-1}$.

Alors $(X_t)_{t \in \mathbb{Z}}$ est un processus stationnaire. On dit que $X_t \rightsquigarrow MA(1)$.

Remarque 1.1.2 En pratique on ne distinguera plus x_t et X_t . (x_t) ou (X_t) désignera toujours le processus et x_1, \dots, x_T ou X_1, \dots, X_T la suite des observations.

Exemple 1.1.2 (Processus non stationnaires) (1) **Marche aléatoire** (*random walk*)

Soit $\varepsilon_t \rightsquigarrow \mathcal{BB}(0, \sigma^2)$.

$(X_t)_{t \in \mathbb{Z}}$ est une marche aléatoire sans dérive si et seulement si

(i) $X_t = X_{t-1} + \varepsilon_t, \forall t \geq 0$

(ii) $\text{Cov}(\varepsilon_t, X_{t-k}) = 0, \forall 0 < k \leq t$.

Même si on a la propriété $\mathbb{E}X_t = \mathbb{E}X_{t-1} \Rightarrow \mathbb{E}X_t = m, \forall t \in \mathbb{Z}$, $(X_t)_t$ n'est pas stationnaire :

$$\left. \begin{aligned} X_t &= X_{t-1} + \varepsilon_t \\ X_{t-1} &= X_{t-2} + \varepsilon_{t-1} \\ &\vdots \\ X_1 &= X_0 + \varepsilon_1 \end{aligned} \right\} \Rightarrow X_t = X_0 + \sum_{k=1}^t \varepsilon_k$$

D'où

$$\begin{aligned} \mathbb{V}(X_t) &= \mathbb{V}(X_0) + 2 \sum_{k=1}^t \text{Cov}(\varepsilon_k, X_0) + \mathbb{V} \left(\sum_{k=1}^t \varepsilon_k \right) \\ &= \mathbb{V}(X_0) + t\sigma^2 \end{aligned}$$

Le processus n'est pas stationnaire en variance.

(2) Processus stationnaire autour d'un *trend* déterministe.

$X_t = a + bt + Y_t$ où $(Y_t)_{t \in \mathbb{Z}}$ est un processus stationnaire.

Par exemple si $Y_t = \varepsilon_t \sim \mathcal{BB}(0, \sigma^2)$, $\mathbb{E}X_t = a + bt$, le processus n'est pas stationnaire en espérance.

Définition 1.1.3 (Fonction d'auto-covariance) L'*auto-covariance* d'un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ est définie par :

$$\begin{aligned} \gamma : \mathbb{Z} &\rightarrow \mathbb{R} \\ h &\mapsto \gamma(h) = \mathbb{Cov}(X_t, *X_{t-h}) \end{aligned}$$

Proposition 1.1.1 (i) γ est une fonction paire :

$$\gamma(-h) = \gamma(h) \quad \forall h$$

(ii) γ est de type positif : $\forall n \in \mathbb{N}, \forall (t_1, \dots, t_n), \forall (a_1, \dots, a_n) \in \mathbb{R}^n$

$$\sum_{1 \leq i, j \leq n} a_i a_j \gamma(t_i - t_j) > 0$$

Démonstration

(i) Parité :

$$\begin{aligned} \gamma(h) &= \mathbb{Cov}(X_t, X_{t+h}) = \mathbb{Cov}(X_{t-h}, X_{(t-h)+h}) = \mathbb{Cov}(X_{t-h}, X_t) \\ &= \mathbb{Cov}(X_t, X_{t-h}) = \gamma(-h) \end{aligned}$$

(ii) Positivité :

$$\begin{aligned} \mathbb{V} \left(\sum a_i X_{t_i} \right) &= \mathbb{Cov} \left(\sum_i a_i X_{t_i}, \sum_j a_j X_{t_j} \right) \\ &= \sum_{i,j} a_i a_j \mathbb{Cov}(X_{t_i}, X_{t_j}) \\ &= \sum_{i,j} a_i a_j \gamma(t_i - t_j) \geq 0 \end{aligned}$$

On suppose toujours qu'il n'y a pas de relations linéaires entre les X_t . En effet, si on avait $\mathbb{V}(\sum a_i X_{t_i}) = 0$ alors $\sum a_i X_{t_i} = \text{constante}$ presque sûrement.

Définition 1.1.4 (Fonction d'auto-corrélation) La fonction d'*auto-corrélation* d'un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ est définie par :

$$\forall h \in \mathbb{Z}, \quad \rho(h) = \frac{\gamma(h)}{\gamma(0)} = \mathbb{Corr}(X_t, X_{t+h})$$

Proposition 1.1.2 $\rho : h \mapsto \rho(h)$ est une fonction paire, de type positif, à valeurs dans $] -1; 1[$.

Démonstration On a

$$\mathbb{C}orr(X_t, X_{t+h}) = \frac{\mathbb{C}ov(X_t, X_{t+h})}{\sqrt{\mathbb{V}ar X_t \mathbb{V}ar X_{t+h}}} = \frac{\gamma(h)}{\gamma(0)}$$

où γ est paire de type positif.

Définition 1.1.5 (Auto-corrélogramme théorique) *L'auto-corrélogramme de $(X_t)_{t \in \mathbb{Z}}$ est le graphe de :*

$$\begin{cases} \mathbb{N} & \rightarrow]-1; 1[\\ h & \mapsto \rho(h) \end{cases}$$

1.1.2 Rappels sur $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$

$\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ est une espace de Hilbert pour le produit scalaire $(X|Y) = \mathbb{E}XY$.

$$X_n \xrightarrow{\mathcal{L}^2} X \iff \lim_{n \rightarrow +\infty} \|X_n - X\|_2 = 0$$

Si

$$\sum_{j \in \mathbb{Z}} \|a_j X_j\|_2 = \sum_{j \in \mathbb{Z}} |a_j| \|X_j\|_2 < +\infty$$

alors la série $\sum_{j \in \mathbb{Z}} a_j X_j$ est définie p.s. et :

$$\sum_{j=-p}^q a_j X_j \xrightarrow{p, q \rightarrow +\infty} \sum_{j \in \mathbb{Z}} a_j X_j$$

Théorème 1.1.2 (Projection sur un s.e.v. fermé H de $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$)

$$\forall X \in \mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P}), \exists ! X^* \in H / \|X - X^*\|_2 = \min_{Y \in H} \|X - Y\|_2$$

$P_H(X) = X^*$ est caractérisé par $X^* \in H$ et $X - X^* \in H^\perp$.

Théorème 1.1.3 (Théorème des trois perpendiculaires) *Soit H un s.e.v. fermé de $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$, G un s.e.v. fermé de H , alors :*

$$\forall X \in \mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P}), \quad P_G(P_H(X)) = P_G(X)$$

1.2 Outils pour l'étude des processus stationnaires

1.2.1 Transformée d'un processus stationnaire par une moyenne mobile infinie

Définition 1.2.1 (Proposition) *Soient $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire et $(a_j)_{j \in \mathbb{Z}}$ une suite de réels tels que $\sum_j |a_j| < +\infty$.*

Alors $Y_t = \sum_{j \in \mathbb{Z}} a_j X_{t-j}$ est défini (p.s.) pour tout t .

On a les propriétés suivantes :

(i) $Y_t \in \mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P}), \forall t \in \mathbb{Z}$

(ii) $(Y_t)_t$ est un processus stationnaire tel que

$$\mathbb{E}Y_t = m_Y = \left(\sum_{j \in \mathbb{Z}} a_j \right) m_X$$

$$\gamma_Y(h) = \sum_{j,k} a_j a_k \gamma(h+k-j) = \sum_{j,k} a_j a_k \gamma(h+j-k), \forall h \in \mathbb{Z}$$

On dit que $(Y_t)_{t \in \mathbb{Z}}$ est la **transformée de** $(X_t)_{t \in \mathbb{Z}}$ **par la moyenne mobile infinie associée aux** $(a_j)_{j \in \mathbb{Z}}$.

Démonstration

(i)

$$\sum_j \|a_j X_{t-j}\|_2 = \sum_j |a_j| \|X_{t-j}\|_2 = \left(\sum_j |a_j| \right) (m_X^2 + \gamma_X(0))^{\frac{1}{2}} < +\infty$$

Y_t est donc défini p.s. et $Y_t \in \mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$

(ii) On a alors :

$$\begin{aligned} \mathbb{E}Y_t &= \int_{\Omega} Y_t dP = \int_{\Omega} \left(\sum_{j \in \mathbb{Z}} a_j X_{t-j} \right) dP \\ &= \sum_{j \in \mathbb{Z}} a_j \left(\int_{\Omega} X_{t-j} dP \right) \text{ (FUBINI)} \\ &= \mathbb{E} \left(\sum_{j \in \mathbb{Z}} a_j X_{t-j} \right) \\ &= \sum_{j \in \mathbb{Z}} a_j \mathbb{E}X_{t-j} \\ &= m_X \left(\sum_j a_j \right) \end{aligned}$$

Enfin :

$$\begin{aligned} \text{Cov}(Y_t, Y_{t-h}) &= \text{Cov} \left(\sum_{j \in \mathbb{Z}} a_j X_{t-j}, \sum_{k \in \mathbb{Z}} a_k X_{t-h-k} \right) \\ &= \sum_j \sum_k a_j a_k \underbrace{\text{Cov}(X_{t-j}, X_{t-h-k})}_{\gamma_X(h+k-j)} \end{aligned}$$

Définition 1.2.2 Si $X_t = \varepsilon_t \sim \mathcal{BB}(0, \sigma^2)$ alors $Y_t = \sum_{j \in \mathbb{Z}} a_j \varepsilon_{t-j}$ et on dit que $Y_t \sim MA(\infty)$.

1.2.2 Régression linéaire ou affine théorique sur un nombre fini de retards

Définition 1.2.3 Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire.

(i) La **régression linéaire théorique** de X_t sur X_{t-1}, \dots, X_{t-p} est la projection orthogonale dans $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ de X_t sur $H = \text{Vect}(X_{t-1}, \dots, X_{t-p})$.

On note généralement $EL(X_t | X_{t-1}, \dots, X_{t-p})$ la régression linéaire théorique de X_t sur X_{t-1}, \dots, X_{t-p} .

(ii) La **régression affine théorique** de X_t sur X_{t-1}, \dots, X_{t-p} est la projection orthogonale dans $\mathcal{L}^2(\Omega, \mathcal{A}, P)$ de X_t sur $H^* = \text{Vect}(1, X_{t-1}, \dots, X_{t-p})$.

On note généralement $EL(X_t | 1, X_{t-1}, \dots, X_{t-p})$ la régression affine théorique de X_t sur X_{t-1}, \dots, X_{t-p} .

Proposition 1.2.1 (i) et (ii) coïncident si et seulement si $\mathbb{E}X_t = 0$.

Remarque 1.2.1 (1) Si $\mathbb{E}X_t \neq 0$, on calculera toujours la régression affine. On la note aussi souvent $EL(X_t | X_{t-1}, \dots, X_{t-p})$.

(2) $\text{Vect}(X_t | X_{t-1}, \dots, X_{t-p})$ et $\text{Vect}(X_t | X_{t-1}, \dots, X_{t-n})$ sont des s.e.v. de dimension finie de \mathcal{L}^2 donc fermés.

(3) Si $(X_t)_t$ est gaussien, alors $EL(X_t | \cdot) = \mathbb{E}(X_t | \cdot)$

Rappel : Calcul de la régression affine théorique (ii)

$H = \text{Vect}(1, X_{t-1}, \dots, X_{t-p})$ et $X_t^* = p_H(X_t)$ est caractérisé par $X_t^* \in H$ et $X_t - X_t^* \perp H$.

$$X_t^* \in H \Leftrightarrow \exists a_0, a_1, \dots, a_p / X_t^* = a_0 + \sum_{j=1}^p a_j X_{t-j}$$

$$\begin{aligned}
X_t - X_t^* \perp H &\Leftrightarrow \begin{cases} (X_t - X_t^* | 1) = 0 \\ (X_t - X_t^* | X_{t-j}) = 0 \quad \forall j = 1, \dots, p \end{cases} \\
&\Leftrightarrow \begin{cases} \mathbb{E}(X_t - X_t^*) = 0 \\ \mathbb{E}[(X_t - X_t^*)X_{t-j}] = 0 \quad \forall j = 1, \dots, p \end{cases} \\
&\Leftrightarrow \begin{cases} \mathbb{E}X_t = m_X = \mathbb{E} \left(a_0 + \sum_{j=1}^p a_j X_{t-j} \right) = a_0 + m_X \sum_{j=1}^p a_j \\ \mathbb{E}(X_t X_{t-j}) = \mathbb{E} \left[\left(a_0 + \sum_{k=1}^p a_k X_{t-k} \right) X_{t-j} \right] \quad \forall j = 1, \dots, p \end{cases} \\
&\Leftrightarrow \begin{cases} a_0 = m_X \left(1 - \sum_{j=1}^p a_j \right) \\ \mathbb{E}(X_t X_{t-j}) = \mathbb{E} \left[m_X \left(1 - \sum_{k=1}^p a_k \right) X_{t-j} \right] + \sum_{k=1}^p a_k \mathbb{E}(X_{t-k} X_{t-j}) \quad \forall j = 1, \dots, p \end{cases} \\
&\Leftrightarrow \begin{cases} a_0 = m_X \left(1 - \sum_{j=1}^p a_j \right) \\ \mathbb{E}(X_{t-k} X_{t-j}) = m_X^2 \left(1 - \sum_{k=1}^p a_k \right) + \sum_{k=1}^p a_k \mathbb{E}(X_t X_{t-k}) \quad \forall j = 1, \dots, p \end{cases} \\
&\Leftrightarrow \begin{cases} a_0 = m_X \left(1 - \sum_{j=1}^p a_j \right) \\ \mathbb{E}(X_{t-k} X_{t-j}) = m_X^2 + \sum_{k=1}^p a_k [\mathbb{E}(X_t X_{t-k}) - m_X^2] \quad \forall j = 1, \dots, p \end{cases}
\end{aligned}$$

On a donc

$$\forall j = 1, \dots, p \quad \mathbb{E}(X_t X_{t-j}) - m_X^2 = \sum_{k=1}^p a_k [\mathbb{E}(X_{t-k} X_{t-j}) - m_X^2]$$

Soit encore

$$\forall j = 1, \dots, p \quad \text{Cov}(X_t, X_{t-j}) = \sum_{k=1}^p a_k \text{Cov}(X_{t-k+j}, X_{t-j})$$

$$\gamma(j) = \sum_{k=1}^p a_k \gamma(k-j)$$

Et

$$a_0 = m_X \left(1 - \sum_{j=1}^p a_j \right)$$

$$\begin{pmatrix} 1 & \gamma(1) & \dots & \gamma(p-1) \\ \gamma(1) & 1 & \dots & \gamma(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(p-1) & \gamma(p-2) & \dots & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix} = \begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(p) \end{pmatrix}$$

Puis en divisant par $\gamma(0)$

$$\Leftrightarrow \begin{pmatrix} 1 & \rho(1) & \dots & \rho(p-1) \\ \rho(1) & 1 & \dots & \rho(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(p-1) & \rho(p-2) & \dots & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \vdots \\ \rho(p) \end{pmatrix}$$

Cette dernière matrice étant inversible si les X_t sont indépendants.

Définition 1.2.4 (Propriété)

On appelle **auto-corrélation partielle** d'ordre p

$$\begin{aligned} r(p) &= \text{Corr}(X_t - EL(X_t|X_{t-1}, \dots, X_{t-p+1}), X_{t-p} - EL(X_{t-p}|X_{t-1}, \dots, X_{t-p+1})) \\ &= \frac{\text{Cov}(X_t - EL(X_t|X_{t-1}, \dots, X_{t-p+1}), X_{t-p} - EL(X_{t-p}|X_{t-1}, \dots, X_{t-p+1}))}{[\text{Var}(X_t - EL(X_t|X_{t-1}, \dots, X_{t-p+1}))\text{Var}(X_{t-p} - EL(X_{t-p}|X_{t-1}, \dots, X_{t-p+1}))]^{1/2}} \end{aligned}$$

On montre que $r(p) = a_p$ coefficient de X_{t-p} dans $EL(X_t|X_{t-1}, \dots, X_{t-p})$.

$$EL(X_{t-p}|X_{t-1}, \dots, X_{t-p+1}) = \sum_{j=1}^p a_j X_{t-j}$$

Pour la démonstration, on utilise le théorème de FRISH-WAUGH.

Définition 1.2.5 (Auto-corrélogramme partiel) L' **auto-corrélogramme partiel** de $(X_t)_{t \in \mathbb{Z}}$ est le graphe de :

$$\begin{cases} \mathbb{N} & \rightarrow]-1; 1[\\ p & \mapsto r(p) \end{cases}$$

1.2.3 Régression linéaire théorique sur un nombre infini de retards

Définition 1.2.6 Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire.

- (i) La **régression linéaire théorique** de X_t sur $X_{t-1}, \dots, X_{t-p}, \dots$ est la projection orthogonale dans $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ de X_t sur $H = \text{Vect}(X_{t-1}, \dots, X_{t-p}, \dots)$.
- (ii) La **régression affine théorique** de X_t sur $1, X_{t-1}, \dots, X_{t-p}, \dots$ est la projection orthogonale dans $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ de X_t sur $H^* = \text{Vect}(1, X_{t-1}, \dots, X_{t-p}, \dots)$.

On note aussi $\bar{\mathcal{L}}(\underline{X}_{t-1})$ l'espace $\bar{\mathcal{L}}(1, X_{t-1}, \dots, X_{t-k}, \dots)$ et :

$$\begin{aligned} EL(X_t|\underline{X}_{t-1}) &= EL(X_t|X_{t-1}, \dots, X_{t-k}, \dots) \\ &(\quad = EL(X_t|1, X_{t-1}, \dots, X_{t-k}, \dots)) \end{aligned}$$

la régression linéaire (ou affine) sur $\bar{\mathcal{L}}(\underline{X}_{t-1})$.

Proposition 1.2.2 Les deux notions coïncident si et seulement si $\mathbb{E}X_t = 0, \forall t$.

Remarque 1.2.2 $X_t^* = EL(X_t|X_{t-1}, \dots, X_{t-p})$

$$\begin{aligned} \|X_t - X_t^*\|_2 &= \min_{a_0, \dots, a_p} \left\| X_t - \left(a_0 + \sum_{j=1}^p a_j X_{t-j} \right) \right\|_2 \\ &= \min_{Y \in H} \|X_t - Y\|_2 \end{aligned}$$

Proposition 1.2.3 $EL(X_t|\underline{X}_{t-1}) = \lim_{n \rightarrow +\infty} EL(X_t|X_{t-1}, \dots, X_{t-n})$ au sens de \mathcal{L}^2 .

Théorème 1.2.1 (Admis) Soient $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire et $X_t^* = EL(X_t|\underline{X}_{t-1})$ la régression affine de X_t sur $\overline{\mathcal{L}}(1, X_{t-1}, \dots, X_{t-k}, \dots)$ et $\varepsilon_t = X_t - X_t^*$, alors

- $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc
- $\mathbb{C}ov(\varepsilon_t, \varepsilon_{t-k}) = 0 \quad \forall k > 0$

Définition 1.2.7 (Processus des innovations) Avec les notations du théorème ci-dessus :

- (i) $(\varepsilon_t)_{t \in \mathbb{Z}}$ est le **processus des innovations** de $(X_t)_{t \in \mathbb{Z}}$
- (ii) ε_t est l'**innovation** de X_t
- (iii) X_t^* est la **prévision optimale** de X_t à la date $t-1$

Remarque 1.2.3 $\varepsilon_t = X_t - X_t^* = X_t - EL(X_t|\underline{X}_{t-1})$

donc $\varepsilon_t \perp 1$ et $\varepsilon_t \perp X_{t-k}$, $\forall k > 0$, ce qui peut aussi s'interpréter comme :

$$\mathbb{E}(\varepsilon_t) = 0$$

$$\forall k > 0, \mathbb{E}(\varepsilon_t X_{t-k}) = \mathbb{C}ov(\varepsilon_t, X_{t-k}) = 0$$

Théorème 1.2.2 (De Wold) Soient $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire et $(\varepsilon_t)_{t \in \mathbb{Z}}$ le processus des innovations correspondant.

Alors

$$\exists (a_k)_{k \in \mathbb{Z}} / \sum_{k=0}^{+\infty} |a_k| < +\infty \text{ et } X_t = m + \sum_{k=0}^{+\infty} a_k \varepsilon_{t-k}$$

1.2.4 Densité spectrale et auto-corrélations inverses

Proposition 1.2.4 (Densité spectrale) Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire de la forme :

$$X_t = m + \sum_{j=0}^{+\infty} a_j \varepsilon_{t-j} \text{ où } (\varepsilon_t)_{t \in \mathbb{Z}} \rightsquigarrow \mathcal{BB} \text{ et } \sum_{j=0}^{+\infty} |a_j| < +\infty$$

Alors

$$(i) \sum_{h \in \mathbb{Z}} |\gamma_X(h)| < +\infty$$

$$(ii) \forall \omega \in [-\pi; \pi], \quad \boxed{f_X(\omega) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma_X(h) e^{i\omega h}}$$

f_X est la **densité spectrale** de $(X_t)_{t \in \mathbb{Z}}$.

Démonstration

$$\sum_{h \in \mathbb{Z}} |\gamma_X(h)| = \sum_{h \in \mathbb{Z}} \left| \sum_{j,k} a_j a_k \gamma_\varepsilon(h+j-k) \right|$$

Et on a

$$\gamma_\varepsilon(h+j-k) = \begin{cases} 0 & \text{si } h+j-k \neq 0 \\ \sigma_\varepsilon^2 & \text{si } h+j-k = 0 \end{cases}$$

Donc

$$\begin{aligned} \sum_{h \in \mathbb{Z}} |\gamma_X(h)| &= \sum_{h \in \mathbb{Z}} \left| \sigma_\varepsilon^2 \sum_j a_j a_{h+j} \right| \\ &\leq \sigma_\varepsilon^2 \sum_{h,j} |a_j| |a_{h+j}| = \sigma_\varepsilon^2 \left(\sum_j a_j \right)^2 < +\infty \end{aligned}$$

Proposition 1.2.5 *Sous les hypothèses précédentes,*

$$f_X(\omega) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma_X(h) \cos(\omega h)$$

Démonstration

$$\begin{aligned} f_X(\omega) &= \frac{1}{2\pi} \left[\gamma_X(0) + \sum_{h>0} \gamma_X(h) e^{i\omega h} + \sum_{h<0} \gamma_X(h) e^{i\omega h} \right] \\ &= \frac{1}{2\pi} \left[\gamma_X(0) + \sum_{h>0} \gamma_X(h) e^{i\omega h} + \sum_{h>0} \underbrace{\gamma_X(-h)}_{=\gamma_X(h)} e^{-i\omega h} \right] \\ &= \frac{1}{2\pi} \left[\gamma_X(0) + \sum_{h>0} \gamma_X(h) \underbrace{(e^{i\omega h} + e^{-i\omega h})}_{=2 \cos(\omega h)} \right] \\ &= \frac{1}{2\pi} \left[\gamma_X(0) + \sum_{h \neq 0} \gamma_X(h) \cos(\omega h) \right] \\ &= \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma_X(h) \cos(\omega h) \end{aligned}$$

Exemple 1.2.1 (1) $(\varepsilon_t)_{t \in \mathbb{Z}} \sim \mathcal{BB}(0, \sigma^2) \Rightarrow f_\varepsilon(\omega) = \frac{\sigma_\varepsilon^2}{2\pi}$

(2) $(X_t)_{t \in \mathbb{Z}} \sim MA(1) \Rightarrow f_\varepsilon(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} (1 + \theta^2 - 2\theta \cos \omega)$

Théorème 1.2.3 (Injectivité) *Avec les notations précédentes,*

$$\forall h \in \mathbb{Z}, \quad \gamma_X(h) = \int_{[-\pi; \pi]} f_X(\omega) e^{-i\omega h} d\omega = \int_{[-\pi; \pi]} f_X(\omega) \cos(\omega h) d\omega$$

Démonstration

$$\begin{aligned}
\int_{[-\pi;\pi]} f_X(\omega) e^{-i\omega h} d\omega &= \frac{1}{2\pi} \int_{[-\pi;\pi]} \left(\sum_{k \in \mathbb{Z}} \gamma_X(k) e^{i\omega k} \right) e^{-i\omega h} d\omega \\
&= \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \gamma_X(k) \underbrace{\left(\int_{[-\pi;\pi]} e^{i\omega(k-h)} d\omega \right)}_{= \begin{cases} 0 & \text{si } k \neq h \\ 2\pi & \text{si } k = h \end{cases}} \quad (\text{d'après FUBINI}) \\
&= \gamma_X(h)
\end{aligned}$$

Conséquence La fonction $f_X \mapsto \gamma_X$ est une bijection donc $(X_t)_t$ est caractérisé complètement par f_X .

Exemple 1.2.2 (1) Si $f_X(\omega) = \text{constante}$ alors $X \sim \mathcal{BB}$

(2) Si $f_X(\omega) = a + 2b \cos \omega$ alors $X \sim MA(1)$

$$\text{où } \begin{cases} a = \frac{\sigma^2}{2\pi} (1 + \theta^2) \\ b = -\frac{\sigma^2}{2\pi} \end{cases}$$

Proposition 1.2.6 Soit $(X_t)_{t \in \mathbb{Z}}$ tel que $X_t = \sum_{j \in \mathbb{Z}} a_j \varepsilon_{t-j}$ où $\varepsilon_t \sim \mathcal{BB}$ et $\sum_j |a_j| < +\infty$.

Soit $Y_t = \sum_{k \in \mathbb{Z}} b_k X_{t-k}$ avec $\sum_k |b_k| < +\infty$

Alors

$$(i) \quad Y_t = \sum_{k \in \mathbb{Z}} c_k \varepsilon_{t-k}$$

$$(ii) \quad f_Y(\omega) = f_X(\omega) \left| \sum_{k \in \mathbb{Z}} b_k e^{i\omega k} \right|^2$$

Démonstration

(i)

$$Y_t = \sum_{k \in \mathbb{Z}} b_k X_{t-k} = \sum_{k \in \mathbb{Z}} b_k \left(\sum_{j \in \mathbb{Z}} a_j \varepsilon_{t-k-j} \right) = \sum_{j, k \in \mathbb{Z}} a_j b_k \varepsilon_{t-(k+j)} = \sum_{j, h \in \mathbb{Z}} a_j b_{h-j} \varepsilon_{t-h} = \sum_{h \in \mathbb{Z}} \underbrace{\left(\sum_{j \in \mathbb{Z}} a_j b_{h-j} \right)}_{c_h} \varepsilon_{t-h}$$

(ii)

$$\begin{aligned}
f_Y(\omega) &= \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma_Y(h) e^{i\omega h} \\
&= \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \left(\sum_{j, k \in \mathbb{Z}} b_j b_k \gamma_X(h + j - k) \right) e^{i\omega h} \\
&= \frac{1}{2\pi} \sum_{h, j, k \in \mathbb{Z}} b_j b_k \gamma_X(h + j - k) e^{i\omega(h+j-k)} e^{-i\omega j} e^{i\omega k} \\
&= \frac{1}{2\pi} \left(\sum_{l \in \mathbb{Z}} \gamma_X(l) e^{i\omega l} \right) \left(\sum_{j \in \mathbb{Z}} b_j e^{i\omega j} \right) \left(\sum_{k \in \mathbb{Z}} b_k e^{-i\omega k} \right) \\
&= f_X(\omega) \left| \sum_{k \in \mathbb{Z}} b_k e^{i\omega k} \right|^2
\end{aligned}$$

Définition 1.2.8 (Auto-corrélations inverses) Soit $(X_t)_{t \in \mathbb{Z}}$ tel que $X_t = \sum_{j \in \mathbb{Z}} a_j \varepsilon_{t-j}$ où $\varepsilon_t \sim \mathcal{BB}$ et $\sum_j |a_j| < +\infty$.

On suppose que $\omega \mapsto \frac{1}{f_X(\omega)} e^{-i\omega h}$ est intégrable sur $[-\pi; \pi]$.

On appelle **auto-covariance inverse** d'ordre h de $(X_t)_{t \in \mathbb{Z}}$

$$\gamma_X^i(h) = \int_{[-\pi; \pi]} \frac{1}{f_X(\omega)} e^{-i\omega h} d\omega$$

L'**auto-corrélation inverse** d'ordre h est alors définie comme

$$\rho_X^i(h) = \frac{\gamma_X^i(h)}{\gamma_X^i(0)}$$

Définition 1.2.9 (Auto-corrélogramme inverse) L'**auto-corrélogramme inverse** de $(X_t)_{t \in \mathbb{Z}}$ est le graphe de :

$$\begin{cases} \mathbb{N} & \rightarrow]-1; 1[\\ h & \mapsto \rho_X^i(h) \end{cases}$$

1.2.5 Estimateurs associés et lois limites

On considère un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ tel que pour tout $t \in \mathbb{Z}$, $\mathbb{E}X_t = m$. On cherche à estimer les grandeurs associées $\gamma_X(h)$, $\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}$, $r_X(h) = a_h^h$, $f_X(\omega)$, $\gamma_X^i(h)$ et $\rho_X^i(h)$ et ceci sachant qu'on observe X_1, \dots, X_T .

On prend comme estimateurs :

- $\hat{m} = \frac{1}{T} \sum_{t=1}^T X_t = \bar{X}_T$ moyenne empirique,
- $\hat{\gamma}_X(h) = \frac{1}{T-h} \sum_{t=h+1}^T (X_t - \bar{X}_T)(X_{t-h} - \bar{X}_T)$ auto-covariance empirique d'ordre h (estimation acceptable si h n'est pas trop grand),
- $\hat{\rho}_X(h) = \frac{\hat{\gamma}_X(h)}{\hat{\gamma}_X(0)}$,

- $\hat{r}(h) = \hat{a}_h^h$ dans la régression empirique (m.c.o.) de x_t sur $1, x_{t-1}, \dots, x_{t-h}$,
- $\hat{f}_X(\omega) = \frac{1}{2\pi} \sum_{h=-H}^H \hat{\gamma}_X(h) e^{i\omega h}$, le problème ici étant que l'on voudrait un H suffisamment grand mais prendre un H trop grand est risqué pour l'estimation de $\hat{\gamma}_X(h)$. On prend alors un estimateur corrigé :

$$\hat{f}_X(\omega) = \frac{1}{2\pi} \sum_{h=-H}^H \underbrace{\left(1 - \frac{|h|}{H+1}\right)}_{\text{coefficient de NEWEY-WEST}} \hat{\gamma}_X(h) e^{i\omega h}$$

On donne moins de poids aux $\hat{\gamma}_X(h) e^{i\omega h}$ avec un grand h .

- on ne prend pas $\hat{\rho}_i = \frac{\hat{\gamma}^i(h)}{\hat{\gamma}^i(0)}$ où $\hat{\gamma}^i(h) = \int_{[-\pi; \pi]} \frac{1}{\hat{f}_X(\omega)} e^{-i\omega h} d\omega$, il existe d'autres façons de l'obtenir.

Proposition 1.2.7 *Si (X_t) est un processus stationnaire alors tous les estimateurs présentés ci-dessus sont convergents.*

Démonstration C'est la loi des grands nombres.

Proposition 1.2.8 *Si $X_t = m + \sum_{j=0}^{+\infty} a_j \varepsilon_{t-j}$ où $\mathbb{E}(\varepsilon_t^4) = \eta < +\infty$, alors tous ces estimateurs ont des lois jointes asymptotiquement gaussiennes :*

$$\begin{aligned} \sqrt{T}(\hat{m} - m) &\xrightarrow{L} \mathcal{N}(0, \sum_{h \in \mathbb{Z}} \gamma_X(h)) \\ \sqrt{T} \begin{pmatrix} \hat{\gamma}(0) - \gamma(0) \\ \vdots \\ \hat{\gamma}(h) - \gamma(h) \end{pmatrix} &\xrightarrow{L} \mathcal{N}(0, \Omega_h) \\ \sqrt{T} \begin{pmatrix} \hat{\rho}(0) - \rho(0) \\ \vdots \\ \hat{\rho}(h) - \rho(h) \end{pmatrix} &\xrightarrow{L} \mathcal{N}(0, \mathcal{W}_h) \\ \sqrt{T} \begin{pmatrix} \hat{r}(0) - r(0) \\ \vdots \\ \hat{r}(h) - r(h) \end{pmatrix} &\xrightarrow{L} \mathcal{N}(0, \Sigma_h) \end{aligned}$$

Ω_h, \mathcal{W}_h et Σ_h étant calculables.

Remarque 1.2.4 (1) Les auto-corrélogrammes (direct, partiel et inverse) associés aux valeurs estimées sont appelés **auto-corrélogrammes empiriques**.

(2) On en déduit des intervalles de confiance asymptotiques.

1.3 Polynômes retard et avance

1.3.1 Définitions et propositions

Définition 1.3.1 (i) *L'opérateur retard L (lag) ou B (backward) est défini sur la classe des processus stationnaires comme étant :*

$$L : (X_t)_{t \in \mathbb{Z}} \mapsto (Y_t)_{t \in \mathbb{Z}} \text{ tel que } Y_t = X_{t-1}$$

On note : $LX_t = X_{t-1}$.

(ii) De la même façon, l'**opérateur avance** F (forward) correspond à

$$F : (X_t)_{t \in \mathbb{Z}} \mapsto (Y_t)_{t \in \mathbb{Z}} \text{ tel que } Y_t = X_{t+1}$$

On note : $FX_t = X_{t+1}$.

Proposition 1.3.1 (i) $L^k = \underbrace{L \circ \dots \circ L}_{k \text{ fois}}$ vérifie $L^k X_t = X_{t-k}$.

(ii) $F^k = \underbrace{F \circ \dots \circ F}_{k \text{ fois}}$ vérifie $F^k X_t = X_{t+k}$.

Notation : $L^0 = Id$ est noté $L^0 = 1$ ($L^0 X_t = X_t$).

Définition 1.3.2 Soit P un polynôme, $P(z) = \sum_{k=0}^p a_k z^k$, $a_k \in \mathbb{R}$, on lui associe le **polynôme retard** $P(L)$ défini comme suit :

$$P(L) = \sum_{k=0}^p a_k L^k$$

Et

$$P(L)X_t = \left(\sum_{k=0}^p a_k L^k \right) X_t = \sum_{k=0}^p a_k X_{t-k}$$

De façon similaire on obtient le **polynôme avance** $P(F)$:

$$P(F)X_t = \left(\sum_{k=0}^p a_k F^k \right) X_t = \sum_{k=0}^p a_k X_{t+k}$$

Définition 1.3.3 (Séries en L (ou en F)) (polynômes de degré infini) Soit $A(z) = \sum_{k=0}^{+\infty} a_k z^k$ et $Y_t = A(L)X_t = \sum_{k=0}^{+\infty} a_k X_{t-k}$.

Alors $(Y_t)_{t \in \mathbb{Z}}$ est bien un processus stationnaire car $\sum_{k=0}^{+\infty} |a_k| < +\infty$.

Proposition 1.3.2 On suppose $\sum_{k=0}^{+\infty} |a_k| < +\infty$ et $\sum_{k=0}^{+\infty} |b_k| < +\infty$ et

$$A(L) = \sum_{k=0}^{+\infty} a_k L^k \quad B(L) = \sum_{k=0}^{+\infty} b_k L^k$$

Alors

$$(i) \forall \alpha \in \mathbb{R}, \quad \alpha A(L) = (\alpha A)(L) = \alpha \left(\sum_{k=0}^{+\infty} a_k L^k \right) = \sum_{k=0}^{+\infty} (\alpha a_k) L^k$$

$$(ii) A(L) + B(L) = (A + B)(L) = \left(\sum_{k=0}^{+\infty} a_k L^k \right) + \left(\sum_{k=0}^{+\infty} b_k L^k \right) = \sum_{k=0}^{+\infty} (a_k + b_k) L^k$$

$$(iii) A(L) \circ B(L) = (AB)(L) = B(L) \circ A(L) \text{ avec } (AB)(L) = (BA)(L) = C(L) = \sum_{k=0}^{+\infty} c_k L^k \text{ et}$$

$$c_k = \sum_{j=0}^{+\infty} a_j b_{k-j}$$

1.3.2 Inversibilité des polynômes en L

Définition 1.3.4 (Inversibilité) $A(L)$ est inversible $\Leftrightarrow \exists B(L)$ tel que $A(L) \circ B(L) = Id$

On suppose $P(L) = \sum_{k=0}^p a_k L^k$ et $Y_t = P(L)X_t$ et on désire savoir si X_t peut s'exprimer en fonction de Y_t ($X_t = P(L)^{-1}Y_t$).

On peut décomposer notre polynôme de la façon suivante :

$$P(z) = \prod_{i=1}^p (z - z_i) = \prod_{i=1}^p (-z_i) \prod_{i=1}^p \left(1 - \frac{z}{z_i}\right) = \alpha \prod_{i=1}^p (1 - \lambda_i z) \text{ avec } \lambda_i = \frac{1}{z_i}$$

où les $z_i \in \mathbb{C}$ sont les racines de P .

$$\text{Finalement } P(L) = \alpha \prod_{i=1}^p (1 - \lambda_i L)$$

Inversibilité de $1 - \lambda L$

Proposition 1.3.3 (i) Si $|\lambda| < 1$, alors $1 - \lambda L$ est inversible et $(1 - \lambda L)^{-1} = \sum_{k=0}^{+\infty} \lambda^k L^k$

(ii) Si $|\lambda| > 1$, $1 - \lambda L$ est inversible et $(1 - \lambda L)^{-1} = -\sum_{k=1}^{+\infty} \frac{1}{\lambda^k} F^k$

(iii) Si $|\lambda| = 1$, $1 - \lambda L$ n'est pas inversible

Démonstration

(i) Si $|\lambda| < 1$ alors $|(1 - \lambda L)^{-1}| \leq \sum_{k=0}^{+\infty} |\lambda^k| = \frac{1}{1-|\lambda|} < +\infty$, donc $A(L) = \sum_{k=0}^{+\infty} \lambda^k L^k$ est bien défini.

On a ainsi $(1 - \lambda L)A(L) = C(L) = \sum_{j=0}^{+\infty} c_j L^j$ et $c_j = \sum_{k=0}^{+\infty} a_k b_{k-j}$ avec $b_0 = 1$, $b_1 = -\lambda$, $b_k = 0$ si $k > 1$ et $a_k = \lambda^k$. On trouve $c_0 = 1$ et $c_j = 0$ si $j \neq 0$, soit encore $C(L) = 1$. On en déduit que $(1 - \lambda L)$ est inversible et $(1 - \lambda L)^{-1} = A(L)$.

On pouvait aussi montrer ce résultat en écrivant :

$$(1 - \lambda L)A(L) = \lim_{k \rightarrow +\infty} (1 - \lambda L) \left(\sum_{j=0}^k \lambda^j L^j \right) = \lim_{k \rightarrow +\infty} 1 - \lambda^{k+1} L^{k+1} = 1$$

(ii) Si $|\lambda| > 1$ alors $1 - \lambda L = -\lambda \left(L - \frac{1}{\lambda} \right) = -\lambda L \left(1 - \frac{F}{\lambda} \right)$.

On a alors

$$(\lambda L)^{-1} = \frac{1}{\lambda} F \text{ et } \left(1 - \frac{F}{\lambda} \right)^{-1} = \sum_{k=0}^{+\infty} \frac{1}{\lambda^k} F^k \text{ car } |\lambda| > 1$$

En combinant ces deux résultats, on obtient

$$(1 - \lambda L) = (-\lambda L)^{-1} \left(1 - \frac{F}{\lambda} \right)^{-1} = -\frac{1}{\lambda} F \left(\sum_{k=0}^{+\infty} \frac{1}{\lambda^k} F^k \right) = -\sum_{k=1}^{+\infty} \frac{1}{\lambda^k} F^k = -\sum_{k=-\infty}^{-1} \lambda^k L^k$$

Dans ce cas, $X_t = (1 - \lambda L)^{-1}Y_t = -\sum_{k=1}^{+\infty} \frac{1}{\lambda^k} Y_{t+k}$.

(iii) Cas $\lambda = 1$.

Alors $1 - L$ n'est pas inversible. Montrons-le par l'absurde.

Supposons $Y_t = (1 - L)X_t = X_t - X_{t-1}$. On a vu que si $(Y_t)_{t \in \mathbb{Z}}$ est stationnaire alors $(X_t)_{t \in \mathbb{Z}}$ ne l'est pas (cf. page 4, l'exemple où $Y_t = \varepsilon_t \sim \mathcal{BB}(0, \sigma^2)$).

- On n'a pas $X_t = \sum_{k \in \mathbb{Z}} a_k Y_{t-k}$ avec $\sum_{k \in \mathbb{Z}} |a_k| < +\infty$;
- Il n'existe pas $A(L) = \sum_{k \in \mathbb{Z}} a_k L^k$, $\sum_{k \in \mathbb{Z}} |a_k| < +\infty$ tel que $(1 - L)A(L) = 1$.

On peut le voir à la main

$$(1 - L)A(L) = 1 \Rightarrow |a_k| = |a_{k-1}| \text{ et donc ne tend pas vers } 0$$

Dans ces conditions $\sum_{k \in \mathbb{Z}} |a_k| = +\infty$.

Inversion d'un polynôme en L

Soit ϕ un polynôme de degré p à coefficients réels :

$$\phi(z) = 1 + \varphi_1 z + \dots + \varphi_p z^p$$

$$\phi(L) = 1 + \varphi_1 L + \dots + \varphi_p L^p \quad (\phi(0) = 1)$$

ϕ possède p racines (z_1, \dots, z_p) dans complexes ou réelles, on peut donc le décomposer en

$$\phi(z) = \varphi_p \prod_{j=1}^p (z - z_j) = \varphi_p \prod_{j=1}^p (-z_j) \prod_{j=1}^p \left(1 - \frac{z}{z_j}\right) = \alpha \prod_{j=1}^p (1 - \lambda_j z)$$

où $\lambda_j = \frac{1}{z_j}$

Par conséquent, on peut se ramener à :

$$\phi(L) = \prod_{j=1}^p (1 - \lambda_j L)$$

On a alors 2 cas possibles :

- si $z_i \in \mathbb{R}$ alors $\lambda_i \in \mathbb{R}$,
- si $z_i \in \mathbb{C} - \mathbb{R}$ alors \bar{z}_i racine de ϕ de même ordre de multiplicité que z_i .

$$\phi(z) = (1 - \lambda_i z)(1 - \bar{\lambda}_i z)\psi(z)$$

(i) Si $|\lambda_i| < 1$ alors $|\bar{\lambda}_i| < 1$ et

$$\begin{aligned} ((1 - \lambda_i L)(1 - \bar{\lambda}_i L))^{-1} &= (1 - \lambda_i L)^{-1}(1 - \bar{\lambda}_i L)^{-1} \\ &= \left(\sum_k \lambda_i^k L^k \right) \left(\sum_k \bar{\lambda}_i^k L^k \right) \\ &= A(L) \times \bar{A}(L) \end{aligned}$$

(ii) $|\lambda_i| > 1$, idem avec $(1 - \lambda_i L)^{-1} = -\sum_{-\infty}^{-1} \lambda_i^k L^k$

(iii) Si $|\lambda_i| = 1$ alors $\phi(L)$ n'est pas inversible.

Proposition 1.3.4 Avec les notations précédentes

- (i) ϕ est inversible si et seulement si ses racines sont de module distinct de 1.
(ii) Si $|\lambda_j| < 1, \forall j \in [1, p]$, alors $\phi(L)$ est inversible et

$$\phi(L)^{-1} = \sum_{k=0}^{+\infty} a_k L^k \text{ où } a_0 = 1, a_k \in \mathbb{R} \text{ et } \sum_{k=0}^{+\infty} |a_k| < +\infty$$

Remarque 1.3.1 $|\lambda_j| < 1 \Leftrightarrow |z_j| = \frac{1}{\lambda_j} > 1$

Démonstration

- (i) $\forall j, (1 - \lambda_j L)^{-1}$ est bien défini, de la forme $\sum_{k \in \mathbb{Z}} a_{j,k} L^k$ et $\phi(L)^{-1} = \prod_{j=1}^p (1 - \lambda_j L)^{-1}$ est donc

aussi défini.

Mais $\phi(L)^{-1}$ peut contenir des termes en $L^k, k > 0$ qui sont des termes concernant le futur et donc peu utilisables en pratique.

- (ii) Si $|\lambda_j| < 1$ pour tout j alors $(1 - \lambda_j L)^{-1} = \sum_{k=0}^{+\infty} \lambda_j^k L^k$ et

$$\phi(L)^{-1} = \prod_{j=1}^p (1 - \lambda_j L)^{-1} = \sum_{k=0}^{+\infty} a_k L^k \text{ tel que } \sum_{k=0}^{+\infty} |a_k| < +\infty$$

Par ailleurs

$$\phi(z) = \prod_{j=1}^p (1 - \lambda_j z) \text{ et } \phi(z)\phi(z)^{-1} = 1 \Leftrightarrow \prod_{j=1}^p (1 - \lambda_j z) \left(\sum_{k=0}^{+\infty} a_k z^k \right) = 1$$

Donc

$$\phi(0)\phi(0)^{-1} = 1 \times a_0 = 1 \Rightarrow a_0 = 1$$

S'il existe j tel que $\lambda_j \in \mathbb{C} \setminus \mathbb{R}$ alors $\phi(L) = (1 - \lambda_j)(1 - \bar{\lambda}_j)P(L)$ et :

$$\begin{aligned} (1 - \lambda_j)^{-1}(1 - \bar{\lambda}_j)^{-1} &= \left(\sum_{k=0}^{+\infty} \lambda_j^k L^k \right) \left(\sum_{k=0}^{+\infty} \bar{\lambda}_j^k L^k \right) \\ &= \sum_{k=0}^{+\infty} \alpha_k L^k \text{ où } \alpha_k \in \mathbb{R}, \alpha_0 = 1, \sum_{k=0}^{+\infty} |\alpha_k| < +\infty \end{aligned}$$

Méthodes pratiques d'inversion de $\phi(L)$

On se place dans le cadre défini précédemment où :

$$\phi(L) = \prod_{j=1}^p (1 - \lambda_j L)$$

a) Quand $p \leq 2$, $\phi(L)^{-1} = \prod_{j=1}^p \left(\sum_{k=0}^{+\infty} \lambda_j^k L^k \right)$
 Cette méthode s'avère fastidieuse en général.

b) Par identification : on écrit que

$$\phi(L) \left(\sum_{k=0}^{+\infty} a_k L^k \right) = (1 + \varphi_1 L + \dots + \varphi_p L^p) \left(\sum_{k=0}^{+\infty} a_k L^k \right) = 1$$

Les a_k sont obtenus par récurrence puis identification.

c) Décomposition en éléments simples :

$$\phi(L)^{-1} = \prod_{j=1}^p \frac{1}{1 - \lambda_j L} = \sum_{j=1}^p a_j \frac{1}{1 - \lambda_j L}$$

On décompose cette fraction rationnelle en éléments simples. Dans la pratique on l'utilise quand les racines sont simples.

d) Division selon les puissances croissantes de 1 par $\phi(z)$:

$$1 = \phi(z)Q_r(z) + z^{r+1}R_r(z)$$

$$\text{tel que } \lim_{r \rightarrow +\infty} Q_r(z) = \phi^{-1}(z)$$

Chapitre 2

Processus $ARMA$ et $ARIMA$

Les $ARMA$ sont des processus stationnaires et les $ARIMA$ des processus non stationnaires intégrés, c'est-à-dire qu'on les rend stationnaires par différenciation.

2.1 Processus auto-régressifs d'ordre p ($AR(p)$)

2.1.1 Définition et représentation canonique

Définition 2.1.1 (Processus AR) $(X_t)_{t \in \mathbb{Z}}$ est un processus $AR(p)$ si

(i) (X_t) est stationnaire

(ii) (X_t) vérifie une équation $X_t = \mu + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + \varepsilon_t$ avec $\varphi_p \neq 0$ et $\varepsilon_t \sim \mathcal{BB}(0, \sigma^2)$

On note $\phi(L)X_t = \mu + \varepsilon_t$ où $\phi(L) = 1 - (\varphi_1 L + \dots + \varphi_p L^p)$

Exemple 2.1.1 $X_t \sim AR(1)$ i.e. $(1 - \rho L)X_t = \mu + \varepsilon_t$ où $\varepsilon_t \sim \mathcal{BB}(0, \sigma^2)$ et $|\rho| < 1$

Remarque 2.1.1 Il existe des solutions non stationnaires (en espérance) de la même équation.

Soit Y_t tel que $(1 - \rho L)Y_t = 0 \Rightarrow Y_t = \rho Y_{t-1} \Rightarrow Y_t = \rho^t Y_0$

Soit (X_t) un processus stationnaire.

On définit (Z_t) par $Z_t = X_t + Y_t$. On a alors

$$(1 - \rho L)Z_t = (1 - \rho L)X_t + (1 - \rho L)Y_t = \varepsilon_t + 0$$

$$\mathbb{E}Z_t = \mathbb{E}X_t + \mathbb{E}Y_t = m_X + \rho^t \mathbb{E}Y_0 \neq cte$$

Donc (Z_t) n'est pas un processus stationnaire.

Proposition 2.1.1 Si $X_t \sim AR(p)$ tel que $\phi(L)X_t = \mu + \varepsilon_t$, alors

$$\mathbb{E}X_t = \frac{\mu}{\phi(1)} = \frac{\mu}{1 - (\varphi_1 + \dots + \varphi_p)}$$

Démonstration

$$\begin{aligned} X_t &= \mu + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + \varepsilon_t \\ \mathbb{E}X_t &= \mu + \varphi_1 \mathbb{E}X_{t-1} + \dots + \varphi_p \mathbb{E}X_{t-p} + \mathbb{E}\varepsilon_t \\ m &= \mu + \varphi_1 m + \dots + \varphi_p m \\ m &= \frac{\mu}{1 - (\varphi_1 + \dots + \varphi_p)} = \frac{\mu}{\phi(1)} \end{aligned}$$

On retrouve bien le résultat annoncé.

Proposition 2.1.2 Si $X_t \rightsquigarrow AR(p)$ est tel que $\phi(L)X_t = \mu + \varepsilon_t$ et si l'on pose $Y_t = X_t - m$ (où $m = \mathbb{E}X_t$), on a alors

$$\phi(L)Y_t = \varepsilon_t \text{ et } \mathbb{E}Y_t = 0$$

Démonstration $\mathbb{E}X_t = m = \frac{\mu}{\phi(1)}$ et $\phi(L)(X_t - m) = \phi(L)X_t - \phi(L)m$
Or $Lm = m$ et par conséquent :

$$\phi(L)m = (1 - (\varphi_1 + \dots + \varphi_p))m = \phi(1)m$$

Finalement :

$$\phi(L)(X_t - m) = \phi(L)X_t - \mu = \varepsilon_t$$

Écriture $MA(\infty)$ quand les racines de ϕ sont de module strictement supérieur à 1

On suppose que $\phi(L)X_t = \mu + \varepsilon_t$ où $\phi(L) = 1 - (\varphi_1 L + \dots + \varphi_p L^p)$ et aussi que $|z| \leq 1 \Rightarrow \phi(z) \neq 0$.

On suppose que $\phi(z) = \prod_{i=1}^p (1 - \lambda_i z)$ où $|\lambda_i| = \frac{1}{|z_i|} < 1$.

Alors $\phi(L)$ est inversible et $\phi(L)^{-1} = \sum_0^\infty a_k L^k = A(L)$ tel que $\sum |a_k| < \infty$ et $a_0 = 1$.

On en déduit

$$\begin{aligned} X_t &= A(L)\mu + A(L)\varepsilon_t \\ &= A(1)\mu + \left(\sum_0^\infty a_k L^k \right) \varepsilon_t \\ &= m + \sum_0^\infty a_k L^k \varepsilon_{t-k} \end{aligned}$$

car $\phi(1)^{-1}\mu = m$

Proposition 2.1.3 Sous les hypothèses précédentes, $(X_t)_{t \in \mathbb{Z}}$ admet une représentation $MA(\infty)$ i.e. :

$$X_t = m + \sum_{k=0}^{+\infty} a_k \varepsilon_{t-k}, \text{ où } a_0 = 1, a_k \in \mathbb{R}, \sum_{k=0}^{+\infty} |a_k| < +\infty$$

Proposition 2.1.4 Sous les hypothèses précédentes :

(i) $\bar{\mathcal{L}}(X_t) = \bar{\mathcal{L}}(\varepsilon_t)$

(ii) ε_t est l'innovation de X_t

Rappel de notation

$$\begin{aligned} \bar{\mathcal{L}}(X_t) &= \bar{\mathcal{L}}(1, X_t, X_{t-1}, \dots, X_{t-p}, \dots) \\ \bar{\mathcal{L}}(\varepsilon_t) &= \bar{\mathcal{L}}(1, \varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-p}, \dots) \end{aligned}$$

Démonstration

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

On a vu que $X_t = \eta + \sum_{k=0}^{+\infty} a_t \varepsilon_{t-k}$

$$\Rightarrow X_t \in \overline{\mathcal{L}}(\underline{\varepsilon}_t) = \overline{\mathcal{L}}(1, \varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-k}, \dots)$$

Donc $\forall k \geq 0, X_{t-k} \in \overline{\mathcal{L}}(\underline{\varepsilon}_{t-k}) \subset \overline{\mathcal{L}}(\underline{\varepsilon}_t)$

$$\Rightarrow \mathcal{L}(1, X_t, X_{t-1}, \dots, X_{t-k}, \dots) \subset \overline{\mathcal{L}}(\underline{\varepsilon}_t)$$

$$\Rightarrow \mathcal{L}(X_t) \subset \overline{\mathcal{L}}(\underline{\varepsilon}_t)$$

$$\Rightarrow \overline{\mathcal{L}}(X_t) \subset \overline{\mathcal{L}}(\underline{\varepsilon}_t)$$

De la même façon, comme

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

on obtient l'inclusion réciproque et finalement $\overline{\mathcal{L}}(X_t) = \overline{\mathcal{L}}(\underline{\varepsilon}_t)$.

(ii) L'innovation de X_t vaut, par définition, $X_t - X_t^*$, or :

$$\begin{aligned} X_t^* &= EL(X_t | X_{t-1}) = EL(X_t | 1, X_{t-1}, \dots, X_{t-k}, \dots) \\ &= EL(\underbrace{\mu + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + \varepsilon_t}_{\in \overline{\mathcal{L}}(X_{t-1})} | X_{t-1}) \\ &= \mu + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + EL(\varepsilon_t | X_{t-1}) \end{aligned}$$

Comme $\overline{\mathcal{L}}(X_{t-1}) = \overline{\mathcal{L}}(\underline{\varepsilon}_{t-1})$, on a :

$$EL(\varepsilon_t | X_{t-1}) = EL(\varepsilon_t | \underline{\varepsilon}_{t-1}) = 0 \text{ car } \varepsilon_t \sim \mathcal{BB}$$

Finalement $X_t^* = \mu + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p}$ et $X_t - X_t^* = \varepsilon_t$, ε_t est bien l'innovation de X_t .

Définition 2.1.2 Soient $X_t \sim AR(p)$ et ϕ un polynôme vérifiant :

- $\phi(L)X_t = \mu + \varepsilon_t$
- $|z| \leq 1 \Rightarrow \phi(z) \neq 0$

On dit que la représentation $\phi(L)X_t = \mu + \varepsilon_t$ est la **représentation canonique** de $(X_t)_{t \in \mathbb{Z}}$.

Cas où ϕ admet des racines de module inférieur à 1

Remarque 2.1.2 (1) Si $(X_t)_{t \in \mathbb{Z}}$ est supposé stationnaire alors ϕ n'a pas de racines de module égal à 1.

(2) On sait que $\phi(L)$ est inversible, $\phi(L)^{-1} = \sum_{\mathbb{Z}} a_k L^k$. Mais on n'a plus l'égalité $\overline{\mathcal{L}}(X_t) = \overline{\mathcal{L}}(\underline{\varepsilon}_t)$.

L'écriture $\phi(L)X_t = \mu + \varepsilon_t$ ne met pas en évidence l'innovation de X_t . On cherche une autre représentation de (X_t) .

On peut écrire

$$\phi(L) = \prod_{j=1}^p (1 - \lambda_j L) = \left[\prod_{j/|\lambda_j| < 1} (1 - \lambda_j L) \right] \left[\prod_{j/|\lambda_j| > 1} (1 - \lambda_j L) \right]$$

On définit

$$\phi^*(z) = \left[\prod_{j/|\lambda_j| < 1} (1 - \lambda_j z) \right] \left[\prod_{j/|\lambda_j| > 1} \left(1 - \frac{z}{\lambda_j}\right) \right]$$

de telle sorte que ϕ^* a toutes ses racines de module strictement supérieur à 1.

On définit ensuite le processus $(\eta_t)_{t \in \mathbb{Z}}$ tel que $\eta_t = \phi^*(L)(X_t - m)$ où $m = \frac{\mu}{\phi(1)}$.

On montre alors que $\eta_t \sim \mathcal{BB}(0, \sigma_\eta^2)$ en calculant $f_\eta(\omega)$:

$$f_\eta(\omega) = f_X(\omega) |\phi^*(e^{i\omega})|^2$$

Comme $\phi(L)X_t = \varepsilon_t$, on a aussi :

$$f_\varepsilon(\omega) = f_X(\omega) |\phi(e^{i\omega})|^2 = \frac{\sigma_\varepsilon^2}{2\pi}$$

Ceci nous mène à :

$$\begin{aligned} f_\eta(\omega) &= \frac{\sigma_\varepsilon^2}{2\pi} \frac{1}{|\phi(e^{i\omega})|^2} - |\phi^*(e^{i\omega})|^2 \\ &= \frac{\sigma_\varepsilon^2}{2\pi} \frac{\left[\prod_{j/|\lambda_j| < 1} |1 - \lambda_j e^{i\omega}|^2 \right] \left[\prod_{j/|\lambda_j| > 1} \left|1 - \frac{e^{i\omega}}{\lambda_j}\right|^2 \right]}{\left[\prod_{j/|\lambda_j| < 1} |1 - \lambda_j e^{i\omega}|^2 \right] \left[\prod_{j/|\lambda_j| > 1} |1 - \lambda_j e^{i\omega}|^2 \right]} \\ &= \frac{\sigma_\varepsilon^2}{2\pi} \prod_{j, |\lambda_j| > 1} \frac{1}{|\lambda_j|^2} \frac{|\lambda_j - e^{i\omega}|^2}{|1 - \lambda_j e^{i\omega}|^2} \end{aligned}$$

Or

$$\prod_{j/|\lambda_j| > 1} \frac{|\lambda_j - e^{i\omega}|^2}{|1 - \lambda_j e^{i\omega}|^2} = 1$$

En effet :

– Si $\lambda_j \in \mathbb{R}$, $|1 - \lambda_j e^{i\omega}|^2 = |1 - \lambda_j e^{-i\omega}|^2 = |1 - \lambda_j e^{i\omega}|^2$

– Si $\lambda_j \in \mathbb{R} \setminus \mathbb{C}$, $\frac{|\lambda_j - e^{i\omega}|^2 |\bar{\lambda}_j - e^{i\omega}|^2}{|1 - \lambda_j e^{i\omega}|^2 |1 - \bar{\lambda}_j e^{i\omega}|^2} = 1$, $\bar{\lambda}_j$ étant aussi une racine de ϕ puisque celui-ci est à coefficients réels.

On a donc $f_\eta(\omega) = \alpha \frac{\sigma_\varepsilon^2}{2\pi} = \frac{\sigma_\eta^2}{2\pi}$ avec $\alpha = \prod_{j, |\lambda_j| > 1} \frac{1}{|\lambda_j|^2} < 1$ et finalement $\eta_t \sim \mathcal{BB}(0, \sigma^2)$ car

sa transformée de FOURIER est une constante.

Bilan La représentation $\phi^*(L)X_t = \phi^*(1)m + \eta_t = \mu^* + \eta_t$ est la représentation canonique de $(X_t)_{t \in \mathbb{Z}}$ car ϕ^* a toutes ses racines de module strictement supérieur à 1 et η_t est l'innovation de X_t .

2.1.2 Propriétés des processus AR(p)

On suppose que $\phi(L)X_t = \mu + \varepsilon_t$ où

- les racines de ϕ sont de module strictement supérieur à 1,
- ε_t suit un bruit blanc.

On peut se ramener ensuite à $\mu = 0$ par centrage car $\phi(L)(X_t - m) = \varepsilon_t$ où $m = \mu/\phi(1)$.

On considère donc le cas où $\phi(L)X_t = \varepsilon_t$ (et $\mathbb{E}X_t = 0$).

Auto-covariance, auto-corrélations et équivalence de Yule-Walker

- L'auto-covariance :

$\gamma(h) = \text{Cov}(X_t, X_{t-h}) = \mathbb{E}(X_t X_{t-h})$ pour $h \geq 0$ (car $m_X = 0$), et

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

Donc

$$\begin{aligned} X_t^2 &= \varphi_1 X_t X_{t-1} + \cdots + \varphi_p X_t X_{t-p} + X_t \varepsilon_t \\ \gamma(0) &= \varphi_1 \gamma(1) + \cdots + \varphi_p \gamma(p) + \mathbb{E}(X_t \varepsilon_t) \end{aligned}$$

Or

$$\mathbb{E}(X_t \varepsilon_t) = \underbrace{\mathbb{E}[(\varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p}) \varepsilon_t]}_{=0 \text{ car } \varepsilon_t \perp \overline{\mathcal{L}}(X_{t-1})} + \mathbb{E}(\varepsilon_t^2)$$

D'où

$$\gamma(0) = \varphi_1 \gamma(1) + \cdots + \varphi_p \gamma(p) + \sigma_\varepsilon^2$$

Si $h > 0$, on procède de la même façon :

$$\begin{aligned} X_t X_{t-h} &= \varphi_1 X_{t-1} X_{t-h} + \cdots + \varphi_p X_{t-p} X_{t-h} + \varepsilon_t X_{t-h} \\ \gamma(h) &= \varphi_1 \gamma(h-1) + \cdots + \varphi_p \gamma(h-p) + \underbrace{\mathbb{E}(\varepsilon_t X_{t-h})}_{=0 \text{ car } \varepsilon_t \perp X_{t-h}} \end{aligned}$$

- Les auto-corrélations :

A partir de la relation de récurrence de $\gamma(h)$ on déduit celle sur $\rho(h) = \frac{\gamma(h)}{\gamma(0)}$.

$$\rho(h) = \varphi_1 \rho(h-1) + \cdots + \varphi_p \rho(h-p), \forall h \geq 0$$

- Ces dernières équations sont appelées **équations de Yule-Walker**.

Pour $h > 0$, les $\gamma(h)$ et les $\rho(h)$ vérifient une relation de récurrence d'ordre p et

$$\begin{aligned} 1 &= \varphi_1 \rho(1) + \cdots + \varphi_p \rho(p) + \frac{\sigma_\varepsilon^2}{\gamma(0)} \\ \Rightarrow \gamma(0) &= \sigma_\varepsilon^2 \frac{1}{1 - (\varphi_1 \rho(1) + \cdots + \varphi_p \rho(p))} \end{aligned}$$

Les équations de YULE-WALKER pour $h = 1, \dots, p$ peuvent s'écrire :

$$\begin{pmatrix} 1 & \rho(1) & \cdots & \rho(p-1) \\ \rho(1) & 1 & \rho(1) & \rho(p-2) \\ \vdots & & \ddots & \vdots \\ \rho(p-1) & & \cdots & 1 \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_p \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \vdots \\ \rho(p) \end{pmatrix}$$

Les solutions de l'équation de récurrence sont complètement déterminées par la donnée de conditions initiales $\rho(1), \dots, \rho(p)$: elles permettent d'obtenir $\varphi_1, \dots, \varphi_p$. En particulier elles donneront une estimation préliminaire de $\hat{\varphi}_1, \dots, \hat{\varphi}_p$ en fonction de $\hat{\rho}_T(1), \dots, \hat{\rho}_T(p)$.

$$\begin{cases} \rho(1) & = & \varphi_1 + \varphi_2\rho(1) + \cdots + \varphi_p\rho(p-1) \\ & \cdots & \\ \rho(p) & = & \varphi_1\rho(p-1) + \cdots + \varphi_{p-1}\rho(1) + \varphi_p \end{cases}$$

$$\Leftrightarrow \begin{cases} \varphi_1 & = & (1 - \varphi_2)\rho(1) - \cdots - \varphi_p\rho(p-1) \\ & \cdots & \\ \varphi_p & = & \rho(p) - \varphi_1\rho(p-1) - \cdots + \varphi_{p-1}\rho(1) \end{cases}$$

On peut donc aussi obtenir $\rho(1), \dots, \rho(p)$ en fonction de $\varphi_1, \dots, \varphi_p$.

Proposition 2.1.5 *Si $X_t \rightsquigarrow AR(p)$ alors les $|\rho(h)|$ et les $\gamma(h)$ décroissent vers 0 exponentiellement avec h .*

Démonstration

$$\forall h > 0, \rho(h) - \varphi_1\rho(h-1) - \cdots - \varphi_p\rho(h-p) = 0$$

Le polynôme caractéristique de cette relation de récurrence est :

$$z^p - \varphi_1z^{p-1} - \cdots - \varphi_{p-1}z - \varphi_p = z^p \left(1 - \frac{\varphi_1}{z} - \cdots - \frac{\varphi_{p-1}}{z^{p-1}} - \frac{\varphi_p}{z^p} \right) = z^p \phi\left(\frac{1}{z}\right)$$

Avec $\phi(L)X_t = \varepsilon_t$ et $\phi(L) = 1 - \varphi_1L - \cdots - \varphi_pL^p$. Les racines du polynôme caractéristique sont les $\lambda_i = \frac{1}{z_i}$ (les z_i étant les racines de ϕ) avec $|\lambda_i| < 1$.

La forme générale de la solution est, si z_1, \dots, z_n sont des racines distinctes de ϕ de multiplicité respective m_1, \dots, m_n :

$$\rho(h) = \sum_{i=1}^n \sum_{k=0}^{m_i-1} \alpha_{ik} \lambda_i^k h^k$$

$|\lambda_i| < 1$ donc $\rho(h)$ décroît vers 0 exponentiellement avec h .

2.1.3 Auto-corrélations partielle et inverse d'un processus $AR(p)$

Proposition 2.1.6 *Si $(X_t)_{t \in \mathbb{Z}} \rightsquigarrow AR(p)$ et si $\phi(L)X_t = \mu + \varepsilon_t$ est sa représentation canonique, alors :*

$$r(h) = \begin{cases} 0 & \text{si } h > p \\ \neq 0 & \text{sinon} \end{cases}$$

Démonstration $r(h)$ est le coefficient de X_{t-h} dans $EL(X_t|X_{t-1}, \dots, X_{t-h})$ et

$$X_t = \mu + \underbrace{\varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p}}_{\in \mathcal{L}(1, X_t, \dots, X_{t-p}) \subset \mathcal{L}(1, X_t, \dots, X_{t-h})} + \varepsilon_t$$

$$\begin{aligned} \Rightarrow EL(X_t|X_{t-1}, \dots, X_{t-h}) &= \mu + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + EL(\varepsilon_t|X_{t-1}, \dots, X_{t-h}) \\ &= \mu + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + 0 \end{aligned}$$

Si $h > p$, le coefficient de X_{t-h} est 0.

Si $h = p$, le coefficient de X_{t-p} est $\varphi_p \neq 0$.

Proposition 2.1.7 Si $(X_t)_{t \in \mathbb{Z}} \rightsquigarrow AR(p)$, alors :

$$\gamma(h) = \begin{cases} 0 & \text{si } |h| > p \\ \neq 0 & \text{si } |h| = p \end{cases}$$

$$\rho(h) = \begin{cases} 0 & \text{si } |h| > p \\ \neq 0 & \text{si } |h| = p \end{cases}$$

Démonstration $\rho_i(h) = \frac{\gamma_i(h)}{\gamma_i(0)}$ où :

$$\gamma_i(h) = \int_{-\pi}^{\pi} \frac{1}{f_X(\omega)} e^{i\omega h} d\omega$$

$\phi(L)X_t = \varepsilon_t$ (en remplaçant éventuellement X_t par $X_t - m$) :

$$f_X(\omega) |\phi(e^{i\omega})|^2 = f_\varepsilon(\omega) = \frac{\sigma_\varepsilon^2}{2\pi}$$

$$\Rightarrow f_X(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{1}{|\phi(e^{i\omega})|^2}$$

Et par conséquent :

$$\frac{1}{f_X(\omega)} = \frac{2\pi}{\sigma_\varepsilon^2} |\phi(e^{i\omega})|^2$$

$$\phi(z) = 1 - \varphi_1 z - \dots - \varphi_p z^p = \sum_{k=0}^p \psi_k z^k$$

avec $\psi_0 = 1$ et $\psi_k = -\varphi_k$, $k > 0$,

$$\begin{aligned} \frac{1}{f_X(\omega)} &= \frac{2\pi}{\sigma_\varepsilon^2} \left(\sum_{k=0}^p \psi_k e^{i\omega k} \right) \left(\sum_{k=0}^p \psi_k e^{-i\omega k} \right) \\ &= \frac{2\pi}{\sigma_\varepsilon^2} \sum_{0 \leq k, l \leq p} \psi_k \psi_l e^{i\omega(k-l)} \end{aligned}$$

Ainsi

$$\gamma_i(h) = \frac{2\pi}{\sigma_\varepsilon^2} \sum_{0 \leq k, l \leq p} \psi_k \psi_l \underbrace{\int_{-\pi}^{\pi} e^{i\omega(k-l+h)} d\omega}_{=0 \text{ sauf si } k-l+h=0}$$

Or $k - l \in \llbracket -p; p \rrbracket$ donc si $h > p$, $\gamma_i(h) = 0$. En revanche si $h = p$:

$$\begin{aligned} \int_{-\pi}^{\pi} e^{i\omega(k-l+h)} d\omega &\Leftrightarrow p = l - k \\ &\Leftrightarrow \begin{cases} l = p \\ k = 0 \end{cases} \end{aligned}$$

Donc

$$\gamma_i(p) = \frac{4\pi^2}{\sigma_\varepsilon^2} \psi_0 \psi_p = -\frac{4\pi^2}{\sigma_\varepsilon^2} \varphi_p \neq 0$$

2.2 Processus moyenne mobile d'ordre q ($MA(q)$)

2.2.1 Définition et représentation canonique

Définition 2.2.1 $(X_t)_{t \in \mathbb{Z}} \rightsquigarrow MA(q)$ s'il existe $\varepsilon_t \rightsquigarrow \mathcal{BB}(0, \sigma^2)$ et $\theta_1, \dots, \theta_q$ tels que

$$X_t = m + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

Proposition 2.2.1 $\mathbb{E}X_t = m$

Remarque 2.2.1 (1) $(X_t)_{t \in \mathbb{Z}}$ est nécessairement stationnaire.

(2) On note $X_t = m + \theta(L)\varepsilon_t$ où $\theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q$.

(3) Comme $X_t - m = \theta(L)\varepsilon_t$, on peut centrer X_t , $\mathbb{E}X_t = 0$.

Ecriture $AR(\infty)$ quand les racines de θ sont de module > 1

Sous ces hypothèses $\theta(L)$ est inversible et $\theta(L)^{-1} = \sum_{k=0}^{+\infty} a_k L^k$ avec $a_0 = 1$ et $\sum |a_k| < +\infty$.

Il s'en suit que

$$X_t - m = \theta(L)\varepsilon_t \iff \theta(L)^{-1}(X_t - m) = \varepsilon_t \iff \theta(L)^{-1}X_t - \frac{m}{\theta(1)} = \varepsilon_t$$

Soit encore

$$\sum_{k=0}^{+\infty} a_k X_{t-k} - \mu = \varepsilon_t \text{ où } \mu = \frac{m}{\theta(1)}$$

D'où la **représentation canonique $AR(\infty)$**

$$\boxed{X_t = \sum_{k=1}^{+\infty} a_k X_{t-k} + \frac{m}{\theta(1)} + \varepsilon_t}$$

Proposition 2.2.2 *Sous les hypothèses précédentes*

(i) $\bar{\mathcal{L}}(X_t) = \bar{\mathcal{L}}(\varepsilon_t)$

(ii) ε_t est l'innovation de X_t

Démonstration

- (i) Comme $X_t = m + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \Rightarrow X_t \in \mathcal{L}(1, \varepsilon_t, \dots, \varepsilon_{t-q}) \subset \overline{\mathcal{L}}(\varepsilon_t)$
 Ceci nous amène à (cf. cas AR page 23 pour plus de détails) :

$$\left. \begin{array}{l} \overline{\mathcal{L}}(X_t) \subset \overline{\mathcal{L}}(\varepsilon_t) \\ \varepsilon_t = \sum_{k=0}^{+\infty} a_k X_{t-k} - \mu \Rightarrow \overline{\mathcal{L}}(\varepsilon_t) \subset \overline{\mathcal{L}}(X_t) \end{array} \right\} \Rightarrow \overline{\mathcal{L}}(X_t) = \overline{\mathcal{L}}(\varepsilon_t)$$

- (ii)

$$\begin{aligned} X_t^* &= EL(X_t | X_{t-1}) \\ &= EL(m + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} | X_{t-1}) \\ &= EL(m + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} | \varepsilon_{t-1}) \\ &= m + 0 - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \\ &= X_t - \varepsilon_t \end{aligned}$$

Donc $X_t - X_t^* = \varepsilon_t$, X_t^* est bien l'innovation de X_t .

Cas où des racines de θ sont de module < 1

On suppose qu'on s'est ramené à $X_t = \theta(L)\varepsilon_t$ par centrage :

$$X_t = \left[\prod_{i/|\lambda_i| < 1} (1 - \lambda_i L) \right] \left[\prod_{i/|\lambda_i| > 1} (1 - \lambda_i L) \right] \varepsilon_t$$

Comme précédemment, on définit :

$$\theta^*(L) = \left[\prod_{i/|\lambda_i| < 1} (1 - \lambda_i L) \right] \left[\prod_{i/|\lambda_i| > 1} \left(1 - \frac{1}{\lambda_i} L \right) \right]$$

On définit aussi (η_t) par $X_t = \theta^*(L)\eta_t$, d'où

$$\eta_t = \theta^*(L)^{-1} X_t$$

On montre que $f_\eta(\omega) = cte \Rightarrow (\eta_t) \rightsquigarrow \mathcal{BB}$.

On a donc

$$\left\{ \begin{array}{l} X_t = \theta^*(L)\eta_t \\ \text{toutes les racines de } \theta^* \text{ sont de module } > 1 \\ \eta_t \rightsquigarrow \mathcal{BB} \end{array} \right.$$

C'est la représentation canonique de (X_t) et (η_t) est le processus des innovations.

Cas où certaines racines de θ sont de module égal à 1

On montre que (X_t) est stationnaire. Par exemple

$$X_t = (1 - L)\varepsilon_t$$

On ne peut plus écrire $\theta(L)$ inversible avec $\theta(L)^{-1} = \sum_0^\infty a_k L^k$.

(ε_t) reste le processus des innovations de (X_t) mais la démonstration est difficile.

2.2.2 Propriétés des processus $MA(q)$

On suppose que la représentation étudiée est la représentation canonique

$$\begin{cases} X_t = m + \theta(L)\varepsilon_t \\ \text{toutes les racines de } \theta \text{ sont de module } > 1 \\ \theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q \\ \varepsilon_t \rightsquigarrow \mathcal{BB} \end{cases}$$

Proposition 2.2.3 (Auto-covariance) *Sous les hypothèses précédentes*

$$\gamma(h) = \begin{cases} 0 & \text{si } |h| > q \\ -\theta_q \sigma_\varepsilon^2 \neq 0 & \text{si } |h| = q \\ \sigma_\varepsilon^2 (-\theta_h + \sum_{i=h+1}^q \theta_i \theta_{i-h}) & \text{si } 1 \leq |h| < q \\ \sigma_\varepsilon^2 (1 + \sum_{i=1}^q \theta_i^2) & \text{si } h = 0 \end{cases}$$

On en déduit $\rho(h) = 0$ si $|h| > q$ et $\rho(q) \neq 0$.

Démonstration $X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$ après centrage.

– Si $h = 0$

$$\gamma(0) = \text{Var} X_t = \sigma_\varepsilon^2 (1 + \theta_1^2 + \dots + \theta_q^2) \neq 0$$

Car $\text{Cov}(\varepsilon_{t-j}, \varepsilon_{t-k}) = 0$ si $j \neq k$.

– Si $h > q$

$$\begin{aligned} \gamma(h) &= \text{Cov}(X_t, X_{t-h}) \\ &= \text{Cov}\left[\underbrace{(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q})}_{\varepsilon_{t-j}, j \in [0, q]}, \underbrace{(\varepsilon_{t-h} - \theta_1 \varepsilon_{t-h-1} - \dots - \theta_q \varepsilon_{t-h-q})}_{\varepsilon_{t-k}, k \in [h, q+h]}\right] \\ &\Rightarrow t - j \neq t - k \\ &\Rightarrow \gamma(h) = 0 \end{aligned}$$

– Si $|h| = q$

$$\begin{aligned} \gamma(q) &= \text{Cov}[(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}), (\varepsilon_{t-q} - \theta_1 \varepsilon_{t-q-1} - \dots - \theta_q \varepsilon_{t-2q})] \\ &= -\theta_q \sigma_\varepsilon^2 \end{aligned}$$

– Si $1 \leq |h| < q$

$$\begin{aligned} \gamma(h) &= \text{Cov}[(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}), (\varepsilon_{t-h} - \theta_1 \varepsilon_{t-h-1} - \dots - \theta_q \varepsilon_{t-h-q})] \\ &= -\sum_{i=1}^q \theta_i \text{Cov}[\varepsilon_{t-i}, \varepsilon_{t-h} - \sum_{k=1}^q \theta_k \varepsilon_{t-h-k}] \quad \text{car } \text{Cov}(\varepsilon_t, \varepsilon_{t-i}) = 0 \quad \forall i > 0 \\ &= -\theta_h \sigma_\varepsilon^2 + \sum_{i=1}^q \sum_{k=1}^q \theta_i \theta_k \underbrace{\text{Cov}(\varepsilon_{t-i}, \varepsilon_{t-h-k})}_{=0 \text{ si } i \neq h+k} \\ &= -\theta_h \sigma_\varepsilon^2 + \left(\sum_{i=h+1}^q \theta_i \theta_{i-h} \right) \sigma_\varepsilon^2 \end{aligned}$$

Remarque 2.2.2 On n'a pas de résultat particulier pour les auto-corrélations partielles.

Proposition 2.2.4 $\rho^i(h)$ décroît exponentiellement avec h .

Démonstration $\rho^i(h) = \frac{\gamma^i(h)}{\gamma^i(0)}$ avec, $\gamma^i(h) = \int_{-\pi}^{\pi} \frac{1}{f_X(\omega)} e^{i\omega h} d\omega$ et

$$\begin{aligned} X_t = \theta(L)\varepsilon_t &\Rightarrow f_X(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} |\theta(e^{i\omega})|^2 \\ &\Rightarrow \frac{1}{f_X(\omega)} = \frac{2\pi}{\sigma_\varepsilon^2 |\theta(e^{i\omega})|^2} \end{aligned}$$

Soit $(Y_t)_{t \in \mathbb{Z}}$ un processus tel que $\theta(L)Y_t = \eta_t$ et $Y_t \rightsquigarrow AR(q)$:

$$\frac{\sigma_\eta^2}{2\pi} = f_Y(\omega) |\theta(e^{i\omega})|^2$$

Donc

$$f_Y(\omega) = \frac{\sigma_\eta^2}{2\pi} \frac{1}{|\theta(e^{i\omega})|^2}$$

On a ainsi :

$$f_Y(\omega) = \frac{1}{f_X(\omega)} \iff \frac{2\pi}{\sigma_\varepsilon^2} = \frac{\sigma_\eta^2}{2\pi} \iff \sigma_\eta^2 = \frac{4\pi^2}{\sigma_\varepsilon^2}$$

Tableau récapitulatif des différentes situations Les auto-corrélations inverses d'un processus $MA(q)$ ont les mêmes propriétés que les auto-corrélations d'un $AR(q)$:

	$AR(p)$	$MA(q)$
$\rho(h)$	décroît exponentiellement vers 0 avec h	0 si $ h > q$ et non nul si $h = q$
$r(h)$	0 si $h > p$ et non nul si $h = p$	-
$\rho^i(h)$	0 si $h > p$ et non nul si $h = p$	décroît exponentiellement vers 0 avec h

2.3 Processus ARMA(p, q)

2.3.1 Définition et représentation canonique minimale

Définition 2.3.1 Un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ admet une **représentation ARMA(p, q) canonique minimale** s'il vérifie une équation :

$$\boxed{\phi(L)X_t = \mu + \theta(L)\varepsilon_t}$$

où

- (i) $\varepsilon_t \rightsquigarrow \mathcal{BB}(0, \sigma^2)$
- (ii) $\phi(L) = 1 - \varphi_1 L - \dots - \varphi_p L^p$, avec $\varphi_p \neq 0$
- (iii) $\theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q$, avec $\theta_q \neq 0$
- (iv) ϕ et θ ont toutes leurs racines de module strictement supérieur à 1 (représentation canonique).
- (v) ϕ et θ n'ont pas de racines communes (représentation minimale).

Remarque 2.3.1 (1) Il existe des solutions non stationnaires : soit (X_t) un processus stationnaire et (Y_t) déterministe tel que $\phi(L)Y = 0$. On définit $Z_t = X_t + Y_t$ qui vérifie l'équation.

(2) Retour sur la représentation canonique :

- Si (X_t) est stationnaire, alors les racines de ϕ sont de module distinct de 1. On pourrait considérer le cas où θ a des racines de module 1 (c'est compatible avec la stationnarité).
 → Si on suppose que ϕ et θ ont des racines de module distinct de 1, on peut toujours se ramener à la représentation

$$\phi^*(L)X_t = \mu^* + \theta^*(L)\eta_t$$

où ϕ^* et θ^* ont des racines de module > 1 .

- Si ϕ et θ ont des racines de module strictement supérieur à 1 mais admettent une racine commune, alors

$$\phi(L) = (1 - \lambda L)\varphi_0(L) \text{ et } \theta(L) = (1 - \lambda L)\theta_0(L)$$

D'où

$$\varphi_0(L)X_t = \frac{\mu}{1 - \lambda} + \theta_0(L)\varepsilon_t \Rightarrow X_t \sim \text{ARMA}(p - 1, q - 1)$$

Proposition 2.3.1 (i) $\mathbb{E}X_t = \frac{\mu}{\phi(1)} = m$

(ii) $\phi(L)(X_t - m) = \theta(L)\varepsilon_t$

Remarque 2.3.2 Par centrage on peut donc se ramener au cas où $\mu = 0$.

Démonstration

(i) On a

$$\mathbb{E}(X_t - \varphi_1 X_{t-1} - \cdots - \varphi_p X_{t-p}) = \mathbb{E}(\mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q})$$

Et comme (X_t) est supposé stationnaire

$$m(1 - \varphi_1 - \cdots - \varphi_p) = \mu + 0 \Rightarrow m = \frac{\mu}{\phi(1)}$$

(ii)

$$\begin{aligned} \phi(L)X_t &= \phi(1)m + \theta(L)\varepsilon_t \\ &= \phi(L)m + \theta(L)\varepsilon_t \end{aligned}$$

Donc $\phi(L)(X_t - m) = \theta(L)\varepsilon_t$.

Proposition 2.3.2 Sous les hypothèses précédentes,

- (i) (X_t) admet une représentation $AR(\infty)$, $\sum_{k=0}^{+\infty} a_k X_{t-k} = \mu + \varepsilon_t$ où $a_0 = 1$ et $\sum_k |a_k| < +\infty$
 (ii) (X_t) admet une représentation $MA(\infty)$, $X_t = m + \sum_{k=0}^{+\infty} b_k \varepsilon_{t-k}$ où $b_0 = 1$ et $\sum_k |b_k| < +\infty$
 (iii) $\bar{\mathcal{L}}(X_t) = \bar{\mathcal{L}}(\varepsilon_t)$
 (iv) ε_t est l'innovation de X_t

Démonstration

(i) On sait que $\phi(L)(X_t - m) = \theta(L)\varepsilon_t$, cela nous permet d'écrire

$$\begin{aligned} \underbrace{\theta(L)^{-1}\phi(L)}_{A(L)}(X_t - m) &= \varepsilon_t \\ \Rightarrow A(L)X_t - A(1)m &= \varepsilon_t \end{aligned}$$

Et ce avec $A(1)m = \frac{\phi(1)}{\theta(1)} = \frac{\mu}{\theta(1)}$.

(ii) De la même façon $\phi(L)X_t = \mu + \theta(L)\varepsilon_t$ amène

$$X_t = \frac{\mu}{\phi(1)} + \underbrace{\phi(L)^{-1}\theta(L)}_{B(L)=\sum b_k L^k} \varepsilon_t$$

(iii) Etant donné que (X_t) est de la forme $AR(\infty)$, on a :

$$\forall t, \varepsilon_t \in \bar{\mathcal{L}}(X_t) \Rightarrow \mathcal{L}(\varepsilon_t) \subset \bar{\mathcal{L}}(X_t) \Rightarrow \bar{\mathcal{L}}(\varepsilon_t) \subset \bar{\mathcal{L}}(X_t)$$

Par un raisonnement identique et tenant compte du fait que (X_t) est également de la forme $MA(\infty)$ on obtient :

$$\bar{\mathcal{L}}(X_t) \subset \bar{\mathcal{L}}(\varepsilon_t)$$

Les deux résultats nous permettent alors de dire que :

$$\bar{\mathcal{L}}(X_t) = \bar{\mathcal{L}}(\varepsilon_t)$$

(iv) Calculons l'innovation de X_t :

$$\begin{aligned} X_t - X_t^* &= X_t - EL(X_t | X_{t-1}) \\ &= X_t - EL\left(-\sum_{k=0}^{+\infty} a_k X_{t-k} + \mu + \varepsilon_t | X_{t-1}\right) \\ &= X_t + \sum_{k=0}^{+\infty} a_k X_{t-k} - \mu - \underbrace{EL(\varepsilon_t | \varepsilon_{t-1})}_{=0} \\ &= \varepsilon_t \end{aligned}$$

Remarque 2.3.3 – $AR(p) \equiv ARMA(p, 0)$

- $MA(q) \equiv ARMA(0, q)$
- $ARMA(p, q) \equiv AR(\infty) \# AR(P)$ si P grand
- $\equiv MA(\infty) \# MA(Q)$ si Q grand

Souvent l'un des paramètres (p ou q) est petit alors que l'autre est grand. Avec l'approximation précédente on a alors moins de paramètres à estimer.

- En vertu du théorème de WOLD, $X_t = m + B(L)\varepsilon_t$, où (ε_t) est le processus des innovations, si de plus $X_t \rightsquigarrow ARMA(p, q)$ alors $B(L) = \frac{\theta(L)}{\phi(L)}$.

2.3.2 Propriétés des processus ARMA(p, q)

On considère un processus $ARMA(p, q)$ tel que :

- $\phi(L)X_t = \theta(L)\varepsilon_t$, éventuellement après centrage,
- $\phi(L) = 1 - \varphi_1 L - \dots - \varphi_p L^p$,
- $\theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q$.

C'est la représentation canonique minimale.

Proposition 2.3.3 (Auto-covariance et auto-corrélation) (i) Pour $h > q$, les $\gamma(h)$ et les $\rho(h)$ vérifient les équations de récurrence d'ordre p :

$$\begin{aligned} \gamma(h) - \varphi_1 \gamma(h-1) - \dots - \varphi_p \gamma(h-p) &= 0 \\ \rho(h) - \varphi_1 \rho(h-1) - \dots - \varphi_p \rho(h-p) &= 0 \end{aligned}$$

(ii) Elles décroissent donc vers 0 exponentiellement avec h , pour $h > q$.

Démonstration

(i) $X_t = \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$ et par conséquent :

$$\begin{aligned} \gamma(h) &= \mathbb{E}(X_t X_{t-h}) \\ &= \varphi_1 \mathbb{E}(X_{t-1} X_{t-h}) + \dots + \varphi_p \mathbb{E}(X_{t-p} X_{t-h}) - \theta_1 \underbrace{\mathbb{E}(\varepsilon_{t-1} X_{t-h})}_{=0} - \dots - \theta_q \underbrace{\mathbb{E}(\varepsilon_{t-q} X_{t-h})}_{=0} \\ &= \varphi_1 \gamma(h-1) + \dots + \varphi_p \gamma(h-p) \end{aligned}$$

Il s'en suit que :

$$\rho(h) = \varphi_1 \rho(h-1) + \dots + \varphi_p \rho(h-p)$$

(ii) Les $\gamma(h)$ et les $\rho(h)$ vérifient une équation de récurrence dont le polynôme caractéristique est $z^{p+1} \phi\left(\frac{1}{z}\right)$.

Les conditions initiales sont $\gamma(q), \gamma(q-1), \dots, \gamma(q-p+1)$ et $\rho(q), \rho(q-1), \dots, \rho(q-p+1)$.

Equations de Yule-Walker L'équation précédente pour $k = q+1, \dots, q+p$ donne :

$$\begin{pmatrix} \rho(q) & \dots & \rho(q+p-1) \\ \vdots & \ddots & \vdots \\ \rho(q+p-1) & \dots & \rho(q) \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_p \end{pmatrix} = \begin{pmatrix} \rho(p+1) \\ \vdots \\ \rho(p+q) \end{pmatrix}$$

Quand ρ est connu ou estimé, on peut alors calculer les ϕ_j .

Ou inversement, quand les φ_j sont connus, on calcule $\rho(q+1), \dots, \rho(q+p)$ qui seront les conditions initiales pour le calcul de $\rho(h)$ tel $h > q$.

2.4 Processus ARIMA(p, d, q)

Ces processus sont non stationnaires dès que $d \leq 1$. Les séries économiques sont souvent non stationnaires, tel le PIB.

Exemple 2.4.1 On considère un processus $(X_t)_{t \in \mathbb{Z}}$ correspondant à une marche aléatoire c'est-à-dire

$$\forall t \geq 0, X_t = X_{t-1} + \varepsilon_t$$

tel que $\varepsilon_t \sim \mathcal{BB}(0, \sigma^2)$ et $\forall t > 0, \text{Cov}(\varepsilon_t, X_0) = 0$.

Alors

$$\begin{aligned} X_t &= X_0 + \sum_{k=1}^t \varepsilon_k = X_0 + \sum_{j=0}^{t-1} \varepsilon_{t-j} \\ &= X_{-1} + \sum_{k=0}^t \varepsilon_k = X_{-1} + \sum_{j=0}^t \varepsilon_{t-j} \end{aligned}$$

On ne peut pas itérer le procédé car $\sum_{j=0}^{+\infty} \varepsilon_{t-j}$ **n'est pas défini**. On ne peut pas supposer le processus démarre à $-\infty$. La condition initiale est $\text{Cov}(X_0, \varepsilon_k) = 0$ pour $k > 0$.

On peut alors penser à considérer :

$$(1-L)X_t = X_t - X_{t-1} = \Delta X_t = \varepsilon_t$$

Idée générale : $X_t \rightsquigarrow ARIMA(p, d, q)$ si et seulement si $(1 - L)^d X_t$ est stationnaire alors que $(1 - L)^{d-1} X_t$ ne l'est pas (dans le cas de la marche aléatoire, $d = 1$).

Définition 2.4.1 (Représentation canonique minimale) $(X_t)_{t \geq -pd}$ est un processus ARIMA(p, d, q) en **représentation canonique minimale** s'il vérifie une équation du type :

$$\forall t \geq 0, (1 - L)^d \phi(L) X_t = \mu + \theta(L) \varepsilon_t$$

Et ceci avec :

(i) $\varepsilon_t \rightsquigarrow \mathcal{BB}(0, \sigma^2)$

(ii) $\phi(L) = 1 - \varphi_1 L - \dots - \varphi_p L^p$ où $\varphi_p \neq 0$
 $\theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q$ où $\theta_q \neq 0$

(iii) ϕ et θ ont leurs racines de module > 1 et n'ont pas de racines communes

(iv) conditions initiales

$$Z = (X_{-1}, \dots, X_{-p-d}, \varepsilon_{-1}, \dots, \varepsilon_{-q})$$

telles que $\text{Cov}(\varepsilon_t, Z) = 0$

Exemple 2.4.2 Soit le processus défini par $(1 - L)X_t = \varepsilon_t$. On a donc $d^o \phi = d^o \theta = 0$.

Si $Z = X_{-1}$, $\text{Cov}(Z_t, \varepsilon_t) = 0 \forall t \leq 0$.

Remarque 2.4.1 Comme $(1 - L)^d \phi(L) X_t = \phi(L)(1 - L)^d X_t$, on pose $Y_t = (1 - L)^d X_t$.

(Y_t) suit alors le processus :

$$\phi(L) Y_t = \mu + \theta(L) \varepsilon_t$$

Proposition 2.4.1 Sous les hypothèses précédentes, $(1 - L)^d X_t = Y_t$ est alors asymptotiquement équivalent à un processus ARMA(p, q).

Démonstration Ce qui signifie qu'il existe un processus stationnaire $(Z_t)_{t \in \mathbb{Z}}$ tel que :

$$\phi(L) Z_t = \mu + \theta(L) \varepsilon_t$$

$$\lim_{t \rightarrow +\infty} \|Y_t - Z_t\|_2 = 0$$

Notations

– Si $d = 0$, $X_t \rightsquigarrow ARMA(p, q)$ qui est un processus stationnaire.

On note $X_t \rightsquigarrow I(0)$.

– Si $d = 1$, (X_t) est un **processus intégré** d'ordre 1.

On note $X_t \rightsquigarrow I(1)$.

– Si $d = 2$, (X_t) n'est pas stationnaire, $Y_t = (1 - L)X_t$ non plus, $Z_t = (1 - L)Y_t = (1 - L)^2 X_t$ est asymptotiquement équivalent à un processus stationnaire.

On note $X_t \rightsquigarrow I(2)$.

Définition 2.4.2 Si $(1 - L)^d X_t$ est asymptotiquement équivalent à un processus stationnaire et si $(1 - L)^{d-1} X_t$ ne l'est pas alors on dit que (X_t) est intégré d'ordre d et on note $X_t \rightsquigarrow I(d)$.

2.4.1 Approximation auto-régressive d'un ARIMA(p, d, q)

Proposition 2.4.2 Avec les notations précédentes,

$$\left. \begin{array}{l} \exists A_t(L), A_t(L) = \sum_{j=0}^t a_j^t L^j \text{ et } a_t^0 = 1 \\ \exists \mu_0 \\ \exists h(t) \in \mathbb{R}^{p+d+q} \text{ et } \lim_{t \rightarrow +\infty} h(t) = 0 \end{array} \right\} \text{ tels que } A_t(L)X_t = \mu_0 + \varepsilon_t + h(t)'Z$$

$$\iff X_t = - \sum_{j=1}^t a_j^t X_{t-j} + \varepsilon_t + h(t)'Z$$

Démonstration On pose $\psi(L) = (1-L)^d \phi(L)$, avec cette notation :

$$\psi(L)X_t = \mu + \theta(L)\varepsilon_t \quad d^o \psi = p + d, \quad d^o \theta = q$$

On effectue la division selon les puissances croissantes à l'ordre t de 1 par $\theta(z)$:

$$1 = \theta(z)Q_t(z) + z^{t+1}R_t(z) \text{ où } d^o Q_t = t, \quad d^o R_t = q - 1$$

Ce qui implique :

$$1 = \theta(L)Q_t(L) + L^{t+1}R_t(L)$$

Or

$$\begin{aligned} \underbrace{\psi(L)Q_t(L)}_{d^o = p+d+t} X_t &= Q_t(1)\mu + Q_t(L)\theta(L)\varepsilon_t \\ &= Q_t(1)\mu + (1 - L^{t+1}R_t(L))\varepsilon_t \end{aligned}$$

Ainsi

$$\sum_{j=0}^{p+d+t} a_j^{(t)} X_{t-j} = \mu_0 + \varepsilon_t - R_t(L)\varepsilon_{-1}$$

En décomposant la somme

$$\sum_{j=0}^t a_j^{(t)} X_{t-j} = \mu_0 + \varepsilon_t - \sum_{j=t+1}^{p+d+t} a_j^{(t)} X_{t-j} - \sum_{k=0}^{q-1} r_k^{(t)} \varepsilon_{-1-k}$$

On effectue le changement d'indice $k = t - j$ dans $\sum_{j=t+1}^{p+d+t} a_j^{(t)} X_{t-j}$:

$$\sum_{j=0}^t a_j^{(t)} X_{t-j} = \mu_0 + \varepsilon_t - \underbrace{\sum_{k=-p-d}^{-1} a_{t-k}^{(t)} X_k - \sum_{k=0}^{q-1} r_k^{(t)} \varepsilon_{-1-k}}_{h(t)'Z}$$

2.4.2 Approximation moyenne mobile d'un ARIMA(p, d, q)

Proposition 2.4.3 Sous les mêmes hypothèses,

$$\left. \begin{array}{l} \exists B_t(L), B_t(L) = \sum_{j=0}^t b_j^{(t)} L^j \text{ et } b_t^{(0)} = 1 \\ \exists \mu_1 \\ \exists \tilde{h}(t) \in \mathbb{R}^{p+d+q} \text{ et } \lim_{t \rightarrow +\infty} \tilde{h}(t) = 0 \end{array} \right\} \text{ tels que } X_t = \mu_1 + B_t(L)\varepsilon_t + \tilde{h}(t)'Z$$

Corollaire $\forall t, \varepsilon_t \in \mathcal{L}(X_0, \dots, X_t, 1, Z)$

$$X_t \in \mathcal{L}(\varepsilon_0, \dots, \varepsilon_t, 1, Z)$$

$\Rightarrow \varepsilon_t = X_t - EL(X_t|X_0, \dots, X_t, 1, Z)$ est le processus des innovations

Proposition 2.4.4 (Calcul de $\mathbb{E}X_t$) Si l'on note $m_t = \mathbb{E}X_t$ alors m_t vérifie $\psi(L)m_t = \mu$. On obtient ainsi :

\rightarrow une équation de récurrence dont le polynôme caractéristique est $z^{p+d+1}\psi\left(\frac{1}{z}\right)$,

\rightarrow une forme générale de la solution (pour $\mu = 0$ et $\mu \neq 0$).

Exemple 2.4.3 (i) Marche aléatoire sans dérive : $(1 - L)X_t = \varepsilon_t$

$$(1 - L)m_t = 0 \Rightarrow m_t = cte$$

(ii) Marche aléatoire avec dérive : $(1 - L)X_t = \mu + \varepsilon_t$ alors $m_t - m_{t-1} = \mu$

$$(1 - L)m_t = \mu \Rightarrow m_t = m_0 + \mu t$$

(iii) $(1 - L)(1 - \varphi L)X_t = \varepsilon_t$ et alors $m_t = \alpha + \beta\varphi t$

(iv) On a vu que : $X_t = X_0 + \sum_{k=1}^t \varepsilon_k$ si $\mu = 0 \Rightarrow \mathbb{E}X_t = \mathbb{E}X_0$

$$X_t = X_0 + \mu t + \sum_{k=1}^t \varepsilon_k \text{ si } \mu \neq 0 \Rightarrow \mathbb{E}X_t = \mathbb{E}X_0 + \mu t$$

Chapitre 3

Identification et estimation d'un modèle *ARMA* ou *ARIMA*

Introduction

On dispose d'observations x_1, \dots, x_T de X_1, \dots, X_T . Comment modéliser par un *ARMA* ou un *ARIMA* ?

On a 2 types de choix :

- $(X_t)_{t \in \mathbb{Z}}$ est un processus stationnaire auquel cas il faut estimer un *ARMA*(p, q),
- ou $(X_t) \rightsquigarrow I(d)$ est donc non stationnaire mais $(1 - L)^d X_t$ est stationnaire, dans ce cas il faut estimer un *ARIMA*(p, d, q).

La démarche pour l'identification est la suivante :

- Choix de d ,
- Choix de (p, q) ,
- Estimer $\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q$ (ce qui peut se faire par le maximum de vraisemblance sous l'hypothèse que les $\varepsilon_t \rightsquigarrow \mathcal{N}(0, \sigma^2)$ sont i.i.d.) et σ^2 ,
- Phase de vérification :
 - $\rightarrow \varphi_p \neq 0$?
 - $\rightarrow \theta_q \neq 0$?
 - $\rightarrow \varepsilon_t \rightsquigarrow \mathcal{BB}(0, \sigma^2)$?

Les deux premières étapes constituent la phase d'identification du processus et pour vérifier la non nullité des coefficients lors de la phase de vérification il faudra définir les tests auxquels on aura recours.

En ce qui concerne le choix de d on peut procéder de façon empirique (en observant les auto-corrélogrammes) ou en effectuant des tests de racine unité :

$$\begin{cases} H_0 : d = 1 & H_1 : d = 0 & \rightarrow \text{DF (ADF), PP, SP} \\ H_0 : d = 0 & H_1 : d = 1 & \rightarrow \text{KPSS} \end{cases}$$

3.1 Première phase de l'identification : choix de d

3.1.1 Approche empirique : l'auto-corrélogramme

On a vu que :

- si $X_t \rightsquigarrow \text{ARMA}(p, q)$, les $\rho(h)$ décroissent exponentiellement vers 0 avec h (pour $h > q$),

- si (X_t) est stationnaire $\hat{\rho}_T(h) \xrightarrow{P} \rho(h)$,
- sous des hypothèses suffisantes ($\mathbb{E}(\varepsilon_t^4) = cte$) :

$$\sqrt{T} \begin{pmatrix} \hat{\rho}_T(1) - \rho(1) \\ \hat{\rho}_T(h) - \rho(h) \end{pmatrix} \xrightarrow{L} \mathcal{N}(0, *), \quad \forall h$$

Remarque 3.1.1 Si (X_t) admet une racine unité, la proposition $\rho(h)$ décroît exponentiellement vers 0 avec h n'est plus vraie : c'est la **persistance des chocs**.

Exemple 3.1.1 On considère un processus (X_t) tel que $X_t - X_{t-1} = \varepsilon_t$ où $\varepsilon_t \sim \mathcal{BB}(0, \sigma^2)$ et $\text{Cov}(\varepsilon_t, X_0) = 0$ si $t \geq 0$.

$$X_t = X_0 + \sum_{k=1}^t \varepsilon_k$$

$$X_{t+h} = X_0 + \sum_{k=1}^{t+h} \varepsilon_k$$

$$\begin{aligned} \rho(h) &= \frac{\text{Cov}(X_t, X_{t+h})}{\sqrt{\mathbb{V}(X_t)\mathbb{V}(X_{t+h})}} \\ &= \frac{\text{Cov}\left(X_0 + \sum_{k=1}^t \varepsilon_k, X_0 + \sum_{j=1}^{t+h} \varepsilon_j\right)}{\sqrt{(\mathbb{V}X_0 + t\sigma^2)^{1/2}(\mathbb{V}X_0 + (t+h)\sigma^2)}} \\ &= \frac{\mathbb{V}X_0 + t\sigma^2}{\sqrt{(\mathbb{V}X_0 + t\sigma^2)(\mathbb{V}X_0 + (t+h)\sigma^2)}} \end{aligned}$$

Pour t grand et $h \ll t$,

$$\rho(h) \# \frac{t\sigma^2}{\sigma^2 \sqrt{t(t+h)}} = \frac{1}{\sqrt{1 + \frac{h}{t}}} \# 1 - \frac{h}{2t}$$

La décroissance est lente et linéaire en h . D'où une règle pratique :

si les $\hat{\rho}_T(h)$ restent proches de 1 ou décroissent linéairement avec h alors le processus est sans doute non stationnaire.

Remarque 3.1.2 (1) Si l'auto-corrélogramme fait penser que (X_t) est non stationnaire, alors on étudie l'auto-corrélogramme de $Y_t = (1 - L)X_t$.

(2) On étudie l'auto-corrélogramme inverse pour étudier une sur-différentiation éventuelle.

(3) Rappel : si $\phi(L)X_t = \theta(L)\varepsilon_t$ et $\theta(L)Z_t = \phi(L)\eta_t$, alors

$$\rho_X^i(h) = \rho_Z(h)$$

Si (X_t) est stationnaire, alors $\phi(1) \neq 0$.

$$\Rightarrow \phi(L)(1 - L)X_t = \theta(L)(1 - L)\varepsilon_t$$

$$\rho_X^i(h) = \rho_W(h)$$

avec (W_t) d'équation $\theta(L)(1 - L)W_t = \phi(L)\eta_t$

Si on a sur-différentiation les $\rho_X^i(h)$ ne décroissent pas vers 0 exponentiellement avec h .

3.1.2 Approche par les tests de racine unité

On présente ci-dessous les principaux tests de racine unité dans la littérature. Dans les trois premiers paragraphes (tests de DICKEY-FULLER, PHILLIPS-PERRON, SCHMIDT-PHILLIPS) ; l'hypothèse nulle est l'hypothèse de non-stationnarité dans la série étudiée ; dans le dernier paragraphe, (tests KPPS), l'hypothèse nulle est celle de stationnarité.

La présentation qui est donnée ici des tests de DICKEY-FULLER et de PHILLIPS-PERRON s'inspire largement de celle de J.D. HAMILTON, *Time Series Analysis*, Princeton University Press, 1994.

Il faut d'emblée signaler que les tests présentés ici sont peu puissants. Par ailleurs, les tests de DICKEY-FULLER sont présentés en détail à cause de la place qu'ils tiennent dans la littérature, mais leur mise en œuvre pratique s'avère souvent problématique : nécessité de procéder à des tests emboîtés d'une part, cadre mal adapté aux séries présentant une tendance d'autre part. Dans ce dernier cas notamment, on leur préfère le test de SCHMIDT-PHILLIPS.

Les tests de Dickey-Fuller

Dans tous les modèles présentés ci-dessous, (η_t) désigne un bruit blanc et ρ un réel tel que $|\rho| \leq 1$.

Le cadre général des tests DF et ADF Ces tests peuvent être regroupés en quatre cas :

Pour les tests DF

1. $y_t = \rho y_{t-1} + \eta$, avec $H_0 : \rho = 1$, marche aléatoire sans dérive ;
2. $y_t = \alpha + \rho y_{t-1} + \eta$, avec $H_0 : \alpha = 0, \rho = 1$, marche aléatoire sans dérive ;
3. $y_t = \alpha + \rho y_{t-1} + \eta$, avec $H_0 : \alpha \neq 0, \rho = 1$, marche aléatoire avec dérive ;
4. $y_t = \alpha + \beta t + \rho y_{t-1} + \eta$, avec $H_0 : \alpha = 0, \beta = 0, \rho = 1$, marche aléatoire sans dérive, ou $H_{01} : \beta = 0, \rho = 1$, marche aléatoire avec dérive.

Pour les tests ADF Soit $\Phi(L)$ polynôme de degré $p \geq 2$, dont les racines sont supposées de module supérieur à 1, et ayant au plus une racine égale à 1 :

$$\Phi(L) = \prod_{i=1}^p (1 - \lambda_i L)$$

avec éventuellement $\exists! i_0 / \lambda_{i_0} = 1$ et $\forall i \neq i_0, |\lambda_i| < 1$.

D'où la réécriture des cas :

1. $\Phi(L)y_t = \eta$, $H_0 : \Phi(1) = 0$;
2. $\Phi(L)y_t = \alpha + \eta$, $H_0 : \Phi(1) = 0, \alpha = 0$;
3. $\Phi(L)y_t = \alpha + \eta$, $H_0 : \Phi(1) = 0, \alpha \neq 0$;
4. $\Phi(L)y_t = \alpha + \beta t + \eta$, $H_0 : \Phi(1) = 0, \alpha = 0, \beta = 0$, ou $H_{01} : \Phi(1) = 0, \beta = 0$.

L'écriture des quatre modèles ci-dessus peut être transformée en utilisant la démarche suivante :

On décompose $\Phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$ sous la forme

$$\Phi(L) = \Phi(1) + (1 - L)\Phi^*(L) = \Phi(1) - (1 - L) \sum_{i=0}^{p-1} \alpha_i L^i$$

avec $\alpha_0 = -(\phi_1 + \dots + \phi_p) = \Phi(1) - 1$ et $\forall 1 \leq i \leq p-1, \alpha_i = \alpha_{i-1} + \phi_i = -(\phi_i + 1 \dots + \phi_p)$.

On obtient :

$$\begin{aligned} \Phi(L)y_t &= \Phi(1)y_t - \alpha_0 \Delta y_t - \sum_{i=1}^{p-1} \alpha_i \Delta y_{t-i} \\ &= \Phi(1)y_t - (\Phi(1) - 1)(y_t - y_{t-1}) - \sum_{i=1}^{p-1} \alpha_i \Delta y_{t-i} \\ &= y_t + (\Phi(1) - 1)y_{t-1} - \sum_{i=1}^p \alpha_i \Delta y_{t-i} \end{aligned}$$

En posant $\rho = 1 - \Phi(1)$, on obtient :

1. $y_t = \rho y_{t-1} + \sum_{i=1}^{p-1} \alpha_i \Delta y_{t-i} = \eta, H_0 : \rho = 1$
2. $y_t = \alpha + \rho y_{t-1} + \sum_{i=1}^{p-1} \alpha_i \Delta y_{t-i} = \eta, H_0 : \alpha = 0, \rho = 1$
3. $y_t = \alpha + \rho y_{t-1} + \sum_{i=1}^{p-1} \alpha_i \Delta y_{t-i} = \eta, H_0 : \alpha \neq 0, \rho = 1$
4. $y_t = \alpha + \beta t \rho y_{t-1} + \sum_{i=1}^{p-1} \alpha_i \Delta y_{t-i} = \eta, H_0 : \alpha = \beta = 0, \rho = 1$ ou $H_{01} : \beta = 0, \rho = 1$

De plus, comme

$$\Phi(1) = \prod_{\lambda_i \in \mathbb{R}} (1 - \lambda_i) \prod_{\lambda_i \in \mathbb{C} - \mathbb{R}} (1 - \lambda_i)(1 - \bar{\lambda}_i) > 0$$

on a, comme précédemment, $\rho \leq 1$.

Les tests DF apparaissent comme des cas particuliers des tests ADF, dans lesquels $p = 1$ et $\sum_{i=1}^{p-1} \alpha_i \Delta y_{t-i} = 0$.

Tous ces modèles sont estimés par les MCO. Pour simplifier, on les écrit souvent sous la forme :

- Cas 1 :

$$\Delta y_t = \phi y_{t-1} + \sum_{i=1}^{p-1} \alpha_i \Delta y_{t-i} + \eta, \phi = \rho - 1$$

- Cas 2 et 3 :

$$\Delta y_t = \alpha + \phi y_{t-1} + \sum_{i=1}^{p-1} \alpha_i \Delta y_{t-i} + \eta$$

- Cas 4 :

$$\Delta y_t = \alpha + \beta t + \phi y_{t-1} + \sum_{i=1}^{p-1} \alpha_i \Delta y_{t-i} + \eta$$

Les statistiques de tests et leurs lois Les résultats sont les suivants :

- les $\hat{\alpha}_i$ et les $t_{\hat{\alpha}_i}$ ont des lois limites standard, même sous l'hypothèse de non-stationnarité, ce qui permet de fixer p par des tests de Fisher, et donc de partir avec p grand ;
- les coefficients qui caractérisent la nature stochastique de la série, $\hat{\alpha}, \hat{\beta}, \hat{\phi} = \hat{\rho} - 1$, ont les mêmes lois dans le cadre DF et ADF. Ces lois sont non standard, mais elles sont tabulées. Il faut noter que les lois asymptotiques sont valables quelle que soit la loi des η , alors que les lois à distance finie sont valables seulement si les η sont gaussiens.

1. $H_0 : \rho = 1 \Leftrightarrow H_0 : \phi = 0$. On dispose des lois sous H_0 de :

- $T_{\hat{\phi}_T} = T_{\hat{\rho}-1} \rightarrow$ table B5 cas 1 ;
 - $t_{\hat{\phi}_T} = t_{\hat{\rho}-1} \rightarrow$ table B6 cas 1.
- N.B. : il s'agit d'un test unilatéral puisque $\rho \leq 1$. On rejette H_0 au seuil a si $T_{\hat{\phi}_T} < c_1^a$ ou $t_{\hat{\phi}_T} < c_2^a$.
2. $H_0 : \alpha = 0, \rho = 1 \Leftrightarrow H_0 : \alpha = 0, \phi = 0$. On dispose des lois sous H_0 de :
 - $T_{\hat{\phi}_T} = T_{\hat{\rho}-1} \rightarrow$ table B5 cas 2 ;
 - $t_{\hat{\phi}_T} = t_{\hat{\rho}-1} \rightarrow$ table B6 cas 2 ;
 - $\hat{\Phi}_1$, statistique de FISHER pour l'hypothèse : table IV ;
 - $t_{\hat{\alpha}}$, statistique de STUDENT associée à α : table I.
 3. $H_0 : \alpha \neq 0, \rho = 1 \Leftrightarrow H_0 : \alpha \neq 0, \phi = 0$. La loi limite sous H_0 de $t_{\hat{\phi}_T} = t_{\hat{\rho}-1}$ est $\mathcal{N}(0, 1)$.
 4. $H_0 : \alpha = 0, \beta = 0, \rho = 1 \Leftrightarrow H_0 : \alpha = 0, \beta = 0, \phi = 0$ ou $H_{01} : \beta = 0, \rho = 1 \Leftrightarrow H_{01} : \beta = 0, \phi = 0$.
 - lois sous H_0
 - $T_{\hat{\phi}_T} = T_{\hat{\rho}-1} \rightarrow$ table B5 cas 4 ;
 - $t_{\hat{\phi}_T} = t_{\hat{\rho}-1} \rightarrow$ table B6 cas 4 ;
 - $\hat{\Phi}_1$, statistique de FISHER pour l'hypothèse : table v ;
 - $t_{\hat{\alpha}}$: table II ;
 - $t_{\hat{\beta}}$: table III.
 - lois sous H_{01}
 - $T_{\hat{\phi}_T} = T_{\hat{\rho}-1} \rightarrow$ table B5 cas 4 ;
 - $t_{\hat{\phi}_T} = t_{\hat{\rho}-1} \rightarrow$ table B6 cas 4 ;
 - $\hat{\Phi}_1$, statistique de FISHER pour l'hypothèse : table VI.

Mise en œuvre pratique des tests On choisit d'abord entre les cadres donnés par le cas 2 ou le cas 4 suivant que le graphique présente une tendance (cas 4) ou non (cas 2).

On se place dans le cadre ADF en choisissant p suffisamment grand pour avoir $\varepsilon_t \rightsquigarrow \mathcal{BB}$. Puisque la loi des $\hat{\alpha}_i$ est standard dans tous les cas, on commence par réduire (éventuellement) p en menant des tests de nullité des derniers retards (FISHER ou STUDENT).

Cas 2 La difficulté de la construction d'une procédure rigoureuse de tests emboîtés provient du fait que la loi de $t_{\hat{\rho}-1} = t_{\hat{\phi}_T}$ dépend de la vraie valeur de α , qui est elle-même inconnue. Cependant, on peut remarquer que, pour un seuil de test donné, la valeur critique c_2^a associée à $t_{\hat{\phi}_T}$ dans le cas où $\alpha = 0$ est inférieure à la valeur critique c_3^a qui lui est associée quand $\alpha \neq 0$ (Cas 3). Par exemple, pour $T = +\infty$ et $a = 0,05$, ces valeurs critiques sont $c_2^a = -2,86$ et $c_3^a = -1,645$ (quantile à 5% de $\mathcal{N}(0, 1)$).

On peut donc proposer la démarche suivante :

- Si $t_{\hat{\phi}_T} < c_2^a$, on rejette l'hypothèse $\rho = 1$ au seuil a , quelle que soit la vraie valeur de α ;
 - Si $t_{\hat{\phi}_T} < c_3^a$, on accepte l'hypothèse $\rho = 1$ au seuil a , quelle que soit la vraie valeur de α (plus exactement, on ne rejette pas cette hypothèse). On peut ensuite mener un test de l'hypothèse jointe $H_0 : \alpha = 0, \rho = 1$ en utilisant la statistique $\hat{\Phi}_1$ et la valeur critique k_2^a associée (table IV) :
 - si $\hat{\Phi}_1 < k_2^a$, on accepte H_0 au seuil a ;
 - si $\hat{\Phi}_1 < k_3^a$, on refuse H_0 , donc on considère que le vrai modèle est celui du cas 3.
- Par exemple, pour $T = +\infty$ et $a = 0,05$, $k_2^a = 4,59$.
- Si $c_2^a < t_{\hat{\phi}_T} < c_3^a$, on ne peut rien conclure au vu de la statistique $t_{\hat{\phi}_T}$. On mène donc le test de l'hypothèse jointe $H_0 : \alpha = 0, \rho = 1$.

- si $\widehat{\Phi}_1 < k_2^a$, on accepte H_0 au seuil a ;
- si $\widehat{\Phi}_1 < k_2^a$, on refuse H_0 au seuil a ; on se trouve vraisemblablement dans le cas où $\alpha \neq 0$ et $\rho < 1$; ceci peut être contrôlé en examinant la statistique de STUDENT associée à α .

Cas 4 Le problème est ici que les lois limites ne sont connues que lorsque $\beta = 0$, alors que la vraie valeur de β est inconnue. Ceci provient du fait que le modèle est mal adapté au cas de séries présentant une tendance déterministe linéaire, comme on le verra ci-dessous. On choisira donc plutôt, dans ce cas, de recourir au test de SCHMIDT-PHILLIPS.

Dans le cadre des tests de DICKEY-FULLER, la seule procédure de tests emboîtés qui puisse être proposée est la suivante :

- si $\widehat{\Phi}_3 < k_3^a$ (table VI), on accepte l'hypothèse $H_{01} : (\beta = 0, \rho = 1)$ au seuil a , quelle que soit la vraie valeur de α . Par exemple, pour $T = +\infty$, $a = 0,05$, $k_3^a = 6,25$. On teste ensuite l'hypothèse $H_0 : (\alpha = 0, \beta = 0, \rho = 1)$ à l'aide de la statistique $\widehat{\Phi}_2$:
- si $\widehat{\Phi}_2 < k_4^a$ (table V), on accepte H_0 ;
- si $\widehat{\Phi}_2 > k_4^a$, on refuse H_0 .
- si $\widehat{\Phi}_3 > k_3^a$, on refuse H_{01} , et donc aussi H_0 .

Les tests de Phillips-Perron

L'idée sous-jacente aux tests ADF est qu'en remplaçant les modèles du cadre DF :

$$\Delta y_t = d_t + \phi y_{t-1} + \eta, \begin{cases} d_t = 0 & \text{cas 1} \\ d_t = \alpha & \text{cas 2 et 3} \\ d_t = \alpha + \beta t & \text{cas 4} \end{cases}$$

par des modèles du type :

$$\Delta y_t = d_t + \phi y_{t-1} + \sum_{i=1}^{p-1} \alpha_i \Delta y_{t-i} + \eta$$

On peut toujours choisir p assez grand pour conserver l'hypothèse de bruit blanc sur η . Ceci entraîne que les lois limites des estimateurs des paramètres caractérisant la nature stochastique de la série sont identiques à celles du cadre DF.

PHILLIPS et PERRON ont proposé une autre façon de traiter l'auto-corrélation éventuelle du processus (Δy_t) . Les modèles considérés ont la même forme que ceux du cadre DF :

$$\Delta y_t = d_t + \phi y_{t-1} + u_t$$

mais on admet la possibilité que les u_t soient auto-corrélés. Les auteurs montrent que, sous réserve d'introduire un terme correctif adapté, les lois des statistiques $T_{\widehat{\phi}_T} = T_{\widehat{\rho}_T-1}$ et $t_{\widehat{\phi}_T} = t_{\widehat{\rho}_T-1}$ sont asymptotiquement identiques à celles qui sont observées dans le cadre DF. Ces termes correctifs sont fondés sur des estimateurs convergents de $\sigma_u^2 = \gamma_u(0)$ et de $\omega^2 = 2\pi f_u(0)$, c'est-à-dire :

$$\omega^2 = \sum_{k \in \mathbb{Z}} \gamma_u(k) = \lim_T V \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T u_t \right)$$

Ces estimateurs sont calculés comme suit :

- on estime le modèle par les MCO et on calcule les résidus estimés \widehat{u}_t ;

– on pose :

$$\forall k \geq 0, \hat{\gamma}_u(k) = \frac{1}{T} \sum_{t=k+1}^T \hat{u}_t \hat{u}_{t-k}$$

et

$$\widehat{\omega_{TK}^2} = \hat{\gamma}_u(0) + 2 \sum_{k=1}^K \left(1 - \frac{k}{K+1}\right) \hat{\gamma}_u(k)$$

avec K suffisamment grand (estimateur de NEWEY-WEST). En général, on choisit K de l'ordre de \sqrt{T} .

Pour les différentes valeurs possibles de d_t ($d_t = 0, d_t = \alpha, d_t = \alpha + \beta t$), on obtient des lois identiques à celles du cadre DF en remplaçant :

– $T_{\hat{\phi}_T} = T_{\hat{\rho}_{T-1}}$ par :

$$T_{\hat{\phi}_T} - \frac{1}{2} T^2 \frac{\widehat{\sigma_{\hat{\phi}_T}^2}}{\widehat{\sigma_u^2}} (\widehat{\omega_{TK}^2} - \hat{\gamma}_0)$$

– $t_{\hat{\phi}_T} = t_{\hat{\rho}_{T-1}}$ par :

$$\sqrt{\frac{\hat{\gamma}_0}{\widehat{\omega_{TK}^2}}} t_{\hat{\phi}_T} - \frac{1}{2} T \frac{\widehat{\sigma_{\hat{\phi}_T}^2}}{\widehat{\sigma_u^2}} \frac{(\widehat{\omega_{TK}^2} - \hat{\gamma}_0)}{\widehat{\omega_{TK}^2}}$$

Le test de Schmidt-Phillips

Le problème posé par le cas 4 des tests DF et ADF est que les paramètres n'ont pas la même interprétation sous l'hypothèse nulle et sous l'hypothèse alternative. Considérons en effet le modèle : $y_t = \alpha + \beta t + \rho y_{t-1} + \eta$ et l'hypothèse : $H_{01} : \beta = 0, \rho = 1$.

– Sous l'hypothèse alternative, on a :

$$\begin{aligned} (1 - \rho L)y_t &= \alpha + \beta t + \eta \Leftrightarrow y_t = (1 - \rho L)^{-1}(\alpha + \beta t + \eta) \\ \Leftrightarrow y_t &= \frac{\alpha}{1-\rho} + \beta(t - \rho(t-1)) + \sum_{k=0}^{\infty} \rho^k \varepsilon_{t-k} \\ \Leftrightarrow y_t &= a + bt + u_t \text{ avec } b = \beta(1 - \rho), a = \beta\rho + \frac{\alpha}{1-\rho} \\ u_t &= \sum_{k=0}^{\infty} \rho^k \varepsilon_{t-k} \sim I(0) \end{aligned}$$

– Sous l'hypothèse H_{01} , on a :

$$y_t = y_0 + \alpha t + \sum_{k=0}^t \varepsilon_{t-k}$$

Donc, sous H_a , y_t est stationnaire autour d'une tendance déterministe de pente $b = \beta(1 - \rho)$, et sous H_{01} , y_t est non stationnaire autour d'une tendance déterministe de pente α .

La formulation du modèle n'est donc pas satisfaisante. SCHMIDT et PHILLIPS ont proposé un modèle et un test beaucoup mieux adapté au cas des séries présentant une tendance. Dans ce modèle, on suppose que $y_t = \alpha + \beta t + u_t$, avec (u_t) non stationnaire sous H_0 et (u_t) stationnaire sous H_1 .

Méthode de test pour le modèle de base Dans le modèle de base, on suppose que :

$$\begin{aligned} y_t &= \alpha + \beta t + u_t \\ u_t &= \rho u_{t-1} + \eta \text{ avec } |\rho| \leq 1, \eta \sim \mathcal{BB}(0, \sigma^2) \end{aligned}$$

On pose : $H_0 : \rho = 1$.

On calcule :

$$\widehat{\beta} = \frac{y_T - y_1}{T - 1}, \widehat{\alpha} = y_1 - \widehat{\beta}, \widehat{u}_t = y_t - \widehat{\alpha} - \widehat{\beta}t$$

Comme le modèle s'écrit aussi :

$$\Delta y_t = b + u_t - u_{t-1} = b + (\rho - 1)u_{t-1} + \eta = b + \phi u_{t-1} + \eta$$

on estime par les MCO le modèle : $\Delta y_t = \mu + \phi \widehat{u}_{t-1} + \eta_t$ et on teste : $H_0 : \phi = 0$ contre $H_1 : \phi < 0$. Soit $\widehat{\phi}_T$ l'estimateur des MCO de ϕ , et $t_{\widehat{\phi}_T}$ la statistique de STUDENT associée, on refuse H_0 au seuil a si $T_{\widehat{\phi}_T} = T_{\widehat{\rho}_T - 1} < c^a$ ou si $t_{\widehat{\phi}_T} < c_1^a$ avec c^a et c_1^a obtenus dans la table 1A (par exemple, pour $T = 100$, $a = 0,05$: $c_a = -3,04$).

Cas général On suppose toujours :

$$\begin{cases} y_t = \alpha + \beta t + u_t \\ u_t = \rho u_{t-1} + \eta \end{cases}$$

mais on ne fait plus l'hypothèse que (η) est un \mathcal{BB} . On effectue le même type de correction que dans le test de PHILLIPS-PERRON pour prendre en compte l'auto-corrélation éventuelle des η . La procédure de test proposée par les auteurs est cependant un peu différente de la précédente. Tenant compte du fait que :

$$\begin{aligned} (1 - \rho L)y_t &= (1 - \rho L)(\alpha + \beta t) + (1 - \rho L)u_t \\ &= a + bt + \eta \end{aligned}$$

Ils estiment directement le modèle :

$$\Delta y_t = a + bt + \phi y_{t-1} + \eta$$

par les MCO, et calcule $\widehat{\sigma}_\varepsilon^2$ et $\widehat{\omega}_{KT}^2$ pour les résidus $\widehat{\varepsilon}_t$ comme cela a été fait pour \widehat{u}_t au paragraphe 2, et $\lambda^2 = \frac{\widehat{\sigma}_\varepsilon^2}{\widehat{\omega}_{KT}^2}$.

On reprend ensuite la démarche exposée au 3.1.2 pour calculer $\widehat{\phi}_T = \widehat{\rho}_T - 1$ et $t_{\widehat{\phi}_T}$. Les auteurs montrent que les lois limites de $\frac{T_{\widehat{\phi}_T}}{\lambda^2}$ et $\frac{t_{\widehat{\phi}_T}}{\lambda^2}$ sont identiques respectivement aux lois obtenues pour $T_{\widehat{\phi}_T}$ et $t_{\widehat{\phi}_T}$ au 3.1.2.

Le test KPSS (Kwiatowski, Phillips, Schmidt, Shin)

Comme on l'a dit en introduction, l'hypothèse nulle de ce test est celle de la stationnarité (autour d'une constante ou d'une tendance déterministe linéaire), contrairement à tous les cas précédents. Deux cas sont donc étudiés :

1. $y_t = r_t + \eta$ où $\eta \rightsquigarrow I(0)$, $r_t = r_{t-1} + u_t$, $u_t \rightsquigarrow \mathcal{BB}(0, \sigma_u^2)$
2. $y_t = \beta t + r_t + \eta$ avec $\eta \rightsquigarrow I(0)$, $r_t = r_{t-1} + u_t$, $u_t \rightsquigarrow \mathcal{BB}(0, \sigma_u^2)$

La statistique de test utilisée correspond à la statistique du test du score lorsque les η sont i.i.d. de loi $\mathcal{N}(0, \sigma_u^2)$. Cependant, elle est corrigée de façon à tenir compte de l'auto-corrélation des η dans le cas général.

La procédure employée est alors la suivante :

- on régresse y_t sur une constante (cas 1) ou sur une constante et un trend (cas 2) et on calcule les résidus \hat{u}_t de la régression ($\hat{u}_t = y_t - \bar{y}$ dans le cas 1, $\hat{u}_t = y_t - \hat{\alpha} - \hat{\beta}t$ dans le cas 2) ;
- on calcule :

$$\hat{S}_t = \sum_{k=1}^t \hat{u}_k$$

et

$$\hat{\omega}_{TK}^2 = \frac{1}{T} \sum_{t=1}^T \hat{u}_t^2 + 2 \sum_{k=1}^K \left(1 - \frac{k}{K+1}\right) \frac{1}{T} \sum_{t=k+1}^T \hat{u}_t \hat{u}_{t-k}$$

avec K de l'ordre de \sqrt{T} .

- la statistique de test est :

$$\eta = \frac{\frac{1}{T^2} \sum_{t=1}^T \hat{S}_t^2}{\hat{\omega}_{TK}^2}$$

La loi limite de η est tabulée dans le cas 1 (η_μ dans la table) et dans le cas 2 (η_τ dans la table).

On refuse $H_0 : \sigma_u^2 = 0$ au seuil α lorsque la valeur obtenue de η est supérieure à la valeur critique correspondante.

3.2 Deuxième phase de l'identification : choix de p et q

On suppose que l'on a déjà déterminé d et on travaille éventuellement sur $Y_t = (1 - L)^d X_t$. On assimile alors (Y_t) à un processus $ARMA(p, q)$. On se propose donc de déterminer les valeurs de p et q .

3.2.1 Résultats préliminaires

Soit $Y_t \sim I(d)$ et $X_t = (1 - L)^d Y_t$.

On a vu que si $X_t \sim AR(p)$ alors

$$\begin{cases} r(h) = 0 \text{ si } h > p \text{ et } r(p) \neq 0 \\ \rho^i(h) = 0 \text{ si } h > p \text{ et } \rho^i(p) \neq 0 \end{cases}$$

Et si $X_t \sim MA(q)$, alors $\rho(h) = 0$ si $h > q$ et $\rho(q) \neq 0$.

Enfin on sait que si $Y_t \sim I(0)$, alors

$$\begin{aligned} \hat{r}_T(h) &\xrightarrow{P} r(h) \\ \hat{\rho}_T^i(h) &\xrightarrow{P} \rho^i(h) \\ \hat{\rho}_T(h) &\xrightarrow{P} \rho(h) \end{aligned}$$

De plus si (ε_t) est stationnaire à l'ordre 4 ($\mathbb{E}(\varepsilon_t^4) = \mu < +\infty$) alors tous ces estimateurs sont asymptotiquement normaux.

Remarque 3.2.1 (Calcul de $\hat{\rho}_T^i$ pour $Y_t \rightsquigarrow ARMA(p, q)$) Si $\phi(L)Y_t = \theta(L)\varepsilon_t$, on a vu que pour $\eta_t \rightsquigarrow \mathcal{BB}$ bien choisi, le processus (Z_t) respectant l'équation $\theta(L)Z_t = \phi(L)\eta_t$ vérifie $\rho_Z(h) = \rho_Y^i(h)$.

On a

$$Z_t = \theta(L)^{-1}\phi(L)\eta_t = A(L)\eta_t = \sum_{j=0}^{+\infty} a_j \eta_{t-j}$$

et

$$\rho_Z(h) = \frac{\sum_{j=0}^{+\infty} a_j a_{j+h}}{\sum_{j=0}^{+\infty} a_j^2} = \rho_{iY}(h)$$

On prendra

$$\hat{\rho}_Y^i(h) = \hat{\rho}_Z(h) = \frac{\sum_{j=0}^K \hat{a}_j \hat{a}_{j+h}}{\sum_{j=0}^K \hat{a}_j^2}$$

où K est suffisamment grand.

$$\begin{aligned} \text{Par ailleurs : } \theta(L)^{-1}\phi(L)Y_t &= \varepsilon_t \\ \Rightarrow A(L)Y_t &= \varepsilon_t \\ \Rightarrow Y_t &= -\sum_{j=1}^{+\infty} a_j Y_{t-j} + \varepsilon_t \end{aligned}$$

Il est possible d'obtenir $\hat{a}_1, \dots, \hat{a}_K$ en régressant Y_t sur Y_{t-1}, \dots, Y_{t-K} .

3.2.2 Choix de p pour un $AR(p)$

On montre que

$$\begin{aligned} \sqrt{T}(\hat{r}_T(h) - r(h)) &\xrightarrow{loi} \mathcal{N}(0, 1) \\ \sqrt{T}(\hat{\rho}_T^i(h) - \rho^i(h)) &\xrightarrow{loi} \mathcal{N}(0, 1) \end{aligned}$$

On teste $H_0 : r(h) = 0$ contre $H_0 : r(h) \neq 0$. On refuse H_0 si $\sqrt{T}|\hat{r}_T(h)| > 1,96$ au seuil de 5%. L'intervalle de confiance à 95% pour $r(h) : \left[\hat{r}_T(h) - \frac{1,96}{\sqrt{T}}, \hat{r}_T(h) + \frac{1,96}{\sqrt{T}} \right]$.

Si $\hat{r}_T(h)$ est non significativement différent de zéro pour $h > p$ et $\hat{r}_T(p) \neq 0$, alors p est l'ordre de l' AR .

Sur l'auto-corrélogramme c'est le rang du dernier pic significatif.

3.2.3 Choix de q pour un $MA(q)$

Si $X_t \rightsquigarrow MA(q)$, on montre que

$$\sqrt{T} \frac{\hat{\rho}_T(h) - \rho(h)}{\sqrt{\sum_1^q \rho^2(k)}} \xrightarrow{loi} \mathcal{N}(0, 1)$$

Si $h > q$,

$$\sqrt{T} \frac{\hat{\rho}_T(h)}{\sqrt{\sum_1^q \rho^2(k)}} \xrightarrow{loi} \mathcal{N}(0, 1)$$

Comme $\hat{\rho}_T(k) \rightarrow \rho(k)$, on a

$$\sqrt{T} \frac{\hat{\rho}_T(h)}{\sqrt{\sum_1^q \hat{\rho}^2(k)}} \xrightarrow{loi} Student$$

3.2.4 Choix de (p, q) pour un $ARMA(p, q)$

Rappel : $AR(p) = ARMA(p, 0)$ et $MA(q) = ARMA(0, q)$.

Si $X_t \rightsquigarrow ARMA(p, q)$ tel que $\phi(L)X_t = \theta(L)\varepsilon_t$ alors (X_t) admet une représentation $AR(\infty)$ donnée par :

$$\theta(L)^{-1}\phi(L)X_t = \sum_{k=0}^{+\infty} a_k X_{t-k} = \varepsilon_t$$

où $a_0 = 1$ et $\sum_{k=0}^{+\infty} |a_k| < +\infty$.

Cette représentation peut être approximée par un $AR(P)$ pour P assez grand :

$$\sum_{k=0}^P a_k X_{t-k} \# \varepsilon_t$$

De la même façon X_t admet une représentation $MA(\infty)$ donnée par :

$$X_t = \phi(L)^{-1}\theta(L)\varepsilon_t = \sum_{k=0}^{+\infty} b_k \varepsilon_{t-k}$$

qui peut aussi être approximée par un $MA(Q)$ pour Q assez grand :

$$\sum_{k=0}^Q b_k \varepsilon_{t-k} \# X_t$$

Exemple 3.2.1 Si $\hat{r}_T(h)$ et $\hat{\rho}_T^i(h)$ sont significativement non nuls pour $h \leq 3$ seulement et si $\hat{\rho}_T(h)$ est significativement non nul pour $h \leq 4$ seulement, on est amené à estimer un $AR(3)$ et un $MA(4)$ ce qui pousse à choisir un $ARMA(p, q)$ avec $p \leq 3$ et $q \leq 4$.

3.3 Estimation

A l'issue des phases précédentes, on a choisi d et divers couples (p, q) compatibles avec les données. Le modèle s'écrit :

$$(1 - L)^d \phi(L)X_t = \mu + \theta(L)\varepsilon_t$$

On suppose que $\varepsilon_t \text{ iid } \rightsquigarrow \mathcal{N}(0, \sigma_\varepsilon^2)$.

Les paramètres à estimer sont $\omega = (\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q)$ et σ_ε^2 .

On calcule l'estimateur du maximum de vraisemblance :

$$\begin{aligned} & \ln(x_{-(p+d)}, \dots, x_{-1}, x_1, x_0, \dots, x_T) \\ & \ln(x_1, \dots, x_T | x_{-(p+d)}, \dots, x_{-1}) \\ & \ln(x_{-(p+d)}, \dots, x_{-1}) \end{aligned}$$

On a recours à des procédures numériques de maximisation de la vraisemblance :

– la valeur initiale $\frac{\mu}{\phi(1)}$ est estimée par la moyenne empirique de $(1 - L)^d X_t$.

– les équations de YULE-WALKER donnent un estimateur initial pour $(\varphi_1, \dots, \varphi_p)$:

$$\begin{pmatrix} \rho(q+1) \\ \vdots \\ \rho(q+p) \end{pmatrix} = \begin{pmatrix} \hat{\rho}(q) & \dots & \hat{\rho}(p+q-1) \\ \vdots & \ddots & \vdots \\ \hat{\rho}(p+q-1) & \dots & \hat{\rho}(q) \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_p \end{pmatrix}$$

– on obtient de la même façon $(\theta_1, \dots, \theta_q)$ à partir des $\hat{\rho}_T^i(h)$.

Propriété de l'EMV Soit $\omega = (\theta, \sigma_\varepsilon^2)'$

On a alors

$$\sqrt{T}(\hat{\omega}_T - \omega) \xrightarrow{loi} \mathcal{N}\left(0, \begin{pmatrix} \Omega & 0 \\ 0 & \alpha \end{pmatrix}\right)$$

On en déduit des intervalles de confiance asymptotiques pour les paramètres. On effectue des tests du type :

- $H_0 : \phi_p = 0$ contre $H_1 : \phi_p \neq 0$
- $H_0 : \theta_p = 0$ contre $H_1 : \theta_p \neq 0$
- $H_0 : \mu = 0$ contre $H_1 : \mu \neq 0$

3.3.1 Cas d'un $AR(p)$

On se place dans le cadre suivant :

$$\begin{aligned} y_t &= (1-L)^d(x_t - m) \\ \phi(L)y_t &= \varepsilon_t \\ y_t &= \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \varepsilon_t \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2) \text{ i.i.d.} \end{aligned}$$

$\begin{pmatrix} y_{p+1} \\ \vdots \\ y_T \end{pmatrix}$ suit, conditionnellement à $\begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix}$, une loi normale dont la densité s'écrit :

$$\begin{aligned} \ell(y_{p+1}, \dots, y_T | y_1, \dots, y_p) &= \prod_{t=p+1}^T \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{y_t - (\varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p})}{2\sigma^2}\right) \\ \Rightarrow \ln \ell(y_{p+1}, \dots, y_T | y_1, \dots, y_p) &= -\frac{T}{2} \ln 2\pi - \frac{T}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=p+1}^T [y_t - (\varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p})] \end{aligned}$$

La maximisation de la vraisemblance conditionnelle donne $\hat{\varphi}_1, \dots, \hat{\varphi}_p$ qui peuvent être calculés en régressant (MCO) y_t sur y_{t-1}, \dots, y_{t-p} .

CLS (méthode approximative) On a

$$\ell(y_1, \dots, y_T) = \ell(y_{p+1}, \dots, y_T | y_1, \dots, y_p) \times \ell(y_1, \dots, y_p)$$

où

$$\begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} \sim \mathcal{N}\left(0, \underbrace{\begin{pmatrix} \gamma(0) & \dots & \gamma(p-1) \\ \vdots & \ddots & \vdots \\ \gamma(p-1) & \dots & \gamma(0) \end{pmatrix}}_{V_p}\right)$$

V_p étant calculable en fonction de $\varphi_1, \dots, \varphi_p$, il en découle :

$$\underbrace{\ell(y_1, \dots, y_p)}_z = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\det V_p}} e^{-\frac{1}{2} z' V_p z}$$

$$\max_{\sigma^2, \varphi_1, \dots, \varphi_p} \ln \ell(y_1, \dots, y_p) = \max_{\sigma^2, \varphi_1, \dots, \varphi_p} -\frac{p}{2} \ln 2\pi - \frac{1}{2} \ln \det V_p - \frac{1}{2} z' V_p z$$

Pour plus de détails sur le cas $AR(p)$ on pourra consulter [3].

3.3.2 Cas d'un $MA(q)$

On considère :

$$\begin{aligned} (1-L)^d Y_t &= \theta(L) \varepsilon_t \\ Y_t &= X_t - m_t \text{ où } \mathbb{E}X_t = m_t \\ \varepsilon_t &\sim \mathcal{N}(0, \sigma^2) \text{ i.i.d.} \\ \theta(L) \varepsilon_t &= \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \end{aligned}$$

La vraisemblance du modèle est $\ell(y_1, \dots, y_T | \varepsilon_0, \varepsilon_{-1}, \dots, \varepsilon_{-q+1})$, fonction dont le maximum doit être calculé de façon numérique.

- Soit on suppose $\varepsilon_0 = \varepsilon_{-1} = \dots = \varepsilon_{-q+1} = 0 \rightarrow$ méthode numérique
 $\rightarrow 2^{\text{e}}$ approximation \rightarrow CLS
- Soit on ne fait pas cette hypothèse \rightarrow méthode numérique
 \rightarrow approximation \rightarrow ULLS

Pour plus de précisions *cf.* [2].

3.3.3 Cas d'un $ARMA(p, q)$

Diverses approximations sont possibles : MLE, CLS, ULLS (*cf.* [2]).

3.4 Vérifications *a posteriori*

3.4.1 Tests sur les paramètres

Exemple 3.4.1 On désire tester $H_0 : \varphi_p = 0$ contre l'hypothèse alternative $H_1 : \varphi_p \neq 0$, on a :

$$\sqrt{T}(\hat{\varphi}_{T,p} - \varphi_p) \xrightarrow{loi} \mathcal{N}\left(0, \mathbb{V}_{as}\left(\sqrt{T}\hat{\varphi}_{T,p}\right)\right)$$

D'où

$$\sqrt{T} \frac{\hat{\varphi}_{T,p} - \varphi_p}{\mathbb{V}_{as}\left(\sqrt{T}\hat{\varphi}_{T,p}\right)} \xrightarrow{loi} \mathcal{N}(0, 1)$$

φ_p est significatif au seuil de 5% (asymptotiquement) si

$$\underbrace{\left| \frac{\hat{\varphi}_{T,p}}{(\mathbb{V}_{as}(\hat{\varphi}_{T,p})^{1/2})} \right|}_{|t_{\hat{\varphi}_{T,p}}|} > 1,96$$

Exemple 3.4.2 Le même test s'applique avec θ_q à la place de φ_p , on refuse alors H_0 (i.e. θ_q est significatif) si $|t_{\hat{\varphi}_{T,p}}| > 1,96$.

Remarque 3.4.1 (Econométrie asymptotique)

$$\hat{\beta} = \beta + \frac{(X'X)^{-1} X'}{T} \varepsilon \text{ et } \frac{X'X}{T} \rightarrow Q$$

$$\sqrt{T} \frac{X' \varepsilon}{T} \xrightarrow{\text{loi}} \mathcal{N} \left(0, \underbrace{\lim \left(\frac{X'X}{T} \right)}_Q \right)$$

$$y_t = a + bx_t + \varepsilon_t$$

$$\frac{\hat{b} - b}{\hat{\sigma}_{\hat{b}}} \underset{H_0}{\rightsquigarrow} \text{Student}(N - 2)$$

Si $\hat{\varphi}_p$ n'est pas significatif (ou θ_q) on relance l'estimation en remplaçant p par $p - 1$ (ou q par $q - 1$).

Si $\hat{\mu}$ n'est pas significatif, on relance l'estimation sans terme constant.

3.4.2 Tests sur les résidus

On se place toujours dans les mêmes conditions :

$$\begin{aligned} \phi(L)Y_t &= \mu + \theta(L)\varepsilon_t \\ \Rightarrow Y_t - \varphi_1 Y_{t-1} - \dots - \varphi_p Y_{t-p} &= \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \\ \Rightarrow \hat{\varepsilon}_t &= Y_t - \hat{\varphi}_1 Y_{t-1} - \dots - \hat{\varphi}_p Y_{t-p} - \hat{\mu} + \hat{\theta}_1 \varepsilon_{t-1} + \dots + \hat{\theta}_q \varepsilon_{t-q} \end{aligned}$$

Les résidus estimés (à savoir $\hat{\varepsilon}_t$) sont-ils compatibles avec l'hypothèse de bruit blanc de ε_t ?
Pour cela on effectue le **test du porte-manteau** :

$$\hat{\rho}_k(\hat{\varepsilon}) = \frac{1}{T-1} \sum_{t=k+1}^T \hat{\varepsilon}_t \hat{\varepsilon}_{t-k}$$

L'auto-corrélation empirique d'ordre k de $\hat{\varepsilon}_t$, est alors :

$$Q_K = T \sum_{k=1}^K \hat{\rho}_{\hat{\varepsilon}}^2(k)$$

Pour K assez grand ($K \geq 12$), on montre que si $\varepsilon_t \sim \mathcal{BB}(0, \sigma^2)$ alors

$$Q_K \xrightarrow{L} \chi^2(K - p - q)$$

On refuse $H_0 : \varepsilon_t \sim \mathcal{BB}$ au seuil α si $Q_K > \chi_{1-\alpha}^2(K - p - q)$.

Remarque 3.4.2 Q_K peut être éventuellement remplacé par :

$$Q'_K = T(T+2) \sum_{k=1}^K \frac{1}{T-k} \hat{\rho}_{\hat{\varepsilon}}^2(k)$$

3.5 Choix du modèle

A l'issue des phases d'estimation et de vérification il reste en général plusieurs modèles possibles pour représenter les données. Pour choisir un modèle on peut se fier à plusieurs critères :

- $\hat{\sigma}^2$ petit,
- critère de parcimonie : $p + q$ minimal,
- critère de qualité de la prévision (*cf.* plus loin),
- critère d'information .

On suppose $\varepsilon_t \sim \mathcal{BB}(0, \sigma^2)$ i.i.d. On considère $f_0(x) = f(x, \omega_0, \sigma_0^2)$ la vraie loi (inconnue) du processus et $f(x) = f(x, \omega, \sigma^2)$ la famille de loi correspondant au modèle $ARMA(p, q)$ estimé. L'écart entre ces lois est mesuré par :

$$\Delta(f, f_0) = \mathbb{E}_0 \left[-2 \ln \frac{f(x)}{f_0(x)} \right]$$

Cette quantité est positive (d'après JENSEN) et est nulle si et seulement si $f(x) = f_0(x)$ p.s.

En pratique on cherche à minimiser l'écart entre f_0 et f . Il existe différentes façons d'approximer ce **critère d'information** :

- $AIC(p, q) = T \ln \hat{\sigma}^2 + 2(p + q)$, critère d'information d'AKAÏKE
- $SBC(p, q) = T \ln \hat{\sigma}^2 + (p + q) \ln T$
- $HQ(p, q) = T \ln \hat{\sigma}^2 + (p + q)c \ln(\ln T)$ avec $c > 2$, critère d'information HANNAN-QUINN

On cherche donc à minimiser la quantité d'informations.

Chapitre 4

Prévision dans les *ARMA* et les *ARIMA*

Introduction

On suppose (p, d, q) connus. On dispose d'observations x_1, \dots, x_T et on veut faire une prévision à l'horizon H , c'est-à-dire prévoir x_{T+1}, \dots, x_{T+H} .

On remplacera $(\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q, \mu, \sigma^2)$ par leurs estimateurs.

4.1 Prévisions dans un *AR*(p)

On suppose que l'on a le modèle :

$$\phi(L)X_t = \mu + \varepsilon_t$$

où $\phi(L) = \varphi_1 L + \varphi_2 + \dots + \varphi_p L^p$

Soit $x_{t+1} = \mu + \varphi_1 x_t + \dots + \varphi_p x_{t-p+1} + \varepsilon_{t+1}$.

Définition 4.1.1 (Prévision optimale)

$${}_t X_{t+j}^* = EL(X_{t+j} | \underline{X}_t)$$

On trouve ensuite :

$$\begin{aligned} {}_t X_{t+1}^* &= EL(x_{t+1} | \underline{x}_t) = \mu + \varphi_1 x_t + \dots + \varphi_p x_{t+1-p} + EL(\varepsilon_{t+1} | \underline{x}_t) \\ {}_t X_{t+2}^* &= EL(x_{t+2} | \underline{x}_t) = \mu + \varphi_1 {}_t X_{t+1}^* + \varphi_2 x_t + \dots + \varphi_p x_{t+1-p} + EL(\varepsilon_{t+1} | \underline{x}_t) \\ &\vdots \\ {}_t X_{t+h}^* &= \mu EL(x_{t+h} | \underline{x}_t) = \sum_{j=1}^p \varphi_j {}_t x_{t+h-j}^* \end{aligned}$$

avec ${}_t x_{t+h-j}^* = x_{t+h-j}$ si $t+h-j \leq t$ et $h \leq j$.

D'où l'écriture

$$\boxed{{}_t X_{t+h}^* = \mu + \sum_{j=1}^{h-1} \varphi_j {}_t x_{t+h-j}^* + \sum_{j=h}^p \varphi_j {}_t x_{t+h-j}}$$

On utilise ensuite l'équation de récurrence :

$$\phi(L)_t X_{t+h}^* = \mu \Leftrightarrow \phi(L) \underbrace{({}_t X_{t+h}^* - m)}_{{}_t Y_{t+h}^*} = 0$$

où $m = \frac{\mu}{\phi(1)} = \mathbb{E}X_t$

${}_t Y_{t+h}^*$ est la solution de l'équation de récurrence de polynôme caractéristique $Z^p \phi(\frac{1}{Z})$. On en déduit que ${}_t Y_{t+h}^*$ est de la forme :

$${}_t Y_{t+h}^* = \sum_{i=1}^r \sum_{j=1}^{m_i-1} \alpha_{ij} \lambda_i^j h^j$$

avec les $\frac{1}{\lambda_i}$ racines distinctes de $\phi(Z)$ avec la multiplicité m_i .

α_{ij} est obtenu à partir des p conditions initiales (observations ou prévisions).

Exemple 4.1.1 On considère un *AR* tel que $\phi(L) = (1 - \varphi L)^2$ où $|\varphi| < 1$. On a donc le processus :

$$\phi(L)X_t = \mu + \varepsilon_t$$

On veut prévoir ${}_t X_{t+h}^*$.

On commence par centrer : $\phi(L)Y_t = \varepsilon_t$

$$\phi(L) \underbrace{({}_t X_{t+h}^* - m)}_{{}_t Y_{t+h}^*} = 0$$

${}_t Y_{t+h}^*$ s'écrit : ${}_t Y_{t+h}^* = \varphi^h (ah + b)$.

Si $h = 0$, ${}_t Y_t^* = Y_t = b$

Si $h = 1$, ${}_t Y_{t+1}^* = 2\varphi Y_t - \varphi^2 Y_{t-1} = \varphi(a + b)$

Par identification $a = Y_t - \varphi Y_{t-1}$ et $b = Y_t$.

On en déduit

$${}_t Y_{t+h}^* = \varphi^h ((Y_t - \varphi Y_{t-1})h + Y_t)$$

4.2 Prévision dans un *MA*(q)

On considère le processus

$$\begin{aligned} X_t &= m + \theta(L)\varepsilon_t \\ &= m + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j} \\ X_{t+h} &= m + \varepsilon_{t+h} - \sum_{j=1}^q \theta_j \varepsilon_{t+h-j} \end{aligned}$$

On a la prévision optimale :

$$\begin{aligned} {}_t X_{t+h}^* &= EL(X_{t+h} | \underline{X}_t) \\ &= EL(X_{t+h} | \underline{\varepsilon}_t) \end{aligned}$$

Si $h > q$, ${}_t X_{t+h}^* = m$.

Si $h \leq q$,

$$\begin{aligned} X_{t+h} &= m + \varepsilon_{t+h} - \sum_{j=1}^{h-1} \theta_j \varepsilon_{t+h-j} - \sum_{j=h}^q \theta_j \varepsilon_{t+h-j} \\ EL(X_{t+h} | \underline{X}_t) &= m + 0 - 0 - \sum_{j=h}^q \theta_j \varepsilon_{t+h-j} \\ {}_t X_{t+h}^* &= m - \sum_{j=h}^q \theta_j \varepsilon_{t+h-j} \end{aligned}$$

Cette forme est exacte mais n'est pas utilisable en pratique car les ε_{t-k} ne sont pas observés pour $k \geq 0$. Mais on peut les calculer à partir des observations en utilisant la forme $AR(\infty)$:

$$\theta(L)^{-1}(X_t - m) = \varepsilon_t \Leftrightarrow \theta(L)^{-1}X_t = \mu + \varepsilon_t$$

où $\mu = \frac{m}{\theta(1)}$

Or $\theta(L)^{-1} = \sum_{k=0}^{\infty} a_k L^k$ où $a_0 = 1$ et $\sum_{k=0}^{\infty} |a_k| < +\infty$.

On a donc

$$\begin{aligned} X_t &= \mu - \sum_{k=1}^{\infty} a_k X_{t-k} + \varepsilon_t \\ X_{t+h} &= \mu - \sum_{k=1}^{\infty} a_k X_{t+h-k} + \varepsilon_{t+h} \end{aligned}$$

Pour les prévisions optimales :

$$\begin{aligned} {}_t X_{t+1}^* &= \mu - \sum_{k=1}^{\infty} a_k X_{t+1-k} \\ &\vdots \\ {}_t X_{t+h}^* &= \mu - \sum_{k=1}^{h-1} a_k X_{t+h-k}^* - \sum_{k=h}^{\infty} a_k X_{t+h-k}^* \end{aligned}$$

En pratique, on n'observe pas les X_t pour $t < 0$. On n'a qu'une prévision approchée en tronquant :

$${}_t \hat{X}_{t+h} = \mu - \sum_{k=1}^{h-1} a_k \hat{X}_{t+h-k}^* - \sum_{k=h}^{t+h} a_k X_{t+h-k}^*$$

En fait on néglige le terme

$$\left\| \sum_{k=t+h+1}^{\infty} a_k X_{t+h-k} \right\|_2 \leq \underbrace{\left(\sum_{k=t+h+1}^{\infty} |a_k| \right)}_{\xrightarrow{h \rightarrow \infty} 0} \|X_i\|_2$$

4.3 Cas d'un ARMA(p, q)

On considère le modèle :

$$\phi(L)X_t = \mu + \theta(L)\varepsilon_t \Leftrightarrow \phi(L)(X_t - m) = \theta(L)\varepsilon_t$$

En posant $Y_t = X_t - m$, on se ramène à $\mu = 0$.

$$Y_t = \varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

4.3.1 Forme AR(∞)

On a en multipliant par $\theta(L)^{-1}$:

$$\underbrace{\theta(L)^{-1}\phi(L)}_{A(L)} Y_t = \varepsilon_t$$

$$\begin{aligned} Y_t &= -\sum_{k=1}^{\infty} a_k Y_{t-k} + \varepsilon_t \\ {}_t Y_{t+h}^* &= -\sum_{k=1}^{h-1} a_k Y_{t+h-k} - \sum_{k=h}^{\infty} a_k Y_{t+h-k} \\ {}_t \hat{Y}_{t+h} &= -\sum_{k=1}^{h-1} a_k \hat{Y}_{t+h-k} - \sum_{k=h}^{\infty} a_k Y_{t+h-k} \end{aligned}$$

4.3.2 Utilisation d'une équation de récurrence

On connaît l'équation de récurrence :

$$Y_{t+h} = \sum_{j=1}^p \varphi_j Y_{t+h-j} + \varepsilon_{t+h} - \sum_{k=1}^q \theta_k \varepsilon_{t+h-k}$$

Si $t+h-q > t$, i.e. $h > q$, alors

$$EL \left(\varepsilon_{t+h} - \sum_{j=1}^p \theta_j \varepsilon_{t+h-j} \middle| \underline{Y}_t \right) = 0$$

D'où

$${}_t Y_{t+h}^* = \sum_{j=1}^p \varphi_j {}_t Y_{t+h-j}^*$$

où ${}_t Y_{t+h-j}^* = Y_{t+h-j}$ si $t+h-j \geq t$ i.e. $j \leq h$.

Pour $h > q$, les ${}_t Y_{t+h}^*$ vérifient l'équation de récurrence de polynôme caractéristique $Z^p \phi\left(\frac{1}{Z}\right)$. on en déduit que

$${}_t Y_{t+h}^* = \sum_{i=1}^r \sum_{j=0}^{m_i-1} \alpha_{ij} \lambda_i^j h^j$$

où on déduit les α_{ij} des conditions initiales ${}_t Y_{t+q}^*, \dots, {}_t Y_{t+q-p}^*$ observés ou prévus.

4.4 Cas d'un ARIMA(p, d, q)

On considère le modèle :

$$\underbrace{(1-L)^d \phi(L)}_{\psi(L)} X_t = \mu + \theta(L)\varepsilon_t$$

On a les conditions initiales $Z' = (X_{-1}, \dots, X_{-p-d}, \varepsilon_{-1}, \dots, \varepsilon_{-q})$.

On se ramène au cas $\mu = 0$ en posant $\mathbb{E}X_t = m_t$. m_t est une solution déterministe de l'équation de récurrence $\psi(L)m_t = \mu$. On pose alors $Y_t = X_t - m$ et on a l'équation

$$\psi(L)Y_t = \theta(L)\varepsilon_t$$

En $t+h$

$$\begin{aligned} \psi(L)Y_{t+h} &= \theta(L)\varepsilon_{t+h} \\ {}_t Y_{t+h}^* &= EL(Y_{t+h}|Y_t, \dots, Y_1, Y_0, Z) \end{aligned}$$

Utilisation de l'approximation auto-régressive

$$Y_t = - \sum_{j=1}^t a_j Y_{t-j} + H'(t)Z + \varepsilon_t$$

avec $H(t) \in \mathbb{R}^{p+d+q}$ et $\lim_{t \rightarrow \infty} \|H(t)\| = 0$

$$\begin{aligned} Y_{t+h} &= \sum_{j=1}^{t+h} a_j Y_{t+h-j} + H'(t+h)Z + \varepsilon_{t+h} \\ {}_t Y_{t+h}^* &= \sum_{j=1}^{t+h} a_j {}_t Y_{t+h-j}^* + H'(t+h)Z \text{ avec } {}_t Y_{t+h-j}^* = Y_{t+h-j} \text{ si } j \leq h \\ {}_t \hat{Y}_{t+h} &= \sum_{j=1}^{t+h} a_j {}_t \hat{Y}_{t+h-j} \end{aligned}$$

Utilisation d'une équation de récurrence $\psi(L)Y_t = \theta(L)\varepsilon_t$

$$\psi(L)Y_{t+h} = \theta(L)\varepsilon_{t+h}$$

$$\psi(L) {}_t Y_{t+h}^* = 0 \text{ si } t+h-q > t \text{ i.e. } h > q$$

Les ${}_t Y_{t+h}^*$ pour $h > q$ sont solutions de l'équation de récurrence de polynôme caractéristique $Z^p + d\psi\left(\frac{1}{Z}\right)$. D'où

$$\begin{aligned} {}_t Y_{t+h}^* &= \sum_{i=1}^r \sum_{j=0}^{m_i-1} \alpha_{ij} \lambda_i^j h^j \\ &= \sum_{j=1}^{d-1} \alpha_{1j} h^j \lambda_1^j h^j + \sum_{i=2}^r \sum_{j=0}^{m_i-1} \alpha_{ij} \lambda_i^j h^j \end{aligned}$$

Remarque 4.4.1 On a la même équation de récurrence pour les ${}_t \hat{Y}_{t+h}$, $h > q$.

4.5 Intervalles de précision

Dans les cas AR , MA , $ARMA$ on sait qu'il existe une représentation $MA(\infty)$:

$$X_t = m + B(L)\varepsilon_t = m + \sum_{k=0}^{\infty} b_k \varepsilon_{t-k}$$

où $\sum |b_k| < \infty$.

On alors en $t+h$

$$\begin{aligned} X_{t+h} &= m + \varepsilon_{t+h} + \sum_{k=1}^{\infty} b_k \varepsilon_{t+h-k} \\ X_{t+h}^* &= EL(X_{t+h} | X_t) = EL(X_{t+h} | \underline{\varepsilon}_t) \\ &= m + \varepsilon_{t+h} + \sum_{k=h}^{\infty} b_k \varepsilon_{t+h-k}, \text{ si } t+h-k \leq t \end{aligned}$$

L'erreur de prévision à l'horizon h est :

$$\begin{aligned} e_t(h) &= X_{t+h} - {}_t X_{t+h}^* \\ &= \varepsilon_{t+h} + \sum_{k=1}^{h-1} b_k \varepsilon_{t+h-k} \end{aligned}$$

Dans les cas $ARMA$, on a vu une approximation MA , en posant $Y_t = X_t - m$:

$$\begin{aligned} Y_t &= \sum_{j=0}^t b_j \varepsilon_{t-j} + \tilde{H}(t)' Z \\ Y_{t+h} &= \sum_{j=0}^t b_j \varepsilon_{t-j} + \tilde{H}(t+h)' Z \\ {}_t Y_{t+h}^* &= EL(Y_{t+h} | Y_t, \dots, Y_0, Z) \\ &= EL(Y_{t+h} | \varepsilon_t, \dots, \varepsilon_0, Z) \\ &= \sum_{j=h}^{t+j} b_j \varepsilon_{t+h-j} + \tilde{H}(t+h)' Z \end{aligned}$$

L'erreur de prévision est donnée par :

$$\begin{aligned} e_t(h) &= X_{t+h} - {}_t X_{t+h}^* \\ &= Y_{t+h} - {}_t Y_{t+h}^* \\ &= \sum_{j=0}^{h-1} b_j \varepsilon_{t+h-j} \end{aligned}$$

Si les ε_t sont iid $\rightsquigarrow \mathcal{N}(0, \sigma^2)$, alors

$$e_t(h) \rightsquigarrow \mathcal{N} \left(0, \sigma^2 \sum_{j=0}^{h-1} b_j^2 \right)$$

Soient $\hat{\sigma}^2$ et \hat{b}_j des estimateurs convergents de σ^2 et b_j . On en déduit que :

$$\frac{e_t(h)}{\hat{\sigma}\sqrt{\sum_{j=0}^{h-1}\hat{b}_j^2}} \rightsquigarrow \mathcal{N}(0, 1)$$

On en déduit un intervalle de prévision au niveau $1 - \alpha$ pour X_{t+h} :

$$\left[{}_tX_{t+h}^* - u_{1-\frac{\alpha}{2}}\hat{\sigma}\sqrt{\sum_{j=0}^{h-1}\hat{b}_j^2} ; {}_tX_{t+h}^* + u_{1-\frac{\alpha}{2}}\hat{\sigma}\sqrt{\sum_{j=0}^{h-1}\hat{b}_j^2} \right]$$

Chapitre 5

Processus vectoriels stationnaires - Processus VAR stationnaires

Introduction

Définition 5.0.1 (Processus vectoriel) (X_t) est un processus à valeurs dans \mathbb{R}^n si

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

avec $(x_{i,t})$ processus à valeurs dans \mathbb{R} .

Pour une étude complète, il faudrait étudier :

- VAR : modèle parlant car on explique à partir des x_{it} passés,
- VMA : moins parlant car on explique à partir des ε_{it} passés,
- VARMA.

Les modèles VAR ont germé avec l'économétrie avec SIMS en 1980 suite à la critique de LUCAS. On teste des modèles structurels, l'exogénéité.

5.1 Processus vectoriels stationnaires du second ordre

5.1.1 Définition et proposition

Définition 5.1.1 (Processus vectoriel du second ordre) (X_t) est un processus du second ordre si et seulement si

$$\begin{aligned} \Leftrightarrow & \quad \forall t, X_t \in L^2_{\mathbb{R}^n}(\Omega, \mathcal{A}, \mathbb{P}) \\ \Leftrightarrow & \quad \forall i, \forall t, x_{it} \in L^2(\Omega, \mathcal{A}, \mathbb{P}) \\ \Leftrightarrow & \quad X'_t X_t = \|X_t\|_2^2 = \sum_{i=1}^n x_{it}^2 \in L^1(\Omega, \mathcal{A}, \mathbb{P}) \end{aligned}$$

$$\text{On a } \mathbb{E}X_t = \begin{pmatrix} \mathbb{E}x_{1t} \\ \vdots \\ \mathbb{E}x_{nt} \end{pmatrix} = m \text{ et } \mathbb{V}X_t = \mathbb{E}(X_t - m)(X_t - m)' = (\text{Cov}(x_{it}, x_{jt}))_{1 \leq i, j \leq n}.$$

Définition 5.1.2 (Processus vectoriel stationnaire au second ordre) Soit (X_t) est un processus du second ordre.

On dit que (X_t) est **stationnaire au second ordre** si et seulement si :

- (i) $\mathbb{E}X_t = m$
- (ii) $\mathbb{V}X_t = \Sigma = \Gamma_X(0)$
- (iii) $\mathbb{E}(X_t - m)(X_{t+h} - m)' = \Gamma_X(h)$

Remarque 5.1.1 (1) (iii) \Rightarrow (ii)

(2) (iii) $\Rightarrow \text{Cov}(x_{it}, x_{j,t-h}) = \gamma_{ij}(h), \forall i, j, h$

(3) (X_t) stationnaire $\Rightarrow (x_{it})$ stationnaire $\forall i$

Mais la réciproque est fautive en général car les conditions :

$$\mathbb{E}X_t = m \Leftrightarrow \mathbb{E}x_{it} = m_i, \forall i$$

$$\text{Cov}(x_{it}, x_{i,t-h}) = \gamma_i(h)$$

ne suffisent pas pour avoir la stationnarité de (X_t) .

Exemple 5.1.1 (1) $\varepsilon_t \sim \mathcal{BB}(0, \Omega)$ où $\varepsilon_t = \begin{pmatrix} \varepsilon_{1,t} \\ \vdots \\ \varepsilon_{n,t} \end{pmatrix}$

Stationnarité?

- (i) $\mathbb{E}X_t = 0$
- (ii) $\mathbb{V}X_t = \Omega$
- (iii) $\mathbb{E}\varepsilon_t \varepsilon_\tau' = 0$ si $t \neq \tau$

\rightarrow OK

(2) $X_t \sim VMA(1)$

Soit $\varepsilon_t \sim \mathcal{BB}(0, \Omega)$, $A \in \mathcal{M}_n(\mathbb{R})$ fixée.

Alors $X_t = \varepsilon_t - A\varepsilon_{t-1}$ est stationnaire :

- (i) $\mathbb{E}X_t = 0$
- (ii)

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

(iii) $\mathbb{E}X_t X_{t-1}' = \mathbb{E}(\varepsilon_t - A\varepsilon_{t-1})(\varepsilon_{t-1} - A\varepsilon_{t-2})'$

Pour $h > 1$, $\mathbb{E}X_t X_{t-h}' = 0$

$\mathbb{E}X_t X_{t+1}' = -\Omega A'$

$\mathbb{E}X_t X_{t+h}' = 0$, si $h > 1$

Proposition 5.1.1 X_t stationnaire $\Rightarrow \forall h \in \mathbb{Z}, \Gamma(-h) = \Gamma(h)'$

Démonstration

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

Proposition 5.1.2 Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire.

Soit $(A_j)_{j \in \mathbb{Z}}$ une suite de matrices telle que $\sum_{j \in \mathbb{Z}} \|A_j\| < +\infty$.

Alors

(i) $Y_t = \sum_j A_j X_{t-j} \in L^2_{\mathbb{R}^n}(\Omega, \mathcal{A}, \mathbb{P})$

(ii) Y_t est stationnaire avec :

$$\begin{aligned}
- \mathbb{E}Y_t &= m_Y = \left(\sum_j A_j\right) m_X \\
- \Gamma_Y(h) &= \sum_{j,k} A_j \Gamma_X(h+k-j) A'_k
\end{aligned}$$

Démonstration

Dans le cas général

$$\begin{aligned}
\Gamma_Y(h) &= \mathbb{E}(Y_t Y'_{t-h}) \\
&= \mathbb{E} \left[\left(\sum_j A_j X_{t-j} \right) \left(\sum_k A_k X_{t-h-k} \right)' \right] \\
&= \sum_{j,k} \mathbb{E}(A_j X_{t-j} X'_{t-h-k} A'_k) \\
&= \sum_{j,k} \mathbb{E}(A_j \Gamma_X(h+k-j) A'_k)
\end{aligned}$$

Cas particulier Si $X_t = \varepsilon_t \sim \mathcal{BB}$, alors

$$Y_t = \sum_{j \in \mathbb{Z}} A_j \varepsilon_{t-j}$$

C'est un processus $VMA(\infty)$.

5.1.2 Densité spectrale d'un processus vectoriel stationnaire

Définition 5.1.3 (Densité spectrale) Si $X_t = \sum_{k \in \mathbb{Z}} A_k \varepsilon_{t-k}$ où $\varepsilon_t \sim \mathcal{BB}$ et $\sum_{k \in \mathbb{Z}} \|A_k\| < +\infty$

alors

(i) $\sum_{h \in \mathbb{Z}} \|\Gamma_X(h)\| < +\infty$

(ii) $S_X(\omega) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \Gamma_X(h) e^{i\omega h}$ est la **matrice de densité spectrale** de (X_t)

Démonstration

$$\begin{aligned}
\Gamma_X(h) &= \mathbb{E}X_t X'_{t-h} \text{ car } \mathbb{E}X_t = 0 \\
&= \mathbb{E} \left[\left(\sum_j A_j \varepsilon_{t-j} \right) \left(\sum_k A_k \varepsilon_{t-k-h} \right)' \right] \\
&= \sum_{j,k} A_k \mathbb{E} \underbrace{(\varepsilon_{t-j} \varepsilon_{t-k-h})}_{= \begin{cases} 0 & \text{si } j \neq h+k \\ \Omega & \text{sinon} \end{cases}} A'_j \\
&= \sum_j A_j \Omega A'_j
\end{aligned}$$

On majore ensuite

$$\begin{aligned}
\sum_{h \in \mathbb{Z}} \|\Gamma_X(h)\| &\leq \sum_h \sum_k \|A_k \Omega A'_{k-h}\| \\
&\leq \sum_{k,h} \|A_k\| \|\Omega\| \|A'_{k-h}\| \\
&\leq \|\Omega\| \left(\sum_k \|A_k\| \right)^2 < +\infty
\end{aligned}$$

Donc $S_X(\omega)$ a un sens.

Proposition 5.1.3 (i) Si $\varepsilon \sim \mathcal{BB}(0, \Omega)$ alors

$$S_\varepsilon(\omega) = \frac{1}{2\pi} \Omega$$

(ii) Soit $X_t = \sum_k A_k \varepsilon_{t-k} = A(L)\varepsilon_t$ tel que $\sum_k \|A_k\| < +\infty$
et $Y_t = \sum_k B_k X_{t-k} = B(L)X_t$ tel que $\sum_k \|B_k\| < +\infty$
alors $Y_t = \sum_k C_k \varepsilon_{t-k} = C(L)\varepsilon_t$
et $S_Y(\omega) = B(e^{i\omega})S_X(\omega)B(e^{-i\omega})'$

Démonstration

- (i) $C(L) = B(L).A(L) = (B.A)(L)$
(ii) $S_Y(\omega)$: cf. le cas réel.

Théorème 5.1.1 (Injectivité) On considère le modèle $X_t = A(L)\varepsilon_t$ où $\varepsilon_t \sim \mathcal{BB}$.

Alors

$$\forall h \in \mathbb{Z}, \Gamma_X(h) = \int_{-\pi}^{\pi} S_X(\omega) e^{-i\omega h} d\omega$$

Démonstration

$$\begin{aligned}
\int_{-\pi}^{\pi} S_X(\omega) e^{-i\omega h} d\omega &= \int_{-\pi}^{\pi} \left(\sum_k \Gamma_x(k) e^{i\omega k} \right) e^{-i\omega h} d\omega \\
&= \sum_k \Gamma_X(k) \underbrace{\int_{-\pi}^{\pi} e^{-i\omega(k-h)} d\omega}_{\delta_{kh} \text{ (symbole de Kronecker)}} \\
&= \Gamma_X(h)
\end{aligned}$$

5.1.3 Innovation d'un processus vectoriel

$$\text{Soit } X_t = \begin{pmatrix} x_{1t} \\ \vdots \\ x_{nt} \end{pmatrix}$$

$$\text{Définition 5.1.4 (Prévision)} \quad X_t^* = EL(X_t | X_{t-1}) = \begin{pmatrix} EL(x_{1t} | X_{t-1}) \\ \vdots \\ EL(x_{nt} | X_{t-1}) \end{pmatrix}$$

où $\forall j$

$$\begin{aligned}
EL(x_{jt} | X_{t-1}) &= EL(x_{jt} | x_{1,t-1}, \dots, x_{n,t-1}) \\
&= EL(x_{jt} | \{x_{1s}, \dots, x_{ns}, s \leq t-1\}) \\
&= EL(x_{jt} | \{x_{is}, i = 1..n, s \leq t-1\})
\end{aligned}$$

Proposition 5.1.4 Soit X_t^* la **prévision linéaire optimale** de X_t à la date $t-1$.
Si \hat{X}_t est une autre prévision linéaire de X_t à la date $t-1$, alors

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

au sens des matrices symétriques positives

On a égalité si et seulement si $\hat{X}_t = X_t^*$.

Remarque 5.1.2 (1) \hat{X}_t est une prévision linéaire de X_t en $t-1$, i.e. \hat{X}_t est une fonction linéaire de x_{js} , $s \leq t-1$, $j = 1..n$.

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

$$\Rightarrow \forall u \in \mathbb{R}^n - \{0\}, u' \mathbb{V}(X_t - X_t^*) u < u' \mathbb{V}(X_t - \hat{X}_t) u$$

En particulier pour $u = (0, \dots, 0, \underbrace{1}_{i\text{ème position}}, 0, \dots, 0)$, on a :

$$\mathbb{V}(x_{it} - x_{it}^*) \leq \mathbb{V}(x_{it} - \hat{x}_{it})$$

x_{it}^* est la prévision optimale de x_{it} comme fonction linéaire des x_{js} , $s \leq t-1$.

- (2) $\forall u \in \mathbb{R}^n - \{0\}, \mathbb{V}u'(X_t - X_t^*) < \mathbb{V}u'(X_t - \hat{X}_t)$
 $\Rightarrow u'X_t^*$ est une prévision de $u'X_t$ plus précise que $u'\hat{X}_t$.
- (3) Si on note

$$\begin{aligned}\bar{\mathcal{L}}(\underline{X}_{t-1}) &= \bar{\mathcal{L}}(x_{1,t-1}, \dots, x_{n,t-1}) \\ &= \bar{\mathcal{L}}(\{x_{1s}, \dots, x_{ns}, s \leq t-1\})\end{aligned}$$

On a $\bar{\mathcal{L}}(x_{i,t-1}) \subsetneq \bar{\mathcal{L}}(\underline{X}_{t-1})$

$$\tilde{x}_{it} = EL(x_{it}|x_{t-1})$$

$$\mathbb{V}(x_{it} - \tilde{x}_{it}) > \mathbb{V}(x_{it} - x_{it}^*)$$

C'est normal car on a plus d'informations dans $\bar{\mathcal{L}}(\underline{X}_{t-1})$ que dans $\bar{\mathcal{L}}(x_{i,t-1})$.

Proposition 5.1.5 (Processus des innovations) Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire et $X_t^* = EL(X_t|x_{t-1})$.

L'innovation de X_t est

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

On a plus précisément

$$\begin{aligned}\varepsilon_t &= \begin{pmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{nt} \end{pmatrix} \\ &= \begin{pmatrix} x_{1t} - EL(x_{1t}|\underline{X}_{t-1}) \\ \vdots \\ x_{nt} - EL(x_{nt}|\underline{X}_{t-1}) \end{pmatrix} \\ &= \begin{pmatrix} x_{1t} - EL(x_{1t}|\{x_{1s}, \dots, x_{ns}, s \leq t-1\}) \\ \vdots \\ x_{nt} - EL(x_{nt}|\{x_{1s}, \dots, x_{ns}, s \leq t-1\}) \end{pmatrix}\end{aligned}$$

On montre que $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc.

Remarque 5.1.3 (1) ce qui a été présenté est valable dans le cas où $\mathbb{E}X_t = 0$.

Quand $\mathbb{E}X_t \neq 0$, on définit

$$\bar{\mathcal{L}}(\underline{X}_{t-1}) = \bar{\mathcal{L}}(1, \underline{x}_{1,t-1}, \dots, \underline{x}_{n,t-1})$$

Puis

$$X_t^* = \begin{pmatrix} EL(x_{1t}|1, \underline{x}_{1,t-1}, \dots, \underline{x}_{n,t-1}) \\ \vdots \\ EL(x_{nt}|1, \underline{x}_{1,t-1}, \dots, \underline{x}_{n,t-1}) \end{pmatrix}$$

- (2) $\varepsilon_t = \begin{pmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{nt} \end{pmatrix}$ où $\varepsilon_{it} \sim \mathcal{BB}, \forall i$

Mais ε_{it} n'est pas l'innovation du processus univarié (x_{it}) . En effet

$$x_{it} - EL(x_{it}|1, \underline{x}_{1,t-1}, \dots, \underline{x}_{n,t-1}) \neq x_{it} - EL(x_{it}|\underline{x}_{1,t-1})$$

Démonstration $\mathbb{E}\varepsilon_t = \begin{pmatrix} \mathbb{E}\varepsilon_{1t} \\ \vdots \\ \mathbb{E}\varepsilon_{nt} \end{pmatrix}$ où $\mathbb{E}\varepsilon_{it} = \mathbb{E}(x_{it} - EL(x_{it}|\underline{X}_{t-1}))$

– Si $\forall j, \mathbb{E}x_{jt} = 0$, alors

$$\mathbb{E}(EL(x_{it}|\underline{X}_{t-1})) = \mathbb{E}(EL(x_{it}|x_{1,t-1}, \dots, x_{n,t-1}))$$

– Si $\mathbb{E}X_t \neq 0$, alors

$$\mathbb{E}(EL(x_{it}|\underline{X}_{t-1})) = \mathbb{E}(EL(x_{it}|1, x_{1,t-1}, \dots, x_{n,t-1}))$$

$$\begin{aligned} \forall i, x_{it} - EL(x_{it}|\underline{X}_{t-1}) \perp 1 &\Leftrightarrow \mathbb{E}x_{it} - \mathbb{E}(EL(x_{it}|\underline{X}_{t-1})) = 0 \\ &\Leftrightarrow \mathbb{E}\varepsilon_{it} = 0 \end{aligned}$$

On peut toujours se ramener à $\mathbb{E}X_t = 0$.

On a $\mathbb{E}(\varepsilon_t \varepsilon'_t) = (\mathbb{E}(\varepsilon_{it} \varepsilon_{jt}))_{1 \leq i, j \leq n}$ avec

$$\begin{aligned} \varepsilon_{it} \varepsilon_{jt} &= (x_{it} - \underbrace{EL(x_{it}|\underline{X}_{t-1})}_{\in \bar{\mathcal{L}}(\underline{X}_{t-1})})(x_{jt} - EL(x_{jt}|\underline{X}_{t-1})) \\ \mathbb{E}(\varepsilon_{it} \varepsilon_{jt}) &= \mathbb{E}[x_{it}(x_{jt} - EL(x_{jt}|\underline{X}_{t-1}))] \\ &= \mathbb{E}(x_{it}x_{jt}) - \mathbb{E}(x_{it}EL(x_{jt}|\underline{X}_{t-1})) \end{aligned}$$

Si $s < t$, alors $\mathbb{E}(\varepsilon_t \varepsilon'_s) = (\mathbb{E}(\varepsilon_{it} \varepsilon_{js}))_{1 \leq i, j \leq n}$ où

$$\begin{aligned} \varepsilon_{it} &= x_{it} - EL(x_{it}|\underline{X}_{t-1}) \perp \bar{\mathcal{L}}(\underline{X}_{t-1}) \\ \varepsilon_{js} &= x_{js} - EL(x_{js}|\underline{X}_{t-1}) \in \bar{\mathcal{L}}(\underline{X}_s) \subset \bar{\mathcal{L}}(\underline{X}_{t-1}) \end{aligned}$$

Donc $\mathbb{E}(\varepsilon_{it} \varepsilon_{js}) = 0$ si $s < t$.

Théorème 5.1.2 (Théorème de Wold) Soit (X_t) un processus vectoriel stationnaire et (ε_t) le processus des innovations.

On a alors $\exists (A_k)_{k \in \mathbb{Z}}$ tel que

$$X_t = m_t + \sum_{\mathbb{Z}} A_k \varepsilon_{t-k}$$

avec $A_0 = I$ et $\sum_{\mathbb{Z}} \|A_k\| < +\infty$

5.1.4 Convergence des moments empiriques

Soit $X_t = \begin{pmatrix} x_{1t} \\ \vdots \\ x_{nt} \end{pmatrix}$ et $\mathbb{E}X_t = \begin{pmatrix} m_1 \\ \vdots \\ m_n \end{pmatrix}$.

On définit les estimateurs :

$$\begin{aligned}\hat{m} &= \frac{1}{T} \sum_{t=1}^T X_t = \bar{X}_T = \begin{pmatrix} \bar{x}_{1t} \\ \vdots \\ \bar{x}_{nt} \end{pmatrix} \\ \hat{\Gamma}(0) &= \frac{1}{T \text{ ou } T-1} \sum_{t=1}^T (X_t - \bar{X}_T)(X_t - \bar{X}_T)' \\ \hat{\Gamma}(h) &= \frac{1}{T-h \text{ ou } T-h-1} \sum_{t=1}^T (X_t - \bar{X}_T)(X_{t-h} - \bar{X}_T)'\end{aligned}$$

Proposition 5.1.6 (i) *Convergence* : \hat{m} et $\hat{\Gamma}(h)$ sont des estimateurs convergents respectivement de m et $\Gamma(h)$.

(ii) *Normalité asymptotique* :

$$\sqrt{T}(\hat{m} - m) \xrightarrow{loi} \mathcal{N}\left(0, \sum_{h \in \mathbb{Z}} \Gamma_X(h)\right)$$

On note $\hat{\Gamma}(h) = (\hat{\gamma}_{ij}(h))_{1 \leq i, j \leq n}$. On a la normalité jointe de toute famille finie de $\hat{\gamma}_{ij}(h)$, $\forall i, j, h$.

Exemple 5.1.2

$$\sqrt{T} \begin{pmatrix} \hat{\gamma}_{11}(0) - \gamma_{11}(0) \\ \vdots \\ \hat{\gamma}_{nn}(0) - \gamma_{nn}(0) \\ \hat{\gamma}_{11}(1) - \gamma_{11}(1) \\ \vdots \\ \hat{\gamma}_{nn}(1) - \gamma_{nn}(1) \end{pmatrix} \left\{ \begin{array}{l} \frac{n(n+1)}{2} \text{ termes de } \hat{\Gamma}(0) - \Gamma(0) \\ n^2 \text{ termes de } \hat{\Gamma}(1) - \Gamma(1) \end{array} \right. \xrightarrow{loi} \mathcal{N}(0, *)$$

5.2 Processus VAR stationnaires

5.2.1 Définition et proposition générale

Définition 5.2.1 Soit $X_t \rightsquigarrow VAR(p)$, $\mu \in \mathbb{R}^n$, $(\phi_j)_{1 \leq j \leq p} \in (\mathcal{M}_n(\mathbb{R}))^p$, $\phi_p \neq 0$, $\varepsilon_t \rightsquigarrow \mathcal{BB}(0, \Omega)$, $\Omega \in \mathcal{M}_n(\mathbb{R})$ tel que

$$X_t = \mu + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t \Leftrightarrow \phi(L)X_t = \mu + \varepsilon_t$$

où $\phi(L) = I_n - \phi_1 L - \dots - \phi_p L^p$

Remarque 5.2.1 $\phi_k = \left(\varphi_{ij}^k\right)_{1 \leq i, j \leq n}$

$$\forall i, x_{it} = \mu_i + \sum_{k=1}^p \sum_{j=1}^n \left(\varphi_{ij}^k x_{j,t-k}\right) + \varepsilon_{it}$$

Proposition 5.2.1 *Si $\det \phi(Z)$ a toutes ses racines de module strictement plus grand que 1, alors*

- (i) $\phi(L)$ est inversible
- (ii) $\phi(L)^{-1} = \sum_{k=0}^{\infty} A_k L^k$ tel que $A_0 = I$ et $\sum \|A_k\| < +\infty$

Démonstration $\phi(L) = I_n - \phi_1 L - \dots - \phi_p L^p = (\varphi_{ij}(L))_{1 \leq i, j \leq n}$ avec $\varphi_{ij}(L) = \delta_{ij} - \varphi_{ij}^1 L - \dots - \varphi_{ij}^p L^p$

On note $\phi(Z) = I - \phi_1 Z - \phi_p Z^p = (\varphi_{ij}(Z))_{1 \leq i, j \leq n}$, $\forall Z \in \mathbb{C}$ avec $\varphi_{ij}(Z) = \delta_{ij} - \varphi_{ij}^1 Z - \dots - \varphi_{ij}^p Z^p$.

On note $\tilde{\phi}(L)$ la comatrice de $\phi(L)$.

On sait que

$$\begin{aligned} \phi(Z)\tilde{\phi}(Z) &= \tilde{\phi}(Z)\phi(Z) = \det \phi(Z) I_n \\ \phi(L)\tilde{\phi}(L) &= \tilde{\phi}(L)\phi(L) = \det \phi(L) I_n \end{aligned}$$

On aura donc $\phi(L)$ inversible $\Leftrightarrow \det \phi(L)$ inversible.

On sait que si les racines de $\det \phi(Z)$ sont de module strictement plus grand que 1, alors $\det \phi(L)$. Dans ce cas, $\det \phi(L)$ a pour inverse

$$(\det \phi(L))^{-1} = \sum_0^{\infty} a_k L^k$$

tel que $a_0 = 1$ et $\sum |a_k| < +\infty$

On a donc

$$\begin{aligned} \phi(L)\tilde{\phi}(L) &= \det \phi(L) I_n \\ \Rightarrow \phi(L) \underbrace{\tilde{\phi}(L)(\det \phi(L))^{-1}}_{\phi(L)^{-1}} &= I_n \end{aligned}$$

avec $\tilde{\phi}(L) = (\tilde{\varphi}_{ij}(L))_{1 \leq i, j \leq n}$ et $d^o \tilde{\varphi}_{ij} \leq p^{n-1}$

$$\phi(L)^{-1} = \left(\sum_0^{\infty} a_k L^k \right) \tilde{\phi}(L) = \sum_0^{\infty} A_k L^k$$

avec $A_0 = a_0 \tilde{\phi}(1) I_n$ et $\sum_k \|A_k\| < +\infty$ car $\sum_k |a_k| < +\infty$

Remarque 5.2.2 Si $\det \phi(1) = 0$ alors (X_t) est non stationnaire.

$$\begin{aligned} \phi(L)X_t &= \mu + \varepsilon_t \\ X_t &= \phi(L)^{-1}(\mu + \varepsilon_t) \\ X_t &= \tilde{\phi}(L)(\det \phi(L))^{-1}(\mu + \varepsilon_t) \\ \det \phi(L) X_t &= \tilde{\phi}(L)(\mu + \varepsilon_t) \\ (1 - L)\psi(L)X_t &= \tilde{\phi}(L)(\mu + \varepsilon_t) \end{aligned}$$

Proposition 5.2.2 *Si $\phi(L)X_t = \mu + \varepsilon_t$ et si les racines de $\det \phi(Z)$ sont de module strictement supérieur à 1, alors*

- (i) $X_t = m + \sum_0^{\infty} A_k \varepsilon_{t-k}$ avec $A_0 = I$ et $\sum \|A_k\| < +\infty$
- (ii) (ε_t) est le processus des innovations de (X_t)

Démonstration

(i) Soit le modèle $\phi(L)X_t = \mu + \varepsilon_t$ avec $\phi(L)$ inversible.

L'inverse de $\phi(L)$ est donné par

$$\phi(L)^{-1} = \det \phi(L)^{-1} \tilde{\phi}(L) = \sum_0^{\infty} A_l L^k$$

où $A_0 = I_n$ et $\sum \|A_k\| < +\infty$

On a

$$\begin{aligned} X_t &= \phi(L)^{-1}(\mu + \varepsilon_t) \\ &= \phi(L)^{-1}\mu + \phi(L)^{-1}\varepsilon_t \\ &= m + \sum_0^{\infty} A_k \varepsilon_{t-k} \end{aligned}$$

(ii)

$$\begin{aligned} \phi(L)X_t &= \mu + \varepsilon_t \\ \Leftrightarrow X_t &= \mu + \sum_1^p \phi_k X_{t-k} + \varepsilon_t \\ \Leftrightarrow x_{it} &= \mu_i + \sum_{k=1}^p \left(\sum_{j=1}^n \varphi_{ij}^k x_{j,t-k} \right) + \varepsilon_{it} \end{aligned}$$

D'où $\varepsilon_{it} \in \mathcal{L}(x_{it}, 1, x_{1,t-1}, \dots, x_{1,t-p}, \dots, x_{n,t-p})$

Puis

$$\bar{\mathcal{L}}(1, \underline{\varepsilon}_t) = \bar{\mathcal{L}}(1, \{\varepsilon_{is}, i = 1..n, s \leq t\}) \subset \bar{\mathcal{L}}(\underline{X}_t)$$

$$\forall i, x_{it} = \mu_i + \sum_{k=0}^{\infty} \left(\sum_{j=1}^n a_{ij}^k \varepsilon_{j,t-k} \right) \in \bar{\mathcal{L}}(1, \{\varepsilon_{is}, i = 1..n, s \leq t\}) = \bar{\mathcal{L}}(1, \underline{\varepsilon}_t)$$

$$\Rightarrow \bar{\mathcal{L}}(1, \underline{X}_t) \subset \bar{\mathcal{L}}(1, \underline{\varepsilon}_t)$$

$$\Rightarrow \bar{\mathcal{L}}(1, \underline{X}_t) = \bar{\mathcal{L}}(1, \underline{\varepsilon}_t)$$

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

$$\begin{aligned}
X_t^* &= EL(X_t | 1, \underline{X}_{t-1}) \\
&= EL\left(\mu + \sum_1^p \phi_k X_{t-k} + \varepsilon_t | \underline{X}_{t-1}\right) \\
&= \begin{pmatrix} EL\left(\mu_1 + \sum_{k=1}^p \left(\sum_{j=1}^n \varphi_{1j}^k x_{j,t-k}\right) \varepsilon_{1t} | \underline{X}_{t-1}\right) \\ \vdots \\ EL\left(\mu_n + \sum_{k=1}^p \left(\sum_{j=1}^n \varphi_{nj}^k x_{j,t-k}\right) \varepsilon_{nt} | \underline{X}_{t-1}\right) \end{pmatrix} \\
&= \underbrace{\mu + \sum_1^p \phi_k X_{t-k}}_{X_t - \varepsilon_t} + \underbrace{EL(\varepsilon_t | 1, \underline{X}_{t-1})}_{\substack{EL(\varepsilon_t | 1, \varepsilon_{t-1}) \\ 0 \text{ car } \varepsilon_t \sim \mathcal{BB}}}
\end{aligned}$$

CQFD

Remarque 5.2.3 (1) $X_t = m + \sum_0^\infty A_k \varepsilon_{t-k} = m + \phi(L)^{-1} \varepsilon_t$ d'après le théorème de WOLD avec $m = \mathbb{E}X_t = \phi(L)^{-1} \mu$.

(2) Soit le modèle $\phi(L)X_t = \mu + \varepsilon_t$.

- Si $\det \phi(Z)$ a ses racines de module strictement plus grand que 1, alors $X_t = m + \phi(L)^{-1} \varepsilon_t$. On en déduit que $(X_t)_\mathbb{Z}$ est stationnaire (représentation $VMA(\infty)$).
- Si $\det \phi(1) = 0$, alors (X_t) ne peut pas être stationnaire car $\det \phi(Z) = (1 - Z)^d \psi(Z)$ tel que les racines de ψ sont de module strictement plus grand que 1.

$$(1 - L)^d X_t = \psi(L)^{-1} \tilde{\phi}(L) (\mu + \varepsilon_t)$$

- Si $\det \phi(Z)$ a toutes ses racines de module strictement supérieur à 1, alors
 - il existe une solution (X_t) stationnaire,
 - il existe aussi des solutions non stationnaires en espérance : $Z_t = X_t + Y_t$ où Y_t est déterministe tel que $\phi(L)Y_t$.

5.2.2 Prévision dans un VAR stationnaire

A la date t , on souhaite effectuer une prévision de X_{t+h} et déterminer une région de confiance de la prévision de X_{t+h} .

On considère le modèle

$$\phi(L)X_t = \mu + \varepsilon_t$$

avec $\det \phi(Z)$ ayant toutes ses racines de module strictement supérieur à 1.

On s'intéresse à la solution stationnaire (X_t) .

Prévision

$$\begin{aligned}
{}_t X_{t+1}^* &= EL(X_{t+1} | 1, \underline{X}_t) \\
&\vdots \\
{}_t X_{t+h}^* &= EL(X_{t+h} | 1, \underline{X}_t)
\end{aligned}$$

Plus précisément on a

$$\begin{aligned}
{}_t X_{t+1}^* &= \mu + \phi_1 X_t + \cdots + \phi_p X_{t+1+p} + \underbrace{EL(\varepsilon_{t+1} | X_t)}_0 \\
X_{t+2} &= \mu + \phi_1 X_{t+1} + \cdots + \phi_p X_{t+2+p} + \varepsilon_{t+2} \\
{}_t X_{t+2}^* &= \mu + \phi_1 {}_t X_{t+1}^* + \cdots + \phi_p X_{t+2+p} + \underbrace{EL(\varepsilon_{t+2} | X_t)}_0 \\
&\vdots \\
X_{t+h} &= \mu + \sum_{k=1}^p \phi_k X_{t+h-k} + \varepsilon_{t+h} \\
{}_t X_{t+h}^* &= \mu + \sum_{k=1}^p \phi_k {}_t X_{t+h-k}^* + 0
\end{aligned}$$

avec ${}_t X_{t+h-k}^* = X_{t+h-k}$ si $k \geq h$.

C'est la méthode itérative de calcul de la prévision de X_{t+h} .

Région de confiance sous hypothèse de normalité

- Si $h = 1$, $X_{t+1} - {}_t X_{t+1}^* = \varepsilon_{t+1} \rightsquigarrow \mathcal{N}(0, \Omega)$ car les ε_t sont iid et $\rightsquigarrow \mathcal{N}(0, \Omega)$.

La région de confiance de \mathbb{R}^n dans laquelle X_{t+1} a une probabilité $1 - \alpha$ de se trouver est

$$\{(X_{t+1} - {}_t X_{t+1}^*)' \Omega^{-1} (X_{t+1} - {}_t X_{t+1}^*) \leq q_{1-\alpha}(\chi^2(n))\}$$

C'est un ellipsoïde en dimension n .

- Si $h = 2$, $X_{t+2} - {}_t X_{t+2}^* = \phi_1 (X_{t+1} - {}_t X_{t+1}^*) \varepsilon_{t+2}$

$$\mathbb{V}(X_{t+2} - {}_t X_{t+2}^*) = \phi_1 \Omega \phi_1' + \Omega$$

D'où

$$(X_{t+2} - {}_t X_{t+2}^*)' (\phi_1 \Omega \phi_1' + \Omega)^{-1} (X_{t+2} - {}_t X_{t+2}^*) \rightsquigarrow \chi^2(n)$$

On peut en déduire comme précédemment une région de confiance au niveau $1 - \alpha$.

5.3 Estimation d'un modèle VAR sous hypothèse de normalité

5.3.1 Ecriture empilée du modèle

Exemple 5.3.1 (Cas $n = 2$, $p = 2$)

$$\begin{aligned}
X_t &= \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} \\
&= \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} \begin{pmatrix} x_{1,t-1} \\ x_{2,t-1} \end{pmatrix} + \begin{pmatrix} c_1 & d_1 \\ c_2 & d_2 \end{pmatrix} \begin{pmatrix} x_{1,t-2} \\ x_{2,t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \\
&= \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t
\end{aligned}$$

Modèle empilé avec T observations :

$$\begin{pmatrix} x_{13} \\ \vdots \\ x_{1T} \\ x_{23} \\ \vdots \\ x_{2T} \end{pmatrix} = \left(\begin{array}{ccccc|ccccc} 1 & x_{12} & x_{22} & x_{11} & x_{21} & 0 & \dots & & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & & \vdots \\ 1 & x_{1,T-1} & x_{2,T-1} & x_{1,T-2} & x_{2,T-2} & 0 & \dots & & 0 \\ \hline 0 & & \dots & & 0 & 1 & x_{12} & x_{22} & x_{11} & x_{21} \\ \vdots & & \ddots & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & & \dots & & 0 & 1 & x_{1,T-1} & x_{2,T-1} & x_{1,T-2} & x_{2,T-2} \end{array} \right) \begin{pmatrix} \mu_1 \\ a_1 \\ b_1 \\ c_1 \\ d_1 \\ \mu_2 \\ a_2 \\ b_2 \\ c_2 \\ d_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{13} \\ \vdots \\ \varepsilon_{1T} \\ \varepsilon_{23} \\ \vdots \\ \varepsilon_{2T} \end{pmatrix}$$

$$\Leftrightarrow \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} Z & 0 \\ 0 & Z \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} Z & 0 \\ 0 & Z \end{pmatrix} \begin{pmatrix} \varepsilon^1 \\ \varepsilon^2 \end{pmatrix}$$

On a donc le modèle (M) $(I_2 \otimes Z)\beta + \varepsilon$ où \otimes désigne le symbole de KRONECKER. (M) se décompose en les 2 sous-modèles :

$$\begin{aligned} (M_1) \quad x_1 &= Z\beta_1 + \varepsilon^1 \\ (M_2) \quad x_2 &= Z\beta_2 + \varepsilon^2 \end{aligned}$$

On veut calculer $\mathbb{V}\varepsilon$, en supposant que $\varepsilon_t \sim \mathcal{N}(0, \Omega)$ où $\Omega = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$ avec $\sigma_{12} = \sigma_{21}$

$$\mathbb{V} \begin{pmatrix} \varepsilon_{13} \\ \vdots \\ \varepsilon_{1T} \\ \varepsilon_{23} \\ \vdots \\ \varepsilon_{2T} \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \dots & 0 & \sigma_{12} & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{11} & 0 & \dots & \sigma_{12} \\ \hline \sigma_{21} & \dots & 0 & \sigma_{22} & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{21} & 0 & \dots & \sigma_{22} \end{pmatrix} = \begin{pmatrix} \sigma_{11}I_{T-2} & \sigma_{12}I_{T-2} \\ \sigma_{11}I_{T-2} & \sigma_{12}I_{T-2} \end{pmatrix} = \Omega \otimes I_{T-2}$$

Cas général On étudie le modèle : $X_t = \mu + \sum_{k=1}^p \phi_k X_{t-k} + \varepsilon_t$ où $\varepsilon_t \sim \mathcal{BB}(0, \Omega)$

$$\Leftrightarrow \begin{pmatrix} x_{1t} \\ \vdots \\ x_{nt} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} + \sum_{k=1}^p \begin{pmatrix} \varphi_{11}^k & \dots & \varphi_{1n}^k \\ \vdots & \ddots & \vdots \\ \varphi_{n1}^k & \dots & \varphi_{nn}^k \end{pmatrix} \begin{pmatrix} x_{1,t-k} \\ \vdots \\ x_{n,t-k} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{nt} \end{pmatrix}$$

Pour le modèle empilé avec T observations :

$$\begin{pmatrix} x_{1,p+1} \\ \vdots \\ x_{1,T} \\ \vdots \\ x_{n,p+1} \\ \vdots \\ x_{n,T} \end{pmatrix} = \begin{pmatrix} & & & 0 & \dots & 0 & & 0 & \dots & 0 \\ & & & \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots \\ & & & 0 & \dots & 0 & & 0 & \dots & 0 \\ \hline & & Z & & & & & \vdots & \ddots & \vdots \\ & & & & & Z & & \vdots & \ddots & \vdots \\ & & & 0 & \dots & 0 & & 0 & \dots & 0 \\ \hline & & & 0 & \dots & 0 & & 0 & \dots & 0 \\ & & & \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots \\ & & & 0 & \dots & 0 & & & & Z \end{pmatrix} \begin{pmatrix} \mu_1 \\ \vdots \\ \phi_{1,n}^p \\ \vdots \\ \mu_n \\ \vdots \\ \phi_{n,n}^p \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,p+1} \\ \vdots \\ \varepsilon_{1,T} \\ \vdots \\ \varepsilon_{n,p+1} \\ \vdots \\ \varepsilon_{n,T} \end{pmatrix}$$

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

$$\Leftrightarrow \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = (I_n \otimes Z) \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} + \begin{pmatrix} \varepsilon^1 \\ \vdots \\ \varepsilon^n \end{pmatrix}$$

Ce qui équivaut au modèle $(M) \quad x = (I_n \otimes Z)\beta + \varepsilon$ qui se décompose en les n sous-modèles $(M_i) \quad x_i = Z\beta_i + \varepsilon^i$

Remarque 5.3.1 (1) Si on suppose que les observations démarrent à $t = -p + 1$ (au lieu de $t = 1$), on peut écrire : $(M_i) \quad x_i = Z\beta_i + \varepsilon^i$ avec

$$Z = \begin{pmatrix} 1 & x_{1,0} & \dots & x_{p,0} & \dots & x_{1,-p+1} & \dots & x_{n,-p+1} \\ \vdots & \vdots \\ 1 & x_{1,T-1} & \dots & x_{n,T-1} & \dots & x_{1,T-p} & \dots & x_{n,T-p} \end{pmatrix}$$

Dans ce cas :

$$\mathbb{V} \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1T} \\ \vdots \\ \varepsilon_{n1} \\ \vdots \\ \varepsilon_{nT} \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \dots & 0 & \dots & 0 & \sigma_{1n} & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{11} & 0 & \dots & 0 & \dots & \sigma_{1n} \\ \dots & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & 0 \\ \dots & \sigma_{n1} & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & \sigma_{n1} & 0 & \dots & 0 & \dots & \sigma_{nn} \end{pmatrix} = \begin{pmatrix} \sigma_{11}I_T & \dots & \sigma_{1n}I_T \\ \vdots & \ddots & \vdots \\ \sigma_{1n}I_T & \dots & \sigma_{nn}I_T \end{pmatrix} = \Omega \otimes I_T$$

$$\text{où } \Omega = \begin{pmatrix} \sigma_{11} & \dots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \dots & \sigma_{nn} \end{pmatrix}$$

$$(2) \quad (M_i) \quad x_i = Z\beta_i + \varepsilon^i \text{ avec } \mathbb{V}(\varepsilon^i) = \mathbb{V} \begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iT} \end{pmatrix} = \sigma_{ii}I_T$$

Bilan

$$\begin{cases} x = (I_n \otimes Z)\beta + \varepsilon \\ \mathbb{V}\varepsilon = \Omega \otimes I_T \end{cases}$$

5.3.2 Estimation par les MCQG

Théorème 5.3.1 (Théorème de Zellner) *Pour le modèle précédent,*

$$\hat{\beta}_{MCQG} = \hat{\beta}_{MCG} = \hat{\beta}_{MCO} = \begin{pmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_n \end{pmatrix}$$

Proposition 5.3.1 *L'estimateur est sans biais : $\mathbb{E}\hat{\beta} = \beta$ et a pour variance $\mathbb{V}(\hat{\beta}) = \Omega \otimes (Z'Z)^{-1}$.*

Démonstration

$$(i) \quad \hat{\beta} = (I_n \otimes (Z'Z)^{-1}Z')x = (I_n \otimes (Z'Z)^{-1}Z')((I_n \otimes Z)\beta + \varepsilon)$$

$$\hat{\beta} = (I_n \otimes I_{n(np+1)})\beta + (I_n \otimes (Z'Z)^{-1}Z')\varepsilon = \beta + (I_n \otimes (Z'Z)^{-1}Z')\varepsilon$$

$$\text{D'où } \mathbb{E}\hat{\beta} = \underbrace{\mathbb{E}\beta}_{\beta} + (I_n \otimes (Z'Z)^{-1}Z') \underbrace{\mathbb{E}\varepsilon}_0$$

$$(ii) \quad \mathbb{V}\hat{\beta} = (I_n \otimes (Z'Z)^{-1}Z')(\Omega \otimes I_T)(I_n \otimes (Z'Z)^{-1}Z')$$

Remarque 5.3.2 (1) $\mathbb{V} \begin{pmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_n \end{pmatrix} = \Omega(Z'Z)^{-1} = \begin{pmatrix} \sigma_{11}(Z'Z)^{-1} & \dots & \sigma_{1n}(Z'Z)^{-1} \\ \vdots & \ddots & \vdots \\ \sigma_{1n}(Z'Z)^{-1} & \dots & \sigma_{nn}(Z'Z)^{-1} \end{pmatrix}$

Dans le modèle (M_i) $x_i = Z\beta_i + \varepsilon^i$, $\mathbb{V}\hat{\beta}_i = \sigma_{ii}(Z'Z)^{-1}$

En outre

$$\mathbb{E}[(\hat{\beta}_i - \beta_i)(\hat{\beta}_j - \beta_j)'] = \sigma_{ij}(Z'Z)^{-1}$$

(2)

$$\begin{aligned} \hat{\varepsilon} &= x - (I_n \otimes Z)\hat{\beta} \\ &= \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} - \begin{pmatrix} Z & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & Z \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_n \end{pmatrix} \\ &= \begin{pmatrix} x_1 - Z\hat{\beta}_1 \\ \vdots \\ x_n - Z\hat{\beta}_n \end{pmatrix} \\ &= \begin{pmatrix} \hat{\varepsilon}_1 \\ \vdots \\ \hat{\varepsilon}_n \end{pmatrix} \\ \hat{\sigma}_{ii}^2 &= \frac{1}{np+1} \sum_1^T (\varepsilon_{it}^1)^2 = \mathbb{V}_{emp}(\hat{\varepsilon}_{it}) \\ \hat{\sigma}_{ij} &= \mathbb{Cov}_{emp}(\hat{\varepsilon}_{it}, \hat{\varepsilon}_{jt}) = \frac{1}{T} \sum_1^T \hat{\varepsilon}_{it}\hat{\varepsilon}_{jt} \end{aligned}$$

Définition 5.3.1 (Matrice de covariance estimée) $\hat{\Sigma} = (\hat{\sigma}_{ij})_{1 \leq i, j \leq n}$ où $\hat{\sigma}_{ij} = \frac{1}{T} \sum_1^T \hat{\varepsilon}_{it}\hat{\varepsilon}_{jt}$
 $\hat{\Sigma} = \frac{1}{T} \sum_1^T \hat{\varepsilon}_t \hat{\varepsilon}_t'$

5.3.3 EMV sous l'hypothèse de normalité

On suppose ε_t iid $\sim \mathcal{N}(0, \Omega)$. On considère le processus

$$X_t = \begin{pmatrix} x_{1t} \\ \vdots \\ x_{nt} \end{pmatrix} = \mu + \sum_{k=1}^p \phi_k X_{t-k} + \varepsilon_t$$

Calcul de la vraisemblance

$$X_t | \underline{X}_{t-1} \sim \mathcal{N} \left(\mu + \sum_1^p \phi_k X_{t-k}, \Omega \right)$$

$$l(X_t | \underline{X}_{t-1}, \theta) = \frac{1}{\sqrt{2\pi}^n} \frac{1}{\sqrt{\det \Omega}} \exp \left[-\frac{1}{2} \left(X_t - \left(\mu + \sum_1^p \phi_k X_{t-k} \right) \right)' \Omega^{-1} \left(X_t - \left(\mu + \sum_1^p \phi_k X_{t-k} \right) \right) \right]$$

avec le paramètre $\theta = (\mu, \phi_1, \dots, \phi_p, \Omega) = \{\mu_i, \varphi_{ij}^k, \sigma_{ij} \text{ tel que } 1 \leq k \leq p \text{ et } 1 \leq i, j \leq n\}$

On en déduit

$$l(X_1, \dots, X_T | X_{-p+1}, \dots, X_0, \theta) = \prod_1^T l(X_t | \underline{X}_{t-1}, \theta) = \frac{1}{\sqrt{2\pi}^{nT}} \frac{1}{\sqrt{\det \Omega}^T} \exp \left[-\frac{1}{2} \sum_1^T \varepsilon_t' \Omega^{-1} \varepsilon_t \right]$$

où $\varepsilon_t = X_t - (\mu + \sum_1^p \phi_k X_{t-k})$

$$\Rightarrow \ln l(X_1, \dots, X_T | X_{-p+1}, \dots, X_0, \theta) = \frac{-nT}{2} \ln 2\pi - \frac{T}{2} \ln \det \Omega - \frac{1}{2} \sum_1^T \varepsilon_t' \Omega^{-1} \varepsilon_t$$

Calcul de l'EMV On écrit les conditions du premier ordre :

$$\left. \begin{array}{l} \frac{\partial \ln l}{\partial \Sigma} = 0 \\ \frac{\partial \ln l}{\partial \beta} = 0 \end{array} \right\} \Rightarrow \hat{\Sigma} = \frac{1}{T} \sum_1^T \hat{\varepsilon}_t' \hat{\varepsilon}_t$$

où $\hat{\varepsilon}_t = X_t - \hat{\mu} - \sum_1^n \hat{\phi}_k X_{t-k}$ avec $\hat{\mu}$ et $\hat{\phi}_k$ EMV respectifs de μ et ϕ_k

D'où par concentration de la vraisemblance

$$\begin{aligned} \hat{\beta}_{EMV} &= \hat{\beta}_{MCG} \text{ (d'après le sur-modèle)} \\ &= \hat{\beta}_{MCO} \text{ (d'après le théorème de ZELLNER)} \end{aligned}$$

Valeur de la vraisemblance au maximum $\ln l(X, \hat{\theta}) = \frac{-nT}{2} \ln 2\pi - \frac{T}{2} \ln \det \hat{\Omega} - \frac{1}{2} \sum_1^T \varepsilon_t' \hat{\Omega}^{-1} \varepsilon_t$
avec

$$\sum_1^T \varepsilon_t' \hat{\Omega}^{-1} \varepsilon_t = \text{tr} \left(\sum_1^T \varepsilon_t' \hat{\Omega}^{-1} \varepsilon_t \right) = \sum_1^T \text{tr} \left(\varepsilon_t' \hat{\Omega}^{-1} \varepsilon_t \right) = \sum_1^T \text{tr} \left(\hat{\Omega}^{-1} \varepsilon_t' \varepsilon_t \right) = \text{tr} \left(\hat{\Omega}^{-1} \underbrace{\sum_1^T \varepsilon_t' \varepsilon_t}_{T\hat{\Omega}} \right) = T \text{tr}(I_n) = nT$$

D'où

$$\begin{aligned} \ln l(X, \hat{\theta}) &= \frac{-nT}{2} \ln 2\pi - \frac{T}{2} \ln \det \hat{\Omega} - \frac{nT}{2} \\ &= \frac{-nT}{2} (1 + \ln 2\pi) - \frac{T}{2} \ln \det \hat{\Omega} \end{aligned}$$

5.3.4 Propriétés de l'EMV sous l'hypothèse de normalité

On sait que $\hat{\beta}_{EMV} = \hat{\beta}_{MCG} = \hat{\beta}_{MCO} = \hat{\beta}$ donc $\mathbb{E}\hat{\beta} = \beta$ et $\mathbb{V}\hat{\beta} = \Omega \otimes (Z'Z)^{-1}$

Proposition 5.3.2 (Sous hypothèse de normalité des résidus)

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

Proposition 5.3.3 (Sans l'hypothèse de normalité) (i) $\hat{\beta} \xrightarrow{P} \beta$

(ii) $\hat{\Omega} \xrightarrow{P} \Omega$

(iii) $\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{loi} \mathcal{N}\left(\Omega \otimes \left(p \lim \frac{Z'Z}{T}\right)^{-1}\right)$

Remarque 5.3.3 (1) (i) et (ii) sont valables pour $\hat{\beta}$ et $\hat{\Omega}$ avec ou sans l'hypothèse de normalité (seule condition : processus stationnaire à l'ordre 4 de ε_t pour avoir la covariance de Ω).

(2) $p \lim \frac{Z'Z}{T}$ existe déjà grâce à l'hypothèse de stationnarité de X_t .

Dans $\frac{Z'Z}{T}$ interviennent des termes de la forme

$$\frac{1}{T} \sum_{t=1}^T x_{it}x_{j,t-k} \xrightarrow{T \rightarrow +\infty} \underbrace{\mathbb{E}(x_{it}x_{j,t-k})}_{\text{existe et dépend de } (i,j,k) \text{ seulement}}$$

avec $\mathbb{E}(x_{it}x_{j,t-k}) = \gamma_{ij}(k) + m_i m_j$ et $\Gamma(k) = \mathbb{E}[(X_t - m)(X_{t-k} - m)']$

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

$$\sqrt{T}(\hat{\beta} - \beta) \rightsquigarrow \mathcal{N}(0, T\Omega \otimes (Z'Z)^{-1})$$

Comme $p \lim \frac{Z'Z}{T}$ existe et est positive, alors $p \lim \left(\frac{Z'Z}{T}\right)^{-1}$ existe. On a donc bien

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{loi} \mathcal{N}\left(0, \Omega \otimes p \lim \left(\frac{Z'Z}{T}\right)^{-1}\right)$$

Cette dernière propriété reste vraie même si on ne suppose pas les ε_t gaussiens. Elle résulte alors du TCL et de la stationnarité de (X_t) .

5.3.5 Tests de restrictions linéaires sur les paramètres du modèle sous hypothèse de normalité

Test de Wald

On teste l'hypothèse $H_0 : R\beta = r$ contre l'hypothèse alternative $H_0 : R\beta \neq r$ avec $R \in \mathcal{M}(q, n(np+1))$, $\beta \in \mathbb{R}^{n(np+1)}$ tel que le $\text{rang } R = q < n(np+1)$.

– Sous l'hypothèse $\varepsilon_t \text{ iid } \rightsquigarrow \mathcal{N}(0, \Omega)$, on a

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

$$R\hat{\beta} - r \rightsquigarrow \mathcal{N}(0, R(\Omega \otimes (Z'Z)^{-1})R') \text{ sous } H_0$$

$$\Rightarrow (R\hat{\beta} - r)' [R(\Omega \otimes (Z'Z)^{-1})R']^{-1} (R\hat{\beta} - r) \rightsquigarrow \chi^2(q) \text{ sous } H_0$$

Comme $p \lim \hat{\Omega} = \Omega$

$$\Rightarrow \xi_N = (R\hat{\beta} - r)' [R(\hat{\Omega} \otimes (Z'Z)^{-1})R']^{-1} (R\hat{\beta} - r) \rightsquigarrow \chi^2(q) \text{ sous } H_0$$

On refuse H_0 au seuil α si $\xi_N > \chi_{1-\alpha/2}^2(q)$.

– Sans l'hypothèse de normalité sur ε , on a $\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{loi} \mathcal{N}\left(\Omega \otimes \left(p \lim \frac{Z'Z}{T}\right)^{-1}\right)$.

$$\Rightarrow \xi_N = (R\hat{\beta} - r)' \left[R \left(\Omega \otimes p \lim \left(\frac{Z'Z}{T} \right)^{-1} \right) R' \right]^{-1} (R\hat{\beta} - r) \xrightarrow{loi} \chi^2(q) \text{ sous } H_0$$

Puis on fait comme dans le cas précédent.

Exemples d'application ($n = 2, p = 2$)

On considère le modèle :

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

Significativité d'un paramètre On teste $H_0 : d_1 = 0$ contre $H_1 : d_1 \neq 0$.

$$(M_1) \quad x_1 = Z\beta_1 + \varepsilon^1 \text{ et } \mathbb{V}(\varepsilon^1) = \sigma_u^2 I_T$$

$$(M_2) \quad x_2 = Z\beta_2 + \varepsilon^2$$

$$\hat{d}_1 \rightsquigarrow \mathcal{N}(d_1, \sigma_{11}^2 z^{55}) \text{ où } (z_{ij})_{i,j} = (Z'Z)^{-1}$$

$$\text{D'où } \frac{\hat{d}_1}{\hat{\sigma}_{11}\sqrt{z^{55}}} \rightsquigarrow St(2)$$

Tests sur les paramètres d'un seul sous-modèle On teste $H_0 : b_1 = d_1$ contre $H_1 : b_1 \neq d_1$.

$$\hat{b}_1 - \hat{d}_1 \rightsquigarrow \mathcal{N}(0, \sigma_{11}^2)$$

On fait un test de STUDENT ou de FISHER dans le modèle (M_1) .

Tests sur plusieurs sous-modèles On tient compte de $\mathbb{V}\hat{\beta} = \Omega \otimes (Z'Z)^{-1}$ et $\hat{\mathbb{V}}\hat{\beta} = \Omega \otimes (Z'Z)^{-1}$.

On teste $H_0 : \phi_p = 0$ contre $H_1 : \phi_p \neq 0$. Ici $c_1 = d_1 = c_2 = d_2 = 0$ ($q = 4$).

$$\begin{pmatrix} \hat{c}_1 \\ \hat{d}_1 \\ \hat{c}_2 \\ \hat{d}_2 \end{pmatrix} \rightsquigarrow \mathcal{N}\left(0, \begin{pmatrix} \text{matrice } (4, 4) \\ \text{fonction de } \Omega \end{pmatrix}\right)$$

On doit en tenir compte dans la statistique de FISHER.

Test du rapport de vraisemblances (LR test)

Exemple 5.3.2 On teste $H_0 : \phi_p = 0$ contre $H_1 : \phi_p \neq 0$. On teste $H_0 : \phi_p = 0$ contre $H_1 : \phi_p \neq 0$.

$$\begin{aligned} - (M) \quad X_t &= \mu + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t \text{ de paramètre } \theta = (\beta, \Omega) \rightarrow \hat{\beta}, \hat{\Omega} = \frac{1}{T} \sum_t \hat{\varepsilon}_t \hat{\varepsilon}_t' \\ \ln l(X, \hat{\theta}) &= \frac{-nT}{2} (1 + \ln 2\pi) - \frac{T}{2} \ln \det \hat{\Omega} \end{aligned}$$

$$\begin{aligned}
- (M^0) \quad X_t &= \mu + \phi_1 X_{t-1} + \cdots + \phi_{p-1} X_{t-p+1} + \varepsilon_t \text{ de paramètre } \theta^0 = (\beta^0, \Omega^0) \rightarrow \hat{\beta}^0, \hat{\Omega}^0 = \\
&\frac{1}{T} \sum_t \hat{\varepsilon}_t^0 \hat{\varepsilon}_t^{0'} \\
\ln l(X, \hat{\theta}^0) &= \frac{-nT}{2} (1 + \ln 2\pi) - \frac{T}{2} \ln \det \hat{\Omega}^0
\end{aligned}$$

$$\begin{aligned}
\xi_{LR} &= 2 \left(\ln l(X, \hat{\theta}) - \ln l(X, \hat{\theta}^0) \right) \\
&= -T \left(\ln \det \hat{\Omega} - \ln \det \hat{\Omega}^0 \right) \\
&= T \ln \left(\frac{\det \hat{\Omega}^0}{\det \hat{\Omega}} \right)
\end{aligned}$$

Test de causalité

Quand on rajoute le passé d'une variable et d'une autre, la précision de la prévision est-elle meilleure ?

On doit comparer la variance de $EL(x_{t+1} | \underline{x}_t, \underline{y}_t)$ à celle de $EL(x_{t+1} | \underline{x}_t)$.

Bibliographie

- [1] W. HÄRDLE and G. KERKYACHARIAN and D. PICARD and A. TSYBAKOV. *Wavelets, Approximation and Statistical Applications*. Springer Verlag, 1998
- [2] C. GOURIEROUX and A. MONFORT. *Séries temporelles et modèles dynamiques*. Economica
- [3] J. HAMILTON. *Time Series Analysis*. Princeton University Press
- [4] BROCKWELL and DAVIS. *Times Series : Theory and Methods*. Springer Verlag
- [5] BOSQ and LECOUTRE. *Analyse et prévision des séries chronologiques ; méthodes paramétriques et non paramétriques*. Masson
- [6] CARBON and DELECROIX. *Non Parametric versus Parametric Forecasting in Time Series : a Computational Point of View, Applied Stochastic Models and Data Analysis*. vol. 9, p 215-229. 1993
- [7] ABADIE and MESLIER. *Etude de l'utilisation des modèles ARIMA pour la prévision à très court terme de l'énergie journalière produite par Electricité de France*. RAIRO Recherche opérationnelle/ Operations research n°1, vol. 13 p 37-54, février 1979
- [8] MARTIN. *Filtrage de Kalman d'une série temporelle saisonnière. Application à la prévision de consommation d'électricité*. Revue de statistique appliquée v. XLVII p 69-86, 1999
- [9] J.M. POGGI. *Prévision non paramétrique de la consommation électrique*. Revue de Statistique appliquée vol. XLII p 83-98, 1994
- [10] Y. MISITI and M. MISITI and G. OPPENHEIM and J.M. POGGI. *Analyse de signaux classiques par décomposition en ondelettes*, Revue de Statistique Appliquée n°4 vol. XLI p 5-35, 1993
- [11] Y. MISITI and M. MISITI and G. OPPENHEIM and J.M. POGGI. *Ondelettes en statistique et traitement du signal*, Revue de Statistique Appliquée n°4 vol. XLI p 34-43, 1993
- [12] Y. MISITI and M. MISITI and G. OPPENHEIM and J.M. POGGI. *Décomposition en ondelettes et méthodes comparatives : étude d'une courbe de charge électrique*, Revue de Statistique Appliquée n°2 vol. XLII p 57-77, 1994

Index

- AKAÏKE (critère d'information d'), 53
- auto-corrélation, 5, 25, 33
 - inverse, 14, 27
 - partielle, 10, 26
- auto-corrélogramme, 6
 - empirique, 15
 - inverse, 14
 - partiel, 10
- auto-covariance, 3, 5, 25, 30, 33
 - inverse, 14
- bruit blanc
 - faible, 4
 - fort, 4
- critère
 - d'information, 53
 - de parcimonie, 53
 - de qualité de la prévision, 53
- critère d'information, 53
- densité spectrale, 11
- densité spectrale (processus vectoriel), 65
- DICKEY-FULLER (test de), 41
- estimation, 74
- HANNAN-QUINN (critère d'information d'), 53
- injectivité (théorème d'), 12
- innovation, 11
- intervalle de précision, 60
- inversibilité, 17
- KOLMOGOROV (théorème de), 3
- KPPS (test), 46
- marche aléatoire, 4
 - sans dérive, 4
- moyenne mobile, 4
 - infinie, 7
- NEWBY-WEST (estimateur de), 45
- opérateur, 15
 - avance, 16
 - retard, 15
- persistance des chocs, 40
- PHILLIPS-PERRON (test de), 44
- polynôme, 16
 - avance, 16
 - retard, 16
- porte-manteau, 52
- prévision, 67, 73
 - linéaire optimale, 67
- prévision optimale, 11, 55
- processus *AR*, 21
- processus *AR*(∞), 28
- processus *ARIMA*, 35
- processus *ARMA*, 31
- processus *MA*, 28
- processus *MA*(∞), 22
- processus des innovations, 11, 68
- processus intégré, 35
- processus stationnaire, 3
 - du second ordre, 3
 - strict, 3
- processus stochastique, 3
- processus vectoriel, 63
 - du second ordre, 63
 - stationnaire du second ordre, 64
- régression, 8
 - affine théorique (retards finis), 8
 - affine théorique (retards infinis), 10
 - linéaire théorique (retards finis), 8
 - linéaire théorique (retards infinis), 10
- représentation *VMA*(∞), 73
- représentation canonique, 23, 24, 28
 - minimale, 31
- représentation canonique minimale, 35

SCHMIDT-PHILLIPS (test de), 45

trajectoire, 3

WALD (test de), 79

WOLD (théorème de), 11, 69

YULE-WALKER (équation de), 49

YULE-WALKER (équations de), 25, 34

ZELLNER (théorème de), 76