

SERIES CHRONOLOGIQUES

Quelques éléments du cours¹

Année 2004-2005

Corinne Perraudin

**Chapitre 1:
Les modèles ARMA stationnaires**

Contents

1	Processus aléatoires stationnaires	4
1.1	Variables aléatoires réelles de carré intégrable	4
1.2	La stationnarité	5
2	Autocorrélations simple et partielle	6
2.1	La fonction d'autocovariance et d'autocorrélation	6
2.2	La fonction d'autocorrélation partielle	8
3	Représentation de Wold	10
3.1	Présentation	10
3.2	Prévision à partir de la représentation de Wold	12
3.3	Opérateur retard	12
4	Le spectre et la densité spectrale	13
5	Modèles ARMA et représentations canoniques	16
5.1	Processus MA	18
5.1.1	Définition	18
5.1.2	Représentation stationnaire et causale	18
5.1.3	Représentation inversible	18

¹Pour toutes remarques, n'hésitez pas à me contacter : Corinne.Perraudin@univ-paris1.fr

5.2	Processus AR	21
5.2.1	Définition	21
5.2.2	Représentation stationnaire	21
5.2.3	Représentation inversible	21
5.2.4	Représentation causale	21
5.3	Processus ARMA	23
5.3.1	Définition	23
5.3.2	Propriété	24
6	Caractéristiques des processus ARMA: autocorrélation simple, partielle et densité spectrale	24
6.1	Corrélogramme simple	25
6.1.1	Cas de l'AR(p)	25
6.1.2	Cas du MA(q)	26
6.1.3	Cas de l'ARMA(p,q)	26
6.2	Corrélogramme partiel	27
6.2.1	Cas de l'AR(p)	27
6.2.2	Cas du MA(q)	28
6.2.3	Cas de l'ARMA(p,q)	28
6.3	L'autocorrélation inverse	28
6.4	Tableau récapitulatif	28
6.5	Densité spectrale	28
7	Identification, estimation, validation, prévision	29
7.1	Identification du processus ARMA	29
7.2	Estimation	30
7.2.1	Maximum de vraisemblance	30
7.2.2	Estimation de Yule-Walker	31
7.2.3	Estimation par l'algorithme de Durbin-Levinson	31
7.2.4	Moindres carrés conditionnels	32
7.3	Validation	32
7.3.1	Tests sur les résidus	32
7.3.2	Tests sur les paramètres	34
7.3.3	Choix d'un modèle parmi plusieurs	34
7.4	Prévision	34
7.4.1	Calcul des prévisions optimales et mise à jour	34
7.4.2	Fonction de prévision	36

Introduction

Quand on cherche à modéliser une série temporelle, on a généralement recours à la classe des modèles ARMA, qui permet de rendre compte d'un assez grand nombre de cas économiques. Dans cette classe de modèle, on distingue :

- les modèles AR(p) : $Y_t = \sum_{i=1}^p \phi_i Y_{t-i} + \varepsilon_t$
- les modèles MA(q) : $Y_t = \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i}$
- les modèles ARMA(p,q) : $Y_t - \sum_{i=1}^p \phi_i Y_{t-i} = \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i}$

où ε_t est un bruit blanc centré de variance σ^2 .

Depuis les travaux de Box et Jenkins en 1970 dans le cas univarié (Y_t correspond à une seule variable), la méthodologie utilisée pour modéliser une série économique repose sur l'algorithme de Box et Jenkins, qui consiste à :

- Transformer la série Y_t afin d'éliminer d'éventuelles non-stationnarités (tendance, saisonnalité).²
- Identifier, sur la base des caractéristiques temporelles de la série étudiée, le modèle ARMA pertinent : il s'agit de choisir un modèle dans la classe des modèles ARMA, ainsi que de déterminer p et q , en comparant les caractéristiques temporelles de la série observée avec les caractéristiques théoriques des modèles ARMA
- Estimer les paramètres ϕ_i, θ_i et σ^2
- Tester la validité et l'adéquation du modèle aux données
- Re-spécifier le modèle si besoin est et faire de la prévision

L'objectif de ce chapitre est de présenter précisément les modèles ARMA et leur caractéristiques, ainsi que la méthodologie adoptée pour spécifier ces modèles, cela dans le cadre univarié. Nous supposons que la série étudiée est stationnaire (ou qu'elle a été rendue stationnaire). Dans le chapitre 2, nous étudierons comment détecter la non stationnarité et comment transformer la série si elle n'est pas stationnaire.

²Nous verrons cela dans le chapitre 2.

1 Processus aléatoires stationnaires

On utilise le terme de processus aléatoire pour décrire une variable dont le comportement ne peut pas être exprimé entièrement par une relation déterministe.

Un processus aléatoire est une suite de variables aléatoires indexées dans le temps et définies sur un espace des états de la nature. Ainsi, pour chaque instant du temps, la valeur de la quantité étudiée Y_t est appelée **variable aléatoire** et l'ensemble des valeurs Y_t quand t varie est appelé **processus aléatoire**.

L'indice t (t appartenant à un ensemble \mathcal{T}) s'interprète comme la date à laquelle est faite l'observation ou comme la période sur laquelle elle porte. Les observations étant en nombre fini, il pourrait paraître naturel de choisir pour \mathcal{T} un ensemble fini. On préfère cependant retenir $\mathcal{T} = \mathbb{N}$ ou $\mathcal{T} = \mathbb{Z}$. Prolonger l'ensemble des indices vers $+\infty$ permet en effet de prendre en compte la possibilité d'observations nouvelles et d'étudier les propriétés asymptotiques des diverses procédures statistiques. L'intérêt du prolongement vers $-\infty$ est principalement mathématique : il permet en effet dans certains cas des écritures et des résultats plus simples.

Une réalisation du processus décrit une histoire possible du processus. Il faut bien remarquer qu'une série chronologique observée, comme par exemple la suite des valeurs de l'indice des prix du pétrole entre 1970 et 1990, est relative à une réalisation et une seule du processus. L'une des difficultés de l'étude des séries temporelles est de reconstituer, au moins partiellement, la loi du processus à partir de l'observation d'une seule de ses réalisations.

Pour que ceci soit possible, il faut bien entendu imposer certaines conditions sur la loi de probabilité de Y_t .

1.1 Variables aléatoires réelles de carré intégrable

On se placera dans l'ensemble L^2 des variables aléatoires réelles (v.a.r.) admettant un moment d'ordre 2 ($E(Y_t)^2 < \infty$) et où 2 variables X et Y sont considérées comme égales si $E[(X - Y)^2] = 0$.

L'ensemble L^2 des v.a.r. de carré intégrable ($E(Y_t^2) < +\infty$) est un espace vectoriel normé sur \mathbb{R} , la norme étant $\|Y\|^2 = \langle Y, Y \rangle = [E(Y^2)]$ et le produit scalaire est $\langle X, Y \rangle = E(XY)$.

On dit qu'il a une structure d'espace de Hilbert, généralisation en dimension infinie de l'espace euclidien \mathbb{R}^n muni du produit scalaire usuel $\langle X, Y \rangle = \sum_i x_i y_i$.

Ainsi, pour deux v.a.r. X et Y de L^2 , il est possible de calculer l'espérance de leur

produit $E(XY)$. X et Y sont dites orthogonales si et seulement si $E(XY) = 0$

L'existence sur L^2 d'une notion d'orthogonalité et d'une notion de convergence permet d'introduire la notion de projection orthogonale. Ainsi, on peut trouver un sous espace fermé de variables de carré intégrable contenant toutes les combinaisons linéaires de variables Y_i de L^2 .

On se restreindra dans la suite du cours aux variables de L^2 . On parlera aussi de processus du second ordre.

Les calculs des moyennes $E(Y_t) = m_t$, des variances et des covariances ont alors un sens. La loi d'un processus du 2nd ordre peut donc être partiellement résumée par :

- la suite des moyennes m_t
- la suite des covariances temporelles (évolution moyenne, modification des dispersions, liaisons dans le temps)

Pour chaque instant du temps, Y_t a une distribution de probabilité. Si on ne fait aucune hypothèse particulière sur la nature du processus aléatoire, alors la fonction de densité de probabilité de Y_t dépend du temps. Ainsi, la moyenne et la variance varient donc également : m_t et σ_t^2 sont des fonctions du temps. Il faudrait alors étudier la distribution de probabilité de Y_t pour chaque valeur de t sachant qu'on ne disposera (en économie) que d'une seule observation de Y_t .

On étudie donc une classe particulière de processus aléatoires appelés processus aléatoires stationnaires. Ces processus sont caractérisés par le fait que leurs propriétés ne changent pas au cours du temps.

1.2 La stationnarité

La stationnarité au sens strict

On dit que le processus Y_t , $t \in \mathcal{T}$ est stationnaire au sens strict (ou fortement stationnaire) si la loi de $\{Y_{t_1}, \dots, Y_{t_n}\}$ est la même que la loi de $\{Y_{t_1+\tau}, \dots, Y_{t_n+\tau}\}$ pour tout (t_1, t_2, \dots, t_n) avec $t_i \in \mathcal{T}$, pour $i = 1, \dots, n$ et pour tout $\tau \in \mathcal{T}$ avec $t_{i+\tau} \in \mathcal{T}$.

Ainsi, un processus aléatoire est strictement stationnaire si toutes ces caractéristiques, c'est-à-dire tous ces moments sont invariants pour tout changement de l'origine du temps.

Mais la stationnarité au sens strict est trop restrictive et on assouplit cette condition en définissant la stationnarité du second ordre.

La stationnarité du second ordre

Un processus $Y_t, t \in \mathcal{T}$ est dit stationnaire du second ordre (ou faiblement stationnaire) si $Y_t, t \in \mathcal{T}$ est du 2nd ordre et si les deux premiers moments sont invariants dans le temps :

- $E(Y_t) = m = \text{Cste} \quad \forall t \in \mathcal{T}$
- $Var(Y_t) = \sigma^2 = \gamma(0) < \infty$
- $Cov(Y_t, Y_{t-h}) = E(Y_t Y_{t-h}) - E(Y_t)E(Y_{t-h}) = \gamma(h) \quad \forall t \in \mathcal{T}, \forall h \in \mathcal{T}$

En résumé, un processus Y_t est dit stationnaire du second ordre si sa moyenne, sa variance et sa covariance sont indépendantes du temps et si sa variance est finie. Un tel processus est sans tendance en moyenne et sans tendance en variance.

Un tel processus admet donc une loi, qui pour ses deux premiers moments, est invariante par changement de l'origine des temps. En particulier, les variables Y_t ont une même variance égale à $\gamma(0)$: propriété d'homoscédasticité.

L'exemple le plus connu de processus stationnaire est le processus bruit blanc (noté BB, ou *White noise*). Un **Bruit Blanc** est une suite de v.a.r. $\varepsilon_t, t \in \mathcal{T}$ telle que :

$$\begin{aligned} E(\varepsilon_t) &= 0 \quad \forall t \in \mathcal{T} \\ \gamma(h) &= E(\varepsilon_t \varepsilon_{t-h}) = \begin{cases} \sigma^2 & h = 0 \\ 0 & h \neq 0 \end{cases} \end{aligned}$$

Il s'agit d'une suite de v.a.r. homoscédastiques et non autocorrélées (et même indépendantes, c'est pourquoi on parle aussi de processus i.i.d. pour identiquement et indépendamment distribué).

2 Autocorrélations simple et partielle

Les principales caractéristiques temporelles d'un processus sont données par l'autocorrélation (simple) et l'autocorrélation partielle.

2.1 La fonction d'autocovariance et d'autocorrélation

La fonction d'autocovariance $\{\gamma(h)\}_{h \in \mathbb{Z}}$ mesure la covariance entre une variable et cette même variable à des dates différentes, pour un délai h :

$$\gamma(h) = Cov(Y_t, Y_{t-h}) = E[(Y_t - E(Y_t))(Y_{t-h} - E(Y_{t-h}))]$$

Ainsi $\gamma(0) = Var(Y_t) = E[(Y_t - E(Y_t))^2] = \sigma_Y^2$.

Elle fournit une information sur la variabilité de la série et sur les liaisons temporelles qui existent entre les diverses composantes de la série Y_t .

La fonction d'autocovariance d'un processus stationnaire est une fonction :

* paire : $\gamma(-h) = \gamma(h) \forall h$

* semi-définie positive :

$$\sum_{j=1}^n \sum_{k=1}^n a_j a_k \gamma(t_j - t_k) > 0 \quad \forall n \in \mathbb{N}, \forall a_j \in \mathbb{R}, \forall t_j \in \mathbb{Z}$$

puisque cette quantité est égale à $V\left(\sum_{j=1}^n a_j Y_{t_j}\right)$.

La fonction d'autocorrélation est définie par :

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}, \quad h \in \mathbb{Z}$$

avec $\rho(0) = 1$ et $|\rho(h)| < 1$ (donc $|\gamma(h)| \leq \gamma(0)$)

On appelle coefficient d'autocorrélation d'ordre 1 (resp. d'ordre k) le coefficient de corrélation linéaire $\rho(1)$ (resp. $\rho(k)$) calculé entre la série et cette série décalée d'une période (resp. k périodes).

On définit la matrice de corrélation (de dimension m) de la manière suivante :

$$R(m) = \begin{pmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(m-1) \\ \rho(1) & 1 & \rho(1) & \dots & \rho(m-2) \\ & & \vdots & & \\ \rho(m-1) & \rho(m-2) & \dots & \rho(1) & 1 \end{pmatrix}$$

Puisque la fonction $\rho(h)$, $h \in \mathbb{Z}$ est de type positif, on a la propriété suivante :

$$\det R(m) \geq 0, \quad \forall m \in \mathbb{N}^*$$

Ainsi, on a les contraintes suivantes :

- $\det R(1) \geq 0$
- $\det R(2) \geq 0 \iff \rho(1)^2 \leq 1$
- $\det R(3) \geq 0 \iff [1 - \rho(2)][1 + \rho(2) - 2\rho(1)^2] \geq 0$.

Ainsi, comme $\rho(2) < 1$, on a $\rho(2) \geq 2\rho(1)^2 - 1$. Si la corrélation d'ordre 1 est élevée, il en est de même de la corrélation d'ordre 2. Il ne peut donc y avoir de chute brutale de valeur entre $\rho(1)$ et $\rho(2)$ lorsque $\rho(1)$ est grand.

L'équivalent empirique de la fonction d'autocorrélation, noté $\hat{\rho}(h)$, est obtenu à partir de l'estimateur suivant pour l'autocovariance $\hat{\gamma}(h)$ à l'ordre h :

$$\hat{\gamma}(h) = \frac{1}{T-h-1} \sum_{t=h+1}^T (y_t - \bar{y})(y_{t-h} - \bar{y})$$

où

$$\bar{y} = \frac{1}{T} \sum_t y_t$$

Afin de tester la nullité du coefficient d'autocorrélation d'ordre h , on calcule la variance de ce coefficient. On peut montrer qu'elle est donnée par :

$$V(\hat{\rho}(h)) = \frac{1}{T} \sum_{j=-(h-1)}^{h-1} \hat{\rho}(j)^2$$

soit en utilisant la symétrie des $\rho(j)$, on obtient:

$$V(\hat{\rho}(h)) = \frac{1}{T} \left(1 + 2 \sum_{j=1}^{h-1} \hat{\rho}(j)^2 \right)$$

La statistique de test de nullité du coefficient d'autocorrélation est :

$$t = \frac{\hat{\rho}(h)}{\sqrt{\hat{V}(\hat{\rho}(h))}}$$

Elle suit une loi de Student. La variance dépendant de h , l'intervalle de confiance associé au corrélogramme (ensemble des coefficients d'autocorrélation quand h varie) augmente avec h .

2.2 La fonction d'autocorrélation partielle

Elle mesure la liaison (linéaire) entre Y_t et Y_{t-h} une fois retirés les liens transitant par les variables intermédiaires $Y_{t-1}, \dots, Y_{t-h+1}$.

Le coefficient d'autocorrélation partielle d'ordre h , noté $r(h)$, est le coefficient de corrélation entre :

* $Y_t - E(Y_t/Y_{t-1}, \dots, Y_{t-h+1})$

* et $Y_{t-h} - E(Y_{t-h}/Y_{t-1}, \dots, Y_{t-h+1})$

On a donc :

$$r(h) = \text{corr}(Y_t, Y_{t-h}/Y_{t-1}, \dots, Y_{t-h+1})$$

C'est donc le coefficient de Y_{t-h} dans la régression de Y_t sur $Y_{t-1}, \dots, Y_{t-h+1}, Y_{t-h}$.

Si Y_t est un processus stationnaire centré, la prédiction optimale de Y_t sachant son passé jusqu'à $t - h$ est donnée par :

$$E(Y_t/Y_{t-1}, \dots, Y_{t-h}) = a_1 Y_{t-1} + \dots + a_h Y_{t-h}$$

que l'on peut réécrire matriciellement :

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_h \end{pmatrix} = \begin{pmatrix} \gamma_0 & \gamma_1 & \dots & \gamma_{h-1} \\ \gamma_1 & \gamma_0 & \dots & \gamma_{h-2} \\ \vdots & & \ddots & \\ \gamma_{h-1} & \gamma_{h-2} & \dots & \gamma_0 \end{pmatrix}^{-1} \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_h \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & \dots & \rho_{h-1} \\ \rho_1 & 1 & \dots & \rho_{h-2} \\ \vdots & & \ddots & \\ \rho_{h-1} & \rho_{h-2} & \dots & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_h \end{pmatrix}$$

Le coefficient d'autocorrélation partielle d'ordre h d'un processus stationnaire est alors a_h et se calcule de la manière suivante :

$$r(h) = \frac{|R(h)^*|}{|R(h)|}$$

avec

$$R(h) = \begin{pmatrix} 1 & \rho_1 & \dots & \rho_{h-1} \\ \rho_1 & 1 & \dots & \rho_{h-2} \\ \vdots & & \ddots & \\ \rho_{h-1} & \rho_{h-2} & \dots & 1 \end{pmatrix}$$

et $R(h)^*$ la matrice $R(h)$ dans laquelle on a remplacé la colonne h par $\begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_h \end{pmatrix}$, soit :

$$R(h)^* = \begin{pmatrix} 1 & \rho_1 & \dots & \rho_1 \\ \rho_1 & 1 & \dots & \rho_2 \\ \vdots & & \ddots & \\ \rho_{h-1} & \rho_{h-2} & \dots & \rho_h \end{pmatrix}$$

Ainsi,

$$r(1) = \rho(1) \quad r(2) = \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2} \dots$$

De manière empirique, les autocorrélations partielles s'estiment soit :

- à partir de la régression MCO de Y_t sur les h retards et en prenant le dernier coefficient,
- en estimant les autocorrélations simples et en calculant $\hat{r}(h)$ à partir de la formule ci-dessus,

- à partir de l'algorithme de Durbin: il permet de calculer récursivement les divers coefficients de régression en évitant l'inversion des matrices de corrélation $R(h)$. Il est basé sur une formule de calcul des coefficients $a_h(H)$ à partir des $a_h(H-1)$ et $a_1(1) = \rho(1)$.

Afin de tester la nullité du coefficient d'autocorrélation partielle d'ordre h , on donne la variance de l'autocorrélation partielle estimée :

$$V(\hat{r}(h)) = \frac{1}{T} \quad \forall h$$

Ainsi, l'intervalle de confiance du corrélogramme partiel est le même pour tout h .

3 Représentation de Wold

3.1 Présentation

Propriété : Si $Y_t, t \in \mathbb{Z}$ est un processus stationnaire, et si $(a_i, i \in \mathbb{Z})$ est une suite de nombres réels absolument sommable ($\sum_{i=-\infty}^{+\infty} |a_i| < +\infty$), alors :

$$Z_t = \sum_{i=-\infty}^{+\infty} a_i Y_{t-i}, \quad t \in \mathbb{Z}$$

est un nouveau processus stationnaire. On parle de représentation moyenne mobile infinie, notée $MA(\infty)$.

En effet, la série $\sum_{i \in \mathbb{Z}} a_i Y_{t-i}$ est convergente dans L^2 , car :

$$\sum_{i=-\infty}^{+\infty} \|a_i Y_{t-i}\| = \sum_{i=-\infty}^{+\infty} |a_i| \|Y_{t-i}\| = [\gamma(0) + m^2]^{1/2} \sum_{i=-\infty}^{+\infty} |a_i| < +\infty$$

L'écriture $\sum_{i=-\infty}^{+\infty} a_i Y_{t-i}$ a alors un sens dans L^2 et la variable Y_t est de carré intégrable.

Les moments de Z_t sont :

$$\begin{aligned} E(Z_t) &= E\left(\sum_{i=-\infty}^{+\infty} a_i Y_{t-i}\right) = \sum_{i=-\infty}^{+\infty} a_i E(Y_{t-i}) = m_Y \sum_{i=-\infty}^{+\infty} a_i \text{ (indépendant de } t) \\ cov(Z_t, Z_{t+h}) &= cov\left(\sum_{i=-\infty}^{+\infty} a_i Y_{t-i}, \sum_{j=-\infty}^{+\infty} a_j Y_{t+h-j}\right) = \sum_{i=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} a_i a_j cov(Y_{t-i}, Y_{t+h-j}) \\ &= \sum_{i=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} a_i a_j \gamma_Y(h+i-j) = \gamma_Z(h) \text{ (indépendant de } t) \end{aligned}$$

L'exemple le plus simple de processus stationnaire est fourni par le processus bruit blanc ε_t , $t \in \mathbb{Z}$. Ainsi, tout processus de la forme :

$$Y_t = \sum_{k \in \mathbb{Z}} a_k \varepsilon_{t-k} \text{ où } \sum_{k \in \mathbb{Z}} |a_k| < \infty$$

est stationnaire du 2nd ordre, avec :

$$E(Y_t) = 0 \quad \gamma_y(0) = \sigma^2 \sum_k a_k^2 \quad \gamma_y(h) = \sigma^2 \sum_k a_k a_{k-h}$$

Les processus pouvant s'écrire comme moyennes mobiles infinies de bruits blancs sont appelés **processus linéaires**. Le théorème qui suit justifie l'utilisation des représentations moyennes mobiles infinies dites unilatérales vers le passé, c'est-à-dire dans lesquelles $a_k = 0$ pour tout $k < 0$.

Le Théorème de Wold (ou décomposition de Wold) :

Tout processus stationnaire du second ordre peut être représenté sous la forme :

$$Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + m \tag{1}$$

où les paramètres ψ_j satisfont $\psi_0 = 1$, $\psi_j \in \mathbb{R}$, $\forall j \in \mathbb{N}^*$, $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ et où ε_t est un bruit blanc *i.i.d.* $(0, \sigma_\varepsilon^2)$.

La somme des chocs passés correspond à la composante linéaire stochastique de Y_t et le terme m désigne la moyenne du processus

Ainsi, tout processus stationnaire peut s'écrire comme une somme pondérée infinie de chocs passés, ces chocs étant représentés par un BB de variance finie.

La condition $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ est très importante! Elle assure l'existence des moments d'ordre 2.

On montre que le processus BB ε_t est le processus d'innovation de Y_t .

On appelle **innovation** d'un processus du 2nd ordre Y_t , la variable :

$$\varepsilon_t = Y_t - Y_t^* = Y_t - E(Y_t | 1, Y_{t-1}, Y_{t-2}, \dots)$$

Y_t^* est la prévision optimale de Y_t fonction de son passé (meilleure approximation linéaire) et l'erreur de prévision correspondante à 1 pas est appelée innovation.

Ainsi, l'innovation est la partie de Y_t non corrélée au passé de la série.

En effet, d'après l'équation (1), Y_{t-1} dépend de $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots$ (non corrélés à ε_t); Y_{t-2} dépend de $\varepsilon_{t-2}, \varepsilon_{t-3}, \dots$ (non corrélés à ε_t). Donc ε_t est non corrélé à Y_{t-1}, Y_{t-2}, \dots , le passé de Y_t , c'est donc le processus d'innovation de Y_t .

On peut aussi montrer que le processus d'innovation constitue un BB lorsque Y_t est stationnaire.

3.2 Prédiction à partir de la représentation de Wold

A partir de la représentation de Wold, on peut prévoir les valeurs futures de Y_T jusqu'en $T+h$ sachant l'ensemble d'information jusqu'en T , prédiction notée $\hat{Y}_T(h)$.

La meilleure prédiction possible de la réalisation de Y_{T+h} connaissant les valeurs jusqu'en T , $Y_T, Y_{T-1}, Y_{T-2}, Y_{T-3}, \dots$ est donnée par l'espérance conditionnelle (idée de projection) :

$$\hat{Y}_T(h) = E(Y_{T+h} | Y_T, Y_{T-1}, Y_{T-2}, Y_{T-3}, \dots)$$

Or d'après le théorème de Wold, la connaissance des valeurs passées de Y_T est équivalente à la connaissance des valeurs passées des chocs $\varepsilon_T, \varepsilon_{T-1}, \dots$

Ainsi, on peut considérer la représentation de Wold pour calculer la prédiction optimale :

$$\hat{Y}_T(h) = E(Y_{T+h} | Y_T, Y_{T-1}, Y_{T-2}, Y_{T-3}, \dots) = E\left(\sum_{j=0}^{\infty} \psi_j \varepsilon_{T+h-j} + m | \varepsilon_T, \varepsilon_{T-1}, \dots\right)$$

On constate alors que l'erreur de prédiction à un pas ($h=1$) est

$$Y_{T+1} - \hat{Y}_T(1) = \varepsilon_{T+1}$$

Il s'agit de l'innovation (fondamentale au choc). Elle a de bonnes propriétés puisque cette erreur de prédiction est indépendante de l'erreur de prédiction que l'on a pu commettre aux dates précédentes puisque $E(\varepsilon_T, \varepsilon_{T-j}) = 0$.

L'erreur de prédiction quand on prévoit à h périodes est donnée par :

$$Y_{T+h} - \hat{Y}_T(h) = \sum_{j=0}^{h-1} \psi_j \varepsilon_{T+h-j}$$

3.3 Opérateur retard

On introduit un opérateur, opérateur retard, afin d'écrire de manière plus synthétique les processus ARMA et la forme MA(∞).

L'opérateur retard L (pour *Lag* ou parfois noté aussi B pour *Backward*) est tel que :

$$Ly_t = y_{t-1}$$

Cet opérateur permet de définir une application qui à toute variable y_t associe la variable retardée. Cet opérateur a les propriétés suivantes :

$$\begin{aligned}
L^j y_t &= y_{t-j} \quad \forall j \in \mathbb{Z} \\
L^0 y_t &= y_t \\
L^j c &= c \quad \text{cste} \\
L^i(L^j y_t) &= L^{i+j} y_t = L^j(L^i y_t) \\
L^{-j} y_t &= y_{t+j} \\
(L^i + L^j) y_t &= L^i y_t + L^j y_t \\
(1 - aL)^{-1} y_t &= \frac{1}{1-aL} y_t = \lim_{j \rightarrow \infty} (1 + aL + a^2 L^2 + a^3 L^3 + \dots + a^j L^j) y_t = \sum_{j=0}^{\infty} a^j L^j y_t \quad \text{si } |a| < 1
\end{aligned}$$

Ainsi, la forme MA(∞) peut s'écrire :

$$y_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j} = \sum_{j=-\infty}^{\infty} \psi_j L^j \varepsilon_t = (\dots + \psi_{-2} L^{-2} + \psi_{-1} L^{-1} + 1 + \psi_1 L + \psi_2 L^2 + \dots) \varepsilon_t = \Psi(L) \varepsilon_t$$

et la représentation de Wold (pour tout processus Y_t stationnaire) :

$$y_t = m + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} = m + \sum_{j=0}^{\infty} \psi_j L^j \varepsilon_t = m + (1 + \psi_1 L + \psi_2 L^2 + \dots) \varepsilon_t = m + \Psi(L) \varepsilon_t$$

4 Le spectre et la densité spectrale

La densité spectrale contient la même information que la fonction d'autocovariance, mais elle est définie dans le domaine des fréquences plutôt que dans le domaine du temps. Elle permet notamment d'identifier les principales composantes d'une série (tendance, saison, cycle).

L'idée est que les différentes composantes d'une série peuvent être vues comme des oscillations ou des composantes périodiques de fréquence plus ou moins élevée. Si la fréquence des oscillations est faible, c'est qu'il s'agit d'un cycle de période très longue et donc d'une composante plutôt tendancielle, alors que si la fréquence est élevée, la composante est un cycle de période faible (par exemple une saisonnalité).

Dans l'approche temporelle, les questions sont directement formulées en termes de temps et peuvent être exprimées sous forme autorégressive. Dans l'approche fréquentielle, l'objectif est la décomposition (spectrale) d'une série (stationnaire), c'est-à-dire l'identification des oscillations aux principales fréquences. Si un processus comporte plusieurs composantes alors sa densité spectrale donnera la répartition des amplitudes des composantes aux différentes fréquences.

Dans ce cadre, on interprète la densité spectrale en étudiant les pics les plus importants. S'il n'y a pas de pics, il n'y a pas de composantes cycliques dominantes, donc il

s'agit d'un BB. S'il y a des pics, on peut alors ré-interpréter les fréquences en termes de temps pour déterminer la durée du cycle.

Les fréquences sont mesurées en fréquence angulaire, notée ω . On a les relations suivantes entre la fréquence angulaire ω et la période de temps T :

$$\omega = \frac{2\pi}{T} \Leftrightarrow T = \frac{2\pi}{\omega}$$

- Pour des données mensuelles, la période est l'année ($T=12$) et la fréquence angulaire correspondante est $\omega = 2\pi/12 = \pi/6$. Ainsi, si l'on observe un pic à la fréquence $\omega = \pi/6$, on pourra conclure que la série a une composante périodique de période 12, c'est-à-dire un effet saisonnier mensuel.
- Pour des données trimestrielles, la période est $T = 4$ et donc la fréquence correspondante est $\omega = \pi/2$. Si on observe un pic dans la densité spectrale à cette fréquence, c'est qu'il y a un effet trimestriel.
- S'il y a un pic pour des basses fréquences, cela signifie qu'il y a un cycle de long-terme. Si la densité spectrale est élevée voir infinie à l'origine, cela signale qu'il y a un problème de non stationnarité.

Comment calculer la densité spectrale d'un processus et quelles sont ses propriétés?

Nous considérons les processus $MA(\infty)$ pour lesquels la définition de la densité spectrale est "naturelle".

Un processus stationnaire linéaire est par définition un processus Y_t admettant une représentation $MA(\infty)$:

$$Y_t = \sum_{j=-\infty}^{\infty} a_j \varepsilon_{t-j}$$

avec ε_t BB et $\sum_j |a_j| < \infty$.

Sa fonction d'autocovariance est :

$$\gamma_Y(h) = cov \left(\sum_j a_j \varepsilon_{t-j}, \sum_k a_k \varepsilon_{t-h-k} \right) = \sigma^2 \sum_j a_j a_{j-h}$$

La **densité spectrale** du processus Y_t , $t \in \mathbb{Z}$ est la fonction définie sur \mathbb{R} par :

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{i\omega h} \quad \forall \omega \in \mathbb{R}$$

où $e^{i\omega h} = \cos \omega h + i \sin \omega h$.

On appelle **fonction génératrice des moments** la fonction de $z \in C$:

$$\Gamma(z) = \sum_h \gamma(h)z^h$$

Ainsi, on peut réécrire

$$f(\omega) = \frac{1}{2\pi} \Gamma(e^{i\omega})$$

Propriétés :

- $f(\omega)$ existe car la série de terme général $\gamma(h)e^{i\omega h}$ est absolument convergente puisque $\sum_h |\gamma(h)| < \infty$ (car Y_t est stationnaire)
- $f(\omega)$ est réelle car, puisque $\gamma(-h) = \gamma(h)$, on a :

$$\begin{aligned} f(\omega) &= \frac{1}{2\pi} \left[\gamma(0) + \sum_{h=1}^{\infty} \gamma(h)(e^{i\omega h} + e^{-i\omega h}) \right] \\ &= \frac{\gamma(0)}{2\pi} + \frac{1}{\pi} \sum_{h=1}^{\infty} \gamma(h) \cos \omega h \\ &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) \cos \omega h \end{aligned}$$

De plus, la parité de $\gamma(h)$ implique aussi

$$\Gamma\left(\frac{1}{z}\right) = \sum_{h=-\infty}^{\infty} \gamma(h)z^{-h} = \sum_{h=-\infty}^{\infty} \gamma(-h)z^{-h} = \sum_{h=-\infty}^{\infty} \gamma(h)z^h = \Gamma(z)$$

ainsi Γ est symétrique (et au moins défini sur le cercle unité $|z| = 1$)

- $f(\omega)$ est une fonction paire, continue, périodique de période 2π (on la représente sur $[0, \pi]$) et positive.

Propriété : Il est équivalent de connaître $\gamma(h)$ ($h \in \mathbb{Z}$) ou $f(\omega)$ ($\omega \in \mathbb{R}$) :

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) \cos \omega h \qquad \gamma(h) = \int_{-\pi}^{\pi} f(\omega) \cos \omega h d\omega$$

Densité spectrale d'un BB:

Soit ε_t un BB(0, σ^2) avec $\gamma_\varepsilon(0) = \sigma^2$ et $\gamma_\varepsilon(h) = 0$ pour $h \neq 0$. Sa densité spectrale est donnée par :

$$f(\omega) = \frac{\gamma(0)}{2\pi} = \frac{\sigma^2}{2\pi} \text{ constante}$$

Inversement si Y_t est un processus stationnaire dont la densité spectrale est constante $f(\omega) = c$, alors ce processus est un BB :

$$\gamma(h) = \int_{-\pi}^{\pi} c \cos \omega h d\omega = 0 \text{ si } h \neq 0$$

Propriété : Soit Y_t un processus $MA(\infty)$:

$$Y_t = \sum_{j \in \mathbb{Z}} a_j \varepsilon_{t-j} = A(L)\varepsilon_t$$

avec $\varepsilon_t \text{ BB}(0, \sigma^2)$ et $\sum_j |a_j| < \infty$, on a alors :

$$\Gamma_Y(z) = \sigma^2 A(z)A(1/z) \qquad f_Y(\omega) = \frac{\sigma^2}{2\pi} |A(e^{i\omega})|^2 = \frac{\sigma^2}{2\pi} A(e^{i\omega})A(e^{-i\omega})$$

Ceci nous donne des manières alternatives de calculer la densité spectrale.

Le **Gain du filtre** est donné par :

$$G(\omega) = |A(e^{i\omega})|^2$$

Idée : Si on applique le filtre $A(L)$ à la série ε_t , la série filtrée est $Y_t = A(L)\varepsilon_t$ et le gain du filtre $G(\omega)$ indique les fréquences qu'il pondère davantage ou les fréquences qu'il atténue.

Propriété : Si $Z_t = \sum_{j \in \mathbb{Z}} c_j Y_{t-j}$ où $\sum_j |c_j| < \infty$, alors

$$f_Z(\omega) = f_Y(\omega) \left| \sum_{j \in \mathbb{Z}} c_j e^{i\omega j} \right|^2$$

5 Modèles ARMA et représentations canoniques

On s'intéresse aux processus ARMA, ceux-ci donnant des représentations plus parcimonieuses que la représentation de Wold (qui nécessite un nombre infini de paramètres!) :

$$\begin{aligned} Y_t - \sum_{i=1}^p \phi_i Y_{t-i} &= \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i} \\ (1 - \sum_{i=1}^p \phi_i L^i) Y_t &= (1 - \sum_{i=1}^q \theta_i L^i) \varepsilon_t \\ \Phi(L) Y_t &= \Theta(L) \varepsilon_t \end{aligned}$$

On ne s'intéressera qu'aux représentations canoniques, c'est-à-dire aux écritures des modèles ARMA telles que ε_t soit le processus d'innovation de $Y_t \iff$ la prédiction optimale est $\hat{Y}_T(k) = E(Y_{T+k}/Y_T, Y_{T-1}, \dots)$ et l'erreur de prévision à 1 pas est l'innovation \Rightarrow on ne peut pas faire mieux que la prédiction optimale.

Il s'agit donc de trouver des contraintes sur les paramètres des modèles ARMA.

Quelles sont ces conditions?

- il faut que la représentation soit stationnaire, donc que Φ admette une inverse (\Rightarrow module des racines de Φ différent de 1) :

$$Y_t = \Phi(L)^{-1}\Theta(L)\varepsilon_t = \sum_{j=-\infty}^{\infty} h_j \varepsilon_{t-j} \text{ avec } \sum |h_j| < \infty$$

- il faut que la représentation soit inversible, c'est-à-dire que la forme $AR(\infty)$ soit tournée vers le passé (\Rightarrow module des racines de Θ strictement supérieur à 1) :

$$\Theta(L)^{-1}\Phi(L)Y_t = \varepsilon_t = \sum_{j=0}^{\infty} \pi_j Y_{t-j}$$

Y_t peut s'écrire en fonction de son passé et de ε_t

- il faut que la représentation soit causale, c'est-à-dire que la forme $MA(\infty)$ soit tournée vers le passé (\Rightarrow module des racines de Φ strictement supérieur à 1) :

$$Y_t = \Phi(L)^{-1}\Theta(L)\varepsilon_t = \sum_{j=0}^{\infty} h_j \varepsilon_{t-j}$$

Y_{t-i} non corrélé à ε_t

\implies Alors, ε_t est le processus d'innovation de Y_t .

Remarques préliminaires: inversion d'un polynôme

Soit le polynôme $1 - aL$. Comment calculer son inverse $(1 - aL)^{-1}$?

- Si $|a| < 1$, donc si la racine de $1 - az = 0$ est supérieure à 1 en valeur absolue $|z| = \frac{1}{|a|} > 1$, alors l'inverse est donnée par :

$$(1 - aL)^{-1} = \sum_{i=0}^{\infty} a^i L^i = 1 + aL + a^2 L^2 + \dots$$

puisque

$$\sum_{i=0}^{\infty} a^i L^i (1 - aL) = \sum_{i=0}^{\infty} a^i L^i - \sum_{i=1}^{\infty} a^i L^i = a^0 L^0 = 1$$

• Si $|a| > 1$, donc si la racine de $1 - aL = 0$ est inférieure à 1 en valeur absolue $|z| = \frac{1}{|a|} < 1$, alors l'inverse est donnée par :

$$(1 - aL)^{-1} = - \sum_{i=1}^{\infty} a^{-i} L^{-i}$$

puisque

$$- \sum_{i=1}^{\infty} a^{-i} L^{-i} (1 - aL) = - \sum_{i=1}^{\infty} a^{-i} L^{-i} + \sum_{i=1}^{\infty} a^{1-i} L^{1-i} = a^0 L^0 = 1$$

5.1 Processus MA

5.1.1 Définition

On appelle processus moyenne mobile d'ordre q , noté $MA(q)$ pour *Moving Average*, un processus $\{Y_t\}_{t \in \mathbb{Z}}$ défini par :

$$Y_t = \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad \forall t \in \mathbb{Z}$$

où les θ_i sont des réels, $\theta_q \neq 0$ et $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ est un processus bruit blanc de variance σ^2 .

$$\iff Y_t = (1 - \theta_1 L - \dots - \theta_q L^q) \varepsilon_t \quad \iff Y_t = \Theta(L) \varepsilon_t$$

5.1.2 Représentation stationnaire et causale

La définition d'un $MA(q)$ est explicite et ne pose donc pas de problème : le processus Y_t est parfaitement défini et est automatiquement stationnaire.

La représentation est causale par définition.

5.1.3 Représentation inversible

$\Theta(L)$ est un polynôme en L de degré q , que l'on peut factoriser en ayant calculé ses racines $z_i = 1/\lambda_i$, $i = 1, \dots, q$:

$$\Theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q = (1 - \lambda_1 L)(1 - \lambda_2 L) \dots (1 - \lambda_q L)$$

Si $\Theta(z)$ n'a pas de racine de module égal à 1, on peut calculer l'inverse de $\Theta(L)$, qui est alors donnée par :

$$\Theta(L)^{-1} = (1 - \lambda_1 L)^{-1} (1 - \lambda_2 L)^{-1} \dots (1 - \lambda_q L)^{-1} = \frac{k_1}{1 - \lambda_1 L} + \dots + \frac{k_q}{1 - \lambda_q L}$$

si toutes les racines de Θ (c'est-à-dire $1/\lambda_i$, $i = 1, \dots, q$) sont distinctes et où k_1, \dots, k_q sont des paramètres qui dépendent de $\lambda_1, \dots, \lambda_q$.

On obtient alors l'expression suivante :

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

où les π sont fonctions des paramètres θ_j , et on peut montrer que $\sum_{i=-\infty}^{\infty} |\pi_i| < \infty$.

- Si les racines de $\Theta(z) = 0$ sont toutes de module supérieur à 1, on peut montrer que $\pi_i = 0 \quad \forall i < 0$ et ε_t s'interprète comme l'innovation du processus. On dit que le processus est inversible.
- Si les racines de $\Theta(z) = 0$ sont inférieures à 1 en module, mais qu'il n'y a pas de racines de module égal à 1, on peut inverser les racines, quitte à changer de bruit blanc, et supposer que le processus est inversible.

On a alors la forme $AR(\infty)$ suivante :

$$Y_t = - \sum_{i=1}^{\infty} \pi_i Y_{t-i} + \varepsilon_t \quad \sum_{i=1}^{\infty} |\pi_i| < \infty$$

les π_i s'obtenant par division croissante de 1 par Θ (avec $\pi_0 = 1$).

Ainsi, Y_t dépend de ε_t et de son passé, donc Y_{t-1} dépend de ε_{t-1} et de son passé, cet ensemble étant indépendant de ε_t (puisque ε_t est un BB), donc le passé de Y_t est indépendant de ε_t , donc ε_t est l'innovation de Y_t .

Exemple du MA(1) : $Y_t = \varepsilon_t - \theta\varepsilon_{t-1} = (1 - \theta L)\varepsilon_t$ avec ε_t BB $(0, \sigma_\varepsilon^2)$

La racine de $1 - \theta z = 0$ est $z = 1/\theta$.

- Si $|\theta| = 1$ alors la forme $AR(\infty)$ n'existe pas, la représentation est stationnaire, causale mais non inversible, donc non canonique.
- Si $|z| > 1$ et donc $|\theta| < 1$ alors :

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

et la forme $AR(\infty)$ s'écrit :

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

La représentation est alors inversible, en plus d'être stationnaire et causale \implies elle est donc canonique.

- Si $|z| < 1$ et donc $|\theta| > 1$ alors :

$$\varepsilon_t = (1 - \theta L)^{-1} Y_t = - \sum_{i=1}^{\infty} \theta^{-i} L^{-i} Y_t = - \left(\frac{Y_{t+1}}{\theta} + \frac{Y_{t+2}}{\theta^2} + \dots \right)$$

Dans ce cas, on peut toujours (tant que $|\theta| \neq 1$) se ramener à une représentation canonique quitte à changer la représentation et surtout à changer de BB et à inverser les racines.

En effet, la fonction génératrice des moments du MA(1) initial est donnée par :

$$\begin{aligned} \Gamma_Y(z) &= (1 - \theta z)(1 - \theta z^{-1}) \\ &= \theta z \left(\frac{1}{\theta z} - 1 \right) \left(\frac{1}{\theta z^{-1}} - 1 \right) \theta z^{-1} \\ &= \theta^2 \left(1 - \frac{1}{\theta} z^{-1} \right) \left(1 - \frac{1}{\theta} z \right) \end{aligned}$$

soit la fonction génératrice des moments d'un MA(1) où la racine est $1/\theta$ et le nouveau BB (η_t) est de variance $\theta^2 \sigma_\varepsilon^2$, soit une représentation de la forme :

$$Y_t = \eta_t - \frac{1}{\theta} \eta_{t-1}$$

Une autre manière de faire est de passer par la fonction d'autocovariance, en identifiant les paramètres de la représentation canonique :

$$Y_t = \varepsilon_t - \theta \varepsilon_{t-1} = \eta_t - \alpha \eta_{t-1}$$

Ainsi,

$$\begin{aligned} \gamma(0) &= (1 + \theta^2) \sigma_\varepsilon^2 = (1 + \alpha^2) \sigma_\eta^2 \\ \gamma(1) &= -\theta \sigma_\varepsilon^2 = -\alpha \sigma_\eta^2 \end{aligned}$$

Les valeurs solutions de α vérifient alors :

$$\alpha^2 \theta - \alpha(1 + \theta^2) + \theta = 0$$

on obtient alors $\alpha = 1/\theta$ ou $\alpha = \theta$. On choisit la solution qui est inférieure à 1 en module, soit $\alpha = 1/\theta$ puisqu'on est dans le cas $|\theta| > 1$. On a alors $\alpha = 1/\theta$ et $\sigma_\eta^2 = \theta^2 \sigma_\varepsilon^2$.

Conclusion : la représentation $Y_t = \Theta(L)\varepsilon_t$ est canonique si les racines de $\Theta(z)$ sont supérieures à 1 en module.

5.2 Processus AR

5.2.1 Définition

On appelle processus autorégressif d'ordre p , noté $AR(p)$, un processus stationnaire $\{Y_t\}_{t \in \mathbb{Z}}$ vérifiant une relation du type :

$$Y_t - \sum_{i=1}^p \phi_i Y_{t-i} = \varepsilon_t \quad \forall t \in \mathbb{Z}$$

où les ϕ_i sont des réels, $\phi_p \neq 0$ et $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ est un bruit blanc de variance σ^2 .

$$\iff (1 - \phi_1 L - \dots - \phi_p L^p) Y_t = \varepsilon_t \quad \iff \Phi(L) Y_t = \varepsilon_t$$

5.2.2 Représentation stationnaire

Ce processus est pour l'instant défini sous forme implicite et en particulier il n'est pas certain que cette dernière équation admette toujours une solution stationnaire.

Si le polynôme Φ a toutes ses racines de module différent de 1, on peut inverser l'opérateur $\Phi(L)$. On en déduit que l'équation admet une solution unique, avec une écriture $MA(\infty)$:

$$Y_t = \Phi(L)^{-1} \varepsilon_t = \sum_{i=-\infty}^{\infty} h_i \varepsilon_{t-i}$$

On peut alors montrer que l'on a $\sum_{i=-\infty}^{\infty} |h_i| < \infty$ et donc que la représentation est stationnaire.

5.2.3 Représentation inversible

La représentation $AR(p)$ est inversible par définition.

5.2.4 Représentation causale

• Si le polynôme Φ a toutes ses racines de module strictement supérieur à 1, l'opérateur inverse $\Phi(L)^{-1}$ admet un développement ne faisant intervenir que les puissances positives de L . On a alors :

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

Dans ce cas, on montre que l'on a

$$\sum_{i=0}^{\infty} |h_i| < \infty, \quad h_0 = 1$$

Dans ce cas, Y_{t-1}, Y_{t-2}, \dots sont fonctions linéaires de $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots$ et en particulier sont non corrélés avec ε_t .

Projetant la relation AR sur le passé de X , on obtient :

$$E(Y_t/Y_{t-1}, Y_{t-2}, \dots) = \sum_{i=1}^p \phi_i Y_{t-i}$$

ainsi, le bruit blanc est aussi l'innovation puisque :

$$Y_t - E(Y_t/Y_{t-1}, Y_{t-2}, \dots) = \varepsilon_t$$

- Lorsque les racines de $\Phi(z)$ sont de module différent de 1, on peut montrer que, quitte à changer de bruit blanc, on peut toujours supposer que ces racines sont de module supérieur à 1.

Mais si certaines racines de $\Phi(z) = 0$ sont de module inférieur à 1, alors ε_t n'est pas l'innovation, puisque la forme $MA(\infty)$ sera tournée vers le futur (et peut-être aussi vers le passé si certaines racines étaient bien supérieur à 1 en module). Dans ce cas, le passé de Y_t dépend du passé et du futur de ε_t : on ne peut plus dire qu'il n'y a pas de corrélation entre le passé de Y_t et ε_t ! Ainsi, ce n'est pas ε_t qui est l'innovation du processus.

Exemple de l'AR(1) : $Y_t = \phi Y_{t-1} + \varepsilon_t$ avec ε_t BB $(0, \sigma_\varepsilon^2)$

La racine de $\Phi(z) = 1 - \phi z$ est $z = 1/\phi$.

- Si $|z| = 1$ et donc $|\phi| = 1$, par exemple $\phi = 1$, alors :

$$Y_t = Y_{t-1} + \varepsilon_t = \dots = Y_{t-n} + \varepsilon_{t-n+1} + \dots + \varepsilon_{t-1} + \varepsilon_t$$

et la variance dépend de t ! donc il n'existe pas de solution stationnaire.

Il s'agit d'un processus non stationnaire de type stochastique, appelé marche aléatoire (*random walk*).

- Si $|z| > 1$ et donc $|\phi| < 1$ alors la forme $MA(\infty)$ est donnée par :

$$Y_t = (1 - \phi L)^{-1} \varepsilon_t = \sum_{i=0}^{\infty} \phi^i L^i \varepsilon_t = \varepsilon_t + \sum_{i=1}^{\infty} \phi^i \varepsilon_{t-i}$$

La représentation est alors causale, en plus d'être stationnaire et inversible \implies elle est donc canonique. ε_t est le processus d'innovation de Y_t , puisque le passé de Y_t dépend du passé de ε_t et donc n'est pas corrélé à ε_t .

- Si $|z| < 1$ et donc $|\phi| > 1$ alors :

$$Y_t = (1 - \phi L)^{-1} \varepsilon_t = - \sum_{i=1}^{\infty} \phi^{-i} L^{-i} \varepsilon_t = - \left(\frac{\varepsilon_{t+1}}{\phi} + \frac{\varepsilon_{t+2}}{\phi^2} + \dots \right)$$

La forme $MA(\infty)$ n'est pas tournée vers le passé et la représentation n'est donc pas canonique.

Cependant, dans ce cas, on peut toujours (tant que $|\phi| \neq 1$) se ramener à une représentation canonique, en changeant la représentation, c'est-à-dire en changeant de BB et en inversant les racines.

En effet, la fonction génératrice des moments de l'AR(1) initial est donnée par :

$$\begin{aligned}\Gamma_Y(z) &= \frac{1}{(1 - \phi z)(1 - \phi z^{-1})} \\ &= \frac{1}{\phi z \left(\frac{1}{\phi z} - 1\right) \left(\frac{1}{\phi z^{-1}} - 1\right) \phi z^{-1}} \\ &= \frac{1}{\phi^2 \left(1 - \frac{1}{\phi} z^{-1}\right) \left(1 - \frac{1}{\phi} z\right)}\end{aligned}$$

soit la fonction génératrice des moments d'un AR(1) où la racine est $1/\phi$ et le nouveau BB (η_t) est de variance $\frac{\sigma^2}{\phi^2}$, soit une représentation de la forme :

$$Y_t = \frac{1}{\phi} Y_{t-1} + \eta_t$$

Conclusion : la représentation $\Phi(L)Y_t = \varepsilon_t$ est canonique si les racines de $\Phi(L) = 0$ sont supérieures à 1 en module.

5.3 Processus ARMA

5.3.1 Définition

Les processus ARMA généralisent simultanément les modèles AR purs et les MA purs. Ces modèles présentent l'avantage d'être plus souples d'utilisation et de fournir généralement de bonnes approximations des séries réelles avec moins de paramètres que les modèles AR ou MA purs.

Définition : Un processus stationnaire Y_t admet une représentation $ARMA(p, q)$ minimale :

$$Y_t - \sum_{i=1}^p \phi_i Y_{t-i} = \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i} \iff \Phi(L)Y_t = \Theta(L)\varepsilon_t$$

s'il satisfait les conditions suivantes :

(i) $\phi_p \neq 0, \theta_q \neq 0$

(ii) Les polynômes Φ et Θ ont toutes leurs racines de module strictement supérieur à 1

(iii) Φ et Θ n'ont pas de racine commune

(iv) ε_t est un bruit blanc, de variance $\sigma^2 \neq 0$

Remarquons qu'on aurait pu définir une représentation plus générale permettant de considérer des processus stationnaires non centrés. Par exemple :

$$Y_t - \sum_{i=1}^p \phi_i Y_{t-i} = \theta^* + \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i}$$

On se ramène directement au cas défini précédemment en remplaçant Y_t par $Y_t - E(Y_t) = Y_t - \frac{\theta^*}{1 - \sum_j \phi_j} = \frac{\theta^*}{\Phi(1)}$.

La condition (ii) assure tout d'abord que la représentation ARMA admet une solution stationnaire (si les racines de Φ sont de module différent de 1), que cette solution stationnaire $MA(\infty)$ fait intervenir que des valeurs passées du bruit (les racines de Φ sont à l'extérieur du disque unité), que la représentation $AR(\infty)$ ne fait intervenir que des valeurs présentes et passées de Y_t (les racines de Θ sont de module strictement supérieur à 1). Ainsi, ε_t est le processus d'innovation du processus Y_t .

La condition (iii) assure que la représentation est unique, sinon il y aurait des simplifications possibles.

5.3.2 Propriété

Si Y est un processus stationnaire de représentation $ARMA(p, q)$ minimale $\Phi(L)Y_t = \Theta(L)\varepsilon_t$, alors :

(i) Y_t admet la représentation $MA(\infty)$:

$$Y_t = \frac{\Theta(L)}{\Phi(L)}\varepsilon_t = \sum_{j=0}^{\infty} h_j \varepsilon_{t-j}, \quad h_0 = 1$$

(ii) Y_t admet la représentation $AR(\infty)$:

$$\frac{\Phi(L)}{\Theta(L)}Y_t = \sum_{j=0}^{\infty} \pi_j Y_{t-j} = \varepsilon_t, \quad \pi_0 = 1$$

(iii) Y_t admet pour innovation ε_t

6 Caractéristiques des processus ARMA: autocorrélation simple, partielle et densité spectrale

On va étudier les caractéristiques propres aux processus AR, MA et ARMA en termes d'autocorrélations simple et partielle afin de se donner un schéma d'identification.

6.1 Corrélogramme simple

On rappelle que la fonction d'autocorrélation simple est donnée par :

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{E(Y_t Y_{t-h})}{E(Y_t^2)}$$

6.1.1 Cas de l'AR(p)

Soit le processus AR(p) :

$$Y_t - \sum_{i=1}^p \phi_i Y_{t-i} = \varepsilon_t \quad \forall t \in \mathbb{Z} \quad (2)$$

Pour calculer sa variance, on multiplie cette équation par Y_t et on en prend son espérance :

$$E(Y_t^2) - \sum_{i=1}^p \phi_i E(Y_t Y_{t-i}) = E(Y_t \varepsilon_t)$$

on obtient alors :

$$\gamma(0) = \sum_{i=1}^p \phi_i \gamma(i) + \sigma^2$$

Ainsi, puisque $\gamma(i) = \gamma(0)\rho(i)$

$$\gamma(0) = \frac{\sigma^2}{1 - \sum_{i=1}^p \phi_i \rho(i)}$$

Pour obtenir les autocovariances, on multiplie l'équation (2) par Y_{t-h} et on en prend son espérance. On obtient alors les équations suivantes :

$$\gamma(h) - \sum_{i=1}^p \phi_i \gamma(h-i) = 0 \quad \forall h > 0$$

Les $p+1$ premières équations sont appelées équations de Yule-Walker.

En divisant par $\gamma(0)$, on obtient alors :

$$\rho(h) - \sum_{i=1}^p \phi_i \rho(h-i) = 0 \quad \forall h > 0$$

soit une équation de récurrence linéaire homogène d'ordre p dont le polynôme caractéristique est :

$$z^h - \sum_{i=1}^p \phi_i z^{h-i} = z^h (1 - \phi_1 z^{-1} - \dots - \phi_p z^{-p}) = 0 \quad (3)$$

Si le polynôme $\Phi(L) = 1 - \sum_{i=1}^p \phi_i L^i$ admet p racines distinctes $1/\lambda_1, 1/\lambda_2, \dots, 1/\lambda_p$, alors les racines de l'équation (3) sont $\lambda_1, \lambda_2, \dots, \lambda_p$ et le coefficient d'autocorrélation d'ordre h est alors donné par :

$$\rho(h) = a_1(\lambda_1)^h + a_2(\lambda_2)^h + \dots + a_p(\lambda_p)^h$$

où les $a_i, i = 1, \dots, p$ sont des constantes déterminées par les conditions initiales.

Puisque la représentation est canonique, les racines $1/\lambda_i, i = 1, \dots, p$ sont alors de module supérieur à 1 donc $|\lambda_i| < 1$.

Ainsi, la fonction d'autocorrélation (d'un processus stationnaire) décroît, soit de manière exponentielle (si les racines sont réelles), soit selon des cycles amortis (si les racines sont complexes).

6.1.2 Cas du MA(q)

Calculons les autocovariances pour le processus MA(q) :

$$Y_t = \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad \forall t \in \mathbb{Z}$$

On obtient :

$$\begin{aligned} \gamma(h) &= E(Y_t Y_{t-h}) \\ &= E\left(\varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i}\right) \left(\varepsilon_{t-h} - \sum_{i=1}^q \theta_i \varepsilon_{t-h-i}\right) \\ &= \begin{cases} (1 + \sum_{i=1}^q \theta_i^2) \sigma^2 & \text{si } h = 0 \\ (-\theta_h + \sum_{i=h+1}^q \theta_i \theta_{i-h}) \sigma^2 & \text{si } 1 \leq h \leq q-1 \\ -\theta_q \sigma^2 & \text{si } h = q \\ 0 & \text{si } h > q \end{cases} \end{aligned}$$

Ainsi, les autocorrélations simples s'annulent après l'ordre du MA :

$$\rho(h) = 0 \text{ si } h > q$$

et pour $h < q$, sont telles que :

$$\rho(h) = \frac{-\theta_h + \sum_{i=h+1}^q \theta_i \theta_{i-h}}{1 + \sum_{i=1}^q \theta_i^2}$$

6.1.3 Cas de l'ARMA(p,q)

Soit le processus ARMA(p,q) suivant :

$$Y_t - \sum_{i=1}^p \phi_i Y_{t-i} = \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i}$$

On obtient, en multipliant par Y_{t-h} et en prenant l'espérance :

$$\gamma(h) - \sum_{i=1}^p \phi_i \gamma(h-i) = E(Y_{t-h} \varepsilon_t) - \sum_{i=1}^q \theta_i E(Y_{t-h} \varepsilon_{t-i})$$

Or, $E(Y_{t-h}\varepsilon_{t-i}) = 0$ si $t - h < t - i$ pour $i = 1, 2, \dots, q$ donc pour $h > q$. Ainsi, pour $h > q$, on retrouve l'équation de récurrence d'ordre p comme dans le cas de l'AR(p) :

$$\gamma(h) - \sum_{i=1}^p \phi_i \gamma(h-i) = 0 \quad h > q$$

Les p équations $h = q+1, \dots, q+p$ sont appelées équations de Yule-Walker. Elles permettent de calculer les p coefficients $\gamma(h)$ en fonction des p valeurs initiales $\gamma(q), \dots, \gamma(q-p+1)$. Ces équations ne permettent pas de déterminer toutes les valeurs $\gamma(h)$ puisqu'il faut disposer de conditions initiales.

Les premières valeurs de $\gamma(h)$ $h = 0, 1, \dots, q$ sont déterminées par :

$$\gamma(h) - \sum_{i=1}^p \phi_i \gamma(h-i) = E(Y_{t-h}\varepsilon_t) - \sum_{i=1}^q \theta_i E(Y_{t-h}\varepsilon_{t-i})$$

pour $h = 0, 1, \dots, q$. Les termes de droite de l'égalité sont calculés à partir de l'expression $MA(\infty)$ de Y_t .

Ainsi, le corrélogramme d'un processus ARMA(p,q) est le même que celui d'un AR à partir de l'ordre $q+1$.

6.2 Corrélogramme partiel

L'autocorrélation partielle d'ordre h est la corrélation partielle entre Y_t et Y_{t-h} , soit :

$$r(h) = \text{corr}(Y_t, Y_{t-h} | Y_{t-1}, \dots, Y_{t-h+1})$$

C'est le dernier coefficient de la régression de Y_t sur Y_{t-1}, \dots, Y_{t-h} :

$$Y_t = a_1 Y_{t-1} + a_2 Y_{t-2} + \dots + a_h Y_{t-h} + \varepsilon_t$$

6.2.1 Cas de l'AR(p)

Le processus AR(p) peut s'écrire :

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{i=p+1}^{\infty} 0 \times Y_{t-i} + \varepsilon_t$$

Ainsi, les autocorrélations partielles d'un AR(p) sont telles que :

$$r(p) = \phi_p \quad r(h) = 0 \quad \forall h > p$$

Pour les premières autocorrélations, il suffit de les calculer en fonction des $\rho(i)$, $i = 1, \dots, p$:

$$r(1) = \rho(1) \quad r(2) = \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2} \quad \dots$$

6.2.2 Cas du MA(q)

Les autocorrélations partielles décroissent soit de manière exponentielle (si les racines sont réelles) soit selon des cycles amortis (si les racines sont complexes).

6.2.3 Cas de l'ARMA(p,q)

Le corrélogramme partiel d'un ARMA(p,q) est le même que celui d'un MA à partir de l'ordre $p + 1$.

6.3 L'autocorrélation inverse

C'est l'autocorrélation associée à la fonction inverse de la densité spectrale d'un processus.

Soit Y_t un processus stationnaire, sa fonction génératrice est donnée par :

$$\Gamma(L) = \sum_h \gamma(h)L^h$$

La fonction génératrice d'autocorrélation inverse $\Gamma^I(L)$ est telle que :

$$\begin{aligned} \Gamma^I(L)\Gamma(L) &= 1 \\ \Gamma^I(L) &= \sum_h \Gamma^I(h)L^h \end{aligned}$$

Ainsi, pour un AR(p) :

$$\Gamma^I(h) = 0 \quad h > p$$

et pour un MA(q), on retrouve le même corrélogramme simple que celui d'un AR.

6.4 Tableau récapitulatif

	AR(p)	MA(q)
autocorrélations simples	décroissent de manière exponentielle	$\rho(h) = 0$ si $ h > q$ $\rho(h) \neq 0$ si $ h \leq q$
autocorrélations partielles	$r(h) = 0$ si $ h > p$ $r(h) \neq 0$ si $ h \leq p$	décroissent de manière exponentielle
autocorrélations inverses	$\rho^i(h) = 0$ si $ h > p$ $\rho^i(h) \neq 0$ si $ h \leq p$	décroissent de manière exponentielle

6.5 Densité spectrale

Soit Y_t un processus ARMA(p,q) :

$$\Phi(L)Y_t = \Theta(L)\varepsilon_t$$

la densité spectrale est :

$$f(\omega) = \frac{\sigma^2 \Theta(z)\Theta(z^{-1})}{2\pi \Phi(z)\Phi(z^{-1})}$$

où $z = e^{i\omega}$.

Soit Y_t un processus MA(1) définit par :

$$Y_t = \varepsilon_t - \theta\varepsilon_{t-1} \quad |\theta| < 1$$

On a $\Theta(z) = 1 - \theta z$. La densité spectrale est :

$$\begin{aligned} f(\omega) &= \frac{\sigma^2}{2\pi} (1 - \theta e^{i\omega})(1 - \theta e^{-i\omega}) \\ &= \frac{\sigma^2}{2\pi} (1 - \theta(e^{i\omega} + e^{-i\omega}) + \theta^2) \\ &= \frac{\sigma^2}{2\pi} (1 - 2\theta \cos \omega + \theta^2) \end{aligned}$$

La densité spectrale est donc une fonction croissante si $\theta > 0$ et décroissante si $\theta < 0$ puisque :

$$f'(\omega) = \frac{\sigma^2}{2\pi} (2\theta \sin \omega)$$

Soit Y_t AR(1) définit par :

$$Y_t = \phi Y_{t-1} + \varepsilon_t \quad |\phi| < 1$$

On a $\Phi(z) = (1 - \phi z)$. La densité spectrale est donc donnée par :

$$\begin{aligned} f(\omega) &= \frac{\sigma^2}{2\pi} (1 - \phi e^{i\omega})^{-1} (1 - \phi e^{-i\omega})^{-1} \\ &= \frac{\sigma^2}{2\pi} (1 - 2\phi \cos \omega + \phi^2)^{-1} \end{aligned}$$

Ainsi, la densité spectrale est croissante si $\phi < 0$ et décroissante si $\phi > 0$ puisque :

$$f'(\omega) = \frac{\sigma^2}{2\pi} \frac{-2\phi \sin \omega}{(1 - 2\phi \cos \omega + \phi^2)^2}$$

7 Identification, estimation, validation, prévision

7.1 Identification du processus ARMA

La méthode d'identification d'un processus ARMA (choix entre AR, MA et ARMA, et choix de p et q) de Box et Jenkins est basée sur la comparaison de caractéristiques

théoriques des processus ARMA à leurs équivalents empiriques (c'est-à-dire calculées sur la série observée). Les caractéristiques utilisées sont les autocorrélations simple et partielle (et éventuellement autocorrélation inverse), étudiées dans la partie précédente.

On peut aussi utiliser des critères de choix de modèle, couramment appelé critères d'information. Les plus couramment utilisés sont le critère de Akaike :

$$AIC = -2 \log \mathcal{L} + 2(p + q)$$

et le critère de Schwarz :

$$BIC = -2 \log \mathcal{L} + (p + q) \log T$$

où $\log \mathcal{L}$ est la log-vraisemblance du modèle ARMA(p,q) estimé et T est le nombre d'observations.

On choisit alors le modèle et p et q qui minimisent ces critères.

7.2 Estimation

Etant donné le processus Y_t , $t \in \mathbb{Z}$ admettant une représentation ARMA(p,q) :

$$Y_t - \sum_{i=1}^p \phi_i Y_{t-i} = \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad \varepsilon_t \sim BB(0, \sigma^2)$$

le problème est d'estimer les paramètres ϕ_i ($i = 1, \dots, p$), θ_j ($j = 1, \dots, q$) et σ^2 à partir d'observations Y_1, \dots, Y_T .

7.2.1 Maximum de vraisemblance

Si le processus ε_t est gaussien $N(0, \sigma^2)$, alors $\varepsilon \sim N(0, \sigma^2 I)$ et $Y \sim N(0, \sigma^2 \Omega)$. Ainsi, la vraisemblance (densité du vecteur $Y = (Y_1, \dots, Y_T)'$) est donnée par :

$$\mathcal{L}(Y_1, \dots, Y_T; \phi, \theta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{T/2}} \frac{1}{(\det\Omega)^{1/2}} \exp\left(-\frac{1}{2\sigma^2} Y' \Omega^{-1} Y\right)$$

La log-vraisemblance est alors donnée par :

$$\ln \mathcal{L}(Y_1, \dots, Y_T; \phi, \theta, \sigma^2) = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln \sigma^2 - \frac{1}{2} \ln(\det\Omega) - \frac{1}{2\sigma^2} Y' \Omega^{-1} Y$$

Le problème est que cette log-vraisemblance est difficile à calculer et donc à maximiser à cause de $(\det\Omega)$ et de Ω^{-1} (matrice $T \times T$). De plus, il faut se donner des valeurs préliminaires pour les paramètres, puisque la maximisation de la log-vraisemblance utilise des algorithmes de maximisation itératifs.

Les h_j sont estimés à partir d'un algorithme récursif appelé algorithme de Durbin-Levinson ou algorithme des innovations. A partir de ces estimations, on peut alors obtenir des estimations de ϕ et de θ .

Ceci se généralise au cas de p et q quelconques.

7.2.4 Moindres carrés conditionnels

Cette estimation (estimation par défaut sous le logiciel SAS) est conditionnelle à l'hypothèse que les erreurs non observées passées sont égales à 0. L'estimation repose sur la forme $AR(\infty)$:

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

où les π_i dépendent des paramètres ϕ et θ .

L'estimation CLS consiste à minimiser la somme des carrés suivante :

$$\sum_{t=1}^T \varepsilon_t^2 = \sum_{t=1}^T (Y_t - \sum_{i=1}^{\infty} \pi_i Y_{t-i})^2$$

où les valeurs passées de Y_t inconnues sont posées à 0 et les valeurs de π sont calculées à chaque étape à partir des estimations de ϕ et de θ .

7.3 Validation

Il s'agit de vérifier notamment que les résidus du modèle ARMA estimé, résidus notés $\hat{\varepsilon}_t$, vérifient les propriétés requises pour que l'estimation soit valide, à savoir qu'ils suivent un processus BB, non autocorrélé et de même variance, et qu'ils suivent une loi normale. Si ces hypothèses ne sont pas rejetées, on peut alors mener des tests sur les paramètres.

7.3.1 Tests sur les résidus

- Regarder le graphique des résidus estimés pour voir s'il apparaît des points aberrants, une tendance, une rupture, de l'autocorrélation, etc. Ceci n'est évidemment qu'indicatif.
- Regarder les autocorrélations simples et partielles. Elles doivent être significativement nulles si les résidus sont un bruit blanc.
- Test du *portemanteau*

Afin de tester que les résidus estimés suivent un BB, on teste l'hypothèse d'absence d'autocorrélation jusqu'à l'ordre m . On utilise la statistique de Ljung-Box, donnée par :

$$LB = T(T+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{T-k}$$

où les coefficients d'autocorrélation $\hat{\rho}_k$ sont calculés sur les résidus estimés $\hat{\varepsilon}_t$.

Cette statistique, sous l'hypothèse que les résidus suivent un BB, suit une loi du χ_m^2 .

- Test d'homoscédasticité

Un test couramment utilisé en séries temporelles est le test d'homoscédasticité contre une alternative ARCH (hétérosécédasticité conditionnelle dans la variance). On teste alors la nullité des paramètres a_1, \dots, a_q dans le modèle :

$$\hat{\varepsilon}_t^2 = a_0 + a_1 \hat{\varepsilon}_{t-1}^2 + \dots + a_q \hat{\varepsilon}_{t-q}^2 + \eta_t$$

La statistique de test est TR^2 , T étant le nombre d'observations et R^2 le coefficient de détermination du modèle ci-dessus. Cette statistique suit, sous l'hypothèse nulle d'homoscédasticité une loi du χ_q^2 .

- Test de normalité

Il s'agit de tester que les résidus estimés $\hat{\varepsilon}_t$ suivent une loi normale, c'est-à-dire ne présentent pas d'asymétrie (*Skewness*) ni d'applatissage (*kurtosis*).

Le coefficient de Skewness est donné par :

$$\beta_1^{1/2} = \frac{\mu_3}{\mu_2^{3/2}}$$

et le coefficient de kurtosis est donné par :

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

où

$$\mu_k = \frac{1}{T} \sum_{t=1}^T (\hat{\varepsilon}_t - \bar{\hat{\varepsilon}})^k$$

est le moment centré d'ordre k de la variable $\hat{\varepsilon}_t$.

Si la distribution est normale et le nombre d'observations grand, alors :

$$\beta_1^{1/2} \sim N\left(0, \sqrt{6/T}\right) \quad \beta_2 \sim N\left(3, \sqrt{24/T}\right)$$

On construit alors les statistiques :

$$\nu_1 = \frac{\beta_1^{1/2}}{\sqrt{6/T}} \quad \text{et} \quad \nu_2 = \frac{\beta_2 - 3}{\sqrt{24/T}}$$

qui suivent chacune une $N(0, 1)$.

Le test de Jarque Bera permet de tester simultanément l'absence d'asymétrie et l'absence d'applatissage. La statistique de test est donnée par :

$$JB = \frac{T}{6} \beta_1 + \frac{T}{24} (\beta_2 - 3)^2$$

Cette statistique suit, sous l'hypothèse nulle de normalité, une loi du χ_2^2 .

7.3.2 Tests sur les paramètres

On vérifie tout d'abord que les racines des polynômes AR et MA ne sont pas égales à 1.

Si les hypothèses testées sur les résidus ne sont pas rejetées, on teste la significativité des retards du modèle ARMA par des tests de Student.

7.3.3 Choix d'un modèle parmi plusieurs

Si, à la suite de ces étapes, il reste plusieurs modèles valides, on peut choisir parmi ces modèles, soit celui qui donne les meilleurs critères d'ajustement, soit celui qui donne les meilleurs performances en prévision.

Concernant les critères d'ajustement, on retient le modèle qui minimise les critères d'information AIC et BIC.

Concernant les performances en prévision des modèles, on utilise couramment les critères suivants (que l'on cherche bien entendu à minimiser) :

$$RMSE(h) = \sqrt{\frac{1}{T - K + 1} \sum_{t=K}^T (Y_t - \hat{Y}_t(h))^2}$$

et

$$MAE(h) = \frac{1}{T - K + 1} \sum_{t=K}^T |Y_t - \hat{Y}_t(h)|$$

où K est le nombre d'observations minimales pour mener une estimation du modèle.

On peut calculer ces critères, soit sur la base de prévisions *in-sample* (toutes les observations ont été utilisées pour estimer le modèle et on calcule les prévisions sur cet même ensemble d'observations), soit sur la base de prévisions *out-of-sample* (on estime le modèle sur un ensemble d'observations et on mène la prévision sur le reste).

7.4 Prévision

Il s'agit de calculer les prévisions optimales du modèle ARMA estimé, à savoir $\hat{Y}_t(k)$ la prévision de Y_{t+k} sachant l'ensemble d'information disponible en t , noté $I_t = \{Y_1, Y_2, \dots, Y_t, Z_{-1}\}$ où $Z_{-1} = \{Y_{-1}, \dots, Y_{-p}, \varepsilon_{-1}, \dots, \varepsilon_{-q}\}$:

$$\hat{Y}_t(k) = E(Y_{t+k} | I_t)$$

7.4.1 Calcul des prévisions optimales et mise à jour

A partir de la forme $MA(\infty)$:

On peut réécrire la forme MA(∞) de la manière suivante :

$$Y_{t+k} = \sum_{j=0}^{t+k} h_j \varepsilon_{t+k-j} + h^*(t+k)Z_{-1}$$

alors la prévision optimale est donnée par :

$$\hat{Y}_t(k) = \sum_{j=k}^{t+k} h_j \varepsilon_{t+k-j} + h^*(t+k)Z_{-1}$$

Ainsi, l'erreur de prévision est :

$$e_t(k) = Y_{t+k} - \hat{Y}_t(k) = \sum_{j=0}^{k-1} h_j \varepsilon_{t+k-j}$$

Cette formule de calcul de la prévision optimale n'est pas directement utilisable puisqu'elle fait intervenir les erreurs non observables, mais elle permet d'obtenir une formule de mise à jour intéressante :

$$\hat{Y}_{t+1}(k-1) - \hat{Y}_t(k) = \sum_{j=k-1}^{t+k} h_j \varepsilon_{t+k-j} - \sum_{j=k}^{t+k} h_j \varepsilon_{t+k-j} = h_{k-1} \varepsilon_{t+1} = h_{k-1} (Y_{t+1} - \hat{Y}_t(1)) \quad (4)$$

A partir de la forme AR(∞) :

On peut réécrire la forme AR(∞) de la manière suivante :

$$Y_{t+k} = - \sum_{j=1}^{t+k} \pi_j Y_{t+k-j} + \pi^*(t+k)Z_{-1} + \varepsilon_{t+k}$$

alors la prévision optimale est donnée par :

$$\hat{Y}_t(k) = - \sum_{j=1}^{t+k} \pi_j \hat{Y}_{t+k-j} \quad (5)$$

où

$$\hat{Y}_{t+k-j} = \begin{cases} \hat{Y}_t(k-j) & \text{si } k > j \\ Y_{t+k-j} & \text{si } k \leq j \end{cases}$$

A partir de la forme ARMA :

$$Y_{t+k} = \sum_{i=1}^p \phi_i Y_{t+k-i} + \varepsilon_{t+k} - \sum_{i=1}^q \theta_i \varepsilon_{t+k-i}$$

On obtient :

$$\hat{Y}_t(k) = \sum_{i=1}^p \phi_i \hat{Y}_{t+k-i} - \sum_{i=1}^q \theta_i \hat{\varepsilon}_{t+k-i} \quad (6)$$

où

$$\hat{\varepsilon}_{t+k-i} = \begin{cases} 0 & \text{si } k > i \\ \varepsilon_{t+k-i} & \text{si } k \leq i \end{cases}$$

On peut utiliser ces différentes manières de calculer la prévision optimale de manière jointe :

- a) En T , on calcule $\hat{Y}_T(k)$, $k = 1, \dots, K$ avec l'équation (5)
- b) En $T + 1$, on connaît Y_{T+1} , on peut alors calculer $\hat{Y}_{T+1}(k)$, $k = 1, \dots, K - 1$ avec l'équation (4) :

$$\hat{Y}_{T+1}(1) \text{ en fonction de } Y_{T+1}, \hat{Y}_T(1), \hat{Y}_T(2)$$

$$\hat{Y}_{T+1}(2) \text{ en fonction de } Y_{T+1}, \hat{Y}_T(1), \hat{Y}_T(3)$$

$$\hat{Y}_{T+1}(K - 1) \text{ en fonction de } Y_{T+1}, \hat{Y}_T(1), \hat{Y}_T(K)$$

- c) On calcule $\hat{Y}_{T+1}(K)$ avec l'équation (6) si $K > q$ $\hat{\varepsilon}_{t+K-i} = 0$ $i = 1, \dots, q$ et si $K > p$ $\hat{Y}_{t+K-i} = \hat{Y}_t(K - i)$ $i = 1, \dots, p$

Connaissant $\hat{Y}_{T+1}(k)$, $k = 1, \dots, K$, on peut répéter b) et c) pour $T + 2$, etc.

7.4.2 Fonction de prévision

La fonction de prévision décrit comment la prévision $\hat{Y}_t(k)$ varie, pour t fixé, en fonction de k .

D'après l'équation (6), on a :

$$\hat{Y}_t(k) - \sum_{i=1}^p \phi_i \hat{Y}_{k-i} = 0 \quad \forall k > q$$

avec $\hat{Y}_{k-i} = Y_{t+k-i}$ si $k \leq i$.

Il s'agit d'une équation de récurrence d'ordre p dont le polynôme caractéristique est :

$$1 - \sum_{i=1}^p \phi_i z^{-i} = \Phi\left(\frac{1}{z}\right) = 0$$

Avec les mêmes notations que dans la partie présentant les autocorrélations (pour les racines $z_i = 1/\lambda_i$), on obtient :

$$\hat{Y}_t(k) = \sum_{i=1}^p b_i(t) \lambda_i^k$$

avec les valeurs de $b_i(t)$ déterminées par les conditions initiales, appelées dans ce contexte valeurs pivotales, à savoir $\hat{Y}_t(k)$, $k = q, q - 1, \dots, q - p + 1$.

Ainsi, la fonction de prévision d'un processus stationnaire tend vers 0 quand $k \rightarrow \infty$ puisque les racines $z_i = 1/\lambda_i$ sont supérieures à 1 en module.

Exemple de l'AR(1) :

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

Ainsi,

$$\begin{aligned}\hat{Y}_t(1) &= \phi Y_t \\ \hat{Y}_t(k) &= \phi \hat{Y}_t(k-1)\end{aligned}$$

L'équation caractéristique est $1 - \phi z^{-1} = 0$, on obtient $z = \phi$ et donc :

$$\hat{Y}_t(k) = b(t)\phi^k$$

Dans ce cas, la valeur pivotale nous permettant de déterminer $b(t)$ est $\hat{Y}_t(0) = Y_t$. On obtient alors $b(t) = Y_t$ et la fonction de prévision est alors :

$$\hat{Y}_t(k) = \phi^k Y_t$$

La prévision converge vers 0 (qui est la moyenne du processus).

Les fonctions de réponses aux chocs :

Le calcul de la fonction de prévision nous permet d'étudier l'effet d'un choc à une date donnée sur l'évolution du processus Y_t . On parle d'IRF (*Impulse Response Function*) : il s'agit de la réponse de Y_t à un choc à une date donnée.

Dans l'exemple de l'AR(1), supposons qu'avant la date t , le processus était à son état stationnaire (ici 0) et qu'à la date t , on fait un choc de taille ε , soit $Y_t = \varepsilon$. La réponse de Y_t à ce choc temporaire est alors donnée par la fonction de prévision :

$$\hat{Y}_t(k) = \phi^k \varepsilon$$

En général, on calcule l'écart entre la réponse à un choc et la réponse sans choc afin d'appréhender l'effet du choc. On définit alors la fonction de réponse au choc de la manière suivante :

$$IRF(Y_t; \varepsilon) = \phi^k \varepsilon - \phi^k \times 0 = \phi^k \varepsilon$$

La valeur de ϕ est déterminante dans la plus ou moins forte persistance du choc sur le processus Y .