

N° d'ordre : 10 / 2006-M / MT

République Algérienne Démocratique Et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique  
Université des Sciences et de la Technologie Houari Boumediene

Faculté de Mathématiques



MEMOIRE

Présenté pour l'obtention du diplôme de Magistère  
En Mathématiques

Spécialité : Recherche Opérationnelle (Méthodes Stochastiques)

Par

Leila DJEDDOU

Thème

**Mélange de Modèles Autorégressifs  
Conditionnellement Hétéroscédastiques**

Soutenu le 30 septembre 2006, devant le jury composé de :

<b>H. Ait Haddadene</b>	<b>Maître de conférence à l'USTHB.</b>	<b>Président</b>
<b>M. Bentarzi</b>	<b>Professeur à l'USTHB.</b>	<b>Directeur de thèse</b>
<b>A. Aissani</b>	<b>Professeur à l'USTHB.</b>	<b>Examineur</b>
<b>M. Moulai</b>	<b>Maître de conférence, à l'USTHB.</b>	<b>Examineur</b>
<b>A. Aknouche</b>	<b>Maître assistant, à l'USTHB.</b>	<b>Examineur</b>

---

# Remerciements

Je remercie avant tout, Dieu le tout puissant, pour m'avoir permis de présenter ce modeste travail.

Je tiens à exprimer ma profonde gratitude à mon directeur de thèse Professeur Mohamed Bentarzi pour m'avoir fait l'honneur de diriger cette thèse. Qu'il trouve ici mes respectueux remerciements pour sa patience, sa compréhension, et le soutien sans relâche qu'il m'a toujours accordé.

Mes remerciements s'adressent aussi à H. Aït-Haddadene, Maître de conférence à l'USTHB pour avoir accepté de présider le jury.

Mes remerciements vont à M. Moulai, Maître de conférence à l'USTHB, A. Aïssani, Professeur à l'USTHB et A. Aknouche, Maître assistant à l'USTHB qui ont bien voulu me faire l'honneur d'être membre du jury.

Je remercie particulièrement Mustapha Djeddou, Docteur à l'EMP, pour son aide et sa patience.

Je tiens à remercier Youcef Bekara, docteur à SUPELEC (FRANCE), pour son aide.

Je remercie aussi Chun Shan Wong, Professeur à l'université de Hong Kong, pour ses explications.

Je remercie tout les membres de ma famille pour leur soutien durant toute la durée qu'a prit la thèse.

*Je dédie cette thèse  
A ma mère  
A mon père*

---

# Table des matières

Résumé	i
Remerciements	ii
Abréviations	vii
Notations	viii
Table des figures	ix
Introduction	1
<b>1 Mélange de modèles autorégressifs</b>	<b>4</b>
1.1 Introduction . . . . .	5
1.2 Mélange de modèles autorégressifs . . . . .	5
1.3 Étude des modèles MAR . . . . .	6
1.3.1 Étude de la stationnarité . . . . .	6
1.3.2 Fonction d'autocorrélation . . . . .	9
1.3.3 La fonction prédictive au pas $m$ . . . . .	9
1.4 Discussion . . . . .	11
<b>2 Les modèles autorégressifs conditionnellement hétéroscédastiques</b>	<b>13</b>
2.1 L'approche <i>ARCH-GARCH</i> et la modélisation de l'incertitude . . . . .	14
2.2 Les grandes classes de modèles non linéaires . . . . .	16
2.2.1 Modèles bilinéaires . . . . .	17
2.2.2 Modèles autorégressifs exponentiels (modèles <i>EXPAR</i> ) . . . . .	17
2.2.3 Modèles autorégressifs à seuil (modèles <i>TAR</i> ) . . . . .	18
2.3 Modèles <i>ARCH-GARCH</i> . . . . .	18
2.3.1 Présentation des modèles <i>ARCH</i> . . . . .	18

2.3.2	Modèle avec erreur $ARCH(q)$ . . . . .	21
2.3.3	Modèles $GARCH(p, q)$ . . . . .	23
2.4	Quelques extensions des modèles $ARCH-GARCH$ . . . . .	24
2.4.1	Modèles $ARMA - GARCH$ . . . . .	24
2.4.2	Modèles $GARCH - M$ . . . . .	24
2.4.3	Modèles $IGARCH$ . . . . .	25
2.4.4	Modèles $ARCH-GARCH$ asymétriques . . . . .	25
2.4.5	Modèles $EGARCH$ . . . . .	25
2.4.6	Les modèles $MAR - ARCH$ . . . . .	26
<b>3</b>	<b>Mélange de modèles autorégressifs conditionnellement hétéroscédastiques</b>	<b>27</b>
3.1	Introduction . . . . .	28
3.2	Mélange de modèles autorégressifs conditionnellement hétéroscédastiques . . . . .	28
3.3	Etude des modèles $MAR - ARCH$ . . . . .	30
3.3.1	Étude de la stationnarité . . . . .	30
3.3.2	Fonction d'autocorrélation . . . . .	35
3.3.3	Moments de modèles $MAR-ARCH$ . . . . .	36
3.3.4	Fonction prédictive au pas $m$ . . . . .	40
3.4	conclusion . . . . .	40
<b>4</b>	<b>Sélection de modèle et estimation des paramètres dans les modèles <math>MAR-ARCH</math></b>	<b>42</b>
4.1	Introduction . . . . .	43
4.2	Problématique du maximum de la vraisemblance . . . . .	43
4.2.1	Racines multiples et choix de la plus grande racine . . . . .	43
4.2.2	Problèmes associés à la méthode . . . . .	44
4.2.3	Difficultés techniques pour obtenir une racine adéquate . . . . .	44
4.3	L'algorithme EM (Expectation-Maximization) . . . . .	45
4.3.1	La théorie fondamentale de l'algorithme EM . . . . .	45
4.3.2	La convergence de l'algorithme EM . . . . .	49
4.3.3	L'algorithme EM généralisé (GEM) . . . . .	50
4.3.4	L'algorithme GEM basé sur la seule étape de Newton - Raphson . . . . .	50
4.3.5	Propriétés et inconvénients de l'algorithme EM . . . . .	51
4.4	L'application de l'algorithme EM dans l'estimation des paramètres dans le modèle $MAR-ARCH$ . . . . .	51

---

4.5	L'estimation de l'erreur standard . . . . .	56
4.5.1	L'utilisation du <i>bootstrap</i> pour les problèmes statistique liées à l'estimation des paramètres . . . . .	56
4.5.2	L'application du <i>bootstrap</i> dans l'estimation de l'erreur standard des estimations des paramètres obtenues par l'algorithme EM . . . . .	60
4.6	Critère de la sélection du modèle <i>MAR – ARCH</i> . . . . .	60
4.7	Etude de simulation . . . . .	61
4.7.1	Etude de simulation pour le BIC . . . . .	61
4.7.2	Etude de simulation pour l'algorithme EM . . . . .	62
<b>5</b>	<b>Application sur une série de données réelles</b>	<b>67</b>
5.1	Présentation de la série test . . . . .	68
5.2	Présentation d'un autre modèle candidat . . . . .	69
5.2.1	Ajustement du modèle MAR-ARCH . . . . .	69
5.2.2	La variance conditionnelle . . . . .	70
5.2.3	Présentation de quelques distributions prédictives . . . . .	70
5.3	Comparaison du modèle MAR-ARCH avec d'autre modèles ajustés . . . . .	70
	<b>Conclusion générale</b>	<b>76</b>
	<b>Bibliographie</b>	<b>78</b>
<b>A</b>	<b>Les modèles GMTD</b>	<b>83</b>
A.1	Description du modèle GMTD . . . . .	83
A.2	Stationnarité . . . . .	84
A.3	La fonction d'autocorrélation . . . . .	84
A.4	Estimation et sélection du modèle . . . . .	85
A.5	Exemples . . . . .	85
A.5.1	Exemple 1 . . . . .	85
A.5.2	Exemple 2 . . . . .	86
<b>B</b>	<b>La série de données</b>	<b>88</b>
<b>C</b>	<b>Programmes</b>	<b>89</b>

---

# Abréviations

ARCH	AutoRegressive Conditional Heteroscedastic
AR-ARCH	AutoRégressif a erreur ARCH
ARMA	AutoRegressive Moving Average
ARIMA	AutoRegressive Integrated Moving Average
BIC	Bayesian Information Criterion
BL'	Bilinear Models
EM	Expectation-Maximisation
EMV	Estimateur de Maximum de Vraisemblance
ESE	Emperical Standard Errors
EXPAR	Exponential Autoregressive
GARCH	Generalized Autoregressive Conditional Heteroscedastic
GEM	Generalized Expectation-Maximisation
GMTD	Gaussian mixture Transition Distribution
MAR	Mixture Autoregressive
MAR-ARCH	Mixture Autoregressive Conditional Heteroscedastic
PDF	Probability Density Function
CDF	cumulative Distribution Function
SETAR	SElf Exciting Threshold Autoregressive Models
TAR	Threshold Autoregressive Models
TSE	Theoritcal Standard Errors

---

# Notations

Dans cette thèse, le sens des symboles sont ceux indiqués ci-dessous, si une autre signification est voulue, elle est explicitement donnée :

$\mathbf{A}^T$	Opérateur de transposition.
$\mathbf{K}$	Le nombre de composantes.
$\mathbf{k}$	La kurtosis.
$\epsilon$	Processus bruit blanc.
$\Theta$	L'ensemble des paramètres.
$\Omega_t$	L'information évaluée à l'instant $t$ .
$\rho_j$	L'autocorrélation du retard d'ordre $j$ .
$\mathbf{L}$	La fonction de vraisemblance.
$\mathbf{l}$	La fonction de log vraisemblance normalisée.
$\mathbf{C}^+$	Le cône positif de l'espace de Banach.
$\hat{\theta}$	Estimée du paramètre $\theta$
$I_m$	Matrice de l'information.
$\ \mathbf{A}\ $	Norme matricielle.
$[\mathbf{A}]_{ij}$ ou $a_{ij}$	Le $(i, j)^{ieme}$ élément de la matrice $\mathbf{A}$ .
$\arg \max_x f(x)$	L'argument maximisant la fonction $f(x)$ par rapport à $x$
$\arg \min_x f(x)$	L'argument minimisant la fonction $f(x)$ par rapport à $x$
$\delta$	Un nombre réel choisi petit.
$E[.]$	Espérance mathématique d'une variable aléatoire.
$Var[.]$	La variance d'une variable aléatoire.
$corr[.]$	L'autocorrélation entre deux variables aléatoires.
$F(.)$	La fonction de distribution conditionnelle.
$\Phi(.)$	La fonction de distribution cumulative d'une loi Gaussienne centrée réduite.
$\varphi(.)$	La fonction de densité de probabilité d'une loi Gaussienne centrée réduite.



---

# Table des figures

1.1	Les séries simulées des modèles (A) et (B). . . . .	9
1.2	Les autocorrélations estimées de la série générées à partir du modèle (A). 10	
1.3	Les autocorrélations estimées de la série générée à partir du modèle (B). 10	
3.1	La série simulée à partir du modèle (C). . . . .	36
3.2	Les autocorrélations estimées de la série simulée à partir du modèle (C). 37	
4.1	la variation de ESE et de TSE en fonction de la taille de l'échantillon pour les paramètres $\alpha_1, \alpha_2, \phi_{10}, \phi_{20}, \phi_{11}$ et $\phi_{21}$ du modèle (1). . . . .	65
4.2	la variation de ESE et de TSE en fonction de la taille de l'échantillon pour les paramètres $\beta_{10}, \beta_{20}, \beta_{11}, \beta_{21}$ du modèle (1). . . . .	66
5.1	La série originale. . . . .	68
5.2	la série différenciée au premier ordre. . . . .	68
5.3	La série différenciée au deuxième ordre. . . . .	69
5.4	La volatilité calculée à partir du modèle MAR-ARCH ajusté. . . . .	70
5.5	La distribution prédictive d'horizon 1 à l'instant $t = 199$ . . . . .	71
5.6	La distribution prédictive d'horizon 1 à l'instant $t = 220$ . . . . .	71
5.7	La distribution prédictive d'horizon 1 à l'instant $t = 221$ . . . . .	72
5.8	(a),(b), (c), (d) et (e) donnent les MSE calculées des prévisions données par les quatres modèles ajustés d'horizon 1, 2, 3, 4 et 5 resp. . . . .	73
5.9	(a), (b), (c), (d), (e) et (f) : La comparaison entre les couvertures empiriques d'horizon 1 données par les modèles ajustés par rapport aux couvertures présumées :95%, 90%, 80%, 70%, 60% et 50% resp. . . . .	74
5.10	(a), (b), (c), (d), (e) et (f) : La comparaison entre les couvertures empiriques d'horizon 2 données par les modèles ajustés par rapport aux couvertures présumées :95%, 90%, 80%, 70%, 60% et 50% resp. . . . .	75

---

# Introduction

CETTE thèse traite le problème de la modélisation des séries temporelles non linéaires ayant des propriétés de non linéarité ou/et non normalité comme les valeurs aberrantes, les valeurs explosives, la distribution conditionnelle prédictive multimodale, etc.

Quand l'objectif d'une analyse est la prévision ou la simulation, il est important de savoir comment modéliser explicitement la série de données. Il est important d'avoir une bonne distribution prédictive qui prend en considération la possibilité qu'il y' aurait dans le futur des valeurs aberrantes ou explosives, etc. Le, Martin, et Raftery (1996) ont introduit la classe de mélanges de distributions de transition Gaussiennes (GMTD). Ils ont montré à l'aide de données concrètes que cette classe de modèles peut vraiment représenter ces propriétés de non linéarité et non normalité.

Wong et Li (2000), ont généralisé le modèle GMTD au mélange de modèles autorégressifs (MAR). Les modèles MAR possèdent les mêmes propriétés que les modèles GMTD, mais avec plus de flexibilité. Ces modèles sont simples, faciles à traiter analytiquement, faciles à estimer et à simuler.

Plusieurs propriétés importantes permettent aux modèles MAR d'être un moyen potentiellement utile dans la modélisation des séries temporelles non linéaires. Wong, et Li (2000, section 5) ont montré à l'aide de données financières la performance du modèle MAR par rapport aux autres modèles (ARIMA, AR, SETAR). Néanmoins, la modélisation des séries non linéaires utilisant ces modèles, montre une limitation. La structure d'autocorrélation du processus carré est assez simple et elle est analogue à celle du modèle AR, ce qui nécessitent une autre étude et une autre extension.

Une autre classe de modèles non linéaires qui possède des propriétés que les modèles linéaires ARMA n'ont pas, et permettent de modéliser de façon plus réaliste un grand nombre de données, en particulier les données financières. Il s'agit de la classe de modèles ARCH. Cette classe de modèles a tout d'abord été introduite par Engle en (1982), dans le cadre d'étude des données macroéconomiques. Engle est parti de la constatation suivante : Les travaux économiques effectués sur ces séries, qui supposent une volatilité constante, laissent certaines données mal expliquées, et celles-ci sont souvent regroupées. Elles correspondent à des périodes où les variations deviennent plus fortes que celles constatées sur l'ensemble de la série, et considérées aberrantes par un modèle usuel. Le modèle ARCH apparaît comme une spécification de l'hétéroscédasticité. L'intérêt de l'approche réside dans le fait qu'il est souvent difficile d'identifier les variables explicatives entrant dans la variance du processus d'innovations. Le modèle considéré par Engle est en fait un modèle où la variance dépend d'une certaine fonction des valeurs passées du processus. L'intuition derrière un tel modèle

est que les chocs aléatoires peuvent avoir une probabilité plus importante à certaines périodes qu'à d'autres. Ces changements dans la variance pouvant avoir dans certains cas un effet persistant sur plusieurs périodes. La variation de la variance est aléatoire et n'est pas un événement provoqué de façon exogène.

L'intérêt du modèle ARCH en économétrie et dans le domaine de la finance montre un intérêt grandissant. Du point de vue statistique, ce modèle constitue une classe spécifique de modèles non linéaires dont l'étude est maintenant bien connue.

Il est intéressant de bien voir qu'un autre modèle peut capter l'hétéroscédasticité mieux que les modèles usuels (ARCH, GARCH, AR-ARCH). Wong et Li (2001) ont généralisé les modèles MAR aux mélanges de modèles autorégressifs conditionnellement hétéroscédastiques (MAR-ARCH). Ce type de modèles est composé d'un mélange de  $K$  modèles autorégressifs à erreurs ARCH, c'est-à-dire la moyenne conditionnelle suit un modèle MAR et la variance conditionnelle suit un mélange de processus ARCH. Les modèles MAR-ARCH permettent plus de flexibilité dans la structure d'autocorrélation du processus carré. Les conditions de stationnarité, la fonction d'autocorrélation du processus, et la fonction d'autocorrélation de carré du processus sont élaborées. La distribution prédictive au pas multiple est discutée. L'estimation peut être facilement faite à travers un simple algorithme EM, et en utilisant le critère BIC pour sélectionner le modèle. La propriété de changement de la forme des distributions conditionnelles rend ces modèles capables de modéliser les séries temporelles avec des distributions conditionnelles multimodales et avec une hétéroscédasticité.

Dans le premier chapitre, nous présentons le modèle MAR, en citant toutefois ses importantes propriétés qui le permettent d'être un candidat prometteur dans l'analyse des séries temporelles non linéaires. Les conditions de stationnarité du premier et second ordre et la fonction d'autocorrélation sont données par la suite. Finalement, nous discutons la construction de la fonction prédictive au pas multiple.

Dans le deuxième chapitre, nous donnons tout d'abord un aperçu sur les grandes classes des modèles non linéaires, puis nous nous intéressons, particulièrement au modèle ARCH. Ces propriétés importantes, le calcul des moments dans les modèles ARCH, les conditions de stationnarité et l'estimation de la fonction d'autocorrélation sont tous abordés dans ce chapitre. Et enfin, nous terminons par des extensions proposées pour ce type de modèles.

Dans le troisième chapitre, nous présentons le mélange de modèles autorégressifs conditionnellement hétéroscédastique (MAR-ARCH). Les démonstrations de plusieurs résultats importants sont présentées, concernant : les conditions de stationnarité au premier ordre pour ces modèles, et les conditions de stationnarité au second ordre pour quelques modèles MAR-ARCH particuliers, la condition nécessaire pour l'existence des moments d'ordre quatre pour des modèles MAR-ARCH particuliers, et la fonction d'autocorrélation du processus carré, sous certaines conditions, pour certains modèles. Nous citons aussi plusieurs propriétés importantes. Le changement de la forme de la distribution conditionnelle prédictive, et la structure d'autocorrélation du processus carré sont les propriétés les plus attrayantes.

Le quatrième chapitre traite le problème de l'estimation des paramètres et la

sélection du modèle dans la classe des modèles MAR-ARCH. Au début, nous discutons des problèmes associés à la méthode du maximum de la vraisemblance, pour passer, par la suite, à la présentation de la procédure la plus utilisée dans l'estimation des paramètres dans les modèles de mélanges, à savoir l'algorithme EM. Nous donnons la théorie fondamentale de cet algorithme, ses propriétés, et son application dans l'estimation des paramètres dans les modèles MAR-ARCH. Afin de savoir comment estimer l'erreur standard théorique, nous donnons un aperçu sur la méthode du Bootstrap. Ensuite, une analyse par simulation est proposée, afin de montrer la performance de cette approche dans l'estimation des paramètres dans les modèles MAR-ARCH. Et enfin nous terminons ce chapitre par l'application du critère BIC dans la sélection du modèle MAR-ARCH, suivie par une application numérique.

Dans le cinquième chapitre, nous illustrons l'utilité, la performance et la qualité de prévision des modèles MAR-ARCH. À l'aide d'une série de données réelles, largement analysée par les statisticiens (série C, Box & al. 1994), nous comparons entre la modélisation utilisant les modèles MAR-ARCH et la modélisation utilisant les modèles proposés par Box, Jenkins, et autres.

CHAPITRE

# 1

---

## Mélange de modèles autorégressifs

Dans ce chapitre nous présentons la généralisation des modèles GMTD aux modèles MAR, introduit par Wong et Li (2000), pour la modélisation des séries temporelles non linéaires. Le modèle MAR composé d'un mélange de  $K$  modèles autorégressifs stationnaires ou non stationnaires. Le changement de la forme de la distribution prédictive conditionnelle permet à ces modèles d'être capable de modéliser les séries temporelles ayant des distributions conditionnelles multimodales et avec hétéroscédasticité. Les conditions de stationnarité, la fonction d'autocorrélation, et la fonction prédictives au pas multiple sont données dans ce chapitre.

## 1.1 Introduction

Dans l'analyse des séries temporelles linéaires, la supposition que les termes des innovations sont gaussiens est souvent adoptée. Sous cette supposition, les distributions conditionnelles et marginales de la série sont aussi gaussiennes. Cependant dans la réalité plusieurs séries chronologiques montrent des propriétés qui contredisent la supposition gaussienne. Parfois ils existent de bonnes raisons de croire que les distributions conditionnelles pour quelques séries temporelles sont multimodales. Voir l'exemple donné par Tong (1990) , Chan et Tong (1998). Les modèles de séries temporelles non linéaires sont les candidats naturels pour modéliser les séries dans de telles situations

Très récemment, Le et al (1996) introduisirent le modèle de mélange de distributions gaussiennes de transition (GMTD)(pour une brève présentation des modèles GMTD, voir l'annexe A), pour capturer les propriétés de non linéarité et non normalité comme les valeurs aberrantes, les valeurs explosives et les distributions conditionnelles multimodales etc..

Dans ce chapitre, nous présentons la généralisation des modèles GMTD aux mélanges de modèles autorégressifs (MAR), dûe à Wong et Li (2000). Les modèles MAR possèdent les mêmes propriétés que les modèles GMTD, mais avec plus de flexibilité.

Ce modèle consiste en un mélange de  $K$  composantes, chaque composante est un modèle autorégressif gaussien. Les propriétés de la stationnarité, la fonction d'auto-corrélation peuvent être facilement dérivées. Wong et Li ont utilisé l'algorithme EM dans l'estimation des paramètres et le critère BIC pour sélectionner le modèle.

Plusieurs propriétés importantes permettent aux modèles MAR d'être candidats prometteurs dans la modélisation des séries temporelles non linéaires. Il est possible qu'un mélange d'un modèle autorégressif stationnaire avec un modèle autorégressif non stationnaire résulte un processus stationnaire. Les distributions conditionnelles de la série donnant le passé varient dans le temps. Ces distributions peuvent être multimodales.

La description de ce modèle et une étude brève sera présentée dans ce chapitre.

## 1.2 Mélange de modèles autorégressifs

Le modèle MAR de  $K$  composantes est défini par :

$$F(y_t | \Omega_{t-1}) = \sum_{k=1}^K \alpha_k \Phi \left( \frac{y_t - \phi_{k0} - \phi_{k1}y_{t-1} - \cdots - \phi_{kp_k}y_{t-p_k}}{\sigma_k} \right) \quad (1.1)$$

Notons ce modèle par  $MAR(K; p_1, \dots, p_K)$ . Tel que  $F(y_t | \Omega_{t-1})$  est la fonction de distribution cumulative conditionnelle de  $Y_t$  sachant l'information passée,  $\Phi(\cdot)$  est la fonction de distribution cumulative de la loi normale standard,  $\sigma_k$  est la variance conditionnelle de la  $k$ -ème composante et  $\alpha_1 + \cdots + \alpha_K = 1$ ,  $\alpha_k > 0$ ,  $k = 1, \dots, K$

Le modèle MAR considéré est un mélange de  $K$  modèles autorégressifs gaussiens.

Puisque les moyennes conditionnelles des composantes dépendent des valeurs passées de la série, alors la forme des distributions conditionnelles de la série varie dans le temps. Par exemple elle peut changer d'unimodale au multimodale. Comme la distribution conditionnelle de la série peut être multimodale, alors l'espérance conditionnelle de  $Y_t$  :

$$\begin{aligned} E(y_t | \Omega_{t-1}) &= \sum_{k=1}^K \alpha_k (\phi_{k0} - \phi_{k1}y_{t-1} - \cdots - \phi_{kp_k}y_{t-p_k}) \\ &= \sum_{k=1}^K \alpha_k \mu_{k,t} \end{aligned} \quad (1.2)$$

Peut perdre sa performance dans l'estimation des valeurs futurs. Le mérite des modèles MAR se trouve dans leur possibilité de décrire la distribution conditionnelle de la série temporelle. Une autre propriété importante des modèles *MAR* concerne leur capacité de modéliser la variance conditionnelle changeante de la série. La variance conditionnelle de  $Y_t$ , qui dépend des moyennes conditionnelles des composantes est donnée par :

$$\text{var}(y_t | \Omega_{t-1}) = \sum_{k=1}^K \alpha_k \sigma_k^2 + \sum_{k=1}^K \alpha_k \mu_{k,t}^2 - \left( \sum_{k=1}^K \alpha_k \mu_{k,t} \right)^2 \quad (1.3)$$

Le terme  $(\sum_{k=1}^K \alpha_k \mu_{k,t}^2 - (\sum_{k=1}^K \alpha_k \mu_{k,t})^2)$  n'est pas négatif et égale à zéro uniquement si  $\mu_{1,t} = \mu_{2,t} = \cdots = \mu_{K,t}$ . La valeur de la variance conditionnelle est importante quand la différence entre les  $\mu_{k,t}$  est significative. Dans ce cas la distribution conditionnelle devra être multimodale. La valeur minimale de la variance conditionnelle est atteinte quand les  $\mu_{k,t}$  sont égales. Nous appelons cette valeur minimale  $\sum_{k=1}^K \alpha_k \sigma_k^2$ , la ligne basse de la variance conditionnelle ou volatilité.

## 1.3 Étude des modèles MAR

### 1.3.1 Étude de la stationnarité

Avant d'étudier la stationnarité dans la classe des modèles *MAR*, nous rappelons quelques notions de stationnarité.

#### Définition d'un processus stationnaire au sens strict (stationnarité forte)

un processus  $(X_t)$  est dit stationnaire au sens fort si  $\forall$  le  $n$ -uplet du temps tel que  $t_i \in \mathbb{Z}$  et pour tout temps  $h \in \mathbb{Z}$  avec  $t_i + h \in \mathbb{Z}, \forall i, i = 1, \dots, n$ , la suite  $(X_{t_1+h}, \dots, X_{t_n+h})$  à la même loi de probabilité que la suite  $(X_{t_1}, \dots, X_{t_n})$ .

**Définition d'un processus stationnaire au second ordre (stationnarité faible)**

Un processus  $(X_t, t \in \mathbb{Z})$  est dit stationnaire au second ordre, ou stationnaire au sens faible, ou stationnaire d'ordre deux si les trois conditions suivantes sont satisfaites :

1.  $\forall t \in \mathbb{Z}, E(X_t^2) < \infty$
2.  $\forall t \in \mathbb{Z}, E(X_t) = m$ , indépendant de  $t$
3.  $\forall (t, h) \in \mathbb{Z}^2, cov(X_t, X_{t+h}) = E[(X_{t+h} - m)(X_t - m)] = \gamma(h)$ , indépendant de  $t$

La première condition  $E(X_t^2) < \infty$  garantit tout simplement l'existence (ou la convergence) des moments d'ordre deux. La seconde condition  $E(X_t) = m, \forall t \in \mathbb{Z}$  porte sur les moments d'ordre 1 et signifie tout simplement que les variables aléatoires  $X_t$  doivent avoir la même espérance quelle que soit la date  $t$ . Autrement dit, l'espérance du processus  $X_t$  doit être indépendante du temps. Enfin, la troisième condition,  $\gamma(h)$  indépendante de  $t$  porte sur les moments d'ordre deux résumés par la fonction d'autocovariance. Cette condition implique que ces moments doivent être indépendants de la date considérée et ne doivent dépendre uniquement que de l'ordre des retards. Par conséquent, il convient de noter que la stationnarité implique que la variance  $\gamma(0)$  du processus  $X_t$  est constante au cours du temps.

Les conditions de stationnarité au premier ordre pour les modèles MAR sont données dans le théorème (1), présenté ci-dessous. La dérivation des conditions de stationnarité au deuxième ordre pour un modèle MAR général est plus compliquée et nécessite les termes d'autocovariance. Nous présentons uniquement les conditions de stationnarité au second ordre aux modèles  $MAR(K; 1, \dots, 1)$  et  $MAR(K; 2, \dots, 2)$  dans les théorèmes (2) et (3) respectivement. Les preuves des théorèmes suivent celles dans Le, Martin et Raftery (1996), qui utilisent un résultat de Benes (1967). La dérivation des conditions de stationnarité au second ordre pour un modèle MAR général est similaire à la dérivation des conditions pour le modèle  $MAR(K; 2, \dots, 2)$ . Pour les détails des preuves voir Wong (1998).

**Théorème 1 (Wong et Li (2000))**

*Une condition nécessaire et suffisante pour qu'un processus  $Y_t$  suivant un modèle  $MAR(K; p_1, \dots, p_K)$  soit stationnaire dans la moyenne est que les racines  $z_1, \dots, z_p$  de l'équation :*

$$1 - \sum_{i=1}^p \left( \sum_{k=1}^K \alpha_k \phi_{ki} \right) z^{-i} = 0 \quad (1.4)$$

*Sont toutes à l'intérieur du cercle unité, où  $\phi_{ki} = 0$  pour  $i > p_k$ .*

**Théorème 2 (Wong et Li (2000))**

*Supposons que le processus  $Y_t$  suivant un modèle  $MAR(K; 1, \dots, 1)$  est stationnaire au premier ordre. Une condition nécessaire et suffisante pour que le processus  $Y_t$  soit stationnaire au second ordre est :*



$$|\alpha_1\phi_{11}^2 + \alpha_2\phi_{21}^2 + \cdots + \alpha_K\phi_{K1}^2| < 1 \quad (1.5)$$

**Théorème 3** (Wong et Li (2000))

Supposons que le processus  $Y_t$  suivant un modèle  $MAR(K; 2, \dots, 2)$  est stationnaire au premier ordre. Une condition nécessaire et suffisante pour que le processus  $Y_t$  soit stationnaire au second ordre est :

$$\begin{aligned} \beta_1 + \beta_2 &< 1 \\ \beta_2 - \beta_1 &< 1 \\ |\beta_2| &< 1 \end{aligned}$$

où

$$\beta_1 = \sum_{k=1}^K \alpha_k \phi_{k1}^2 + \frac{2 \left( \sum_{k=1}^K \alpha_k \phi_{k1} \phi_{k2} \right) \left( \sum_{k=1}^K \alpha_k \phi_{k1} \right)}{1 - \sum_{k=1}^K \alpha_k \phi_{k2}} \quad (1.6)$$

et

$$\beta_2 = \sum_{k=1}^K \alpha_k \phi_{k2}^2 \quad (1.7)$$

Il est remarquable à partir des théorèmes précédents que les conditions de stationnarité dépendent largement des proportions des composantes,  $i$ , et les valeurs des  $\alpha_k$ .

**Proposition 4** (Wong et Li (2000))

Un mélange d'un modèle autorégressif non stationnaire avec un modèle autorégressif stationnaire peut donner un processus stationnaire.

Nous illustrons ce point avec deux séries simulées, le modèle de simulation est un  $MAR(2; 1, 1)$ .

Modèle (A) :

$$F(y_t | \Omega_{t-1}) = 0.5\Phi\left(\frac{y_t - 0.5y_{t-1}}{5.0}\right) + 0.5\Phi\left(\frac{y_t - 1.1y_{t-1}}{1.0}\right) \quad (1.8)$$

Modèle (B) :

$$F(y_t | \Omega_{t-1}) = 0.75\Phi\left(\frac{y_t - 0.5y_{t-1}}{5.0}\right) + 0.25\Phi\left(\frac{y_t - 1.4y_{t-1}}{1.0}\right) \quad (1.9)$$

Les deux séries simulées sont données dans la figure (1.1). Nous remarquons que les deux séries simulées apparaissent stationnaires.

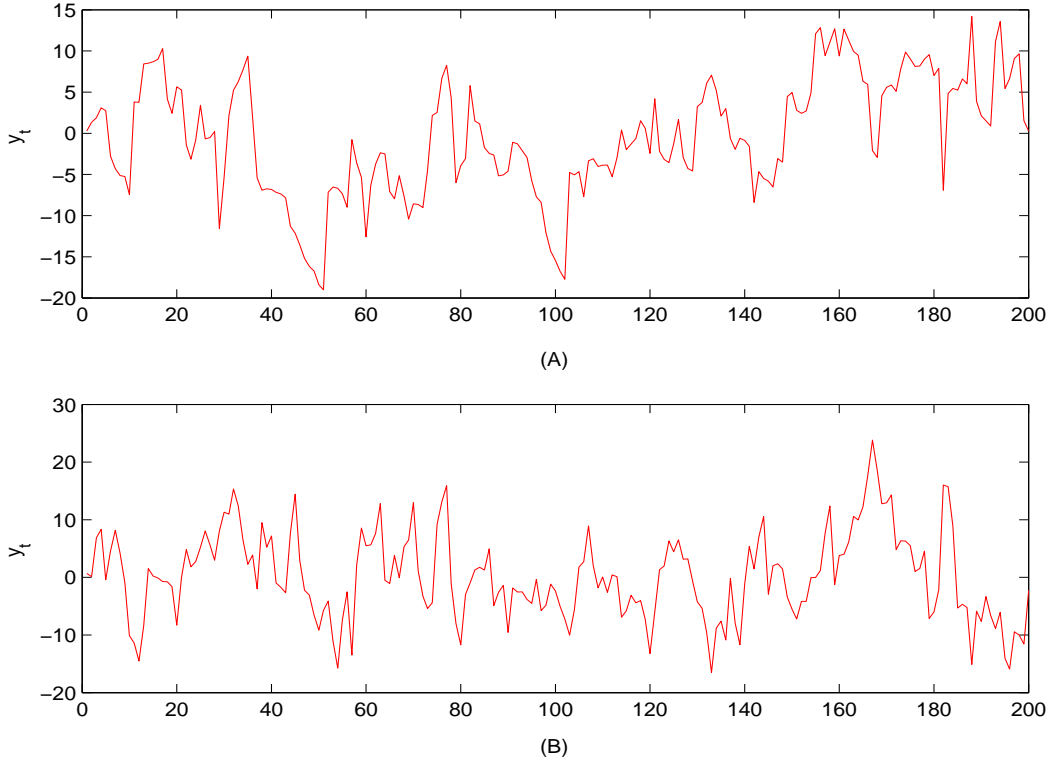


Fig. 1.1 — Les séries simulées des modèles (A) et (B).

### 1.3.2 Fonction d'autocorrélation

Pour un processus stationnaire au second ordre  $Y_t$  suivant un modèle  $MAR(K; p_1, \dots, p_K)$ , il est facile de voir que

$$\rho_j = \sum_{i=1}^p \left( \sum_{k=1}^K \alpha_k \phi_{ki} \right) \rho_{|j-i|} \quad (1.10)$$

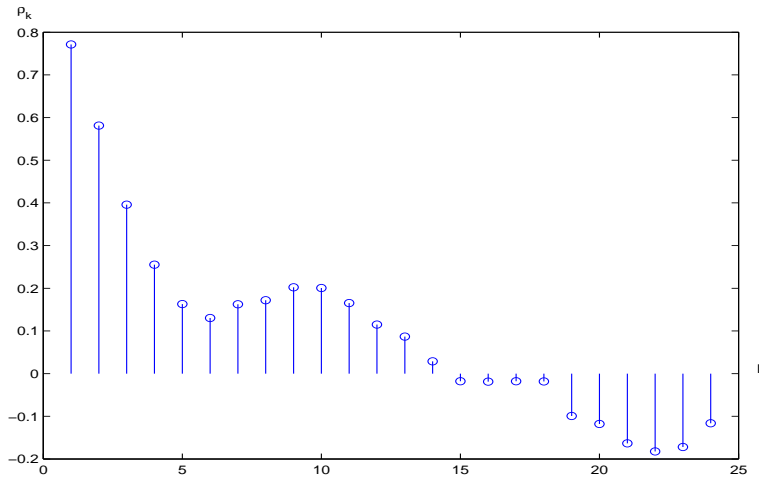
où  $\rho_j$  est l'autocorrélation de retard  $j$ , tel que  $\phi_{ki} = 0$  pour  $i > p_k$ . Notons que ces équations sont similaires à celles des équations de Yule-Walker pour un processus AR(p) ordinaire, où les coefficients  $\sum \alpha_k \phi_{ki}$  remplacent les coefficients du  $i$ -ème retard du processus AR(p).

Les figures (1.2) et (1.3) donnent les autocorrélations estimées pour les deux séries simulées.

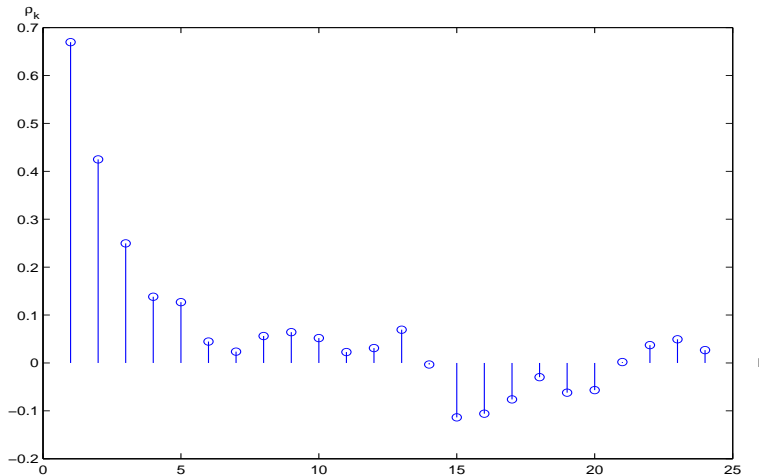
**Remarque 5** *Le rayon des autocorrélations possibles est plus grand que celui du processus AR standard, puisque le modèle AR est juste un cas spécial du modèle MAR.*

### 1.3.3 La fonction prédictive au pas $m$

La fonction prédictive au premier pas  $F(y_{t+1} | \Omega_t)$  est facile à calculer en se basant sur le modèle (1.1). Le calcul de la fonction prédictive au pas  $m$   $F(y_{t+m} | \Omega_t)$  n'est



**Fig. 1.2** — Les autocorrélations estimées de la série générées à partir du modèle (A).



**Fig. 1.3** — Les autocorrélations estimées de la série générée à partir du modèle (B).

pas direct. Granger et trasvirta (1993) donnèrent une importance aux prévisions au pas  $m$  basées sur les modèles non linéaires générales. Nous proposons trois approches pour le calcul des distributions prédictives au pas  $m$ , appelées approches naïve, exacte et Monte Carlo. Nous utilisons la distribution prédictive au deuxième pas  $F(y_{t+2} | \Omega_t)$  comme une illustration de ces approches.

Dans l’approche naïve, nous utilisons tout simplement la prévision au premier pas  $\hat{y}_{t+1} = E(y_{t+1} | \Omega_t)$  comme si la vraie valeur de  $y_{t+1}$ . Nous avons alors

$$F(y_{t+2} | \Omega_t) = F(y_{t+2} | \Omega_t, y_{t+1} = \hat{y}_{t+1}) \tag{1.11}$$

Cette approche est commode mais elle ignore toute information fournie par la forme de la distribution  $F(y_{t+1} | \Omega_t)$ . Concernant le modèle MAR l’information perdue peut être importante puisque  $F(y_{t+1} | \Omega_t)$  peut être multimodale. Une meilleure approche

consiste à calculer la distribution prédictive au deuxième pas exacte par :

$$F(y_{t+2} | \Omega_t) = \int F(y_{t+2} | \Omega_t, y_{t+1}) \partial F(y_{t+1} | \Omega_t) \quad (1.12)$$

Cette intégration peut être difficile à traiter analytiquement mais nous pouvons cependant utiliser des méthodes numériques. Alternativement, l'approximation de Monte Carlo de la distribution prédictive peut être plus adéquate. L'approximation de Monte Carlo de la distribution prédictive au deuxième pas est donnée par :

$$F(y_{t+2} | \Omega_t) = \frac{1}{N} \sum_{j=1}^N F(y_{t+2} | \Omega_t, y_{t+1}^j) \quad (1.13)$$

où  $\{y_{t+1}^j\}$  sont des échantillons générés à partir de  $F(y_{t+1} | \Omega_t)$ .

## 1.4 Discussion

Nous avons discuté une nouvelle classe de modèles, les modèles MAR, qui sont potentiellement utiles dans la modélisation des séries temporelles non linéaires. Ces modèles permettent plus de flexibilité par rapport aux autres modèles (section 5 Wong, Li (2000)). L'algorithme EM permet une estimation facile des paramètres dans ce modèle et en utilisant le critère BIC pour la sélection de l'ordre. Néanmoins, il y a des aspects qui nécessitent une autre étude et d'autres extensions.

Les tests d'identifications du modèle peuvent être effectués avec les résidus normalisés

$$e_t = \{y_t - E(y_t | \Omega_{t-1})\} / \sqrt{\text{var}(y_t | \Omega_{t-1})} \quad (1.14)$$

Les autocorrélations des résidus normalisés et autocorrélations des résidus normalisés carrés peuvent être aussi utilisées pour tester si les modèles proposés sont adéquats et la normalité du graphe des résidus normalisés peut être examinée puisque elle devrait avoir une distribution conditionnelle normale. Néanmoins, il y a un inconvénient à cette approche. Comme nous avons discutés dans la section 2, les moyennes conditionnelles peuvent ne pas être les meilleurs estimateurs quand les distributions conditionnelles sont multimodales. Ainsi les résidus définis dans le sens usuel peuvent ne pas être appropriés aux modèles MAR. Néanmoins l'identification du modèle avec les résidus normalisés peut donner une bonne indication sur le modèle, est ce qu'il est adéquat ou pas.

Les modèles MAR peuvent être étendus dans plusieurs directions. La distribution conditionnelle normale dans ce modèle peut être remplacée par d'autres distributions. Par exemple, si elle est remplacée par les distributions gamma, exponentiel ou Weibull, ces modèles peuvent être utilisés dans la représentation des séries temporelles à valeurs positives. Voir Gaver et Lewis (1980) et Lawrance et Lewis (1985) pour leurs efforts dans cette direction.

---

Autre direction est de considérer un mélange de modèles AR avec un autorégressif conditionnellement hétéroscédastique (ARCH) ou avec un Autorégressif conditionnellement hétéroscédastique généralisé (GARCH). Voir Engle (1982) et Bollerslev (1986). Il est intéressant de bien voir qu'un modèle pareil peut capter l'hétéroscédasticité mieux que les modèles usuels ARCH ou GARCH. Dans notre travail de mémoire nous allons essayer de présenter une étude sur les mélanges de modèles autorégressifs à erreurs ARCH (*MAR – ARCH*). Ce modèle est une extension des modèles MAR et qui peut capturer l'hétéroscédasticité mieux que les modèles MAR et les modèles ARCH classiques. Avant d'entamer cette étude, nous évoquons dans le chapitre suivant la théorie fondamentale des modèles ARCH.

CHAPITRE

# 2

---

## Les modèles autorégressifs conditionnellement hétéroscédastiques

Dans ce chapitre, nous commençons par donner un aperçu sur les grandes classes de modèles de série chronologiques non linéaires. Ensuite, nous étudions la classe de modèles ARCH. En citant ses quelques propriétés importantes. Et enfin nous terminons ce chapitre par des extensions des modèles ARCH

## 2.1 L'approche *ARCH-GARCH* et la modélisation de l'incertitude

L'apparition des modèles *ARCH-GARCH* doit être replacé dans le contexte plus vaste du débat sur la représentation linéaire ou non linéaire des processus stochastiques temporels. L'approche *ARCH-GARCH*, a été proposée pour prendre en compte des variances conditionnelles dépendant du temps. Le principe général consiste donc à remettre en cause la propriété d'homoscédasticité que l'on retient généralement dans le cadre du modèle linéaire. " Commençons par un petit exemple introductif. Dans l'analyse traditionnelle de la prévision (cf. Box et Jenkins), la construction des valeurs futurs prévues est fondée sur la moyenne conditionnelle de la série utilisée. Ainsi, la prévision de la valeur d'une série temporelle  $y_t$  à la date  $t + 1$  compte tenu de l'information disponible à la date  $t$  est donnée par  $E[y_{t+1} | y_t, y_{t-1}, \dots]$ . Supposons que  $y_t$  suivent un processus *AR(1)* stationnaire  $y_t = \theta y_{t-1} + \varepsilon_t$ , avec  $\varepsilon_t \sim i, i, d(0, \sigma_\varepsilon^2)$ , alors la moyenne conditionnelle de  $y_{t+1}$  est  $\theta y_t$  tandis que son espérance nonconditionnelle est nulle. Comme le fait remarquer Engle (1982), l'amélioration des prévisions issues de modèles de séries temporelles provient clairement de l'exploitation de l'information contenue dans l'espérance conditionnelle du processus. L'idée serait de tenir compte des autres moments conditionnels de ce processus. Or, la variance conditionnelle du processus *AR(1)* est égale à  $\sigma_\varepsilon^2$ , tandis que sa variance nonconditionnelle est égale à  $\frac{\sigma_\varepsilon^2}{(1-\theta^2)}$ . Ainsi, les variances des prévisions conditionnelles et nonconditionnelles sont constantes quelle que soit la date de la prévision. Avec de tels modèles, on est donc incapables de mesurer d'éventuels changement dans les variances des erreurs de prévision même si l'on souhaite que celles-ci soient affectées par l'évolution passée.

Si l'on dit à l'économètre classique qu'il doit estimer les paramètres d'un modèle dont la variance des perturbations évolue avec le temps, il évoquera un problème d'hétéroscédasticité, qui signifie tout simplement que la matrice de variance covariances des erreurs n'est pas définie à un scalaire près par la matrice identité. Autrement dit, les termes de la diagonale principale de la matrice de variance covariance (c'est-à-dire les variances) ne sont pas tous identiques pour les différentes perturbations intervenant à des dates différentes. Il convient alors de traiter ce problème pour obtenir des estimateurs efficaces : généralement, la solution consiste à introduire une variable exogène  $X_t$  qui permet de prévoir l'évolution de la variance. Si par exemple  $y_t = \varepsilon_t x_{t-1}$  alors la variance conditionnelle de  $y_t$  est égale à  $\sigma_\varepsilon^2 x_{t-1}^2$ . Les intervalles de prévision sur  $y_t$  dépendront alors de l'évolution de la variable exogène  $x_t$ . Toutes fois, cette première approche est peu satisfaisante puisqu'elle nécessite de spécifier a priori une cause à l'évolution de la variance.

Le principe général proposé par Engle (1982) consiste à supposer que la variance dépend de l'ensemble informationnel dont on dispose. Il propose une spécification *ARCH(q)* où le carré des perturbations suit un processus autorégressif d'ordre  $q$ . Les modèles *ARCH* sont donc des modèles autorégressifs conditionnellement hétéroscédastiques. Engle (1982) a donc proposé ces processus pour pallier aux insuffisances de la classe des représentations *ARMA*, notamment en ce qui concerne les séries financières qui présentent une volatilité (ou variabilité instantanée mesurée par

la variance conditionnelle) fonction du temps et par des ajustements asymétriques. Ainsi, les modèles *ARCH* sont basés sur une paramétrisation endogène de la variance conditionnelle. La famille des modèles *ARCH* peut se décomposer en deux sous-ensembles : les modèles *ARCH* linéaires et les modèles *ARCH* non linéaires. Les premiers reposent sur une spécification quadratique de la variance conditionnelle des perturbations : modèles *ARCH* ( $q$ ), *GARCH* ( $p, q$ ) et *IGARCH* ( $p, q$ ). Les modèles *ARCH* non linéaires sont caractérisés par des spécifications asymétriques des perturbations, ce sont les modèles : *EGARCH* ( $p, q$ ), *TARCH* ( $q$ ) et *TGARCH* ( $p, q$ ) (Bresson et pirotte, série temporelles). Avant de présenter les modèles *ARCH-GARCH*, nous rappelons le théorème fondamental dans l'analyse des séries temporelles, ensuite nous donnons un aperçu sur les grandes classes de modèles de séries temporelles structurellement non linéaires.

**Théorème 6** (*Décomposition de Wold*) : tout processus stationnaire d'ordre deux ( $X_t, t \in \mathbb{Z}$ ) peut être représenté sous la forme :

$$X_t = \sum_{j=0}^{\infty} \phi_j \varepsilon_{t-j} + k_t \quad (2.1)$$

où les paramètres  $\phi_j$  satisfont  $\phi_0 = 1, \phi_j \in \mathbb{R}, \forall j \in \mathbb{N}, \sum_{j=0}^{\infty} \phi_j^2 < \infty$  et où  $\varepsilon_t$  est un bruit blanc *i.i.d.* ( $0, \sigma_\varepsilon^2$ ). On dit que la somme des chocs passés correspond à la composante linéaire stochastique de  $X_t$ . Le terme  $k_t$  désigne la composante linéaire déterministe telle que  $cov(k_t, \varepsilon_t) = 0, \forall j \in \mathbb{Z}$ .

La démonstration de ce théorème est donnée dans Wold (1938). Ainsi, d'après le théorème de Wold, si l'on omet la composante déterministe  $k_t$ , tout processus stationnaire peut s'écrire comme une somme pondérée infinie de chocs passés, ces chocs étant représentés par un bruit blanc de variance finie. L'implication forte de ce théorème est que, si l'on connaît les pondérations  $\phi_j, \forall j \in \mathbb{N}$ , et si l'on connaît la variance  $\sigma_\varepsilon^2$  du bruit blanc, on est en mesure de proposer une représentation de n'importe quel processus stationnaire. Cette représentation est aussi qualifiée de *représentation moyenne mobile infinie*.

Reste à comprendre ce que peut être cette composante linéaire déterministe  $k_t$ . La condition  $cov(k_t, \varepsilon_t) = 0$ , implique que ce terme est, par définition (déterministe), indépendant des chocs. Alors le cas le plus simple est celui d'un processus stationnaire ( $X_t, t \in \mathbb{Z}$ ) d'espérance non nulle. Puisque le bruit blanc est par définition un processus centré, une somme pondérée de ces chocs est elle-même centrée. Par conséquent, la représentation de Wold du processus ( $X_t, t \in \mathbb{Z}$ ) suppose que l'on ajoute à cette somme pondérée des chocs passés, une composante déterministe qui n'est autre que l'espérance du processus :  $k_t = m$ . On a donc :

$$X_t = \sum_{j=0}^{\infty} \phi_j \varepsilon_{t-j} + m \quad (2.2)$$



## 2.2 Les grandes classes de modèles non linéaires

“Naturellement ces propriétés sont difficiles, voir impossibles, à reproduire à partir de modèle *ARMA* linéaire classiques. Ces modèles linéaires de séries temporelles n'étaient finalement fondés que sur des combinaisons linéaires de valeurs présentes et passés de chocs. En effet, le théorème central de l'analyse des séries temporelles qui est le théorème de Wold (1938), indique que tout processus faiblement stationnaire peut être réécrit sous la forme d'une moyenne mobile infinie de processus de type bruits blancs, c'est-à-dire sous la forme d'une combinaison linéaire d'une séquence de variable aléatoire non corrélées dans le temps (voir le théorème 6). Par conséquent, l'hypothèse de processus *ARMA* stationnaire ne permet pas de prendre en compte d'une part les mécanismes d'asymétrie et d'autre part les ruptures de forte amplitude. D'où la nécessité d'aller vers des modélisations non linéaires.

Rappelons la définition générale d'un processus autorégressif linéaire (Gourieroux, 1992) :

### Définition 1

*Un processus stochastique  $X_t$  est un processus autorégressif d'ordre  $k$  si et seulement si :*

$$E(X_t | X_{t-1}) = E(X_t | X_{t-1}, X_{t-2}, \dots, X_{t-k}) \quad (2.3)$$

*Un processus stochastique  $X_t$  est un processus autorégressif linéaire d'ordre  $k$  si et seulement si :*

$$EL(X_t | X_{t-1}) = EL(X_t | X_{t-1}, X_{t-2}, \dots, X_{t-k}) \quad (2.4)$$

où  $EL(\cdot)$  désigne l'espérance linéaire.

L'espérance conditionnelle  $E(X_t | X_{t-1})$  est la meilleure approximation au sens de l'erreur quadratique moyenne de  $X_t$  par une fonction des valeurs passées  $X_{t-1}, X_{t-2}, \dots$ . Cette approximation est en général une fonction non linéaire de ces valeurs. La régression linéaire  $EL(X_t | X_{t-1})$  est la meilleure approximation de  $X_t$  par fonction linéaire affine de  $X_{t-1}, X_{t-2}, \dots$ . Au regard de cette définition, il existe une infinité de processus non linéaire susceptible de représenter les séries temporelles.

Campbell, Lo et Mackinlay (1997) ont proposé le cadre suivant pour décrire un processus non linéaire :

$$X_t = g(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots) + \varepsilon_t h(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots) \quad (2.5)$$

où la fonction  $g(\cdot)$  correspond à la moyenne conditionnelle du processus  $X_t$  et où la fonction  $h(\cdot)$  correspond à un coefficient de proportionnalité entre  $X_t$  et le choc  $\varepsilon_t$ . Cela permet de classer les processus non linéaires en deux parties :

1. Processus non linéaires en moyenne pour les quels  $g(\cdot)$  est non linéaire
2. Processus non linéaires en variance pour les quels  $h(\cdot)$  est non linéaire.

Une autre façon d’appréhender cette littérature sur les processus non linéaires consiste à opposer deux types d’approches. La première approche fondée sur des extensions non linéaires de processus *ARMA* qui permettent notamment d’appréhender les mécanismes d’asymétrie et de seuils, les économètres ont développé toute une panoplie de spécifications :

- Modèles bilinéaires (Granger et Anderson, 1978)
- Modèles exponentiels autorégressifs (modèles *EXPAR*)
- Modèles à seuils (*TAR*, *SETAR*, *STAR*, *MA* asymétriques etc...) développés depuis les travaux pionniers de Tong (1978)
- Modèles *MA* non linéaires.

La seconde voie consiste à proposer une représentation autorégressive de la variance conditionnellement à son information passée permettant de tenir compte des phénomènes de volatilité. Dans ce domaine, le papier de Engle de 1982, "AutoRegressive Conditional Heteroskedasticity with estimates of variance of UK inflation", *Econometrica* (1982), a ouvert la voie à la modélisation *ARCH* et ses nombreuses développements (Hurlin). C’est précisément ce que on va évoquer dans ce chapitre.

Mais avant cela, nous allons présenter un certain nombre de modèles non linéaires appartenant à la première approche dont certains sont pourtant proches des modèles *ARCH* de part leurs implications.

### 2.2.1 Modèles bilinéaires

Le modèle bilinéaire, introduit par Granger et Anderson (1978), présente la particularité d’être à la fois linéaire en  $X_t$  et  $\varepsilon_t$  mais ne pas l’être par rapport à ces deux variables prises conjointement. Un modèle bilinéaire, noté *BL* ( $p, q, P, Q$ ), s’écrira ainsi sous la forme :

$$X_t = \mu + \sum_{i=1}^p \phi_i X_{t-i} + \sum_{i=0}^q \theta_i \varepsilon_{t-i} + \sum_{i=1}^P \sum_{j=1}^Q \lambda_{ij} X_{t-i} \varepsilon_{t-j} \quad (2.6)$$

Avec  $\theta_0 = 1$  et  $(\phi_p, \theta_q, \lambda_{iQ}, \lambda_{Pj}) \in \mathbb{R}^{*4}, \forall (i, j)$  et où  $\varepsilon_t$  désigne un bruit blanc, éventuellement gaussien.

### 2.2.2 Modèles autorégressifs exponentiels (modèles *EXPAR*)

Ces modèles ont une structure autorégressive de type *AR*, mais permettent de rendre en compte des phénomènes de cluster de volatilité de la série. Un modèle EXPonentiel Autorégressif (*EXPAR*) est défini par la relation :

$$X_t = \mu + \sum_{i=1}^p [\alpha_i + \beta_i \exp(-\gamma X_{t-1}^2)] X_{t-i} + \varepsilon_t \quad (2.7)$$

On reconnaît ici une structure multiplicative sensiblement proche des modèles bilinéaires.

### 2.2.3 Modèles autorégressifs à seuil (modèles *TAR*)

Les modèles autorégressifs à seuil constituent l'une des spécifications possibles de la grande famille des modèles à régime. L'idée consiste à postuler l'existence de plusieurs dynamique pour une même série (plusieurs régimes) et à spécifier un mécanisme de transition d'un régime à l'autre. Deux grands types de mécanismes de transition existent :

- Des mécanismes de transition stochastiques et exogènes régis par des processus de type chaîne de Markov : on parle alors de Markov Switching Models (Hamilton, 1989).
- Des mécanismes de transition endogènes où la fonction de transition dépend de la variable dépendante et d'un seuil : on parle alors de modèles à seuil ou Threshold Autorégressives models (*TAR*).

Les modèles à seuil ont été introduits par Tong (1978). Il existe toute une classe de modèles suivant la définition retenue de la fonction de transition : *TAR*, *MTAR*, *STAR*, *ESTAR*, *LSTAR*, *MSTAR* etc .. le modèle le plus simple est le modèle *SETAR* introduit par Tong. Considérons le cas le plus simple d'un modèle à deux régimes :

$$X_t = \Phi_1(L) X_t I_{X_{t-d} > \gamma} + \Phi_2(L) X_t I_{X_{t-d} \leq \gamma} + \varepsilon_t \quad (2.8)$$

où  $\Phi_j(L)$ ,  $j = 1, 2$  désignent deux polynômes retard d'ordre fini et où  $\varepsilon_t$  est  $i, i, d.(0, \sigma_\varepsilon^2)$  la fonction  $I$  désigne l'indicatrice telle que :

$$I_{X_{t-d} > \gamma} = 1 \text{ si } X_{t-d} > \gamma \text{ et } I_{X_{t-d} > \gamma} = 0 \text{ si non.} \quad (2.9)$$

Le paramètre  $d \in \mathbb{N}$ , appelé délai, est nécessairement supérieur à l'unité pour éviter des problèmes de simultanéité. Le paramètre  $\gamma \in \mathbb{R}$ , est appelé paramètre de seuil. Ce type de modèle permet très facilement de modéliser des phénomènes tels que l'asymétrie : pour un même choc, les mécanismes de propagation diffèrent suivant les valeurs passées de la variable dépendante  $X_{t-d}$ . Ce type de processus permet aussi d'obtenir des distributions leptokurtiques de la variable dépendante.

Ainsi, cette présentation de quelques grandes classes de modèles non linéaires montre qu'il existe un grand nombre de voies de recherche possibles.

## 2.3 Modèles *ARCH-GARCH*

### 2.3.1 Présentation des modèles *ARCH*

Commençons par présenter le modèle *ARCH* (1) introduit par Engle (1982). On considère un processus  $X_t$  tel que :

$$X_t = z_t \sqrt{\alpha_0 + \alpha_1 X_{t-1}^2} \quad (2.10)$$

où  $z_t$  est une suite de v.a (i.i.d) gaussiennes centrées, tel que  $E(z_t^2) = \sigma_z^2$  et les paramètres  $\alpha_0$  et  $\alpha_1$  sont tels que  $\alpha_0 > 0, \alpha_1 \geq 0$ . On note généralement  $h_t = \alpha_0 + \alpha_1 X_{t-1}^2$ . De façon générale,  $z_t$  désigne un ensemble de variables aléatoires indépendantes, identiquement distribuées, centrées, réduites. La composante  $h_t$  désigne une variable qui, conditionnellement à l'ensemble d'information des valeurs passées de  $X_t$ , i.e. à  $X_{t-1} = \{X_{t-1}, X_{t-2}, \dots, X_{t-j}, \dots\}$ , est déterministe. Dans ce système, le processus  $X_t$  est caractérisé par des autocorrélations nulles et une variance conditionnellement variable dans le temps en fonction de l'ampleur de l'innovation passée. Notons que dorénavant, sans perdre de généralité, on prend  $\sigma_z^2 = 1$ .

On peut établir des résultats intéressants en considérant le processus autorégressif sur  $X_t^2$ . Pour simplifier, on se limite au cas du *ARCH* (1). Dans ces conditions :

$$h_t = \alpha_0 + \alpha_1 X_{t-1}^2 \iff X_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + (X_t^2 - h_t) \quad (2.11)$$

soit encore :

$$X_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \varepsilon_t \quad (2.12)$$

où  $\varepsilon_t = X_t^2 - h_t$  vérifiant  $E(\varepsilon_t | X_{t-1}) = 0$  est un processus d'innovation pour  $X_t^2$ .

Ainsi, cette écriture précédente correspond à celle d'un *AR*(1) sur le carré  $X_t^2$ . On sait que ce processus  $X_t^2$  est stationnaire au second ordre si et seulement  $\alpha_1 < 1$ . On peut déduire de ces différentes écritures, un certain nombre de propriétés qui pourront être étendues au cas des processus *ARCH*( $q$ ).

**Proposition 7** *Le processus  $X_t$  suivant un modèle *ARCH*(1) est une différence de martingale homoscédastique :*

$$E(X_t | X_{t-1}) = 0 \quad V(X_t) = \frac{\alpha_0}{1 - \alpha_1} \quad \forall t \quad (2.13)$$

Cette propriété signifie que le processus  $X_t$  peut s'apparenter à un processus de bruit (faible), ce qui explique notamment que l'on spécifiera des erreurs de modèles sous la forme *ARCH*. On retrouve alors toutes les propriétés de modèles établies sous la propriété de bruit blanc des erreurs. Mais cette propriété signifie en outre que le processus  $X_t$  est non conditionnellement homoscédastique.

**Proposition 8** *La variance conditionnelle du processus  $X_t$  suivant un modèle *ARCH*(1) est non constante dans le temps et vérifie :*

$$V(X_t | X_{t-h}) = \alpha_0 \left( \frac{1 - \alpha_1^h}{1 - \alpha_1} \right) + \alpha_1^h X_{t-h}^2 \quad \forall t \quad (2.14)$$

C'est la propriété centrale des processus *ARCH* : le processus  $X_t$  possède les propriétés d'un bruit blanc homoscédastique, mais sa variance conditionnelle dépend du temps. Lorsque  $h$  tend vers l'infini, ces variances conditionnelles convergent vers la variance non conditionnelle, et l'on retrouve alors la formule de la propriété 7 :

$$\begin{aligned} V(X_t) &= \lim_{h \rightarrow \infty} V(X_t | X_{t-h}) = \lim_{h \rightarrow \infty} \left[ \alpha_0 \left( \frac{1 - \alpha_1^h}{1 - \alpha_1} \right) + \alpha_1^h X_{t-h}^2 \right] \\ &= \frac{\alpha_0}{1 - \alpha_1} \end{aligned} \quad (2.15)$$

**Proposition 9** *Les autocovariances conditionnelles du processus  $X_t$  suivant un modèle ARCH(1) sont nulles :*

$$\text{cov}(X_t, X_{t+k} | X_{t-h}) = 0, \forall h \geq 1, \forall k \geq 1. \quad (2.16)$$

*Un processus est donc un processus qui conditionnellement à  $X_{t-h}$  est un processus sans mémoire.*

**Proposition 10** *les conditions suffisantes qui garantissent la positivité du processus  $X_t^2$  sont  $\alpha_1 > 0$  et  $\alpha_0 + \varepsilon_t \geq 0$  pour toute valeur admissible de  $\varepsilon_t$ . Ceci implique notamment des contraintes sur le support de la loi de  $\varepsilon_t$ . De plus, la variance marginale du processus  $X_t$  existe si et seulement si  $\alpha_0 > 0$  et  $0 < \alpha_1 < 1$ .*

En effet, il convient de vérifier notamment que  $V(X_t^2)$  et  $V(X_t)$  sont définies de façon positive. Sous les conditions  $\alpha_0 > 0$  et  $0 < \alpha_1 < 1$ , la variance de  $X_t$  existe et est constante dans le temps : le processus  $X_t$  est stationnaire au second ordre.

On peut en outre établir les moments conditionnels et non conditionnels d'ordre 4 du processus  $X_t$ .

**Proposition 11** *Le moment conditionnel centré d'ordre 4 du processus  $X_t$  vérifie :*

$$E(X_t^4 | X_{t-1}) = 3(\alpha_0 + \alpha_1 X_{t-1}^2)^2 \quad (2.17)$$

*Sous l'hypothèse  $3\alpha_1^2 < 1$ , le moment non conditionnel centré d'ordre 4 du processus  $X_t$  est égal à :*

$$E(X_t^4) = 3 \left[ \alpha_0^2 + \frac{2\alpha_1\alpha_0^2}{1 - \alpha_1} + \alpha_1^2 E(X_{t-1}^4) \right]$$

*En tenant compte de la stationnarité, il s'ensuit que :*

$$E(X_t^4) = \frac{3\alpha_0^2(1 + \alpha_1)}{(1 - 3\alpha_1^2)(1 - \alpha_1)} \quad (2.18)$$

*La kurtosis non conditionnelle associée au processus ARCH(1) est égale à :*

$$\begin{aligned} \text{Kurtosis} &= \frac{E(X_t^4)}{(E(X_t^2))^2} \\ &= 3 \left( \frac{1 - \alpha_1^2}{1 - 3\alpha_1^2} \right) \\ &= 3 \left( 1 + \frac{2\alpha_1^2}{1 - 3\alpha_1^2} \right) > 3 \end{aligned} \quad (2.19)$$

où  $3\alpha_1^2 < 1$

Sous l'hypothèse de positivité du paramètre  $\alpha_1$ , la kurtosis non conditionnelle est toujours supérieure à celle de la loi normale : elle traduit l'aspect leptokurtique de la distribution du processus  $X_t$ . Toutes ces propriétés peuvent être généralisées au cas d'un processus  $ARCH(q)$

#### Définition 4

Un processus  $X_t$  satisfait une représentation  $ARCH(q)$  si

$$X_t = z_t \sqrt{h_t} \quad (2.20)$$

avec

$$h_t = \alpha_0 + \sum_{i=1}^q \alpha_i X_{t-i}^2 \quad (2.21)$$

où  $z_t$  est une suite de v.a (i.i.d) gaussiennes centrées réduites et les paramètres  $\alpha_0$  et  $\alpha_i$  sont tels que  $\alpha_0 > 0, \alpha_i \geq 0$ . Pour ce type de processus, on retrouve les deux propriétés essentielles vues précédemment, à savoir la propriété de différence de martingale (ou bruit blanc)  $E(X_t | X_{t-1}) = 0$  et la propriété de variance conditionnelle variable dans le temps puisque :

$$V(X_t | X_{t-1}) = h_t = \alpha_0 + \sum_{i=1}^q \alpha_i X_{t-i}^2 \quad (2.22)$$

### 2.3.2 Modèle avec erreur $ARCH(q)$

On considère dorénavant non plus un processus  $ARCH$  pour modéliser directement une série temporelle, mais le résidu d'un modèle linéaire. Prenons l'exemple d'un modèle linéaire autorégressif avec résidus de type  $ARCH(q)$

#### Définition 5

On considère un modèle linéaire autorégressif de la forme

$$Y_t = EL(Y_t | Y_{t-1}) + \varepsilon_t \quad (2.23)$$

tel que  $\varepsilon_t$  admet une représentation de type  $ARCH(q)$

$$\varepsilon_t = z_t \sqrt{h_t} \text{ avec } h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 \quad (2.24)$$

où  $z_t$  est une suite de v.a (i.i.d) gaussiennes centrées réduites. On a donc un modèle qui décrit à la fois l'évolution de l'espérance conditionnelle et la variance conditionnelle du processus  $Y_t$  dans le temps. Envisageons la cas le plus simple d'un processus de type  $AR(1)$  avec erreur  $ARCH(1)$  :

$$Y_t = \mu + \phi Y_{t-1} + \varepsilon_t \quad (2.25)$$

$$\varepsilon_t = z_t \sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2} \quad (2.26)$$

avec  $|\rho| < 1$ , et les paramètres  $\alpha_0$  et  $\alpha_1$  sont tels que  $\alpha_0 > 0, \alpha_1 \geq 0$ . Dans ce cas, les résidus satisfont les 4 principales propriétés étudiées précédemment :

- Propriété de différence de martingale :  $E(\varepsilon_t | \varepsilon_{t-1}) = 0$  et de façon générale

$$E(\varepsilon_t | \varepsilon_{t-h}) = 0, \quad \forall h \geq 1 \quad (2.27)$$

- Variance conditionnelle dépendante du temps puisque :

$$V(\varepsilon_t | \varepsilon_{t-h}) = \alpha_0 \left( \frac{1 - \alpha_1^h}{1 - \alpha_1} \right) + \alpha_1^h \varepsilon_{t-h}^2 \quad V(\varepsilon_t) = \frac{\alpha_0}{1 - \alpha_1} \quad (2.28)$$

- Les autocovariances conditionnelles sont nulles  $cov(\varepsilon_t, \varepsilon_{t+k} | \varepsilon_{t-h}) = 0, \forall h \geq 1, \forall k \geq 1$
- Sous l'hypothèse  $\alpha_1^2 < \frac{1}{3}$ , la distribution des résidus est leptokurtique puisque :

$$Kurtosis = 3 \left( \frac{1 - \alpha_1^2}{1 - 3\alpha_1^2} \right) > 3 \quad (2.29)$$

On peut en outre en déduire un certain nombre de conclusions quant au processus  $Y_t$  lui-même. On peut montrer tout abord que l'espérance conditionnelle de  $Y_t$  vérifie :

$$\begin{aligned} E(Y_t | Y_{t-h}) &= \mu + \rho E(Y_{t-1} | Y_{t-h}) \\ &= \mu \left( \frac{1 - \rho^h}{1 - \rho} \right) + \rho^h Y_{t-h} \end{aligned} \quad (2.30)$$

De la même façon, on montre que la variance conditionnelle de  $Y_t$  dépend du temps. En effet, on peut montrer qu'elle dépend du processus  $\varepsilon_{t-h}^2$  de la façon suivante

**Proposition 12** *La variance conditionnelle du processus AR(1) avec erreur ARCH(1),  $Y_t$  s'écrit :*

$$\begin{aligned} V(Y_t | Y_{t-h}) &= \left( \frac{\mu}{1 - \alpha_1} \right) \left[ \left( \frac{1 - \rho^{2h}}{1 - \rho^2} \right) - \alpha_1 \left( \frac{\alpha_1^h - \rho^{2h}}{\alpha_1 - \rho^2} \right) \right] \\ &+ \alpha_1 \left( \frac{\alpha_1^h - \rho^{2h}}{\alpha_1 - \rho^2} \right) \varepsilon_{t-h}^2 \end{aligned} \quad (2.31)$$

Ainsi la variance d'une erreur de prévision à l'horizon 1, s'écrit :

$$V(Y_t | Y_{t-h}) = \mu + \alpha_1 \varepsilon_{t-1}^2 \quad (2.32)$$

En conclusion, si l'on désire le processus  $Y_t$  dans le cas d'erreur  $ARCH(1)$ , l'erreur de prévision à un horizon d'une période admet une variance  $V(Y_t | Y_{t-1})$  qui varie dans le temps en fonction de la valeur de  $\varepsilon_{t-1}^2$ . Dès lors, les intervalles de confiance sur cet prévision, contrairement aux  $AR(1)$ , ne sont plus constants dans le temps.

**Remarque 13** *La variance d'une erreur de prévision sur un processus avec erreur ARCH dépend du temps.*

$$V(Y_t | Y_{t-h}) = g(\varepsilon_{t-h}) \quad (2.33)$$

*L'amplitude des intervalles de confiance associés à cette prévision n'est donc pas constante dans le temps.*

### 2.3.3 Modèles $GARCH(p, q)$

#### Définition 6

*Un processus  $\varepsilon_t$  satisfait une représentation  $GARCH(p, q)$  si*

$$\varepsilon_t = z_t \sqrt{h_t} \quad (2.34)$$

$$h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i} \quad (2.35)$$

où  $z_t$  est une suite de v.a (i.i.d)gaussiennes centrées réduites et  $\alpha_0 > 0, \alpha_i \geq 0, i = 1, \dots, q$  et  $\beta_i \geq 0, i = 1, \dots, p$ . Ainsi, l'erreur du processus  $Y_t$ , définie par le processus  $GARCH(p, q)$   $\varepsilon_t$  admet pour moments conditionnels :

$$E(\varepsilon_t | \varepsilon_{t-1}) = 0 \quad (2.36)$$

$$V(\varepsilon_t | \varepsilon_{t-1}) = h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i} \quad (2.37)$$

**Proposition 14** *Si le processus  $\varepsilon_t$  satisfait une représentation  $GARCH$  conditionnellement Gaussienne, telle que :*

$$\varepsilon_t = z_t \sqrt{h_t}$$

$$V(\varepsilon_t | \varepsilon_{t-1}) = h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i} \quad (2.38)$$

où  $z_t$  est une suite de v.a (i.i.d) gaussiennes centrées réduites, alors (i) la loi marginale de  $\varepsilon_t$  a des queues plus épaisses que celles d'une loi normale (distribution leptokurtique) :



$$E(\varepsilon_t^4) \geq 3 [E(\varepsilon_t^2)]^2 \quad (2.39)$$

et (ii) son coefficient d'excès de kurtosis peut s'exprimer sous la forme suivante :

$$\text{Excès de kurtosis} = \frac{E(\varepsilon_t^4)}{E(\varepsilon_t^2)^2} - 3 = 3 \frac{\text{Var} \left[ E(\varepsilon_t^2 | \varepsilon_{t-1}) \right]}{E(\varepsilon_t^2)^2} \quad (2.40)$$

## 2.4 Quelques extensions des modèles ARCH-GARCH

### 2.4.1 Modèles ARMA – GARCH

C'est Weiss (1986) qui a introduit dans la variance conditionnelle des effets additionnels de variables expliquées. En effet, la modélisation *GARCH* peut être appliquée non au processus initial, mais au processus d'innovation. Ceci permet alors d'introduire divers effets additionnels de variables explicatives soit dans la moyenne conditionnelle, soit dans la variance conditionnelle. Par exemple, on peut considérer un modèle de régression linéaire avec erreurs *GARCH* :

$$y_t = x_t b + \varepsilon_t \quad \varepsilon_t \sim \text{GARCH}(p, q) \quad (2.41)$$

On peut aussi considérer un modèle *ARMA* avec erreur *GARCH*( $p, q$ ) :

$$\Phi(L) y_t = \Theta(L) \varepsilon_t \quad \varepsilon_t \sim \text{GARCH}(p, q) \quad (2.42)$$

Enfin, on peut concevoir un modèle *ARMA* dans lequel la variance non conditionnelle de  $y$  peut avoir un effet sur la variance conditionnelle :

$$\Phi(L) y_t = \Theta(L) \varepsilon_t \quad (2.43)$$

et

$$E(\varepsilon_t | \varepsilon_{t-1}) = 0 \quad (2.44)$$

$$V(\varepsilon_t | \varepsilon_{t-1}) = c + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \gamma_0 [E(y_t | y_{t-1})]^2 + \sum_{j=1}^p \gamma_j y_{t-j}^2 \quad (2.45)$$

### 2.4.2 Modèles GARCH – M

Un processus  $y_t$  satisfait une représentation *GARCH – M* linéaire en variance conditionnelle si et seulement si :

$$\begin{aligned} y_t &= x_t b + \delta h_t + \varepsilon_t \\ &= x_t b + \delta V(\varepsilon_t | \varepsilon_{t-1}) + \varepsilon_t \end{aligned} \quad (2.46)$$

où

$$\varepsilon_t = z_t \sqrt{h_t} \quad (2.47)$$

où  $z_t \sim \mathcal{N}(0,1)$  et

$$h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i} \quad (2.48)$$

avec  $E(\varepsilon_t | \varepsilon_{t-1}) = 0$  et  $V(\varepsilon_t | \varepsilon_{t-1}) = h_t$

### 2.4.3 Modèles *IGARCH*

Les processus *IGARCH* proposés par Engle et Bollerslev (1986), correspondent au cas d'une racine unitaire dans le processus de variance conditionnelle. Ces modèles sont alors par un effet de persistance dans la variance. C'est-à-dire qu'un choc sur la variance conditionnelle actuelle se répercute sur toutes les valeurs futures prévues. Un processus  $\varepsilon_t$  satisfait une représentation *IGARCH* si et seulement si :

$$\begin{aligned} V(\varepsilon_t | \varepsilon_{t-1}) &= h_t \\ &= \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i} \end{aligned} \quad (2.49)$$

avec  $\alpha_0 \geq 0, \alpha_i \geq 0$  pour  $i = 1, \dots, q$  et  $\beta_j \geq 0, j = 1, \dots, p$  et

$$\sum_{i=1}^q \alpha_i + \sum_{i=1}^p \beta_i = 1 \quad (2.50)$$

### 2.4.4 Modèles *ARCH-GARCH* asymétriques

La seconde grande approche couvre les modèles *ARCH* non linéaire et plus particulièrement la prise en compte des phénomènes d'asymétries. Deux grandes classes de modèles ont été proposées :

- Nelson (1990) s'est intéressé aux évolutions asymétriques de la variance à l'aide des modèles *EGARCH*.
- Engle et Bollerslev (1986) ont étudié les modèles à seuil (*TARCH*) où la variance est une fonction linéaire définie par morceaux qui permet différentes fonctions de volatilité selon le signe et la valeur des chocs.

### 2.4.5 Modèles *EGARCH*.

Proposé par Nelson (1991), le processus Exponential *GARCH* ou *EGARCH* ( $p, q$ ), donne à la variance conditionnelle la définition suivante :

Un processus  $\varepsilon_t$  satisfait une représentation *EGARCH*  $(p, q)$  si et seulement si :

$$\varepsilon_t = z_t \sqrt{h_t} \quad (2.51)$$

$$\log(h_t) = \alpha_0 + \sum_{i=1}^q \alpha_i g(z_{t-i}) + \sum_{i=1}^p \beta_i \log(h_{t-i}) \quad (2.52)$$

où  $z_t$  est une suite de v.a, i.i.d gaussiennes centées réduites et où la fonction  $g(\cdot)$  vérifie :

$$g(z_{t-i}) = \theta z_{t-i} + \gamma (|z_{t-i}| - E|z_{t-i}|) \quad (2.53)$$

### 2.4.6 Les modèles *MAR* – *ARCH*

Introduit par Wong et Li (2001), pour une meilleure capture de l'hétéroscédasticité et une meilleure modélisation des séries temporelles ayant des valeurs aberrantes, des valeurs explosives, des distributions conditionnelles multimodales, et d'autres propriétés de non linéarité et non Gaussiennes. Ce type de modèles est composé d'un mélange de  $K$  modèles autorégressifs à erreur *ARCH* c'est-à-dire la moyenne conditionnelle suit un mélange de modèles *AR* et la variance conditionnelle suit un mélange de modèles *ARCH*. Les détails de la description de ce modèle, ses importantes propriétés, l'estimation de ses paramètres et la sélection du modèle sont traités dans les chapitres suivants.

CHAPITRE

# 3

---

## Mélange de modèles autorégressifs conditionnellement hétéroscédastiques

Dans ce chapitre, nous présentons la généralisation du modèle MAR, étudié brièvement dans le chapitre 1, au mélange de modèles autorégressifs conditionnellement hétéroscédastiques (MAR-ARCH), proposé par wong et Li (2001) pour la modélisation des séries temporelles non linéaires. Le modèle MAR-ARCH constitué d'un mélange de  $K$  modèles autorégressifs conditionnellement hétéroscédastiques, tel que la moyenne conditionnelle du processus suit un modèle MAR et la variance conditionnelle du processus suit un mélange de modèles ARCH. En plus de la meilleur description de la distribution conditionnelle obtenue par le modèle MAR, le modèle MAR-ARCH permet une structure d'autocorrélation du processus carré plus flexible. Les conditions de stationnarité, la fonction d'autocorrélation, la fonction d'autocorrélation du processus carré sont dérivées. Et enfin, nous utilisons l'approche de Monte Carlo pour le calcul de la distribution prédictive au pas multiple.

### 3.1 Introduction

Dans la section précédente, nous avons présenté les modèles MAR comme étant un moyen utile pour modéliser les séries temporelles non linéaires. Rappelons qu'un modèle MAR est défini par :

$$F(y_t | \Omega_t) = \sum_{k=1}^K \alpha_k \Phi \left( \frac{y_t - \phi_{k0} - \dots - \phi_{kp_k} y_{t-p_k}}{\sigma_k} \right) \quad (3.1)$$

où  $\Omega(\cdot)$  est la fonction de distribution cummulative gaussienne standard. Certaines propriétés permettent à ces modèles d'être prometteurs : la possibilité de mélanger un processus AR non stationnaire avec un processus AR stationnaire, la capacité de modéliser les distributions conditionnelles multimodales, et l'aptitude aussi à capturer l'hétéroscédasticité conditionnelle. Wong et Li (2000) ont montré l'utilité des modèles MAR avec quelques données financières.

Malgré les propriétés intéressantes des modèles MAR, la modélisation des séries temporelles non linéaires utilisant ces modèles montre une limitation. Précisément aux données financières car, la structure d'autocorrélation carré du modèle MAR est assez simple, et elle est analogue à celle du modèle AR (Granger et Newbold 1986 p 309; Wong 1998).

Dans ce chapitre, nous présentons la généralisation des modèle MAR, prposée par Wong et Li (2001), mélanges de modèles autorégressifs conditionnellement hétéroscédastique (MAR-ARCH). Ce type de modèle constitué de K composantes de modèles autorégressifs à erreur ARCH. C'est-à-dire la moyenne conditionnelle suit un mélange de processus autorégressifs (MAR) et la variance conditionnelle suit un mélange de processus ARCH.

Les modèles MAR-ARCH permettent plus de flexibilité dans la structure d'auto-corrélation du carré du processus. Les conditions de stationnarité, la fonction d'auto-corrélation, la distribution de prédictive au pas multiple et la fonction d'autocorrélation du carré du processus sont élaborées. L'estimation peut être facilement faite à travers un simple algorithme EM, concernant l'identification, utilisant le critère BIC pour sélectionner le modèle. La propriété de changement de la forme des distributions conditionnelles permet à ces modèles d'être capables de modéliser les séries temporelles avec distributions conditionnelles multimodales et avec hétéroscédasticité.

### 3.2 Mélange de modèles autorégressifs conditionnellement hétéroscédastiques

Le modèle MAR-ARCH de K composantes est défini par

$$F(y_t | \Omega_{t-1}) = \sum_{k=1}^K \alpha_k \Phi \left( \frac{e_{k,t}}{\sqrt{h_{k,t}}} \right) \quad (3.2)$$

où

$$e_{k,t} = y_t - \phi_{k0} - \cdots - \phi_{kp_k} y_{t-p_k}$$

et

$$h_{k,t} = \beta_{k0} + \beta_{k1} e_{k,t-1}^2 + \cdots + \beta_{kq_k} e_{k,t-q_k}^2$$

Notons ce modèle par  $MAR-ARCH(K; p_1, \dots, p_K; q_1, \dots, q_K)$ .  $F(y_t | \Omega_{t-1})$  est la fonction de distribution cummulative conditionnelle de  $y_t$  donnant l'information passées, évaluée à  $y_t$  ; établie à l'instant  $t$  ;  $\Phi(\cdot)$  est la fonction de distribution cummulative de la distribution gaussienne standard ;  $h_{k,t}$  est la variance conditionnelle de  $e_{k,t}$  et  $\alpha_1 + \alpha_2 + \cdots + \alpha_K = 1, \alpha_k > 0 (k = 1, \dots, K)$ . Pour éviter la possibilité de la négation ou nullité de la variance, les conditions suivantes sur les  $\beta_{ki}$  doivent être imposées :  $\beta_{k0} > 0 (k = 1, \dots, K)$ , et  $\beta_{ki} \geq 0 (i = q_1, \dots, q_K; k = 1, \dots, K)$ . Le modèle MAR-ARCH considéré est un mélange de  $K$  modèles autorégressifs à innovations ARCH. Les propriétés du modèle MAR-ARCH sont similaires à celles du modèle MAR de Wong et Li (2000). Par exemple, les moyennes conditionnelles et les variances des composantes changent dans le temps, elles dépendent des valeurs passées de la série temporelle. La forme de la distribution conditionnelle de la série change dans le temps, et elle peut changer d'unimodale au multimodale. Ce qui implique que l'espérance conditionnelle de  $y_t$  donnant le passé,

$$\begin{aligned} E(y_t | \Omega_{t-1}) &= \sum_{k=1}^K \alpha_k (\phi_{k0} + \cdots + \phi_{kp_k} y_{t-p_k}) \\ &= \sum_{k=1}^K \alpha_k \mu_{k,t} \end{aligned} \quad (3.3)$$

où  $\mu_{k,t} = \phi_{k0} + \cdots + \phi_{kp_k} y_{t-p_k}$ , peut perdre sa performance dans l'estimation des valeurs futurs. Le modèle MAR-ARCH possède plus de flexibilité dans la modélisation de la variance conditionnelle changeante. La variance conditionnelle de  $y_t$  donnant le passé est donnée par :

$$var(y_t | \Omega_{t-1}) = \sum_{k=1}^K \alpha_k h_{k,t} + \sum_{k=1}^K \alpha_k \mu_{k,t}^2 - \left( \sum_{k=1}^K \alpha_k \mu_{k,t} \right)^2 \quad (3.4)$$

Le premier terme permet la modélisation de la dépendance de la variance conditionnelle des erreurs passées. Le second et le troisième terme modélisent le changement de la variance conditionnelle résultant de la différence dans les moyennes conditionnelles des composantes.

### 3.3 Etude des modèles MAR -ARCH

#### 3.3.1 Étude de la stationnarité

Les conditions de stationnarité au premier ordre pour les modèles  $MAR - ARCH$  sont données dans le théorème (15). La dérivation des conditions de stationnarité au deuxième ordre pour un modèle  $MAR - ARCH$  général est plus compliquée et nécessite les termes d'autocovariance et l'espérance de  $e_{k,t}$ . Nous présentons uniquement les conditions de stationnarité au second ordre pour les modèles  $MAR - ARCH(K; 1, \dots, 1; q_1, \dots, q_K)$  dans le théorème (16). Les conditions de stationnarité au second ordre pour les modèles  $MAR - ARCH(K; 0, \dots, 0; q_1, \dots, q_K)$  peuvent être obtenues par la supposition que  $\phi_{k1} = 0, k = 1, \dots, K$  dans le théorème (16). Soit  $p = \max(p_1, \dots, p_K), q = \max(q_1, \dots, q_K)$ .

**Théorème 15** (*Wong et Li (2001)*)

*Une condition nécessaire et suffisante pour que le processus  $Y_t$  soit stationnaire en moyenne est que les racines de l'équation :*

$$1 - \sum_{i=1}^p \left( \sum_{k=1}^K \alpha_k \phi_{ki} \right) z^{-i} = 0 \quad (3.5)$$

sont toutes à l'intérieur du cercle unité, où  $\phi_{ki} = 0$  pour  $i > p_k$ .

**Preuve.** Soient :

$\mu_t = E(y_t), \mu_{t,k} = \phi_{k0} + \phi_{k1}y_{t-1} + \dots + \phi_{kp_k}y_{t-p_k}, g_k(y_t) = \left( \frac{1}{\sqrt{h_{k,t}}} \right) \varphi \left( \frac{e_{k,t}}{\sqrt{h_{k,t}}} \right)$  et  $C$  est une constante indépendante de  $t$ . Tel que  $\varphi$  est la fonction de densité de probabilité de la distribution normale standard et  $k = 1, \dots, K$ . La moyenne conditionnelle de  $y_t$  est donnée par :

$$\begin{aligned} E(y_t | \Omega_{t-1}) &= \int y_t f(y_t | \Omega_{t-1}) dy_t \\ &= \int y_t \sum_{k=1}^K \alpha_k g_k(y_t) dy_t \\ &= \sum_{k=1}^K \alpha_k \int y_t \left( \frac{1}{\sqrt{h_{k,t}}} \right) \varphi \left( \frac{e_{k,t}}{\sqrt{h_{k,t}}} \right) dy_t \\ &= \sum_{k=1}^K \alpha_k \int (y_t - \mu_{k,t} + \mu_{k,t}) \left( \frac{1}{\sqrt{h_{k,t}}} \right) \varphi \left( \frac{e_{k,t}}{\sqrt{h_{k,t}}} \right) dy_t \\ &= \sum_{k=1}^K \alpha_k \mu_{k,t} \int \left( \frac{1}{\sqrt{h_{k,t}}} \right) \varphi \left( \frac{e_{k,t}}{\sqrt{h_{k,t}}} \right) dy_t \\ &+ \sum_{k=1}^K \alpha_k \int (y_t - \mu_{k,t}) \left( \frac{1}{\sqrt{h_{k,t}}} \right) \varphi \left( \frac{e_{k,t}}{\sqrt{h_{k,t}}} \right) dy_t \end{aligned} \quad (3.6)$$

Le premier terme égale à  $\sum_{k=1}^K \alpha_k \mu_{k,t}$  car :

$$\int \left( \frac{1}{\sqrt{h_{k,t}}} \right) \varphi \left( \frac{e_{k,t}}{\sqrt{h_{k,t}}} \right) dy_t = 1 \quad (3.7)$$

(faisant le changement de variable  $z_t = \frac{e_{k,t}}{\sqrt{h_{k,t}}}$ ). Le deuxième terme égale à zéro, en effet : Posons  $z_t = \frac{e_{k,t}}{\sqrt{h_{k,t}}}$  ( $Z_t \sim \mathcal{N}(0, 1)$ ) alors

$$\begin{aligned} \sum_{k=1}^K \alpha_k \int (y_t - \mu_{k,t}) \left( \frac{1}{\sqrt{h_{k,t}}} \right) \varphi \left( \frac{e_{k,t}}{\sqrt{h_{k,t}}} \right) dy_t &= \sum_{k=1}^K \alpha_k \int \sqrt{h_{k,t}} z_t \varphi(z_t) dz_t \\ &= \sum_{k=1}^K \alpha_k \sqrt{h_{k,t}} E(z_t) \\ &= 0 \end{aligned} \quad (3.8)$$

d'où

$$E(y_t | \Omega_{t-1}) = \sum_{k=1}^K \alpha_k \mu_{k,t} \quad (3.9)$$

$$\begin{aligned} \mu_t &= E(y_t) \\ &= E(E(y_t | \Omega_{t-1})) \\ &= E \left( \sum_{k=1}^K \alpha_k (\phi_{k0} + \phi_{k1} y_{t-1} + \cdots + \phi_{kp_k} y_{t-p_k}) \right) \\ &= \sum_{k=1}^K \alpha_k E(\phi_{k0} + \phi_{k1} y_{t-1} + \cdots + \phi_{kp_k} y_{t-p_k}) \\ &= \sum_{k=1}^K \alpha_k (\phi_{k0} + \phi_{k1} \mu_{t-1} + \cdots + \phi_{kp_k} \mu_{t-p_k}) \end{aligned} \quad (3.10)$$

Notons  $p = \max(p_1, \dots, p_K)$  et supposant que les  $\phi_{k,i} = 0$  si  $i > p_k$ . Alors nous pouvons écrire sous la forme suivante :

$$\mu_t = \sum_{i=1}^p \left( \sum_{k=1}^K \alpha_k \phi_{k,i} \right) \mu_{t-i} + C \quad (3.11)$$

tel que

$$C = \sum_{k=1}^K \alpha_k \phi_{k,0} \quad (3.12)$$

La condition nécessaire et suffisante pour que cette équation aux différences non homogènes possède une solution stable, finie et indépendante du temps est que les racines de l'équation :



$$1 - \sum_{i=1}^p \left( \sum_{k=1}^K \alpha_k \phi_{ki} \right) z^{-i} = 0 \quad (3.13)$$

où  $z_1, \dots, z_p$  sont toutes à l'intérieur du cercle unité (Goldberg 1958). Soit  $\mu$  la mesure correspondante à la distribution du processus à l'instant 0 ; qui est le mélange fini de distributions gaussiennes, et alors  $\mu \in \mathbb{C}^+$ . Ici  $\mathbb{C}^+$  est le cône strictement positif de l'espace de Banach comme il est défini par Benes (1967). Donc, d'après le résultat de Benes il existe une mesure finie et invariante pour le processus  $y_t$ . ■

**Théorème 16** (Wong et Li (2001))

Supposons que le processus  $y_t$  suivant le modèle MAR – ARCH ( $K; 1, \dots, 1; q_1, \dots, q_K$ ) est stationnaire au premier ordre. Une condition nécessaire et suffisante pour que le processus soit stationnaire au second ordre est que les racines  $z_1, \dots, z_p$  de l'équation suivante sont toutes à l'intérieur du cercle unité :

$$\begin{aligned} & 1 - \sum_{k=1}^K \alpha_k (\beta_{k1} + \phi_{k1}^2) z^{-1} \\ & - \sum_{i=2}^q \sum_{k=1}^K \alpha_k \left[ \beta_{ki} + \beta_{k,i-1} \phi_{k1} \left\{ \phi_{k1} - 2 \left( \sum_{j=1}^K \alpha_j \phi_{j1} \right) \right\} \right] z^{-i} \\ & - \sum_{k=1}^K \alpha_k \beta_{kq} \phi_{k1} \left\{ \phi_{k1} - 2 \left( \sum_{j=1}^K \alpha_j \phi_{j1} \right) \right\} z^{-q-1} = 0 \end{aligned} \quad (3.14)$$

où  $\beta_{ki} = 0$  pour  $i > q_k$

**Preuve.** Soient :

$$\phi_{k0} = 0 \ (k = 1, \dots, K), \gamma_{it} = E(y_t y_{t-i}), g_k(y_t) = \left( \frac{1}{\sqrt{h_{k,t}}} \right) \varphi \left\{ \frac{(y_t - \phi_{k,1} y_{t-1})}{\sqrt{h_{k,t}}} \right\}.$$

Le moment conditionnelle d'ordre deux de  $y_t$  est donné par :

$$\begin{aligned}
E(y_t^2 | \Omega_{t-1}) &= \int y_t^2 f(y_t | \Omega_{t-1}) dy_t \\
&= \int y_t^2 \sum_{k=1}^K \alpha_k g_k(y_t) dy_t \\
&= \sum_{k=1}^K \alpha_k \int (y_t - \phi_{k1} y_{t-1} + \phi_{k1} y_{t-1})^2 g_k(y_t) dy_t \\
&= \sum_{k=1}^K \alpha_k \int \{(y_t - \phi_{k1} y_{t-1})^2 + 2(y_t - \phi_{k1} y_{t-1}) \phi_{k1} y_{t-1} + (\phi_{k1} y_{t-1})^2\} g_k(y_t) dy_t \\
&= \sum_{k=1}^K \alpha_k \int (y_t - \phi_{k1} y_{t-1})^2 g_k(y_t) dy_t + \sum_{k=1}^K \alpha_k \int 2(y_t - \phi_{k1} y_{t-1}) \phi_{k1} y_{t-1} g_k(y_t) dy_t \\
&\quad + \sum_{k=1}^K \alpha_k \int (\phi_{k1} y_{t-1})^2 g_k(y_t) dy_t \tag{3.15}
\end{aligned}$$

Le premier terme égale à  $\sum_{k=1}^K \alpha_k h_{k,t}$  en effet

$$\begin{aligned}
&\sum_{k=1}^K \alpha_k \int (y_t - \phi_{k1} y_{t-1})^2 g_k(y_t) dy_t \\
&= \sum_{k=1}^K \alpha_k \int h_{k,t} \left( \frac{y_t - \phi_{k1} y_{t-1}}{\sqrt{h_{k,t}}} \right)^2 \frac{1}{\sqrt{h_{k,t}}} \varphi \left( \frac{y_t - \phi_{k1} y_{t-1}}{\sqrt{h_{k,t}}} \right) dy_t \\
&= \sum_{k=1}^K \alpha_k h_{k,t} E \left( \frac{y_t - \phi_{k1} y_{t-1}}{\sqrt{h_{k,t}}} \right)^2 \\
&= \sum_{k=1}^K \alpha_k h_{k,t} \tag{3.16}
\end{aligned}$$

Car  $\frac{y_t - \phi_{k1} y_{t-1}}{\sqrt{h_{k,t}}} \sim \mathcal{N}(0, 1)$ . Le deuxième terme égale à zéro, en effet :

Posons  $z_t = \frac{y_t - \phi_{k1} y_{t-1}}{\sqrt{h_{k,t}}}$  on obtient alors

$$\begin{aligned}
\int (y_t - \phi_{k1} y_{t-1}) \phi_{k1} y_{t-1} g_k(y_t) dy_t &= \int \phi_{k1} y_{t-1} \sqrt{h_{k,t}} z_t \varphi(z_t) dz_t \\
&= \phi_{k1} y_{t-1} \sqrt{h_{k,t}} E(z_t) \\
&= 0 \tag{3.17}
\end{aligned}$$

Le troisième terme égale à  $\sum_{k=1}^K \alpha_k (\phi_{k1} y_{t-1})^2$  car  $\int g_k(y_t) dy_t = 1$ . d'où

$$E(y_t^2 | \Omega_{t-1}) = \sum_{k=1}^K \alpha_k h_{k,t} + \sum_{k=1}^K \alpha_k (\phi_{k1} y_{t-1})^2 \tag{3.18}$$

Pour obtenir le moment d'ordre deux du processus  $y_t$ , nous considérons tout d'abord :

$$\begin{aligned}
 E(e_{k,t}^2) &= E((y_t - \phi_{k1}y_{t-1})^2) \\
 &= E(y_t^2) - 2E(\phi_{k1}y_t y_{t-1}) + \phi_{k1}^2 E(y_{t-1}^2) \\
 &= \gamma_{0,t} + \left\{ \phi_{k1}^2 - 2\phi_{k1} \left( \sum_{k=1}^K \alpha_k \phi_{k1} \right) \right\} \gamma_{0,t-1}
 \end{aligned} \tag{3.19}$$

en effet

$$\begin{aligned}
 E(y_t y_{t-1}) &= E(y_{t-1} E(y_t | \Omega_{t-1})) \\
 &= E\left(y_{t-1} \left( \sum_{k=1}^K \alpha_k \phi_{k1} y_{t-1} \right)\right) \\
 &= \left( \sum_{k=1}^K \alpha_k \phi_{k1} \right) E(y_{t-1}^2)
 \end{aligned} \tag{3.20}$$

Le moment d'ordre deux de  $Y_t$  est donné par :

$$\begin{aligned}
 \gamma_{0,t} &= E\{E(y_t^2 | \Omega_{t-1})\} \\
 &= E\left\{ \sum_{k=1}^K \alpha_k (\beta_{k0} + \beta_{k1} e_{k,t-1}^2 + \dots + \beta_{kq} e_{k,t-q}^2) + \sum_{k=1}^K \alpha_k \phi_{k1}^2 y_{t-1}^2 \right\} \\
 &= \sum_{k=1}^K \alpha_k \beta_{k0} + \sum_{i=1}^q \sum_{k=1}^K \alpha_k \beta_{ki} E(e_{k,t-i}^2) + \sum_{k=1}^K \alpha_k \phi_{k1}^2 E(y_{t-1}^2) \\
 &= \sum_{k=1}^K \alpha_k \beta_{k0} + \sum_{i=1}^q \sum_{k=1}^K \alpha_k \beta_{ki} \left[ \gamma_{0,t-i} + \left\{ \phi_{k1}^2 - 2\phi_{k1} \left( \sum_{j=1}^K \alpha_j \phi_{j1} \right) \right\} \gamma_{0,t-i-1} \right] + \sum_{k=1}^K \alpha_k \phi_{k1}^2 \gamma_{0,t-1} \\
 &= C + \sum_{k=1}^K \alpha_k (\beta_{k1} + \phi_{k1}^2) \gamma_{0,t-1} + \sum_{i=2}^q \sum_{k=1}^K \alpha_k \left[ \beta_{ki} + \beta_{ki-1} \left\{ \phi_{k1}^2 - 2\phi_{k1} \left( \sum_{j=1}^K \alpha_j \phi_{j1} \right) \right\} \right] \gamma_{0,t-i} \\
 &+ \sum_{k=1}^K \alpha_k \beta_{kq} \left\{ \phi_{k1}^2 - 2\phi_{k1} \left( \sum_{j=1}^K \alpha_j \phi_{j1} \right) \right\} \gamma_{0,t-1-q}
 \end{aligned} \tag{3.21}$$

tel que C est une constante indépendante du temps. La condition nécessaire et suffisante pour que cette équation aux différences non homogènes possède une solution stable, finie et indépendante du temps est que les racines de l'équation :

$$\begin{aligned}
 & 1 - \sum_{k=1}^K \alpha_k (\beta_{k1} + \phi_{k1}^2) z^{-1} \\
 & - \sum_{i=2}^q \sum_{k=1}^K \alpha_k \left[ \beta_{ki} + \beta_{ki-1} \phi_{k1} \left\{ \phi_{k1} - 2 \left( \sum_{j=1}^K \alpha_j \phi_{j1} \right) \right\} \right] z^{-i} \\
 & - \sum_{k=1}^K \alpha_k \beta_{kq} \phi_{k1} \left\{ \phi_{k1} - 2 \left( \sum_{j=1}^K \alpha_j \phi_{j1} \right) \right\} z^{-q-1} = 0
 \end{aligned} \tag{3.22}$$

où  $z_1, \dots, z_{q+1}$  sont toutes à l'intérieur du cercle unité (Goldberg 1958). Soit  $\mu$  est la mesure correspondant à la distribution du processus à l'instant 0; qui est le mélange fini de distributions gaussiennes, et alors  $\mu \in \mathbb{C}^+$ . Donc, par le résultat de Benes il existe une mesure finie et invariante pour le processus  $Y_t$ . ■

**Remarque 17** *On remarque à partir de ces théorèmes que les conditions de stationnarité dépendent largement des proportions des composantes, i.e les valeurs des  $\alpha_k$ .*

**Proposition 18** *Un mélange d'un modèle AR – ARCH non stationnaire avec un modèle AR – ARCH stationnaire peut donner un processus stationnaire.*

Nous illustrons ce point avec une série simulée à partir de modèle (C) : *MAR – ARCH* (2; 1, 1; 1, 1) définit par :

$$F(y_t | \Omega_{t-1}) = 0.75\Phi(e_{1,t} | \sqrt{h_{1,t}}) + 0.25\Phi(e_{2,t} | \sqrt{h_{2,t}}) \tag{3.23}$$

tel que

$$\begin{aligned}
 e_{1,t} &= y_t - 0.5y_{t-1} \\
 h_{1,t} &= 1 + 0.5e_{1,t-1}^2 \\
 e_{2,t} &= y_t - 1.1y_{t-1} \\
 h_{2,t} &= 1 + 1.2e_{2,t-1}^2
 \end{aligned}$$

La série simulée est présentée dans la figure (3.1)

### 3.3.2 Fonction d'autocorrélation

Pour un processus stationnaire au second ordre  $Y_t$  suivant un modèle MAR-ARCH, les autocorrélations sont similaires à celles du modèle MAR. Pour  $j = 1, \dots, p$

$$\rho_j = \sum_{i=1}^p \left( \sum_{k=1}^K \alpha_k \phi_{ki} \right) \rho |j - i| \tag{3.24}$$

où  $\rho_j$  est l'autocorrélation de retard  $j$ . En effet :  $\rho(j) = \frac{\gamma(j)}{\gamma(0)}$  tel que  $\gamma(j) = E(y_t y_{t-j})$ ,  $p = \max(p_1, \dots, p_K)$ ,  $\phi_{ki} = 0$  si  $i > p_k$  et  $\phi_{k0} = 0$  pour  $k = 1, \dots, K$

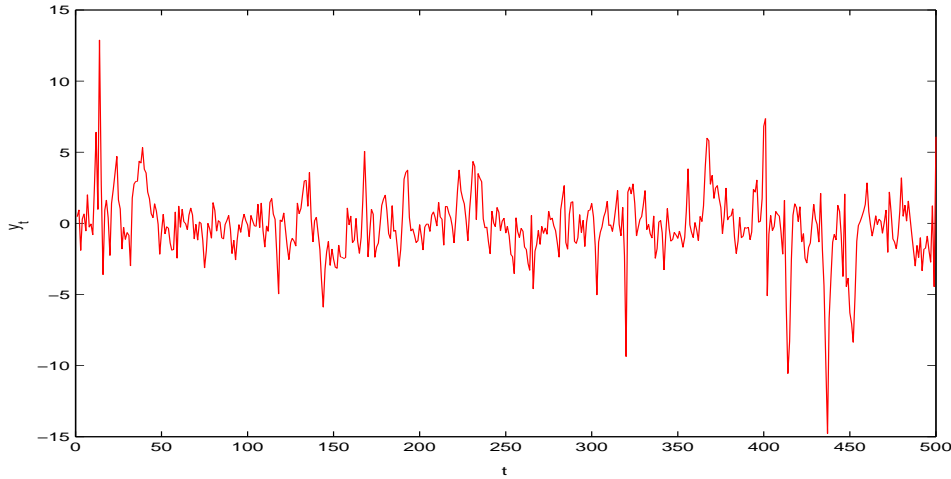


Fig. 3.1 — La série simulée à partir du modèle (C).

$$\begin{aligned}
 \gamma(j) &= E(y_t y_{t-j}) \\
 &= E(y_{t-j} E(y_t | \Omega_{t-1})) \\
 &= E\left(y_{t-j} \left(\sum_{k=1}^K \alpha_k (\phi_{k0} + \phi_{k1} y_{t-1} + \cdots + \phi_{kp} y_{t-p})\right)\right) \\
 &= \sum_{k=1}^K \alpha_k E(\phi_{k0} y_{t-j} + \phi_{k1} y_{t-1} y_{t-j} + \cdots + \phi_{kp} y_{t-p} y_{t-j}) \\
 &= \sum_{k=1}^K \alpha_k (0 + \phi_{k1} E(y_{t-1} y_{t-j}) + \cdots + \phi_{kp} E(y_{t-p} y_{t-j})) \\
 &= \sum_{k=1}^K \alpha_k \sum_{i=1}^p \gamma(|j-i|) \phi_{ki}
 \end{aligned} \tag{3.25}$$

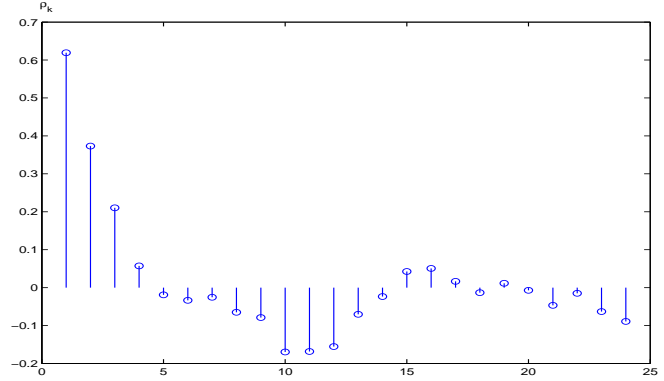
**Remarque 19** Notons que ces équations sont similaires à celles des équations de Yule-Walker pour un processus AR ( $p$ ) ordinaire, où les coefficients  $\sum_{k=1}^K \alpha_k \phi_{ki}$  remplacent les coefficients du  $i^{\text{ème}}$  retard du processus AR ( $p$ ).

Les autocorrélations de la série simulée à partir du modèle (C) sont présentées dans la figure (3.2)

Le rayon des autocorrélations possibles est plus grand que celui du processus AR standard, puisque le modèle AR est juste un cas spécial de modèle MAR-ARCH. (Voir Le et al, 1996 et Wong et Li, 2000).

### 3.3.3 Moments de modèles MAR-ARCH

Nous pouvons obtenir la condition d'existence des moments d'ordre élevé pour un modèle MAR-ARCH général. Néanmoins la dérivation est très compliquée, par ce que



**Fig. 3.2** — Les autocorrélations estimées de la série simulée à partir du modèle (C).

elle nécessite l'espérance du produit des  $y_t^2$ . Ici nous donnons uniquement la condition d'existence du moment d'ordre quatre du modèle  $MAR - ARCH(K; 0, \dots, 0; 1, \dots, 1)$  avec  $\phi_{k0} = 0$  ( $k = 1, \dots, K$ ) dans le théorème (20). Notons que les moments d'ordre un et trois sont nulles, par contre le moment d'ordre deux est donné par :

$$\frac{\sum_{k=1}^K \alpha_k \beta_{k0}}{\left(1 - \sum_{k=1}^K \alpha_k \beta_{k1}\right)} \quad (3.26)$$

**Théorème 20** (*Wong et Li (2001)*)

Le moment d'ordre 4 pour un modèle  $MAR - ARCH(K; 0, \dots, 0; 1, \dots, 1)$  stationnaire avec  $\phi_{k0} = 0$  ( $k = 1, \dots, K$ ) existe si et seulement si

$$\sum_{k=1}^K \alpha_k \beta_{k1}^2 < \frac{1}{3} \quad (3.27)$$

**Preuve.**

Il est facile de voir que  $\mu_{k,t} = 0$ , ( $k = 1, \dots, K$ ) et par conséquent  $E(y_t) = 0$ . Soit  $g_k(y_t) = \left(\frac{1}{\sqrt{h_{k,t}}}\right) \varphi\left(\frac{y_t}{\sqrt{h_{k,t}}}\right)$ . Le moment d'ordre quatre de  $y_t$  est donné par :

$$\begin{aligned} E(y_t^4) &= E(E(y_t^4 | \Omega_{t-1})) \\ &= E\left(\int y_t^4 \sum_{k=1}^K \alpha_k g_k(y_t) dy_t\right) \\ &= \sum_{k=1}^K \alpha_k E\left(\int y_t^4 g_k(y_t) dy_t\right) \\ &= \sum_{k=1}^K \alpha_k E\left(\int \left(\frac{y_t}{\sqrt{h_{k,t}}}\right)^4 h_{k,t}^2 g_k(y_t) dy_t\right) \end{aligned} \quad (3.28)$$

Faisant le changement de variable :  $z_t = \frac{y_t}{\sqrt{h_{k,t}}}$   $z_t \sim \mathcal{N}(0, 1)$ .

Utilisant la fonction génératrice des moments pour le calcul de  $E(z_t^4)$ . Par conséquent :

$$\begin{aligned}
 E(y_t^4) &= \sum_{k=1}^K \alpha_k E(3h_{k,t}^2) \\
 &= 3 \sum_{k=1}^K \alpha_k E(\beta_{k0}^2 + 2\beta_{k0}\beta_{k1}y_{t-1}^2 + \beta_{k1}^2y_{t-1}^4) \\
 &= 3 \sum_{k=1}^K \alpha_k (\beta_{k0}^2 + 2\beta_{k0}\beta_{k1}E(y_{t-1}^2) + \beta_{k1}^2E(y_{t-1}^4)) \tag{3.29} \\
 &= 3 \left[ \sum_{k=1}^K \alpha_k \beta_{k0}^2 + 2 \sum_{k=1}^K \alpha_k \beta_{k0} \beta_{k1} \frac{\sum_{k=1}^K \alpha_k \beta_{k0}}{\left(1 - \sum_{k=1}^K \alpha_k \beta_{k1}\right)} + \sum_{k=1}^K \alpha_k \beta_{k1}^2 E(y_{t-1}^4) \right]
 \end{aligned}$$

en effet

$$\begin{aligned}
 E(y_{t-1}^2) &= E(E(y_{t-1}^2 | \Omega_{t-2})) \\
 &= E\left(\int y_{t-1}^2 \sum_{k=1}^K \alpha_k g_k(y_{t-1}) \partial y_{t-1}\right) \\
 &= E\left(\sum_{k=1}^K \alpha_k \int y_{t-1}^2 g_k(y_{t-1}) \partial y_{t-1}\right) \\
 &= E\left(\sum_{k=1}^K \alpha_k h_{k,t-1} \int \left(\frac{y_{t-1}}{\sqrt{h_{k,t-1}}}\right)^2 g_k(y_{t-1}) \partial y_{t-1}\right) \\
 &= E\left(\sum_{k=1}^K \alpha_k h_{k,t-1}\right) \\
 &= \sum_{k=1}^K \alpha_k E(\beta_{k0} + \beta_{k1}y_{t-2}^2) \\
 &= \sum_{k=1}^K \alpha_k \beta_{k0} + \sum_{k=1}^K \alpha_k \beta_{k1} E(y_{t-2}^2) \\
 &= \sum_{k=1}^K \alpha_k \beta_{k0} + \sum_{k=1}^K \alpha_k \beta_{k1} \left(\sum_{k=1}^K \alpha_k \beta_{k0} + \sum_{k=1}^K \alpha_k \beta_{k1} E(y_{t-3}^2)\right) \\
 &\vdots \\
 &= \sum_{k=1}^K \alpha_k \beta_{k0} \left[1 + \sum_{k=1}^K \alpha_k \beta_{k1} + \left(\sum_{k=1}^K \alpha_k \beta_{k1}\right)^2 + \cdots + \left(\sum_{k=1}^K \alpha_k \beta_{k1}\right)^{t-1}\right] \\
 &= \frac{\sum_{k=1}^K \alpha_k \beta_{k0}}{\left(1 - \sum_{k=1}^K \alpha_k \beta_{k1}\right)} \tag{3.30}
 \end{aligned}$$

La condition nécessaire et suffisante pour que cette équation aux différences non homogènes possède une solution stable, finie et indépendante du temps est la condition (3.28); d'où le moment d'ordre quatre du processus est donné par

$$3 \frac{\sum_{k=1}^K \alpha_k \beta_{k0}^2 + 2 \left( \sum_{k=1}^K \alpha_k \beta_{k0} \beta_{k1} \right) \left( \sum_{k=1}^K \alpha_k \beta_{k0} \right) + \left( 1 - \left( \sum_{k=1}^K \alpha_k \beta_{k1} \right) \right)}{1 - 3 \sum_{k=1}^K \alpha_k \beta_{k1}^2} \quad (3.31)$$

■

La condition d'existence du moment d'ordre quatre du modèle stationnaire  $MAR-ARCH(K; 0, \dots, 0; 1, \dots, 1)$  est similaire à celle du modèle ARCH (1) en remplaçant  $\alpha_1^2$  (voir chapitre 1) par  $\sum_{k=1}^K \alpha_k \beta_{k1}^2$ . Et notons aussi que La kurtosis de  $Y_t$  est généralement supérieur à trois. Les autocorrélations du carré d'un processus suivant un modèle  $MAR-ARCH$  stationnaire de moment d'ordre quatre fini doivent être similaires à celles du modèle ARCH. Par exemple, nous considérons le modèle stationnaire  $MAR-ARCH(K; 0, \dots, 0; 1, \dots, 1)$  avec  $\phi_{k0} = 0$  ( $k = 1, \dots, K$ ) et qui satisfait la condition donnée dans le théorème (20). L'espérance du produit  $y_t^2 y_{t-l}^2$  est donné par :

$$\begin{aligned} E(y_t^2 y_{t-l}^2) &= E(y_{t-l}^2 E(y_t^2 | \Omega_{t-1})) \\ &= E\left(y_{t-l}^2 \left( \sum_{k=1}^K \alpha_k \beta_{k0} + \sum_{k=1}^K \alpha_k \beta_{k1} y_{t-1}^2 \right)\right) \\ &= E(y_{t-l}^2) \sum_{k=1}^K \alpha_k \beta_{k0} + \sum_{k=1}^K \alpha_k \beta_{k1} E(y_{t-1}^2 y_{t-l}^2) \\ &= E(y_{t-l}^2) \sum_{k=1}^K \alpha_k \beta_{k0} + \sum_{k=1}^K \alpha_k \beta_{k1} E\left(y_{t-l}^2 \left( \sum_{k=1}^K \alpha_k \beta_{k0} + \sum_{k=1}^K \alpha_k \beta_{k1} y_{t-2}^2 \right)\right) \\ &= E\left[ y_{t-l}^2 \left( \sum_{k=1}^K \alpha_k \beta_{k0} + \sum_{k=1}^K \alpha_k \beta_{k1} \left( \sum_{k=1}^K \alpha_k \beta_{k0} + \sum_{k=1}^K \alpha_k \beta_{k1} y_{t-3}^2 \right) \right) \right] \\ &\quad \vdots \\ &= E\left[ y_{t-l}^2 \left( \sum_{k=1}^K \alpha_k \beta_{k0} \left( 1 + \sum_{k=1}^K \alpha_k \beta_{k1} + \left( \sum_{k=1}^K \alpha_k \beta_{k1} \right)^2 \right) + \left( \sum_{k=1}^K \alpha_k \beta_{k1} \right)^l y_{t-l}^2 \right) \right] \\ &= \frac{\sum_{k=1}^K \alpha_k \beta_{k0}}{1 - \sum_{k=1}^K \alpha_k \beta_{k1}} \left( 1 - \left( \sum_{k=1}^K \alpha_k \beta_{k1} \right)^l \right) E(y_{t-l}^2) + \left( \sum_{k=1}^K \alpha_k \beta_{k1} \right)^l E(y_{t-l}^4) \end{aligned} \quad (3.32)$$

En tenant compte du fait que :

$$E(y_{t-1}^2) = \frac{\sum_{k=1}^K \alpha_k \beta_{k0}}{1 - \sum_{k=1}^K \alpha_k \beta_{k1}} \quad (3.33)$$

et

$$E(y_t^2) = E(y_{t-1}^2) \quad (3.34)$$



il s'ensuit que

$$E(y_t^2 y_{t-l}^2) = (E(y_t^2))^2 + \left( \sum_{k=1}^K \alpha_k \beta_{k1} \right)^l \left( E(y_{t-l}^4) - (E(y_t^2))^2 \right) \quad (3.35)$$

d'autre part

$$\begin{aligned} \text{corr}(y_t^2, y_{t-l}^2) &= \frac{\text{cov}(y_t^2, y_{t-l}^2)}{(\text{var}(y_t^2) \text{var}(y_{t-l}^2))^{\frac{1}{2}}} \\ &= \frac{E(y_t^2 y_{t-l}^2) - (E(y_t^2))^2}{E(y_{t-l}^4) - (E(y_t^2))^2} \end{aligned} \quad (3.36)$$

$$\implies E(y_t^2 y_{t-l}^2) = (E(y_t^2))^2 + \text{corr}(y_t^2, y_{t-l}^2) \left( E(y_{t-l}^4) - (E(y_t^2))^2 \right)$$

d'où on a

$$\text{corr}(y_t^2, y_{t-l}^2) = \left( \sum_{k=1}^K \alpha_k \beta_{k1} \right)^l \quad (3.37)$$

Remarquons que cette fonction d'autocorrélation est similaire à celle du carré d'un processus suivant le modèle ARCH (1), en remplaçant le coefficient du premier retard par  $\sum_{k=1}^K \alpha_k \beta_{k1}$ .

### 3.3.4 Fonction prédictive au pas $m$

Le calcul de la fonction de prédiction au pas  $m$  d'un modèle  $MAR - ARCH$  sera traité comme dans le cas des modèles  $MAR$ .

La fonction de prédiction au premier pas  $F(y_{t+1} | \Omega_t)$  est facile à calculée en se basant sur le modèle (3.2). Le calcul de la fonction de prédiction au pas  $m$   $F(y_{t+m} | \Omega_t)$  n'est pas aussi direct. Nous utilisons l'approche de Monte Carlo pour estimer la fonction de distribution prédictive. L'approximation de Monte Carlo de la distribution prédictive au deuxième pas est donnée par :

$$F(y_{t+2} | \Omega_t) = \frac{1}{N} \sum_{j=1}^N F(y_{t+2} | \Omega_t, y_{t+1}^j) \quad (3.38)$$

où  $\{y_{t+1}^j\}$  sont des échantillons générés à partir de  $F(y_{t+1} | \Omega_t)$ .

## 3.4 conclusion

Dans ce chapitre nous avons discuté une nouvelle classe de modèles non linéaires, les modèles  $MAR-ARCH$ . Wong et Li ont généralisé les modèles  $MAR$  aux modèles  $MAR-ARCH$  pour avoir une meilleure modélisation des séries temporelles ayant les propriétés de non linéarité et non normalité, et pour une meilleure capture de l'hétéroscédasticité

par rapport les modèles classiques. Nous avons discuté les conditions de stationnarité au premier ordre et au deuxième ordre pour un processus suivant un modèle MAR-ARCH. Nous avons vu que les équations d'autocorrélations du processus sont similaires à celles de Yule-Walker pour un processus AR(p) ordinaire, et notons aussi que sous certaines conditions, les autocorrélations du carré d'un processus suivant un modèle MAR-ARCH doivent être similaires à celles obtenues par le carré d'un processus suivant un modèle ARCH. Nous avons vérifié cette proposition avec le modèle  $MAR-ARCH(K; 0, \dots, 0; 1, \dots, 1)$  avec  $\phi_{k0} = 0$  ( $k = 1, \dots, K$ ). Et notons aussi que la kurtosis de processus est généralement supérieur à trois.

CHAPITRE

# 4

---

## Sélection de modèle et estimation des paramètres dans les modèles MAR-ARCH

Dans ce chapitre, nous discutons de quelques problèmes associés à la méthode du maximum de la vraisemblance. Nous présentons ensuite une procédure très utilisée dans l'estimation des paramètres dans les modèles de mélange, l'algorithme EM. La description de l'algorithme EM, ses propriétés, son application dans l'estimation des paramètres dans les modèles MAR-ARCH et l'estimation de l'erreur standard théorique sont détaillées dans ce chapitre. Ensuite, une analyse par simulation est proposée, afin de prouver la performance de cette approche dans l'estimation des paramètres dans les modèles MAR-ARCH. Et en fin, nous terminons ce chapitre par l'application du BIC dans la sélection du modèle MAR-ARCH, suivie par une application numérique.

## 4.1 Introduction

Dans le chapitre précédent nous avons généralisé les modèles MAR aux modèles MAR-ARCH. Nous avons présenté plusieurs résultats importants concernant cette classe de modèles. Pour terminer notre étude, il reste à proposer une méthode pour estimer les paramètres dans cette classe de modèle et une approche pour sélectionner le modèle. Nous avons choisit la méthode la plus populaire pour estimer les paramètres dans la classe de modèles de mélanges où il figure le problème de présence de données manquantes, l'algorithme EM. Le critère BIC est aussi le critère le plus utilisé dans la sélection du modèle dans ce type de classe.

Nous commençons ce chapitre par un rappel de certaines difficultés liées à la technique du maximum de vraisemblance, en balayant plusieurs aspects statistiques et algorithmiques. La section deux est consacrées à l'algorithme EM comme technique renommée d'optimisation de la vraisemblance ainsi qu'à ses propriétés et certaines de ses variantes. La section trois propose d'appliquer l'algorithme EM dans l'estimation des paramètres dans la classe des modèles MAR-ARCH. La section quatre traite le problème de l'estimation de l'erreur standard. Nous proposons l'approche du Bootstrap pour résoudre ce problème. Dans la section cinq nous appliquons le critère BIC dans la sélection du modèle dans la classe des modèles MAR-ARCH. Dans la section six nous présentons une analyse par simulation pour l'algorithme EM et le critère BIC, afin de montrer leurs performances.

## 4.2 Problématique du maximum de la vraisemblance

### 4.2.1 Racines multiples et choix de la plus grande racine

Sous le nom générique de méthode du « maximum de la vraisemblance » se dissimulent un certain nombre de problèmes, propre ou non d'ailleurs au cas particulier des mélanges. En effet, dans de nombreuses applications où cette technique d'estimation est utilisée, les statisticiens sont souvent confrontés à l'existence des racines multiples de l'équation normale de la log-vraisemblance, c'est-à-dire aux différentes valeurs de  $\theta$  qui vérifient, sous quelques conditions de régularité, l'équation

$$\frac{\partial \log(\theta)}{\partial \theta} = 0 \quad (4.1)$$

Bien entendu, sous certaines conditions de régularité, la théorie affirme l'existence d'une unique racine convergente de cette équation (voir par exemple Cramer 1946 ou encore sa généralisation dans le cas multivarié par Tarone et Gruenhagen 1965), mais généralement peu d'indications sont données sur le choix de la racine convergente dans le cas de plusieurs racines. Le papier bibliographique de Small et al. (2000) discute d'approches variées pour sélectionner une des racines (voir aussi une discussion antérieure dans Lehmann 1983), incluant par exemple l'emploi d'une technique de bootstrap ou

encore l'examen des propriétés asymptotiques des racines lorsque leur formulation est explicite. Une autre possibilité consiste simplement à sélectionner la racine associée à la valeur maximum de la vraisemblance puisque Wald (1949) a établi la convergence du maximum de vraisemblance sous certaines conditions. Les propriétés de Wald concernant cet estimateur du maximum de vraisemblance ont été ensuite généralisées par White (1982) dans le cas plus réaliste d'un modèle éventuellement mal spécifié. Ainsi, la stratégie qui consiste à retenir la valeur maximum de la vraisemblance est souvent adoptée.

### 4.2.2 Problèmes associés à la méthode

Néanmoins, il existe des situations pour lesquelles ce maximum global est inconsistant et on pourra en trouver des exemples dans Neyman et Scott (1948), Ferguson (1982) ou bien encore Stefanski et Carroll (1987). Dans le cadre des mélanges gaussiens, essentiellement dans le cas hétéroscédastiques, le maximum global de la vraisemblance n'existe pas puisque la vraisemblance n'est pas bornée (Mclachlan et Peel, 2000, chap. 3). Par ailleurs, parmi les autres maxima locaux de la vraisemblance, certains d'entre eux peuvent correspondre à des «*spurious maximizers*» comme dénommés par Mclachlan et Peel, 2000, c'est-à-dire à des solutions où une ou plusieurs matrices de variance sont proches de la matrice nulle, solutions conduisant parfois à une assez grande vraisemblance mais ne représentant pourtant aucune réalité par rapport au paramètre recherché.

### 4.2.3 Difficultés techniques pour obtenir une racine adéquate

On pourra se reporter à Redner et Walker (1984) ou Mclachlan et Peel, (2000) pour une revue détaillée sur l'historique des méthodes visant à maximiser la vraisemblance dans les mélanges de distributions.

Dans ce cadre des mélanges, la résolution de l'équation (4.1) conduit à des équations non linéaires, donc généralement impossibles à résoudre sous forme explicite. L'évolution des moyens de calculs, grâce à l'informatique, a permis de surmonter graduellement cette difficulté. Dans un premier temps, certains cas simples de mélanges ont été résolus avec succès par des méthodes itératives. Par exemple, Rao (1948) utilisa la méthode du *scoring* de Fisher pour étudier un mélange de deux gaussiennes homoscédastiques univariées, Mendenhall et Hader (1958) utilisèrent une méthode de Newton pour un cas encore plus simple puisque ayant une unique inconnue. Par la suite, Day (1969) dans le cas d'un mélange multivarié de deux gaussiennes homoscédastiques, et Wolfe (1971) (ainsi que d'autres références antérieures de ce même auteur) avec un nombre quelconque de composantes hétéroscédastiques, ont employé à des périodes similaires des méthodes d'optimisation, l'algorithme EM de Dempster et al. (1977). L'algorithme EM, voire certaines de ses nombreuses variantes depuis les travaux de Dempster *et al.* (1977), est certainement aujourd'hui la méthode d'estimation de la vraisemblance la plus utilisée. Bien que de tels algorithmes permettent ainsi d'apporter des solutions relativement simples et satisfaisantes pour maximiser la vraisemblance, ils sont tout de même confrontés, en général, aux problèmes théoriques

inhérents à la méthode du maximum de vraisemblance : multiplicité des points stationnaires, dégénérescence de la vraisemblance et solutions «*spurious*» dans les mélanges gaussiens. S'agrègent parfois aux problèmes de la méthode du maximum de la vraisemblance des difficultés propres à la méthode d'optimisation retenue, comme une certaine lenteur de convergence comme cela peut être le cas pour EM. Ainsi, l'ensemble de ces questions ouvre la voie à un certain nombre de recherches dont certaines sont déjà débattues (Biernacki (2004)).

### 4.3 L'algorithme EM (Expectation-Maximization)

L'algorithme EM de Dempster *et al.* (1977) est une procédure générale pour maximiser la vraisemblance. Elle est adaptée à de nombreuses situations décrites sous forme de problèmes avec *données incomplètes*. L'algorithme EM formalise l'idée intuitive :

1. Remplacer les données manquantes par leurs valeurs estimées ;
2. Estimer les paramètres ;
3. Répéter :
  - L'étape (1) utilisant les paramètres estimés comme étant des vraies valeurs, et
  - L'étape (2) utilisant les valeurs estimées comme des valeurs observées, itérant jusqu'à la convergence.

Cette idée a été en usage pendant plusieurs années avant Orchard et Woodbury (1972) dans leurs "*missing information principle*" dont ils pourvurent la théorie fondamentale de l'idée sous-jacente. Le terme EM a été introduit dans Dempster, Laird, et Rubin (1977) où les résultats généraux sur le fonctionnement de l'algorithme ont été tout d'abord prouvées par un très grand nombre d'applications. Depuis cet article de référence, de nombreux auteurs ont décrit cet algorithme, ses propriétés et parfois ses variantes, par exemples McLachlan et Krishnan (1997) ou encore, dans le cas des mélanges de lois Redner et Walker (1984). Cet algorithme procède en deux étapes successives, l'étape E (pour Expectation) et l'étape M (pour Maximization). La description de cet algorithme, ses propriétés, et son application dans les modèles MAR-ARCH, sera détaillée dans ce chapitre.

#### 4.3.1 La théorie fondamentale de l'algorithme EM

Supposons que nous avons un vecteur aléatoire  $y$  de fonction densité conjointe  $f(y; \theta)$  tel que  $\theta$  est un paramètre de dimension  $p$ ,  $\theta \in \Theta \subseteq \mathbb{R}^p$ . Si le vecteur de donnée complète  $y$  a été observé il est intéressant de calculer l'estimateur de maximum de vraisemblance de  $\theta$  sur la distribution de  $y$ . La fonction log-vraisemblance de  $y$  qui doit être maximisée est :

$$\log L(y; \theta) = l(y; \theta) = \log f(y; \theta) \tag{4.2}$$

Cependant, en présence de données manquantes, uniquement une fonction du vecteur des données complètes est observée. Nous dénotons celle-ci par  $y = (y_{obs}, y_{mis})$  où  $y_{obs}$  dénote l'observée mais donnée « incomplète », et  $y_{mis}$  dénote la non observée ou la donnée « manquante ». Pour simplifier la description, nous supposons que la donnée manquante est manquante à un aléatoire (Rubin, 1977), par conséquent

$$\begin{aligned} f(y; \theta) &= f(y_{obs}, y_{mis}; \theta) \\ &= f_1(y_{obs}; \theta) f_2(y_{mis} | y_{obs}; \theta) \end{aligned} \quad (4.3)$$

où  $f_1$  est la densité conjointe de  $y_{obs}$  et  $f_2$  est la densité conjointe de  $y_{mis}$  conditionnée par la donnée observée  $y_{obs}$ , respectivement. Ceci implique

$$l_{obs}(y_{obs}; \theta) = l(y; \theta) - \log f_2(y_{mis} | y_{obs}; \theta) \quad (4.4)$$

où  $l_{obs}(y_{obs}; \theta)$  est la log de la vraisemblance de donnée observée.

L'algorithme EM est utile quand la maximisation de  $l_{obs}$  est difficile mais la maximisation du log de vraisemblance de donnée complète est simple. Mais, puisque  $y$  n'est pas observée,  $l$  ne peut pas être évaluée, par conséquent ne peut être maximisée. L'algorithme EM essaye de maximiser itérativement, en remplaçant  $l(y; \theta)$  par son espérance conditionnelle donnant  $y_{obs}$ . Explicitement, si  $\theta^{(0)}$  est la valeur initiale de  $\theta$ , il est nécessaire dans la première itération de calculer

$$Q(\theta; \theta^{(0)}) = E_{\theta^{(0)}}(l(y; \theta) | y_{obs}) \quad (4.5)$$

Cherchons  $\theta$  qui maximise  $Q(\theta; \theta^{(0)})$ , on la note  $\theta^{(1)}$ , qui vérifie

$$Q(\theta^{(1)}; \theta^{(0)}) \geq Q(\theta; \theta^{(0)}) \quad (4.6)$$

pour tout  $\theta \in \Theta$ , ainsi l'algorithme EM nécessite deux étapes, l'étape E (l'étape de l'estimation) suivie par l'étape M (l'étape de maximisation) définis comme suit :

**L'étape E** : Calculer  $Q(\theta; \theta^{(t)})$  où

$$Q(\theta; \theta^{(t)}) = E_{\theta^{(t)}}(l(y; \theta) | y_{obs}) \quad (4.7)$$

**L'étape M** : Trouver  $\theta^{(t+1)}$  dans  $\Theta$  tel que

$$Q(\theta^{(t+1)}; \theta^{(t)}) \geq Q(\theta; \theta^{(t)}) \quad (4.8)$$

pour tout  $\theta \in \Theta$ .

L'étape E et l'étape M sont répétées alternativement jusqu'à ce que la différence  $L(\theta^{(t+1)}) - L(\theta^{(t)})$  soit inférieur à  $\delta$  tel que  $\delta$  est choisi petit ; ou jusqu'à ce que la quantité  $\|\theta^{(k+1)} - \theta^{(k)}\|$  soit inférieur à  $\varepsilon$  tel que  $\varepsilon > 0$  et  $\|\cdot\|$  est une norme appropriée.

Le calcul de ses deux étapes sera simplifié si la log-vraisemblance est linéaire en statistique exhaustive de  $\theta$ . En particulier, ceci est possible, si la distribution du vecteur de donnée complète appartient à la famille des lois exponentielles. Dans ce cas l'étape E se réduit à calculer l'espérance de la statistique exhaustive de la donnée complète donnant la donnée observée. Lorsque le vecteur de donnée complète provient de la famille des lois exponentielles i,e :

$$f(y; \theta) = \frac{b(y) \exp \{c^T(\theta) t(y)\}}{a(\theta)} \quad (4.9)$$

où  $t(y)$  est le vecteur de statistiques exhaustives de donnée complète,  $c^T(\theta)$  est un vecteur fonction de vecteur des paramètres,  $a(\theta)$  et  $b(y)$  sont des fonctions scalaires. Nous pouvons écrire l'étape E de la manière suivante :

$$Q(\theta; \theta^{(k)}) = E_{\theta^k}(\log b(y) | y_{obs}) + c^T(\theta) t^{(k)} - \log a(\theta) \quad (4.10)$$

où  $t^{(k)} = E_{\theta^k}(t(y) | y_{obs})$  est un estimateur pour la statistique exhaustive. L'étape M maximise la fonction  $Q$  suivant  $\theta$ ; mais  $E_{\theta^k}(\log b(y) | y_{obs})$  ne dépend pas de  $\theta$ , alors il est suffisant d'écrire :

**L'étape E :** Calculer :

$$t^{(k)} = E_{\theta^k}(t(y) | y_{obs}) \quad (4.11)$$

**L'étape M :** Calculer :

$$\theta^{(k+1)} = \arg \max_{\theta} [c^T(\theta) t^{(k)} - \log a(\theta)] \quad (4.12)$$

Un exemple explicatif est donné pour illustrer les deux étapes dans le cas de la famille des lois exponentielles.

Soient le vecteur de donnée complète  $y = (y_1, \dots, y_n)'$  un échantillon suivant  $\mathcal{N}(\mu, \sigma^2)$ . Alors

$$\begin{aligned} f(y, \mu, \sigma^2) &= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mu)^2}{\sigma^2} \right\} \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp \left\{ \frac{-1}{2\sigma^2} \left( \sum y_i^2 - 2\mu \sum y_i - n\mu^2 \right) \right\} \end{aligned} \quad (4.13)$$

Ceci implique que  $(\sum y_i, \sum y_i^2)$  sont des statistiques exhaustives pour  $\theta = (\mu, \sigma^2)'$ . La fonction de log-vraisemblance de donnée complète est :



$$\begin{aligned}
 l(\mu, \sigma^2; y) &= -\frac{n}{2} \log(\sigma^2) - \frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mu)^2}{\sigma^2} + c \\
 &= -\frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n y_i^2 + \frac{\mu}{\sigma^2} \sum_{i=1}^n y_i - \frac{n\mu^2}{\sigma^2} + c \quad (4.14)
 \end{aligned}$$

Il s'ensuit que la log-vraisemblance basée sur la donnée complète est linéaire en statistiques exhaustives. Supposons que  $y_i, i = 1, \dots, m$  sont observées et  $y_i, i = m+1, \dots, n$  sont manquantes (aléatoirement) tel que les  $y_i$  sont supposés i.i.d. On dénote le vecteur de donnée observée par  $y_{obs} = (y_1, \dots, y_m)'$ . Comme la donnée complète provienne d'une famille de lois exponentielles, l'étape E nécessite le calcul de  $E_\theta(\sum_{i=1}^n y_i | y_{obs})$  et  $E_\theta(\sum_{i=1}^n y_i^2 | y_{obs})$ .

Au lieu de calculer  $E_{\theta^{(t)}}(l(\theta; y) | y_{obs})$  dans la  $t$ ème itération de l'étape E, nous calculons

$$\begin{aligned}
 S_1^{(t)} &= E_{\mu^{(t)}\sigma^{2(t)}}\left(\sum_{i=1}^n y_i | y_{obs}\right) \\
 &= \sum_{i=1}^m y_i + (n-m)\mu^{(t)} \quad (4.15)
 \end{aligned}$$

puisque  $E_{\mu^{(t)}\sigma^{2(t)}}(y_i) = \mu^{(t)}$  ou  $\mu^{(t)}$  et  $\sigma^{2(t)}$  sont les estimations courantes de  $\mu$  et de  $\sigma^2$ , et

$$\begin{aligned}
 S_2^{(t)} &= E_{\mu^{(t)}\sigma^{2(t)}}\left(\sum_{i=1}^n y_i^2 | y_{obs}\right) \\
 &= \sum_{i=1}^m y_i^2 + (n-m)\left(\mu^{(t)^2} + \sigma^{(t)^2}\right) \quad (4.16)
 \end{aligned}$$

Puisque  $E_{\mu^{(t)}\sigma^{2(t)}}(y_i^2) = \mu^{(t)^2} + \sigma^{(t)^2}$ .

Pour l'étape M, notons que les estimateurs du maximum de vraisemblance de  $\mu$  et de  $\sigma^2$  sont :

$$\hat{\mu} = \frac{\sum_{i=1}^n y_i}{n} \quad (4.17)$$

et

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n y_i^2}{n} - \left(\frac{\sum_{i=1}^n y_i}{n}\right)^2 \quad (4.18)$$

Notons aussi que les statistiques exhaustives de donnée complète elles même ne peuvent être calculées directement puisque  $y_i, i = m+1, \dots, n$  ne sont pas observées. Nous acceptons les expressions

$$\mu^{(t+1)} = \frac{S_1^{(t)}}{n} \quad (4.19)$$

et

$$\sigma^{2(t+1)} = \frac{S_2^{(t)}}{n} + \mu^{(t+1)^2} \quad (4.20)$$

Ainsi, l'étape E nécessite pour le calcul des expressions (4.15) et (4.16) de commencer avec des valeurs initiales  $\mu^{(0)}$  et  $\sigma^{2(0)}$ . L'étape M nécessite de remplacer ces valeurs dans (4.19) et (4.20) pour calculer la nouvelle valeur  $\mu^{(1)}$  et  $\sigma^{2(1)}$ , etc. Ainsi l'algorithme EM réitère successivement entre (4.15), (4.16), (4.19) et (4.20). Bien sur dans cet exemple, il n'est pas nécessaire d'utiliser l'algorithme EM puisque les estimateurs du maximum de vraisemblance pour  $\mu$  et  $\sigma^2$  sont clairement donnés par  $\hat{\mu} = \frac{\sum_{i=1}^n y_i}{n}$  et  $\hat{\sigma}^2 = \frac{\sum_{i=1}^n y_i^2}{n} - \left( \frac{\sum_{i=1}^n y_i}{n} \right)^2$ .

### 4.3.2 La convergence de l'algorithme EM

L'algorithme EM essaie de maximiser  $l_{obs}(y_{obs}|\theta)$  par la maximisation de  $l(\theta; y)$ . Chaque itération de l'algorithme EM a deux étapes : l'étape E et l'étape M. la *tème* étape E calcule  $Q(\theta; \theta^t) = E_{\theta^t}(l(\theta; y) | y_{obs})$  comme une fonction de  $\theta$  en donnant  $y_{obs}$  et  $\theta^t$ . La *tème* étape M trouve  $\theta^{(t+1)}$  qui maximise  $Q(\theta; \theta^t)$ , i.e ; trouver  $\theta^{(t+1)}$  tel que

$$Q(\theta^{(t+1)}; \theta^t) \geq Q(\theta; \theta^t) \quad (4.21)$$

Pour tous  $\theta \in \Theta$ . Pour vérifier que cette itération produit une suite d'itérations convergente vers un maximum de  $l_{obs}(y_{obs}|\theta)$ , écrivons tout d'abord :

$$l_{obs}(\theta; y_{obs}) = l(y; \theta) - \log f_2(y_{mis} | \theta; y_{obs}) \quad (4.22)$$

A partir de la distribution de  $y$  donnant  $y_{obs}$  à  $\theta^t$  (l'estimation courante),  $l(\theta; y_{obs})$  peut être exprimée sous la forme

$$\begin{aligned} l_{obs}(\theta; y_{obs}) &= \int l(\theta; y) f(y | \theta^t; y_{obs}) dy - \int \log f_2(y_{mis} | \theta; y_{obs}) f(y | \theta^t; y_{obs}) dy \\ &= E_{\theta^t}(l(\theta; y) | y_{obs}) - E_{\theta^t}(\log f_2(y_{mis} | \theta; y_{obs}) | y_{obs}) \\ &= Q(\theta; \theta^t) - H(\theta; \theta^t) \end{aligned} \quad (4.23)$$

où

$$Q(\theta; \theta^t) = E_{\theta^t}(l(\theta; y) | y_{obs}) \quad (4.24)$$

et

$$H(\theta; \theta^t) = E_{\theta^t}(\log f_2(y_{mis} | \theta; y_{obs}) | y_{obs}) \quad (4.25)$$

Pour prouver que la suite  $\theta^{(t)}$  évaluée par l'algorithme EM converge vers le maximum local de  $l_{obs}(\theta; y_{obs})$  nous utilisons le lemme suivant :

**Lemme**

Pour tous  $\theta \in \Theta$

$$H(\theta; \theta^{(t)}) \leq H(\theta^{(t)}; \theta^{(t)}) \quad (4.26)$$

**Théorème 21** *L'algorithme EM fait augmenter  $l_{obs}(\theta; y_{obs})$  dans chaque itération  $i, e$ ;*

$$l_{obs}(\theta^{(t+1)}; y_{obs}) \geq l_{obs}(\theta^{(t)}; y_{obs}) \quad (4.27)$$

*Nous aurons l'égalité si et seulement si :*

$$Q(\theta^{(t+1)}; \theta^{(t)}) = Q(\theta^{(t)}; \theta^{(t)}) \quad (4.28)$$

Ce théorème montre que la croissance de  $Q(\theta; \theta^{(t)})$  dans chaque itération implique la maximisation ou au moins la croissance sans cesse de  $l_{obs}(\theta; y_{obs})$ .

### 4.3.3 L'algorithme EM généralisé (GEM)

Rappelons que dans l'étape M, nous maximisons  $Q(\theta; \theta^{(t)})$  i e; trouvons  $\theta^{(t+1)}$  tel que :

$$Q(\theta^{(t+1)}; \theta^{(t)}) \geq Q(\theta; \theta^{(t)}) \quad (4.29)$$

Pour tout  $\theta \in \Theta$ . Dans la version généralisée de l'algorithme EM nous allons exiger que  $\theta^{(t+1)}$  vérifie

$$Q(\theta^{(t+1)}; \theta^{(t)}) \geq Q(\theta^{(t)}; \theta^{(t)}) \quad (4.30)$$

$\theta^{(t+1)}$  est choisit pour faire augmenter  $Q(\theta; \theta^{(t)})$  par rapport à sa valeur au point  $\theta^{(t)}$  dans chaque itération. Ceci est suffisant pour assurer que

$$l(\theta^{(t+1)}; y) \geq l(\theta^{(t)}; y) \quad (4.31)$$

Dans chaque itération; la suite d'itération (GEM) converge aussi vers un maximum local.

### 4.3.4 L'algorithme GEM basé sur la seule étape de Newton - Raphson

Nous utilisons l'algorithme de type GEM quand le maximum global n'est pas facile à trouver, par conséquent nous pouvons pas effectuer un calcul dans l'étape M. Dans ce cas, il est possible qu'une méthode itérative soit nécessaire pour accomplir l'étape M. Puisque il n'est pas essentiel de maximiser  $Q$  dans un GEM, mais uniquement augmenter la vraisemblance, nous pouvons remplacer l'étape M par une autre étape qui réalise le même but. Nous proposons la seule itération de l'algorithme de Newton-Raphson.

Soit

$$\theta^{(t+1)} = \theta^{(t)} + a^{(t)}\delta^{(t)} \quad (4.32)$$

où

$$\delta^{(t)} = - \left[ \frac{\partial^2 Q(\theta; \theta^{(t)})}{\partial \theta \partial \theta'} \right]_{\theta=\theta^{(t)}}^{-1} \left[ \frac{\partial Q(\theta; \theta^{(t)})}{\partial \theta} \right]_{\theta=\theta^{(t)}} \quad (4.33)$$

où,  $\delta^{(t)}$  est la direction de Newton-Raphson au point  $\theta^{(t)}$  et  $0 < a^{(t)} \leq 1$ . Si  $a^{(t)} = 1$  cela va définir l'étape exacte de Newton-Raphson.

### 4.3.5 Propriétés et inconvénients de l'algorithme EM

Nous proposons maintenant d'énoncer certaines propriétés importantes de l'algorithme EM.

- L'algorithme EM est souvent reconnu comme un algorithme convergeant assez lentement. Le taux de convergence est linéaire au voisinage d'un point stationnaire  $\theta^*$  de la vraisemblance (voir McLachlan et Krishnan (1997) chap. 3), contrairement à des méthodes de type Newton qui bénéficient d'une convergence localement quadratique. Comme le taux de convergence est généralement donné par :

$$\gamma = \lim_{t \rightarrow \infty} \frac{\|\theta^{t+1} - \theta^*\|}{\|\theta^t - \theta^*\|} \quad (4.34)$$

pour n'importe quelle norme  $\|\cdot\|$  de  $\mathbb{R}^n$ . La vitesse de convergence de EM sera donc dépendante de la valeur de  $\gamma$ , une grande valeur  $\gamma$  imposant une convergence lente

- Les étapes E et M peuvent être analytiquement intraitables ;
- La vraisemblance augmente de manière monotone à chaque itération de l'algorithme EM, c'est-à-dire  $L(\theta^{t+1}; y) \geq L(\theta^t; y)$  pour  $t \geq 0$ . Pour montrer ce point (voir par exemple McLachlan et Krishnan (1997) chap. 3).
- L'algorithme EM est largement apprécié par sa simplicité d'implémentation, ses itérations sont généralement peu gourmandes en temps de calcul, le peu de mémoire nécessaire pour le faire fonctionner (il nécessite peu de stockage) et enfin son principe assez naturel heuristiquement.
- Avec l'algorithme EM nous ne pouvons pas avoir automatiquement une estimation de la matrice de covariance des estimations des paramètres. Néanmoins, cet inconvénient peut être enlevé en utilisant des méthodes appropriées associées à l'algorithme EM.

## 4.4 L'application de l'algorithme EM dans l'estimation des paramètres dans le modèle MAR-ARCH

Vue que les modèles de mélanges sont typiquement des modèles ayant une structure de donnée incomplète, nous proposons l'algorithme EM pour estimer les paramètres

dans les modèles MAR-ARCH.

L'estimation des paramètres dans le modèle MAR-ARCH est similaire à celle du modèle MAR mais avec plus de complications.

Supposons que le vecteur d'observations  $Y = (y_1, \dots, y_n)$  est généré à partir d'un modèle *MAR-ARCH*. Soit  $Z = (z_1, \dots, z_n)$  où  $z_t$  ( $t = 1 : n$ ) est un vecteur aléatoire non observable de dimension  $K$ , tel que la  $k^{\text{ème}}$  composante  $z_{t,k} = 1$  si  $y_t$  provient de la fonction de distribution de la  $k^{\text{ème}}$  composante et  $z_{t,k} = 0$  autrement. Notons  $\alpha_k$  la probabilité pour que  $z_{t,k} = 1$ . Soient  $\alpha = (\alpha_1, \dots, \alpha_{K-1})$  avec  $\sum_{k=1}^K \alpha_k = 1$ ,  $\theta_k = (\phi_{k0}, \dots, \phi_{kp_k})$  ( $k = 1, \dots, K$ ),  $\beta_k = (\beta_{k0}, \dots, \beta_{kq_k})$  ( $k = 1, \dots, K$ ), et  $\theta = (\alpha', \theta'_1, \beta'_1, \dots, \theta'_K, \beta'_K)$ .

$$\begin{aligned} L(Y, Z, | \theta) &= \prod_{t=p+q+1}^n f(y_t, z_t | \theta) \\ &= \prod_{t=p+q+1}^n \prod_{k=1}^K f(y_t, z_{t,k} | \theta'_k, \beta_k, \alpha_k) \\ &= \prod_{t=p+q+1}^n \prod_{k=1}^K \alpha_k^{z_{tk}} \left[ \varphi(y_t | \theta'_k, \beta_k, \alpha_k) \right]^{z_{tk}} \end{aligned} \quad (4.35)$$

D'où la fonction de log-vraisemblance normalisée est donnée par :

$$l = \frac{1}{N} \sum_{t=p+q+1}^n \left( \sum_{k=1}^K z_{k,t} \log \alpha_k - \sum_{k=1}^K \frac{z_{k,t}}{2} \log h_{k,t} - \sum_{k=1}^K \frac{z_{k,t} e_{k,t}^2}{2h_{k,t}} \right) \quad (4.36)$$

où  $N = n - p - q$ .

L'algorithme EM nécessite deux étapes, l'étape E et l'étape M.

### L'étape E :

Puisque la fonction de vraisemblance est de la famille des lois exponentielles, donc il suffit de calculer l'espérance conditionnelle de la statistique exhaustive  $\sum_t z_{k,t}$  conditionnée par les observations observées et les paramètres. Dans ce cas l'espérance conditionnelle de la  $k^{\text{ème}}$  composante est exactement la probabilité conditionnelle que l'observation provient de la  $k^{\text{ème}}$  composante du mélange de distributions, conditionnée par  $\theta$  et  $Y$ . Soit  $\tau_{k,t}$  est l'espérance conditionnelle de  $z_{k,t}$ . D'où les équations de l'étape E sont :

$$\tau_{k,t} = \frac{\alpha_k}{\sqrt{h_{k,t}}} \phi\left(\frac{e_{k,t}}{\sqrt{h_{k,t}}}\right) \left( \sum_{l=1}^K \frac{\alpha_l}{\sqrt{h_{l,t}}} \phi\left(\frac{e_{l,t}}{\sqrt{h_{l,t}}}\right) \right)^{-1} \quad (4.37)$$

Dans la pratique,  $\theta$  est la valeur obtenue de l'étape M précédente dans la procédure EM.

### L'étape M :

Dans cette étape nous devons trouver les expressions des estimateurs du maximum de vraisemblance pour chaque paramètre. Nous commençons tout d'abord par le calcul des dérivées partielles.

$$\begin{aligned} \frac{\partial l}{\partial \alpha_k} &= \frac{\partial}{\partial \alpha_k} \left[ \frac{1}{N} \sum_{t=p+q+1}^n \sum_{k=1}^K z_{k,t} \log \alpha_k \right] \\ &= \frac{\partial}{\partial \alpha_k} \left[ \frac{1}{N} \sum_{t=p+q+1}^n \sum_{k=1}^{K-1} z_{k,t} \log \alpha_k - z_{k,t} \log \left( 1 - \sum_{k=1}^{K-1} \alpha_k \right) \right] \end{aligned} \quad (4.38)$$

pour  $(k = 1, \dots, K - 1)$ .

$$\frac{\partial l}{\partial \alpha_k} = \frac{1}{N} \sum_t \left( \frac{z_{k,t}}{\alpha_k} - \frac{z_{K,t}}{\alpha_K} \right) \quad (4.39)$$

pour  $(k = 1, \dots, K), i = 0, \dots, p_k$

$$\frac{\partial l}{\partial \phi_{ki}} = \frac{1}{N} \left\{ \sum_t \frac{z_{k,t}}{2h_{k,t}} \frac{\partial h_{k,t}}{\partial \phi_{ki}} \left( \frac{e_{k,t}^2}{h_{k,t}} - 1 \right) - \sum_t \frac{z_{k,t}}{h_{k,t}} \frac{e_{k,t}}{\partial \phi_{ki}} \frac{\partial e_{k,t}}{\partial \phi_{ki}} \right\} \quad (4.40)$$

pour  $(k = 1, \dots, K), i = 0, \dots, q_k$

$$\frac{\partial l}{\partial \beta_{ki}} = \frac{1}{N} \sum_t \frac{z_{k,t}}{2h_{k,t}} \frac{\partial h_{k,t}}{\partial \beta_{ki}} \left( \frac{e_{k,t}^2}{h_{k,t}} - 1 \right) \quad (4.41)$$

ici

$$\begin{aligned} \frac{\partial e_{k,t}}{\partial \phi_{ki}} &= -1 \text{ pour } i = 0, \\ \frac{\partial e_{k,t}}{\partial \phi_{ki}} &= -y_{t-i} \text{ pour } 0 < i \leq p_k, \\ \frac{\partial h_{k,t}}{\partial \phi_{ki}} &= 2 \sum_{j=1}^{q_k} \partial \beta_{kj} e_{k,t-j} \left( \frac{\partial e_{k,t-j}}{\partial \phi_{ki}} \right) \text{ pour } i = 0, \\ \frac{\partial h_{k,t}}{\partial \beta_{ki}} &= 1 \text{ pour } i = 0, \\ \frac{\partial h_{k,t}}{\partial \beta_{ki}} &= e_{k,t-i}^2 \text{ pour } 0 < i \leq q_k. \end{aligned}$$

pour  $k = 1, \dots, K$ . Dans la pratique  $\theta$  est la valeur obtenue dans l'étape M précédente de la procédure EM.

$$\frac{\partial l}{\partial \alpha_k} = 0 \implies \hat{\alpha}_k = \frac{\sum_{t=p+q+1}^n z_{k,t}}{n - p - q} \quad (4.42)$$

Remplaçant la statistique exhaustive par son espérance. d'où pour  $k = 1, \dots, K$

$$\hat{\alpha}_k = \frac{\sum_{t=p+q+1}^n \tau_{k,t}}{n-p-q} \quad (4.43)$$

Concernant les paramètres  $\theta_k, \beta_k$ , ils n'existent pas des solutions explicites pour les équations dérivées de vraisemblance. Des méthodes numériques sont utilisées pour trouver les estimations des paramètres. Ici, nous utilisons la méthode de Newton-Raphson. Notons ici que nous avons pour  $k, h = 1, \dots, K$  et  $k \neq h$

$$\begin{aligned} \frac{\partial^2 l}{(\partial \theta_k, \partial \theta_h)} &= \frac{\partial^2 l}{(\partial \beta_k, \partial \beta_h)} \\ &= \frac{\partial^2 l}{(\partial \theta_k, \partial \beta_h)} \\ &= 0 \end{aligned} \quad (4.44)$$

Nous pouvons considérer l'estimation des paramètres  $\theta_k$  et  $\beta_k$  dans chaque composante séparément. Nous pouvons approximer les dérivées de seconde ordre de log-vraisemblance par rapport aux paramètres dans chaque composante par les quantités suivantes :

pour  $i, j = 0, \dots, p_k$

$$-\frac{\partial^2 l}{\partial \phi_{ki} \partial \phi_{kj}} \approx \frac{1}{N} \sum_t z_{k,t} \left( \frac{1}{2h_{k,t}^2} \frac{\partial h_{k,t}}{\partial \phi_{k,i}} \frac{\partial h_{k,t}}{\partial \phi_{k,j}} + \frac{1}{h_{k,t}} \frac{\partial e_{k,t}}{\partial \phi_{k,i}} \frac{\partial e_{k,t}}{\partial \phi_{k,j}} \right) \quad (4.45)$$

et pour  $i, j = 0, \dots, q_k$

$$-\frac{\partial^2 l}{\partial \beta_{ki} \partial \beta_{kj}} \approx \frac{1}{N} \sum_t z_{k,t} \left( \frac{1}{2h_{k,t}^2} \frac{\partial h_{k,t}}{\partial \beta_{k,i}} \frac{\partial h_{k,t}}{\partial \beta_{k,j}} \right) \quad (4.46)$$

$$-\frac{\partial^2 l}{\partial \phi_{ki} \partial \beta_{kj}} \approx 0 \quad (4.47)$$

Nous donnons la justification des ces approximations :

pour tous  $i, j = 0, \dots, p_k$

$$\begin{aligned} \frac{\partial^2 l_t}{\partial \phi_{ki} \partial \phi_{kj}} &= -\frac{z_{k,t}}{2h_{k,t}^2} \frac{\partial h_{k,t}}{\partial \phi_{k,i}} \frac{\partial h_{k,t}}{\partial \phi_{k,j}} \left( \frac{e_{k,t}^2}{h_{k,t}} - 1 \right) \\ &\quad + \frac{z_{k,t}}{h_{k,t}} \frac{\partial^2 h_{k,t}}{\partial \phi_{k,i} \partial \phi_{k,j}} \left( \frac{e_{k,t}^2}{h_{k,t}} - 1 \right) \\ &\quad + \frac{z_{k,t} e_{k,t}}{2h_{k,t}^2} \frac{\partial h_{k,t}}{\partial \phi_{k,i}} \frac{\partial e_{k,t}}{\partial \phi_{k,j}} - \frac{z_{k,t} e_{k,t}^2}{2h_{k,t}^3} \frac{\partial h_{k,t}}{\partial \phi_{k,i}} \frac{\partial h_{k,t}}{\partial \phi_{k,j}} \\ &\quad - \frac{z_{k,t}}{h_{k,t}} \frac{\partial e_{k,t}}{\partial \phi_{k,i}} \frac{\partial e_{k,t}}{\partial \phi_{k,j}} + \frac{z_{k,t} e_{k,t}}{h_{k,t}^2} \frac{\partial e_{k,t}}{\partial \phi_{k,i}} \frac{\partial h_{k,t}}{\partial \phi_{k,j}} \\ &\quad - \frac{z_{k,t} e_{k,t}}{2h_{k,t}} \frac{\partial^2 e_{k,t}}{\partial \phi_{k,i} \partial \phi_{k,j}} \end{aligned} \quad (4.48)$$

pour tous  $i, j = 0, \dots, q_k$

$$\begin{aligned} \frac{\partial^2 l_t}{\partial \beta_{ki} \partial \beta_{kj}} &= -\frac{z_{k,t}}{2h_{k,t}^2} \frac{\partial h_{k,t}}{\partial \beta_{k,i}} \frac{\partial h_{k,t}}{\partial \beta_{k,j}} \left( \frac{e_{k,t}^2}{h_{k,t}} - 1 \right) \\ &\quad + \frac{z_{k,t}}{2h_{k,t}} \frac{\partial^2 h_{k,t}}{\partial \beta_{k,i} \partial \beta_{k,j}} \left( \frac{e_{k,t}^2}{h_{k,t}} - 1 \right) \\ &\quad - \frac{z_{k,t} e_{k,t}^2}{2h_{k,t}^3} \frac{\partial h_{k,t}}{\partial \beta_{k,i}} \frac{\partial h_{k,t}}{\partial \beta_{k,j}} \end{aligned} \quad (4.49)$$

pour tous  $i = 0, \dots, p_k$  et  $j = 0, \dots, q_k$

$$\begin{aligned} \frac{\partial^2 l_t}{\partial \phi_{ki} \partial \beta_{kj}} &= -\frac{z_{k,t}}{2h_{k,t}^2} \frac{\partial h_{k,t}}{\partial \phi_{k,i}} \frac{\partial h_{k,t}}{\partial \beta_{k,j}} \left( \frac{e_{k,t}^2}{h_{k,t}} - 1 \right) \\ &\quad + \frac{z_{k,t}}{2h_{k,t}} \frac{\partial^2 h_{k,t}}{\partial \phi_{k,i} \partial \beta_{k,j}} \left( \frac{e_{k,t}^2}{h_{k,t}} - 1 \right) \\ &\quad + \frac{z_{k,t} e_{k,t}}{h_{k,t}^2} \frac{\partial h_{k,t}}{\partial \beta_{k,j}} \frac{\partial e_{k,t}}{\partial \phi_{k,i}} - \frac{z_{k,t} e_{k,t}^2}{2h_{k,t}^3} \frac{\partial h_{k,t}}{\partial \phi_{k,i}} \frac{\partial h_{k,t}}{\partial \beta_{k,j}} \end{aligned} \quad (4.50)$$

Dans l'étape M de l'algorithme EM, la réalisation de la variable aléatoire  $z_{k,t}$  est considérée connue. Si  $z_{k,t}$  égale à zéro, toutes les dérivées au second ordre du log-vraisemblance sont nulles. D'autre part, si  $z_{k,t}$  égale à un, alors la densité conditionnelle de  $y_t$  est donnée par

$$f(y_t \mid \Omega_{t-1}, z_{t,k} = 1) = \left( \frac{1}{\sqrt{h_{k,t}}} \right) \varphi \left( \frac{e_{k,t}}{\sqrt{h_{k,t}}} \right) \quad (4.51)$$

D'ici les espérances conditionnelles des dérivées du second ordre sont :

$$E \left( -\frac{\partial^2 l_t}{\partial \phi_{ki} \partial \phi_{kj}} \mid \Omega_{t-1}, z_{t,k} \right) = \frac{z_{k,t}}{2h_{k,t}^2} \frac{\partial h_{k,t}}{\partial \phi_{k,i}} \frac{\partial h_{k,t}}{\partial \phi_{k,j}} + \frac{z_{k,t}}{h_{k,t}} \frac{\partial e_{k,t}}{\partial \phi_{k,i}} \frac{\partial e_{k,t}}{\partial \phi_{k,j}} \quad (4.52)$$

$$E \left( -\frac{\partial^2 l_t}{\partial \beta_{ki} \partial \beta_{kj}} \mid \Omega_{t-1}, z_{t,k} \right) = \frac{z_{k,t}}{2h_{k,t}^2} \frac{\partial h_{k,t}}{\partial \beta_{k,i}} \frac{\partial h_{k,t}}{\partial \beta_{k,j}} \quad (4.53)$$

$$E \left( -\frac{\partial^2 l_t}{\partial \phi_{ki} \partial \beta_{kj}} \mid \Omega_{t-1}, z_{t,k} \right) = \frac{z_{k,t}}{2h_{k,t}^2} \frac{\partial h_{k,t}}{\partial \phi_{k,i}} \frac{\partial h_{k,t}}{\partial \beta_{k,j}} \quad (4.54)$$

et les espérances des dérivées de  $l_t$  donnant  $z_{k,t}$  sont :

$$E \left( -\frac{\partial^2 l_t}{\partial \phi_{ki} \partial \phi_{kj}} \mid z_{t,k} \right) = \frac{z_{k,t}}{2} E \left( \frac{1}{h_{k,t}^2} \frac{\partial h_{k,t}}{\partial \phi_{k,i}} \frac{\partial h_{k,t}}{\partial \phi_{k,j}} \right) + z_{k,t} E \left( \frac{1}{h_{k,t}} \frac{\partial e_{k,t}}{\partial \phi_{k,i}} \frac{\partial e_{k,t}}{\partial \phi_{k,j}} \right) \quad (4.55)$$



$$E \left( -\frac{\partial^2 l_t}{\partial \beta_{ki} \partial \beta_{kj}} \mid z_{t,k} \right) = \frac{z_{k,t}}{2} E \left( \frac{1}{h_{k,t}^2} \frac{\partial h_{k,t}}{\partial \beta_{k,i}} \frac{\partial h_{k,t}}{\partial \beta_{k,j}} \right) \quad (4.56)$$

$$E \left( -\frac{\partial^2 l_t}{\partial \phi_{ki} \partial \beta_{kj}} \mid \Omega_{t-1}, z_{t,k} \right) = \frac{z_{k,t}}{2} E \left( \frac{1}{h_{k,t}^2} \frac{\partial h_{k,t}}{\partial \phi_{k,i}} \frac{\partial h_{k,t}}{\partial \beta_{k,j}} \right) \quad (4.57)$$

et ceci justifie les approximations utilisées.

Les résultats précédentes nous permettent d'écrire les valeurs de  $\theta_k$  et de  $\beta_k$  dans les sous suites d'itérations suivantes :

$$\theta_k^{(i+1)} = \theta_k^{(i)} + \left( -\frac{\partial^2 l}{\partial \theta_k \partial \theta_k} \right)_{\theta_k^{(i)}, \beta_k^{(i)}}^{-1} \left( \frac{\partial l}{\partial \theta_k} \right)_{\theta_k^{(i)}, \beta_k^{(i)}} \quad (4.58)$$

et

$$\beta_k^{(i+1)} = \beta_k^{(i)} + \left( -\frac{\partial^2 l}{\partial \beta_k \partial \beta_k} \right)_{\theta_k^{(i+1)}, \beta_k^{(i)}}^{-1} \left( \frac{\partial l}{\partial \beta_k} \right)_{\theta_k^{(i+1)}, \beta_k^{(i)}} \quad (4.59)$$

où  $\theta_k^{(i)}$  et  $\beta_k^{(i)}$  sont les valeurs obtenues dans la  $i^{\text{ème}}$  itération. Les estimations des paramètres  $\hat{\theta}_k$  et  $\hat{\beta}_k$  dans l'itération particulière de l'étape M sont obtenues par la réitération de (4.58) et (4.59) jusqu' à la convergence. Dans la pratique, les  $z_{k,t}$  sont obtenues à partir de  $\tau_{k,t}$  calculées dans l'étape E précédente de la procédure EM.

## 4.5 L'estimation de l'erreur standard

Plusieurs méthodes ont été suggérées dans la littérature de l'algorithme EM, afin de le renforcer pour pouvoir estimer la matrice de covariance de l'EMV calculé. La plupart de ces méthodes essayent d'exploiter le calcul dans les étapes de l'algorithme EM. Ces méthodes sont basées sur la matrice de l'information observée  $I(\hat{\theta}; y_{obs})$  ou estimée, (la méthode de Louis (1982), Meilijson (1989), Meng et Rubin (1991), McLachlan et Krishnan (1997)) ou basées sur le rééchantillonnage (l'approche de Bootstrap).

Basford et al (1997) ont appliqué la méthode utilisant Bootstrap et les approches basées sur l'information pour quelques modèles de mélanges de distributions gaussiennes. Ils ont déduit que la stabilité de l'approche basée sur l'information nécessite une taille de l'échantillon très élevée. Bootstrap est une technique puissante et elle nécessite dans l'estimation uniquement la donnée disponible. Avant de donner les étapes d'utilisation du *Bootstrap* pour l'estimation de l'erreur standard des estimations des paramètres obtenues par l'algorithme *EM*, quelques définitions sont essentielles

### 4.5.1 L'utilisation du *bootstrap* pour les problèmes statistique liées à l'estimation des paramètres

Les méthodes classiques d'inférence statistiques ne permettent pas d'obtenir des réponses correctes à tous les problèmes concrets que se pose l'utilisateur. Elles ne

sont en effet valables que sous des conditions d'application particulières. Des solutions proposées permettent incontestablement d'élargir l'éventail des problèmes auxquels une solution peut être rapportée. Elles ne permettent cependant pas de résoudre tous les problèmes. L'accès généralisé à des moyens de calcul puissant a permis le développement de méthodes d'inférence statistique basées sur l'utilisation intensive de l'ordinateur. Le *bootstrap* fait partie de ces méthodes.

“Le *bootstrap* est une méthode d'inférence statistique basée sur l'utilisation de l'ordinateur qui peut répondre sans formules à beaucoup de questions statistiques réelles”. C'est en ces termes qu'Efron et Tibshirani (1993) présentent le *bootstrap* dans la préface de leur ouvrage.

Le principe général de la méthode est de rééchantillonner un grand nombre de fois l'échantillon initial qui a été réellement prélevé dans la population, l'inférence statistique étant basée sur les résultats des échantillons ainsi obtenus.

L'objectif de cette section est de décrire comment le *bootstrap* peut être utilisé pour résoudre les problèmes d'inférence statistique en relation avec l'estimation des paramètres. Nous présentons d'abord les méthodes de rééchantillonnage. Ensuite, nous examinons l'estimation de l'erreur standard et du biais d'un estimateur. Des informations plus détaillées concernant le bootstrap et d'autres méthodes de rééchantillonnage sont données, notamment, par Chernick (1999), Efron et Tibshirani (1993) et Manly (1997).

## Méthode de rééchantillonnage

### Bootstrap des individus

On considère un échantillon de  $n$  observations :  $x_1, x_2, \dots, x_n$ , prélevé de manière aléatoire et simple dans une population. Ces observations peuvent concerner une seule variable, ou, au contraire, être relatives à plusieurs variables. Dans ce cas, les  $x_i$  représentent des vecteurs de dimension  $p$ ,  $p$  étant le nombre de variables. Afin de ne pas alourdir les notations, nous ne distinguerons pas ces deux situations et, de manière plus condensée, nous désignerons l'échantillon initial par le symbole  $x$ , qu'il s'agisse d'un vecteur ou d'une matrice. Le principe de la méthode du *bootstrap* est de prélever une série d'échantillons aléatoires et simples avec remise de  $n$  observations dans l'échantillon initial, considéré comme une population. Ces échantillons successifs seront notés :

$x_1^*, x_2^*, \dots, x_B^*$ ,  $B$  étant le nombre de rééchantillonnages effectués.

Pour l'ensemble des  $B$  échantillons obtenus par *bootstrap* (*bootstrap sample*), les observations  $x_i$  n'apparaissent pas en nombre égal et on peut définir les proportions d'apparition  $P_i^*$  de chacune des observations,  $P_i^*$  étant égal au nombre de fois que l'observation  $x_i$  a été prélevée pour l'ensemble des  $B$  échantillons, divisé par le nombre total de prélèvement, qui est égal à  $nB$ . Ces proportions  $P_i^*$  interviennent dans certaines estimations. Des méthodes de rééchantillonnages assurant l'égalité de ces proportions sont également proposées. Cette approche porte le nom de rééchantillonnage balancé.

### Erreur-Standard et biais d'un paramètre

#### L'estimation de l'erreur standard :

Soit un paramètre  $\theta$  de la population et soit :

$$\begin{aligned}\hat{\theta} &= f(x_1, \dots, x_n) \\ &= f(x)\end{aligned}\tag{4.60}$$

Une estimation de ce paramètre, obtenue à partir des données de l'échantillon initial  $x$ .

Chaque échantillon obtenu par rééchantillonnage permet de calculer une répétition du *bootstrap* (*bootstrap replication*) de l'estimation  $\hat{\theta}$  :

$$\hat{\theta}_k^* = f(x_k^*) \quad (k = 1, \dots, B)\tag{4.61}$$

La fonction étant la même que celle utilisée pour la définition de  $\hat{\theta}$ . Supposons qu'on s'intéresse à la moyenne et la variance de la distribution et qu'on se propose d'estimer ces deux paramètres à partir de l'échantillon  $x$ . Si on utilise les estimateurs classiques, le paramètre  $\hat{\theta}$  s'écrit, successivement pour les deux paramètres considérés :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i\tag{4.62}$$

et

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2\tag{4.63}$$

Disposant de  $B$  répétitions, pour chaque rééchantillonnage  $x_k^*$  nous calculons l'estimation  $\hat{\theta}_k^*$ .

On peut déterminer la moyenne des estimations obtenues :

$$\hat{\theta}^* = \frac{1}{B} \sum_{k=1}^B \hat{\theta}_k^*\tag{4.64}$$

et l'écart-type des  $\hat{\theta}_k^*$  :

$$\hat{\sigma}_{\hat{\theta}^*} = \sqrt{\frac{\sum_{k=1}^B (\hat{\theta}_k^* - \hat{\theta}^*)^2}{(B-1)}}\tag{4.65}$$

L'écart type  $\hat{\sigma}_{\hat{\theta}^*}$  est une estimation de l'erreur standard de l'estimateur du paramètre  $\theta$ . Pour les situations où on dispose d'un estimateur de cette erreur standard, et pour autant que les conditions d'application soient remplies, on peut montrer que l'écart type des  $\hat{\theta}_k^*$  tend vers le résultat analytique, lorsque  $B$  tend vers l'infini.

Ainsi, pour la moyenne d'un échantillon aléatoire et simple, on sait que l'erreur standard de la moyenne est égale à  $\frac{\sigma}{\sqrt{n}}$ . Si  $B$  tend vers l'infini, l'écart-type  $\hat{\sigma}_{\hat{\theta}^*}$  tend vers  $\frac{\hat{\sigma}_{plug}}{\sqrt{n}}$  avec :

$$\widehat{\sigma}_{plug} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad (4.66)$$

L'estimation  $\widehat{\sigma}_{plug}$  de l'écart-type de la population donnée ci-dessus est appelée estimation par insertion (*plug-in estimator*). Pour un estimateur par insertion, la formule conduisant à l'estimation est la même que celle utilisée pour la définition du paramètre de la population finie de taille  $N$  et de moyenne  $m_X$ , on a :

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - m_X)^2}{N}} \quad (4.67)$$

En d'autres termes, pour un estimateur par insertion, on considère l'échantillon comme une population particulière et on utilise la formule relative au paramètre de la population.

D'une manière générale, lorsque  $B$  tend vers l'infini, la valeur  $\widehat{\sigma}_{\widehat{\theta}^*}$  tend vers une valeur fixée qui correspond à l'estimation de l'erreur standard du *bootstrap* idéal. Efron et Tibshirani (1993) proposent les règles empiriques suivantes pour le choix de  $B$  :

- Un nombre réduit de répétition ( $B = 25$ , par exemple) permet de d'obtenir une première information et  $B = 50$  est généralement suffisant pour avoir une bonne estimation de l'erreur standard ;
- Il est très rare que plus de 200 répétitions soient nécessaires pour estimer une erreur standard.

On peut noter que le choix de  $B$  n'est pas fonction de la taille  $n$  de l'échantillon.

### Estimation du biais :

Le biais d'un paramètre peut être estimé par la méthode du *bootstrap* de la manière suivante :

$$bias_B(\widehat{\theta}) = \widehat{\theta}^* - \widehat{\theta}_{plug} \quad (4.68)$$

$\widehat{\theta}_{plug}$  étant l'estimation par *plug-in* du paramètre  $\theta$ , même si  $\widehat{\theta}$  n'est pas une estimation par *plug-in*. Disposant d'une estimation du biais, on peut éventuellement corriger l'estimation initiale. On obtient alors :

$$\widehat{\theta}_C = \widehat{\theta} - bias_B(\widehat{\theta}) \quad (4.69)$$

On notera cependant que la correction symétrique du biais, par la relation ci-dessus, peut s'avérer dangereuse dans la mesure où cette correction peut augmenter l'erreur standard de manière importante. En pratique, lorsque le rapport du biais à l'erreur standard est inférieur à 0.25, il est souvent préférable de ne pas corriger l'estimateur pour le biais. D'autre part, un rapport supérieur à 0.25 peut être une indication que la statistique  $\widehat{\theta} = f(x_1, \dots, x_n)$  est inappropriée pour estimer  $\theta$ .

### 4.5.2 L'application du *bootstrap* dans l'estimation de l'erreur standard des estimations des paramètres obtenues par l'algorithme EM

L'estimation de l'erreur standard de  $\hat{\theta}$  peut être calculée suivant les étapes suivantes :

1. Une nouvelle série de donnée  $Y^*$  appelé l'échantillon de Bootstrap, est généré à partir de  $\hat{F}$ , tel que  $\hat{F}$  est une estimation de la fonction de distribution de  $Y$  formée à partir de la donnée observée originale  $Y$  ;
2. L'algorithme EM est appliqué à la série de donnée observée  $Y^*$  de Bootstrap pour calculer l'EMV pour cette série de donnée  $\hat{\theta}^*$  ;
3. La matrice de covariance de  $\hat{\theta}^*$  est donnée par.

$$cov^* (\hat{\theta}^*) = E^* \left[ \left\{ \hat{\theta}^* - E^* (\hat{\theta}^*) \right\} \left\{ \hat{\theta}^* - E^* (\hat{\theta}^*) \right\}^T \right] \quad (4.70)$$

où  $E^*$  dénote l'espérance sur la distribution de Bootstrap spécifié par  $\hat{F}$ . La matrice de covariance peut être approximée par la méthode de Monte Carlo. Les étapes 1 et 2 sont répétées indépendamment  $B$  fois. Pour donner  $B$  réalisations indépendantes de  $\hat{\theta}^*$ , dénotées par  $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$ . Puis, nous pouvons approximer l'équation (4.70) par :

$$cov^* (\hat{\theta}^*) \simeq \frac{\sum_{b=1}^B (\hat{\theta}_b^* - \bar{\theta}^*) (\hat{\theta}_b^* - \bar{\theta}^*)^T}{(B-1)} \quad (4.71)$$

où

$$\bar{\theta}^* = \frac{\sum_{b=1}^B \hat{\theta}_b^*}{B} \quad (4.72)$$

L'erreur standard de  $i$ ème élément de  $\hat{\theta}$  peut être estimée par la racine carrée de  $i$ ème élément de la diagonale de (4.71). Il est prouvé que 50 à 100 réplication de Bootstrap sont suffisantes pour donner la valeur de l'estimation de l'erreur standard (Efron et Tishirani, 1993).

## 4.6 Critère de la sélection du modèle *MAR-ARCH*

Dans cette section nous allons étudier le problème de la sélection du modèle MAR-ARCH. Nous considérons le critère de l'information bayésienne (BIC) (Shwarz 1978) qui a été utile dans le cas des modèles MAR (Wong et Li 2000).

Il existe deux aspects pour sélectionner le modèle, le nombre de composantes  $K$  et l'ordre de chaque composante AR-ARCH. Dans cette section, on s'intéresse à la sélection de l'ordre de chaque composante, car l'utilisation du BIC pour choisir  $K$  est un problème non standard. En effet, il correspond aux problèmes de tests avec paramètres de nuisance qui n'existent pas sous l'hypothèse nulle (Davies 1977,1987).

Nous utilisons la définition du BIC suivante :

$$BIC = -2Nl^* + \log(n - p_{\max} - q_{\max}) \left( 3K - 1 + \sum_{k=1}^K p_k + \sum_{k=1}^K q_k \right) \quad (4.73)$$

ici  $l^*$  est la log-vraisemblance maximisé calculé à partir de la fonction densité de probabilité (conditionnelle) du modèle  $MAR - ARCH$ .

$$\begin{aligned} l^* &= \frac{1}{N} \sum_{t=p+q+1}^n \log f(y_t | \Omega_{t-1}) \\ &= \frac{1}{N} \sum_{t=p+q+1}^n \log \frac{\partial}{\partial y_t} F(y_t | \Omega_{t-1}) \end{aligned} \quad (4.74)$$

Les ordres des composantes sont choisis de sorte que la valeur du  $BIC$  soit minimale.

## 4.7 Etude de simulation

### 4.7.1 Etude de simulation pour le BIC

Nous avons fait une étude de simulation pour illustrer la performance de la procédure de la sélection de l'ordre BIC. Nous avons généré 50 échantillons de taille 500 à partir du modèle (1) :  $MAR - ARCH(2; 1, 1; 1, 1)$  avec les valeurs des paramètres  $(\alpha_1, \alpha_2, \phi_{10}, \phi_{11}, \beta_{10}, \beta_{11}, \phi_{20}, \phi_{21}, \beta_{20}, \beta_{21}) = (0.5, 0.5, 1.0, 0.7, 1.0, 0.5, -1.0, -0.7, 1.0, 0.5)$ . Pour chaque échantillon nous avons sélectionné les ordres de chaque composante utilisant le BIC. Les résultats obtenus sont présentés dans la table 4.1 :

paramètre	la vraie valeur	le nombre de fois que la vraie valeur est sélectionnée
$p_1$	1	46
$p_2$	1	37
$q_1$	1	41
$q_2$	1	33

**Tab. 4.1** — La sélection du modèle (1) utilisant le BIC

#### Interprétation des résultats :

- Nous remarquons que le BIC a bien estimé les ordres ;
- Pour chaque composante  $p = \max(p_1, p_2)$  n'a pas dépassé 3 et  $q = \max(q_1, q_2)$  n'a pas dépassé 5 ;
- Nous signalons que le temps d'exécution est très élevé.

**Remarque 22** *Nous pouvons compter sur la qualité de l'estimation via l'algorithme EM si le BIC a choisit des ordres égales ou supérieur aux vrais ordres, mais la qualité*

de l'estimation peut se détérioré si le BIC a choisit dans quelques composantes un ordre inférieur au vrai ordre.

### 4.7.2 Etude de simulation pour l'algorithme EM

Dans cette section nous allons présenter une étude de simulation, afin de prouver la performance de l'algorithme EM dans l'estimation des paramètres dans la classe des modèles  $MAR - ARCH$ . Nous allons appliquer l'algorithme EM des séries générées par les trois modèles suivants :

Modèle (1) :

$MAR - ARCH(2; 1, 1; 1, 1)$  avec les valeurs des paramètres

$$(\alpha_1, \alpha_2, \phi_{10}, \phi_{11}, \beta_{10}, \beta_{11}, \phi_{20}, \phi_{21}, \beta_{20}, \beta_{21}) = (0.5, 0.5, 1.0, 0.7, 1.0, 0.5, -1.0, -0.7, 1.0, 0.5).$$

Modèle (2) :

$MAR - ARCH(2; 1, 1; 1, 1)$  avec les valeurs des paramètres

$$(\alpha_1, \alpha_2, \phi_{10}, \phi_{11}, \beta_{10}, \beta_{11}, \phi_{20}, \phi_{21}, \beta_{20}, \beta_{21}) = (0.75, 0.25, 0, 0.5, 1.0, 0.5, 0, 1.1, 1.0, 1.2).$$

Modèle (3) :

$MAR - ARCH(3; 1, 1, 1; 1, 2, 1)$  avec les valeurs des paramètres

$$(\alpha_1, \alpha_2, \alpha_3, \phi_{10}, \phi_{11}, \beta_{10}, \beta_{11}, \phi_{20}, \phi_{21}, \beta_{20}, \beta_{21}, \beta_{22}, \phi_{30}, \phi_{31}, \beta_{30}, \beta_{31}) = (0.4, 0.3, 0.3, 1.0, 0.9, 1.0, 0.3, 0, -0.7, 1.0, 0.2, 0.4, -1.0, 0.2, 1.0, 0.5).$$

Nous avons généré des échantillons indépendants de différentes tailles. Pour chaque taille, l'algorithme EM est répété 500 fois. Et notons aussi que les estimations obtenus vérifient les conditions de stationnarité au second ordre. Les résultats d'estimations sont présentés dans les tables (4.2), (4.3) et (4.4)

paramètre	V.I	la vraie valeur	N=150	N=200	N=300	N=500
$\alpha_1$	0.75	0.5	0.5093	0.4896	0.04948	0.4951
$\alpha_2$	0.25	0.5	0.4907	0.5004	0.4986	0.5009
$\phi_{10}$	0.75	1	1.0173	1.0162	0.9978	0.9775
$\phi_{20}$	-0.75	-1	-0.9917	-1.0071	-1.0024	-0.9976
$\phi_{11}$	0.5	0.7	0.6588	0.6778	0.6933	0.6842
$\phi_{21}$	-0.55	-0.7	-0.6898	-0.6971	-0.6852	-0.6889
$\beta_{10}$	0.75	1	1.1332	0.9257	0.9864	1.0069
$\beta_{20}$	0.75	1	0.9128	0.9241	0.9808	0.9881
$\beta_{11}$	0.25	0.5	0.4947	0.4983	0.4985	0.5163
$\beta_{21}$	0.35	0.5	0.4741	0.4871	0.5134	0.5094

Tab. 4.2 — Résultats d'estimation pour le modèle 1

#### Interprétation des résultats :

1. Nous remarquons que tous les paramètres sont bien estimés par l'algorithme EM.
2. La table (4.2) illustre la performance de l'algorithme EM pour un modèle MAR-ARCH avec deux processus AR-ARCH stationnaires.

paramètre	V.I	la vraie valeur	N=150	N=200	N=300	N=500
$\alpha_1$	0.5	0.75	0.7133	0.7266	0.7359	0.7414
$\alpha_2$	0.5	0.25	0.2767	0.2744	0.2641	0.2586
$\phi_{10}$	0.1	0	0.0420	-0.0227	0.0331	0.0039
$\phi_{20}$	0.1	0	0.1	-0.0196	0.0097	0.0014
$\phi_{11}$	0.4	0.5	0.4480	0.4609	0.4768	0.5046
$\phi_{21}$	1	1.1	1.0079	1.0093	1.0225	1.0616
$\beta_{10}$	0.75	1	1.0419	1.0279	0.9644	0.9880
$\beta_{20}$	0.75	1	0.6892	0.6965	1.0459	0.9953
$\beta_{11}$	0.35	0.5	0.3831	0.4289	0.4320	0.4568
$\beta_{21}$	1	1.2	1.4171	1.1444	0.9702	1.2266

Tab. 4.3 — Résultats d'estimation pour le modèle 2

paramètre	la vraie valeur	N=1000
$\alpha_1$	0.4	0.3474
$\alpha_2$	0.3	0.3216
$\alpha_3$	0.3	0.3280
$\phi_{10}$	1	0.8524
$\phi_{20}$	0	-0.3247
$\phi_{30}$	-1	-0.7379
$\phi_{11}$	0.9	0.8792
$\phi_{21}$	-0.7	-1.0117
$\phi_{31}$	0.2	0.1914
$\beta_{10}$	1	1.6893
$\beta_{20}$	1	1.4300
$\beta_{30}$	1	0.9231
$\beta_{11}$	0.3	0.3375
$\beta_{21}$	0.2	0.5354
$\beta_{31}$	0.5	0.6535
$\beta_{22}$	0.4	0.4008

Tab. 4.4 — Résultats d'estimation pour le modèle 3

3. La table (4.3) illustre le cas d'un mélange de processus AR-ARCH stationnaires avec un processus AR-ARCH non stationnaire.
4. La table (4.4) illustre le cas d'un MAR-ARCH avec trois processus AR-ARCH stationnaires. Dans ce cas, nous avons augmenté la taille de l'échantillon, pour assurer un nombre d'observations suffisant pour estimer les paramètres pour chaque composante.
5. Nous signalons que le temps d'exécution est très élevé.
6. Nous constatons également que plus la taille de l'échantillon augmente plus les estimateurs des paramètres s'améliorent. Nous en déduisons donc que les estimateurs des paramètres sont consistants.



7. Le problème de dégénérescence de la vraisemblance figure souvent.

Pour vérifier l'efficacité de cette méthode, nous utilisons le critère ESE (Emperical Standard Errors), défini par :

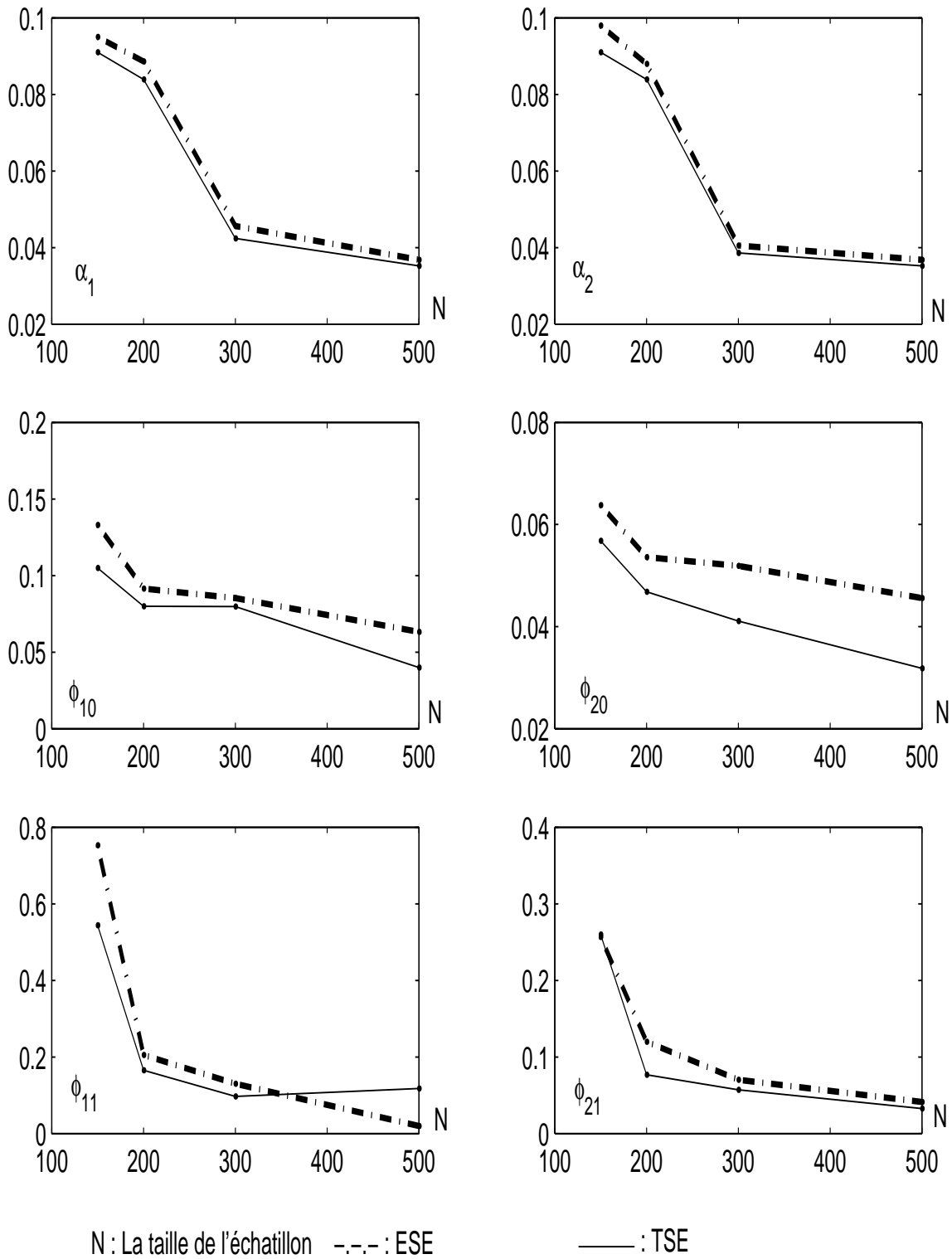
$$ESE(\hat{\theta} | y_{obs}) = \sqrt{\frac{1}{m} \sum_{j=1}^m \left( \hat{\theta}_j - \frac{1}{m} \sum_{j=1}^m \hat{\theta}_j \right)^2} \quad (4.75)$$

qui permet d'évaluer la précision de l'estimateur au sens de l'erreur quadratique. Nous avons calculé les ESE des estimations des paramètres du modèle (1). Nous avons aussi appliqué la méthode du Bootstrap pour le calcul de l'ESE (Theoretical Standard Errors) des estimations des paramètres du modèle (1) obtenues par l'algorithme EM

Les graphes dans les figures(4.1) et (4.2) donnent la variation de ESE et de TSE pour chaque estimation du paramètre en fonction de la taille de l'échantillon pour le modèle (1).

Nous remarquons que :

1. Plus la taille de l'échantillon augmente plus la valeur de ESE diminue. Nous déduisons donc que les estimateurs des paramètres de modèle *MAR - ARCH* par l'algorithme EM, sont consistants.
2. Nous remarquons aussi que les erreurs standards théoriques des estimations (TSE) calculées par la méthode du bootstrap sont proches des erreurs standards empiriques (ESE) des estimations obtenues par l'algorithme EM. Ce qui prouve que la méthode de l'estimation que nous avons utilisé est non biaisée, car nous avons obtenu des biais petits et raisonnables.



**Fig. 4.1** — la variation de ESE et de TSE en fonction de la taille de l'échantillon pour les paramètres  $\alpha_1$ ,  $\alpha_2$ ,  $\phi_{10}$ ,  $\phi_{20}$ ,  $\phi_{11}$  et  $\phi_{21}$  du modèle (1).

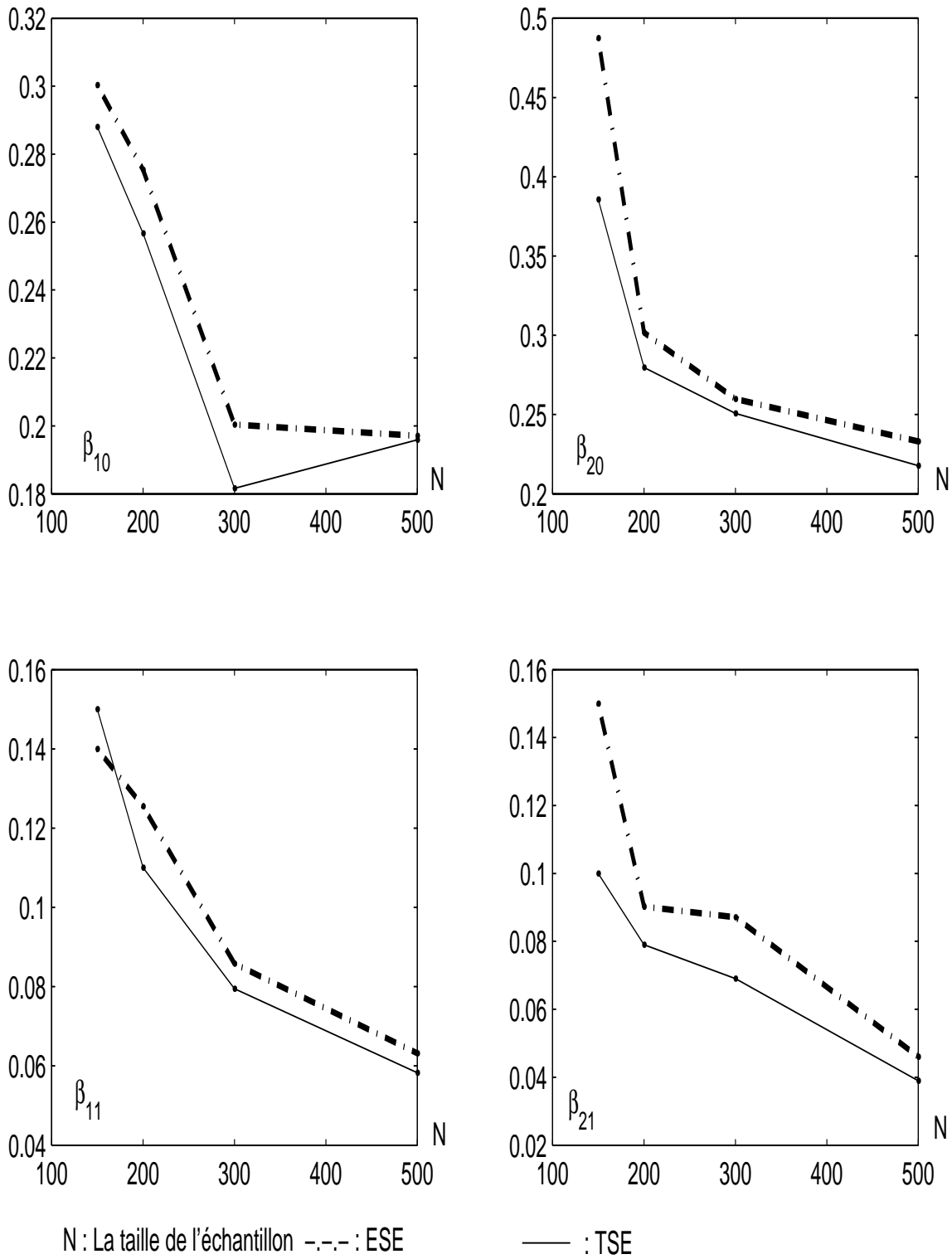


Fig. 4.2 — la variation de ESE et de TSE en fonction de la taille de l'échantillon pour les paramètres  $\beta_{10}$ ,  $\beta_{20}$ ,  $\beta_{11}$ ,  $\beta_{21}$  du modèle (1).

CHAPITRE

# 5

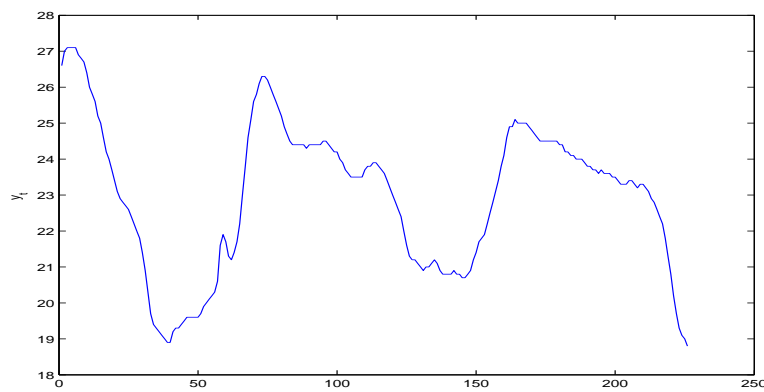
---

## Application sur une série de données réelles

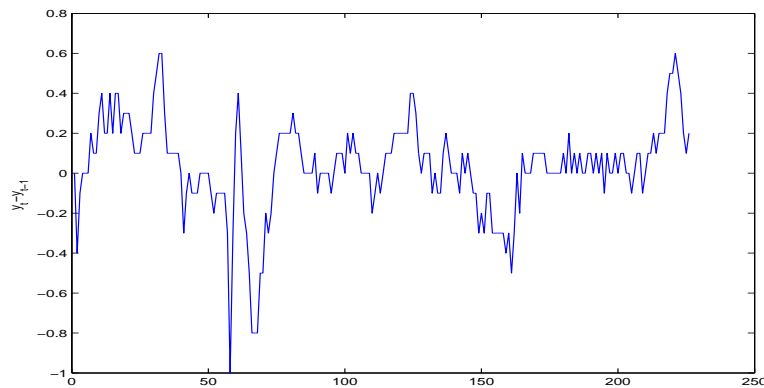
Dans ce chapitre, nous allons illustrer l'utilité, la performance, et la qualité de prévision des modèles MAR-ARCH. A l'aide d'une série de données réelles. Nous allons comparer entre la modélisation, utilisant les modèles MAR-ARCH, et la modélisation utilisant les modèles ajustés par Box, Jenkins, et Reinsel (1994).

## 5.1 Présentation de la série test

Nous proposons la série de températures, lues chaque minute, à partir d'un processus chimique (226 observations). Cette série est rapportée par Box, Jenkins, et Reinsel (1994) (série C). Cette série de donnée a été largement analysée (E, G, Gray, Kelly, et McIntire (1978); Tsay (1987), Tiao (1984)). La série originale, la série différenciée au premier ordre, et la série différenciée au deuxième ordre sont présentées dans les figures (5.1), (5.2), (5.3).



**Fig. 5.1** — La série originale.

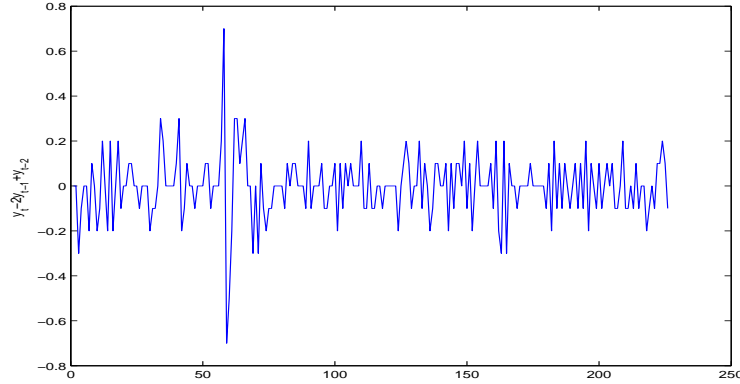


**Fig. 5.2** — la série différenciée au premier ordre.

Nous remarquons, à partir de ces figures que la série n'est pas stationnaire. Il n'est pas aussi confirmé que la série obtenue après la première différenciation est stationnaire, et que la différenciation au deuxième ordre est nécessaire. Nous remarquons aussi que la variabilité de la série différenciée au second ordre n'est pas uniforme.

Box et autres suggèrent deux modèles ARIMA pour modéliser cette série :

Le premier modèle est un ARIMA (1, 1, 0) :  $y_t - 1.82y_{t-1} + 0.82y_{t-2} = e_t$ , où  $e_t \sim \mathcal{N}(0, 0.018)$ ,



**Fig. 5.3** — La série différenciée au deuxième ordre.

Le deuxième modèle est un ARIMA (0, 2, 2) :  $y_t - 2y_{t-1} + y_{t-2} = e_t - 0.13e_{t-1} - 0.12e_{t-2}$ , où  $e_t \sim \mathcal{N}(0, 0.019)$

Mais ces deux modèles ne sont pas adéquats en effet, les autocorrélations du carré des résidus sont larges. Ceci est due à la présence de l'hétéroscédasticité (Tsay. (1987)).

## 5.2 Présentation d'un autre modèle candidat

### 5.2.1 Ajustement du modèle MAR-ARCH

Nous considérons le modèle MAR-ARCH, proposé par Wong et Li, pour modéliser la série différenciée au premier ordre. Les ordres des composantes AR-ARCH sont choisis par la minimisation de la fonction BIC. Le meilleur modèle proposé par Wong et Li est un MAR-ARCH (2; 1, 1; 0, 1), avec  $\phi_{1,0} = \phi_{2,0} = 0$  et le BIC = -700.73. Les estimations et les erreurs standard pour les paramètres  $(\alpha_1, \alpha_2, \phi_{11}, \beta_{10}, \phi_{21}, \beta_{20}, \beta_{21})$  sont (0.2738, 0.7262, 0.5377, 0.0037, 0.9966, 0.0102, 0.4725) et (0.0865, 0.08665, 0.0487, 0.0016, 0.0499, 0.0018, 0.1557).

Le modèle MAR-ARCH a deux composantes, choisit pour la série différenciée au premier ordre, est aussi transformé à un modèle MAR-ARCH à deux composantes, définit par :

$$F(y_t | \Omega_{t-1}) = 0.2738\Phi\left(\frac{e_{1,t}}{\sqrt{h_{1,t}}}\right) + 0.7262\Phi\left(\frac{e_{2,t}}{\sqrt{h_{2,t}}}\right) \quad (5.1)$$

tel que

$$\begin{aligned} e_{1,t} &= y_t - 1.5377y_{t-1} + 0.5377y_{t-2}, h_{1,t} = 0.0037 \\ e_{2,t} &= y_t - 1.9966y_{t-1} + 0.9966y_{t-2}, \\ h_{2,t} &= 0.0102 + 0.4725e_{2,t-1}^2 \end{aligned}$$

Ici la première composante est un processus ARIMA d'ordre d'intégration  $d = 1$ , et la deuxième composante peut être considérée comme étant un processus ARIMA d'ordre  $d = 2$ , car l'estimation du paramètre  $\phi_{21}$  est proche de 1.

Notons que si nous donnons à  $\phi_{21}$  la valeur 1 dans l'estimation, la valeur du  $BIC$  sera égale à -706.13.

## 5.2.2 La variance conditionnelle

La variance conditionnelle  $Var(y_t | \Omega_{t-1})$  est présentée dans la figure (5.4). Nous remarquons que le comportement de la variance conditionnelle calculée à partir du modèle MAR-ARCH ajusté se ressemble avec la variabilité des données de la série différencier au deuxième ordre.

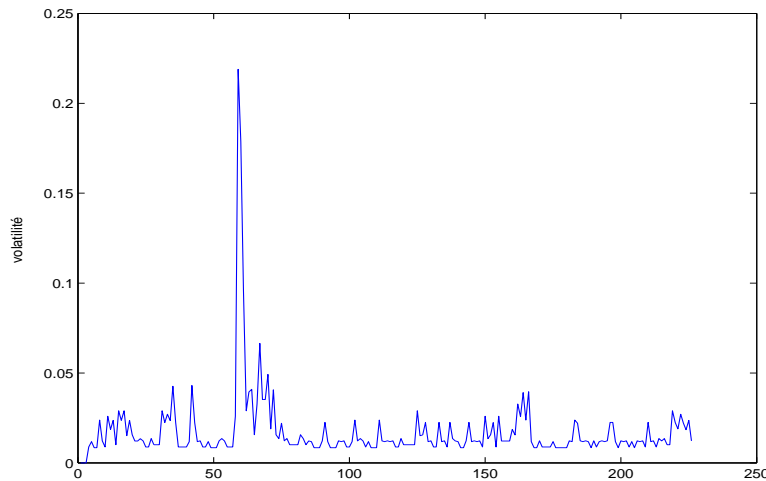


Fig. 5.4 — La volatilité calculée à partir du modèle MAR-ARCH ajusté.

## 5.2.3 Présentation de quelques distributions prédictives

Les distributions prédictives au premier pas  $F(y_{t+1} | \Omega_t)$  aux instants  $t = 199$ ,  $t = 220$ , et  $t = 221$  sont présentées dans les figures (5.5), (5.6), (5.7), et avec les valeurs actuelles, les formes des distributions prédictives sont unimodales ou bimodales.

## 5.3 Comparaison du modèle MAR-ARCH avec d'autres modèles ajustés

Nous comparons les modèles MAR-ARCH avec les modèles ajustés par Box et autres (1994), et un autre modèle AR-ARCH ajusté défini par :

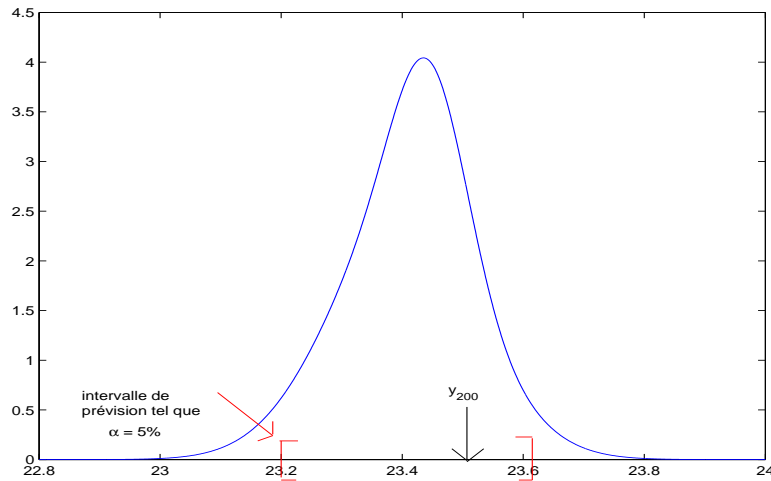


Fig. 5.5 — La distribution prédictive d’horizon 1 à l’instant  $t = 199$ .

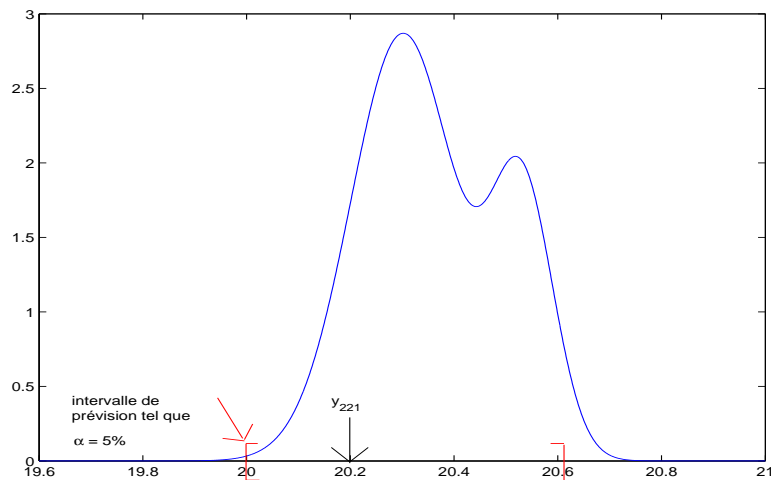


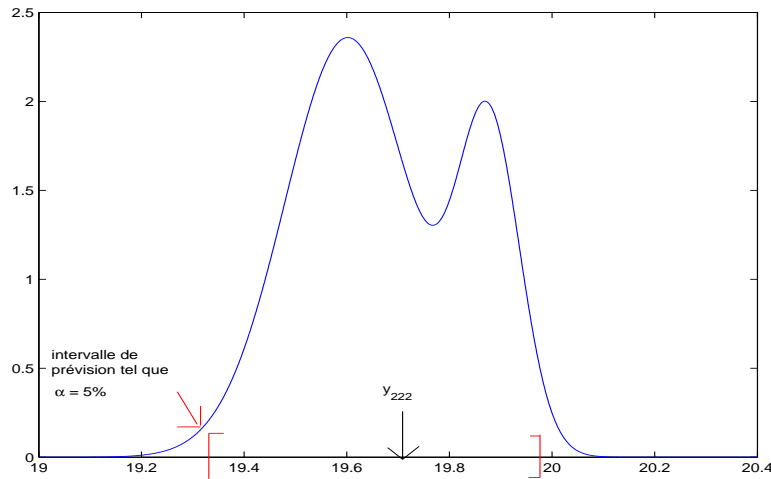
Fig. 5.6 — La distribution prédictive d’horizon 1 à l’instant  $t = 220$ .

$$\begin{aligned}
 y_t &= 1.8427y_{t-1} - 0.8427y_{t-2} + \varepsilon_t, & \varepsilon_t &\sim \mathcal{N}(0, h_t), \\
 h_t &= E(\varepsilon_t^2 | \Omega_{t-1}) = 0.0098 + 0.4104\varepsilon_{t-1}^2.
 \end{aligned}$$

Nous appliquons la procédure suivante pour estimer la prévision exacte obtenues par ces modèles :

Les 200 premières observations sont réservées pour la partie d’initialisation, et les 26 dernières observations sont réservées pour tester les estimations calculées. Nous générons les prévisions du premier horizon au cinquième horizon, à partir des trois modèles proposés pour tester la qualité de prévision. La figure (5.8) donne les moyennes des erreurs carrées de ces prévisions.



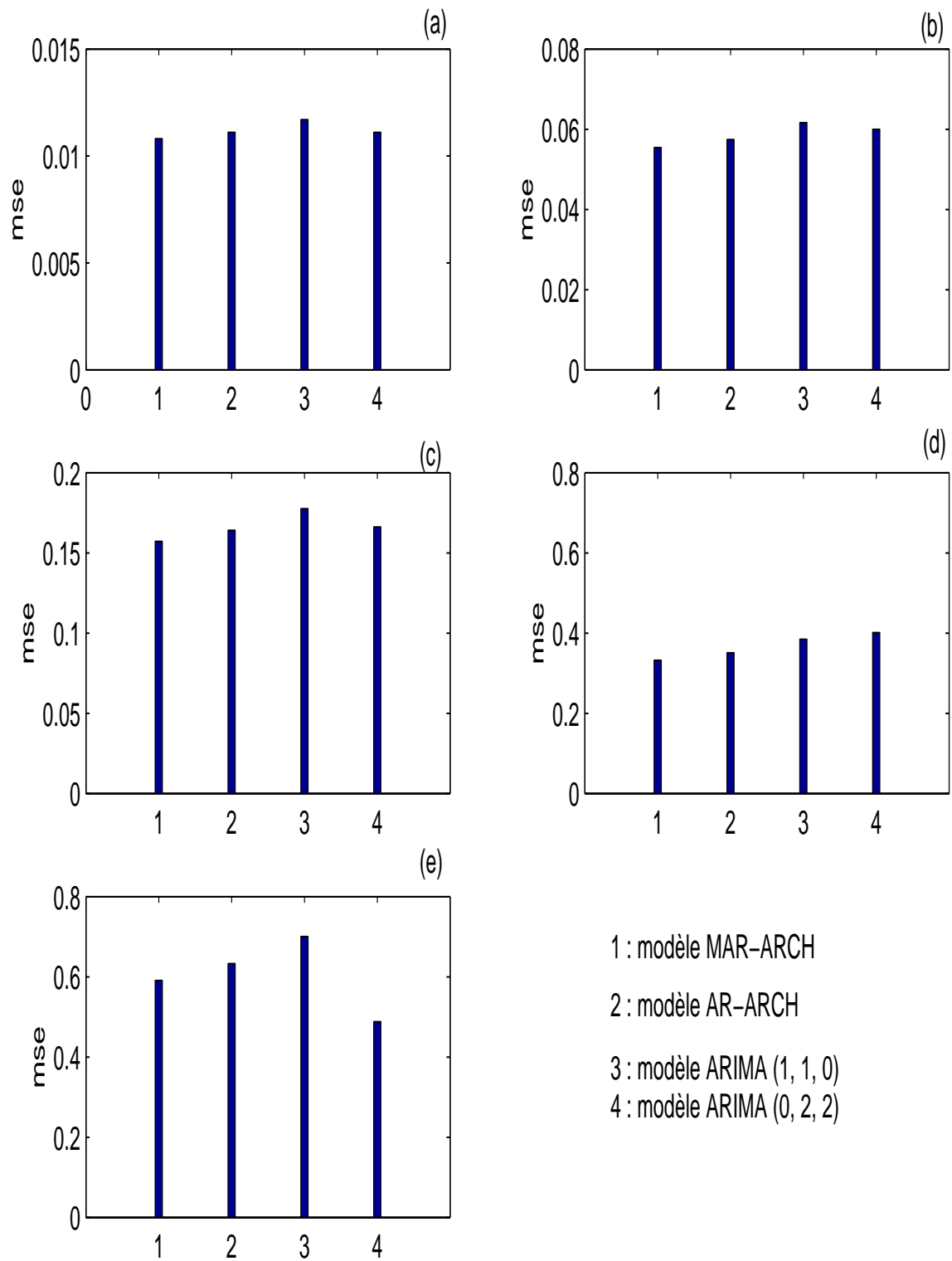


**Fig. 5.7** — La distribution prédictive d’horizon 1 à l’instant  $t = 221$ .

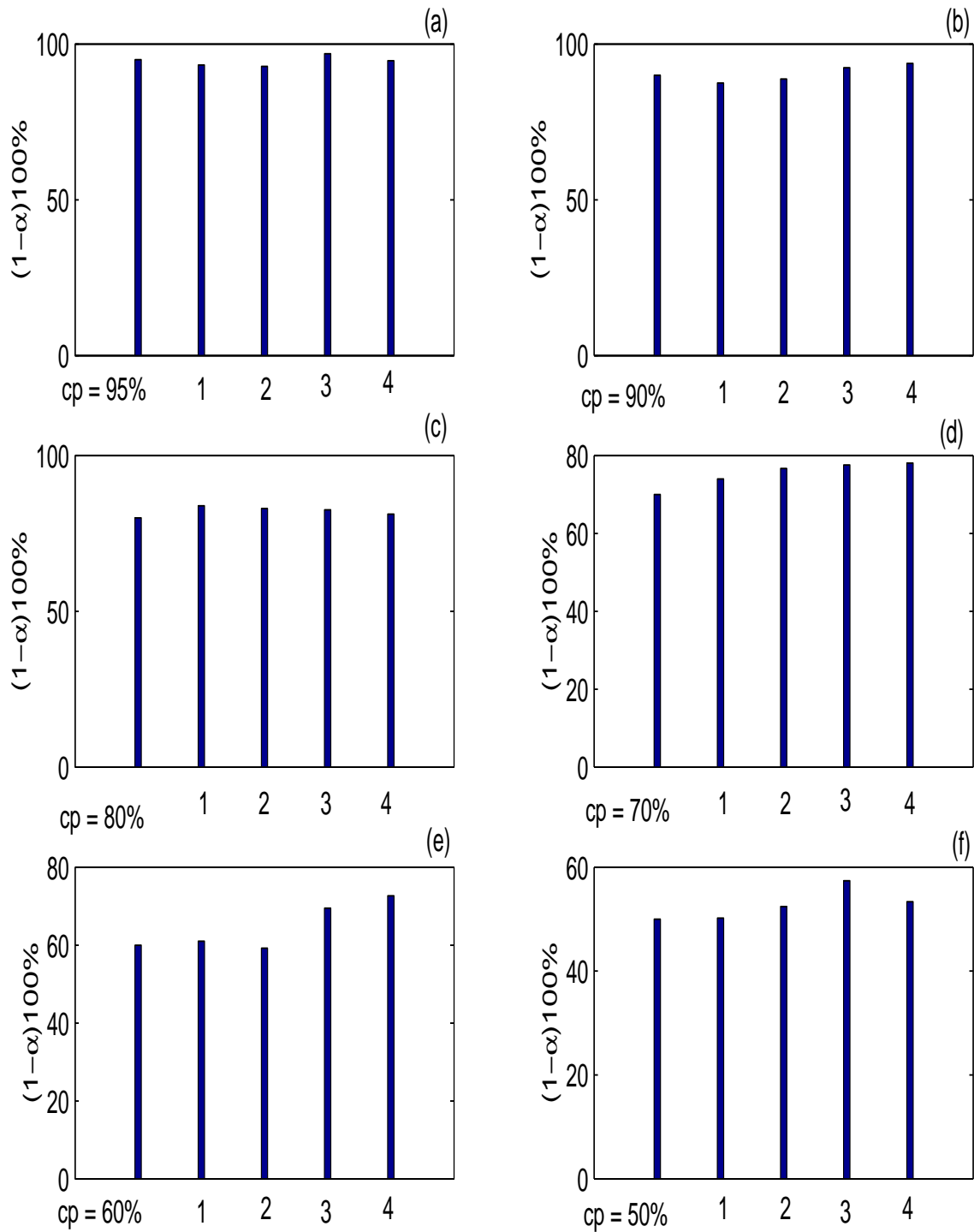
Nous remarquons que les prévisions générées à partir du modèle MAR-ARCH sont plus performantes quand l’horizon de prévision est petit. Par contre les prévisions générées à partir du modèle ARIMA (0, 2, 2) sont plus performantes quand l’horizon de prévision est large.

Bien que le mérite des modèles MAR-ARCH est dans sa capacité de décrire la distribution conditionnelle de la série temporelle, la capacité des modèles dans la description de la distribution prédictive peut être comparée par le calcul de la couverture empirique des intervalles de prévision aux premier et au deuxième pas construisaient par ces modèles. Pour chaque instant  $t$ , nous calculons les distributions prédictives au premier, et au deuxième pas, utilisant les modèles ajustés. En se basant sur ces distributions prédictives nous construisons les intervalles de prévision au premier et au deuxième pas. La couverture empirique de l’intervalle de prévision est défini par la proportion d’observations qui se trouvent à l’intérieur de cet intervalle. Les couvertures empiriques des intervalles de prévision au premier et au deuxième pas sont données dans les figures (5.9) et (5.10).

Nous remarquons que les couvertures obtenues par le modèle MAR-ARCH sont plus proches aux couvertures présumées par rapport aux couvertures basées sur les intervalles calculés à partir des modèles ARIMA, et AR-ARCH. Nous justifions cette particularité par le fait que les modèles ARIMA et AR-ARCH peuvent décrire uniquement les distributions prédictives unimodales. Notons que nous avons utiliser la méthode de Monte Carlo dans le calcul des intervalles de prévision au deuxième pas.

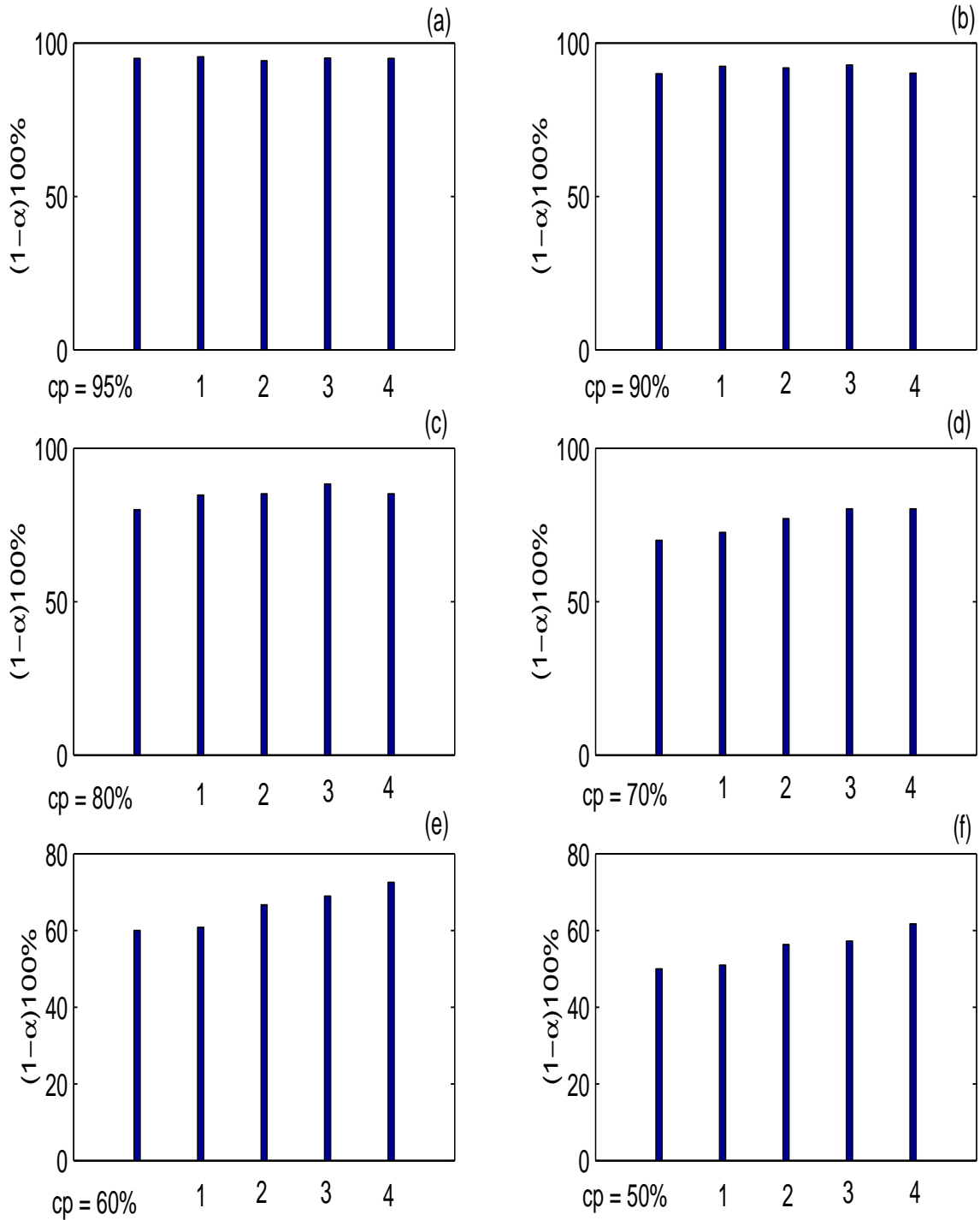


**Fig. 5.8** — (a),(b), (c), (d) et (e) donnent les MSE calculées des prévisions données par les quatres modèles ajustés d’horizon 1, 2, 3, 4 et 5 resp.



Modèles : 1 : MAR-ARCH, 2 : AR-ARCH, 3 : ARIMA (1, 1, 0), 4 : ARIMA (0, 2, 2)

**Fig. 5.9** — (a), (b), (c), (d), (e) et (f) : La comparaison entre les couvertures empiriques d'horizon 1 données par les modèles ajustés par rapport aux couvertures présumées :95%, 90%, 80%, 70%, 60% et 50% resp.



Modèles : 1: MAR-ARCH, 2 : AR-ARCH, 3 : ARIMA (1, 1, 0), 4 : ARIMA (0, 2, 2)

**Fig. 5.10** — (a), (b), (c), (d), (e) et (f) : La comparaison entre les couvertures empiriques d'horizon 2 données par les modèles ajustés par rapport aux couvertures présumées :95%, 90%, 80%, 70%, 60% et 50% resp.

---

# Conclusion générale

DANS ce travail de mémoire, nous considérons le problème de la modélisation des séries temporelles. Nous nous intéressons, en particulier, aux séries temporelles ayant des propriétés de non linéarité et non normalité, comme les distributions conditionnelles multimodales, les déploiements monotones, l'hétéroscédasticité, etc. Nous avons discuté une nouvelle classe de modèles de séries chronologiques non linéaires, la classe de mélanges de modèles autorégressifs à erreurs ARCH (MAR-ARCH). Cette classe permet plus de flexibilité par rapport aux autres modèles proposés pour modéliser ce type de séries. Pour aborder le problème, nous avons donné un aperçu sur la classe des mélanges de modèles autorégressifs. Nous avons cité ses propriétés qui lui rendent un outil potentiel, dans la modélisation des séries temporelles non linéaires, néanmoins, cette classe de modèles possède une structure d'autocorrélation de processus carré assez simple et elle est analogue à celle du modèle AR ordinaire. Cette classe de modèle nécessite une autre étude. Le modèle MAR-ARCH est l'extension proposée par Wong et Li. Cette classe permet plus de flexibilité dans sa structure d'autocorrélation de processus carré et une meilleure capture de l'hétéroscédasticité. Avant de présenter le modèle MAR-ARCH, il a été important de décrire le modèle ARCH. Nous avons montré que le modèle ARCH et ses extensions possèdent des propriétés que les modèles linéaires ARMA n'ont pas et permettent de modéliser les séries temporelles, présentant le phénomène de volatilité, de façon plus réaliste. Ensuite, nous avons présenté le modèle MAR-ARCH. Plusieurs résultats importants ont été prouvés :

- Un mélange de processus stationnaire avec un processus non stationnaire peut résulter un processus stationnaire ;
- Les conditions de stationnarité au premier ordre pour ces modèles, et les conditions de stationnarité au second ordre pour quelques modèles MAR-ARCH particulier ;
- La distribution conditionnelle prédictive change sa forme dans le temps, ce qui prouve la capacité des modèles MAR-ARCH à modéliser les séries temporelles avec hétéroscédasticité et leurs distributions conditionnelles sont multimodales ;
- Les modèles MAR-ARCH possèdent plus de flexibilité dans la modélisation de la variance conditionnelle variante dans le temps ;
- Nous avons présenté la condition nécessaire pour l'existence des moments d'ordre quatre pour des modèles MAR-ARCH particulier, et nous avons marqué que pour ces modèles la kurtosis est généralement supérieur à trois ;
- Sous certaines conditions et pour certains modèles, la fonction d'autocorrélation du processus carré est élaboré.

Le changement de la forme de la distribution conditionnelle prédictive, et la structure d'autocorrélation du processus carré sont les propriétés les plus attrayantes. Nous

avons aussi discuté la construction de la fonction prédictive au pas multiple. pour terminer notre étude, nous avons vu comment peut-on appliquer l'algorithme EM dans l'estimation des paramètres dans les modèles MAR-ARCH. Nous avons vu aussi comment utiliser la technique du Bootstrap dans l'estimation de l'erreur standard des estimations obtenues par l'algorithme EM. Et enfin, nous avons utilisé le critère BIC pour sélectionner le modèle. Afin de prouver la performance de l'algorithme EM et l'efficacité du BIC nous avons présenter une étude simulation. A l'aide d'une séries de données réelles, nous avons pu montré que la classe des modèles MAR-ARCH est puissante par rapport aux autres modèles usuels.

### **Perspectives**

- La comparaison entre l'approche bayésienne et l'algorithme EM dans l'estimation des paramètres du modèle MAR-ARCH
- La généralisation des modèles MAR-ARCH aux modèles MAR-ARCH multivarié,
- Des approches pour sélectionner le nombre de composantes.

---

# Bibliographie

- [1] Basford, K. E, Greenway, D. R., Mclachlan, G.J. and Peel, D. "Standard errors of fitted means under normal mixture models," *Computational Statistics*, 12, 1-17, 1997.
- [2] Benes, V.,E., "Existence of Finite Invariante Mesure for Markove Process ," *Proceeding of the American Mathematical Society*, 18,1058-1061, 1967.
- [3] Biernacki, C, "Initializing EM Using the Properties of its Trajectories in Gaussian Mixtures," *Statistics and Computing*, 14, 3, 267-279, 2004.
- [4] Bollerslev, T., "Generalized Aytoregressive Conditional Heteroscedasticity ," *Jornal of econometric*, 31,307-327, 1986.
- [5] Box G. E. P.,Jenkins. G. M.,and Reinsel. G. C. "Time series Analysis forecasting and Control (3rd ed)," ,Englewood Cliffs, Nj : Prentice-Hall,(1994).
- [6] Campbell J. Y., Lo A. et Mackinaly, A. C "The Econometrics of Finantial Markets," *Princeton University Press* , 1997.
- [7] Chan K. S and Tong H. "A Note on Testing for Multi-modality with Dependent Data." *Unpublished.*, 1998.
- [8] Chernick MR. "Bootstrap methods ," *a practitioner's guide. New YorK,wiley* , 264 p, 1999.
- [9] H. Cramér, "Mathematical Methods of Statistics," *Princeton University Press*, Princeton, NJ, 1946.
- [10] Davies, R. B., "Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative ," *Biometrika*, 64,247-254, 1977.

- 
- [11] Davies, R. B., "Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative ," *Biometrika*, 74,33-43, 1987.
- [12] N. E. Day. "Estimation of the Minimum of a Function Using Order Statistics ," *Biometrika*, No 56 :463 :474. 1969.
- [13] Dempster, A. P.,Laird, N. M.,and Rubin, D. B., "Maximum Likelihood From Incomplete Data via the EM Algorithm ," *Journal of the Royal Statistical Society*, Ser. B, No. 39,1-38, 1977.
- [14] Efron B,Tibshirani RJ, "An introduction to the bootstrap ," *New York : Chapman and hall*,436,1993.
- [15] Engle, R. F., "Autoregressive Conditional Heteroscedasticity With Estimates of Variance of United Kingdom Inflation ," *Econometrica*, No. 50,937-1007, 1982.
- [16] Engle, R. F.and T. Bollerslev "Modelling the persistence of conditional variances ," *Econometric Rev*, No. 5,1-50, 1986.
- [17] T. S. Ferguson."An Inconsistent Maximum Likelihood Estimate ," *Jornal of the American Statistical Association*, No 77(380) : 831-834, 1982.
- [18] Gaver, D. P. and Lewis, P. A. W. "Furst-order autoregressive Gamma sequences and point process ," *Adv. Appl. Probab.*, 12, 727-745, 1980.
- [19] Goldberg, S., "Difference Equations ," *New York ; Wiley*, 1958.
- [20] Gouriéroux, C., "Models ARCH et Applications Financières," *Collection ENSAE, Economica* ,1992.
- [21] Granger, C. W. J.,and Newbold, P., "Forecasting Economic Time Series ," *San Diego ; Acadimic Press*, 1986.
- [22] Granger, C. W. J.,and Travistra, T., "Modeling Nonlinear Economic Relationships ," *New York : Oxford University Prerss*,1993.
- [23] Granger, C.,W. , and Andersen. A. "An Introduction to Bilinear Time Seies Models ," *Vandehhoeck et Ruprechet, Gottingen* , 1978.
- [24] Gray, s. F., "Modling the Conditional Distribution of Interest Rates as a Regime-Switching Process," *Journal of Financial Economics*, B, No. 42,27-62, 1985.



- 
- [25] Gray, H. L., Kelly, G. D., and McIntire, D. D., "A new Approach to ARMA Modeling ," *Communications in Statistics*, Part B, 7, 1-77 No. , 1978.
- [26] Hamilton, J. D., and Susmel, R., "Autoregressive Conditional Heteroscedasticity and Changes in Regime ," *Journal of Econometrics*, 64, 307-333, 1994.
- [27] C. Hurlin, "Econometrie pour la finance ," *Master ESA. Econométrie et Statistique Appliquée* , 2004.
- [28] Lawrance, A, j and Lewis, P, A, W. "Modeling and residual analysis of nonlinear autoregressive time series in exponential variables," *J. R. statist. soc. B*, 47, 165-202., 1985.
- [29] Le, N. D., Martin, R. D., and Raftery, A. E., "Modeling Flat Stretches Bursts, and Outliers in Time Series Using Mixture Transtion Distribution Models ," *Journal of the American Statistical Association* , No. 91,1504-1514, 1996.
- [30] Lehmann. E. L. "Theory of Point Estimation ," *Wiley, New-York* , No. 1983.
- [31] Louis, T. A., "Finding the Observed Information Matrix hen Using the EM Algorithm ," *Journal of the Royal Statistical Society*, Ser. B, No. 44,226-233, 1982.
- [32] Manly BFJ., "Randomization, Bootstrap and Monte Carlo Methods in Biology," *New York : Chapman and Hall*,399 p, 1997.
- [33] McLachlan, G. J., and D. Peel "Finite Mixture Models ," *Wiley, New-York* , 2000.
- [34] McLachlan, G. J., and T. Krishnam "The EM Algorithm and Extensions ," *Wiley, New-York* , 1997.
- [35] W. Mendenhall and R. J. Hader "Estimation of Parameters of Mixed Exponentially Distributed Failure Time Distributions from Censored Life Test Data," *Biometrika* , No 45 :504 :520, 1958.
- [36] Meng, X. L., and Rubin, D. B., "Using EM to Obtain asymptotic Variance-Covariance Matrices : The SEM Algorithm ," *Journal of the American Statistical Association*, No. 86,899-909, 1991.
- [37] Meng, X. L., and Rubin, D. B., "Maximum Likelihood Estimation via the ECM Algorithm : A general framwork," *Biometrika*, No. 80,267-278, 1993.

- [38] Meng, X. L., "On the Rate of Convergence of the ECM Algorithm ," *The Annals of Statistics*, No. 22,326-339, 1994.
- [39] Meilijson, I., "A Fast Improvement to the EM Algorithm on its Own Terms ," *Journal of the Royal Statistical Society*, Ser. B, No. 44,226-233, 1989.
- [40] Nelson, D. B., "Stationarity and persistence in the GARCH (1,1 model)," *Econometric Theory* , No. 6,318-334, 1990.
- [41] Nelson, D. B., "Conditional Heteroskedasticity Asset Returns : A New Approach," *Econometrica*, No. 59,347-370, 1991.
- [42] J. Neyman and E. Scott. "Consistent Estimators Based on Partially Consistent Observations ," *Econometrica*, No. 16 : 1-32, 1948.
- [43] Orchard, T., and Woodbury, M. A., "A Missing Information Principle ; Theory and Application ," *Proc. of the 6th symp. on Math. Stat. and Prob.*, No. 1,679-715, 1972.
- [44] C. R. Rao. "The Utilization of Multiple Measurements in Problems of Biological Classification ," *Journal of the royal Statistical Society, Serie B* , No 10 :159-203, 1948.
- [45] R. A. Redner and H. F. Walker "Mixture Densities, Maximum Likelihood and the EM Algorithm ," *SIAM review* , No 26(2) : 195 :239, 1984.
- [46] Schwarz, G., "Estimating the Dimension of a Model ," *The Annals of Statistics*, No. 6,461-464, 1978.
- [47] C. G. Small, J. Wang, and Z. Yang. S., "Eliminating Multiple Root Problems in estimation ," *Statistical Science*, No. 15(4),313-341, 2000.
- [48] L. A. Stefanski and R. J. Carroll "Conditional Scores and Optimal Scores for Generalized linear Measurement-error Models ," *Biometrika*, No74(4) :703-716. 1987.
- [49] Tarone. R. D and G. Gruenhage , "A note on the Uniqueness of the Roots of the Likelihood equations for Vector-valued Parameters," *Journal of the American Statistical Association*, No.70,903-904, 1975.
- [50] Tong, H. "Non-linear Time Series," *New York : Oxford University Press* , 1990.
- [51] Tsay, R. S., "Conditional Heteroscedastic Time Series Models ," *Journal of the American Statistical Association* , No. 82,590-604, 1987.

- [52] Tsay, R. S., and Tiao, G. C., "Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA model ," *Journal of the American Statistical Association* , No. 79,84-96, 1984.
- [53] H. Wald., "Note on the Consistency of the Maximum Likelihood Estimate ," *Annals of the Mathematical Statistics* , No. 20,595-601, 1949.
- [54] Weiss, A. A., "Asymptotic Theory for ARCH Models : Estimation and Testting," *Econometric Theory*, No. 2,107-131, 1986.
- [55] H. White., "Maximum Likelihood Estimation of Misspecified Models ," *Econometrica*, No. 50(1) :1-25, 1982.
- [56] Wold, H. "A study in the stationnary time series," *Almqvist and Wiksell, Stockholm* , (2 nd edition : 1954) (1938).
- [57] J. H. Wolfe. "A Monte Carlo Study of the Sampling Distribution of the Likelihood Ratio for Mixtures of Multinomial Distributions ," *Technical Bulletin STB 72-2, US Naval Personnel Research Activity, San Diego, California* , 1971.
- [58] Wong, C. S., "Statistical Inference for Some Nonlinear Time series Models ," *doctoral thesis, The Univesity of Hong Kong* , 1998.
- [59] Wong,C. S.,and Li, W. K., "On a Mixture Autoregressive Model ," *Journal of the Royal Statistical Society, Ser. B*, No. 62,95-115, 2000.
- [60] Wong,C. S.,and Li, W. K., "On a Mixture Autoregressive Conditional Heteroscedastic Model ," *Journal of the American Statistical Association* , No. 455,982-995, 2001.

Dans cette annexe nous allons donner un aperçu sur la classe de modèles *GMTD*. Cette classe de modèles a été introduite par Le, Martin, et Raftery (1996) afin de modéliser les séries temporelles ayant des valeurs aberrantes, des valeurs explosives, et possèdent des déploiements monotones. Nous décrivons ce type de modèle, ensuite nous citons quelques ses propriétés. Et enfin nous présentons quelques résultats obtenus à partir de quelques exemples traités.

## A.1 Description du modèle GMTD

Supposons que  $\{y_t; t \in N\}$  est une série temporelle, une suite de variables aléatoires prenant leurs valeurs dans un espace arbitraire.  $\{y_t; t \in N\}$

est générées à partir d'un modèle de mélange de distributions de transition (*MTD*) si :

$$F(y_t | \Omega_{t-1}) = \sum_{i=1}^p \alpha_i G_i(y_t | y_{t-i}) \quad (\text{A.1})$$

Dans (A.1),  $F(y_t | \Omega_{t-1})$  est la fonction de distribution cumulative (*cdf*) conditionnelle de  $y_t$  donnant le passé. Et  $\sum_{i=1}^p \alpha_i = 1, \alpha_i \geq 0, i = 1, \dots, p$  et  $G_i(\cdot | y)$  est la *cdf* conditionnelle pour chaque  $i = 1, \dots, p$  et pour chaque valeur de  $y$ .  $G_i(\cdot | \cdot)$  doit être spécifiée par un paramètre  $\theta_i$ .

Le modèle *MTD* (A.1) a été introduit par Raftery (1985, a, b) dans le cas discret, comme étant un modèle de chaîne de Markov d'ordre élevé. Ensuite Martin et Raftery (1987) indiquèrent que le modèle *MTD* peut être défini pour des séries temporelles définies sur un espace arbitraires et ne restreigne pas pour des séries temporelles définies sur un ensemble de valeurs discrètes.

Le, et al (1996) présentèrent un cas spécial de modèles *MTD*, tel que les *cdf* conditionnelles des composantes de ce modèle sont Gaussiennes et inclut le modèle *AR*( $p$ ) comme un cas spécial. Ceci conduit à :

$$\begin{aligned}
F(y_t \mid \Omega_{t-1}) &= \alpha_0 \Phi \left( \frac{y_t - \sum_{j=1}^p \phi_{0j} y_{t-j}}{\sigma_0} \right) \\
&+ \sum_{i=1}^p \alpha_i \Phi \left( \frac{y_t - \phi_i y_{t-i}}{\sigma_i} \right)
\end{aligned} \tag{A.2}$$

Notons le modèle (A.2) par le modèle *GMTD*.

Le modèle *GMTD* peut contenir la situation où la série est assez approximée par un modèle *AR* et possède des valeurs aberrantes et explosives espacées. Dans ce cas la composante *AR* principale de la série peut être capturée par le premier terme du modèle (A.2) et les autres composantes additionnelles comme les valeurs aberrantes et explosives peuvent être capturées par les autres termes du modèle (A.2). Par exemple, les valeurs aberrantes peuvent être capturées par un terme dans le modèle ayant une variance  $\sigma_i^2$  élevée et un petit  $\alpha_i$ .

## A.2 Stationnarité

**Théorème 23** *Une condition nécessaire et suffisante pour que le processus  $y_t$  soit stationnaire dans la moyenne est que les racines  $z_1, \dots, z_p$  de l'équation :*

$$1 - \sum_{i=1}^p (\alpha_0 \phi_{0i} + \alpha_i \phi_i) z^{-i} = 0 \tag{A.3}$$

Sont toutes à l'intérieur du cercle unité.

**Théorème 24** *Supposons que le processus  $y_t$  définit par (A.1) et (A.2) tel que  $\alpha_0 = 0$ , est stationnaire au premier ordre. Une condition nécessaire et suffisante pour que le processus soit stationnaire au deuxième ordre est que les racines  $z_1, \dots, z_p$  de l'équation :*

$$1 - \sum_{i=1}^p \alpha_i \phi_i^2 z^{-i} = 0 \tag{A.4}$$

Sont toutes à l'intérieur du cercle unité.

Dans le cas où  $p = 2$ , cette condition donne une région de stationnarité définit par :

$$\alpha_1 \phi_1^2 + \alpha_2 \phi_2^2 < 1 \tag{A.5}$$

## A.3 La fonction d'autocorrélation

Nous obtenons maintenant les autocorrélations pour les modèles *GMTD*. Ces autocorrélations satisfont le système d'équations de Yule-Walker.

Supposons que le processus  $y_t$  est stationnaire au second ordre et nous assumons la moyenne est zéro. Soit  $\rho_l$  l'autocorrélation au  $l^{\text{ème}}$  retard. Alors

$$\begin{aligned} E(y_t y_{t-l}) &= E[E(y_t y_{t-l} \mid \Omega_{t-1})] \\ &= E\left(y_{t-l} \sum_{i=1}^p (\alpha_0 \phi_{0i} + \alpha_i \phi_i) y_{t-i}\right) \\ &= \sum_i (\alpha_0 \phi_{0i} + \alpha_i \phi_i) E(y_{t-l} y_{t-i}) \end{aligned} \quad (\text{A.6})$$

Puisque le processus est de second ordre, il s'ensuit que pour  $l = 1, \dots, p$

$$\rho_l = \sum_i (\alpha_0 \phi_{0i} + \alpha_i \phi_i) \rho_{|l-i|}$$

Ces équations sont similaires à celles de Yule-Walker pour un processus  $AR(p)$  ordinaire, où les coefficients  $\alpha_0 \phi_{0i} + \alpha_i \phi_i$  sont remplacés par les coefficients du  $i^{\text{ème}}$  retard du processus  $AR(p)$ .

## A.4 Estimation et sélection du modèle

L'algorithme *EM* est la méthode la plus populaire dans l'estimation des paramètres dans les modèles de mélange. Le, Martin, et Raftery (1996) ont prouvé à l'aide d'une analyse par simulation que l'algorithme *EM* a bien estimé les paramètres du modèle *GMTD* et qu'il est simple et facile à programmer.

Pour comparer entre les modèles concurrents pour modéliser notre série temporelle nous utilisons le critère de l'information de Bayes (*BIC*) défini par :

$$BIC = 2 \log(\text{vraisemblance maximisée}) - k \log(n),$$

où  $k$  est le nombre de paramètres indépendants estimés. C'est la première dérivation de Schwartz (1978).

## A.5 Exemples

Dans cette partie nous allons présenter quelques résultats obtenues par une étude faite par Le, Martin, et Raftery (1996), afin de comparer cette classe de modèles avec d'autres modèles ajustés par Box et Jenkins.

### A.5.1 Exemple 1

Le premier exemple s'agit de la série *B* rapportée par Box et Jenkins (1976), (*IBM stock prices*). Cette série contient 369 observations. Dans cet exemple la comparaison est entre les meilleurs modèles *ARIMA* et *GMTD* choisis.

Le meilleur modèle *GMTD* d'après le *BIC* est le suivant :

$$F(y_t \mid \Omega_{t-1}) = \alpha_0 \Phi \left( \frac{y_t - \phi y_{t-1} - (1 - \phi) y_{t-2}}{\sigma_0} \right) \\ + \alpha_1 \Phi \left( \frac{y_t - y_{t-1}}{\sigma_1} \right) + \alpha_2 \Phi \left( \frac{y_t - y_{t-2}}{\sigma_2} \right)$$

Les estimations des paramètres sont :

$$\phi = 1.94,$$

$$(\alpha_0, \alpha_1, \alpha_2) = (0.24, 0.69, 0.07),$$

et

$$(\sigma_0, \sigma_1, \sigma_2) = (6.38, 5.03, 11.23).$$

Le modèle *ARIMA* favorisé par Box-Jenkins est le modèle *ARIMA*(0, 1, 1)

La table donne les couvertures empiriques des intervalles de prévision aux différents taux de sécurité pour chaque modèle.

$(1 - \alpha) \%$	90	80	70	60	50
<i>GMTD</i>	91	81	71	62	51
<i>ARIMA</i>	90	84	78	73	56

Nous remarquons que les couvertures empiriques calculées par le modèle *GMTD* sont plus proches aux couvertures nominales par rapport au modèle *ARIMA*.

## A.5.2 Exemple 2

La deuxième série proposée est la série *D* rapportée par Box-Jenkins, il s'agit de viscosités lues chaque heure à partir d'un processus chimique. Dans cet exemple la comparaison est entre le modèle *GMTD* sélectionné dans le premier exemple avec les estimations des paramètres suivantes :

$$\phi = 1.06,$$

$$(\alpha_0, \alpha_1, \alpha_2) = (0.63, 0.25, 0.12),$$

et

$$(\sigma_0, \sigma_1, \sigma_2) = (0.1195, 0.0014, 0.0989).$$

Et le modèle *AR* (1) suivant, proposé par Box-Jenkins :

$$y_t = 0.87y_{t-1} + 1.17 + \varepsilon_t$$

tel que  $\sigma_\varepsilon^2 = 0.09$

La table donne les couvertures empiriques des intervalles de prévision aux différents taux de sécurité pour chaque modèle.

$(1 - \alpha) \%$	90	80	70	60	50
<i>GMTD</i>	91	80	67	56	48
<i>AR</i>	91	79	73	68	59

Nous remarquons que les couvertures empiriques calculées par le modèle *GMTD* sont plus proches aux couvertures nominales par rapport au modèle *AR*.



La série de températures, lues chaque minute, à partir d'un processus chimique (226 observations). Cette série est rapportée par Box, Jenkins, et Reinsel (1994) (série C)

26.6	20.9	21.3	24.4	22.8	21.7	24.4	23.2
27	20.3	21.2	24.4	22.6	21.8	24.2	23.1
27.1	19.7	21.4	24.4	22.4	21.9	24.2	22.9
27.1	19.4	21.7	24.4	22	22.2	24.1	22.8
27.1	19.3	22.2	24.5	21.6	22.5	24.1	22.6
27.1	19.2	23	24.5	21.3	22.8	24	22.4
26.9	19.1	23.8	24.4	21.2	23.1	24	22.2
26.8	19	24.6	24.3	21.2	23.4	24	21.8
26.7	18.9	25.1	24.2	21.1	23.8	23.9	21.3
26.4	18.9	25.6	24.2	21	24.1	23.8	20.8
26.	19.2	25.8	24	20.9	24.6	23.8	20.2
25.8	19.3	26.1	23.9	21	24.9	23.7	19.7
25.6	19.3	26.3	23.7	21	24.9	23.7	19.3
25.2	19.4	26.3	23.6	21.1	25.1	23.6	19.1
25	19	26.2	23.5	21.2	25	23.7	19
24.6	19.6	26	23.5	21.1	25	23.6	18.8
24.2	19.6	25.8	23.5	20.9	25	23.6	
24	19.6	25.6	23.5	20.8	25	23.6	
23.7	19.6	25.4	23.5	20.8	24.9	23.5	
23.4	19.6	25.2	23.7	20.8	24.8	23.5	
23.1	19.7	24.9	23.8	20.8	24.7	23.4	
22.9	19.9	24.7	23.8	20.9	24.6	23.3	
22.8	20	24.5	23.9	20.8	24.5	23.3	
22.7	20.1	24.4	23.9	20.8	24.5	23.3	
22.6	20.2	24.4	23.8	20.7	24.5	23.4	
22.4	20.3	24.4	23.7	20.7	24.5	23.4	
22.2	20.6	24.4	23.6	20.8	24.5	23.3	
22	21.6	24.4	23.4	20.9	24.5	23.2	
21.8	21.9	24.4	23.2	21.2	24.5	23.3	
21.4	21.7	24.4	23	21.4	24.4	23.3	

Dans cette annexe, nous allons présenter quelques programmes effectués pour avoir les résultats donnés dans les chapitres précédents. La programmation a été faite sous MATLAB 7.0

## Programme 1 :

Ce programme consiste à générer des données à partir d'un mélange de distributions gauusiennes.

```
function y = data_generation(N, alpha, mu, sigma)
alpha = alpha/sum(alpha);
alpha = alpha(:);
mu = mu(:);
sigma = sigma(:);
K = length(alpha); M(1) = alpha(1);
for k = 2 : K
M(k) = alpha(k) + M(k - 1);
end
for n = 1 : N
t = rand;
if (t <= M(1))
ind(n) = 1;
end
for k = 2 : K
if (t > M(k - 1))&(t <= M(k))
ind(n) = k;
end
end
end
ind = ind(:);
x = randn(N, 1);
y = mu(ind) + sqrt(sigma(ind)).* randn(N, 1)
```

**Programme 2 :**

Ce programme consiste à générer des données à partir d'un modèle MAR-ARCH :

```

function[y] = generdata2(N, p, q, alph, fi0, fi, bita0, bita)
K = length(alph);
M(1) = alph(1);
for k = 2 : K
M(k) = alph(k) + M(k - 1);
end
for n = 1 : N
t = rand;
if (t <= M(1))
ind(n) = 1;
end
for k = 2 : K
if (t > M(k - 1)) & (t <= M(k))
ind(n) = k;
end
end
end
y = zeros(1, N);
form = max(p(k)) + max(q(k)) : N
fork = 1 : K
z = zeros(K, N);
e = zeros(N, 1);
y(1 : max(p(k)) + max(q(k))) = rand(1, max(p(k)) + max(q(k)));
forn = max(p(k)) + max(q(k)) + 1 : N
forl = 1 : p(k)
s(n, l) = fi(k, l) * y(n - l);
end
e(n) = y(n) - fi0(k) - sum(s(n, :));
forj = 1 : q(k)
ss(n, j) = bita(k, j) * (e(n - j)^2);
end
h(n) = bita0(k) + sum(ss(n, :));
z(k, n) = fi0(k) + sum(s(n, :)) + sqrt(h(n)) * randn;
end
ifind(m) == k
y(m) = z(k, m);
end
end
end
end

```

**Programme 3 :**

Ce programme consiste à calculer les ordres d'un modèle MAR-ARCH qui minimisent la fonction BIC

```

functiontarget = biccal
clearall
clc
N = 500;
K = 2;
y = generdata2(N, [1 1], [1 1], [0.5 0.5], [11], [0.7 - 0.7], [1 1], [0.5 0.5]);
alph = [0.50.5];
c1 = [1 - 1];
c2 = [11];
fi = [0.711111; -0.711111];
bta = [0.511111; 0.511111];
p1 = 1 : 5;
q1 = 1 : 5;
p2 = 1 : 5;
q2 = 1 : 5;
pmax = max(p1, p2);
qmax = max(q1, q2);
for p1 = 1 : 5
for q1 = 1 : 5
for p2 = 1 : 5
for q2 = 1 : 5
f1 = alph(1) * foncdensity(y, N, p1, q1, c1(1), c2(1), fi(1, 1 : p1), bta(1, 1 : q1));
f2 = alph(2) * foncdensity(y, N, p2, q2, c1(2), c2(2), fi(2, 1 : p2), bta(2, 1 : q2));
for n = 1 : N
ff(n) = f1(n) + f2(n);
end
f(p1, q1, p2, q2) = sum(log(ff(pmax + qmax + 1 : N)));
BIC(p1, q1, p2, q2) = -2 * f(p1, q1, p2, q2) + log(N - max(p1, p2) - max(q1, q2)) *
(5 + p1 + p1 + q1 + q2);
end
end
end
end
min = 100000;
for ii = 1 : 5
for jj = 1 : 5
for kk = 1 : 5
for nn = 1 : 5
if BIC(ii, jj, kk, nn) < minmin = BIC(ii, jj, kk, nn);
target = [ii jj kk nn max(ii, kk) max(jj, nn)];
end
end
end
end

```

```

end
end
end
target

```

Ce programme nécessite le programme suivant :

#### Programme 4 :

```

function[fonc] = foncdensity(y, N, p, q, c1, c2, fi, bta)
e = zeros(N, 1);
y(1 : p + q) = rand(1, p + q);
for n = p + q + 1 : N
for k = 1 : p
s(n, k) = fi(k) * y(n - k);
end
e(n) = y(n) - c1 - sum(s(n, :));
for j = 1 : q
ss(n, k) = bta(j) * (e(n - j))^2;
end
h(n) = c2 + sum(ss(n, :));
if h(n) == 0
h(n) = eps;
end
fonc(n) = (1/sqrt(h(n))) * normpdf(e(n)/sqrt(h(n)));
end
for n = 1 : N
if fonc(n) == 0
fonc(n) = eps;
end
end
end

```

#### Programme 5 :

Dans ce programme nous répétons le programme 7 "s" fois afin de prouver que le BIC sélectionne bien le vrai ordre  $function[p1, q1, p2, q2] = replicbic(s)$

```

m = [];
for g = 1 : s
target = biccal;
m = [m; target];
end
p1 = length(find(m(:, 1) == 1));
q1 = length(find(m(:, 2) == 1));
p2 = length(find(m(:, 3) == 1));

```

```

q2 = length(find(m(:, 4) == 1));
pmax = length(find(m(:, 5) == 1));
qmax = length(find(m(:, 6) == 1));
mp = max(m(:, 5));
mq = max(m(:, 6));
p1
q1
p2
q2
pmax
qmax
mp
mq

```

### Programme 6 :

Ce programme estime les paramètres d'un mélange de distributions gaussiennes utilisant l'algorithme EM :

```

function[PP, MU, SIGMA, Q, L] = EM(data, K, P0, Mu0, Sigma0)
%data : are the data generated by the function data_generation.m
%K is the number of mixtures
%P0 initial value of the weights
%Mu0 initial value of the mean
%Sigma0 initial value of the variance
%A la sortie :
%PP a matrix (itmax+1)*K of points, take the last of line
%itmax is the max number of iterations; idem pour Mu et SIGMA
%L is the value of likelihood in each iteration
%Q is the value of likelihood with missing data in each iteration
if nargin == 2
hist(data, K) = [P, Mu];
P = P/sum(P);
Sigma(1 : K) = max(abs(diff(Mu)));
P = P(:)';
Mu = Mu(:)';
Sigma = Sigma(:)';
else
P = P0(:)';
Mu = Mu0(:)';
Sigma = Sigma0(:)';
end
N = length(data);
itmax = 100;
%MAXIMUM NUMBER OF ITERATION
z = zeros(N, K);

```

```

for it = 1 : itmax
%E-step -----
-----
for n = 1 : N
T = zeros(1, K);
for k = 1 : K
T(k) = P(it, k) * func(data(n), Mu(it, k), Sigma(it, k));
end
z(n, :) = T/sum(T);
end
%z(n,k) are probabilities that observation i is coming
%from the component j. But the probs are not normalized.
%The following cycle will compute the normalizing constant.
for n = 1 : N
clear H
clear h
H = z(n, :);
h = find(H == max(H));
zm(n) = h(1);
end
%-----
-----
%-----M-step-----
-----
P(it + 1, :) = sum(z)/N;
for k = 1 : K
Mu(it + 1, k) = sum(data .* z(:, k))/sum(z(:, k));
Sigma(it + 1, k) = sum(z(:, k) .* ((data - Mu(it + 1, k)).^2))/sum(z(:, k));
end
%-----
-----
ddQ = 1;
ddL = 1;
for k = n : N
ddQ = ddQ * func(data(n), Mu(it, zm(n)), Sigma(it, zm(n))) * P(it, zm(n));
count = 0;
for k = 1 : K
count = count + P(it, k) * func(data(n), Mu(it, k), Sigma(it, k));
end
ddL = count * ddL;
end
Q(it) = ddQ;
L(it) = ddL;
end
PP = P;
MU = Mu;

```

$SIGMA = Sigma;$

### Programme 7

Ce programme estime les paramètres d'un modèle MAR-ARCH à deux coposantes utilisant l'algorithme EM :

```

function v = EM
clear all
clc
K = 2;
N = 100;
y = generdata2(N, [1 1], [1 1], [0.5 0.5], [1 -1], [0.7; -0.7], [1 1], [0.5; 0.5]);
alph = [0.5 0.5];
phi0 = [1 -1];
phi1 = [0.7 -0.7];
bit0 = [1 1];
bit1 = [0.5 0.5];
z = zeros(N, K);
T = zeros(1 : K);
for it = 1 : 300
e = zeros(N, K);
h = zeros(N, K);
for n = 3 : N
for k = 1 : K
e(n, k) = y(n) - (phi0(it, k) + phi1(it, k) * y(n - 1));
h(n, k) = bit0(it, k) + bit1(it, k) * e(n - 1, k)^2;
if h(n, k) == 0
h(n, k) = eps;
end
T(k) = alph(it, k) * normpdf(e(n, k)/sqrt(h(n, k))) * (1/sqrt(h(n, k)));
end
z(n, :) = T/sum(T);
end
%----- Mstep-----
alph(it + 1, :) = sum(z)/N;
for n = 3 : N
for k = 1 : K
dhdp0(n, k) = -2 * bit1(it, k) * e(n - 1, k);
dhdp1(n, k) = -2 * bit1(it, k) * e(n - 1, k) * y(n - 2);
dedp0(n, k) = -1;
dedp1(n, k) = -y(n - 1);
end
end
for k = 1 : K
dldp0(k) = (1/N) * (sum(z(3 : N, k)) * dhdp0(3 : N, k) * ((e(3 : N, k).^2)./h(3 :

```



```

N, k)) - 1)./
(2. * h(3 : N, k))) - sum(z(3 : N, k). * e(3 : N, k). * dedp0(3 : N, k)./(h(3 : N, k))));
dldp1(k) = (1/N) * (sum(z(3 : N, k). * dhdp1(3 : N, k). * (((e(3 : N, k).2)/h(3 :
N, k)) - 1)./
(2. * h(3 : N, k))) - sum(z(3 : N, k). * e(3 : N, k). * dedp1(3 : N, k)./(h(3 : N, k))));
d2ld2p0(k) = (1/N) * (sum(z(3 : N, k). * (((dhdp0(3 : N, k).2)/(2. * (h(3 : N, k).2))) +
((dedp0(3 : N, k).2)/h(3 : N, k))));
d2ld2p1(k) = (1/N) * (sum(z(3 : N, k). * (((dhdp1(3 : N, k).2)/(2. * (h(3 : N, k).2))) +
((dedp1(3 : N, k).2)/h(3 : N, k))));
d2ldp1dp0(k) = (1/N) * (sum(z(3 : N, k). * ((dhdp0(3 : N, k). * dhdp1(3 :
N, k)./(2. * (h(3 : N, k).2))) +
((dedp0(3 : N, k). * dedp1(3 : N, k))./h(3 : N, k))));
end
for k = 1 : K
A = ([phi0(it, k) phi1(it, k)]' + inv([d2ld2p0(k) d2ldp1dp0(k); d2ldp1dp0(k) d2ld2p1(k)])) *

([dldp0(k) dldp1(k)]');
phi0(it + 1, k) = A(1);
phi1(it + 1, k) = A(2);
end
for n = 3 : N
for k = 1 : K
e1(n, k) = y(n) - (phi0(it + 1, k) + phi1(it + 1, k) * y(n - 1));
h1(n, k) = bit0(it, k) + bit1(it, k) * e(n - 1, k)2;
if h1(n, k) == 0
h1(n, k) = eps;
end
dhdb0(n, k) = 1;
dhdb1(n, k) = e1(n - 1, k)2;
end
end
for k = 1 : 2
dl db0(k) = (1/N) * (sum(z(3 : N, k). * dhdb0(3 : N, k). * (((e1(3 : N, k).2)/h1(3 :
N, k)) - 1)./(2 * h1(3 : N, k))));
dl db1(k) = (1/N) * (sum(z(3 : N, k). * dhdb1(3 : N, k). * (((e1(3 : N, k).2)/h1(3 :
N, k)) - 1)./(2 * h1(3 : N, k))));
d2ld2b0(k) = (1/N) * (sum(z(3 : N, k). * ((dhdb0(3 : N, k).2)/(2 * (h1(3 : N, k).2)))));
d2ld2b1(k) = (1/N) * (sum(z(3 : N, k). * ((dhdb1(3 : N, k).2)/(2 * (h1(3 : N, k).2)))));
d2ldb1db0(k) = (1/N) * (sum(z(3 : N, k). * ((dhdb0(3 : N, k). * dhdb1(3 :
N, k))./(2 * (h1(3 : N, k).2)))));
end
for k = 1 : K
B = ([bit0(it, k) bit1(it, k)]' + inv([d2ld2b0(k) d2ldb1db0(k); d2ldb1db0(k) d2ld2b1(k)])) *

([dl db0(k) dl db1(k)]');
bit0(it + 1, k) = B(1);

```

```

bit1(it + 1, k) = B(2);
end
fork = 1 : K
B = ([bit0(it, k)bit1(it, k)])' + (inv([d2ld2b0(k)d2ldb1db0(k); d2ldb1db0(k)d2ld2b1(k)])) *
([dlb0(k)dlb1(k)])' ;
bit0(it + 1, k) = B(1); bit1(it + 1, k) = B(2);
s(it + 1) = sum(((alph(it + 1, :)) * (bit1(it + 1, :)) * ((phi1(it + 1, 1)).^2))) - 2
* sum((alph(it + 1, :)) * (bit1(it + 1, :)) * (phi1(it + 1, :)))
* sum((alph(it + 1, :)) * (phi1(it + 1, :)));
ifabs(s(it + 1)) < 1
alph(it + 1, :) = alph(it, :) + (0.3) * (alph(it + 1, :) - alph(it, :));
phi1(it + 1, :) = phi1(it, :) + (0.3) * (phi1(it + 1, :) - phi1(it, :));
bit1(it + 1, :) = bit1(it, :) + (0.3) * (bit1(it + 1, :) - bit1(it, :));
end
end
end
alph;
phi0;
phi1;
bit0;
bit1;
v = [alph(end, :) phi0(end, :) phi1(end, :) bit0(end, :) bit1(end, :)];

```

## Programme 8

Ce programme estime les paramètres d'un modèle MAR-ARCH à trois coposantes utilisant l'algorithme EM :

```

function v = algoEM
clear all
clc
K = 3;
N = 500;
y = generdata2(N, [1 1 1], [1 2 1], [0.4 0.3 0.3], [1 0 -1], [0.9
; -0.7; 0.2], [1 1 1], [0.3 0; 0.2 0.4; 0.50]);
alph = [0.4 0.3 0.3];
phi0 = [1 0 -1];
phi1 = [0.9 -0.7 0.2];
bit0 = [1 1 1];
bit1 = [0.3 0.2 0.5];
bit2 = [0.4];
z = zeros(N, K);
T = zeros(1 : K);
for it = 1 : 100
e = zeros(N, K);
h = zeros(N, K);

```

```

%-----Estep-----
for n = 4 : N
for k = 1 : K
e(n, k) = y(n) - (phi0(it, k) + phi1(it, k) * y(n - 1));
end
h(n, 1) = bit0(it, 1) + bit1(it, 1) * e(n - 1, 1)^2;
h(n, 2) = bit0(it, 2) + bit1(it, 2) * e(n - 1, 2)^2 + bit2(it, 1) * e(n - 2, 2)^2;
h(n, 3) = bit0(it, 3) + bit1(it, 3) * e(n - 1, 3)^2;
end
for k = 1 : K
if h(n, k) == 0
h(n, k) = eps;
end
end
for n = 4 : N
for k = 1 : K
T(n, k) = alph(it, k) * normpdf(e(n, k)/sqrt(h(n, k))) * (1/sqrt(h(n, k)));
end
for k = 1 : K
z(n, k) = T(n, k)/sum(T(n, :));
end
end
%-----Mstep-----
alph(it + 1, :) = sum(z)/N;
for n = 4 : N
for k = 1 : K
if k == 2
dhdp0(n, k) = -2 * bit1(it, k) * e(n - 1, k) - 2 * bit2(it, 1) * e(n - 2, k);
else
dhdp0(n, k) = -2 * bit1(it, k) * e(n - 1, k);
end
if k == 2
dhdp1(n, k) = -2 * bit1(it, k) * e(n - 1, k) * y(n - 2) - 2 * bit2(it, 1) * e(n - 2, k) * y(n - 3);
else
dhdp1(n, k) = -2 * bit1(it, k) * e(n - 1, k) * y(n - 2);
end
dedp0(n, k) = -1;
dedp1(n, k) = -y(n - 1);
end
end
for k = 1 : K
lldp0(k) = (1/N) * (sum(z(4 : N, k)) * dhdp0(4 : N, k) * (((e(4 : N, k)).^2)/h(4 : N, k)) - 1) ./
(2 * h(4 : N, k))) - sum(z(4 : N, k) * e(4 : N, k) * dedp0(4 : N, k) ./ (h(4 : N, k)));
lldp1(k) = (1/N) * (sum(z(4 : N, k)) * dhdp1(4 : N, k) * (((e(4 : N, k)).^2)/h(4 : N, k)) - 1) ./

```

```

(2 * h(4 : N, k)) - sum(z(4 : N, k) * e(4 : N, k) * dedp1(4 : N, k) ./ (h(4 : N, k)));
d2ld2p0(k) = (1/N) * (sum(z(4 : N, k) * (((dhdp0(4 : N, k).^2) ./ (2 * (h(4 : N, k).^2))) +
((dedp0(4 : N, k).^2) ./ h(4 : N, k))));
d2ld2p1(k) = (1/N) * (sum(z(4 : N, k) * (((dhdp1(4 : N, k).^2) ./ (2 * (h(4 : N, k).^2))) +
((dedp1(4 : N, k).^2) ./ h(4 : N, k))));
d2ldp1dp0(k) = (1/N) * (sum(z(4 : N, k) * ((dhdp0(4 : N, k) * dhdp1(4 :
N, k) ./ (2 * (h(4 : N, k).^2))) +
((dedp0(4 : N, k) * dedp1(4 : N, k)) ./ h(4 : N, k))));
end
for k = 1 : K
A = ([phi0(it, k) phi1(it, k)]' + (inv([d2ld2p0(k) d2ldp1dp0(k); d2ldp1dp0(k) d2ld2p1(k)])) *
([dldp0(k) dldp1(k)]))';
phi0(it + 1, k) = A(1);
phi1(it + 1, k) = A(2);
end
for n = 4 : N
for k = 1 : K
e1(n, k) = y(n) - (phi0(it + 1, k) + phi1(it + 1, k) * y(n - 1));
if k == 2
h1(n, k) = bit0(it, k) + bit1(it, k) * e1(n - 1, k)^2 + bit2(it, 1) * e1(n - 2, k)^2;
else
h1(n, k) = bit0(it, k) + bit1(it, k) * e1(n - 1, k)^2;
end
if h1(n, k) == 0
h1(n, k) = eps;
end
dhdb0(n, k) = 1;
dhdb1(n, k) = e1(n - 1, k)^2;
end
dhdb2(n, 1) = e1(n - 2, 2)^2;
end
for k = 1 : K
dldb0(k) = (1/N) * (sum(z(4 : N, k) * dhdb0(4 : N, k) *
(((e1(4 : N, k).^2) ./ h1(4 : N, k)) - 1) ./ (2 * h1(4 : N, k))));
dldb1(k) = (1/N) * (sum(z(4 : N, k) * dhdb1(4 : N, k) * (((e1(4 : N, k).^2) ./ h1(4 :
N, k)) - 1) ./
(2 * h1(4 : N, k))));
d2ld2b0(k) = (1/N) * (sum(z(4 : N, k) * ((dhdb0(4 : N, k).^2) ./ (2 * (h1(4 : N, k).^2))));
d2ld2b1(k) = (1/N) * (sum(z(4 : N, k) * ((dhdb1(4 : N, k).^2) ./ (2 * (h1(4 : N, k).^2))));
d2ldb1db0(k) = (1/N) * (sum(z(4 : N, k) * ((dhdb0(4 : N, k) * dhdb1(4 :
N, k)) ./ (2 * (h1(4 : N, k).^2))));
end
dldb2 = (1/N) * (sum(z(4 : N, 2) * dhdb2(4 : N, 1) * (((e1(4 : N, 2).^2) ./ h1(4 :
N, 2)) - 1) ./ (2 * h1(4 : N, 2))));
d2ld2b2 = (1/N) * (sum(z(4 : N, 2) * ((dhdb2(4 : N, 1).^2) ./ (2 * (h1(4 : N, 2).^2))));

```

```

d2ldb1db2 = (1/N) * (sum(z(4 : N, 2) .* ((dhdb2(4 : N, 1) .* dhdb1(4 : N, 2))./(2 * (h1(4 :
N, 2).^2)))));
d2ldb0db2 = (1/N) * (sum(z(4 : N, 2) .* ((dhdb2(4 : N, 1) .* dhdb0(4 : N, 2))./(2 * (h1(4 :
N, 2).^2)))));
B1 = ([bit0(it, 1) bit1(it, 1)])' + (inv([d2ld2b0(1) d2ldb1db0(1); d2ldb1db0(1) d2ld2b1(1)])) *
([dlb0(1) dldb1(1)])' ;
bit0(it + 1, 1) = B1(1);
bit1(it + 1, 1) = B1(2);
B3 = ([bit0(it, 3) bit1(it, 3)])' + (inv([d2ld2b0(3) d2ldb1db0(3); d2ldb1db0(3) d2ld2b1(3)])) *
([dlb0(3) dldb1(3)])' ;
bit0(it + 1, 3) = B3(1);
bit1(it + 1, 3) = B3(2);
B2 = ([bit0(it, 2) bit1(it, 2) bit2(it, 1)])' +
(inv([d2ld2b0(2) d2ldb1db0(2) d2ldb0db2; d2ldb1db0(2) d2ld2b1(2) d2ldb1db2;
d2ldb0db2 d2ldb1db2 d2ld2b2])) * ([dlb0(2) dldb1(2) dldb2])' ;
bit0(it + 1, 2) = B2(1);
bit1(it + 1, 2) = B2(2);
bit2(it + 1, 1) = B2(3);
end
end
alph
phi0
phi1
bit0
bit1
bit2

```

Ces deux programmes nécessitent les deux programmes suivants :

### Programme 9

Ce programme exécute le programme 12 s fois :

```

function m = replic(s)
m = [];
for g = 1 : s
v = EM;
m = [m; v];
end
m;

```

**Programme 10**

Ce programme calcule les rmse des estimations obtenues par le programme 13 :

```
function R = rmse(s, tv)
m = replic(s);
M = mean(m);
V = var(m);
b = M - tv;
R = sqrt(V + b.^2);
M
```

**Programme 11**

Ce programme génère B échantillons bootstrap à partir d'une série de données suivant un modèle MAR-ARCH :

```
function [y] = bootsdata(B, N, p, q, alph, fi0, fi, bita0, bita)
K = length(alph);
M(1) = alph(1);
fork = 2 : K; M(k) = alph(k) + M(k - 1);
end
for n = 1 : N
t = rand;
if (t <= M(1))
ind(n) = 1;
end
fork = 2 : K
if (t > M(k - 1)) & (t <= M(k))
ind(n) = k;
end
end
end
[k, r] = bootstrp(B, 'mean', ind); bots = ind(r);
y = zeros(N, B);
for b = 1 : B
form = max(p(k)) + max(q(k)) : N
fork = 1 : K
z = zeros(K, N);
e = zeros(N, 1);
y(1 : max(p(k)) + max(q(k)), b) = rand(max(p(k)) + max(q(k)), 1);
for n = max(p(k)) + max(q(k)) + 1 : N
for l = 1 : p(k)
s(n, l) = fi(k, l) * y(n - l, b);
end
```

```

e(n) = y(n, b) - fi0(k) - sum(s(n, :));
for j = 1 : q(k)
ss(n, j) = bita(k, j) * (e(n - j)^2);
end
h(n) = bita0(k) + sum(ss(n, :));
z(k, n) = fi0(k) + sum(s(n, :)) + sqrt(h(n)) * randn;
end
if bots(m, b) == k
y(m, b) = z(k, m);
end
end
end
end
end

```

## Programme 12

Ce programme calcule la moyenne des estimations des paramètres d'un modèle MAR-ARCH utilisant les B échantillons de bootstrap :

```

function v = EMbootstrap
clearall
clc
K = 2;
B = 50;
N = 150;
[y] = bootsdata(2, 150, [1 1], [1 1], [0.5 0.5], [1 -1], [0.7; -0.7], [1 1], [0.5; 0.5]);
alph = [0.75 0.25];
phi0 = [0.75 -0.75];
phi1 = [0.5 -0.55];
bit0 = [0.75 0.75];
bit1 = [0.25 0.35];
z = zeros(N, K);
T = zeros(1 : K);
for b = 1 : B
for it = 1 : 100
e = zeros(N, K);
h = zeros(N, K);
for n = 3 : N
fork = 1 : K
e(n, k) = y(n, b) - (phi0(it, k) + phi1(it, k) * y(n - 1, b));
h(n, k) = bit0(it, k) + bit1(it, k) * e(n - 1, k)^2;
if h(n, k) == 0
h(n, k) = eps;
end
T(k) = alph(it, k) * normpdf(e(n, k)/sqrt(h(n, k))) * (1/sqrt(h(n, k)));
end
end
end
end
end

```

```

z(n,:) = T/sum(T);
end
alph(it + 1,:) = sum(z)/N;
for n = 3 : N
  for k = 1 : K
    dhdp0(n, k) = -2 * bit1(it, k) * e(n - 1, k);
    dhdp1(n, k) = -2 * bit1(it, k) * e(n - 1, k) * y(n - 2, b);
    dedp0(n, k) = -1;
    dedp1(n, k) = -y(n - 1, b);
  end
end
for k = 1 : K
  dldp0(k) = (1/N) * (sum(z(3 : N, k). * dhdp0(3 : N, k). * (((e(3 : N, k).2)/h(3 : N, k)) - 1)./(2. * h(3 : N, k))))
  - sum(z(3 : N, k). * e(3 : N, k). * dedp0(3 : N, k)./(h(3 : N, k))));
  dldp1(k) = (1/N) * (sum(z(3 : N, k). * dhdp1(3 : N, k). * (((e(3 : N, k).2)/h(3 : N, k)) - 1)./(2. * h(3 : N, k))))
  - sum(z(3 : N, k). * e(3 : N, k). * dedp1(3 : N, k)./(h(3 : N, k))));
  d2ld2p0(k) = (1/N) * (sum(z(3 : N, k). * (((dhdp0(3 : N, k).2)/(2. * (h(3 : N, k).2))) +
  ((dedp0(3 : N, k).2)/h(3 : N, k))));
  d2ld2p1(k) = (1/N) * (sum(z(3 : N, k). * (((dhdp1(3 : N, k).2)/(2. * (h(3 : N, k).2))) +
  ((dedp1(3 : N, k).2)/h(3 : N, k))));
  d2ldp1dp0(k) = (1/N) * (sum(z(3 : N, k). * ((dhdp0(3 : N, k). * dhdp1(3 : N, k)./(2. * (h(3 : N, k).2)))
  + ((dedp0(3 : N, k). * dedp1(3 : N, k))./h(3 : N, k))));
end
for k = 1 : K
  A = ([phi0(it, k) phi1(it, k)]' + (inv([d2ld2p0(k) d2ldp1dp0(k); d2ldp1dp0(k) d2ld2p1(k)]))
  * ([dldp0(k) dldp1(k)]))';
  phi0(it + 1, k) = A(1)
  phi1(it + 1, k) = A(2)
end
for n = 3 : N
  for k = 1 : K
    e1(n, k) = y(n, b) - (phi0(it + 1, k) + phi1(it + 1, k) * y(n - 1, b));
    h1(n, k) = bit0(it, k) + bit1(it, k) * e(n - 1, k)2;
    ifh1(n, k) == 0
    h1(n, k) = eps;
  end
end
dhdb0(n, k) = 1;
dhdb1(n, k) = e1(n - 1, k)2;
end
for k = 1 : 2
  dlhb0(k) = (1/N) * (sum(z(3 : N, k). * dhdb0(3 : N, k). * (((e1(3 : N, k).2)/h1(3 : N, k)) - 1)

```



```

./(2 * h1(3 : N, k)));
dlldb1(k) = (1/N) * (sum(z(3 : N, k). * dhdb1(3 : N, k). * (((e1(3 : N, k).^2)./h1(3 :
N, k)) - 1)
./(2 * h1(3 : N, k))));
d2ld2b0(k) = (1/N) * (sum(z(3 : N, k). * ((dhdb0(3 : N, k).^2)./(2 * (h1(3 : N, k).^2)))));
d2ld2b1(k) = (1/N) * (sum(z(3 : N, k). * ((dhdb1(3 : N, k).^2)./(2 * (h1(3 : N, k).^2)))));
d2ldb1db0(k) = (1/N) * (sum(z(3 : N, k). * ((dhdb0(3 : N, k). * dhdb1(3 :
N, k))./(2 * (h1(3 : N, k).^2)))));
end
fork = 1 : K
B = ([bit0(it, k) bit1(it, k)]' + inv([d2ld2b0(k) d2ldb1db0(k); d2ldb1db0(k) d2ld2b1(k)])
* ([dlldb0(k) dlldb1(k)]')';
bit0(it + 1, k) = B(1);
bit1(it + 1, k) = B(2);
end
end
alph;
phi0;
phi1;
bit0;
bit1;
v(b, :) = [alph(end, :) phi0(end, :) phi1(end, :) bit0(end, :) bit1(end, :)];
end
mean(v)

```

### Programme 13

Ce programme calcule les autocorrélations données par un modèle MAR-ARCH :

```

clear all
close all
clc
y = generdata2(N, p, q, alph, fi0, fi, bita0, bita);
r = covf(y', 25);
h = r/var(y);
stem(h)

```

### Programme 14

Ce programme calcule les variances conditionnelles données dans le chapitre 5 :

```

clear all
clc
y; %y est la série de données données dans l'annexe B;
alfa1 = 0.2738;

```

```
alfa2 = 0.7262;
fork = 4 : 226
u1(k) = 1.5377 * y(k - 1) - 0.5377 * y(k - 2);
u2(k) = 1.9966 * y(k - 1) - 0.9966 * y(k - 2);
h1(k) = 0.0037;
e2(k) = y(k) - 1.9966 * y(k - 1) + 0.9966 * y(k - 2);
h2(k) = 0.0102 + 0.4725 * (e2(k - 1))^2;
end
for i = 4 : 226
v1(i) = (alfa1 * u1(i)) + (alfa2 * u2(i));
v2(i) = (alfa1 * ((u1(i))^2)) + (alfa2 * (u2(i))^2);
v3(i) = (alfa1 * h1(i)) + (alfa2 * h2(i));
end
for i = 4 : 226
vola(i) = v3(i) + v2(i) - (v1(i))^2;
end
plot(vola)
```