



**HAL**  
open science

# Étude probabiliste et statistique de modèles conditionnellement hétéroscédastiques non linéaires.

Youssef Saidi

► **To cite this version:**

Youssef Saidi. Étude probabiliste et statistique de modèles conditionnellement hétéroscédastiques non linéaires.. Probabilités [math.PR]. Université de Lille 3, 2003. Français. NNT : . tel-01488734

**HAL Id: tel-01488734**

**<https://theses.hal.science/tel-01488734>**

Submitted on 13 Mar 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**UNIVERSITE LILLE 3 - CHARLES DE GAULLE**

UFR de Mathématiques, Sciences Économiques et Sociales

# **THÈSE**

pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITÉ LILLE 3**

Discipline:

**Mathématiques Appliquées aux Sciences Économiques**

présentée et soutenue publiquement par

**Youssef SAIDI**

le vendredi 12 décembre 2003

**Titre:**

**Étude probabiliste et statistique de modèles  
conditionnellement hétéroscédastiques non linéaires**

Sous la direction de

**Laurence BROZE et Jean-Michel ZAKOÏAN**

## **JURY**

Laurence BROZE	Professeur	Université Lille 3	Co-directrice de thèse
Michel CARBON	Professeur	Université Rennes 2	Rapporteur
Christian FRANCO	Professeur	Université Lille 3	Examinateur
Michael ROCKINGER	Professeur	université de Lausanne (suisse)	Rapporteur
J. Michel ZAKOÏAN	Professeur	Université Lille 3	Co-directeur de thèse

## Remerciements

Je voudrais tout d'abord remercier Laurence BROZE et Jean-Michel ZAKOÏAN d'avoir accepté la direction de ma thèse. Ils m'ont guidé durant celle-ci et m'ont apporté le soutien nécessaire. De part leurs qualités pédagogiques, leurs précieux conseils, leurs stimulants encouragements et leur disponibilité, j'ai pu mener à bien ce travail dans des conditions favorables.

Egalement, ils m'ont permis d'approfondir mes connaissances lors de différentes rencontres scientifiques en France et en Belgique (CORE) qu'ils ont organisées.

Bien évidemment, je remercie Michel CARBON et Michael ROCKINGER de l'honneur qu'ils me font d'être les rapporteurs de cette thèse, du temps et du soin qu'ils ont apporté à la rédaction de leurs rapports.

Mes remerciements vont aussi à Christian FRANCO d'avoir accepté la présidence du jury.

Je tiens à remercier Olivier TORRÈS, pour sa disponibilité, ses qualités humaines et avec qui j'ai eu des discussions fructueuses. Il m'a aidé et encouragé.

Mille mercis à Luc CHAMPARNAUD, Jérôme FONCEL, Frédéric JOUNEAU-SION, Ouafae BENRABAH, Elias OULD-SAID pour leurs encouragements et leur amitié.

Un grand merci à tous les membres de GREMARS de Lille 3 pour leur accueil chaleureux.

Je tiens à exprimer ma reconnaissance à toute ma famille, mes amis et tous ceux qui m'ont aidé de près et de loin.

Je dédie ce travail à mon père décédé, ma mère, mes filles, mes frères et mes sœurs.

## Résumé :

Nous étudions une classe de modèles conditionnellement hétéroscédastiques (ARCH) non linéaires. La volatilité de la variable à la date  $t$  dépend de la position relative des variables passées. Nous montrons qu'une famille de tels modèles admet une représentation Markovienne permettant d'en étudier la stabilité. A partir de critères de Lyapounov, nous établissons des conditions d'existence des moments. Les propriétés asymptotiques (convergence forte et en loi) de trois types d'estimateurs des paramètres sont établies. Ces résultats sont illustrés à distance finie à partir de méthodes simulées et sur des séries réelles.

**Mots clés :** modèles ARCH, modèles à seuils, séries temporelles non linéaires, stationnarité, maximum de vraisemblance, méthode des moments, critères d'érgodicité.

## abstract

In this thesis, we study a class of conditional heteroskedastic nonlinear models (ARCH). The volatility of the variable, at time  $t$ , depends on the relative position of the past variables. We show that such a family of models admits a Markovian representation which allows us to study the stability. From Lyapounov's criteria, we establish the conditions under which the moments exist. Furthermore, asymptotic properties (strong convergence and convergence in law) of three kinds of parameters estimators are given. These results are illustrated at finite distance from simulated and real data sets.

**Keywords :** ARCH models, threshold models, nonlinear time series, stationarity, maximum likelihood, method of moments, criteria for ergodicity.

# Table des matières

<b>0</b>	<b>Introduction générale</b>	<b>1</b>
<b>1</b>	<b>Stabilité des chaînes de Markov</b>	<b>7</b>
1.1	Noyau de transition et définition d'une chaîne de Markov . . . . .	9
1.2	Irréductibilité, apériodicité . . . . .	11
1.3	Chaînes récurrentes, chaînes transitoires . . . . .	16
1.4	Existence de mesures invariantes . . . . .	18
1.5	Ergodicité . . . . .	20
1.6	Critères de classification . . . . .	23
<b>2</b>	<b>Etude probabiliste</b>	<b>31</b>
2.1	Introduction . . . . .	31
2.2	Le modèle et ses propriétés élémentaires . . . . .	34
2.3	Stabilité du modèle déterministe . . . . .	41
2.4	Propriétés de chaîne de Markov . . . . .	45
2.5	Ergodicité . . . . .	52
2.6	Stationnarité au second-ordre . . . . .	55

2.7	Moments d'ordre supérieur . . . . .	62
<b>3</b>	<b>Inférence statistique</b>	<b>71</b>
3.1	Introduction . . . . .	71
3.2	Modèle et hypothèses . . . . .	72
3.3	Méthode des moindres carrés . . . . .	74
3.4	Méthode des moments . . . . .	84
3.5	Méthode du maximum de vraisemblance . . . . .	92
<b>4</b>	<b>Applications empiriques</b>	<b>111</b>
4.1	Test de l'existence de plusieurs régimes . . . . .	111
4.2	Étude par simulations . . . . .	115
4.2.1	Test . . . . .	115
4.2.2	Localisation du seuil $k$ et estimation des paramètres . . . . .	116
4.2.3	Méthodes d'estimations . . . . .	118
4.3	Application de la méthode des moments . . . . .	122
4.4	Résultats Empiriques : l'indice CAC 40 . . . . .	126
4.4.1	Test de changements de régime . . . . .	126
4.4.2	Estimation des paramètres . . . . .	127
4.5	Résultats Empiriques : l'indice S&P 500 . . . . .	132
4.5.1	Test de changements de régime . . . . .	133
4.5.2	Estimation des paramètres . . . . .	133
4.6	Annexe . . . . .	140

# Chapitre 0

## Introduction générale

Il est connu depuis longtemps que la modélisation des séries financières pose des problèmes spécifiques. L'analyse traditionnelle, ou méthodologie de Box et Jenkins (1976), essentiellement fondée sur les propriétés du second-ordre des séries, est inopérante car la plupart des séries financières (actions, indices, taux de change) ont, une fois différenciées, les propriétés au second-ordre d'un bruit blanc.

Dans un article paru en 1963, Mandelbrot mettait en évidence un certain nombre de caractéristiques communes à ces séries. Les principales propriétés (compatibles avec celle de bruit blanc) sont : (i) l'autocorrélation des carrés des variations de prix, (ii) les regroupements de variations de grand module (volatility clustering), (iii) la leptokurticité des distributions marginales, traduisant le grand nombre (relativement à une loi normale) de valeurs de grand module ou de module proche de zéro. A ces propriétés, étayées par de nombreuses études empiriques, il faut ajouter celle dite d'effet de levier, consistant en une asymétrie entre variations positives et négatives des prix quant à leur effet aux dates ultérieures. Les variations négatives (baisses) tendent à générer plus de turbulences que les hausses de même ampleur.

La traduction mathématique de ces observations passe par le concept, central en fi-

nance, de volatilité. Pour les économètres, la volatilité est définie comme la variance du processus observé conditionnellement à ses valeurs passées. En théorie financière, le terme de volatilité peut recouvrir des notions un peu différentes. Depuis une vingtaine d'années, de grands efforts ont été faits pour modéliser de manière adéquate la volatilité. La capacité des modèles à imiter les caractéristiques précédentes, la possibilité d'étudier les propriétés probabilistes de leurs solutions, l'existence de méthodes d'estimation permettant de les ajuster aux séries réelles sont autant de critères permettant de juger de leur pertinence.

Depuis l'article de Engle (1982) introduisant les modèles ARCH (autorégressifs conditionnellement hétéroscédastiques) de nombreuses spécifications de la volatilité ont été proposées. En parallèle, divers articles (voir par exemple Pagan et Hong (1990), Gallant et Tauchen (1989)) ont étudié l'apport des méthodes non paramétriques, ou semi-paramétriques, pour la modélisation de l'espérance conditionnelle du carré du processus.

Cette thèse est centrée sur une nouvelle classe de modèles conditionnellement hétéroscédastiques. Celle-ci s'apparente aux modèles à seuils introduits dans le cadre ARMA par Tong et Lim (1980) et dans le cadre GARCH par Nelson (1991), Zakoïan (1994). Dans ces modèles, la dynamique du processus  $X_t$  à la date présente dépend de la position d'une variable spécifique, notée  $\Delta_t$ . On peut donc voir cette dernière comme une variable de changement de régimes, ceux-ci consistant en les divers modèles possibles lorsque  $\Delta_t$  est dans telle ou telle région de l'espace. Un exemple simple est l'autorégressif d'ordre 1 à seuil défini par

$$X_t = a_i X_{t-1} + \epsilon_t, \quad \text{si } \Delta_t \in R_i, \quad i = 1, \dots, m$$

où  $(R_i)$  constitue une partition de la droite réelle,  $(\epsilon_t)$  est un bruit et  $\Delta_t$  est une

variable réelle, généralement supposée indépendante de  $\epsilon_t$ . Les régimes sont ici des AR(1) classiques. On distingue de grandes classes de tels processus, selon les hypothèses faites sur  $\Delta_t$ . On obtient ainsi les modèles dits à coefficients aléatoires si l'on suppose que  $(\Delta_t)$  est un processus indépendant. Les modèles à changement de régime markoviens sont obtenus en prenant pour  $(\Delta_t)$  une chaîne de Markov à espace discret. Les modèles de type SETAR (Self Exciting Threshold AR) font de  $\Delta_t$  une fonction du passé de  $X_t$  (par exemple  $\Delta_t = X_{t-1}$  pour le modèle le plus courant), permettant par exemple de modéliser des cycles asymétriques.

Dans le cadre des modèles GARCH les principaux modèles de type SETAR étudiés (voir références indiquées plus haut) font dépendre la volatilité à la date  $t$  du signe des innovations passées (la partition se limitant à  $\mathbb{R}^-$  et  $\mathbb{R}^+$ ). Cette spécification permet de prendre en compte l'effet de levier décrit précédemment. Une généralisation introduite par Liu, Li et Li (1997) spécifie la volatilité comme une fonction dépendant de la position des innovations passées. On peut ainsi, par exemple, obtenir une forme de volatilité différente dans les périodes turbulentes et dans les périodes calmes. La spécification étudiée dans cette thèse fait dépendre la volatilité de la position relative des innovations passées. Ce n'est donc pas le niveau des variations de prix qui gouverne les changements de régime mais le fait qu'une variation soit grande, ou petite, par rapport aux variations des dates précédentes. L'étude probabiliste de modèles de ce type est complexe et ne peut être menée à partir de représentations linéaires markoviennes, comme c'est le cas, à notre connaissance, pour toutes les classes de modèles GARCH introduites dans la littérature.

La suite de cette thèse comporte trois chapitres. Dans le chapitre 1, intitulé "Stabilité des chaînes de Markov à espace d'états continu", nous introduisons les principaux

outils et concepts probabilistes utilisés dans les chapitres suivants. Ce chapitre s'appuie principalement sur le livre de Meyn et Tweedie (1996). Le chapitre 2, intitulé "Etude probabiliste" présente la classe de modèles considérée et établit des conditions de stationnarité. Nous commençons par considérer le cas particulier du modèle déterministe. Nous étudions ensuite les propriétés de chaîne de Markov (irréductibilité, apériodicité, caractérisation des petits ensembles) du modèle écrit sous une forme vectorielle adéquate. Ceci permet d'établir, à partir de critères de type Foster-Lyapounov, une condition d'ergodicité géométrique dans la partie suivante. Nous étudions ensuite les propriétés de stationnarité au second-ordre du modèle. La dernière partie donne des conditions d'existence des moments d'ordre supérieur, illustrées dans les cas particuliers de lois uniforme et gaussienne. Le chapitre 3, intitulé "Inférence statistique", est essentiellement consacré à l'estimation de modèles ARCH plus généraux que ceux présentés dans le chapitre précédent. Divers estimateurs sont présentés (estimateur des moindres carrés, estimateur des moments, estimateur du maximum de vraisemblance) et leurs propriétés asymptotiques (convergence forte, normalité asymptotique) sont établies sous diverses hypothèses que nous discutons. Des expériences de Monte-Carlo permettent d'évaluer numériquement les variances asymptotiques obtenues pour diverses valeurs des paramètres. Le chapitre 4 intitulé "Applications empiriques" présente, dans une première partie, une méthode de test, inspirée de la méthode développée par Tsay (1989) pour les modèles autorégressifs à seuils. La deuxième partie met en œuvre les méthodes précédentes et obtient des résultats encourageants quant à la pertinence de la modélisation proposée.

# Bibliographie

- [1] **Box, G. E. P. et Jenkins, G. M. (1976)**, *Time Series Analysis : Forecasting and Control*, (revised edn), Holden-Day, San Francisco.
- [2] **Engle, R.F. (1982)**, *Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation*, *Econometrica*, 50, 987-1008.
- [3] **Gallant, R., et G. Tauchen (1989)**, *Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes : Asset Pricing Applications*, *Econometrica*, 57, 1091-1120.
- [4] **Liu J., Li W.K. et C.W. Li (1997)**, *On a Threshold Autoregression with Conditional Heteroskedastic Variances*, *Journal of Statistical Planning and Inference*, 62, 279-300.
- [5] **Mandelbrot, B. (1963)**, *The Variations of Certain Speculative Prices*, *Journal of Business*, 36, 394-419.
- [6] **Meyn et R. L. Tweedie (1996)**, *Markov Chains and Stochastic Stability*, Springer-Verlag, 3rd printing, London.
- [7] **Nelson D.B. (1991)**, *Conditional Heteroskedasticity in Asset Returns : a New Approach*, *Econometrica*, 59, 347-370.

- [8] **Pagan, A., et Y. Hong (1990)**, *Nonparametric Estimation and the Risk Premium*, dans *Semiparametric and Nonparametric Methods in Econometrics and Statistics*, ed. par W. Barnett, J. Powel, et G. Tauchen. Cambridge University Press.
- [9] **Tong, H., et K.S. Lim (1980)**, *Threshold Autoregression, Limit Cycles and Cyclical Data*, *Journal of the Royal Statistical Society B*, 42, 245-292.
- [10] **Tsay R.S (1989)**, *Testing and Modeling Threshold Autoregressive Processes*, *Journal of American Statistical Association*, 84, no. 405, 231-240.
- [11] **Zakoïan, J.M. (1994)**, *Threshold Heteroskedastic Models*, *Journal of Economic Dynamics and Control*, 18, 931-955.

# Chapitre 1

## Stabilité des chaînes de Markov à espace d'états continu

L'étude des modèles de séries temporelles comporte généralement une partie cruciale liée aux conditions d'existence de solutions stationnaires. Il n'est pas toujours simple d'obtenir des conditions suffisantes (portant sur les paramètres, la loi des erreurs ...) pour qu'un modèle admette des solutions stationnaires (au sens strict ou au second ordre). Dans le cas des modèles ARMA (Autoregressive Moving Average) univariés, les conditions portent sur les racines des polynômes retard (autorégressif et moyenne mobile) et ces conditions sont les mêmes pour les deux types de stationnarité (voir par exemple Brockwell et Davis (1991) ou Gouriéroux et Monfort (1989)). Pour d'autres modèles, il existe des représentations Markoviennes linéaires, de la forme

$$X_n = A_n X_{n-1} + B_n,$$

où les matrices  $A_n$  et  $B_n$  sont aléatoires mais constituent une suite  $(A_n, B_n)$  indépendante et identiquement distribuée (i.i.d.) permettant de relier l'étude de la stationnarité au comportement de produits de matrices aléatoires. De nombreux modèles de type GARCH ont ainsi été étudiés, à partir des travaux de Brandt (1986) et Bougerol et Picard (1992a, 1992b).

D'autres modèles conduisent à des représentations markoviennes de la forme :

$$X_n = F(X_{n-1}, \epsilon_n)$$

où  $(\epsilon_n)$  est une suite de vecteurs aléatoires i.i.d. et  $F$  est une application non linéaire. Une approche du problème de la stationnarité pour ce type de modèles, est fondée sur la théorie des chaînes de Markov à espace d'états continu, exposée dans le livre de Meyn et Tweedie (1996). Cette approche est par exemple utilisée lorsque la spécification du modèle comporte des effets de seuils excluant l'existence de représentations linéaires du type précédent.

L'objet de ce chapitre est d'introduire les principaux outils et concepts de cette théorie, que nous utiliserons dans les chapitres suivants.

Les propriétés des chaînes de Markov à temps discret et espace d'états dénombrable sont bien connues et elles figurent dans de nombreux manuels (voir par exemple Meyn et Tweedie (1996)). En particulier, les états d'une chaîne de Markov homogène sur un espace dénombrable sont transitoires ou récurrents, selon que la probabilité que le temps de retour dans cet état soit fini est strictement inférieur à 1 ou non. Chaque état récurrent peut ensuite être classé comme étant "nul" ou "positif" : on dira qu'un état récurrent est nul si la probabilité de retour dans cet état en  $n$  étapes tend vers zéro lorsque  $n$  tend vers l'infini, et positif sinon. Enfin, si la chaîne est irréductible, ce qui de manière heuristique signifie que tous les états communiquent

avec probabilité positive, tous les états ont la même classification.

Même si les notions ne coïncident pas toujours, une classification similaire existe pour les chaînes de Markov à espace d'états général. Nous renvoyons aux ouvrages de Duflo (1990), Meyn et Tweedie (1996), Numelin (1984) pour une étude détaillée.

## 1.1 Noyau de transition et définition d'une chaîne de Markov

La définition suivante introduit la notion de noyau de transition sur un ensemble quelconque  $\mathcal{X}$  muni d'une tribu  $\mathcal{B}$ .

**Définition 1.1.1** *On appelle noyau de transition toute famille  $\mathcal{P} = \{P(x, B), x \in \mathcal{X}, B \in \mathcal{B}\}$  telle que :*

1. *pour tout  $B \in \mathcal{B}$ ,  $x \rightarrow P(x, B)$  est une fonction mesurable de  $(\mathcal{X}, \mathcal{B})$  dans  $([0, 1], \mathcal{B}([0, 1]))$ ;*
2. *pour tout  $x \in \mathcal{X}$ ,  $P(x, \cdot)$  est une mesure de probabilité sur  $(\mathcal{X}, \mathcal{B})$ .*

*Les  $P(x, B)$  sont appelées probabilités de transition (de  $x$  vers  $B$ ).*

Etant donnée une mesure de probabilité initiale  $\mu$  sur  $(\mathcal{X}, \mathcal{B})$ , la définition d'une chaîne de Markov homogène  $(X_n)$  sur  $(\mathcal{X}, \mathcal{B})$  repose sur ses distributions fini dimensionnelles. Pour tout entier  $n$  et tout  $(n + 1)$ -uplet  $(B_0, \dots, B_n)$  de parties de  $\mathcal{B}$ , on pose :

$$\mathbb{P}_\mu[X_0 \in B_0, \dots, X_n \in B_n] = \int_{x_0 \in B_0} \dots \int_{x_{n-1} \in B_{n-1}} \mu(dx_0) P(x_0, dx_1) \dots P(x_{n-1}, B_n) \quad (1.1)$$

Pour le passage des lois fini-dimensionnelles aux probabilités d'évènements impliquant un nombre quelconque de variables  $X_n$ , le théorème de Kolmogorov impose des conditions à l'espace d'états : il s'applique si  $(\mathcal{X}, \mathcal{B})$  est : i)  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ , ou ii) un espace dénombrable muni de la tribu de toutes ses parties, ou iii) un espace métrique complet possédant une base dénombrable d'ouverts muni de sa tribu borélienne. Nous nous placerons toujours dans l'un de ces cas par la suite.

A condition de choisir comme espace  $\Omega$  l'espace  $\mathcal{X}^{\mathbb{N}}$ , muni de la tribu produit  $\mathcal{B}^{\otimes \mathbb{N}}$ , il existe donc un processus  $X = (X_n)_{n \in \mathbb{N}}$  et une probabilité  $\mathbb{P}_\mu$  sur  $\mathcal{B}^{\otimes \mathbb{N}}$  telle que (1.1) soit vraie et pour tout  $B \in \mathcal{B}^{\mathbb{N}}$ ,  $\mathbb{P}_\mu(B)$  soit la probabilité de l'évènement  $[X \in B]$ . Remarquons en particulier que, si  $\{x\} \in \mathcal{B}$  et  $\mu(\{x\}) \neq 0$ ,  $P(x, B) = \mathbb{P}_\mu[X_1 \in B \mid X_0 = x]$ .

**Définition 1.1.2** *Le processus  $X = (X_n)$  dont les lois fini-dimensionnelles satisfont (1.1) est appelé chaîne de Markov homogène de noyau de transition  $\mathcal{P}$  et de mesure initiale  $\mu$ .*

Etant donnés deux noyaux de transition  $\mathcal{P} = \{P(x, B)\}$  et  $\mathcal{Q} = \{Q(x, B)\}$  sur  $(\mathcal{X}, \mathcal{B})$ , il est utile d'introduire leur produit  $\mathcal{P}\mathcal{Q} = \{[PQ](x, B)\}$  défini par :

$$[PQ](x, B) = \int_{y \in \mathcal{X}} P(x, dy)Q(y, B), \quad \forall x \in \mathcal{X}, \forall B \in \mathcal{B}$$

On vérifie facilement que cette égalité définit un nouveau noyau de transition sur  $(\mathcal{X}, \mathcal{B})$ . En particulier, on notera  $\mathcal{P}^n$  le noyau de transition obtenu en effectuant  $(n - 1)$  produits de  $\mathcal{P}$  avec lui même. On a pour tout  $n \geq 1$  :  $\forall x \in \mathcal{X}, \forall B \in \mathcal{B}$ ,  $P^n(x, B) = \int_{y \in \mathcal{X}} P(x, dy)P^{n-1}(y, B)$ , en prenant pour  $P^0(y, B)$  la mesure de Dirac en  $y$  ( $P^0(y, B) = 1$  si  $y \in B$ ,  $P^0(y, B) = 0$  sinon). On vérifie que  $P^n(x, B)$  s'interprète comme la probabilité de l'évènement  $X_n \in B$  conditionnelle à  $X_0 = x$ .

Une généralisation utile est la relation de Chapman-Kolmogorov :

$$\forall n \in \mathbb{N}, \forall m, 0 \leq m \leq n, \forall x \in \mathcal{X}, \forall B \in \mathcal{B}$$

$$P^n(x, B) = \int_{y \in \mathcal{X}} P^m(x, dy) P^{n-m}(y, B)$$

On montre cette relation à partir de (1.1) en choisissant  $B_i = \mathcal{X}$  pour  $1 \leq i \leq n-1$  et  $B_0 = \{x\}$ . Cette équation admet une interprétation intéressante : pour atteindre  $B$  en  $n$  étapes, partant de  $x$ , la chaîne doit nécessairement passer par une valeur  $y$  en  $m$  étapes, puis tout se passe comme si la chaîne démarrait en  $y$ , pour atteindre  $B$  en  $n-m$  étapes. Les valeurs prises antérieurement en  $y$  n'ont donc plus d'importance pour ce qui est postérieur à l'étape  $m$ . Plus généralement, la propriété suivante traduit l'indépendance entre futur et passé conditionnellement au présent :

**Propriété 1.1.1** *Si  $X = (X_n)_{n \in \mathbb{N}}$  est une chaîne de Markov de mesure initiale  $\mu$ , pour toute fonction mesurable bornée  $h : \Omega \rightarrow \mathbb{R}$ , on a*

$$E[h(X_{n+1}, X_{n+2}, \dots) \mid X_0, \dots, X_n; X_n = x] = E[h(X_1, X_2, \dots) \mid X_0 = x]$$

[voir Meyn et Tweedie(1996), Proposition 3.4.3]

## 1.2 Irréductibilité, apériodicité

L'irréductibilité traduit le fait que, partant de n'importe quel point, la chaîne peut atteindre toute partie de mesure positive de l'espace d'états, pour une mesure appropriée. Au préalable, définissons pour tout  $B \in \mathcal{B}$ ,

– le temps de séjour dans  $B$  (après la date initiale) :

$$\eta_B = \sum_{n=1}^{\infty} \mathbb{1}_{X_n \in B}$$

- où  $\mathbb{1}_A$  désigne la variable indicatrice d'un évènement  $A$  ;  
 – le temps de premier passage dans  $B$  (hormis l'état initial) :

$$\tau_B = \min\{n \geq 1, X_n \in B\}.$$

La durée moyenne de séjour dans  $B$  partant de  $x \in \mathcal{X}$  s'obtient comme :

$$E[\eta_B \mid X_0 = x] = \sum_{n=1}^{\infty} P^n[x, B] = U(x, B)$$

La famille  $\{U(x, B)\}$  a des propriétés similaires à celles d'un noyau de transition mais les  $U(x, \cdot)$  ne sont pas des probabilités. On utilisera également la probabilité de passage dans  $B$  partant de  $x$  :

$$L(x, B) = \mathbb{P}(\tau_B < \infty \mid X_0 = x)$$

Nous pouvons maintenant définir le concept de  $\phi$ -irréductibilité.

**Définition 1.2.1** Soit  $\phi$  une mesure non identiquement nulle définie sur  $(\mathcal{X}, \mathcal{B})$ .  
 On dit que  $\phi$  est une mesure d'irréductibilité si

$$\forall x \in \mathcal{X}, \forall B \in \mathcal{B}, \quad \phi(B) > 0 \implies L(x, B) > 0$$

On dit alors que la chaîne  $X = (X_n)$  est  $\phi$ -irréductible.

On vérifie facilement qu'une définition équivalente de l'irréductibilité est obtenue en remplaçant  $L(x, B)$  par  $U(x, B)$ . Le choix de la mesure d'irréductibilité conditionnant la classification de la chaîne, comme nous le verrons ensuite, il importe de choisir une mesure de support le plus large possible. On définit ainsi une mesure d'irréductibilité "maximale", dont l'existence est assurée par la propriété suivante :

**Propriété 1.2.1** *Si  $(X_n)_{n \in \mathbb{N}}$  est  $\varphi$ -irréductible, il existe une mesure  $\psi$  sur  $(\mathcal{X}, \mathcal{B})$  telle que :*

1.  $(X_n)_{n \in \mathbb{N}}$  est  $\psi$ -irréductible ;
2. Pour toute autre mesure  $\varphi'$  sur  $(\mathcal{X}, \mathcal{B})$ ,  $(X_n)$  est  $\varphi'$ -irréductible si et seulement si  $\varphi'$  est absolument continue par rapport à  $\psi$  ( $\psi \succ \varphi'$ ) ;
3. Si  $\psi(B) = 0$ , alors  $\psi(\bar{B}) = 0$  où  $\bar{B} = \{x; L(x, B) > 0\}$ .

La mesure  $\psi$  est alors appelée mesure d'irréductibilité maximale.

[voir Meyn et Tweedie (1996), Proposition 4.2.2]

Nous noterons, pour une chaîne de mesure d'irréductibilité maximale  $\psi$  :

$$\mathcal{B}^+ = \{B \in \mathcal{B} \ ; \ \psi(B) > 0\}.$$

D'après la propriété précédente, cet ensemble ne dépend pas du choix de  $\psi$ , car toutes les mesures d'irréductibilité maximales sont équivalentes.

**Exemple :** Etant donnée une suite  $(\varepsilon_t)_{t \in \mathbb{N}}$  de variables i.i.d centrées, soit  $(X_t)_{t \in \mathbb{N}}$  le processus AR(1) défini par

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

où  $\theta \in \mathbb{R}$ . Si l'on se donne une mesure initiale, la suite  $(X_t)_{t \in \mathbb{N}}$  constitue clairement une chaîne de Markov sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , de probabilités de transition définies par :  $P(x, B) = P_\varepsilon(B - \theta x)$ , où  $P_\varepsilon$  désigne la loi des  $\varepsilon_t$ . Si cette loi est absolument continue par rapport à la mesure de Lebesgue  $\lambda$  sur  $\mathbb{R}$  et admet une densité partout positive sur  $\mathbb{R}$ , il en est de même de  $P(x, \cdot)$  pour tout  $x \in \mathcal{X}$ . Donc toute mesure  $\varphi$

absolument continue par rapport à  $\lambda$  est une mesure d'irréductibilité. De plus  $\lambda$  est une mesure d'irréductibilité maximale.

Si, par contre, la densité des  $\varepsilon_t$  n'est pas partout positive sur  $\mathbb{R}$  l'existence d'une mesure d'irréductibilité n'est pas garantie. Supposons par exemple que les  $\varepsilon_t$  suivent une loi uniforme sur  $[-1, 1]$ . Si  $|\theta| \leq 1$ , on voit facilement que la chaîne est  $\lambda_{[-1,1]}$ -irréductible, où  $\lambda_{[-1,1]}$  désigne la restriction de la mesure de Lebesgue à  $[-1, 1]$ . Mais si par exemple  $\theta > 1$  et  $X_0 = x_0 > \frac{1}{\theta-1}$  on a :  $x_0 < X_1 < X_2 \dots$ , quels que soient les  $\varepsilon_t$ . Donc la chaîne est explosive et il ne peut exister de mesure d'irréductibilité : une telle mesure devrait vérifier  $\varphi(]-\infty, x_0]) = 0, \forall x_0 \in \mathbb{R}$ , ce qui conduirait à  $\varphi = 0$ .

L'existence d'une mesure d'irréductibilité est compatible avec un comportement périodique de la chaîne. Afin d'exclure cette possibilité, définissons précisément un concept de périodicité pour les chaînes à espace d'états quelconque. Dans le cas où  $\mathcal{X}$  est dénombrable et la chaîne irréductible (tout les états communiquent), il suffit que  $P(x, x) > 0$  pour un état  $x$  pour que la chaîne soit apériodique : on parle alors de chaîne fortement apériodique. La définition est moins immédiate pour un espace d'états quelconque.

Nous commençons par définir les notions de "small set" et "petite set".

**Définition 1.2.2** On appelle  $\nu_m$ -small set tout  $C \in \mathcal{B}$  tel qu'il existe  $m > 0$  et une mesure non triviale  $\nu_m$  sur  $\mathcal{B}$  tels que :

$$P^m(x, B) \geq \nu_m(B), \quad \forall x \in C, \forall B \in \mathcal{B}.$$

On appelle  $\nu_a$ -petite set tout  $C \in \mathcal{B}$  tel que :

$$K_a(x, B) := \sum_{m=0}^{\infty} a_m P^m(x, B) \geq \nu_a(B), \quad \forall x \in C, \forall B \in \mathcal{B}$$

où  $a = (a_m)$  est une suite de constantes positives de somme égale à 1 et  $\nu_a$  est une mesure non triviale sur  $\mathcal{B}$ .

**Remarque.** On montre que (i) si  $C$  est  $\nu_a$ -petite, il existe  $(b_m)$  telle que  $C$  soit  $\nu_b$ -petite, avec  $\nu_b$  mesure maximale d'irréductibilité; (ii) l'union de deux petite (resp. small) sets est également un petite (resp. small) set (Meyn et Tweedie (1996), Proposition 5.5.5.).

Soit un  $\nu_M$ -small set  $C$ , où la mesure  $\nu_M := \nu$  est telle que  $\nu(C) > 0$ . Une telle mesure existe toujours si  $C \in \mathcal{B}^+$  (voir Meyn et Tweedie (1996), Proposition 5.2.4.). On considère l'ensemble des entiers  $m$  tels que  $C$  est  $\nu_m$ -small pour une mesure  $\nu_m$  proportionnelle à  $\nu$  :

$$E_C = \{m \geq 1 \mid C \text{ est } \nu_m\text{-small, avec } \nu_m = \delta_m \nu, \text{ pour un } \delta_m > 0\}.$$

**Définition 1.2.3** Soit  $(X_n)_{n \in \mathbb{N}}$  une chaîne  $\varphi$ -irréductible. Soit  $C \in \mathcal{B}^+$ . Alors  $d = \text{p.g.c.d.}(E_C)$  ne dépend pas du choix de  $C$  et est appelé période de la chaîne. Si  $d = 1$ ,  $(X_n)$  est dite apériodique. Si  $1 \in E_C$ ,  $(X_n)$  est dite fortement apériodique.

Notons en particulier que dans le cas fortement apériodique,  $P(x, C) > 0, \forall x \in C$ , ce qui fait le lien avec la définition dans le cas dénombrable.

On a le résultat suivant :

**Théorème 1.2.1** Pour une chaîne apériodique, les petite sets sont des small sets.

[voir Meyn et Tweedie(1996), Théorème 5.5.7.]

Pour les chaînes irréductibles et fortement apériodique, une première classification est fondée sur la notion de récurrence.

### 1.3 Chaînes récurrentes, chaînes transitoires

Lorsque  $\mathcal{X}$  est dénombrable et la chaîne  $(X_n)$  irréductible, un état est dit transitoire si le temps de séjour moyen dans cet état, partant de ce même état, est fini et il est dit récurrent dans le cas contraire. De plus tous les états sont de même nature (récurrents ou transitoires) et la chaîne peut être qualifiée de récurrente ou transitoire.

On retrouve la même dichotomie pour un espace d'états quelconque, mais les définitions sont plus délicates. La classification repose sur le (pseudo) noyau de transition  $\{U(x, B)\}$ ,  $U(x, B)$  représentant le temps de séjour moyen dans  $B$  partant de  $x$ .

**Définition 1.3.1** *L'ensemble  $B \in \mathcal{B}$  est dit*

1. *récurrent si  $U(x, B) = +\infty, \forall x \in B$  ;*
2. *uniformément transitoire si  $\exists M < \infty$  t.q  $U(x, B) < M, \forall x \in B$  ;*
3. *transitoire s'il existe une famille dénombrable  $(B_n)$  d'ensembles uniformément transitoires telle que  $B = \bigcup_n B_n$ .*

Avec cette définition, un ensemble récurrent peut également être transitoire. Ainsi  $\mathcal{X}$  est toujours récurrent et s'il est dénombrable, on peut évidemment l'écrire comme la réunion de tous les singletons. Si  $(X_n)$  est transitoire, tous les états sont transitoires donc les singletons sont uniformément transitoires et  $\mathcal{X}$  est transitoire. Le résultat suivant fournit une première dichotomie de l'ensemble des chaînes irréductibles.

**Définition 1.3.2** On note  $\mathcal{R}$  la classe des chaînes récurrentes, définies par :

$$(X_n) \in \mathcal{R} \iff U(x, B) = +\infty, \quad \forall x \in \mathcal{X}, \quad \forall B \in \mathcal{B}^+$$

On note  $\mathcal{T}$  la classe des chaînes transitoires, définies par :

$$(X_n) \in \mathcal{T} \iff \exists (B_j)_j, \quad \mathcal{X} = \bigcup_j B_j, \quad U(x, B_j) \leq M_j < \infty, \quad \forall x \in \mathcal{X}.$$

**Théorème 1.3.1** Si  $(X_n)$  est une chaîne  $\varphi$ -irréductible, elle est soit récurrente soit transitoire.

[voir Meyn et Tweedie (1996), Théorème 8.3.4]

Une notion de récurrence un peu plus forte, mais permettant d'obtenir des résultats également plus forts, est celle de Harris-récurrence. Au lieu d'être fondée sur l'espérance du temps de séjour dans  $B$ , la définition repose sur la probabilité que ce temps de séjour soit infini.

**Définition 1.3.3** Une chaîne  $(X_n)$   $\varphi$ -irréductible est dite Harris-récurrente si

$$\forall x \in \mathcal{X}, \quad \forall B \in \mathcal{B}^+, \quad Q(x, B) = \mathbb{P}[\eta_B = +\infty \mid X_0 = x] = 1$$

Autrement dit, partant d'un point quelconque, la chaîne visite infiniment souvent tout ensemble de  $\varphi$ -mesure positive. On peut de manière équivalente (voir Meyn et Tweedie (1996), Proposition 9.1.1) définir la Harris-récurrence par :

$$\forall x \in \mathcal{X}, \quad \forall B \in \mathcal{B}^+, \quad L(x, B) = \mathbb{P}[\tau_B < \infty \mid X_0 = x] = 1.$$

Si  $(X_n)$  est récurrente elle n'est pas nécessairement Harris-récurrente mais il existe un ensemble absorbant  $H$  sur laquelle la restriction de la chaîne est Harris-récurrente,

et tel que  $\varphi(H^c) = 0$ . Le critère suivant montre l'importance des petite sets (voir Meyn et Tweedie (1996), Théorème 8.3.6).

**Théorème 1.3.2** *Si  $(X_n)$  est une chaîne  $\varphi$ -irréductible. S'il existe un petite set  $C$  tel que  $\forall x \in \mathcal{X}$ ,  $L(x, C) = 1$  alors  $(X_n)$  est Harris-récurrente.*

## 1.4 Existence de mesures invariantes

Afin d'obtenir des propriétés de stabilité plus fortes, on est amené à pousser plus loin la classification des chaînes récurrentes. La stabilité maximale pour une chaîne de Markov consiste en ce que les variables  $X_n$  soient toutes de même loi. Ceci nous amène à introduire les notions de mesure invariante et sous-invariante

**Définition 1.4.1** *Une mesure non nulle  $\sigma$ -finie  $\pi$  sur  $\mathcal{B}$  telle que*

$$\forall B \in \mathcal{B}, \quad \pi(B) = \int P(x, B)\pi(dx) \quad [\text{resp.} \quad \pi(B) \geq \int P(x, B)\pi(dx)]$$

*est dite invariante [resp. sous-invariante]. S'il existe une mesure de probabilité invariante, la chaîne est dite positive.*

Pour une chaîne récurrente, une mesure invariante peut ou non exister. Lorsqu'une telle mesure existe et est finie, on peut toujours la normaliser pour obtenir une mesure de probabilité invariante. Remarquons qu'étant donnée une mesure de probabilité initiale  $\mu$ , on a

$$\int P(x, B)\mu(dx) = \mathbb{P}_\mu[X_1 \in B].$$

Donc si  $\mu$  est invariante :

$$\mu(B) = \mathbb{P}_\mu[X_1 \in B] = \int \left[ \int P(y, dx)\mu(dy) \right] P(x, B)$$

$$\begin{aligned}
&= \int \mu(dy) \int P(y, dx) P(x, B) \\
&= \int \mu(dy) P^2(y, B) \\
&= \mathbb{P}_\mu[X_2 \in B]
\end{aligned}$$

Donc pour tout  $n$ ,  $\mathbb{P}_\mu[X_n \in B] = \mu(B)$  ( $\forall B \in \mathcal{B}$ ). En utilisant la propriété de Markov, cela équivaut à la stricte stationnarité de la chaîne : la loi du processus  $(X_n, X_{n+1}, \dots, X_{n+k})$  est indépendante de  $n$ , pour tout entier  $k$ .

Inversement si  $(X_n)$  est strictement stationnaire,  $X_1$  et  $X_0$  ont la même loi, ce qui s'écrit  $\forall B \in \mathcal{B}$ ,  $\mu(B) = \int P(x, B) d\mu(x)$ , où  $\mu$  désigne la probabilité initiale. On a donc le résultat suivant.

**Propriété 1.4.1** *La chaîne  $(X_n)$  est strictement stationnaire si et seulement si sa loi de probabilité initiale est invariante.*

Cette propriété rend cruciale la recherche de mesures invariantes de masse finie. Une autre raison de l'importance des mesures de probabilité invariantes est liée au comportement ergodique (ou de long terme) de la chaîne. Supposons en effet qu'il existe une mesure de probabilité limite  $\pi_\mu$ , définie par :  $\forall B \in \mathcal{B}$ ,  $\mathbb{P}_\mu[X_n \in B] \longrightarrow \pi_\mu(B)$  quand  $n \rightarrow +\infty$ .

Alors

$$\begin{aligned}
\pi_\mu(B) &= \lim_{n \rightarrow +\infty} \int \mu(dx) P^n(x, B) \\
&= \lim_{n \rightarrow +\infty} \int \mu(dx) \int P^{n-1}(x, dy) P(y, B) \\
&= \int P(y, B) \lim_{n \rightarrow +\infty} \int P^{n-1}(x, dy) \mu(dx) \\
&= \int P(y, B) \pi_\mu(dy)
\end{aligned}$$

Donc  $\pi_\mu$  est invariante. En particulier, s'il existe une unique mesure de probabilité invariante, la probabilité limite sera indépendante de  $\mu$ . On a le résultat suivant :

**Théorème 1.4.1** *Si  $(X_n)$  est récurrente, elle admet une unique mesure sous-invariante  $\pi$  (à une constante multiplicative près) et cette mesure est invariante. De plus  $\pi$  est équivalente à toute mesure d'irréductibilité maximale  $\psi$ .*

[voir Meyn et Tweedie (1996), théorèmes 10.4.4 et 10.4.9]

Ce théorème nous conduit à une nouvelle dichotomie parmi les chaînes récurrentes. On distingue celles admettant une mesure de probabilité invariante (la chaîne est dite positive récurrente) des autres (la chaîne est dite nulle récurrente). Avec des notations évidentes, cette dichotomie s'écrit :  $\mathcal{R} = \mathcal{P} + \mathcal{N}$ , avec pour  $(X_n) \in \mathcal{R}$  :

$$(X_n) \in \mathcal{P} \quad \Longleftrightarrow \quad \forall B \in \mathcal{B}, \quad \pi(B) = \int P(x, B)\pi(dx)$$

où  $\pi$  est une mesure de probabilité, et

$$(X_n) \in \mathcal{N} \quad \Longleftrightarrow \quad \forall B \in \mathcal{B}, \quad \mu(B) \geq \int P(x, B)\mu(dx)$$

où  $\mu$  est une mesure telle que  $\mu(\mathcal{X}) = +\infty$ . Notons qu'il existe d'autres caractérisations de la positive (ou nulle) récurrence fondées sur le temps de retour : par exemple si l'on peut trouver un compact  $C$  tel que  $\sup_{y \in C} E[\tau_C \mid X_0 = y] < +\infty$ , la chaîne est positive récurrente moyennant quelques conditions supplémentaires [assurant que les compacts sont des petite sets, voir Théorème 1.6.1 plus bas].

## 1.5 Ergodicité

Jusqu'ici nous avons caractérisé la stabilité d'une chaîne en terme de récurrence. Une autre façon d'envisager la stabilité d'une chaîne consiste à étudier si elle converge

vers un état stationnaire quelque soit son point de départ. Nous avons déjà vu que l'existence d'une mesure de probabilité invariante était une condition nécessaire pour qu'une telle convergence ait lieu. Nous allons voir que pour les chaînes  $\psi$ -irréductibles et positives récurrentes la propriété est vérifiée pour une notion forte de la convergence que nous allons préciser. Rappelons que la norme en variation totale d'une mesure  $\mu$  est définie par  $\|\mu\| = \text{Sup} \int f d\mu$ , où le supremum est pris sur  $\{f : \mathcal{X} \rightarrow \mathbb{R}, f \text{ mesurable et } |f| \leq 1\}$ . Plus généralement, on définit  $\|\mu\|_g = \text{Sup}\{\int f d\mu : |f| \leq g\}$ , où  $g : \mathcal{X} \rightarrow \mathbb{R}^+$  est mesurable.

**Théorème 1.5.1** *Si  $(X_t)$  est une chaîne de Markov Harris récurrente, positive de mesure invariante  $\pi$  et apériodique, alors pour toute mesure initiale  $\mu$*

$$\| \int P^n(x, \cdot) \mu(dx) - \pi \| \rightarrow 0 \quad \text{quand} \quad n \rightarrow +\infty$$

[voir Meyn et Tweedie (1996), Théorème 13.3.3]

En prenant pour  $\mu$  la masse de Dirac en  $x$ , un corollaire de ce théorème est que pour toute fonction  $f$  bornée et mesurable :

$$\forall x \in \mathcal{X}, \quad \lim_{n \rightarrow +\infty} E[f(X_n) | X_0 = x] = \int f d\pi. \quad (1.2)$$

En particulier, pour  $f = \mathbf{1}_B$ ,  $B \in \mathcal{B} : P^n(x, B) \rightarrow \pi(B)$ ,  $\forall x \in \mathcal{X}$ .

Les chaînes vérifiant les hypothèses du théorème sont dites *ergodiques*.

Lorsque  $f$  n'est pas supposée bornée, on a le résultat suivant :

**Théorème 1.5.2** *Soit  $(X_t)$  une chaîne de Markov Harris récurrente, positive de mesure invariante  $\pi$ , fortement apériodique, de mesure d'irréductibilité maximale  $\psi$ , et soit  $f : \mathcal{X} \rightarrow [1, +\infty[$ . Alors*

1. si  $\int f d\pi = +\infty$ ,  $E[f(X_n) | X_0 = x] \longrightarrow +\infty$  quand  $n \longrightarrow +\infty$
2. si  $\int f d\pi < +\infty$ , alors pour tout  $x \in A$ , où  $A$  est absorbant et tel que  $\psi(A^c) = 0$ , on a

$$\| P^n(x, \cdot) - \pi \|_f \longrightarrow 0 \quad \text{quand} \quad n \longrightarrow +\infty.$$

[voir Meyn et Tweedie (1996), Théorème 14.3.3]

On obtient en particulier, cette fois-ci pour  $f$  non nécessairement bornée (mais sous des hypothèses un peu plus fortes)

$$\forall x \in A, \quad \lim_{n \rightarrow +\infty} E[f(X_n) | X_0 = x] = \int f d\pi. \quad (1.3)$$

Si dans le théorème, la condition 2. est vraie pour tout  $x$ , la chaîne est dite *f-ergodique*. Il est parfois nécessaire de connaître la vitesse de convergence vers 0 de  $\| P^n(x, \cdot) - \pi \|_f$ . Sous les hypothèses du théorème précédent, nous dirons que la chaîne est *f-géométriquement ergodique* (ou plus simplement géométriquement ergodique si  $f = 1$ ) si : il existe une constante  $r_f > 1$  telle que

$$\sum_{n=1}^{\infty} r_f^n \| P^n(x, \cdot) - \pi \|_f < \infty, \quad \forall x \in \mathcal{X}.$$

Nous donnerons dans la partie suivante un critère assurant une telle propriété.

Enfin  $(X_n)$  est dite *uniformément ergodique* si

$$\sup_{x \in \mathcal{X}} \| P^n(x, \cdot) - \pi \| \longrightarrow 0 \quad \text{quand} \quad n \longrightarrow +\infty$$

Le résultat suivant donne des caractérisations équivalentes de l'ergodicité uniforme.

**Théorème 1.5.3** *Soit  $(X_t)$  une chaîne quelconque. Les propriétés suivantes sont équivalentes :*

1.  $(X_t)$  est uniformément ergodique.
2.  $\exists \rho < 1, \exists R < \infty, \forall x, \quad \| P^n(x, \cdot) - \pi \| \leq R\rho^n.$
3.  $\exists n \in \mathbb{N}, \quad \sup_{x \in \mathcal{X}} \| P^n(x, \cdot) - \pi \| < 1.$

[voir Meyn et Tweedie (1996), Théorème 16.2.1]

Une conséquence importante de l'ergodicité uniforme est le lien avec la notion de mélange. On a le

**Théorème 1.5.4** *Si  $(X_t)$  est uniformément ergodique, il existe des constantes  $R < +\infty$  et  $\rho < 1$  telles que, pour toutes fonctions  $h$  et  $g$  telles que  $|g| \leq 1$  et  $|h| \leq 1, \forall n, k, \forall x \in \mathcal{X}$*

$$|E\{g(X_k)h(X_{n+k}) \mid X_0 = x\} - E\{g(X_k) \mid X_0 = x\}E\{h(X_{n+k}) \mid X_0 = x\}| \leq R\rho^n[1+\rho^k]$$

*La chaîne est donc géométriquement mélangeante.*

[voir Meyn et Tweedie (1996), Théorème 16.1.5]

## 1.6 Critères de classification

Pour des modèles particuliers, il n'est généralement pas facile de vérifier directement les propriétés de chaîne de Markov transitoire, récurrente ou admettant une loi de probabilité invariante. Il existe heureusement des critères s'exprimant simplement à partir de conditions sur les probabilités de transition en une étape. Ces contraintes

sont de deux types : continuité et contrôle de la dérive moyenne. Pour simplifier la présentation nous supposerons dans cette partie que l'espace d'états  $\mathcal{X}$  est égal à  $\mathbb{R}^d$ , où  $d$  est un entier, muni de sa topologie usuelle.

La condition suivante est appelée forte continuité :

$(C_0)$  : L'application  $P(\cdot, B)$  est continue, pour tout  $B \in \mathcal{B}$ .

Cette contrainte se révèle souvent trop forte, par exemple pour les dynamiques avec effets de seuil. Rappelons qu'une fonction  $h : \mathcal{X} \rightarrow \mathbb{R}$  est semi-continue inférieurement si l'ensemble  $\{x : h(x) > c\}$  est ouvert pour toute constante  $c$ . Une chaîne telle que, pour tout ouvert  $B$ , l'application  $P(\cdot, B)$  est semi-continue inférieurement est dite *Fellerienne*. Plus généralement, on a la définition suivante.

**Définition 1.6.1** *S'il existe une application  $T(\cdot, \cdot)$  telle que*

1.  $\forall x \in \mathcal{X}, \forall B \in \mathcal{B}, P(x, B) \geq T(x, B)$ ,
2.  $T(\cdot, B)$  est semi-continue inférieurement,  $\forall B \in \mathcal{B}$ ,
3.  $T(x, \cdot)$  est une mesure non nulle sur  $(\mathcal{X}, \mathcal{B})$ ,  $\forall x \in \mathcal{X}$

$(X_t)$  est appelée *T-chaîne*.

Une chaîne vérifiant  $(C_0)$  est évidemment une *T-chaîne*. La condition 1. peut être remplacée, si nécessaire, par des contraintes moins fortes, portant sur toutes les probabilités de transition, du type

1. Il existe une suite  $(a_n)$  de constantes positives de somme égale à 1 telle que

$$K_a(x, B) = \sum_{n \geq 1} a_n P^n(x, B) \geq T(x, B).$$

$T$  est alors appelée *composante continue* de  $K_a$ .

Ces contraintes de continuité ne sont pas nécessaires pour la stabilité, mais elles permettent d'obtenir des caractérisations en se limitant aux ensembles compacts.

**Théorème 1.6.1** *Si  $(X_t)$  est une  $T$ -chaîne  $\psi$ -irréductible, les compacts sont des petite sets.*

[voir Meyn et Tweedie (1996), Théorème 6.2.5]

L'autre type de conditions porte sur l'opérateur de dérive  $\Delta$ . A toute fonction  $V : \mathcal{X} \rightarrow \mathbb{R}^+$ , intégrable par rapport aux mesures  $P(x, \cdot)$  ( $\forall x \in \mathcal{X}$ ), on associe

$$\Delta V(x) = \int P(x, dy)V(y) - V(x) = E[V(X_{t+1}) - V(X_t) \mid X_t = x]$$

Cet opérateur calcule, pour tout point initial  $x$ , la valeur moyenne de la variation de  $V(x)$  en une étape. On a le critère de récurrence suivant :

**Théorème 1.6.2** *Soit  $(X_n)$  une chaîne de markov  $\psi$ -irréductible,  $V$  une fonction de  $\mathcal{X} \rightarrow \mathbb{R}^+$ , intégrable par rapport aux mesures  $P(x, \cdot)$ ,  $\forall x \in \mathcal{X}$ , et  $C$  un compact (ou un petite set). Alors si  $(C_0)$  est vraie, ou si  $(X_n)$  est une  $T$ -chaîne, et si  $\Delta V(x) \leq 0$ ,  $\forall x \in C^c$ , la chaîne  $(X_n)$  est récurrente.*

[voir Meyn et Tweedie (1996), Théorème 8.4.3]

L'interprétation de l'inégalité est évidente : si la chaîne sort d'un certain compact, il existe une "force de rappel" faisant diminuer, en moyenne, la valeur de  $V(X_t)$ . En pratique, il s'agit de trouver une telle fonction  $V$  (appelée fonction test, ou fonction de Lyapounov) ainsi qu'un compact  $C$  permettant de conclure. L'existence d'une loi

de probabilité invariante peut être obtenue à partir du résultat suivant.

**Théorème 1.6.3** *Sous les hypothèses du Théorème 1.6.2 et si*

$$\begin{aligned}\Delta V(x) &\leq -1, & \forall x \in C^c \\ \Delta V(x) &\leq M, & \forall x \in C\end{aligned}$$

où  $M$  est une constante, la chaîne  $(X_t)$  est positive récurrente.

Dans la première inégalité, la valeur  $-1$  peut être remplacée par n'importe quelle constante strictement négative (quitte à modifier  $V$ ). Le critère suivant permet d'obtenir l'existence de moments de la probabilité invariante.

**Théorème 1.6.4** *Soit  $(X_t)$  une chaîne de Markov positive récurrente de loi de probabilité invariante  $\pi$ . Soient  $V, f$  et  $s$  des fonctions  $\mathcal{X} \rightarrow \mathbb{R}^+$  telles que*

$$\Delta V(x) \leq -f(x) + s(x), \quad \forall x \in \mathcal{X}.$$

Alors  $\int f d\pi \leq \int s d\pi$ .

[voir Meyn et Tweedie (1996), Théorème 14.3.7]

Si dans ce Théorème on peut choisir  $s$  de la forme :  $s(x) = b\mathbb{1}_C(x)$ , où  $b$  est une constante positive et  $C \in \mathcal{B}$ , on obtient

$$\int f d\pi \leq b\pi(C) < +\infty.$$

**Théorème 1.6.5** *Sous les hypothèses du Théorème 1.6.2, si  $(X_t)$  est apériodique et si pour des constantes  $\beta > 0$  et  $b < +\infty$ , un petite set  $C$  et  $V : \mathcal{X} \rightarrow [1, +\infty[$  on a*

pour tout  $x \in \mathcal{X}$

$$\Delta V(x) \leq -\beta V(x) + b\mathbf{1}_C(x), \quad (1.4)$$

alors il existe  $r > 1$  tel que

$$\sum_{n \geq 0} r^n \|\| P^n - \pi \|\|_V < +\infty \quad (1.5)$$

où  $\|\| P^n - \pi \|\|_V = \sup_{x \in \mathcal{X}} \frac{\|P^n(x, \cdot) - \pi\|_V}{V(x)}$ . La chaîne est dite  $V$ -uniformément ergodique.

[voir Meyn et Tweedie (1996), Théorème 16.1.2].

**Remarques :** 1. Sous les conditions du Théorème précédent on a

$$\int V d\pi < +\infty,$$

donc  $E_\pi\{V(X_t)\} < +\infty$ , en utilisant le Théorème 1.6.4 avec  $s(x) = b\mathbf{1}_C(x)$ .

2. Une conséquence immédiate de (1.5) est que

$$\forall x \in \mathcal{X}, \quad \sum_{n \geq 0} r^n \| P^n(x, \cdot) - \pi \| < RV(x) \quad (1.6)$$

pour une constante  $R$  finie. La chaîne est donc géométriquement ergodique.

Il est parfois utile de disposer de critères ne reposant pas sur les probabilités de transition en une étape. Le résultat suivant montre comment adapter ce qui précède.

**Théorème 1.6.6** *Sous les hypothèses du Théorème 1.6.5 et si la condition (1.4) est remplacée par*

$$E[V(X_{t+n})|X_t = x] \leq (1 - \beta)V(x) + b\mathbf{1}_C(x),$$

alors (1.6) est vérifiée. La chaîne est donc géométriquement ergodique.

Ce résultat peut être étendu en faisant dépendre de l'état  $x$  l'entier  $n$  et le réel  $\beta$ . [voir Meyn et Tweedie (1996), Théorème 19.1.3].

Donnons pour conclure cette partie, des critères utilisant l'opérateur de dérive peuvent être vérifiés pour établir des réciproques aux conditions de stabilité.

**Théorème 1.6.7** *Si  $(X_t)$  est une chaîne de Markov  $\psi$ -irréductible et si, pour  $V : \mathcal{X} \rightarrow \mathbb{R}^+$ , on a  $\Delta V(x) \geq 0 \quad \forall x \in C^c$*

*où  $C \in \mathcal{B}^+$ , et*

$$\sup_{x \in \mathcal{X}} \int P(x, dy) |V(x) - V(y)| < \infty,$$

*si de plus  $\{x \in \mathcal{X} \mid V(x) > \sup_{y \in C} V(y)\} \in \mathcal{B}^+$  alors la chaîne n'est pas positive récurrente*

[voir Meyn et Tweedie (1996), Théorème 11.5.1]

**Théorème 1.6.8** *Soit  $(X_t)$  est une chaîne de Markov  $\psi$ -irréductible. Alors  $(X_t)$  est transitoire si et seulement si il existe une fonction,  $V : \mathcal{X} \rightarrow \mathbb{R}^+$  bornée, et  $r \geq 0$  tels que*

1.  $\{x \mid V(x) \leq r\}$  et  $\{x \mid V(x) > r\}$  appartiennent à  $\mathcal{B}^+$
2. Pour tout  $x$  tel que  $V(x) > r$ , on a  $\Delta V(x) > 0$

[voir Meyn et Tweedie (1996), Théorème 8.4.2]

# Bibliographie

- [1] **Bougerol, P., et N. Picard (1992a)**, *Strict Stationarity of Generalized Autoregressive Processes*, Ann. Probab. 20, 1714-1730.
- [2] **Bougerol, P., et N. Picard (1992b)**, *Stationnarity of GARCH Processus and of Some Non-Negative Time Series*, Journal of Econometrics, 52, 115-127.
- [3] **Brandt A. (1986)**, *The Stochastic Equation  $y_{n+1} = a_n y_n + b_n$  With Stationary Coefficients*. Adv. Appl. Probab., 18, 211-220, 1986.
- [4] **Brockwell, P. J. et R. L. Davis (1991)**, *Times Series : Theory and Methods*, Springer-Verlag, New York.
- [5] **Chan, K. S. , J. D. Petrucci, H. Tong, S. W. Woolford (1985)**, *A Multiple-Threshold AR(1) Model*, J. Appl. Prob. 22, 267-279.
- [6] **Duflo, M. (1990)**, *Méthodes Récursives Aléatoires*, Masson.
- [7] **Gouriéroux C. et A. Monfort (1989)**, *Statistique et Modèles Econométriques*, Economica, Paris.
- [8] **Meyn et R. L. Tweedie (1996)**, *Markov Chains and Stochastic Stability*, Springer-Verlag, 3rd printing, London.

- [9] **Nummelin, E. (1984)**, *General Irreducible Markov Chains and Non-Negative Operators*, Cambridge University Press.
- [10] **Petruccelli, J. D. et Woolford, S. W. (1984)**, *A Threshold AR(1) Model*, J. Appl. Prob. 21, 270-286.
- [11] **Revuz D. (1975)**, *Markov Chains*, North-Holland.
- [12] **Tong H. (1990)**, *Non-linear Time Series, A Dynamical System Approach*, Oxford Science Publications.
- [13] **Tweedie, R. L. (1974)**, *R-Theory For Markov Chains On a General State Space I : Solidarity Properties and R-Recurrent Chains*, The Annals of Probabilities, Vol 2 n°5, 840-864.
- [14] **Tweedie, R. L. (1975)**, *Sufficient Conditions for Ergodicity and Recurrence of Markov Chains on General State Space*, Stochastic Processes and their Applications, 3, 385-403.
- [15] **Tweedie, R. L. (1976)**, *Criteria For Classifying General Markov Chains*, Adv. Appl. Probabilities, 8, 737-771.
- [16] **Tweedie, R. L. (1983a)**, *Criteria For Rates of Convergence of Markov Chains, With Application to Queueing and Storage Theory*, Probability, Statistics and Analysis, édité par J. F. C. Kingman et G. E. H. Reuter. London Mathematics Society Lecture Note Series, Cambridge University Press, Cambridge.
- [17] **Tweedie, R. L. (1983b)**, *The Existence of Moments for Stationary Markov Chains*, J. Appl. Prob. 20, 191-196.

# Chapitre 2

## Etude probabiliste

### 2.1 Introduction

Le modèle GARCH( $p, q$ ) standard, introduit par Engle (1982) et généralisé par Bollerslev, se présente sous la forme

$$\begin{cases} \varepsilon_t &= h_t^{1/2} \eta_t \\ h_t &= \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j} \end{cases} \quad (2.1)$$

où  $(\eta_t)$  est une suite de variables indépendantes, identiquement distribuées (i.i.d.) centrées et de variance unité, les  $\alpha_i$  et  $\beta_j$  sont des constantes positives et  $\omega$  est une constante strictement positive. Sous certaines conditions sur les paramètres et la loi de  $\eta_t$ , le modèle admet une unique solution strictement stationnaire non anticipative (*i.e.* telle que  $\varepsilon_t$  soit une fonction mesurable de  $\eta_t$  et de son passé). Sous des conditions plus fortes, le processus solution admet des moments d'ordre pair. En particulier, ce modèle est compatible avec l'existence d'une variance marginale pour le processus  $\varepsilon_t$ . La variable  $h_t$ , appelée volatilité, s'interprète comme la variance

de  $\varepsilon_t$  conditionnellement à son passé. Cette spécification a connu un grand succès car elle permet de reproduire les trajectoires des séries financières caractérisées par des périodes calmes, suivies de périodes agitées. Elle a également la propriété de leptokurticité des distributions marginales (queues plus épaisses que pour les lois gaussiennes), tout en étant suffisamment simple pour permettre une étude étendue des propriétés probabilistes et statistiques du modèle.

Cependant, diverses limitations du modèle initial ont conduit des auteurs à proposer des généralisations. Une première classe d'extensions a visé à la prise en compte d'asymétries (voir Nelson (1991), modèle GARCH exponentiel; Rabemananjara et Zakoïan (1993), modèle GARCH à seuil, TGARCH) dans les modèles GARCH. Par construction, la volatilité du modèle (2.1) ne dépend que du module des variables passées : l'effet sur la volatilité de la date présente des innovations passées positives et négatives est donc identique. Cette propriété est en contradiction avec de nombreuses études sur les séries d'action, qui mettent en évidence une corrélation négative entre le carré des innovations de la date présente et les innovations passées : si la distribution conditionnelle était symétrique en les variables passées, cette corrélation serait nulle. Or, on observe une asymétrie marquée : l'accroissement de volatilité dû à une baisse des prix est généralement supérieur à celui résultant d'une hausse de même ampleur. Une façon de prendre en compte cette asymétrie consiste à faire dépendre la volatilité du signe des valeurs passées comme par exemple (modèle TGARCH  $(p, q)$ )

$$h_t = \omega + \sum_{i=1}^q \alpha_{i,+} \varepsilon_{t-i}^+ - \alpha_{i,-} \varepsilon_{t-i}^- + \sum_{j=1}^p \beta_j h_{t-j}$$

où  $\varepsilon_t^+ = \max(\varepsilon_t, 0)$ ,  $\varepsilon_t^- = \min(\varepsilon_t, 0)$ , les constantes  $\alpha_{i,+}$ ,  $\alpha_{i,-}$ ,  $\beta_j$  sont positives et  $\omega$

strictement positive. Dans ce modèle, comme dans le GARCH standard, la volatilité dépend évidemment du module des valeurs passées mais la dynamique est la même pour les grandes et petites valeurs (en module).

Une classe de généralisations permettant de dissocier les dynamiques des grandes et petites innovations est celle des modèles GARCH à changements de régimes markoviens (voir Hamilton et Susmel (1994), Francq, Roussignol et Zakoïan (2001)). Ils s'obtiennent en faisant dépendre les coefficients  $\omega$ ,  $\alpha_i$  et  $\beta_j$  dans (2.1) d'une chaîne de Markov à nombre fini d'états. Cette spécification permet par exemple de spécifier des volatilités différentes pour les innovations de petit module et pour celles de grand module. On constate empiriquement lorsqu'on estime ces modèles que la volatilité est presque constante dans les régimes correspondant aux petites innovations, et que l'effet ARCH n'apparaît significativement que pour les régimes des innovations de grands modules.

Le modèle que nous allons étudier se rattache aux extensions précédentes par l'existence de plusieurs régimes. Dans le modèle TGARCH, les deux régimes correspondent au signe des valeurs passées. Dans les modèles à changements de régimes markoviens, ceux-ci sont exogènes, les régimes correspondant aux états d'une chaîne de Markov inobservable. Dans le modèle proposé dans ce chapitre, la volatilité dépend de la position relative des innovations passées. Nous commençons par présenter le modèle le plus simple.

## 2.2 Le modèle et ses propriétés élémentaires

On considère dans cette partie le modèle d'ordre 1 défini par

$$\begin{cases} \varepsilon_t = h_t^{1/2} \eta_t, \\ h_t = \omega + \alpha \varepsilon_{t-1}^2 \mathbb{1}_{[\varepsilon_{t-1}^2 > k \varepsilon_{t-2}^2]} \end{cases} \quad (2.2)$$

où  $(\eta_t)_{t \in \mathbb{Z}}$  est une suite de variables i.i.d centrées et de variance 1, définies sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$ ;  $\omega > 0$ ,  $\alpha \geq 0$ ,  $k \geq 0$ . On suppose que  $\eta_t$  est indépendante de  $\{\varepsilon_{t-j}, j > 0\}$ .

Dans ce modèle,  $h_t$  s'interprète comme la variance conditionnelle au passé de  $\varepsilon_t$ . Celle-ci étant non constante, le modèle est de type ARCH à deux régimes dès que  $k \neq 0$ .

On obtient le modèle ARCH(1) standard

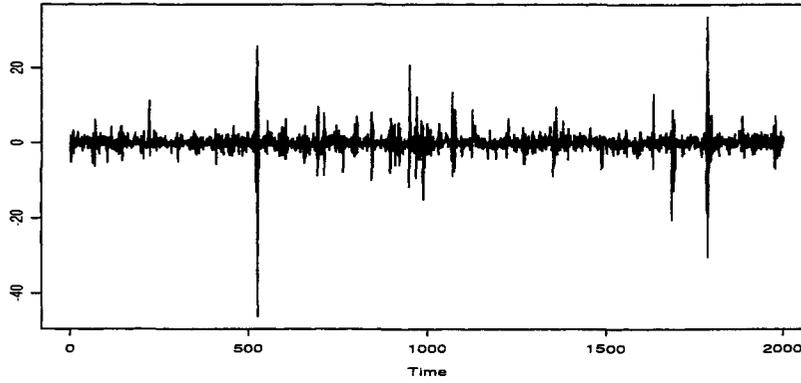
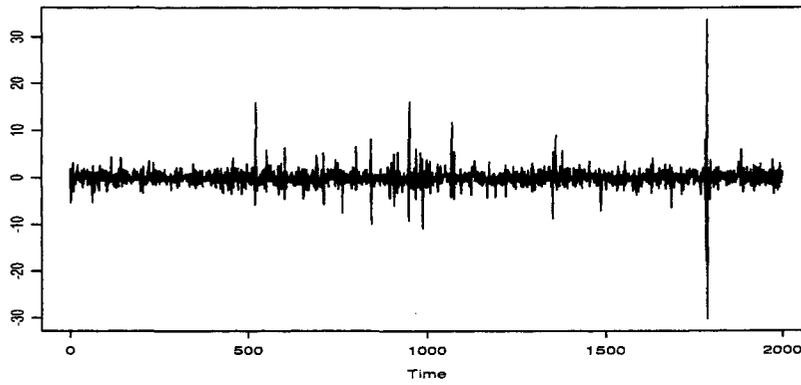
$$\varepsilon_t = \sqrt{\omega + \alpha \varepsilon_{t-1}^2} \eta_t$$

en choisissant  $k = 0$  dans (2.2). Si on prend  $\alpha = 0$ , on obtient le modèle conditionnellement homoscédastique  $\varepsilon_t = \sqrt{\omega} \eta_t$ . Notons que si  $k$  est "grand" on obtient également un modèle proche de l'homoscédasticité. En effet, lorsque  $k = \infty$ , la contrainte n'est jamais vérifiée et le modèle (2.2) se ramène à un bruit blanc fort à volatilité constante.

Pour des valeurs non particulières de  $\alpha$  et  $k$ , la dynamique définie par (2.2) présente deux régimes : conditionnellement homoscédastique et ARCH(1). En effet, lorsque  $0 < k < \infty$ , le processus "saute" entre un régime à volatilité constante (lorsque  $\varepsilon_{t-1}^2$  est "faible" par rapport à  $k \varepsilon_{t-2}^2$ ) et un régime à volatilité ARCH(1) (lorsque  $\varepsilon_{t-1}^2$  est "grand" par rapport à  $k \varepsilon_{t-2}^2$ ).

L'occurrence de l'un des deux régimes à la date  $t$  dépend uniquement de la valeur du ratio  $\frac{\varepsilon_t^2}{\varepsilon_{t-2}^2}$ . Tant que ce ratio reste inférieur à  $k$ , la dynamique reste dans le régime homoscédastique.

Les simulations suivantes illustrent la flexibilité de la spécification (Figures 2.1 à 2.10). Toutes les simulations ont été effectuées avec la même suite  $(\eta_t)$ . La Figure 2.1 correspond à un modèle ARCH(1) standard, admettant des moments d'ordre 2 mais pas de moments d'ordre 4. Ceci explique l'occurrence de valeurs de très grands modules. Dans la Figure 2.2 on a  $k = 1.5$  : l'introduction du régime homoscédastique a pour effet de supprimer certaines valeurs de grand module (autour de l'observation 500) dans les périodes turbulentes. Si l'on augmente  $\alpha$  le modèle peut devenir explosif en variance, comme cela semble être le cas dans la simulation de la Figure 2.3. L'étude de la stationnarité menée plus loin montrera que le processus simulé est strictement stationnaire mais, pour cette série, les conditions (suffisantes) de stationnarité au second-ordre sont violées. En revanche les simulations présentées dans les Figures 2.4 à 2.10 vérifient les conditions de stationnarité au second-ordre. Certaines de ces simulations, par exemple celle de la Figure 2.8, correspondent pourtant à une valeur de  $\alpha$  très supérieure à 1. Pour de telles valeurs les dynamiques sont très différentes dans les deux régimes. La Figure 2.10 permet de voir plus précisément le comportement de ces processus. Les traits en pointillés indiquent l'appartenance au régime homoscédastique (valeur 0) ou ARCH (valeur 10 pour plus de visibilité). On constate que les transitions d'un régime à l'autre sont fréquentes, que la série passe des durées totales voisines dans chaque régime et que les observations de grand module correspondent à un séjour de plusieurs dates dans le régime ARCH.

FIG. 2.1 – Trajectoire du modèle (2.2) avec  $k = 0$ ,  $\alpha = 0.9$ .FIG. 2.2 – Trajectoire du modèle (2.2) avec  $k = 1.5$ ,  $\alpha = 0.9$ .

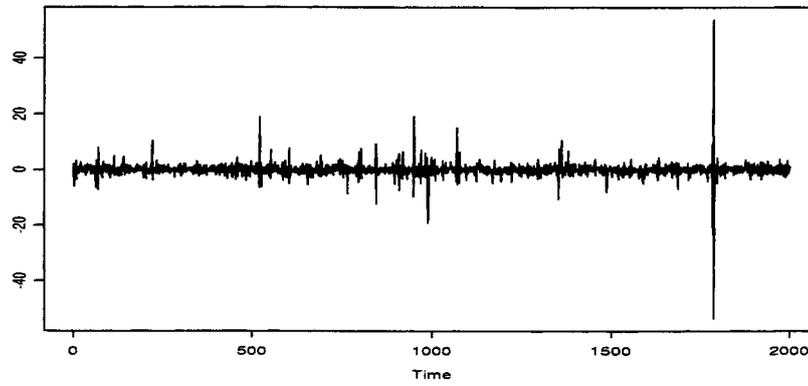


FIG. 2.3 – Trajectoire du modèle (2.2) avec  $k = 1.5$ ,  $\alpha = 1.1$ .

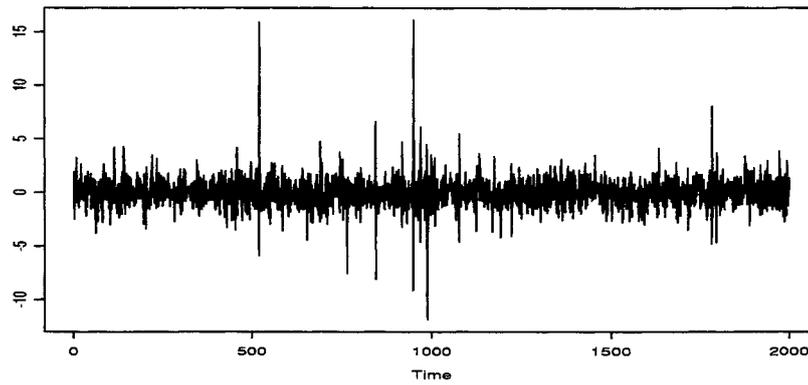
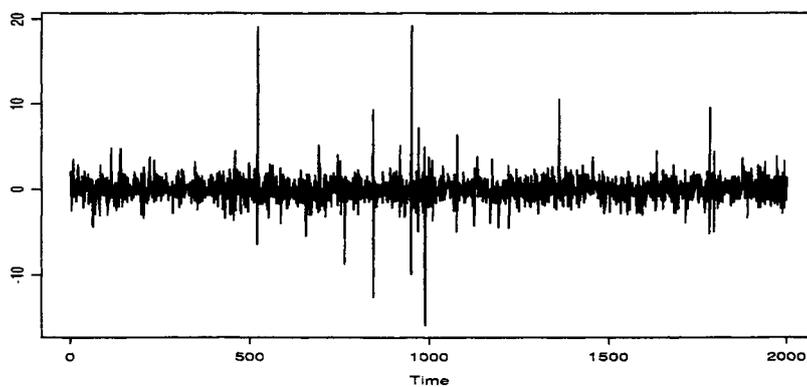
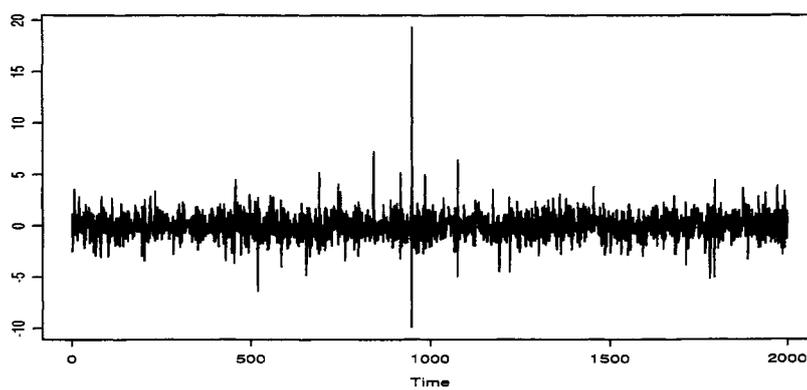


FIG. 2.4 – Trajectoire du modèle (2.2) avec  $k = 4$ ,  $\alpha = 0.9$ .

FIG. 2.5 – Trajectoire du modèle (2.2) avec  $k = 4$ ,  $\alpha = 1.1$ .FIG. 2.6 – Trajectoire du modèle (2.2) avec  $k = 9$ ,  $\alpha = 1.1$ .

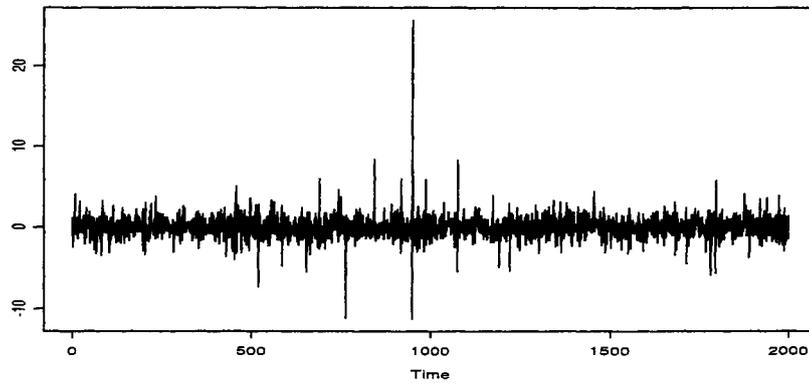


FIG. 2.7 – Trajectoire du modèle (2.2) avec  $k = 9$ ,  $\alpha = 1.5$ .

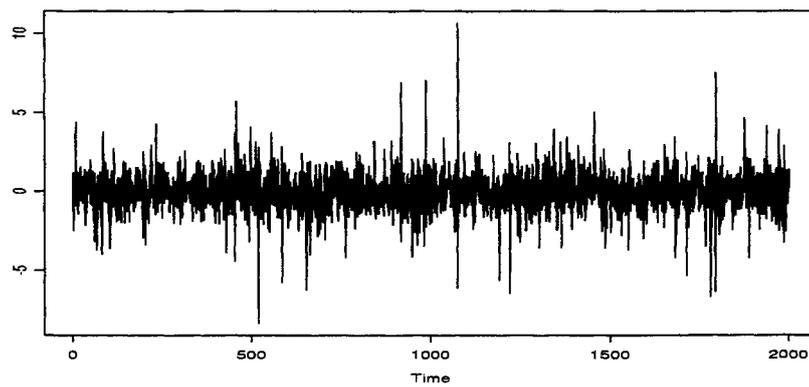
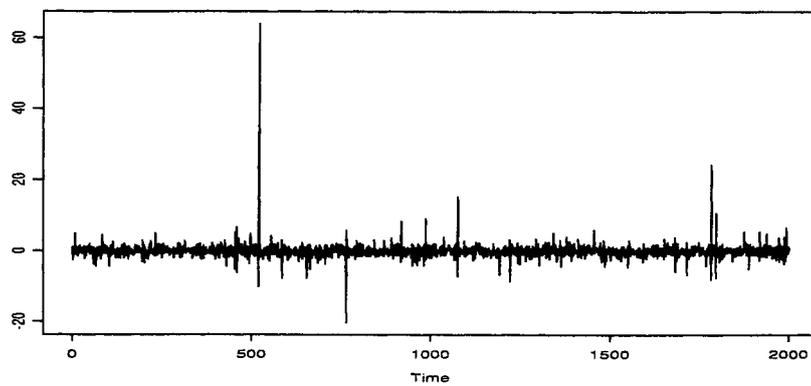
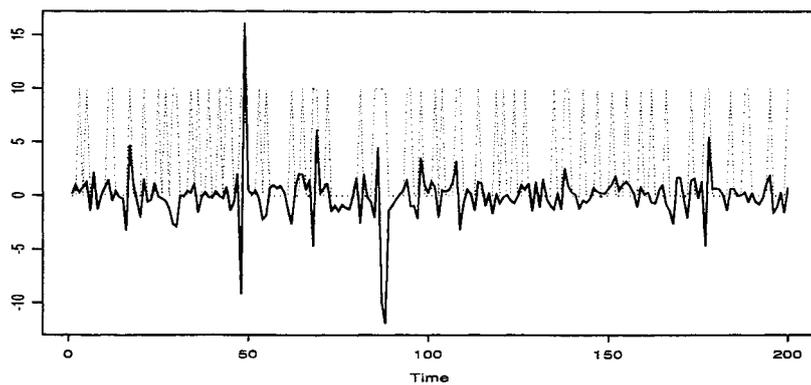


FIG. 2.8 – Trajectoire du modèle (2.2) avec  $k = 15$ ,  $\alpha = 2$ .

FIG. 2.9 – Trajectoire du modèle (2.2) avec  $k = 15$ ,  $\alpha = 3$ .FIG. 2.10 – Trajectoire du modèle (2.2) avec  $k = 15$ ,  $\alpha = 3$ .

Il est clair que d'après les simulations précédentes que si  $k$  est petit et  $\alpha$  est grand, les trajectoires auront tendance à présenter un caractère *explosif*.

Ce modèle s'inscrit dans la ligne des modèles ARCH asymétriques (dont les plus utilisés ont été cités dans l'introduction) où la spécification de la volatilité présente plusieurs régimes en fonction du signe ou de la taille des innovations. L'originalité du modèle réside dans la prise en compte d'une évolution relative des innovations.

Afin d'avoir une intuition plus précise sur les contraintes de stabilité à imposer sur les coefficients, nous considérons pour commencer le modèle déterministe.

## 2.3 Stabilité du modèle déterministe

Supposons que dans (2.2), on ait pour tout  $t$ ,  $\eta_t^2 = 1$ , *p.s.*

Nous appellerons ce modèle déterministe bien que  $\varepsilon_t$  soit une variable aléatoire. Pour la simplicité de l'exposé nous supposons que  $\varepsilon_0 = 0$  mais toute autre valeur initiale conduirait au même résultat.

Le modèle déterministe s'écrit donc

$$\begin{cases} \varepsilon_0 = 0, & \varepsilon_1 = \sqrt{\omega} \\ \varepsilon_t^2 = \omega + \alpha \varepsilon_{t-1}^2 \mathbb{1}_{[\varepsilon_{t-1}^2 > k \varepsilon_{t-2}^2]}, & t > 1 \quad p.s \end{cases} \quad (2.3)$$

Le résultat suivant fournit les conditions de stabilité du modèle, en fonction de  $\alpha$  et  $k$ .

**Théorème 2.3.1** Soit  $(\varepsilon_t)_{t \geq 0}$  défini par (2.3). Alors :

- i) Si  $\max(\alpha, 1) < k$ , il existe  $i \geq 0$  tel que :  $\forall t \geq i$ ,  $\varepsilon_t^2 = \omega$  p.s.;
- ii) Si  $\alpha < 1$  et  $k \leq 1$ ,  $\varepsilon_t^2 \rightarrow \frac{\omega}{1-\alpha}$  p.s. quand  $t \rightarrow +\infty$ ;
- iii) Si  $\alpha \geq \max(1, k)$ ,  $\varepsilon_t^2 \rightarrow +\infty$  p.s. quand  $t \rightarrow +\infty$ .

**Preuve** : On a, presque sûrement  $\varepsilon_1^2 = \omega$  et  $\varepsilon_2^2 = \omega(1 + \alpha)$ . La valeur de  $\varepsilon_3^2$  dépend de la position de  $1 + \alpha$  par rapport à  $k$ . On introduit, pour tout  $i \geq 0$ , l'ensemble

$$E_i = \{(\alpha, k) ; \alpha + \frac{1}{1 + \alpha + \dots + \alpha^i} \leq k\}$$

Puisque  $\alpha \geq 0$ , la suite  $(\alpha + \frac{1}{1 + \alpha + \dots + \alpha^i})_{i \geq 0}$  est décroissante. Donc les ensembles  $E_i$  constituent une suite croissante pour l'inclusion. On a

$$\begin{aligned} E_\infty &:= \bigcup_{i \geq 0} E_i = \{\alpha = 0, k \geq 1\} \cup \{0 < \alpha < 1 < k\} \cup \{1 \leq \alpha < k\} \\ &= \{\max(\alpha, 1) < k\} \cup \{(0, 1)\}. \end{aligned}$$

Examinons les divers cas.

*Cas i).* On a  $(\alpha, k) \in E_\infty$ , donc il existe  $i \geq 0$  tel que  $(\alpha, k) \in E_i$ .

Soit  $i_0 = \min\{i \geq 0, (\alpha, k) \in E_i\}$ . On a pour  $1 \leq i \leq i_0 + 2$ ,  $\varepsilon_i^2 = \omega(1 + \dots + \alpha^{i-1})$ .

Donc

$$\frac{\varepsilon_{i_0+2}^2}{\varepsilon_{i_0+1}^2} = \alpha + \frac{1}{1 + \alpha + \dots + \alpha^{i_0}} \leq k.$$

Par suite  $\varepsilon_{i_0+3}^2 = \omega$  et puisque  $\frac{\varepsilon_{i_0+3}^2}{\varepsilon_{i_0+2}^2} = \frac{1}{1 + \alpha + \dots + \alpha^{i_0+1}} \leq 1 < k$ ,  $\varepsilon_i^2 = \omega$  pour tout  $i \geq i_0 + 3$ .

*Cas ii).* Si  $(\alpha, k) \neq (0, 1)$ , alors  $(\alpha, k) \notin E_\infty$ . Donc pour tout  $i \geq 1$ ,  $\varepsilon_i^2 = \omega(1 + \dots + \alpha^{i-1})$  et on a la convergence annoncée. Lorsque  $(\alpha, k) = (0, 1)$ , la suite  $(\varepsilon_i^2)$  est constante et vaut  $\omega$ .

*Cas iii).* On a  $(\alpha, k) \notin E_\infty$ . Donc pour tout  $i \geq 1$ ,  $\varepsilon_i^2 = \omega(1 + \dots + \alpha^{i-1})$  et la suite  $(\varepsilon_i^2)$  converge vers  $+\infty$ .

□

De ce théorème, il découle que la zone de non explosion du modèle déterministe est donnée par

$$\alpha < \text{Max}(1, k) \quad (2.4)$$

Cette région est tracée dans la Figure 2.11. La Figure 2.12 représente des trajec-

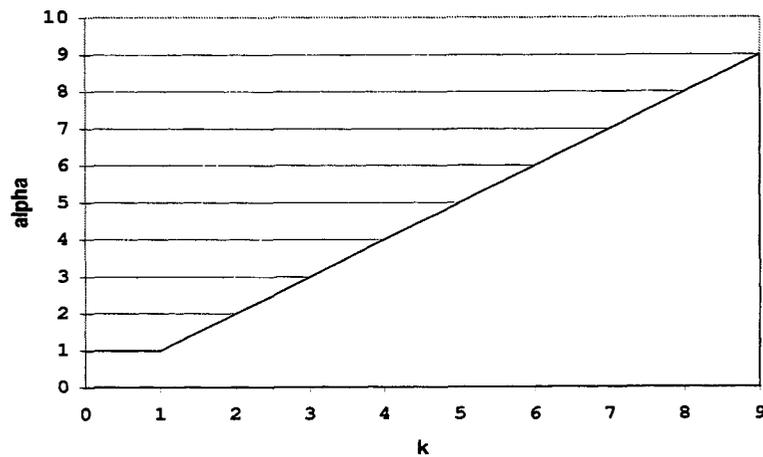


FIG. 2.11 – La région de stabilité du modèle (2.3).

toires du processus déterministe (2.3) avec ou sans contrainte (2.4).

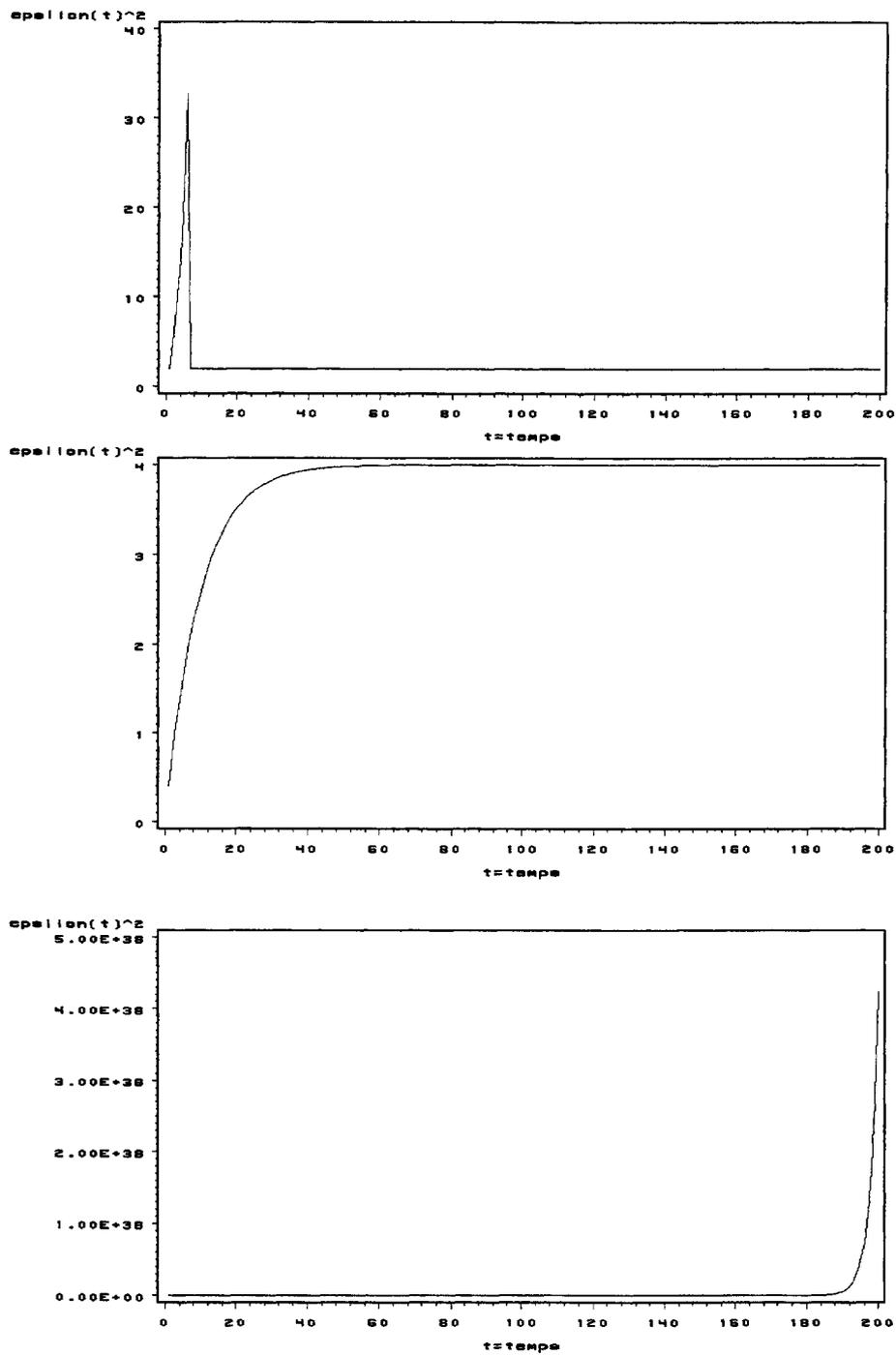


FIG. 2.12 – Trajectoires du modèle (2.3) : (1)  $(\omega=2, \alpha=1.4, k=1.5)$ , (2)  $(\omega=0.4, \alpha=0.9, k=0.3)$ , (3)  $(\omega=2, \alpha=1.55, k=1.5)$ .

## 2.4 Propriétés de chaîne de Markov

Contrairement aux modèles GARCH standard, la recherche de conditions de stationnarité (au sens strict ou au second ordre) ne peut être effectuée à partir des techniques de Bougerol et Picard (1992b) car il n'existe pas de représentation markovienne linéaire de la forme  $X_t = A_t X_{t-1} + b_t$ . Nous utilisons les techniques de chaîne de Markov présentées dans le chapitre introductif.

On note

$$X_t = \begin{pmatrix} \varepsilon_t^2 \\ \varepsilon_{t-1}^2 \end{pmatrix} = \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix}$$

et

$$\forall x \in \mathbb{R}^2, \quad \psi(x) = \omega + \alpha x_1 \mathbb{1}_{x_1 > kx_2}.$$

La représentation vectorielle du modèle (2.2) prend la forme d'une équation de récurrence stochastique non linéaire autorégressive d'ordre 1 :

$$X_t = \begin{pmatrix} \psi(X_{t-1})\eta_t^2 \\ X_{1,t-1} \end{pmatrix} := F(X_{t-1}, \eta_t). \quad (2.5)$$

La fonction  $F : \mathbb{R}^{+2} \times \mathbb{R} \rightarrow \mathbb{R}^{+2}$  n'est pas continue, à cause de la non-continuité de  $\psi$ , ce qui va compliquer l'étude de la stationnarité. Nous commençons par montrer plusieurs lemmes sur le processus  $(X_t)$ .

**Lemme 2.4.1** *Le processus  $(X_t)$  est une chaîne de Markov sur  $\mathbb{R}^{+2}$ , de probabilités de transition définies par :*

$$\forall x = (x_1, x_2) \in \mathbb{R}^{+2}, \forall B = B_1 \times B_2 \in \mathcal{B}(\mathbb{R}^+)^2,$$

$$\mathbf{P}(x, B) = P[\eta_t^2 \in \psi(x)^{-1}B_1] \mathbb{1}_{x_1 \in B_2}. \quad (2.6)$$

**Preuve :** Le caractère markovien découle de l'indépendance entre  $\eta_t$  et le passé de  $\varepsilon_t$ . L'homogénéité de la chaîne par rapport au temps est évidente car les  $\eta_t$  ont même loi.

□

Nous considérons maintenant le problème de l'irréductibilité. Rappelons que  $(X_t)$  est  $\phi$ -irréductible, où  $\phi$  est une mesure sur  $(\mathbb{R}^{+2}, \mathcal{B}(\mathbb{R}^{+2}))$ , si :

$$\forall B \in \mathcal{B}(\mathbb{R}^{+2}), \quad \forall x \in \mathbb{R}^{+2}, \quad \phi(B) > 0 \quad \Longrightarrow \quad \sum_{n=1}^{\infty} \mathbf{P}^n(x, B) > 0$$

où  $\mathbf{P}^t$  désigne le noyau de la transition en  $t$  étapes. Soit  $\lambda_1^+$  la mesure de lebesgue sur  $\mathbb{R}^+$  et  $\lambda_2^+$  la mesure de lebesgue sur  $\mathbb{R}^{+2}$ . Dans tout ce qui suit nous ferons l'hypothèse

**H :** Les variables  $\eta_t^2$  admettent une densité  $f$  par rapport à la mesure de lebesgue sur  $\mathbb{R}^+$ , avec  $f > 0$ .

Comme pour les modèles à seuils, cette hypothèse peut être remplacée par une condition moins forte (par exemple en supposant l'existence d'une densité positive dans un voisinage fixe de 0) mais au prix de complications peu utiles.

**Lemme 2.4.2** *Sous l'hypothèse H,  $(X_t)$  est  $\lambda_2^+$ -irréductible et  $\lambda_2^+$  est une mesure maximale d'irréductibilité.*

**Preuve :** Les probabilités de transition en 2 étapes sont définies par :

$$\forall B = B_1 \times B_2 \in \mathcal{B}(\mathbb{R}^+)^2, \quad \forall x = (x_1, x_2) \in \mathbb{R}^{+2},$$

$$\begin{aligned} \mathbf{P}^2(x, B) &= P[X_t \in B | X_{t-2} = x] \\ &= P[\psi(X_{t-1})\eta_t^2 \in B_1, \psi(x)\eta_{t-1}^2 \in B_2 | X_{t-2} = x] \\ &= P[\psi\{\psi(x)\eta_{t-1}^2, x_1\}\eta_t^2 \in B_1, \psi(x)\eta_{t-1}^2 \in B_2] \end{aligned} \quad (2.7)$$

$$= \int \mathbf{1}_{\psi(x)^{-1}B_2}(y) P[\eta_t^2 \in \psi\{\psi(x)y, x_1\}^{-1}B_1] f(y) d\lambda_1^+(y) \quad (2.8)$$

par le théorème de Fubini, en utilisant l'indépendance entre  $\eta_t$  et  $\eta_{t-1}$  et en notant que  $\psi(\cdot) \geq \omega > 0$ . Si  $\lambda_1^+(B_1) > 0$ , on a pour tout  $y \in \mathbb{R}^+$ ,  $P[\eta_t^2 \in \psi\{\psi(x)y, x_1\}^{-1}B_1] > 0$  d'après l'hypothèse sur la densité de  $\eta_t$ . De même  $\lambda_1^+\{\psi^{-1}(x)B_2\} > 0$  si  $\lambda_1^+(B_2) > 0$ . Par suite  $\mathbf{P}^2(x, B) > 0$  ce qui prouve l'irréductibilité de  $\lambda_2^+$ . La mesure est évidemment maximale,  $(X_t)$  prenant ses valeurs dans le quadrant positif.

□

**Remarque :** Nous avons déjà vu que la probabilité de transition (2.6) n'est généralement pas une fonction continue de  $x$ , pour  $B \in \mathcal{B}(\mathbb{R}^+)^2$ . Elle n'est pas non plus semi-continue inférieurement (s.c.i.) pour tout ouvert  $B$  : en effet, pour  $0 \leq \alpha < 1$ , la propriété que  $P[\eta_t^2 \in \psi(x)^{-1}B_1] > \alpha$ , supposée vérifiée en un  $x \in \mathbb{R}^{+2}$ , n'est généralement pas vraie sur un voisinage de  $x$  d'intersection non vide avec la droite  $x_1 = kx_2$  (en particulier si  $x$  appartient à cette droite). En conséquence, la chaîne  $(X_t)$  n'est pas Fellerienne et on ne peut pas considérer a priori les compacts comme des small sets, utilisés dans les critères d'ergodicité. Cette propriété, les compacts sont des small sets, découlera du résultat suivant.

**Lemme 2.4.3** *Sous les hypothèses du Lemme 2.4.2, le processus  $(X_t)$  est une T-chaîne.*

**Preuve :** Comme dans la preuve précédente, on raisonne sur  $\mathbf{P}^2(x, B)$  pour  $B = B_1 \times B_2 \in \mathcal{B}(\mathbb{R}^{+2})$ . Partitionnons le quadrant positif de  $\mathbb{R}^2$  en trois régions  $D_1 = \{x_1 < kx_2\}$ ,  $D_2 = \{x_1 = kx_2\}$  et  $D_3 = \{x_1 > kx_2\}$ , représentées sur la Figure 4.2 .

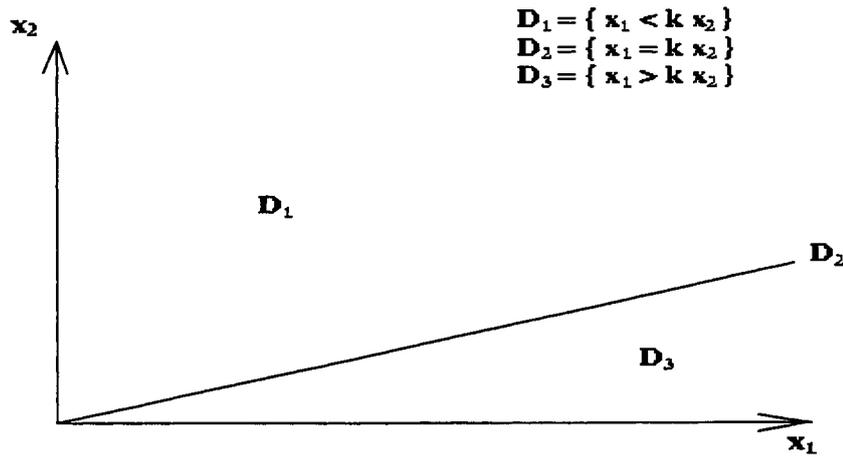


FIG. 2.13 – Les régions  $D_1$ ,  $D_2$  et  $D_3$

Pour  $x \in D_1 \cup D_2$ , on a  $\psi(x) = \omega$  donc, d'après (2.7) et en utilisant Fubini et l'indépendance entre  $\eta_t$  et  $\eta_{t-1}$  :

$$\begin{aligned} \mathbf{P}^2(x, B) &= P[\psi(\omega\eta_{t-1}^2, x_1)\eta_t^2 \in B_1, \quad \omega\eta_{t-1}^2 \in B_2] \\ &= \int \mathbf{1}_{\omega^{-1}B_2 \cap ]-\infty, \omega^{-1}kx_1]}(y) P[\eta_t^2 \in \omega^{-1}B_1] f(y) d\lambda_1^+(y) \\ &\quad + \int \mathbf{1}_{\omega^{-1}B_2 \cap ]\omega^{-1}kx_1, +\infty]}(y) P[\eta_t^2 \in \{\omega(1 + \alpha y)\}^{-1}B_1] f(y) d\lambda_1^+(y). \end{aligned}$$

Par application du théorème de Lebesgue, en utilisant les hypothèses de positivité en tout point de  $\mathbb{R}^+$  des densités de  $\eta_{t-1}^2$  et  $\eta_t^2$ , on obtient que  $\mathbf{P}^2(\cdot, B)$  est continue

sur  $D_1$ . Cela n'est généralement pas le cas pour  $x \in D_2$ . Cependant si  $x_n \rightarrow x$ , avec  $x_n \in D_1 \cup D_2$ , on a  $\mathbf{P}^2(x_n, B) \rightarrow \mathbf{P}^2(x, B)$  par les mêmes arguments. Pour  $x_n = (x_{n_1}, x_{n_2}) \in D_3$ , on a

$$\mathbf{P}^2(x_n, B) = P[\psi\{(\omega + \alpha x_{n_1})\eta_{t-1}^2, x_{n_1}\}\eta_t^2 \in B_1, (\omega + \alpha x_{n_1})\eta_{t-1}^2 \in B_2]$$

car  $\psi(x_n) = \omega + \alpha x_{n_1}$ . Donc en procédant comme précédemment sur  $D_1$ ,

$$\lim_{x_n \rightarrow x, x_n \in D_3} \mathbf{P}^2(x_n, B) = P[\psi\{(\omega + \alpha x_1)\eta_{t-1}^2, x_1\}\eta_t^2 \in B_1, (\omega + \alpha x_1)\eta_{t-1}^2 \in B_2]$$

Par suite, si on pose

$$\begin{aligned} T(x, B) &= P[\psi\{\omega\eta_{t-1}^2, x_1\}\eta_t^2 \in B_1, \omega\eta_{t-1}^2 \in B_2, \\ &\quad \psi\{(\omega + \alpha x_1)\eta_{t-1}^2, x_1\}\eta_t^2 \in B_1, (\omega + \alpha x_1)\eta_{t-1}^2 \in B_2], \end{aligned}$$

on définit une mesure pour tout  $x$ , non triviale car  $T(x, \mathbb{R}^{+2}) = 1$ , une application continue pour tout  $B$  et on a  $\mathbf{P}^2(x, B) \geq T(x, B)$  pour tout  $x$  et tout  $B$ . Donc  $T$  est une composante continue de  $\mathbf{P}^2$ . De plus ce qui prouve que  $(X_t)$  est une T-chaîne.

□

Une conséquence importante de la propriété précédente est que les compacts sont des petite sets. Les démonstrations classiques de l'ergodicité pour les modèles non linéaires (par exemple les modèles TAR, voir Petrucci et Woolford (1984)) reposent sur l'existence d'une "force de rappel" en dehors d'un compact, permettant de vérifier les critères de type Foster-Lyapounov. Dans le modèle étudié ici, cette force n'existe pas dans la région  $\{x_1 \leq kx_2\}$ . Il est donc impossible de vérifier les critères d'ergodicité en raisonnant seulement à partir des compacts et de leurs complémentaires. Nous montrons donc la propriété suivante.

**Lemme 2.4.4** *Sous les hypothèses du Lemme 2.4.2, l'ensemble  $C = \{x_1 \leq kx_2\}$  est un petite set pour la chaîne  $(X_t)$  et celle-ci est apériodique.*

**Preuve :** Dans la première partie de cette preuve nous raisonnons à nouveau sur les probabilités de transition en deux étapes. Pour  $x = (x_1, x_2) \in C$ , on a  $\psi(x) = \omega$ , donc d'après (2.7), pour tout  $B = B_1 \times B_2 \in \mathcal{B}(\mathbb{R}^{+2})$

$$\begin{aligned} \mathbf{P}^2(x, B) &= P[\psi(\omega\eta_{t-1}^2, x_1)\eta_t^2 \in B_1, \omega\eta_{t-1}^2 \in B_2] \\ &= P[\omega\eta_t^2 \in B_1, \omega\eta_{t-1}^2 \leq kx_1, \omega\eta_{t-1}^2 \in B_2] \\ &\quad + P[(\omega + \alpha\omega\eta_{t-1}^2)\eta_t^2 \in B_1, \omega\eta_{t-1}^2 > kx_1, \omega\eta_{t-1}^2 \in B_2] \\ &:= \mathbf{P}_1(x, B) + \mathbf{P}_2(x, B) \end{aligned}$$

Soit  $\varepsilon > 0$ . Pour  $x_1 > \varepsilon$ , on a

$$\begin{aligned} \mathbf{P}_1(x, B) &\geq P[\omega\eta_t^2 \in B_1, \omega\eta_{t-1}^2 \leq k\varepsilon, \omega\eta_{t-1}^2 \in B_2] \\ &:= \mu_1(B) \end{aligned} \tag{2.9}$$

Pour  $x_1 \leq \varepsilon$ , on a

$$\begin{aligned} \mathbf{P}_2(x, B) &\geq P[(\omega + \alpha\omega\eta_{t-1}^2)\eta_t^2 \in B_1, \omega\eta_{t-1}^2 > k\varepsilon, \omega\eta_{t-1}^2 \in B_2] \\ &:= \mu_2(B) \end{aligned}$$

Il est clair que les mesures  $\mu_1$  et  $\mu_2$  sont non triviales.

Par suite, les ensembles  $C_1 = \{x_1 \leq \varepsilon\} \cap C$  et  $C_2 = \{x_1 > \varepsilon\} \cap C$  sont des petite (small avec  $m = 2$ ) sets. En vertu de la remarque suivant la Définition 1.2.2, l'union de deux petite sets est un petite set. Donc  $C = C_1 \cup C_2$  est également un petite (small) set.

Pour montrer l'apériodicité, nous considérons maintenant la probabilité de transition en trois étapes. On choisit comme petite set l'ensemble  $\tilde{C} = C \cap \{x_1 \geq \varepsilon\}$ , pour  $\varepsilon > 0$  quelconque. Pour  $x = (x_1, x_2) \in \tilde{C}$  et pour tout  $B = B_1 \times B_2 \in \mathcal{B}(\mathbb{R}^{+2})$ , on a d'après (2.9), en translatant les dates d'une unité,

$$\begin{aligned} \mathbf{P}^2(x, B) &\geq P[\omega\eta_{t+1}^2 \in B_1, \omega\eta_t^2 \leq k\varepsilon, \omega\eta_t^2 \in B_2] \\ &\geq P[\omega\eta_{t+1}^2 \in B_1, \omega\eta_{t-1}^2 \leq k\varepsilon, \omega\eta_t^2 \leq k\varepsilon, \omega\eta_t^2 \in B_2, \eta_t^2 \leq k\eta_{t-1}^2] \\ &:= \mu(B) \end{aligned}$$

$$\begin{aligned} \mathbf{P}^3(x, B) &= P[\psi\{\psi(\omega\eta_{t-1}^2, x_1)\eta_t^2, \omega\eta_{t-1}^2\}\eta_{t+1}^2 \in B_1, \psi(\omega\eta_{t-1}^2, x_1)\eta_t^2 \in B_2] \\ &= P[\psi\{\omega\eta_t^2, \omega\eta_{t-1}^2\}\eta_{t+1}^2 \in B_1, \omega\eta_{t-1}^2 \leq kx_1, \omega\eta_t^2 \in B_2] \\ &\quad + P[\psi\{(\omega + \alpha\omega\eta_{t-1}^2)\eta_t^2, \omega\eta_{t-1}^2\}\eta_{t+1}^2 \in B_1, \\ &\quad \quad \quad \omega\eta_{t-1}^2 > kx_1, (\omega + \alpha\omega\eta_{t-1}^2)\eta_t^2 \in B_2] \\ &\geq P[\psi\{\omega\eta_t^2, \omega\eta_{t-1}^2\}\eta_{t+1}^2 \in B_1, \omega\eta_{t-1}^2 \leq k\varepsilon, \omega\eta_t^2 \in B_2] \\ &\geq P[\omega\eta_{t+1}^2 \in B_1, \omega\eta_{t-1}^2 \leq k\varepsilon, \omega\eta_t^2 \in B_2, \eta_t^2 \leq k\eta_{t-1}^2] \\ &\geq \mu(B). \end{aligned}$$

L'ensemble  $\tilde{C}$  est donc un  $\nu_2$ -small set et un  $\nu_3$ -small set, pour  $\nu_2 = \nu_3 = \mu$ . Il est clair que cette mesure est non triviale. Le pgcd de l'ensemble  $E_{\tilde{C}}$  intervenant dans la Définition 1.2.3 est donc 1. Ce pgcd étant indépendant du choix du petite set  $\tilde{C}$ , on en déduit l'apériodicité de la chaîne.

□

## 2.5 Ergodicité

Nous pouvons maintenant établir le résultat suivant, qui fournit une condition suffisante de stationnarité stricte pour le modèle (2.2).

**Théorème 2.5.1** *Sous l'hypothèse **H** et si*

$$0 \leq \alpha < \exp \{ -E(\log \eta_t^2) \}, \quad (2.10)$$

le processus  $(\varepsilon_t)_{t \geq 0}$  solution de (2.2) est géométriquement ergodique. De plus il existe un entier  $r > 0$  tel que  $E_\pi(\varepsilon_t^{2r}) < \infty$ , où  $\pi$  désigne la loi invariante de  $(\varepsilon_t)$ .

**Remarques : 1.** L'espérance dans (2.10) existe toujours dans  $[-\infty, +\infty[$  car  $E(\log^+ \eta_t^2) \leq E(\eta_t^2) = 1$ . On a également par Jensen,  $E(\log^+ \eta_t^2) \leq \log E(\eta_t^2) = 0$ . Donc le terme en exponentielle est plus grand que 1. Dans le cas du modèle ARCH(1) standard ( $k = 0$ ) on retrouve la condition usuelle de stationnarité stricte.

**2.** Dans le cas déterministe,  $\eta_t^2 = 1, p.s.$ , la condition devient :

$$0 \leq \alpha < 1. \quad (2.11)$$

On sait d'après le Théorème 2.3.1 que cette condition est trop forte lorsque  $k > 1$ .

**Preuve :** Soit  $r$  un réel strictement positif,  $r \in ]0, 1]$ . Notons  $E(\eta_t^{2r}) = \mu_{2r}$ . Nous utilisons le Théorème 1.6.6 du chapitre introductif avec  $n = 2$ , c'est à dire en travaillant sur les transitions en deux étapes. Grâce aux Lemmes 2.4.1-2.4.4, il suffit d'exhiber une fonction mesurable  $g : \mathcal{X} \rightarrow \mathbb{R}^+$  et un petite set  $C$  tels que

$$\begin{aligned} (i) \quad E[g(X_{t+2})|X_t = x] &\leq (1 - \beta)g(x) - \beta, & x \in C^c, \\ (ii) \quad E[g(X_{t+2})|X_t = x] &\leq b, & x \in C, \end{aligned}$$

pour des constantes finies strictement positives  $\beta$  et  $b$ . On a  $g = V - 1$  avec les notations du Théorème 1.6.6.

Soit  $g(x) = g(x_1, x_2) = x_1^r$ . L'espérance conditionnelle en (i) et (ii) s'écrit

$$\begin{aligned} E[\varepsilon_{t+2}^{2r} | X_t = (x_1, x_2)] &= E[\eta_{t+2}^{2r} \psi\{\psi(x_1, x_2) \eta_{t+1}^2, x_1\}^r | X_t = (x_1, x_2)] \\ &= \mu_{2r} E[\psi\{\psi(x_1, x_2) \eta_{t+1}^2, x_1\}^r] \\ &= \mu_{2r} E \left\{ \omega + \alpha \psi(x_1, x_2) \eta_{t+1}^2 \mathbf{1}_{\eta_{t+1}^2 > \frac{kx_1}{\psi(x_1, x_2)}} \right\}^r. \end{aligned}$$

Pour  $r \in ]0, 1]$  on a l'inégalité élémentaire  $(a + b)^r \leq a^r + b^r$  pour tous  $a, b \geq 0$ . En effet,  $x^r \geq x$  pour tout  $x \in [0, 1]$ , donc si  $a + b > 0$ ,  $(\frac{a}{a+b})^r + (\frac{b}{a+b})^r \geq \frac{a}{a+b} + \frac{b}{a+b} = 1$ . Par suite il vient

$$\begin{aligned} E[\varepsilon_{t+2}^{2r} | X_t = (x_1, x_2)] &\leq \mu_{2r} \omega^r + \mu_{2r} \alpha^r \psi(x_1, x_2)^r E[\eta_{t+1}^{2r} \mathbf{1}_{\eta_{t+1}^2 > \frac{kx_1}{\psi(x_1, x_2)}}] \\ &\leq \mu_{2r} \omega^r + \mu_{2r}^2 \alpha^r \psi(x_1, x_2)^r \end{aligned} \quad (2.12)$$

Donc pour  $x_1 > kx_2$ ,

$$E[\varepsilon_{t+2}^{2r} | X_t = (x_1, x_2)] \leq \mu_{2r} \omega^r + \mu_{2r}^2 \alpha^r (\omega + \alpha x_1)^r$$

Lorsque  $x_1 \rightarrow +\infty$ , le terme de droite dans cette inégalité est équivalent à :

$$\{\alpha^r \mu_{2r}\}^2 x_1^r.$$

Au voisinage de  $r = 0$  on a, par le théorème de Lebesgue, le développement limité :

$$\mu_{2r} = 1 + E(\log \eta_t^2) r + o(r).$$

Donc  $\alpha^r \mu_{2r} = 1 + \{\log \alpha + E(\log \eta_t^2)\} r + o(r)$  au voisinage de  $r = 0$ . Le terme entre accolades est strictement négatif, donc  $\alpha^r \mu_{2r} < 1$  pour  $r$  suffisamment petit. Par

suite il existe  $\beta > 0$ ,  $r > 0$  et  $M > 0$  tels que

$$x_1 > M \text{ et } x_1 > kx_2 \implies E[\varepsilon_{t+2}^{2r} | X_t = (x_1, x_2)] \leq (1 - \beta)x_1^r - \beta.$$

De plus pour  $x_1 \leq M$  on a  $\psi(x_1, x_2) \leq \omega + \alpha M$  et donc d'après (2.12), en utilisant

$$E[\eta_{t+1}^{2r} \mathbb{1}_{\eta_{t+1}^2 > \frac{kx_1}{\psi(x_1, x_2)}}] \leq E[\eta_{t+1}^{2r}] = \mu_{2r},$$

$$E[\varepsilon_{t+2}^{2r} | X_t = (x_1, x_2)] \leq \mu_{2r}\omega^r + \mu_{2r}^2\alpha^r(\omega + \alpha M)^r.$$

Enfin pour  $x_1 \leq kx_2$  on a  $\psi(x_1, x_2) = \omega$  et donc toujours d'après (2.12)

$$E[\varepsilon_{t+2}^{2r} | X_t = (x_1, x_2)] \leq \mu_{2r}\omega^r(1 + \mu_{2r}\alpha^r).$$

On en conclut que (i) et (ii) sont vérifiés, avec  $C = [0, M]^2 \cup \{x_1 \leq kx_2\}$  et  $b = \mu_{2r}\omega^r + \mu_{2r}^2\alpha^r(\omega + \alpha M)^r$ . Le fait que  $C$  soit un petite set résulte des lemmes 2.4.3 et 2.4.4 et de la remarque suivant la définition des petite sets dans le chapitre introductif. On en déduit que le processus  $(X_t)$  est géométriquement ergodique, et par suite que  $(\varepsilon_t)$  est géométriquement ergodique.

Pour montrer l'existence du moment d'ordre  $2r$ , on utilise la Remarque 1 suivant le Théorème 1.5. D'après ce qui précède, la chaîne  $X_{2t}$  est uniformément ergodique, de mesure invariante  $\pi_2$ , et on a  $E_{\pi_2}\{g(X_t)\} = E_{\pi_2}(\varepsilon_{2t}^2) < \infty$ . Puisque  $(X_t)$  est uniformément ergodique, sa mesure invariante  $\pi$  coïncide nécessairement avec  $\pi_2$  et  $E_{\pi}(\varepsilon_t^{2r})$  est une constante finie indépendante de  $t$ .

□

**Exemples :** La région de stationnarité stricte peut être explicitée pour ces lois particulières. Avec une loi normale centrée réduite pour  $\eta_t$  la condition s'écrit :  $0 \leq \alpha < 3.56$ .

Si  $\eta_t$  suit une loi uniforme symétrique centrée de variance 1, c'est-à-dire la loi uniforme sur  $[-\sqrt{3}, \sqrt{3}]$  on a :  $0 \leq \alpha < e^2/3 = 2.46$ .

## 2.6 Stationnarité au second-ordre

Dans le résultat suivant nous établissons une condition d'existence d'une solution strictement stationnaire possédant un moment d'ordre 2.

**Théorème 2.6.1** *Soit  $E(\eta_t^{2m}) = \mu_{2m}$ , pour  $m > 0$ . Alors sous l'hypothèse **H**, si  $k > 0$  et*

$$0 \leq \alpha < \max_{m \in \{1, 2, \dots\}} \left( \frac{k^{m-1}}{\mu_{2m}} \right)^{\frac{1}{m+1}}, \quad (2.13)$$

le processus  $(\varepsilon_t)_{t \geq 0}$  solution de (2.2) est géométriquement ergodique. De plus on a  $E_\pi(\varepsilon_t^2) < \infty$ , où  $\pi$  désigne la loi invariante de  $(\varepsilon_t)$ .

**Remarques : 1.** Si  $k \leq 1$ , la borne supérieure (*i.e.* le membre de droite dans (2.13)) est 1 (obtenu pour  $m = 1$ ), quelle que soit la distribution du processus i.i.d.  $(\eta_t)$ . En particulier pour  $k = 0$  on obtient (avec la convention que  $0^0 = 1$ ) la condition de stationnarité classique pour le modèle ARCH(1) standard, *i.e.*  $\alpha < 1$ .

**2.** Soit  $m$  réalisant le maximum dans (2.13). Un tel  $m$  existe toujours car  $E(\eta_t^2) = \mu_2 = 1$  par hypothèse, donc en utilisant Hölder,  $1 \leq \mu_{2m} \leq \infty$ . Le terme de droite dans (2.13) est donc toujours strictement inférieur à  $\max(1, k)$ . Si  $k > 1$ , la borne supérieure est donc toujours strictement inférieure à  $k$  et supérieure ou égale à 1 (valeur obtenue pour  $m = 1$ ). On obtient donc une zone de stationnarité au second ordre plus large que pour l'ARCH(1) standard.

**3.** Si la distribution de  $(\eta_t^2)$  est la masse de Dirac en 1, on trouve comme bonne supérieure  $\max(1, k)$ , ce qui confirme la borne obtenue dans le cas déterministe.

4. Pour des grandes valeurs de  $k$ , la condition d'ergodicité (2.13) peut se révéler préférable à (2.10). Par exemple pour  $m = 2$ , le second membre de (2.13) est strictement plus grand que celui de (2.10) dès que  $k > \mu_4 \exp\{-3E(\log \eta_t^2)\}$ .

**Preuve :** Soit  $g(x) = g(x_1, x_2) = x_1$ . Comme pour la preuve du Théorème 2.5.1 il s'agit de trouver un petite set  $C$  et des constantes tels que (i) et (ii) soient vérifiées. L'espérance conditionnelle en (i) et (ii) s'écrit

$$\begin{aligned} E[\varepsilon_{t+2}^2 | X_t = (x_1, x_2)] &= E[\eta_{t+2}^2 \psi\{\psi(x_1, x_2)\eta_{t+1}^2, x_1\} | X_t = (x_1, x_2)] \\ &= E[\psi\{\psi(x_1, x_2)\eta_{t+1}^2, x_1\}] \\ &= \omega + \alpha\psi(x_1, x_2)E[\eta_{t+1}^2 \mathbb{1}_{\eta_{t+1}^2 > \frac{kx_1}{\psi(x_1, x_2)}}] \end{aligned} \quad (2.14)$$

Par les inégalités de Hölder et Markov on a, pour  $m \geq 1$

$$\begin{aligned} E(\eta_{t+1}^2 \mathbb{1}_{\eta_{t+1}^2 > \frac{kx_1}{\psi(x_1, x_2)}}) &\leq \{E(\eta_{t+1}^{2m})\}^{\frac{1}{m}} \left\{E(\mathbb{1}_{\eta_{t+1}^2 > \frac{kx_1}{\psi(x_1, x_2)}})\right\}^{\frac{m-1}{m}} \\ &= \{E(\eta_{t+1}^{2m})\}^{\frac{1}{m}} \left\{P\left[\eta_{t+1}^{2m} > \left(\frac{kx_1}{\psi(x_1, x_2)}\right)^m\right]\right\}^{\frac{m-1}{m}} \\ &\leq \{E(\eta_{t+1}^{2m})\}^{\frac{1}{m}} \left\{\frac{E(\eta_{t+1}^{2m})\psi(x_1, x_2)^m}{(kx_1)^m}\right\}^{\frac{m-1}{m}} \\ &= \mu_{2m} \left\{\frac{\psi(x_1, x_2)}{kx_1}\right\}^{m-1}. \end{aligned}$$

Donc pour  $x_1 > kx_2$ , d'après (2.14)

$$E[\varepsilon_{t+2}^2 | X_t = (x_1, x_2)] \leq \omega + \alpha(\omega + \alpha x_1)\mu_{2m} \left\{\frac{\omega + \alpha x_1}{kx_1}\right\}^{m-1}.$$

Lorsque  $x_1 \rightarrow +\infty$ , le terme de droite dans cette inégalité est équivalent à :

$$\left\{\alpha \left(\frac{\mu_{2m}}{k^{m-1}}\right)^{1/m+1}\right\}^{m+1} x_1 \leq (1 - \beta)x_1 - \beta$$

pour une constante  $\beta > 0$  suffisamment petite, le terme entre accolades étant strictement inférieur à 1. Donc il existe  $M > 0$  tel que

$$x_1 > M \text{ et } x_1 > kx_2 \implies E[\varepsilon_{t+2}^2 | X_t = (x_1, x_2)] \leq (1 - \beta)x_1 - \beta.$$

De plus pour  $x_1 \leq M$  on a  $\psi(x_1, x_2) \leq \omega + \alpha M$  et donc d'après (2.14), en utilisant  $E[\eta_{t+1}^2 \mathbf{1}_{\eta_{t+1}^2 > \frac{kx_1}{\psi(x_1, x_2)}}] \leq E[\eta_{t+1}^2] = 1$ ,

$$E[\varepsilon_{t+2}^2 | X_t = (x_1, x_2)] \leq \omega + \alpha(\omega + \alpha M).$$

Enfin pour  $x_1 \leq kx_2$  on a  $\psi(x_1, x_2) = \omega$  et donc d'après (2.14)

$$E[\varepsilon_{t+2}^2 | X_t = (x_1, x_2)] \leq \omega(1 + \alpha).$$

On en conclut que (i) et (ii) sont vérifiés, avec  $C = [0, M]^2 \cup \{x_1 \leq kx_2\}$  et  $b = \omega + \alpha(\omega + \alpha M)$ . On en déduit que le processus  $(X_t)$  est géométriquement ergodique, et par suite que  $(\varepsilon_t)$  est géométriquement ergodique. Comme dans la preuve du Théorème 2.5.1 on en déduit également que  $E_\pi(\varepsilon_t^2)$  est une constante finie indépendante de  $t$ .

□

Pour certaines lois classiques sur  $\eta_t$  la condition de stationnarité peut être déterminée de manière plus explicite.

**Exemple 1 :** Prenons pour loi de  $\eta_t$  une uniforme symétrique centrée de variance 1, c'est-à-dire la loi uniforme sur  $[-\sqrt{3}, \sqrt{3}]$ . Alors  $\mu_{2m} = \frac{3^m}{2m+1}$  et la condition (2.13) s'écrit :

$$0 \leq \alpha < \max_{m \in \{1, 2, \dots\}} \left( \frac{(2m+1)k^{m-1}}{3^m} \right)^{\frac{1}{m+1}} := \max_{m \in \{1, 2, \dots\}} h(m).$$

On a, pour  $m > 0$

$$\frac{h(m)}{h(m+1)} = \left\{ \frac{(2m+1)^{m+2} 3}{(2m+3)^{m+1} k^2} \right\}^{\frac{1}{(m+1)(m+2)}}.$$

La suite apparaissant en facteur de  $3/k^2$  dans l'accolade est croissante. Le maximum dans (2.13) est donc atteint pour

$$m_0 = \min_{m \in \{1, 2, \dots\}} \left\{ m : \frac{k^2}{3} \leq \frac{(2m+1)^{m+2}}{(2m+3)^{m+1}} \right\}$$

la contrainte de stationnarité s'écrivant alors :

$$0 < \alpha < \left( \frac{k^{m_0-1} (2m_0+1)}{3^{m_0}} \right)^{\frac{1}{m_0+1}}.$$

La Table 2.1 donne pour les premières valeurs de  $k$  les valeurs numériques de  $\alpha$  conduisant à des solutions ergodiques. On constate que pour des valeurs de  $k$  assez grandes, les conditions de stationnarité stricte peuvent être beaucoup moins contraignantes sur  $\alpha$  que pour un ARCH(1) classique.

**Exemple 2 :** On choisit la loi gaussienne standard comme distribution de  $(\eta_t)$ . On a ici  $\mu_{2m} = \frac{(2m)!}{2^m m!}$ . Il est facile de vérifier que le maximum dans (2.13) est atteint pour

$$m_0 = \min_{m \in \{2, 3, \dots\}} \left\{ m : k < \left( \frac{(2m-1)^m}{\mu_{2(m-1)}} \right)^{1/2} \right\} - 1$$

la contrainte de stationnarité s'écrivant alors :

$$0 < \alpha < \left( \frac{k^{m_0-1} 2^{m_0} m_0!}{(2m_0)!} \right)^{\frac{1}{m_0+1}}.$$

Pour  $k \geq 3$  des valeurs de  $\alpha$  supérieures à 1 peuvent conduire à des solutions stationnaires, comme le montre la Table 2.2 suivante. Les différences par rapport à

TAB. 2.1 – Contraintes de stationnarité pour la loi uniforme sur  $[-\sqrt{3}, \sqrt{3}]$  en fonction des valeurs de  $k$ . La dernière colonne donne la valeur maximale de  $\alpha$  pour la valeur de  $k$  correspondant à la borne supérieure de l'intervalle.

$k$	$\alpha$	$\alpha_{max}$
$[0, 1.8[$	$[0, 1[$	1
$[1.8, 2.338[$	$[0, \{\frac{5k}{9}\}^{1/3}[$	1.091
$[2.338, 2.772[$	$[0, \{\frac{7k^2}{3^3}\}^{1/4}[$	1.188
$[2.772, 3.146[$	$[0, \{\frac{9k^3}{3^4}\}^{1/5}[$	1.282
$[3.146, 3.480[$	$[0, \{\frac{11k^4}{3^5}\}^{1/6}[$	1.371
$[3.480, 3.785[$	$[0, \{\frac{13k^5}{3^6}\}^{1/7}[$	1.456
$[3.785, 4.066[$	$[0, \{\frac{15k^6}{3^7}\}^{1/8}[$	1.536
$[4.066, 4.329[$	$[0, \{\frac{17k^7}{3^8}\}^{1/9}[$	1.613
$[4.329, 4.577[$	$[0, \{\frac{19k^8}{3^9}\}^{1/10}[$	1.686

TAB. 2.2 – Contraintes de stationnarité pour la loi normale standard en fonction des valeurs de  $k$ . La dernière colonne donne la valeur maximale de  $\alpha$  pour la valeur de  $k$  correspondant à la borne supérieure de l'intervalle.

$k$	$\alpha$	$\alpha_{max}$
$[0, 3[$	$[0, 1[$	1
$[3, 6.455[$	$[0, \{\frac{k}{\mu_4}\}^{1/3}[$	1.291
$[6.455, 12.652[$	$[0, \{\frac{k^2}{\mu_6}\}^{1/4}[$	1.807
$[12.652, 23.714[$	$[0, \{\frac{k^3}{\mu_8}\}^{1/5}[$	2.635
$[23.714, 43.297[$	$[0, \{\frac{k^4}{\mu_{10}}\}^{1/6}[$	3.936
$[43.297, 77.694[$	$[0, \{\frac{k^5}{\mu_{12}}\}^{1/7}[$	5.976
$[77.694, 137.715[$	$[0, \{\frac{k^6}{\mu_{14}}\}^{1/8}[$	9.181

l'ARCH(1) classique sont évidemment plus marquées que dans l'exemple précédent. Des valeurs de  $k$  encore plus grandes conduiraient à une condition plus large que  $\alpha < 3.56$  obtenue à partir de la condition (2.10).

Il est facile d'obtenir une condition nécessaire de stationnarité au second-ordre. En effet, si  $(\varepsilon_t)$  est une solution stationnaire au second-ordre de (2.2), on a

$$\begin{aligned} E(\varepsilon_t^2) &= E\{\eta_t^2(\omega + \alpha\varepsilon_{t-1}^2\mathbb{1}_{\varepsilon_{t-1}^2 > k\varepsilon_{t-2}^2})\} \\ &= \omega + \alpha E\{\varepsilon_{t-1}^2\mathbb{1}_{\varepsilon_{t-1}^2 > k\varepsilon_{t-2}^2}\} \\ &= \omega + \alpha E(\varepsilon_t^2) - \alpha E\{\varepsilon_{t-1}^2\mathbb{1}_{\varepsilon_{t-1}^2 \leq k\varepsilon_{t-2}^2}\} \\ &\geq \omega + \alpha E(\varepsilon_t^2) - \alpha k E(\varepsilon_{t-2}^2), \end{aligned}$$

soit

$$\{1 - \alpha(1 - k)\}E(\varepsilon_t^2) \geq \omega.$$

Remarquons que dans ce qui précède, la troisième égalité implique

$$(1 - \alpha)E(\varepsilon_t^2) \leq \omega.$$

On en déduit le résultat suivant.

**Lemme 2.6.1** *Une condition nécessaire d'existence d'une solution stationnaire au second ordre de (2.2) est*

$$\alpha(1 - k) < 1. \tag{2.15}$$

On a alors

$$E(\varepsilon_t^2) \geq \frac{\omega}{1 - \alpha(1 - k)}.$$

Si  $\alpha < 1$  on a de plus  $E(\varepsilon_t^2) \leq \frac{\omega}{1 - \alpha}$ .

**Remarque.** Dans le cas  $k = 0$ , cet encadrement fournit la valeur de  $E(\varepsilon_t^2)$  pour un ARCH(1) standard. La condition (2.15) n'est intéressante que si  $k < 1$  et particulièrement si  $k$  est proche de 0, c'est à dire au voisinage de l'ARCH(1) standard.

## 2.7 Moments d'ordre supérieur

Nous avons vu dans la partie précédente que (2.13) est une condition suffisante de stationnarité stricte et au second-ordre pour  $(\varepsilon_t)$ . Des conditions d'existence des moments d'ordre supérieur peuvent être obtenues à partir de critères de Foster-Lyapounov. On a le résultat suivant :

**Théorème 2.7.1** *Sous l'hypothèse  $\mathbf{H}$ , si  $k > 0$  et si pour un entier  $r$*

$$0 \leq \alpha < \max_{m \in \{1, 2, \dots\}} \left( \frac{k^{m-1}}{\mu_{2r} \mu_{2m}^{1-\frac{1}{m}} \mu_{2mr}^{\frac{1}{m}}} \right)^{\frac{1}{2r+m-1}}, \quad (2.16)$$

le processus  $(\varepsilon_t)_{t \geq 0}$  solution de (2.2) est géométriquement ergodique. De plus on a  $E_\pi(\varepsilon_t^{2r}) < \infty$ , où  $\pi$  désigne la loi invariante de  $(\varepsilon_t)$ .

**Remarques :** 1. Si  $k \leq 1$ , la borne supérieure obtenue pour  $m = 1$  et la contrainte devient

$$\mu_{2r} \alpha^r < 1 \quad (2.17)$$

comme pour le modèle ARCH(1) standard.

2. Si  $k > 1$ , la borne supérieure est toujours strictement inférieure à  $k$  et supérieure ou égale à 1 (valeur obtenue pour  $m = 1$ ). Si la distribution de  $(\eta_t^2)$  est la masse de Dirac en 1, on retrouve comme borne supérieure  $\max(1, k)$ .

**Preuve :** Si  $\mu_{2r} = \infty$  le terme de droite dans (2.16) vaut zéro et l'ensemble des valeurs de  $\alpha$  satisfaisant la contrainte est vide. Supposons donc  $\mu_{2r} < \infty$ .

Nous procédons comme dans la preuve des Théorèmes 2.5.1 et 2.6.1, en montrant les inégalités (i) et (ii) avec  $g(x) = g(x_1, x_2) = x_1^r$ .

Pour les mêmes raisons que précédemment, le membre de droite dans (2.16) est strictement inférieur à  $\max(1, k)$ . Soit  $m$  réalisant le maximum dans (2.13).

En procédant comme pour la preuve du Théorème 2.6.1 on obtient

$$\begin{aligned}
& E[\varepsilon_{t+2}^{2r} | X_t = (x_1, x_2)] \\
&= E[\eta_{t+2}^{2r} \psi\{\psi(x_1, x_2)\eta_{t+1}^2, x_1\}^r | X_{t-1} = (x_1, x_2)] \\
&= \mu_{2r} E \left\{ \omega + \alpha \psi(x_1, x_2) \eta_{t+1}^2 \mathbb{1}_{\eta_{t+1}^2 > \frac{kx_1}{\psi(x_1, x_2)}} \right\}^r \\
&= \mu_{2r} \sum_{s=0}^r \binom{s}{r} \omega^{r-s} \alpha^s \psi(x_1, x_2)^s E[\eta_{t+1}^{2s} \mathbb{1}_{\eta_{t+1}^2 > \frac{kx_1}{\psi(x_1, x_2)}}] \quad (2.18)
\end{aligned}$$

Par les inégalités de Hölder et Markov on a, pour  $m \geq 1$

$$\begin{aligned}
E(\eta_{t+1}^{2s} \mathbb{1}_{\eta_{t+1}^2 > \frac{kx_1}{\psi(x_1, x_2)}}) &\leq \{E(\eta_{t+1}^{2ms})\}^{\frac{1}{m}} \left\{ P \left[ \eta_{t+1}^{2m} > \left( \frac{kx_1}{\psi(x_1, x_2)} \right)^m \right] \right\}^{\frac{m-1}{m}} \\
&\leq \{E(\eta_{t+1}^{2ms})\}^{\frac{1}{m}} \left\{ \frac{E(\eta_{t+1}^{2m}) \psi(x_1, x_2)^m}{(kx_1)^m} \right\}^{\frac{m-1}{m}} \\
&= \mu_{2ms}^{\frac{1}{m}} \mu_{2m}^{\frac{m-1}{m}} \left\{ \frac{\psi(x_1, x_2)}{kx_1} \right\}^{m-1}. \quad (2.19)
\end{aligned}$$

Lorsque  $x_1 > kx_2$  et  $x_1 \rightarrow +\infty$ , (2.18) est donc majoré par un terme équivalent à :

$$\begin{aligned}
\mu_{2r} \alpha^{2r} \mu_{2mr}^{\frac{1}{m}} \mu_{2m}^{\frac{m-1}{m}} \left\{ \frac{\alpha}{k} \right\}^{m-1} x_1^r &= \left\{ \alpha \frac{\mu_{2r}^{\frac{1}{2r+m-1}} \mu_{2mr}^{\frac{1}{m(2r+m-1)}} \mu_{2m}^{\frac{m-1}{m(2r+m-1)}}}{k^{\frac{m-1}{2r+m-1}}} \right\}^{2r+m-1} x_1^r \\
&\leq (1 - \beta) x_1^r - \beta
\end{aligned}$$

pour une constante  $\beta > 0$  suffisamment petite, le terme entre accolades étant strictement inférieur à 1. Donc il existe  $M > 0$  tel que

$$x_1 > M \text{ et } x_1 > kx_2 \implies E[\varepsilon_{t+2}^{2r} | X_t = (x_1, x_2)] \leq (1 - \beta)x_1^r - \beta.$$

De plus pour  $x_1 \leq M$  on a  $\psi(x_1, x_2) \leq \omega + \alpha M$  et donc d'après (2.18),

$$E[\varepsilon_{t+2}^2 | X_t = (x_1, x_2)] \leq \mu_{2r} \sum_{s=0}^r \binom{s}{r} \omega^{r-s} \alpha^s (\omega + \alpha M)^s = \mu_{2r} \{\omega + \alpha(\omega + \alpha M)\}^r.$$

Enfin pour  $x_1 \leq kx_2$  on a  $\psi(x_1, x_2) = \omega$  et donc d'après (2.18)

$$E[\varepsilon_{t+2}^2 | X_t = (x_1, x_2)] \leq \mu_{2r} \{\omega + \alpha\omega\}^r.$$

On en conclut que (i) et (ii) sont vérifiés, avec  $C = [0, M]^2 \cup \{x_1 \leq kx_2\}$  et  $b = \mu_{2r} \{\omega + \alpha(\omega + \alpha M)\}^r$ . On en déduit que le processus  $(\varepsilon_t)$  est géométriquement ergodique et qu'il admet un moment d'ordre  $2r$ .

□

Reprenons les exemples de la partie précédente.

**Exemple 1 :** Pour la loi uniforme sur  $[-\sqrt{3}, \sqrt{3}]$  des calculs analogues à ceux effectués précédemment montrent que la borne supérieure de la condition (2.16) est atteinte pour

$$m_0 = \min_{m \in \{1, 2, \dots\}} \left\{ m : \ln \left( \frac{k^{2r}}{2r+1} \right) < \frac{(m-1)(2r+m)}{m} \ln(2m+1) - \frac{m(2r+m-1)}{m+1} \ln(2m+3) + \frac{2r+m}{m} \ln(2mr+1) - \frac{2r+m-1}{m+1} \ln(2(m+1)r+1) \right\}.$$

A titre illustratif, la Table 2.3 donne pour les premières zones de valeurs de  $k$  les contraintes sur  $\alpha$  assurant l'existence du moment d'ordre 4 ( $r = 2$ ). Le premier intervalle pour  $k$  conduit à  $m_0 = 1$ . Pour chacun des intervalles suivants  $m_0$  s'accroît d'une unité.

**Exemple 2 :** Dans le cas de la loi  $\mathcal{N}(0, 1)$ , la Table 2.4 fournit les valeurs de  $\alpha$  compatibles avec l'existence d'un moment d'ordre 4, pour différentes plages de valeurs de  $k$ . Les valeurs de  $m$  conduisant au maximum dans (2.16) ont été obtenues numériquement pour les différentes valeurs de  $k$ . Pour  $k < 3.416$ , le maximum est atteint pour  $m = 1$  et la contrainte est celle d'un modèle ARCH(1) standard ( $3\alpha^2 < 1$ ). Lorsque  $k$  augmente, le maximum est atteint pour des  $m > 1$  et des valeurs de plus en plus grandes de  $\alpha$  sont autorisées. Des tables analogues peuvent être construites pour toute valeur de  $r$ .

Afin d'obtenir une condition nécessaire, soit  $(\varepsilon_t)$  une solution de (2.2) admettant un moment d'ordre  $2r$ . On a

$$\begin{aligned} E(\varepsilon_t^{2r}) &\geq \mu_{2r}[\omega^r + \alpha^r E\{\varepsilon_{t-1}^{2r} \mathbf{1}_{\varepsilon_{t-1}^2 > k\varepsilon_{t-2}^2}\}] \\ &= \mu_{2r}[\omega^r + \alpha^r E(\varepsilon_t^{2r}) - \alpha^r E\{\varepsilon_{t-1}^{2r} \mathbf{1}_{\varepsilon_{t-1}^2 \leq k\varepsilon_{t-2}^2}\}] \\ &\geq \mu_{2r}[\omega^r + \alpha^r E(\varepsilon_t^{2r}) - \alpha^r k^r E(\varepsilon_{t-2}^{2r})]. \end{aligned}$$

Par suite

$$\{1 - \mu_{2r}\alpha^r(1 - k^r)\}E(\varepsilon_t^{2r}) \geq \mu_{2r}\omega^r.$$

On en déduit le résultat suivant.

TAB. 2.3 – Contraintes d'existence du moment d'ordre 4 pour la loi uniforme sur  $[-\sqrt{3}, \sqrt{3}]$  en fonction des valeurs de  $k$ . La dernière colonne donne la valeur maximale de  $\alpha$  pour la valeur de  $k$  correspondant à la borne supérieure de l'intervalle.

$k$	$\alpha$	$\alpha_{max}$
$[0, 1.667[$	$[0, \frac{1}{3}5^{1/2}[$	0.745
$[1.667, 1.763[$	$[0, \frac{1}{3}k^{1/5}9^{1/10}5^{3/10}\{[$	0.754
$[1.763, 1.886[$	$[0, \frac{1}{3}k^{1/3}5^{1/6}7^{1/9}13^{1/18}\{[$	0.771
$[1.886, 1.998[$	$[0, \frac{1}{3}k^{3/7}5^{1/7}9^{3/28}17^{1/28}[$	0.790
$[1.998, 2.097[$	$[0, \frac{1}{3}k^{1/2}5^{1/8}11^{1/10}21^{1/40}[$	0.810
$[2.097, 2.186[$	$[0, \frac{1}{3}k^{5/9}5^{1/9}13^{5/54}25^{1/54}[$	0.828
$[2.186, 2.266[$	$[0, \frac{1}{3}k^{3/5}5^{1/10}15^{3/28}29^{1/56}[$	0.908
$[2.266, 2.339[$	$[0, \frac{1}{3}k^{7/11}5^{1/11}17^{7/72}33^{1/72}[$	0.916
$[2.339, 2.407[$	$[0, \frac{1}{3}k^{2/3}5^{1/12}19^{4/45}37^{1/90}[$	0.926

TAB. 2.4 – Contraintes d'existence du moment d'ordre 4 pour la loi normale standard en fonction des valeurs de  $k$ . La dernière colonne donne la valeur maximale de  $\alpha$  pour la valeur de  $k$  correspondant à la borne supérieure de l'intervalle.

$k$	$\alpha$	$\alpha_{max}$
$[0, 3.416[$	$[0, \frac{1}{\mu_4^{1/2}}[$	0.577
$[3.416, 4.579[$	$[0, \{\frac{k}{\mu_4^{3/2} \mu_8^{1/2}}\}^{1/5}[$	0.612
$[4.579, 6.373[$	$[0, \{\frac{k^2}{\mu_4 \mu_6^{2/3} \mu_{12}^{1/3}}\}^{1/6}[$	0.684
$[6.373, 8.846[$	$[0, \{\frac{k^3}{\mu_4 \mu_8^{3/4} \mu_{16}^{1/4}}\}^{1/7}[$	0.787
$[8.846, 12.183[$	$[0, \{\frac{k^4}{\mu_4 \mu_{10}^{4/5} \mu_{20}^{1/5}}\}^{1/8}[$	0.923
$[12.183, 16.656[$	$[0, \{\frac{k^5}{\mu_4 \mu_{12}^{5/6} \mu_{24}^{1/6}}\}^{1/9}[$	1.098
$[16.656, 22.626[$	$[0, \{\frac{k^6}{\mu_4 \mu_{14}^{6/7} \mu_{28}^{1/7}}\}^{1/10}[$	1.320
$[22.626, 30.571[$	$[0, \{\frac{k^7}{\mu_4 \mu_{16}^{7/8} \mu_{32}^{1/8}}\}^{1/11}[$	1.599
$[30.571, 41.122[$	$[0, \{\frac{k^8}{\mu_4 \mu_{18}^{8/9} \mu_{36}^{1/9}}\}^{1/12}[$	1.948

**Lemme 2.7.1** *Une condition nécessaire d'existence d'une solution admettant un moment d'ordre  $2r$ , où  $r$  est un entier positif, est*

$$\mu_{2r}\alpha^r(1 - k^r) < 1. \quad (2.20)$$

*On a alors*

$$E(\varepsilon_t^{2r}) \geq \frac{\mu_{2r}\omega^r}{1 - \mu_{2r}\alpha^r(1 - k^r)}.$$

**Remarque.** Dans le cas  $k = 0$ , (2.20) coïncide avec (2.17) et fournit la condition nécessaire et suffisante d'existence de  $E(\varepsilon_t^{2r})$  pour un ARCH(1) standard.

# Bibliographie

- [1] **Bollerslev, T. (1986)**, *Generalized Autoregressive Conditional Heteroskedasticity*, Journal of Econometrics, 31, pp. 307-327.
- [2] **Bougerol, P., et N. Picard (1992b)**, *Stationnarity of GARCH Processus and of Some Non-Negative Time Series*, Journal of Econometrics, 52, 115-127.
- [3] **Engle, R. F. (1982)**, *Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of U.K Inflation*, Econometrica, 50, pp.987-1008.
- [4] **Feigin, Paul D. et Tweedie Richard L. (1985)**, *Random Coefficient Autoregressive Processes : A Markov Chain Analysis of Stationarity and Finiteness of Moments*, Journal of Times Series Analysis, Vol. 6, No. 1.
- [5] **Francq, C., Roussignol, M. et J.-M. Zakoïan (2001)** *Conditional Heteroskedasticity Driven by Hidden Markov Chains*, Journal of Time Series Analysis,
- [6] **Hamilton, J. D. et R. Susmel (1994)**, *Autoregressive Conditional Heteroskedasticity and Changes in Regime*, Journal of Econometrics, 64 (1994), 307-333.
- [7] **Nelson, Daniel B. (1991)**, *Conditional Heteroskedasticity in Asset Returns : A New Approach*, Economica, Vol. 59, No. 2, 347-370.

- [8] **Petrucelli, J. D. et Woolford, S. W. (1984)**, *A Threshold AR(1) Model*, J. Appl. Prob. 21, 270-286.
- [9] **Rabemananjara R. et J.M. Zakoïan (1993)**, *Threshold ARCH Models and Asymmetries in Volatility*, Journal of Applied Econometrics, 8, 31-49.
- [10] **Revuz D. (1975)**, *Markov Chains*, North-Holland.
- [11] **Tweedie, R. L. (1975)**, *Sufficient Conditions for Ergodicity and Recurrence of Markov Chains on General State Space*, Stochastic Processes and Their Applications, 3, 385-403.
- [12] **Tweedie, R. L. (1983a)**, *Criteria for Rates of Convergence of Markov Chains, with Application to Queueing and Storage Theory*, Probability, Statistics and Analysis édité par J. F. C. Kingman et G. E. H. Reuter. Cambridge University Press : Cambridge.
- [13] **Tweedie, R. L. (1983b)**, *The Existence of Moments for Stationary Markov Chains*, J. Appl. Prob. 20, 191-196.
- [14] **Zakoïan, J.-M. (1990)**, *Threshold Heteroskedastic Models*, Journal of Economics Dynamics and Control, 18 (1994), 931-955.

# Chapitre 3

## Inférence statistique

### 3.1 Introduction

Dans cette partie nous nous intéressons à des problèmes d'estimation sur les modèles ARCH à seuils du type de ceux considérés dans les chapitres précédents.

L'estimation des modèles GARCH( $p, q$ ) standard a donné lieu à une abondante littérature depuis leur introduction par Engle (1982) et Bollerslev (1986). La principale méthode d'estimation est celle du Maximum de Vraisemblance (MV), ou Pseudo-Maximum de Vraisemblance (PMV) lorsque l'hypothèse de normalité du processus i.i.d n'est utilisée que pour construire le critère. Les propriétés asymptotiques de l'estimateur du PMV ont été établies sous des hypothèses de plus en plus faibles portant sur l'espace des paramètres, la vraie valeur du paramètre et le processus i.i.d. Des références récentes sur ce thème sont Lee et Hansen (1994), Lumsdaine (1996), Boussama (1998, 2000), Berkes, Horváth et Kokoszka (2003), Francq et Zakoïan (2003).

La nécessité de disposer de valeurs initiales dans la mise en œuvre de la méthode du PMV, ou celle d'obtenir des estimateurs plus simples à calculer et étudier, ou plus robuste à la non-normalité du processus i.i.d, ont conduit des chercheurs à s'intéresser à d'autres méthodes d'estimation. Les propriétés asymptotiques de l'estimateur des moindres carrés ont été étudiées par Weiss (1986). Récemment, Baillie et Chung (2001) ont proposé un estimateur fondé sur l'écart entre autocorrélations théoriques et empiriques du carré des observations. Horváth et Liese (2003) ont introduit une famille d'estimateurs  $L_p$  pondérés pour les modèles ARCH( $q$ ), permettant de réduire les hypothèses de moments.

Nous commencerons par développer une méthode des moments permettant d'obtenir des estimateurs explicites. Nous considérerons ensuite une méthode des moindres carrés. Puis nous étudierons les propriétés théoriques de la méthode du PMV.

## 3.2 Modèle et hypothèses

Dans ce chapitre nous allons considérer l'estimation du modèle ARCH(1) à seuils de la forme

$$\begin{cases} \varepsilon_t = h_t^{1/2} \eta_t, \\ h_t = \omega_0 + \left( \sum_{i=1}^r \alpha_{0i} \mathbb{1}_{Z_{t-1} \in R_i} \right) \varepsilon_{t-1}^2 \end{cases} \quad (3.1)$$

où  $(\eta_t)_{t \in \mathbb{Z}}$  est une suite de variables i.i.d centrées et de variance 1, définies sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$ ;  $Z_{t-1}$  est une variable à valeurs dans  $\mathbb{R}^d$  ( $d \in \{1, 2, \dots\}$ ) appartenant à la tribu  $\mathcal{F}_{t-1}$  engendrée par  $\{\varepsilon_{t-j}, j > 0\}$ ;  $(R_i)_{i=1, \dots, r}$  est une partition de  $\mathbb{R}^d$ ;  $\omega > 0$  et pour  $i = 1, \dots, r$ ,  $\alpha_i \geq 0$ . On suppose que  $\eta_t$  est indépendante de  $\{\varepsilon_{t-j}, j > 0\}$ . On note  $\mu_{2m} = E(\eta_t^{2m})$ .

Le modèle ARCH(1) standard est obtenu lorsque les  $\alpha_{0i}$  sont égaux ou lorsque  $r = 1$ .

Le modèle TAR(1) introduit par Rabemananjara et Zakoïan (1993), correspond à  $Z_{t-1} = \varepsilon_{t-1}$ ,  $r = 2$ ,  $R_1 = \mathbb{R}^+$  et  $R_2 = \mathbb{R}^-$ . Le modèle ARCH(1) à seuil étudié dans le chapitre 2 est obtenu pour  $Z_{t-1} = \varepsilon_{t-1}^2/\varepsilon_{t-2}^2$ ,  $r = 2$ ,  $R_1 = [0, k]$  et  $R_2 = ]k, +\infty[$ . La formulation plus générale présentée ici permet de prendre en compte des dynamiques plus complexes, l'impact de  $\varepsilon_{t-1}^2$  sur la volatilité de la date  $t$  dépendant de la position (absolue ou relative) de certaines variables antérieures à la date  $t$ .

On observe une réalisation de taille  $n$ ,  $(\varepsilon_1, \dots, \varepsilon_n)$  du processus  $\varepsilon_t$ . On attribue la valeur 0 aux variables antérieures à la date 1. Les variables  $Z_{t-1}$  sont des fonctions connues des  $\varepsilon_{t-j}$ . L'entier  $r$  et la partition  $(R_i)$  seront également supposés connus.

Le vecteur des paramètres inconnus est

$$\theta = (\theta_1, \dots, \theta_{r+1})' := (\omega, \alpha_1, \dots, \alpha_r)' \in \Theta \subset ]0, \infty[ \times [0, +\infty[^r,$$

où  $\Theta$  est l'espace des paramètres. La vraie valeur du paramètre est inconnue et notée  $\theta_0 = (\omega_0, \alpha_{01}, \dots, \alpha_{0r})$ .

Nous serons amenés à considérer, dans un premier temps, les hypothèses suivantes :

**A1** : Le processus  $(\varepsilon_t)$  est une solution non anticipative strictement stationnaire et ergodique du modèle (3.1).

**A2** : pour tout  $i \in \{1, \dots, r\}$ ,  $P_{\theta_0}[Z_t \in R_i] > 0$ .

**A3** : Il n'existe aucun ensemble  $\wedge$  de cardinal  $r$  tel que  $P(\eta_t^2 \in \wedge) = 1$ .

**A4** :  $E_{\theta_0}(\varepsilon_t^4) < +\infty$ .

**Remarque.** Des conditions assurant la validité de l'hypothèse **A1** ont été obtenues pour les modèles spécifiques cités plus haut (voir les références correspondantes ainsi que le chapitre 2). L'hypothèse **A2** est nécessaire pour l'identifiabilité des paramètres, car si  $P_{\theta_0}[Z_t \in R_i] = 0$  le coefficient  $\alpha_i$  ne peut être identifié. L'hypothèse

**A3** est requise pour obtenir des lois limites d'estimateurs non dégénérées. La condition **A4**, enfin, est requise, pour la méthode des MCO, mais pourra être affaiblie pour la méthode des moments ou du maximum de vraisemblance.

### 3.3 Méthode des moindres carrés

Dans cette partie nous considérons l'estimation par la méthode des moindres carrés ordinaires (MCO) du modèle (3.1). Supposons que  $(\varepsilon_t^2)$  soit strictement stationnaire et admette des moments jusqu'à l'ordre 4 au moins. On a le modèle AR(1) non linéaire

$$\varepsilon_t^2 = \omega_0 + \left( \sum_{i=1}^r \alpha_{0i} \mathbb{1}_{Z_{t-1} \in R_i} \right) \varepsilon_{t-1}^2 + u_t \quad (3.2)$$

où  $u_t = \varepsilon_t^2 - h_t = (\eta_t^2 - 1)h_t$ . La suite  $(u_t, \mathcal{F}_{t-1})_t$  constitue donc une différence de martingale. Les estimateurs des moindres carrés des composantes de  $\theta$  sont obtenus en minimisant

$$Q_n(\theta) = \sum_{t=1}^n \left\{ \varepsilon_t^2 - \omega - \left( \sum_{i=1}^r \alpha_i \mathbb{1}_{Z_{t-1} \in R_i} \right) \varepsilon_{t-1}^2 \right\}^2.$$

On note  $J(i), i = 1, \dots, r$  l'ensemble des entiers  $\{1 \leq t \leq n : Z_{t-1} \in R_i\}$ ,

$$\hat{m}_i^{(k)} = \frac{1}{n} \sum_{t \in J(i)} \varepsilon_{t-1}^k, \quad \hat{c}_i = \frac{1}{n} \sum_{t \in J(i)} \varepsilon_t^2 \varepsilon_{t-1}^2, \quad i = 1, \dots, r, \quad k = 2, 4.$$

On obtient facilement en dérivant le critère  $Q_n(\theta)$  les estimateurs des MCO suivants :

$$\begin{aligned} \hat{\omega} &= \left\{ \sum_{i=1}^r \hat{m}_i^{(2)} - \frac{\hat{m}_i^{(2)}}{\hat{m}_i^{(4)}} \hat{c}_i \right\} \left\{ 1 - \sum_{i=1}^r \frac{(\hat{m}_i^{(2)})^2}{\hat{m}_i^{(4)}} \right\}^{-1}, \\ \hat{\alpha}_i &= \frac{\hat{c}_i}{\hat{m}_i^{(4)}} - \hat{\omega} \frac{\hat{m}_i^{(2)}}{\hat{m}_i^{(4)}}, \quad i = 1, \dots, r. \end{aligned}$$

On en déduit un estimateur naturel de  $\sigma_0^2 = \text{Var}_{\theta_0}(u_t)$  :

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n \left\{ \varepsilon_t^2 - \hat{\omega} - \left( \sum_{i=1}^r \hat{\alpha}_i \mathbb{1}_{Z_{t-1} \in R_i} \right) \varepsilon_{t-1}^2 \right\}^2.$$

Notons pour  $k \geq 0$ ,

$$m_i^{(k)} = E_{\theta_0}(\varepsilon_{t-1}^k \mathbb{1}_{Z_{t-1} \in R_i}), \quad 1 \leq i \leq r,$$

$$c_i = E_{\theta_0}(\varepsilon_t^2 \varepsilon_{t-1}^2 \mathbb{1}_{Z_{t-1} \in R_i}), \quad 1 \leq i \leq r.$$

On a le résultat suivant.

**Théorème 3.3.1** *Sous les hypothèse **A1-A4**,  $\hat{\omega}$ ,  $\hat{\alpha}_i$ ,  $i = 1, \dots, r$  et  $\hat{\sigma}^2$  sont des estimateurs fortement convergents de  $\omega$ , des  $\alpha_i$  et de  $\sigma^2$  respectivement.*

**Preuve.** Les processus  $(Z_t)$  est ergodique car  $Z_t$  est une fonction mesurable de  $\varepsilon_{t-s}$ ,  $s > 0$  et  $(\varepsilon_t)$  est ergodique d'après **A1**. Par suite, quand  $n \rightarrow \infty$ , pour  $k = 1, \dots, 4$

$$\hat{m}_i^{(k)} = \frac{1}{n} \sum_{t=1}^n \varepsilon_{t-1}^k \mathbb{1}_{Z_{t-1} \in R_i} \rightarrow m_i^{(k)} \quad p.s., \quad 1 \leq i \leq r,$$

$$\hat{c}_i = \frac{1}{n} \sum_{t=1}^n \varepsilon_t^2 \varepsilon_{t-1}^2 \mathbb{1}_{Z_{t-1} \in R_i} \rightarrow c_i \quad p.s., \quad 1 \leq i \leq r.$$

Pour tout  $i$  la variable  $\varepsilon_{t-1}^k \mathbb{1}_{Z_{t-1} \in R_i}$  est p.s. strictement positive, pour  $k = 2$  ou  $k = 4$ , sous l'hypothèse **A2**. Elle n'est pas p.s. constante, sinon  $\eta_t^2$  appartiendrait p.s. à  $\mathcal{F}_{t-1}$ , ce qui est exclu, ou serait p.s. constante, donc égale à 1, ce qui est également exclu par **A3**. Donc  $0 < (m_i^{(2)})^2 < m_i^{(4)} < +\infty$ , en utilisant **A4** pour la dernière inégalité. Cette remarque permet, en écrivant les relations d'orthogonalité

$$E_{\theta_0}(u_t) = E_{\theta_0}(u_t \varepsilon_{t-1}^2) = 0,$$

d'obtenir les relations

$$\omega = \left\{ \sum_{i=1}^r m_i^{(2)} - \frac{m_i^{(2)}}{m_i^{(4)}} c_i \right\} \left\{ 1 - \sum_{i=1}^r \frac{(m_i^{(2)})^2}{m_i^{(4)}} \right\}^{-1},$$

$$\alpha_i = \frac{c_i}{m_i^{(4)}} - \omega \frac{m_i^{(2)}}{m_i^{(4)}}, \quad i = 1, \dots, r.$$

La convergence forte des estimateurs découle trivialement des convergences précédentes.

□

Pour établir la normalité asymptotique de l'estimateur des MCO, nous introduisons les matrices

$$I = \begin{pmatrix} \xi_1^{(2)} & 0 & 0 & \dots & \xi_1^{(1)} \\ 0 & \xi_2^{(2)} & 0 & 0 & \dots & \xi_2^{(1)} \\ \vdots & \vdots & & \ddots & & \\ 0 & 0 & \dots & \xi_r^{(2)} & \xi_r^{(1)} \\ \xi_1^{(1)} & \dots & \dots & \xi_r^{(1)} & \sum_{i=1}^r \xi_i^{(0)} \end{pmatrix}$$

et

$$J = \begin{pmatrix} m_1^{(2)} & m_1^{(4)} & 0 & \dots & 0 \\ m_2^{(2)} & 0 & m_2^{(4)} & 0 & \dots & 0 \\ \vdots & \vdots & & \ddots & & \\ m_r^{(2)} & 0 & \dots & & & m_r^{(4)} \\ 1 & m_1^{(2)} & \dots & \dots & & m_r^{(2)} \end{pmatrix}$$

où pour  $i = 1, \dots, r$  et  $k = 0, 1, 2$

$$\xi_i^{(k)} = E_{\theta_0}(\varepsilon_{t-1}^{2k} \mathbf{1}_{Z_{t-1} \in R_i} h_t^2) = \omega_0^2 m_i^{(2k)} + 2\omega_0 \alpha_{0i} m_i^{(2k+2)} + \alpha_{0i}^2 m_i^{(2k+4)}. \quad (3.3)$$

Remarquons que  $E_{\theta_0}(\varepsilon_t^4) = \mu_4 \sum_{i=1}^r \xi_i^{(0)}$ .

Le résultat suivant établit la normalité asymptotique de l'estimateur des MCO sous des hypothèses de moments plus fortes que pour la convergence. L'inversibilité de  $J$  sera montrée dans la preuve.

**Théorème 3.3.2** *Sous les hypothèse **A1-A3** et si  $E_{\theta_0}(\varepsilon_t^8) < +\infty$ ,*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \Rightarrow \mathcal{N}(0, (\mu_4 - 1)J^{-1}I(J')^{-1}).$$

**Preuve.** A partir de (3.2) on obtient facilement

$$\begin{aligned} \hat{c}_i &= \omega_0 \hat{m}_i^{(2)} + \sum_{i=1}^r \alpha_{0i} \hat{m}_i^{(4)} + \frac{1}{n} \sum_{t=1}^n u_t \varepsilon_{t-1}^2, \quad i = 1, \dots, r, \\ \frac{1}{n} \sum_{t=1}^n \varepsilon_t^2 &= \omega_0 + \sum_{i=1}^r \alpha_{0i} \hat{m}_i^{(2)} + \frac{1}{n} \sum_{t=1}^n u_t. \end{aligned}$$

Par suite,  $\sum_{t=1}^n \varepsilon_t^2/n$  étant asymptotiquement équivalente à  $\sum_{i=1}^r \hat{m}_i^{(2)}$ , on a

$$\begin{aligned}
& \begin{pmatrix} \hat{m}_1^{(2)} & \hat{m}_1^{(4)} & 0 & \dots & 0 \\ \hat{m}_2^{(2)} & 0 & \hat{m}_2^{(4)} & 0 & \dots & 0 \\ \vdots & \vdots & & \ddots & & \\ \hat{m}_r^{(2)} & 0 & \dots & & \hat{m}_r^{(4)} \\ 1 & \hat{m}_1^{(2)} & \dots & \dots & \hat{m}_r^{(2)} \end{pmatrix} \begin{pmatrix} \sqrt{n}(\hat{\omega} - \omega_0) \\ \sqrt{n}(\hat{\alpha}_1 - \alpha_{01}) \\ \vdots \\ \sqrt{n}(\hat{\alpha}_r - \alpha_{0r}) \end{pmatrix} \\
& := \hat{J}\sqrt{n}(\hat{\theta}_n - \theta_0) = \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{t=1}^n u_t \varepsilon_{t-1}^2 \mathbb{1}_{Z_{t-1} \in R_1} \\ \frac{1}{\sqrt{n}} \sum_{t=1}^n u_t \varepsilon_{t-1}^2 \mathbb{1}_{Z_{t-1} \in R_2} \\ \vdots \\ \frac{1}{\sqrt{n}} \sum_{t=1}^n u_t \varepsilon_{t-1}^2 \mathbb{1}_{Z_{t-1} \in R_r} \\ \frac{1}{\sqrt{n}} \sum_{t=1}^n u_t \end{pmatrix} + o_P(1). \quad (3.4)
\end{aligned}$$

Notons  $\underline{\varepsilon}_{t-1}^2 = (\varepsilon_{t-1}^2 \mathbb{1}_{Z_{t-1} \in R_1}, \dots, \varepsilon_{t-1}^2 \mathbb{1}_{Z_{t-1} \in R_r}, 1)'$ . Pour tout vecteur  $\beta = (\beta_1, \dots, \beta_{r+1})'$ , la suite  $(\frac{1}{\sqrt{n}} u_t \beta' \underline{\varepsilon}_{t-1}^2, \mathcal{F}_{t-1})_t$  est une différence de martingale. On a en utilisant  $u_t = (\eta_t^2 - 1)h_t$

$$\begin{aligned}
& \sum_{t=1}^n E_{\theta_0} \left\{ \left( \frac{1}{\sqrt{n}} u_t \beta' \underline{\varepsilon}_{t-1}^2 \right)^2 \mid \mathcal{F}_{t-1} \right\} \\
& = \frac{1}{n} \sum_{t=1}^n E(\eta_t^2 - 1)^2 \left( \omega_0^2 + \sum_{i=1}^r \alpha_{0i}^2 \mathbb{1}_{Z_{t-1} \in R_i} \varepsilon_{t-1}^4 + 2\omega_0 \sum_{i=1}^r \alpha_{0i} \mathbb{1}_{Z_{t-1} \in R_i} \varepsilon_{t-1}^2 \right) \\
& \quad \times \left( \sum_{i=1}^r \beta_i^2 \mathbb{1}_{Z_{t-1} \in R_i} \varepsilon_{t-1}^4 + 2 \sum_{i=1}^r \beta_i \mathbb{1}_{Z_{t-1} \in R_i} \beta_{r+1} \varepsilon_{t-1}^2 + \beta_{r+1}^2 \right) \\
& \rightarrow (\mu_4 - 1) \left\{ \sum_{i=1}^r \beta_i^2 \xi_i^{(2)} + 2\beta_i \beta_{r+1} \xi_i^{(1)} + \beta_{r+1}^2 \xi_i^{(0)} \right\},
\end{aligned}$$

en probabilité quand  $n$  tend vers l'infini. De plus, pour toute constante  $\epsilon > 0$ ,

$$\begin{aligned} & \sum_{t=1}^n E \left\{ \left( \frac{1}{\sqrt{n}} u_t \beta' \underline{\varepsilon}_{t-1}^2 \right)^2 \mathbb{1}_{\{n^{-1/2} |u_t \beta' \underline{\varepsilon}_{t-1}^2| \geq \epsilon\}} \right\} \\ &= E \left\{ (u_t \beta' \underline{\varepsilon}_{t-1}^2)^2 \mathbb{1}_{\{|u_t \beta' \underline{\varepsilon}_{t-1}^2| \geq \sqrt{n} \epsilon\}} \right\} \rightarrow 0, \quad \text{quand } n \rightarrow \infty \end{aligned}$$

car  $\{|u_t \beta' \underline{\varepsilon}_{t-1}^2| \geq \sqrt{n} \epsilon\} \downarrow \emptyset$  et  $E(u_t \beta' \underline{\varepsilon}_{t-1}^2)^2 < \infty$ . Le théorème central limite pour tableaux triangulaires de différences de martingales (voir Billingsley (1995)) permet de déduire de ce qui précède que

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n u_t \beta' \underline{\varepsilon}_{t-1}^2 \Rightarrow \mathcal{N} \left( 0, (\mu_4 - 1) \left\{ \sum_{i=1}^r (\beta_i^2 \xi_i^{(2)} + 2\beta_i \beta_{r+1} \xi_i^{(1)}) + \beta_{r+1}^2 \xi_1^{(0)} \right\} \right)$$

et par suite, en utilisant la propriété de Cramer-Wold, que

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n u_t \underline{\varepsilon}_{t-1}^2 \Rightarrow \mathcal{N}(0, (\mu_4 - 1)I). \quad (3.5)$$

Montrons que cette loi limite est non dégénérée, c'est-à-dire que  $I$  est inversible. Il est clair que les  $\xi_i^{(k)}$  sont non nuls sous l'hypothèse **A2** et que les  $r$  premières colonnes de  $I$  sont linéairement indépendantes. Supposons que la  $r+1$ -ème colonne soit combinaison linéaire des autres. Donc il existe  $(\lambda_1, \dots, \lambda_r)$  tel que :  $\xi_i^{(1)} = \lambda_i \xi_i^{(2)}$ ,  $i = 1, \dots, r$  et  $\sum_{i=1}^r \xi_i^{(0)} = \sum_{i=1}^r \lambda_i \xi_i^{(1)}$ . On en déduit que

$$\sum_{i=1}^r \xi_i^{(0)} = \sum_{i=1}^r \frac{(\xi_i^{(1)})^2}{\xi_i^{(2)}}. \quad (3.6)$$

Or l'inégalité de Schwarz implique que

$$\begin{aligned} (\xi_i^{(1)})^2 &= \{E_{\theta_0} [(\varepsilon_{t-1}^2 \mathbb{1}_{Z_{t-1} \in R_i} h_t) (\mathbb{1}_{Z_{t-1} \in R_i} h_t)]\}^2 \\ &\leq E_{\theta_0} (\varepsilon_{t-1}^4 \mathbb{1}_{Z_{t-1} \in R_i} h_t^2) E_{\theta_0} (\mathbb{1}_{Z_{t-1} \in R_i} h_t^2) = \xi_i^{(2)} \xi_i^{(0)} \end{aligned}$$

L'une de ces inégalités est stricte car sinon il existerait des constantes  $\mu_i$  telles que, p.s.,  $(\varepsilon_{t-1}^4 - \mu_i)\mathbb{1}_{Z_{t-1} \in R_i} h_t^2 = 0$ . On en déduirait en sommant sur  $i$  que la variable  $(\varepsilon_{t-1}^4)$  prendrait p.s. au plus  $r$  valeurs distinctes. Il en serait de même de  $h_t$  et par suite  $\eta_t^2$  prendrait p.s. au plus  $2r$  valeurs distinctes, ce qui est impossible d'après **A3**. Par suite (3.6) est impossible donc  $I$  est bien inversible.

Dans (3.4) la matrice  $\hat{J}$  converge p.s. lorsque  $n$  tend vers l'infini vers  $J$ . On montre que  $J$  est inversible de manière similaire à ce qui a été fait pour  $I$ . Le théorème se déduit de (3.4) et (3.5) par un raisonnement classique.

□

**Remarque :** on obtient un estimateur des coefficients  $\xi_i^{(k)}$  à partir de (3.3) en remplaçant les moments théoriques,  $m_i^{(j)}$ , par les moments empiriques  $\hat{m}_i^{(j)}$ . En utilisant les convergences montrées précédemment, on obtient ainsi des estimateurs convergents p.s. des matrices  $I$  et  $J$ . Si on suppose  $\mu_4$  connu on en déduit un estimateur fortement convergent de la matrice de variance asymptotique.

**Exemples :** 1. Dans le cas où  $r = 1$  on obtient la variance asymptotique de l'estimateur des MCO dans le cas ARCH standard. On a, si  $\mu_8 \alpha^4 < 1$ , en omettant

les indices

$$I = \begin{pmatrix} \xi^{(2)} & \xi^{(1)} \\ \xi^{(1)} & \xi^{(0)} \end{pmatrix}, \quad J = \begin{pmatrix} m^{(2)} & m^{(4)} \\ 1 & m^{(2)} \end{pmatrix},$$

$$m^{(2)} = \frac{\omega}{1-\alpha}, \quad m^{(4)} = \frac{\omega^2(1+\alpha)}{(1-\mu_4\alpha^2)(1-\alpha)} \mu_4,$$

$$m^{(6)} = \frac{\omega^3\{1+2\alpha+\mu_4\alpha^2(2+\alpha)\}}{(1-\mu_6\alpha^3)(1-\mu_4\alpha^2)(1-\alpha)} \mu_6,$$

$$m^{(8)} = \frac{\omega^4\{1+3\alpha+5\mu_4\alpha^2+3(\mu_4+\mu_6)\alpha^3+5\mu_6\alpha^4+\mu_4\mu_6\alpha^5(3+\alpha)\}}{(1-\mu_8\alpha^4)(1-\mu_6\alpha^3)(1-\mu_4\alpha^2)(1-\alpha)} \mu_8,$$

$$\xi^{(k)} = E(\varepsilon_{t-1}^{2k} h_t^2) = \omega^2 m^{(2k)} + 2\alpha\omega m^{(2k+2)} + \alpha^2 m^{(2k+4)}.$$

A titre d'exemple, voici quelques valeurs numériques de cette matrice pour différentes valeurs des paramètres lorsque  $\eta_t$  suit la loi normale (0,1) :

paramètres	$\omega = 1, \alpha = 0.1$	$\omega = 1, \alpha = 0.2$	$\omega = 1, \alpha = 0.3$
$\text{Var}_{as}\{\sqrt{n}(\hat{\theta}_n - \theta_0)\}$	$\begin{pmatrix} 3.976 & -1.850 \\ -1.850 & 2.151 \end{pmatrix}$	$\begin{pmatrix} 8.031 & -5.261 \\ -5.261 & 5.460 \end{pmatrix}$	$\begin{pmatrix} 151.0 & -106.5 \\ -106.5 & 77.6 \end{pmatrix}$

La précision asymptotique du paramètre  $\omega$ , et globalement celle du couple  $\theta$ , devient très faible au voisinage de la frontière d'existence du moment d'ordre 8 ( $\alpha < 0.312$ ). La convergence est cependant obtenue sous existence du moment d'ordre 4 ( $\alpha < 0.577$ ). Pour des valeurs de  $\alpha$  pas trop élevées la méthode peut donc fournir de bonnes valeurs initiales dans l'algorithme de maximisation de la vraisemblance.

2. Dans le cas où  $r = 2$  on obtient

$$I = \begin{pmatrix} \xi_1^{(2)} & 0 & \xi_1^{(1)} \\ 0 & \xi_2^{(2)} & \xi_2^{(1)} \\ \xi_1^{(1)} & \xi_2^{(1)} & \sum_{i=1}^2 \xi_i^{(0)} \end{pmatrix} \quad J = \begin{pmatrix} m_1^{(2)} & m_1^{(4)} & 0 \\ m_2^{(2)} & 0 & m_2^{(4)} \\ 1 & m_1^{(2)} & m_2^{(2)} \end{pmatrix}$$

Les valeurs théoriques des composantes de  $I$  ne peuvent être calculées analytiquement mais elles peuvent être estimées par simulations. A titre illustratif, nous présentons ci-dessous des expériences de Monte-Carlo pour quelques exemples de modèles à deux régimes, avec

$$Z_{t-1} = \varepsilon_{t-1}^2 / \varepsilon_{t-2}^2, \quad R_1 = [0, k], \quad R_2 = ]k, +\infty[.$$

On a effectué  $N = 1000$  simulations de taille  $n = 10000$  du modèle, calculé pour chaque simulation un estimateur  $\hat{\theta}_n$  de  $\theta_0$ . La moyenne sur les  $N$  simulations est notée  $\bar{\theta}_N$ . On donne également la fréquence empirique (sur une simulation) du régime 1, notée  $\hat{p}(R_1)$ . Ces résultats seront comparés ultérieurement avec ceux correspondant à la méthode des moments.

## 3.3.2 - Estimation de la variance asymptotique de l'estimateur des MCO

$k = 1$		
$(\omega_0, \alpha_{01}, \alpha_{02})$	(1, 0.3, 0.5)	(1, 3.0, 0.1)
$\bar{\theta}_N$	(1.062, 0.279, 0.460)	(1.051, 2.924, 0.087)
$\bar{\sigma}_N^2$	15.356	58.942
$\hat{p}(R_1)$	0.490	0.486
$\bar{\Sigma}_N$	$\begin{pmatrix} 0.855 & -1.344 & 0.238 \\ -1.344 & 18.463 & -3.033 \\ 0.238 & -3.033 & 7.230 \end{pmatrix}$	$\begin{pmatrix} 0.717 & -1.463 & 0.263 \\ -1.463 & 64.635 & -7.971 \\ 0.263 & -7.971 & 20.720 \end{pmatrix}$
$k = 3$		
$(\omega_0, \alpha_{01}, \alpha_{02})$	(1, 0.3, 0.5)	(1, 0.5, 0.1)
$\bar{\theta}_N$	(1.019, 0.284, 0.490)	(1.012, 0.482, 0.096)
$\bar{\sigma}_N^2$	8.12	5.013
$\hat{p}(R_1)$	0.679	0.667
$\bar{\Sigma}_N$	$\begin{pmatrix} 1.504 & -1.367 & 0.178 \\ -1.367 & 11.480 & -1.463 \\ 0.178 & -1.464 & 1.561 \end{pmatrix}$	$\begin{pmatrix} 1.259 & -1.116 & 0.253 \\ -1.116 & 6.843 & -1.546 \\ 0.253 & -1.546 & 2.159 \end{pmatrix}$
$k = 5$		
$(\omega_0, \alpha_{01}, \alpha_{02})$	(1, 0.2, 1.1)	(1, 0.7, 0.2)
$\bar{\theta}_N$	(1.138, 0.178, 0.956)	(1.192, 0.570, 0.187)
$\bar{\sigma}_N^2$	76.55	58.00
$\hat{p}(R_1)$	0.745	0.755
$\bar{\Sigma}_N$	$\begin{pmatrix} 1.929 & -1.585 & 0.063 \\ -1.585 & 80.971 & -1.851 \\ 0.063 & -1.851 & 1.355 \end{pmatrix}$	$\begin{pmatrix} 5.924 & -4.231 & 0.108 \\ -4.231 & 63.185 & -1.554 \\ 0.108 & -1.554 & 1.028 \end{pmatrix}$

### 3.4 Méthode des moments

Dans cette partie nous considérons l'estimation par la méthode des moments du modèle (3.1). Supposons à nouveau que  $(\varepsilon_t^2)$  soit strictement stationnaire et admette des moments jusqu'à l'ordre 4 au moins. Les moments et autocovariance usuels du processus  $(\varepsilon_t^2)$  semblent impossibles à calculer analytiquement mais on a les relations suivantes entre les moments de variables tronquées de la forme  $\varepsilon_t^2 \mathbb{1}_{Z_{t-1} \in R_i}$  :

$$\begin{aligned} \forall j = 1, \dots, r, \\ E_{\theta_0}(\varepsilon_t^2 \mathbb{1}_{Z_{t-1} \in R_j}) &= \omega_0 E_{\theta_0}(\mathbb{1}_{Z_{t-1} \in R_j}) + \left\{ \sum_{i=1}^r \alpha_{0i} E_{\theta_0}(\varepsilon_{t-1}^2 \mathbb{1}_{Z_{t-1} \in R_i} \mathbb{1}_{Z_{t-1} \in R_j}) \right\} \\ &= \omega_0 E_{\theta_0}(\mathbb{1}_{Z_{t-1} \in R_j}) + \alpha_{0j} E_{\theta_0}(\varepsilon_{t-1}^2 \mathbb{1}_{Z_{t-1} \in R_j}), \end{aligned} \quad (3.7)$$

$$E_{\theta_0}(\varepsilon_t^2 \varepsilon_{t-1}^2) = \omega_0 E_{\theta_0}(\varepsilon_{t-1}^2) + \left\{ \sum_{i=1}^r \alpha_{0i} E_{\theta_0}(\varepsilon_{t-1}^4 \mathbb{1}_{Z_{t-1} \in R_i}) \right\}, \quad (3.8)$$

soit sous forme vectorielle

$$b_0 = A_0 \theta_0, \quad (3.9)$$

$$b_0 = \begin{pmatrix} m_1^{(2)}(1) \\ m_2^{(2)}(1) \\ \vdots \\ m_r^{(2)}(1) \\ E_{\theta_0}(\varepsilon_t^2 \varepsilon_{t-1}^2) \end{pmatrix}, \quad A_0 = \begin{pmatrix} m_1^{(0)} & m_1^{(2)} & 0 & \dots & 0 \\ m_2^{(0)} & 0 & m_2^{(2)} & 0 & \dots & 0 \\ \vdots & \vdots & & \ddots & & \\ m_r^{(0)} & 0 & \dots & & 0 & m_r^{(2)} \\ E_{\theta_0}(\varepsilon_t^2) & m_1^{(4)} & \dots & \dots & & m_r^{(4)} \end{pmatrix}$$

en posant, pour  $i = 1, \dots, r$ ,  $k = 0, 1, \dots$ ,  $h = 0, 1, \dots$

$$m_i^{(k)}(h) = E_{\theta_0}(\varepsilon_t^k \mathbb{1}_{Z_{t-h} \in R_i})$$

et en remarquant que  $m_i^{(k)}(0) = m_i^{(k)}$ . On a la propriété suivante.

**Lemme 3.4.1** *Sous les hypothèses A1-A4 la matrice  $A_0$  existe et est inversible.*

Il est clair que les  $r$  dernières colonnes de  $A_0$  sont linéairement indépendantes. Supposons que la première colonne, notée  $C_0$ , soit combinaison linéaire des autres, notées  $C_i, i = 1, \dots, r$ . Donc il existe des constantes  $\beta_i$  telles que  $C_0 = \sum_{i=1}^r \beta_i C_i$ . Puisque  $P[Z_t \in R_i] > 0, m_i^{(k)}(h) > 0$  pour  $k$  pair. Donc  $\beta_i = \frac{m_i^{(0)}}{m_i^{(2)}(0)}$ . Par suite, la dernière ligne donne,

$$E_{\theta_0}(\varepsilon_{t-1}^2) = \sum_{i=1}^r \beta_i m_i^{(4)}(0) = \sum_{i=1}^r \frac{m_i^{(0)} m_i^{(4)}(0)}{m_i^{(2)}(0)} \geq \sum_{i=1}^r \frac{\{m_i^{(2)}(0)\}^2}{m_i^{(2)}(0)} = E_{\theta_0}(\varepsilon_{t-1}^2) \quad (3.10)$$

en utilisant Schwarz et le fait que  $\sum_{i=1}^r \mathbb{1}_{Z_t \in R_i} = 1$ . L'inégalité est stricte par les arguments déjà donnés dans la preuve du Théorème 3.3.2. Donc les colonnes de  $A_0$  sont linéairement indépendantes.

□

L'équation (3.9) permet donc d'identifier  $\theta_0$  à partir de  $b_0$  et  $A_0$ .

On définit un estimateur des moments de  $\theta$  en posant, si  $\hat{A}$  est inversible,

$$\tilde{\theta}_n = \hat{A}^{-1} \hat{b} \quad (3.11)$$

où  $\hat{b}$  et  $\hat{A}$  sont obtenus en remplaçant les moments théoriques par leur contrepartie empirique, c'est-à-dire

$$\hat{b} = \begin{pmatrix} \hat{m}_1^{(2)}(1) \\ \hat{m}_2^{(2)}(1) \\ \vdots \\ \hat{m}_r^{(2)}(1) \\ \frac{1}{n} \sum_{t=1}^n \varepsilon_t^2 \varepsilon_{t-1}^2 \end{pmatrix}, \quad \hat{A} = \begin{pmatrix} \hat{m}_1^{(0)} & \hat{m}_1^{(2)}(0) & 0 & \dots & 0 \\ \hat{m}_2^{(0)} & 0 & \hat{m}_2^{(2)}(0) & 0 & \dots & 0 \\ \vdots & \vdots & & \ddots & & \\ \hat{m}_r^{(0)} & 0 & \dots & & 0 & \hat{m}_r^{(2)}(0) \\ \frac{1}{n} \sum_{t=1}^n \varepsilon_{t-1}^2 & \hat{m}_1^{(4)}(0) & \dots & \dots & \dots & \hat{m}_r^{(4)}(0) \end{pmatrix}$$

et

$$\hat{m}_i^{(0)} = \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{Z_t \in R_i}, \quad i = 1, \dots, r,$$

$$\hat{m}_i^{(k)}(h) = \frac{1}{n} \sum_{t=1}^n \varepsilon_t^k \mathbb{1}_{Z_{t-h} \in R_i}, \quad i = 1, \dots, r, \quad k = 0, 1, \dots, \quad h = 0, 1, \dots$$

Les résultats suivants établissent la convergence forte et la normalité asymptotique de l'estimateur  $\tilde{\theta}_n$ .

**Théorème 3.4.1** *Sous les hypothèses A1-A4,  $\tilde{\theta}_n$  est un estimateur fortement convergent de  $\theta$ .*

**Preuve.** Le processus  $(Z_t)$  est ergodique car  $Z_t$  est une fonction mesurable de  $\varepsilon_{t-s}$ ,  $s > 0$  et  $(\varepsilon_t)$  est ergodique. Par suite, quand  $n \rightarrow \infty$

$$\hat{m}_i^{(0)} \rightarrow m_i^{(0)} \quad p.s., \quad 1 \leq i \leq r,$$

$$\hat{m}_i^{(k)}(h) \rightarrow m_i^{(k)}(h) \quad p.s., \quad 1 \leq i \leq r, \quad k = 1, 2, \quad h = 0, 1,$$

$$\frac{1}{n} \sum_{t=1}^n \varepsilon_{t-1}^2 \rightarrow E\varepsilon_t^2, \quad p.s., \quad \frac{1}{n} \sum_{t=1}^n \varepsilon_t^2 \varepsilon_{t-1}^2 \rightarrow E(\varepsilon_t^2 \varepsilon_{t-1}^2) \quad p.s.$$

Par suite  $\hat{A} \rightarrow A_0$  p.s.. Donc, d'après le Lemme 3.4.1,  $\hat{A}$  est p.s. inversible pour  $n$  assez grand et  $\tilde{\theta}_n$  est bien défini. La convergence forte de l'estimateur découle trivialement des convergences précédentes.

□

Le résultat suivant établit la normalité asymptotique de l'estimateur des moments sous les mêmes hypothèses que celles de la convergence. On définit

$$I^* = \begin{pmatrix} \xi_1^{(0)} & 0 & 0 & \dots & \xi_1^{(1)} \\ 0 & \xi_2^{(0)} & 0 & 0 & \dots & \xi_2^{(1)} \\ \vdots & \vdots & & \ddots & & \\ 0 & 0 & \dots & \xi_r^{(0)} & \xi_r^{(1)} \\ \xi_1^{(1)} & \dots & \dots & \xi_r^{(1)} & \sum_{i=1}^r \xi_i^{(2)} \end{pmatrix}.$$

L'inversibilité de  $I^*$  sera montrée dans la preuve du théorème suivant.

**Théorème 3.4.2** *Sous les hypothèse A1-A4*

$$\sqrt{n}(\tilde{\theta}_n - \theta_0) \Rightarrow \mathcal{N}(0, (\mu_4 - 1)A_0^{-1}I^*(A_0')^{-1}).$$

**Preuve.** En utilisant (3.2) on remarque que

$$\begin{aligned} \hat{m}_i^{(2)}(1) &= \omega_0 \hat{m}_i^{(0)} + \alpha_{0i} \hat{m}_i^{(2)} + \frac{1}{n} \sum_{t=1}^n u_t \mathbb{1}_{Z_{t-1} \in R_i}, \quad i = 1, \dots, r, \\ \frac{1}{n} \sum_{t=1}^n \varepsilon_t^2 \varepsilon_{t-1}^2 &= \omega_0 \frac{1}{n} \sum_{t=1}^n \varepsilon_{t-1}^2 + \sum_{i=1}^r \alpha_{0i} \hat{m}_i^{(4)} + \frac{1}{n} \sum_{t=1}^n u_t \varepsilon_{t-1}^2 \end{aligned}$$

ce qui s'écrit vectoriellement

$$\hat{b} = \hat{A}\theta_0 + \begin{pmatrix} \frac{1}{n} \sum_{t=1}^n u_t \mathbb{1}_{Z_{t-1} \in R_1} \\ \frac{1}{n} \sum_{t=1}^n u_t \mathbb{1}_{Z_{t-1} \in R_2} \\ \vdots \\ \frac{1}{n} \sum_{t=1}^n u_t \mathbb{1}_{Z_{t-1} \in R_r} \\ \frac{1}{n} \sum_{t=1}^n u_t \varepsilon_{t-1}^2 \end{pmatrix}.$$

Par suite

$$\hat{A}\sqrt{n}(\tilde{\theta}_n - \theta_0) = \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{t=1}^n u_t \mathbb{1}_{Z_{t-1} \in R_1} \\ \frac{1}{\sqrt{n}} \sum_{t=1}^n u_t \mathbb{1}_{Z_{t-1} \in R_2} \\ \vdots \\ \frac{1}{\sqrt{n}} \sum_{t=1}^n u_t \mathbb{1}_{Z_{t-1} \in R_r} \\ \frac{1}{\sqrt{n}} \sum_{t=1}^n u_t \varepsilon_{t-1}^2 \end{pmatrix}. \quad (3.12)$$

Notons  $W_{t-1} = (\mathbb{1}_{Z_{t-1} \in R_1}, \dots, \mathbb{1}_{Z_{t-1} \in R_r}, \varepsilon_{t-1}^2)'$ . Pour tout vecteur  $\beta = (\beta_1, \dots, \beta_{r+1})'$ , la suite  $(\frac{1}{\sqrt{n}} u_t \beta' W_{t-1}, \mathcal{F}_{t-1})_t$  est une différence de martingale. On a

$$\begin{aligned} & \sum_{t=1}^n E \left\{ \left( \frac{1}{\sqrt{n}} u_t \beta' W_{t-1} \right)^2 \middle| \mathcal{F}_{t-1} \right\} \\ &= \frac{1}{n} \sum_{t=1}^n E(\eta_t^2 - 1)^2 \left( \omega_0^2 + \sum_{i=1}^r \alpha_{0i}^2 \mathbb{1}_{Z_{t-1} \in R_i} \varepsilon_{t-1}^4 + 2\omega_0 \sum_{i=1}^r \alpha_{0i} \mathbb{1}_{Z_{t-1} \in R_i} \varepsilon_{t-1}^2 \right) \\ & \quad \times \left( \sum_{i=1}^r \beta_i^2 \mathbb{1}_{Z_{t-1} \in R_i} + 2 \sum_{i=1}^r \beta_i \mathbb{1}_{Z_{t-1} \in R_i} \beta_{r+1} \varepsilon_{t-1}^2 + \beta_{r+1}^2 \varepsilon_{t-1}^4 \right) \\ & \rightarrow (\mu_4 - 1) \left\{ \sum_{i=1}^r \beta_i^2 \xi_i^{(0)} + 2\beta_i \beta_{r+1} \xi_i^{(1)} + \beta_{r+1}^2 \xi_i^{(2)} \right\}, \end{aligned}$$

en probabilité quand  $n$  tend vers l'infini. De plus, pour toute constante  $\epsilon > 0$ ,

$$\sum_{t=1}^n E \left\{ \left( \frac{1}{\sqrt{n}} u_t \beta' W_{t-1} \right)^2 \mathbb{1}_{\{n^{-1/2} |u_t \beta' W_{t-1}| \geq \epsilon\}} \right\} \rightarrow 0, \quad \text{quand } n \rightarrow \infty$$

par le même raisonnement que dans la preuve du Théorème 3.3.2. En appliquant comme précédemment le théorème central limite pour tableaux triangulaires de différences de martingales on obtient

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n u_t \beta' W_{t-1} \Rightarrow \mathcal{N} \left( 0, (\mu_4 - 1) \left\{ \sum_{i=1}^r \beta_i^2 \xi_i^{(0)} + 2\beta_i \beta_{r+1} \xi_i^{(1)} + \beta_{r+1}^2 \xi_i^{(2)} \right\} \right)$$

et par suite, en utilisant la propriété de Cramer-Wold, que

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n u_t W_{t-1} \Rightarrow \mathcal{N}(0, (\mu_4 - 1) I^*). \quad (3.13)$$

On montre que  $I^*$  est inversible par le même raisonnement que pour la matrice  $I$ . On conclut en utilisant les équations (3.12) et (3.13), le fait que  $\hat{A} \rightarrow A_0$  *p.s.* et l'inversibilité de  $A_0$  d'après le Lemme 3.4.1.

□

**Exemples (suite) :** 1. Dans le cas où  $r = 1$  on obtient, si  $\mu_8\alpha^4 < 1$ ,

$$I^* = \begin{pmatrix} \xi^{(0)} & \xi^{(1)} \\ \xi^{(1)} & \xi^{(2)} \end{pmatrix}, \quad A_0 = \begin{pmatrix} 1 & m^{(2)} \\ m^{(2)} & m^{(4)} \end{pmatrix}.$$

Par suite, on retrouve la matrice de variance obtenue dans le cas des MCO. Les lois asymptotiques des estimateurs des MCO et des moments sont donc les mêmes car les deux méthodes sont, dans ce cas, fondées sur les mêmes relations d'orthogonalité. Numériquement, les estimateurs sont également très proches mais diffèrent par quelques observations.

2. Dans le cas où  $r = 2$  on obtient

$$I^* = \begin{pmatrix} \xi_1^{(0)} & 0 & \xi_1^{(1)} \\ 0 & \xi_2^{(0)} & \xi_2^{(1)} \\ \xi_1^{(1)} & \xi_2^{(1)} & \sum_{i=1}^2 \xi_i^{(2)} \end{pmatrix}, \quad A_0 = \begin{pmatrix} m_1^{(0)} & m_1^{(2)} & 0 \\ m_2^{(0)} & 0 & m_2^{(2)} \\ E_{\theta_0}(\varepsilon_t^2) & m_1^{(4)} & m_2^{(4)} \end{pmatrix}.$$

Il est évidemment possible d'appliquer la méthode des moments à partir d'autres équations que (3.7) et (3.8),

## 3.4.1 : Estimation de la variance asymptotique de l'estimateur des moments

$k = 1$		
$(\omega_0, \alpha_{01}, \alpha_{02})$	(1, 0.3, 0.5)	(1, 3.0, 0.1)
$\bar{\theta}_N$	(1.094, 0.172, 0.465)	(1.080, 2.898, 0.078)
$\bar{\sigma}_N^2$	10.250	22.214
$\hat{p}(R_1)$	0.490	0.486
$\bar{\Sigma}_N$	$\begin{pmatrix} 21.438 & -14.240 & -0.539 \\ -14.240 & 53.820 & -3.941 \\ -0.539 & -3.941 & 1.378 \end{pmatrix}$	$\begin{pmatrix} 94.066 & -93.771 & 0.349 \\ -93.771 & 278.812 & -6.104 \\ 0.349 & -6.104 & 1.106 \end{pmatrix}$
$k = 3$		
$(\omega_0, \alpha_{01}, \alpha_{02})$	(1, 0.3, 0.5)	(1, 0.5, 0.1)
$\bar{\theta}_N$	(1.030, 0.273, 0.487)	(1.013, 0.484, 0.094)
$\bar{\sigma}_N^2$	7.95	5.287
$\hat{p}(R_1)$	0.679	0.667
$\bar{\Sigma}_N$	$\begin{pmatrix} 11.805 & -1.959 & -1.265 \\ -1.959 & 4.189 & -1.939 \\ -1.265 & -1.939 & 3.594 \end{pmatrix}$	$\begin{pmatrix} 7.084 & -2.374 & -0.757 \\ -2.373 & 5.526 & -1.428 \\ -0.757 & -1.428 & 2.179 \end{pmatrix}$
$k = 5$		
$(\omega_0, \alpha_{01}, \alpha_{02})$	(1, 0.2, 1.1)	(1, 0.7, 0.2)
$\bar{\theta}_N$	(1.279, 0.039, 1.012)	(1.227, 0.578, 0.119)
$\bar{\sigma}_N^2$	27.01	27.435
$\hat{p}(R_1)$	0.745	0.755
$\bar{\Sigma}_N$	$\begin{pmatrix} 116.475 & -26.927 & -2.182 \\ -26.927 & 24.419 & -7.995 \\ -2.182 & -7.995 & 15.889 \end{pmatrix}$	$\begin{pmatrix} 76.161 & -5.283 & -15.073 \\ -5.283 & 5.566 & -3.968 \\ -15.073 & -3.968 & 30.117 \end{pmatrix}$

### 3.5 Méthode du maximum de vraisemblance

Nous présentons dans cette partie la méthode du maximum de vraisemblance. Sous l'hypothèse de normalité des variables  $\eta_t$ , la vraisemblance peut être écrite comme pour les modèles ARCH classiques. Nous ne ferons cependant pas l'hypothèse que cette loi constitue la vraie densité des  $\eta_t$ , mais nous l'utiliserons pour définir le critère. Conditionnellement à des valeurs initiales  $\varepsilon_0^2$  et  $\varepsilon_{-1}^2$  la pseudo-vraisemblance gaussienne est donnée par

$$L_n(\theta) = L_n(\theta, \varepsilon_1, \dots, \varepsilon_n) = \prod_{t=1}^n \frac{1}{\sqrt{2\pi\tilde{h}_t}} \exp\left(-\frac{\varepsilon_t^2}{2\tilde{h}_t}\right),$$

avec  $\tilde{h}_t = \tilde{h}_t(\theta)$  est définie par  $\tilde{h}_1 = \tilde{h}_2 = \omega$  et  $\tilde{h}_t = h_t$  pour  $t = 3, \dots, n$ , calculé à partir de la deuxième équation du modèle (3.1). Notons  $\tilde{\mathbf{I}}(\theta)$  la log-vraisemblance associée à  $\varepsilon_t$  conditionnellement au passé. La log-vraisemblance conditionnelle s'écrit, normalisée par  $n$ , à une constante additive près :

$$\tilde{\mathbf{I}}_n(\theta) = -\frac{1}{2n} \sum_{t=1}^n \tilde{\ell}_t = \frac{-1}{2n} \sum_{t=1}^n \left[ \ln \tilde{h}_t + \frac{\varepsilon_t^2}{\tilde{h}_t} \right]$$

On définit un estimateur du maximum de vraisemblance  $\hat{\theta}_n$ , comme une solution mesurable de :

$$\hat{\theta}_n = \arg \max_{\theta \in \Theta} L_n(\theta) = \arg \max_{\theta \in \Theta} \tilde{\mathbf{I}}_n(\theta), \quad (3.14)$$

Afin de démontrer la convergence forte de l'estimateur du maximum de vraisemblance, nous remplacerons l'hypothèse **A4** par :

$$\mathbf{A4}' : \exists s > 0, E_{\theta_0}(\varepsilon_t^{2s}) < +\infty.$$

Nous ferons également l'hypothèse

$$\mathbf{A5} : \theta_0 \in \Theta \text{ et } \Theta \text{ est un compact.}$$

**Théorème 3.5.1** Soit  $\hat{\theta}_n$  solutions de (3.14). Sous les hypothèses **A1**, **A2**, **A3**, **A4'** et **A5**, on a

$$\hat{\theta}_n \longrightarrow \theta_0 \quad \text{quand} \quad n \longrightarrow +\infty$$

**Preuve.** La démonstration de ce théorème est basée sur une adaptation de la démonstration de Francq et Zakoïan (2003), Théorème 2.1. Il sera utile d'introduire

$$\mathbf{l}_n(\theta) = \mathbf{l}_n(\theta, \varepsilon_n, \varepsilon_{n-1}, \dots) = (-2n)^{-1} \sum_{t=1}^n \ell_t, \quad \ell_t = \ell_t(\theta) = \ln[h_t] + \frac{\varepsilon_t^2}{h_t}.$$

Sous les hypothèses du théorème 3.5.1, nous montrons les Lemmes suivants

**Lemme 3.5.1**  $\limsup_{n \rightarrow \infty} \sup_{\theta \in \Theta} |\mathbf{l}_n(\theta) - \tilde{\mathbf{l}}_n(\theta)| = 0$ , *p.s.*

**Lemme 3.5.2**  $h_t(\theta) = h_t(\theta_0)$   $P_{\theta_0}$  *p.s.*  $\implies \theta = \theta_0$ .

**Lemme 3.5.3** la fonction  $\bar{\mathbf{l}}(\theta) = \frac{-1}{2} E[\ell_t(\theta)]$  admet un unique maximum en  $\theta_0$ .

**Lemme 3.5.4**  $E_{\theta_0} |\ell_t(\theta)| < \infty$ ,  $\forall \theta \in \Theta$  et si  $\theta \neq \theta_0$ ,  $E_{\theta_0} \ell_t(\theta) > E_{\theta_0} \ell_t(\theta_0)$ ,

**Lemme 3.5.5** Pour tout  $\theta \neq \theta_0$  il existe un voisinage  $V(\theta)$  tel que  $\liminf_{n \rightarrow \infty} \inf_{\theta^* \in V(\theta)} \tilde{\mathbf{l}}_n(\theta^*) > E_{\theta_0} \ell_1(\theta_0)$ , *p.s.*

**Preuve du Lemme 3.5.1** : Le résultat est immédiat car

$$|\mathbf{l}_n(\theta) - \tilde{\mathbf{l}}_n(\theta)| = \frac{1}{2n} \left| \left[ \ln\left(\frac{h_1}{\tilde{h}_1}\right) + \frac{\varepsilon_1^2}{h_1} - \frac{\varepsilon_1^2}{\tilde{h}_1} + \ln\left(\frac{h_2}{\tilde{h}_2}\right) + \frac{\varepsilon_2^2}{h_2} - \frac{\varepsilon_2^2}{\tilde{h}_2} \right] \right|$$

où tous les numérateurs sont minorés par un réel strictement positif puisque  $\Theta$  est compact et  $\omega > 0$ .

□

**Preuve du Lemme 3.5.2 :**

Supposons que  $h_t(\theta) = h_t(\theta_0)$   $P_{\theta_0}$  presque sûrement. Alors

$$\begin{aligned}
\omega + \sum_{i=1}^r \alpha_i \mathbb{1}_{Z_{t-1} \in R_i} \varepsilon_{t-1}^2 &= \omega_0 + \sum_{i=1}^r \alpha_{0i} \mathbb{1}_{Z_{t-1} \in R_i} \varepsilon_{t-1}^2 \\
\iff (\omega - \omega_0) + \sum_{i=1}^r (\alpha_i - \alpha_{0i}) \mathbb{1}_{Z_{t-1} \in R_i} \varepsilon_{t-1}^2 &= 0 \\
\iff (\omega - \omega_0) + \sum_{i=1}^r (\alpha_i - \alpha_{0i}) h_{t-1} \eta_{t-1}^2 \mathbb{1}_{Z_{t-1} \in R_i} &= 0 \\
\iff (\omega - \omega_0) + \sum_{i=1}^r \alpha_{i,t-2} \eta_{t-1}^2 \mathbb{1}_{Z_{t-1} \in R_i} &= 0
\end{aligned}$$

où  $\alpha_{i,t-2} = (\alpha_i - \alpha_{0i}) h_{t-1}$ . Donc si  $\alpha_i \neq \alpha_{0i}$ ,  $\forall i = 1, \dots, r$

$$\implies \eta_{t-1}^2 = \begin{cases} \frac{\omega_0 - \omega}{\alpha_{1,t-2}} & \text{si } Z_{t-1} \in R_1 \\ \vdots & \\ \frac{\omega_0 - \omega}{\alpha_{r,t-2}} & \text{si } Z_{t-1} \in R_r \end{cases} \quad (3.15)$$

Puisque les  $\{R_i\}_{1,\dots,r}$  constituent une partition de  $\mathbb{R}^d$  et d'après **A1**,  $\eta_{t-1}$  est indépendante du passé  $\{\varepsilon_{t-2}, \varepsilon_{t-3}, \dots\}$  donc  $\eta_{t-1}^2$  n'a que  $r$  valeurs ce qui est incompatible avec l'hypothèse **A3**, d'où  $\alpha_i = \alpha_{0i} \forall i = 1, \dots, r$  et par conséquence  $\omega = \omega_0$  ie  $\theta = \theta_0$ .

□

**Preuve du Lemme 3.5.3 :**

Remarquons tout d'abord que  $\bar{\mathbf{I}}(\theta)$  est bien défini et ne dépend pas de  $t$ , en vertu du Lemme précédent et de l'hypothèse **A1**.

Formons

$$\bar{\mathbf{I}}(\theta_0) - \bar{\mathbf{I}}(\theta) = \frac{1}{2}E_{\theta_0}\left[\ln \frac{h_t(\theta)}{h_t(\theta_0)}\right] - \frac{1}{2}E_{\theta_0}\left[\frac{\varepsilon_t^2}{h_t(\theta_0)} - \frac{\varepsilon_t^2}{h_t(\theta)}\right]$$

On a  $\varepsilon_t = h_t^{1/2}(\theta_0)\eta_t$ . Donc

$$\begin{aligned}\bar{\mathbf{I}}(\theta_0) - \bar{\mathbf{I}}(\theta) &= \frac{1}{2}E_{\theta_0}\left[\ln \frac{h_t(\theta)}{h_t(\theta_0)}\right] - \frac{1}{2}\left[1 - E_{\theta_0}\left(\frac{h_t(\theta_0)}{h_t(\theta)}\right)\right] \\ &\geq \frac{1}{2}\left[1 - E_{\theta_0}\left(\frac{h_t(\theta_0)}{h_t(\theta)}\right)\right] - \frac{1}{2}\left[1 - E_{\theta_0}\left(\frac{h_t(\theta_0)}{h_t(\theta)}\right)\right] = 0\end{aligned}$$

en utilisant la relation  $\ln x \geq 1 - \frac{1}{x}$  pour  $x > 0$ . Cette dernière inégalité est stricte sauf si  $x = 1$ , ce qui permet d'affirmer que  $\bar{\mathbf{I}}(\theta_0) - \bar{\mathbf{I}}(\theta) > 0$  sauf si  $h_t(\theta_0) = h_t(\theta)$  p.s.

□

**Preuve du Lemme 3.5.4 :**

Dans le cas où  $\ell_t(\theta)$  est négative, on a  $E_{\theta_0}\ell_t^-(\theta) \leq E_{\theta_0}\ln^-h_t \leq \max\{0, -\ln\omega\} < \infty$ . Nous montrons aussi que  $E_{\theta_0}\ell_t^+(\theta) < 0$ . En effet, d'après l'hypothèse **A4'** on a  $\exists s > 0$ ,  $E_{\theta_0}(\varepsilon_t^{2s}) < +\infty$ . Or  $E_{\theta_0}(\varepsilon_t^{2s}) = E_{\theta_0}(\eta_t^{2s}h_t^s) = E_{\theta_0}(\eta_t^{2s})E_{\theta_0}(h_t^s) < +\infty$  donc  $E_{\theta_0}(h_t^s) < +\infty$  car  $E_{\theta_0}(\eta_t^{2s}) < +\infty$ . Donc

$$E_{\theta_0}\ln h_t(\theta) = E_{\theta_0}\frac{1}{s}\ln\{h_t(\theta)\}^s \leq \frac{1}{s}\ln E_{\theta_0}\{h_t(\theta)\}^s < \infty$$

et

$$E_{\theta_0}\ell_t(\theta) = E_{\theta_0}\left\{\frac{h_t(\theta)\eta_t^2}{h_t(\theta_0)} + E_{\theta_0}\ln h_t(\theta)\right\} = E_{\theta_0}\left\{\frac{h_t(\theta)}{h_t(\theta_0)}\right\} + \ln h_t(\theta) < \infty$$

D'après le Lemme 3.5.3, si  $\theta \neq \theta_0$  alors  $\bar{\mathbf{I}}(\theta_0) - \bar{\mathbf{I}}(\theta) = \frac{1}{2}[E_{\theta_0}[\ell_t(\theta)] - E_{\theta_0}[\ell_t(\theta_0)]] > 0$  d'ou le résultat. Notons que  $E_{\theta_0}\ell_t^-(\theta) \leq E_{\theta_0}\ln^-h_t \leq \max\{0, -\ln\omega\} < \infty$  et  $E_{\theta_0}\mathbf{1}_n(\theta) = E_{\theta_0}\ell_t(\theta)$ .

□

**Preuve du Lemme 3.5.5 :**

La preuve de ce Lemme est la même que dans Francq et Zakoïan (2003), Théorème 2.1. Elle est basée sur les résultats des Lemmes 3.5.1, 3.5.3. Par l'hypothèse **A5**, nous complétons la preuve du théorème 3.5.1.

□

Pour montrer la normalité asymptotique nous adaptons la démonstration de Francq et Zakoïan (2003), Théorème 2.2. Nous introduisons les deux hypothèses suivantes

**A6** :  $\theta_0 \in \overset{\circ}{\Theta}$ , avec  $\overset{\circ}{\Theta}$ , désigne l'intérieur de  $\Theta$ .

**A7** :  $\kappa_\eta := E(\eta_t^4) < \infty$ .

**Théorème 3.5.2** *Sous les hypothèses du Théorème 3.5.1 et sous **A6-A7**,  $\sqrt{n}(\hat{\theta}_n - \theta_0)$  est asymptotiquement de loi  $\mathcal{N}(0, (\kappa_\eta - 1)J^{-1})$  avec*

$$J := E_{\theta_0} \left( \frac{\partial^2 \ell_t(\theta_0)}{\partial \theta \partial \theta'} \right) = E_{\theta_0} \left( \frac{1}{h_t^2(\theta_0)} \frac{\partial h_t(\theta_0)}{\partial \theta} \frac{\partial h_t(\theta_0)}{\partial \theta'} \right). \quad (3.16)$$

Il est intéressant de comparer la variance asymptotique de l'estimateur du MV avec celles obtenues par les méthodes précédentes.

**Exemples (suite)** : 1. Dans le cas où  $r = 1$  on a

$$J = E_{\theta_0} \left[ \begin{array}{cc} \frac{1}{(\omega_0 + \alpha_0 \epsilon_{t-1}^2)^2} & \frac{\epsilon_{t-1}^2}{(\omega_0 + \alpha_0 \epsilon_{t-1}^2)^2} \\ \frac{\epsilon_{t-1}^2}{(\omega_0 + \alpha_0 \epsilon_{t-1}^2)^2} & \frac{\epsilon_{t-1}^4}{(\omega_0 + \alpha_0 \epsilon_{t-1}^2)^2} \end{array} \right].$$

On obtient par simulation ( $N = 1000$  simulations de séries de taille  $n = 10000$  avec bruit  $\eta_t$  gaussien standard) :

## 3.5.1 : Estimation de la variance asymptotique de l'estimateur du MV

paramètres	$\omega = 1, \alpha = 0.1$	$\omega = 1, \alpha = 0.5$	$\omega = 1, \alpha = 0.95$
$\text{Var}_{as}\{\sqrt{n}(\hat{\theta}_n - \theta_0)\}$	$\begin{pmatrix} 3.46 & -1.34 \\ -1.34 & 1.87 \end{pmatrix}$	$\begin{pmatrix} 4.85 & -2.15 \\ -2.15 & 3.99 \end{pmatrix}$	$\begin{pmatrix} 6.61 & -2.83 \\ -2.83 & 6.67 \end{pmatrix}$

Comme pour l'estimateur des MCO, la précision asymptotique se détériore globalement lorsque  $\alpha$  augmente. Elle se détériore également pour chaque coefficient, contrairement au cas MCO. Pour  $\alpha = 0.1$  ou  $\alpha = 0.5$  aucun des estimateurs ne l'emporte globalement sur l'autre. L'écart entre les précisions asymptotiques est cependant nettement plus important en ce qui concerne  $\omega$  que  $\alpha$ , ceci en faveur du MV. Même lorsque  $\alpha$  est très proche de 1, ce qui correspond aux séries réelles, la précision asymptotique de l'estimateur du MV reste correcte : par exemple pour une valeur estimée de  $(1, 0.95)$ , à partir d'une série de longueur 10000, l'utilisation de cette matrice conduirait à approximer l'intervalle de confiance pour  $\alpha$ , au seuil 95%, par  $[0.90, 1.00]$ . Rappelons que la matrice asymptotique de l'estimateur MCO n'existe pas pour les valeurs de  $\alpha$  supérieures à 0.577.

2. On a lorsque  $r = 2$

$$J = E_{\theta_0} \left\{ \frac{1}{(\omega_0 + \alpha_{01}\epsilon_{t-1}^2 \mathbb{1}_{Z_{t-1} \in R_1} + \alpha_{02}\epsilon_{t-1}^2 \mathbb{1}_{Z_{t-1} \in R_2})^2} \times \begin{bmatrix} 1 & \epsilon_{t-1}^2 \mathbb{1}_{Z_{t-1} \in R_1} & \epsilon_{t-1}^2 \mathbb{1}_{Z_{t-1} \in R_2} \\ \epsilon_{t-1}^2 \mathbb{1}_{Z_{t-1} \in R_1} & \epsilon_{t-1}^4 \mathbb{1}_{Z_{t-1} \in R_1} & 0 \\ \epsilon_{t-1}^2 \mathbb{1}_{Z_{t-1} \in R_2} & 0 & \epsilon_{t-1}^4 \mathbb{1}_{Z_{t-1} \in R_2} \end{bmatrix} \right\}.$$

La variance asymptotique de l'estimateur du MV est présentée dans le Tableau suivant pour les valeurs utilisées pour la méthode MCO et celle des moments. La méthode du MV n'apparaît strictement supérieure aux deux autres dans aucun cas, excepté lorsque  $k = 1$  et  $\theta = (1, 3, 0.1)$  où l'écart entre les variances asymptotiques de l'estimateur des moments et du MV est une matrice définie positive. Cependant, la méthode du MV se distingue des deux autres par le fait que les variances asymptotiques ne prennent généralement pas de très grandes valeurs (sauf dans le cas cité plus haut pour le coefficient  $\alpha_1$ ).

## 3.5.2 : Estimation de la variance asymptotique de l'estimateur du MV

$k = 1$		
$(\omega_0, \alpha_{01}, \alpha_{02})$	(1, 0.3, 0.5)	(1, 3.0, 0.1)
$\hat{p}(R_1)$	0.490	0.486
$\bar{\Sigma}_N$	$\begin{pmatrix} 4.693 & -2.818 & -1.783 \\ -2.818 & 10.401 & 1.071 \\ -1.783 & 1.071 & 4.676 \end{pmatrix}$	$\begin{pmatrix} 4.932 & -8.312 & -0.920 \\ -8.312 & 153.793 & 1.552 \\ -0.920 & 1.552 & 0.759 \end{pmatrix}$
$k = 3$		
$(\omega_0, \alpha_{01}, \alpha_{02})$	(1, 0.3, 0.5)	(1, 0.5, 0.1)
$\hat{p}(R_1)$	0.679	0.667
$\bar{\Sigma}_N$	$\begin{pmatrix} 4.426 & -1.997 & -1.775 \\ -1.997 & 4.612 & 0.801 \\ -1.775 & 0.801 & 7.378 \end{pmatrix}$	$\begin{pmatrix} 4.232 & -2.802 & -1.145 \\ -2.802 & 9.955 & 0.758 \\ -1.145 & 0.758 & 1.730 \end{pmatrix}$
$k = 5$		
$(\omega_0, \alpha_{01}, \alpha_{02})$	(1, 0.2, 1.1)	(1, 0.7, 0.2)
$\hat{p}(R_1)$	0.745	0.755
$\bar{\Sigma}_N$	$\begin{pmatrix} 4.188 & -1.349 & -2.112 \\ -1.349 & 2.186 & 0.680 \\ -2.112 & 0.680 & 25.493 \end{pmatrix}$	$\begin{pmatrix} 4.923 & -2.679 & -1.418 \\ -2.679 & 8.652 & 0.772 \\ -1.418 & 0.772 & 3.334 \end{pmatrix}$

**Preuve :**

Le résultat de ce théorème est basé sur les développements suivants :

$$\begin{aligned} 0 &= n^{-1/2} \sum_{t=1}^n \frac{\partial}{\partial \theta} \tilde{\ell}_t(\hat{\theta}_n) \\ &= n^{-1/2} \sum_{t=1}^n \frac{\partial}{\partial \theta} \tilde{\ell}_t(\theta_0) + \left( \frac{1}{n} \sum_{t=1}^n \frac{\partial^2}{\partial \theta_i \partial \theta_j} \tilde{\ell}_t(\theta_{ij}^*) \right) \sqrt{n}(\hat{\theta}_n - \theta_0), \end{aligned}$$

où les  $\theta_{ij}^*$  sont entre  $\hat{\theta}_n$  et  $\theta_0$ . Les dérivées de  $\ell_t(\theta) = \ln[h_t] + \frac{\varepsilon_t^2}{h_t}$  sont données par

$$\frac{\partial \ell_t(\theta)}{\partial \theta} = \left\{ 1 - \frac{\varepsilon_t^2}{h_t} \right\} \left\{ \frac{1}{h_t} \frac{\partial h_t}{\partial \theta} \right\}, \quad \text{avec} \quad \frac{\partial h_t}{\partial \theta} = \begin{bmatrix} 1 \\ \varepsilon_{t-1}^2 \mathbb{1}_{\frac{\varepsilon_{t-1}^2}{\varepsilon_{t-2}^2} \in R_1} \\ \vdots \\ \varepsilon_{t-1}^2 \mathbb{1}_{\frac{\varepsilon_{t-1}^2}{\varepsilon_{t-2}^2} \in R_r} \end{bmatrix} \quad (3.17)$$

$$\begin{aligned} \frac{\partial^2 \ell_t(\theta_0)}{\partial \theta \partial \theta'} &= \left\{ 1 - \frac{\varepsilon_t^2}{h_t} \right\} \left\{ \frac{1}{h_t} \frac{\partial^2 h_t}{\partial \theta \partial \theta'} \right\} + \left\{ 2 \frac{\varepsilon_t^2}{h_t} - 1 \right\} \left\{ \frac{1}{h_t} \frac{\partial h_t}{\partial \theta} \right\} \left\{ \frac{1}{h_t} \frac{\partial h_t}{\partial \theta'} \right\} \\ &= \left\{ 2 \frac{\varepsilon_t^2}{h_t} - 1 \right\} \left\{ \frac{1}{h_t} \frac{\partial h_t}{\partial \theta} \right\} \left\{ \frac{1}{h_t} \frac{\partial h_t}{\partial \theta'} \right\} \end{aligned} \quad (3.18)$$

Afin de prouver le théorème 3.5.2, nous montrons sous ces hypothèses, les Lemmes suivants

**Lemme 3.5.6**  $E_{\theta_0} \left\| \frac{\partial \ell_t(\theta_0)}{\partial \theta} \frac{\partial \ell_t(\theta_0)}{\partial \theta'} \right\| < \infty$ ,  $E_{\theta_0} \left\| \frac{\partial^2 \ell_t(\theta_0)}{\partial \theta \partial \theta'} \right\| < \infty$ .

**Lemme 3.5.7**  $J$  est inversible et  $\text{Var}_{\theta_0} \left\{ \frac{\partial \ell_t(\theta_0)}{\partial \theta} \right\} = \{\kappa_\eta - 1\} J$ .

**Lemme 3.5.8** Il existe un voisinage  $\mathcal{V}(\theta_0)$  de  $\theta_0$  tel que pour tout  $i, j, k \in \{1, \dots, r+1\}$ ,

$$E_{\theta_0} \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^3 \ell_t(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_k} \right| < \infty.$$

**Lemme 3.5.9**  $\left\| n^{-1/2} \sum_{t=1}^n \left\{ \frac{\partial \ell_t(\theta_0)}{\partial \theta} - \frac{\partial \tilde{\ell}_t(\theta_0)}{\partial \theta} \right\} \right\| \rightarrow 0$

et  $\sup_{\theta \in \mathcal{V}(\theta_0)} \left\| n^{-1} \sum_{t=1}^n \left\{ \frac{\partial^2 \ell_t(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \tilde{\ell}_t(\theta)}{\partial \theta \partial \theta'} \right\} \right\| \rightarrow 0$  en probabilité quand  $n \rightarrow \infty$ .

**Lemme 3.5.10**  $n^{-1/2} \sum_{t=1}^n \frac{\partial}{\partial \theta} \ell_t(\theta_0) \Rightarrow \mathcal{N}(0, (\kappa_\eta - 1)J)$ .

**Lemme 3.5.11**  $n^{-1} \sum_{t=1}^n \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ell_t(\theta_{ij}^*) \rightarrow J(i, j)$  p.s.

**Preuve du Lemme 3.5.6 :**

Nous utilisons pour une matrice  $A = (a_{ij})$  la norme définie par  $\|A\| = \sum |a_{ij}|$ .

Pour montrer les deux inégalités, nous utilisons (3.17) et (3.18). On a

$$\begin{aligned} \left\| \frac{\partial \ell_t(\theta_0)}{\partial \theta} \frac{\partial \ell_t(\theta_0)}{\partial \theta'} \right\| &\leq \left\| \frac{\partial \ell_t(\theta_0)}{\partial \theta} \right\|^2 \\ &= \left(1 - \frac{\varepsilon_t^2}{h_t}\right)^2 \left\| \frac{1}{h_t} \frac{\partial h_t}{\partial \theta} \right\|^2 \end{aligned}$$

Pour  $\theta = \theta_0$ ,  $\frac{\varepsilon_t^2}{h_t} = \eta_t^2$  indépendant de  $h_t$  et de ses dérivées. On a aussi

$$\begin{aligned} \left\| \frac{1}{h_t(\theta_0)} \frac{\partial h_t(\theta_0)}{\partial \theta} \right\|^2 &\leq \frac{1}{h_t(\theta_0)^2} (1 + \varepsilon_{t-1}^2)^2 \\ &= \left( \frac{1}{h_t(\theta_0)} + \frac{\varepsilon_{t-1}^2}{h_t(\theta_0)} \right)^2 \\ &< \left( \frac{1}{\omega_0} + \frac{1}{\min_{1 \leq i \leq r} (\alpha_{0i})} \right)^2 \end{aligned}$$

D'après l'hypothèse **A7**, on a

$$\begin{aligned}
E_{\theta_0} \left\| \frac{\partial \ell_t(\theta_0)}{\partial \theta} \frac{\partial \ell_t(\theta_0)}{\partial \theta'} \right\| &\leq \left( \frac{1}{\omega_0} + \frac{1}{\min_{1 \leq i \leq r} (\alpha_{0i})} \right)^2 E_{\theta_0} (1 - \eta_t^2)^2 \\
&= \left( \frac{1}{\omega_0} + \frac{1}{\min_{1 \leq i \leq r} (\alpha_{0i})} \right)^2 (\kappa_\eta - 1) \\
&< \infty
\end{aligned}$$

En utilisant (3.18) nous pouvons montrer facilement la deuxième inégalité. En effet

$$\begin{aligned}
\left\| \frac{\partial^2 \ell_t(\theta_0)}{\partial \theta \partial \theta'} \right\| &= \left| 2 \frac{\varepsilon_t^2}{h_t(\theta_0)} - 1 \right| \left\| \frac{1}{h_t(\theta_0)^2} \frac{\partial h_t(\theta_0)}{\partial \theta} \frac{\partial h_t(\theta_0)}{\partial \theta'} \right\| \\
&= \left| 2 \frac{\varepsilon_t^2}{h_t(\theta_0)} - 1 \right| \left( \frac{1}{h_t(\theta_0)} + \frac{\varepsilon_{t-1}^2}{h_t(\theta_0)} \right)^2 \\
&\leq |2\eta_t^2 - 1| \left( \frac{1}{\omega_0} + \frac{1}{\min_{1 \leq i \leq r} (\alpha_{0i})} \right)^2
\end{aligned}$$

Alors

$$\begin{aligned}
E_{\theta_0} \left\| \frac{\partial^2 \ell_t(\theta_0)}{\partial \theta \partial \theta'} \right\| &\leq 3 \left( \frac{1}{\omega_0} + \frac{1}{\min_{1 \leq i \leq r} (\alpha_{0i})} \right)^2 \\
&< \infty
\end{aligned}$$

□

**Preuve du Lemme 3.5.7 :**

En utilisant (3.18) et le Lemme 3.5.6, la matrice  $J$  existe et (3.16) est vérifiée et

nous avons aussi

$$\begin{aligned}
 \text{Var}_{\theta_0} \left\{ \frac{\partial \ell_t(\theta_0)}{\partial \theta} \right\} &= E_{\theta_0} \left\{ \frac{\partial \ell_t(\theta_0)}{\partial \theta} \frac{\partial \ell_t(\theta_0)}{\partial \theta'} \right\} \\
 &= E\{(1 - \eta_t^2)^2\} E_{\theta_0} \left\{ \frac{\partial h_t(\theta_0)}{\partial \theta} \frac{\partial h_t(\theta_0)}{\partial \theta'} \right\} \\
 &= (\kappa_\eta - 1)J.
 \end{aligned}$$

Nous supposons que  $\lambda' J \lambda = E\left[\frac{1}{h_t^2(\theta_0)} (\lambda' \frac{\partial h_t(\theta_0)}{\partial \theta})^2\right] = 0$  pour  $\lambda \in \mathbb{R}^{r+1}$ . Donc presque sûrement,  $\lambda' \frac{\partial h_t(\theta_0)}{\partial \theta} = 0$ .

Posons  $\lambda = (\lambda_0, \lambda_1, \dots, \lambda_r)$ . Sous les deux hypothèses **A1**, **A4'** et en utilisant le même raisonnement que pour le Lemme 3.5.2, nous pouvons montrer facilement que  $\lambda_0 = \lambda_1 = \dots = \lambda_r = 0$  d'où la fin de la démonstration du Lemme.

□

### Preuve du Lemme 3.5.8 :

Nous dérivons l'égalité (3.18) :

$$\begin{aligned}
 \frac{\partial^3 \ell_t(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_k} &= \left\{ -2 \frac{\varepsilon_t^2}{h_t^2} \frac{\partial h_t}{\partial \theta_i} \right\} \left\{ \frac{1}{h_t} \frac{\partial h_t}{\partial \theta_j} \right\} \left\{ \frac{1}{h_t} \frac{\partial h_t}{\partial \theta_k} \right\} \\
 &\quad + \left\{ 2 \frac{\varepsilon_t^2}{h_t} - 1 \right\} \left\{ -2 \frac{1}{h_t^3} \frac{\partial h_t}{\partial \theta_k} \frac{\partial h_t}{\partial \theta_i} \frac{\partial h_t}{\partial \theta_j} \right\} \\
 &= \left\{ 2 - 6 \frac{\varepsilon_t^2}{h_t} \right\} \left\{ \frac{1}{h_t^3} \frac{\partial h_t}{\partial \theta_i} \frac{\partial h_t}{\partial \theta_j} \frac{\partial h_t}{\partial \theta_k} \right\}
 \end{aligned} \tag{3.19}$$

Soit  $\Theta^*$  un compact inclus dans  $\overset{\circ}{\Theta}$ . Pour tout  $\delta > 0$ , il existe un voisinage  $\mathcal{V}(\theta_0)$  de  $\theta_0$  inclus dans  $\Theta^*$  tel que  $\omega_0 \leq (1 + \delta)\omega$ ,  $\alpha_{0i} \leq (1 + \delta)\alpha_j$ ,  $\forall 1 \leq i, j \leq r$  pour tout

$\theta \in \mathcal{V}(\theta_0)$ . Nous avons

$$\begin{aligned}
\frac{h_t(\theta_0)}{h_t(\theta)} &= \frac{\omega_0 + \sum_{i=1}^r \alpha_{0i} \mathbf{1}_{Z_{t-1} \in R_i} \varepsilon_{t-1}^2}{\omega + \sum_{i=1}^r \alpha_i \mathbf{1}_{Z_{t-1} \in R_i} \varepsilon_{t-1}^2} \\
&= \frac{\omega_0}{\omega + \sum_{i=1}^r \alpha_i \mathbf{1}_{Z_{t-1} \in R_i} \varepsilon_{t-1}^2} + \frac{\sum_{i=1}^r \alpha_{0i} \mathbf{1}_{Z_{t-1} \in R_i} \varepsilon_{t-1}^2}{\omega + \sum_{i=1}^r \alpha_i \mathbf{1}_{Z_{t-1} \in R_i} \varepsilon_{t-1}^2} \\
&\leq \frac{\omega_0}{\omega} + \frac{\sum_{i=1}^r \alpha_{0i} \mathbf{1}_{Z_{t-1} \in R_i}}{\sum_{i=1}^r \alpha_i \mathbf{1}_{Z_{t-1} \in R_i}} \\
&\leq 2(1 + \delta)
\end{aligned}$$

Notons par  $\|X\|_p := E^{1/p}(X^p)$ . On a

$$\begin{aligned}
\left\| \sup_{\theta \in \mathcal{V}(\theta_0)} \frac{\varepsilon_t^2}{h_t} \right\|_2 &= \kappa_\eta^{1/2} \left\| \sup_{\theta \in \mathcal{V}(\theta_0)} \frac{h_t(\theta_0)}{h_t} \right\|_2 \\
&\leq 2\kappa_\eta^{1/2}(1 + \delta)
\end{aligned} \tag{3.20}$$

En utilisant (3.20), les résultats du Lemme précédent et l'inégalité de Hölder

$$\begin{aligned}
E_{\theta_0} \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\partial^3 \ell_t(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_k} \right| &= E_{\theta_0} \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \left\{ 2 - 6 \frac{\varepsilon_t^2}{h_t} \right\} \left\{ \frac{1}{h_t^3} \frac{\partial h_t}{\partial \theta_i} \frac{\partial h_t}{\partial \theta_j} \frac{\partial h_t}{\partial \theta_k} \right\} \right| \\
&\leq \left\| \sup_{\theta \in \mathcal{V}(\theta_0)} \left| 2 - 6 \frac{\varepsilon_t^2}{h_t} \right| \right\|_2 \max_i \left\| \sup_{\theta \in \Theta^*} \left| \frac{1}{h_t} \frac{\partial h_t}{\partial \theta_i} \right| \right\|_6^2 < \infty.
\end{aligned}$$

□

**Preuve du Lemme 3.5.9 :**

D'après les résultats des Lemmes 3.5.6 et 3.5.7 , on a

$$\left\| n^{1/2} \sum_{t=1}^n \left\{ \frac{\partial \ell_t(\theta_0)}{\partial \theta} - \frac{\partial \tilde{\ell}_t(\theta_0)}{\partial \theta} \right\} \right\|$$

$$\begin{aligned}
&= \left\| n^{1/2} \sum_{t=1}^2 \left\{ 1 - \frac{\varepsilon_t^2}{h_t} \right\} \left\{ \frac{1}{h_t} \frac{\partial h_t}{\partial \theta} \right\} + n^{1/2} \sum_{t=1}^2 \left\{ \frac{\varepsilon_t^2}{\tilde{h}_t} - 1 \right\} \left\{ \frac{1}{\tilde{h}_t} \frac{\partial \tilde{h}_t}{\partial \theta} \right\} \right\| (\theta_0) \\
&\leq n^{1/2} \sum_{t=1}^2 |1 - \eta_t^2| \left\| \frac{1}{h_t} \frac{\partial h_t}{\partial \theta} \right\| + n^{1/2} \sum_{t=1}^2 |1 - \eta_t^2| \frac{1}{\omega_0} \\
&\leq n^{1/2} \sum_{t=1}^2 (1 + \eta_t^2) \left( \frac{2}{\omega_0} + \frac{1}{\min_{1 \leq i \leq r} (\alpha_{0i})} \right)
\end{aligned}$$

En utilisant l'inégalité de Markov, pour tout  $\epsilon > 0$ ,

$$\begin{aligned}
&P \left( n^{1/2} \sum_{t=1}^2 (1 + \eta_t^2) \left( \frac{2}{\omega_0} + \frac{1}{\min_{1 \leq i \leq r} (\alpha_{0i})} \right) > \epsilon \right) \\
&\leq \frac{4}{\epsilon} n^{-1/2} \left( \frac{2}{\omega_0} + \frac{1}{\min_{1 \leq i \leq r} (\alpha_{0i})} \right) \rightarrow 0
\end{aligned}$$

De la même façon, on a

$$\begin{aligned}
&\sup_{\theta \in \mathcal{V}(\theta_0)} \left\| n^{-1} \sum_{t=1}^n \left\{ \frac{\partial^2 \ell_t(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \tilde{\ell}_t(\theta)}{\partial \theta \partial \theta'} \right\} \right\| \\
&\leq n^{-1} \sum_{t=1}^n \sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \left\{ \frac{\partial^2 \ell_t(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \tilde{\ell}_t(\theta)}{\partial \theta \partial \theta'} \right\} \right\| \\
&= n^{-1} \sum_{t=1}^2 \sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \left\{ 2 \frac{\varepsilon_t^2}{h_t} - 1 \right\} \left\{ \frac{1}{h_t} \frac{\partial h_t}{\partial \theta} \right\} \left\{ \frac{1}{h_t} \frac{\partial h_t}{\partial \theta'} \right\} - \left\{ 2 \frac{\varepsilon_t^2}{\tilde{h}_t} - 1 \right\} \left\{ \frac{1}{\tilde{h}_t} \frac{\partial \tilde{h}_t}{\partial \theta} \right\} \left\{ \frac{1}{\tilde{h}_t} \frac{\partial \tilde{h}_t}{\partial \theta'} \right\} \right\| \\
&\leq n^{-1} \sum_{t=1}^2 \sup_{\theta \in \mathcal{V}(\theta_0)} \left| 2 \frac{\varepsilon_t^2}{h_t} - 1 \right| \sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{1}{h_t} \frac{\partial h_t}{\partial \theta} \right\|^2 + n^{-1} \sum_{t=1}^2 \sup_{\theta \in \mathcal{V}(\theta_0)} \left| 2 \frac{\varepsilon_t^2}{\tilde{h}_t} - 1 \right| \sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{1}{\tilde{h}_t} \frac{\partial \tilde{h}_t}{\partial \theta} \right\|^2 \\
&\leq n^{-1} \sum_{t=1}^2 \left[ (4\eta_t^2(1 + \delta) + 1) \sup_{\theta \in \mathcal{V}(\theta_0)} \left( \frac{1}{\omega} + \frac{1}{\min_{1 \leq i \leq r} (\alpha_{0i})} \right)^2 + (2\eta_t^2(1 + \delta) + 1) \sup_{\theta \in \mathcal{V}(\theta_0)} \frac{1}{\omega^2} \right]
\end{aligned}$$

En utilisant l'inégalité de Markov, pour tout  $\epsilon > 0$ ,

$$\begin{aligned} P \left( n^{1/2} \sum_{t=1}^2 \left[ (4\eta_t^2(1+\delta) + 1) \sup_{\theta \in \mathcal{V}(\theta_0)} \left( \frac{1}{\omega} + \frac{1}{\min_{1 \leq i \leq r}(\alpha_{0i})} \right)^2 + (2\eta_t^2(1+\delta) + 1) \sup_{\theta \in \mathcal{V}(\theta_0)} \frac{1}{\omega^2} \right] > \epsilon \right) \\ \leq \frac{2}{\epsilon} \left[ (4\delta + 5) \sup_{\theta \in \mathcal{V}(\theta_0)} \left( \frac{1}{\omega} + \frac{1}{\min_{1 \leq i \leq r}(\alpha_{0i})} \right)^2 + (2\delta + 3) \sup_{\theta \in \mathcal{V}(\theta_0)} \frac{1}{\omega^2} \right] n^{-1} \longrightarrow 0 \end{aligned}$$

□

**Preuve du Lemme 3.5.10 :**

Pour montrer cette convergence en loi, nous utilisons nous aussi le théorème central limite pour les différences martingales. Soit  $\mathcal{F}_t$  la  $\sigma$ -algèbre engendrée par les variables  $\varepsilon_{t-i}$ ,  $i \geq 0$ . On a  $E_{\theta_0}(\frac{\partial}{\partial \theta} \ell_t(\theta_0) | \mathcal{F}_{t-1}) = 0$  et par le Lemme 3.5.7,  $Var_{\theta_0}(\partial \ell_t(\theta_0) / \partial \theta)$  existe. Par le Lemme 3.5.7, les hypothèses **A4'** et **A6** et le fait que  $0 < \kappa_\eta < \infty$ ,  $Var_{\theta_0}(\partial \ell_t(\theta_0) / \partial \theta)$  est non dégénérée. Donc pour tout  $\lambda \in \mathbb{R}^{r+1}$ , la suite  $\{\lambda' \frac{\partial}{\partial \theta} \ell_t(\theta), \mathcal{F}_t\}_t$  est une différence martingale stationnaire et carrée intégrable. En utilisant le théorème central limite de Billingsley (1961) et la technique de Cramer-Wold, nous avons la normalité asymptotique.

□

**Preuve du Lemme 3.5.11 :**

La preuve de ce Lemme est la même que dans Francq et Zakoïan (2003), Théorème 2.2.

□

La méthode de maximum de vraisemblance nous permet d'estimer le modèle (3.1) avec des estimateurs convergents quelque soit le nombre  $r$  de changements de régime.

Cette dernière sera utilisée dans le chapitre suivant pour estimer notre modèle pour des données simulées et réelles.



# Bibliographie

- [1] **Baillie, R.T. et H. Chung (2001)** *Estimation of GARCH Models From the Autocorrelations of the Squares of a Process*, Journal of Time Series Analysis, 22, 631–650.
- [2] **Berkes, I., Horváth, L. et P. Kokoszka (2003)** *GARCH Processes : Structure and Estimation*, A paraître dans Bernoulli.
- [3] **Billingsley, P.(1995)** *Probability and Measure*, John Wiley, New York.
- [4] **Bollerslev, T. (1986)** *Generalized Autoregressive Conditional Heteroskedasticity*, Journal of Econometrics, 31, 307–327.
- [5] **Boussama, F. (1998)** *Ergodicité, Mélange et Estimation dans les Modèles GARCH*, Thèse de l'Université Paris-7.
- [6] **Boussama, F. (2000)** *Normalité Asymptotique de l'Estimateur du Pseudo-Maximum de Vraisemblance d'un Modèle GARCH*, C. R. Acad. Sci. Paris, 331, 81–84.
- [7] **Brockwell, P. J. et R. A. Davis (1991)** *Time Series : Theory and Methods*, Springer-Verlag, New York.

- [8] **Engle, R. F., (1982)** *Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of the United Kingdom Inflation*, *Econometrica*, 50, 987–1007.
- [9] **Francq, C., et J-M. Zakoïan (2003)** *Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes*, A paraître dans *Bernoulli*.
- [10] **Horváth, L. et F. Liese (2003)**  *$L_p$  Estimators in ARCH Models*, A paraître dans *Journal of Statistical Planning and Inference*.
- [11] **Lee, S.-W. et B. E. Hansen (1994)** *Asymptotic Theory For the GARCH(1,1) Quasi-Maximum Likelihood Estimator*, *Econometric Theory*, 10, 29–52.
- [12] **Lumsdaine, R. L. (1996)** *Consistency and Asymptotic Normality of the Quasi-Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models*, *Econometrica*, 64, 575–596.
- [13] **Rabemananjara R. et J.M. Zakoïan (1993)** : *Threshold ARCH Models and Asymmetries in Volatility*, *Journal of Applied Econometrics*, 8, 31–49.
- [14] **Weiss, A. A. (1986)** *Asymptotic Theory For ARCH Models : Estimation and Testing*, *Econometric Theory*, 2, 107–131.

# Chapitre 4

## Applications empiriques

Dans ce chapitre, nous illustrons la pertinence des modèles non linéaires conditionnellement hétéroscédastiques développés dans les parties précédentes à partir des données réelles. Nous commençons par présenter une méthode de test de l'existence de plusieurs régimes.

### 4.1 Test de l'existence de plusieurs régimes

L'inférence des modèles à seuils (inconnus) est composée d'une partie identification des valeurs des seuils et d'une partie estimation classique. Des tests qui permettent de justifier la présence de plusieurs régimes ont été introduits dans la littérature. Ce problème de détection de seuils a attiré l'attention de nombreux auteurs (voir Chan et Tong (1990), Hansen (1996)). Par exemple, Tsay (1989, 1998) propose un test statistique pour des modèles AR à seuils pour tester l'hypothèse de linéarité contre celle de présence de seuils.

Nous proposons d'adapter la méthode de Tsay (1989,1998) développée initialement pour des modèles à seuil (TAR) sur la moyenne. Cette méthode ne nécessite pas d'estimer le modèle contraint (plusieurs régimes) et donc de connaître les seuils. Elle se fonde sur l'orthogonalité des résidus prédictifs par rapport aux variables passées lorsque le modèle linéaire (dans notre cas ARCH(1) classique) est bien spécifié.

Dans cette partie, nous utilisons le modèle (3.1) avec  $r = 2$ ,  $d = 1$  (c à d il existe  $k > 0$ ,  $R_1 = [0, k]$  et  $R_2 = ]k, +\infty[$ ) que nous réécrivons pour faciliter la présentation.

$$\varepsilon_t = \sqrt{h_t}\eta_t \quad , \quad h_t = \begin{cases} \omega_0 + \alpha_{01}\varepsilon_{t-1}^2 & \text{si } Z_{t-1} \leq k \\ \omega_0 + \alpha_{02}\varepsilon_{t-1}^2 & \text{si } Z_{t-1} > k \end{cases} \quad (4.1)$$

Nous rappelons que  $Z_{t-1}$  est une variable à valeurs dans  $\mathbb{R}^+$  appartenant à la tribu  $\mathcal{F}_{t-1}$  engendrée par  $\{\varepsilon_{t-j}, j > 0\}$ . Par exemple,  $Z_t = \frac{\varepsilon_{t-1}^2}{\varepsilon_{t-2}^2}$ , ou  $Z_t = \varepsilon_{t-1}^2$ . Le but est de détecter la présence de changements de régime et de localiser le seuil  $k$ .

On dispose d'observations  $\varepsilon_1, \dots, \varepsilon_n$ ; ainsi que  $n_0$  valeurs initiales, qui permettent d'obtenir les  $Z_t$ ,  $t = 1, \dots, n$ .

On souhaite tester

$$\begin{array}{ccc} H_0 : \alpha_{01} = \alpha_{02} \quad \text{ou} \quad k = 0 & \text{contre} & H_a : \alpha_{01} \neq \alpha_{02} \quad \text{et} \quad k > 0 \\ \text{(un seul régime)} & & \text{(deux régimes)} \end{array}$$

Pour construire le test, nous considérons la statistique d'ordre sur  $(Z_t)$  et notons

$Z_{(i)}$  le  $i^{\text{eme}}$  plus petit élément. Posons  $\tau(i)$  la date (aléatoire) correspondant à  $Z_{(i)}$  :

$$\forall i = 1, 2, \dots, n, \quad \exists! t = \tau(i), \quad Z_t = Z_{(i)}$$

La statistique d'ordre sur  $Z_t$  permet de ranger les  $\varepsilon_t$  dans l'ordre suivant :  $\varepsilon_{\tau(1)}, \dots, \varepsilon_{\tau(n)}$ .

Avec ces notations, les équations du modèle (4.1) s'écrivent, pour ces observations :

$$\forall i = 1, 2, \dots, n, \quad \begin{cases} \varepsilon_{\tau(i)} = h_{\tau(i)}^{\frac{1}{2}} \eta_{\tau(i)}, \\ h_{\tau(i)} = \omega_0 + \alpha_{01} \varepsilon_{\tau(i)-1}^2 & \text{si } i \leq s \\ h_{\tau(i)} = \omega_0 + \alpha_{02} \varepsilon_{\tau(i)-1}^2 & \text{si } i > s \end{cases} \quad (4.2)$$

où  $s$  satisfait  $Z_{(s)} \leq k < Z_{(s+1)}$ . C'est un classement tel que les  $s$  premières observations ( $\varepsilon_{\tau(i)}$ ) sont dans le premier régime et les  $n - s$  suivantes dans le second. Pour ce classement, seul le nombre  $s$  dépend du seuil  $k$ .

Considérons le modèle de régression

$$\varepsilon_{\tau(i)}^2 = \omega + \alpha \varepsilon_{\tau(i)-1}^2 + \nu_{\tau(i)} \quad i = 1, \dots, n \quad (4.3)$$

Soit  $\hat{\theta}^{(m)} = (\hat{\omega}^{(m)}, \hat{\alpha}^{(m)})$  l'estimateur des moindres carrés de  $\theta = (\omega, \alpha)$  obtenu à partir des  $m$  premières valeurs des  $(\varepsilon_{\tau(i)}^2)$ . L'idée du test est que si  $(\varepsilon_t)$  est de type (4.1), quand  $i \leq s$  les résidus  $(\nu_{\tau(i)})$  constituent approximativement un bruit blanc donc sont (approximativement) orthogonaux aux valeurs passées  $\varepsilon_{\tau(i)-1}^2$  ( $\nu_{\tau(i)}$  non corrélées avec  $\varepsilon_{\tau(i)-1}^2$ ). Quand  $m$  excède  $s$ , les résidus  $(\nu_{\tau(i)})$  vont être biaisés car le régime change à l'indice  $(s + 1)$ . Donc l'orthogonalité entre les résidus  $(\nu_{\tau(i)})$  et  $(\varepsilon_{\tau(i)-1}^2)$  va se détruire progressivement à partir du moment où le nombre d'observations  $m$  prend des valeurs supérieures à  $s$ .

Les étapes du test sont les suivantes :

Étape 1 : Reclassement des observations selon la statistique d'ordre sur  $Z_t$  et le calcul des estimateurs  $\hat{\theta}^{(m)} = (\hat{\omega}^{(m)}, \hat{\alpha}^{(m)})$ ,  $m = m_0, m_0 + 1, \dots, n - 1$ , pour une valeur  $m_0$  arbitraire suffisamment grande, par la méthode des moindres carrés.

Étape 2 : Calcul des résidus  $\hat{\nu}_{\tau(m+1)} = \varepsilon_{\tau(m+1)}^2 - \hat{\omega}^{(m)} - \hat{\alpha}^{(m)} \varepsilon_{\tau(m+1)-1}^2$  et calcul des résidus normalisés

$$\hat{\vartheta}_{\tau(m+1)} = \hat{\nu}_{\tau(m+1)} / [1 + \varepsilon_{\tau(m)}^4 (\sum_{i=1}^m \varepsilon_{\tau(i)}^2)^{-1}]^{1/2}, \quad m = m_0, m_0 + 1, \dots, n - 1.$$

Étape 3 : Régression, par la méthode des moindres carrés,

$$\hat{\vartheta}_{\tau(l)} = \Psi_0 + \Psi_1 \varepsilon_{\tau(l)-1}^2 + w_{\tau(l)}, \quad l = m_0 + 1, m_0 + 2, \dots, n. \quad (4.4)$$

et calcul des résidus  $\hat{w}_{\tau(l)}$ .

Étape 4 : Test de l'hypothèse nulle  $H'_0 : \Psi_0 = \Psi_1 = 0$  contre l'hypothèse alternative  $H'_a : \Psi_0 \neq 0$  ou  $\Psi_1 \neq 0$ . Si  $H'_0$  est rejetée on rejette  $H_0$ .

Si, dans l'étape 4, l'hypothèse  $H'_0$  est vraie, alors la statistique :

$$\hat{F} = \frac{\sum \hat{\vartheta}_t^2 - \sum \hat{w}_t^2}{\sum \hat{w}_t^2} \frac{n - m_0 - 2}{2} \quad (4.5)$$

est approximativement, pour  $n$  grand, de loi de Fisher-Snedecor  $F(2, n - m_0 - 2)$ .

D'où la règle de décision

- Si  $\hat{F} > F_{1-\alpha}(2, n - m_0 - 2)$ , alors  $H_0$  est rejetée au profit de  $H_a$  ;
- Si  $\hat{F} \leq F_{1-\alpha}(2, n - m_0 - 2)$ , alors  $H_0$  n'est pas rejetée au profit de  $H_a$  ;

où  $\alpha$  est le niveau de risque du test.

**Remarque :**

La loi asymptotique de  $\hat{F}$  ne change pas si on utilise les résidus  $(\hat{\nu}_t)$  à la place des

$(\hat{\vartheta}_t)$  dans la régression (4.4). Pour les séries de taille faible, les expériences montrent qu'il est préférable d'utiliser les résidus  $(\hat{\vartheta}_t)$ .

Hormis le test précédent, la procédure de régressions successives permet de localiser  $k$ . Dans la partie suivante nous décrivons un procédé graphique permettant cette localisation.

## 4.2 Étude par simulations

Dans cette partie, nous effectuons le test précédent sur des séries simulées et utilisons la classification pour localiser  $k$ . Nous examinons ensuite le comportement des estimateurs présentés au chapitre 3.

### 4.2.1 Test

Nous utilisons 100 simulations, de taille  $n = 2400$ , du modèle (4.1) avec  $Z_t = \frac{\varepsilon_t^2 - 1}{\varepsilon_t^2 - 2}$  et  $(k, \omega, \alpha_1, \alpha_2) = (4, 0.1, 0.1, 0.6)$  (pour une étude de puissance) et d'un ARCH(1) avec  $(\omega, \alpha) = (0.1, 0.2)$  (pour une étude de niveau). Nous appliquons les étapes 1, 2, 3 du test précédent pour  $m_0 = 500$ . Cette valeur est suffisamment grande pour fournir des estimateurs assez précis dès les premières valeurs. Nous donnons les résultats dans le Tableau 4.1.

On constate que la fréquence empirique de rejet sous  $H_0$  dépasse la valeur nominale de 5%. Sous l'alternative  $H_a$ , cette fréquence dépasse la valeurs nominale de 1%.

TAB. 4.1 – Test de changement de régime : fréquences de rejets d'un ARCH(1) sur 100 simulations.

Paramètres	Seuil	Fréquence des rejets de $H_0$
$\omega = 0.1, \alpha = 0.2$ (sous $H_0$ )	1%	3%
	5%	7%
$k = 4, \omega = 0.1, \alpha_1 = 0.1, \alpha_2 = 0.2$ (sous $H_a$ )	1%	98%
	5%	99%

#### 4.2.2 Localisation du seuil $k$ et estimation des paramètres

Nous reprenons l'une des 100 simulations précédentes pour laquelle le changement de régime a été validé par les résultats de test. D'après les résultats de cette simulation, nous notons que le passage de  $Z_t$  par la valeur  $k = 4$  se fait entre  $i_0 = 1718$  et  $i_0 + 1 = 1719$ . Nous devons nous attendre à voir apparaître une rupture dans les coefficients estimés pour des  $i > i_0$ .

Pour localiser  $k$ , nous représentons graphiquement dans la Figure 4.1, les valeurs des estimateurs de  $\alpha$  et  $\omega$  de la régression ( $800 \leq m \leq 2400$ ) de l'étape 1. Sur le graphique des  $\alpha$ , nous remarquons de fortes variations entre  $m = 1700$  et  $2000$ . Nous avons observé ce phénomène chez plusieurs simulations. Pour affiner cette localisation, nous balayons la région  $Z_{(i)} \in [2.4, 21.2]$  (i.e 2.4, 2.5, ..., 21.2) en calculant les estimateurs de  $\alpha_1$  et  $\alpha_2$  par la méthode de maximum de vraisemblance. Nous avons fixé dans notre programme  $(\omega, \alpha_1, \alpha_2) = (1.1, 0.1, 0.1)$ . Nous avons choisi  $\omega$  plus grande que la valeur réelle car dans la pratique les paramètres sont inconnus.

Nous choisissons la valeur  $k$  qui correspond à la plus grande vraisemblance. Nous

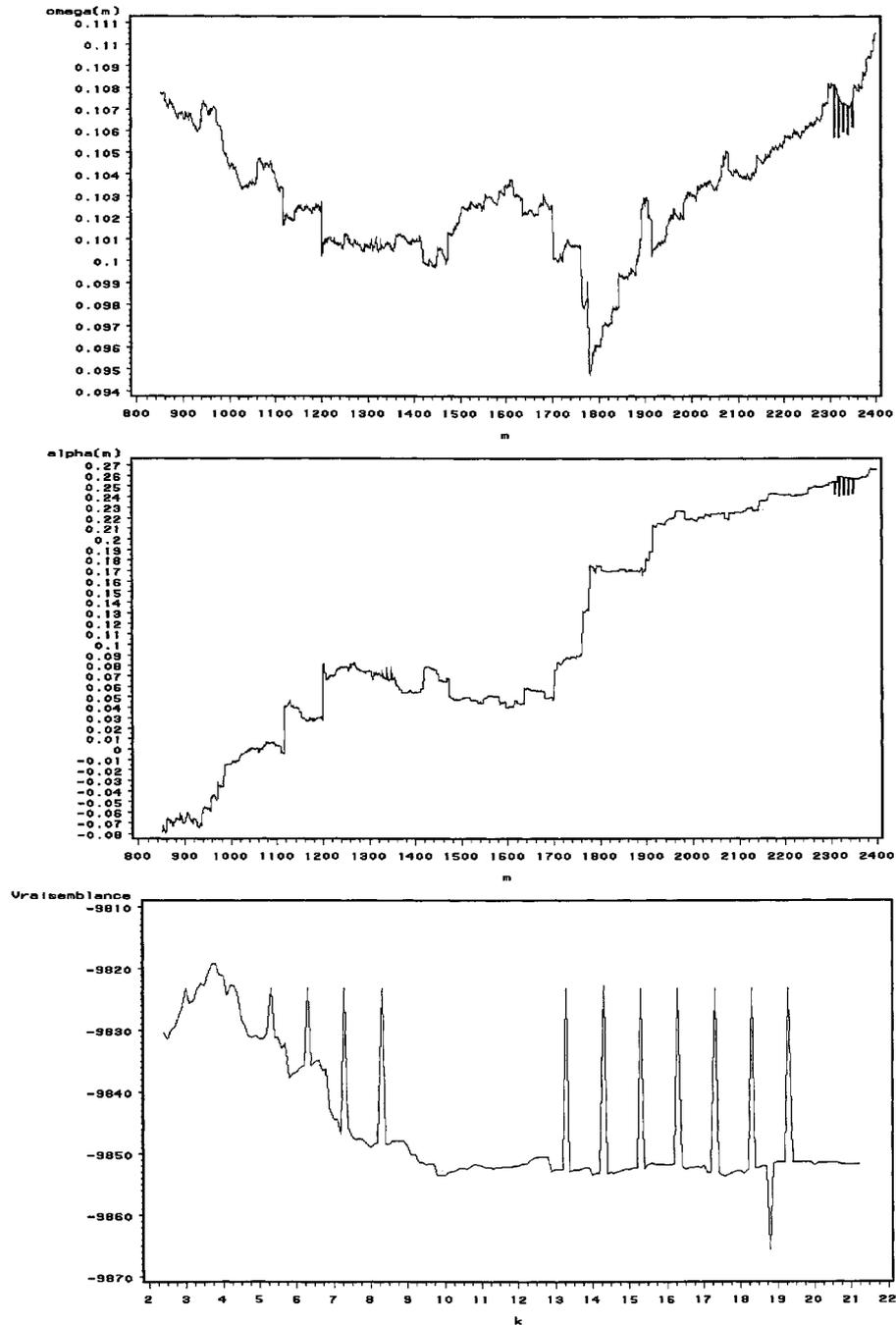


FIG. 4.1 – Localisation graphique de  $k$  pour une simulation de taille 2400 avec  $(k, \omega, \alpha_1, \alpha_2) = (4, 0.1, 0.1, 0.6)$ . Graphique du haut : en abscisses : taille  $m$  de l'échantillon ; en ordonnées : valeurs de  $\hat{\omega}^{(m)}$ . Graphique du milieu : en abscisses : taille  $m$  de l'échantillon ; en ordonnées : valeurs de  $\hat{\alpha}^{(m)}$ . Graphique du bas : vraisemblance en fonction de  $k$ .

affinons, le balayage autour de cette valeur avec une précision de l'ordre  $10^{-2}$  et nous constatons que  $k = 3.8$  a la plus grande vraisemblance dans la région de balayage. Le graphique du bas de la Figure 4.1 nous confirme le choix de  $k = 3.8$ . Nous pouvons expliquer ce biais dans la localisation de  $k$  par le mauvais choix des valeurs initiales des paramètres dans le programme d'estimation. Nous pouvons justement éviter ce problème en estimant ces valeurs par la méthode des moments que nous avons présentée précédemment.

### 4.2.3 Méthodes d'estimations

Dans cette partie numérique, nous comparons sur  $N = 100$  échantillons simulés la méthode des moments et celle du maximum de vraisemblance présentées dans le chapitre 3.

Dans la première application de l'estimation par la méthode de Maximum de vraisemblance, nous testons cette dernière pour les valeurs ( $k = 4$ ,  $\omega = 0.1$ ,  $\alpha_1 = 0.1$ ,  $\alpha_2 = 0.2$ ). Nous étudions dans l'expérience 1, cette méthode pour le cas où la taille de l'échantillon est  $n = 100$ . Dans les expériences 2 et 3, nous traitons successivement les cas où  $n = 1000$  et  $n = 5000$ . A chaque fois  $N = 100$  échantillons indépendants sont générés.

Dans la deuxième application, nous étudions la même méthode mais pour d'autres données qui sont le résultat d'une simulation de notre modèle selon des nouvelles valeurs ( $k = 1$ ,  $\omega = 0.1$ ,  $\alpha_1 = 0.3$ ,  $\alpha_2 = 0.15$ ). Nous nous limitons à une seule taille d'échantillon  $n = 1000$ .

Dans la troisième application, nous testons ainsi la même méthode (MV) pour le cas où  $k$  est assez grand ( $k = 6$ ,  $\omega = 0.1$ ,  $\alpha_1 = 0.3$ ,  $\alpha_2 = 0.15$ ).

Pour tester la méthode des moments, nous réutilisons les 100 échantillons de taille  $n=1000$  pour les différentes valeurs de  $k = 1, 4, 6$ . Nous l'avons comparé avec la méthodes de MV. Les résultats sont donnés dans des tableaux.

### Première application de la méthode de MV

Nous générons suivant le modèle (4.1) un nombre  $N$  d'échantillons de taille  $n$ . Pour ne pas influencer les résultats, nous avons pris, dans le programme d'estimation par la méthode de MV, des valeurs initiales différentes des vraies valeurs ( $\omega=\alpha_2=\alpha_1=0.5$ ). Nous donnons les résultats numériques obtenus selon les expériences suivantes :

**Expérience 1 :  $n=100$ ,  $N=100$ .** Nous avons testé cette méthode sur  $N = 100$  simulations du modèle (4.1) de taille  $n = 100$ , avec ( $k = 4$ ,  $\omega = 0.1$ ,  $\alpha_1 = 0.1$ ,  $\alpha_2 = 0.2$  ). Un exemple de trajectoire est donné dans le graphique 4.2.

Le Tableau 4.2 présente des statistiques descriptives pour ces simulations. Celles-ci étant menées sans contrainte, nous constatons parfois des valeurs négatives pour les paramètres  $\alpha_1$  et  $\alpha_2$ . Le nombre de ces valeurs négatives est noté  $n_0$ .

**Expérience 2 :  $n=1000$ ,  $N=100$ .** Pour les mêmes valeurs réelles de  $\omega$ ,  $\alpha_2$  et  $\alpha_1$ , nous avons augmenté la taille des échantillons à 1000. Nous avons testé à nouveau la procédure d'estimation sur 100 échantillons. Le Tableau 4.3 montre une amélioration des estimations, par comparaison avec le Tableau 4.2, à la fois en terme de biais et de précision. Par ailleurs le nombre d'estimations négatives diminue fortement.

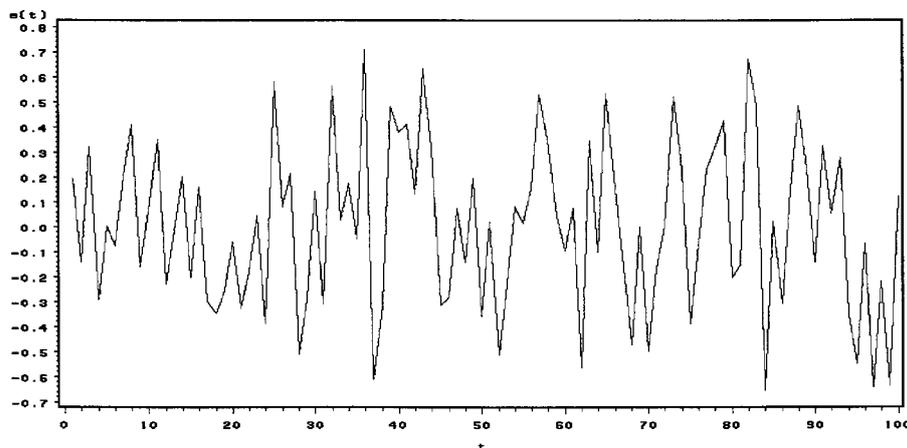


FIG. 4.2 – une simulation du modèle (4.1) [ $k = 4$ ,  $\omega = 0.1$ ,  $\alpha_1 = 0.1$ ,  $\alpha_2 = 0.2$ ]

**Expérience 3 :  $n=5000$ ,  $N=100$ .** La même expérience est conduite pour une taille 5000. On constate (Tableau 4.4) que les valeurs négatives disparaissent complètement, et que tous les paramètres sont bien estimés.

La variabilité est moindre pour  $\omega$ , cependant, ce qui peut s'expliquer par le fait que ce paramètre apparaît dans les deux régimes alors que  $\alpha_1$  et  $\alpha_2$  n'apparaissent que dans un seul.

### Deuxième application de la méthode de MV

Nous utilisons d'autres échantillons qui correspondent à une simulation de notre modèle selon les nouvelles valeurs ( $k = 1$ ,  $\omega = 0.1$ ,  $\alpha_1 = 0.3$ ,  $\alpha_2 = 0.15$ ) en nous limitant à une taille de 1000. Le Tableau 4.5 fait apparaître des résultats voisins de ceux de la Table 4.4.

Estimation des paramètres par la méthode de MV								
$\theta_0$	Moy	E-type	Min	Q1	Med	Q3	Max	$n_0$
$k = 4$								
$\omega = 0.1$	0,104	0,023	0,054	0,0912	0,1025	0,115	0,184	0
$\alpha_1 = 0.1$	0,059	0,1732	-0,293	-0,058	0,052	0,156	0,567	42
$\alpha_2 = 0.2$	0,196	0,212	-0,162	0,032	0,176	0,342	0,884	19

TAB. 4.2 -  $n=100, N=100$ 

Estimation des paramètres par la méthode de MV								
$\theta_0$	Moy	E-type	Min	Q1	Méd	Q3	Max	$n_0$
$k = 4$								
$\omega = 0.1$	0,100	0,006	0,087	0,096	0,100	0,104	0,119	0
$\alpha_1 = 0.1$	0,110	0,047	-0,017	0,075	0,110	0,148	0,224	1
$\alpha_2 = 0.2$	0,198	0,054	0,064	0,160	0,200	0,230	0,353	0

TAB. 4.3 -  $n = 1000 N = 100$ 

### Troisième application de la méthode de MV

Dans cette application, nous augmentons la valeur de  $k$  ( $k = 6$ ) afin de tester la performance de cette méthode dans ce type de situation. Les résultats sont donnés dans le Tableau 4.6. D'après les résultats de la première application, en utilisant une taille d'échantillon plus grand que  $n = 1000$ , nous pourrions améliorer les estimations de  $\alpha_1$ . Nous avons obtenu une seule estimation négative concernant seulement  $\alpha_2$ .

Estimation des paramètres par la méthode de MV

$\theta_0$	Moy	E-type	Min	Q1	Méd	Q3	Max	$n_0$
$k = 4$								
$\omega = 0.1$	0,100	0,003	0,092	0,098	0,100	0,102	0,105	0
$\alpha_1 = 0.1$	0,103	0,025	0,041	0,086	0,101	0,119	0,163	0
$\alpha_2 = 0.2$	0,201	0,028	0,128	0,186	0,199	0,218	0,263	0

TAB. 4.4 -  $n = 5000$   $N = 100$ 

Estimation des paramètres par la méthode de MV

$\theta_0$	Moy	E-type	Min	Q1	Méd	Q3	Max	$n_0$
$k = 1$								
$\omega = 0.1$	0,101	0,006	0,086	0,097	0,100	0,104	0,117	0
$\alpha_1 = 0.3$	0,286	0,118	0,068	0,202	0,279	0,362	0,653	0
$\alpha_2 = 0.15$	0,147	0,041	0,057	0,117	0,145	0,173	0,278	0

TAB. 4.5 -  $n = 1000$   $N = 100$ 

### 4.3 Application de la méthode des moments

Dans cette méthode, les estimateurs de  $\omega$ ,  $\alpha_1$  et  $\alpha_2$  ont une forme explicite. Nous souhaitons comparer cette méthode avec celle de maximum de vraisemblance. Pour cela, nous avons réutilisé ici les mêmes échantillons de taille 1000.

Les résultats d'estimation pour les valeurs ( $k = 4$ ,  $\omega = 0.1$ ,  $\alpha_1 = 0.1$ ,  $\alpha_2 = 0.2$ ) sont donnés dans le Tableau 4.7. Par comparaison avec la Table 4.3, nous remarquons que le paramètre  $\omega$  s'estime mieux par la méthode de maximum de vraisemblance

Estimation des paramètres par la méthode de MV

$\theta_0$	Moy	E-type	Min	Q1	Méd	Q3	Max	$n_0$
$k = 6$								
$\omega = 0.1$	0,103	0,007	0,082	0,097	0,104	0,108	0,120	0
$\alpha_1 = 0.3$	0,180	0,062	0,043	0,143	0,183	0,215	0,334	0
$\alpha_2 = 0.15$	0,141	0,068	-0,035	0,095	0,138	0,177	0,322	1

TAB. 4.6 -  $n = 1000$   $N = 100$ 

avec une variance plus faible. Par contre, la moyenne des estimations du paramètre  $\alpha_1$  est plus proche de la valeur réelle par la méthode des moments; mais le nombre de mauvaises estimations par cette dernière (onze valeurs négatives contre une pour celle de MV) est très important. Les moyennes d'estimation du paramètre  $\alpha_2$  par les deux méthodes sont très proches. Une variance plus petite et une médiane plus proche de la valeur réelle font que la méthode de maximum de vraisemblance est plus précise que celle de moments pour ce cas.

Nous testons la méthode pour les valeurs des paramètres ( $k = 1$ ,  $\omega = 0.1$ ,  $\alpha_1 = 0.3$ ,  $\alpha_2 = 0.15$ ). Nous présentons les résultats des estimations dans le Tableau 4.8. Pour ce qui concerne  $\omega$ , cette méthode l'estime presque avec la même précision que celle de maximum de vraisemblance. Les paramètres  $\alpha_1$  et  $\alpha_2$  ne sont pas mieux estimés en comparant le Tableau 4.8 avec celui de la méthode du maximum de vraisemblance (voir le Tableau 4.5). Nous avons obtenu cinq estimations négatives de  $\alpha_1$  par cette méthode alors que par celle de maximum de vraisemblance toutes les estimations sont positives.

Estimation des paramètres par la méthode des moments

$\theta_0$	Moy	E-type	Min	Q1	Méd	Q3	Max	$n_0$
$k = 4$								
$\omega = 0.1$	0,101	0,007	0,082	0,096	0,101	0,106	0,117	0
$\alpha_1 = 0.1$	0,094	0,073	-0,104	0,047	0,101	0,141	0,270	11
$\alpha_2 = 0.2$	0,198	0,056	0,083	0,157	0,195	0,226	0,362	0

TAB. 4.7 -  $n = 1000$   $N = 100$ 

Les paramètres  $\alpha_1$  et  $\alpha_2$  s'estiment mieux par la méthode de maximum de vraisemblance que par cette méthode (voir le Tableau 4.9). D'après les résultats que nous avons obtenus, nous pouvons juger que la méthode de maximum de vraisemblance donne des meilleurs résultats que celle des moments pour de grandes valeurs de  $k$ .

Estimation des paramètres par la méthode des moments

$\theta_0$	Moy	E-type	Min	Q1	Méd	Q3	Max	$n_0$
$k = 1$								
$\omega = 0.1$	0,101	0,008	0,074	0,097	0,101	0,105	0,120	0
$\alpha_1 = 0.3$	0,274	0,178	-0,110	0,145	0,265	0,370	0,706	5
$\alpha_2 = 0.15$	0,146	0,047	0,048	0,115	0,141	0,174	0,295	0

TAB. 4.8 -  $n = 1000$   $N = 100$

Estimation des paramètres par la méthode des moments
--

$\theta_0$	Moy	E-type	Min	Q1	Méd	Q3	Max	$n_0$
$k = 6$								
$\omega = 0.1$	0.103	0.008	0.082	0.099	0.104	0.108	0.124	0
$\alpha_1 = 0.3$	0.165	0.076	-0.042	0.116	0.162	0.213	0.426	1
$\alpha_2 = 0.15$	0.137	0.071	-0.045	0.084	0.138	0.177	0.351	2

TAB. 4.9 -  $n = 1000$   $N = 100$

## 4.4 Résultats Empiriques : l'indice CAC 40

L'indice CAC 40 est l'indice de la Bourse de Paris qui porte sur les 40 plus grandes ou plus représentatives sociétés françaises cotées au Premier Marché. CAC signifie cotation assistée en continu (avant, Cie des agents de change).

Nous utilisons la série journalière des rendements du CAC 40 (logarithme du rapport du prix à la date  $t$  sur le prix à la date  $t - 1$ ) sur la période 1988-98 qui correspond à 2385 observations. Nous donnons sa trajectoire dans la Figure 4.3.

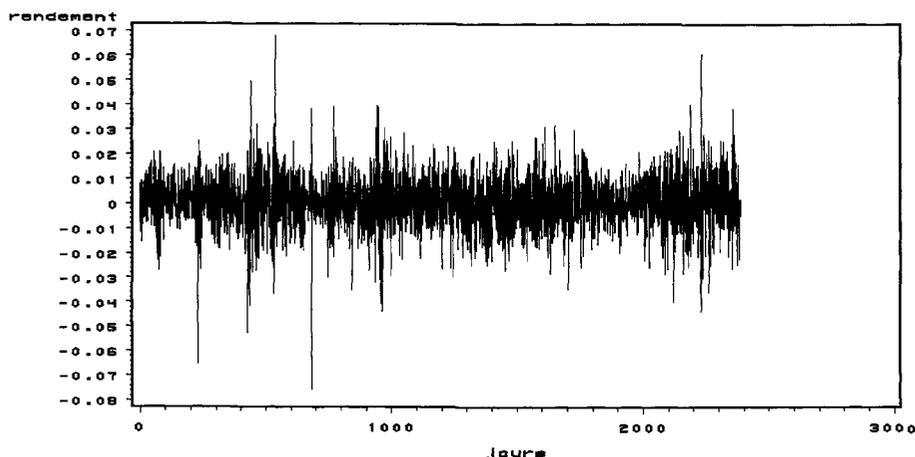


FIG. 4.3 – Série journalière des rendements du CAC 40 sur la période 1988-98 (2385 observations)

### 4.4.1 Test de changements de régime

Nous nous plaçons pour la suite de cette étude dans le cadre du modèle (4.1) avec  $Z_t = \frac{\varepsilon_t^2 - 1}{\varepsilon_t^2 - 2}$ . Nous commençons par appliquer le test de changement de régime introduit dans la section 4.1 avec  $m_0 = 500$ . En utilisant (4.5), nous avons calculé  $\hat{F} = 11.83$ . Pour un niveau de risque  $\alpha = 0.01$  on a :  $F_{1-\alpha}(2, n - m_0 - 2) = F_{0.99}(2, 1781) = 4.61$ .

Donc, d'après la règle de décision,  $H_0$  est largement rejetée au profit de  $H_a$ . Ce qui nous permet de conclure à l'existence de changements de régime de type (4.1).

#### 4.4.2 Estimation des paramètres

Afin de localiser  $k$ , nous représentons graphiquement les estimations de  $\hat{\omega}^{(m)}$  et  $\hat{\alpha}^{(m)}$  de l'étape 1 du test (voir les deux premiers graphiques de la Figure 4.4). Nous observons une nette décroissance des estimations de  $\alpha$  à partir de valeurs de  $m$  proches de 1600, correspondant à  $Z_{(i)} = 3.22$ . Ceci nous permet de localiser  $k$  avant cette valeur.

Afin d'affiner la localisation, nous balayons la région  $Z_{(i)} \in [1.23, 3.21]$  en calculant les estimations de  $\omega$ ,  $\alpha_1$  et  $\alpha_2$  par la méthode du maximum de vraisemblance. Nous choisissons, selon le premier balayage,  $\hat{k} = 2.5$  qui correspond à la plus grande vraisemblance.

Nous affinons le balayage autour de cette valeur avec une précision de l'ordre  $10^{-2}$  et nous constatons que cette valeur a vraiment la plus grande vraisemblance. Le troisième graphique de la Figure 4.4 nous confirme le choix de  $\hat{k} = 2.5$ .

Nous estimons, par la méthode du maximum de vraisemblance, le modèle (4.1) pour  $k = 2.5$ . Nous donnons les estimations des paramètres  $\omega$ ,  $\alpha_1$  et  $\alpha_2$  ainsi que leurs caractéristiques dans le Tableau 4.10.

Afin de vérifier la plausibilité du modèle (4.1) estimé (voir le Tableau 4.10), nous le comparons avec d'autres modèles ARCH en utilisant des simulations.

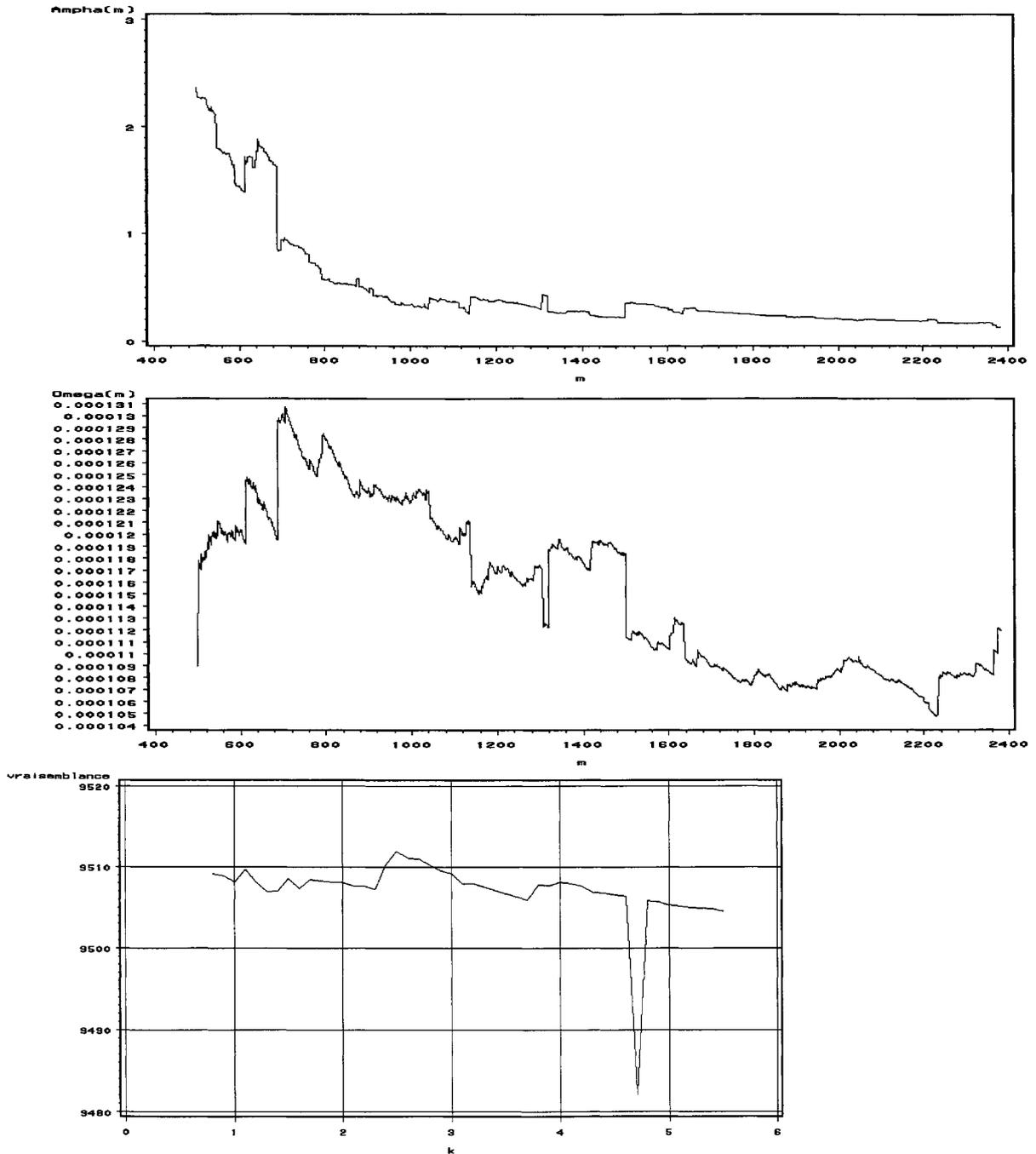


FIG. 4.4 – Rendements du CAC 40 : localisation graphique de  $k$ . Graphique du haut : en abscisses : taille  $m$  de l'échantillon ; en ordonnées : valeurs de  $\hat{\omega}^{(m)}$ . Graphique du milieu : en abscisses : taille  $m$  de l'échantillon ; en ordonnées : valeurs de  $\hat{\alpha}^{(m)}$ . Graphique du bas : vraisemblance en fonction de  $k$ .

TAB. 4.10 – Rendements du CAC 40 : Estimation par MV des paramètres du modèle (4.1)

Paramètres	Estimation	Ecart-type estimé [RATS]	T-stat	Signif
$k = 2.5$				
$\omega$	0.0001	$0.26 * 10^{-5}$	41.97	0.00
$\alpha_1$	0.30	$0.49 * 10^{-1}$	6.12	0.00
$\alpha_2$	0.07	$0.16 * 10^{-1}$	4.43	$9.6 * 10^{-6}$

En utilisant la série des rendements du CAC 40, nous avons estimé ainsi par la méthode du maximum de vraisemblance deux spécifications ARCH(2) et GARCH(1,1), nous avons comparé les séries des simulations des trois spécifications avec celle du CAC 40. Le Tableau 4.11 présente les caractéristiques des distributions marginales des séries. Le Tableau 4.12 présente les autocorrélations empiriques de ces séries (réelles et simulées) et leurs carrés.

Comme on pouvait s'y attendre, les autocorrélations empiriques des quatre séries sont proches de 0. Des différences apparaissent dans les autocorrélations des carrés. Le modèle (4.1) semble plus proche de la série réelle pour l'autocorrélation d'ordre 1 que les autres modèles, mais les résultats sont inversés pour les ordres supérieurs.

TAB. 4.11 – Caractéristiques empiriques de la série des rendements du CAC 40 et de celles des simulées des modèles estimés (4.1), ARCH(2) et GARCH(1,1) (taille  $n = 2385$ ).

Moyenne	E-type	Min	Q1	Méd	Q3	Max
<b>Série du CAC 40</b>						
0.0004	0.0113	-0.0757	-0.0062	0.0004	0.0073	0.0681
<b>Simulé du modèle (4.1) avec <math>[k = 2.5, \omega = 0.0001, \alpha_1 = 0.30, \alpha_2 = 0.07]</math></b>						
0.0002	0.0113	-0.0702	-0.007	0.0002	0.0075	0.0369
<b>Simulé d'un ARCH(2) avec <math>[\omega = 0.0001, \alpha_1 = 0.096, \alpha_2 = 0.112]</math></b>						
$10^{-5}$	0.0114	-0.0451	-0.0075	0.0001	0.0074	0.0439
<b>Simulé d'un GARCH(1,1) avec <math>[\omega = 9 * 10^{-6}, \alpha_1 = 0.09, \beta_1 = 0.84]</math></b>						
-0.0002	0.0112	-0.0430	-0.0079	-0.0004	0.0077	0.0443
<b>Carrés des valeurs de la série du CAC 40</b>						
0.0001	0.0003	$1.04 * 10^{-10}$	$9.68 * 10^{-5}$	$4.48 * 10^{-5}$	0.0001	0.0057
<b>Carrés des valeurs simulées du modèle (4.1)</b>						
0.0001	0.0002	$2.9 * 10^{-12}$	$1.43 * 10^{-5}$	$5.7 * 10^{-5}$	0.0002	0.0017
<b>Carrés des valeurs simulées du modèle ARCH(2)</b>						
0.0001	0.0002	$1.4 * 10^{-10}$	$1.3 * 10^{-5}$	$5.5 * 10^{-5}$	0.0002	0.0020
<b>Carrés des valeurs simulées du modèle GARCH(1,1)</b>						
0.0001	0.0002	$7.1 * 10^{-11}$	$1.4 * 10^{-5}$	$6 * 10^{-5}$	0.0002	0.0020

TAB. 4.12 – Autocorrélations empiriques d'ordre 1,2 et 3 de la série des rendements du CAC 40 et de celles des simulées des modèles estimés : (4.1) [ $k = 2.5, \omega = 0.0001, \alpha_1 = 0.30, \alpha_2 = 0.07$ ], ARCH(2) [ $\omega = 0.0001, \alpha_1 = 0.096, \alpha_2 = 0.112$ ] et GARCH(1,1) [ $\omega = 9 * 10^{-6}, \alpha_1 = 0.09, \beta_1 = 0.84$ ] (taille  $n = 2385$ ).

type	$\rho(1)$	$\rho(2)$	$\rho(3)$
<b>Rendements</b>			
CAC 40	0.03016438	0.005338	-0.0317397
modèle (4.1)	0.00579572	-0.01136637	0.00154891
modèle ARCH(2)	-0.01171023	-0.02221684	0.0136662
modèle GARCH(1,1)	0.01548441	-0.01064883	-0.03003595
<b>Carrés des Rendements</b>			
CAC 40	0.12935535	0.12715937	0.11678609
modèle (4.1)	0.13131777	0.06273975	0.001722466
modèle ARCH(2)	0.1616179	0.11869827	0.02302493
modèle GARCH(1,1)	0.14204794	0.02162795	0.06212848

## 4.5 Résultats Empiriques : l'indice S&P 500

L'indice S&P 500 (Standard & Poor's 500) est constitué à partir du prix de 500 titres de la bourse de New York. Ces titres représentent environ 75 % de la capitalisation totale de la bourse de New York. Considéré comme l'indice le plus représentatif du marché américain, il est la principale référence pour mesurer la performance des gestionnaires sur ce marché. De nombreuses études empiriques ont estimé les modèles GARCH et leurs extensions, sur cette série (voir Noh, Engle et Kane (1994), Rockinger et Jondeau (2001)).

Nous donnons la trajectoire des données (logarithmes des rendements) de cet indice dans la Figure 4.5. Nous avons supprimé les valeurs nulles correspondant à des jours de fermeture de la bourse.

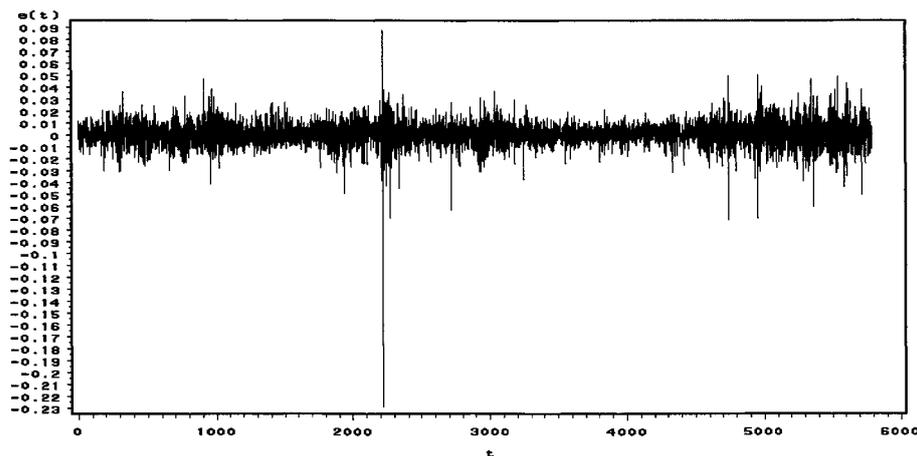


FIG. 4.5 – La série des rendements en logarithme de l'indice S&P 500

### 4.5.1 Test de changements de régime

Nous appliquons le test de changement de régime sur la série des rendements en logarithme du S&P 500 dans le cadre du modèle (4.1) avec  $Z_t = \frac{\varepsilon_t^2 - 1}{\varepsilon_{t-2}^2}$ . Nous choisissons  $m_0 = 500$ . En utilisant (4.5), nous avons calculé  $\hat{F} = 5.33$ . Pour un niveau de risque  $\alpha = 0.01$  on a :  $F_{1-\alpha}(2, n - m_0 - 2) = F_{0.99}(2, 5284) = 4.61$ . Donc, d'après la règle de décision,  $H_0$  est rejetée au profit de  $H_a$ . Ce qui nous permet de conclure à l'existence de changements de régime de type (4.1) avec un niveau de risque  $\alpha = 0.01$ .

### 4.5.2 Estimation des paramètres

Dans la série des données S&P 500, il existe deux valeurs  $\{-0.2289, 8.71 * 10^{-2}\}$  qui sont très grandes par rapport aux autres en valeur absolue (voir la Figure 4.5). Elles correspondent aux dates  $t_1 = 2215$  et  $t_2 = 2217$ . En classant la série par ordre croissant de  $Z_t$ , nous remarquons que le carré de la première valeur  $(-0.2289)^2$  prend la position  $t'_1 = 4109$ . Cette valeur doit changer fortement les valeurs estimées de  $\hat{\alpha}^{(m)}$  et  $\hat{\omega}^{(m)}$ . A cause de cette valeur, les deux premiers graphiques de la Figure 4.6 montrent rupture à partir de  $m = 4150$ .

Pour localiser  $k$ , nous avons balayé la zone  $m \in [3576, 4900]$  qui correspond à  $Z_{(i)} \in [2.4, 21.2]$  en estimant les paramètres  $\omega$ ,  $\alpha_1$  et  $\alpha_2$  par la méthode du maximum de vraisemblance. le Tableau 4.13 donne les estimations des paramètres, ainsi que leurs caractéristiques, pour  $\hat{k} = 5$  correspondant à la plus grande vraisemblance (voir aussi le troisième graphique de la Figure 4.6).

Afin d'améliorer nos résultats, nous reprenons la procédure de localisation de  $k$  et d'estimation des autres paramètres en supprimant les valeurs aberrantes de la série

de rendements de S&P 500. Les deux premiers graphiques de la nouvelle Figure 4.7 montrent plus de régularité. Nous proposons de balayer la zone  $m \in [1946, 4672]$  correspondante à  $Z_{(i)} \in [0.3, 13]$ . La plus grande vraisemblance est obtenue sous la valeur de  $\hat{k} = 1.4$  (voir le troisième graphique de la Figure 4.7). Nous donnons les résultats, de l'estimation du modèle (4.1) sous  $k = 1.4$ , dans le Tableau 4.14.

Une étude empirique du modèle (4.1), similaire au cas de la série du CAC 40, est donnée dans les Tableaux 4.15 et 4.16. Les modèles ARCH(1) et GARCH(1,1) sont estimés à partir de la série des rendements du S&P 500 sans les valeurs aberrantes. D'après les résultats, nous remarquons que ces valeurs influencent l'estimation du modèle (4.1). D'après le Tableau 4.15, nous avons obtenu de meilleures valeurs avec la simulation (II) du modèle (4.1). Les autocorrélations empiriques des carrés de la série simulée (II) du modèle (4.1) sont plus proches de celles de la série réelle que de celles du modèle ARCH(2).

TAB. 4.13 – Rendements du S&P 500 : estimation par MV des paramètres du modèle (4.1)

Paramètres	Estimation	Ecart-type [RATS]	T-stat	Signif
$k = 5$				
$\omega$	0.78	0.009	82.78	0.00
$\alpha_1$	0.08	0.012	7.21	0.00
$\alpha_2$	0.38	0.014	25.93	0.00

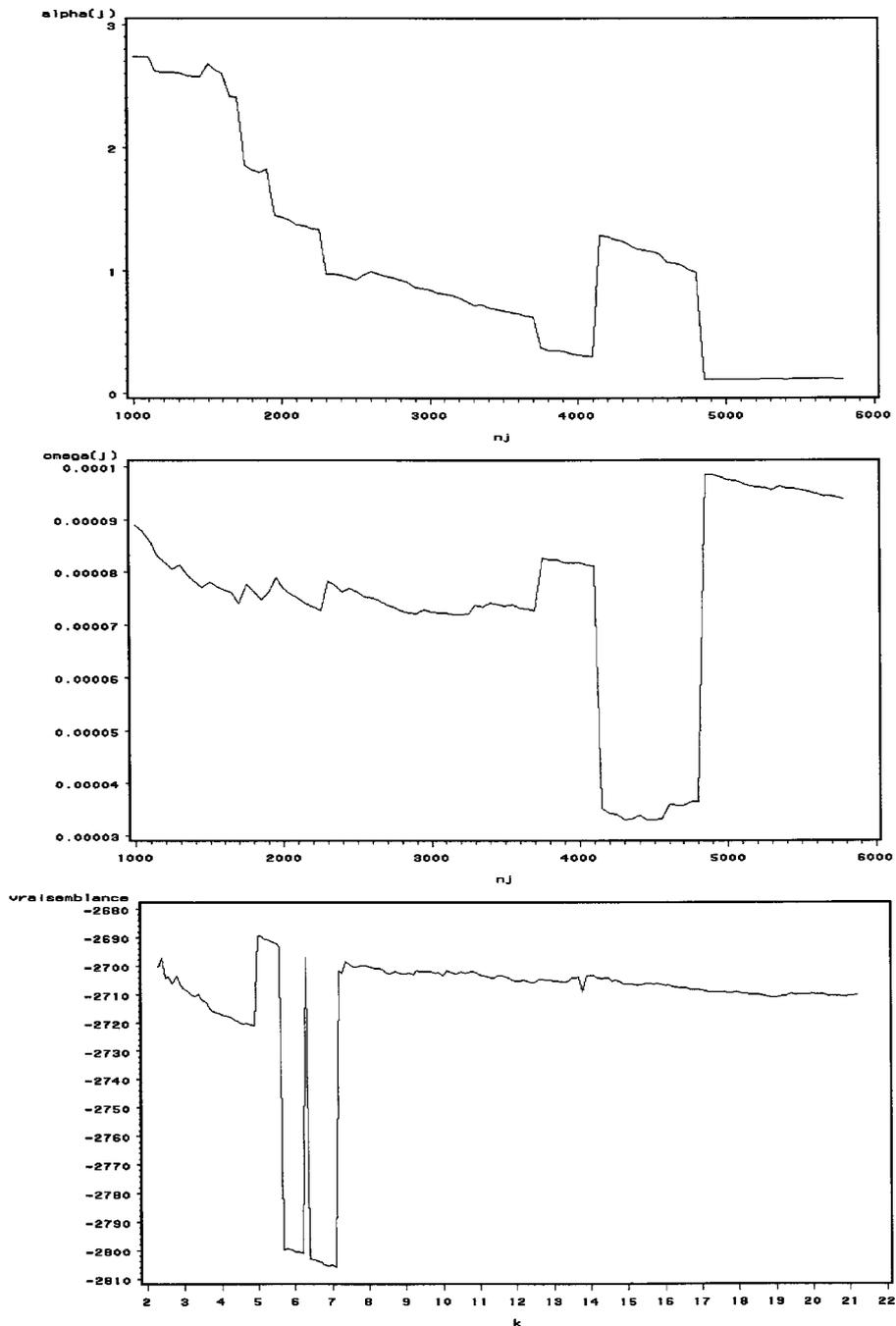


FIG. 4.6 – Rendements du Standard & Poor's 500 : localisation graphique de  $k$ . Graphique du haut : en abscisses : taille  $m$  de l'échantillon ; en ordonnées : valeurs de  $\hat{\omega}^{(m)}$ . Graphique du milieu : en abscisses : taille  $m$  de l'échantillon ; en ordonnées : valeurs de  $\hat{\alpha}^{(m)}$ . Graphique du bas : vraisemblance en fonction de  $k$ .

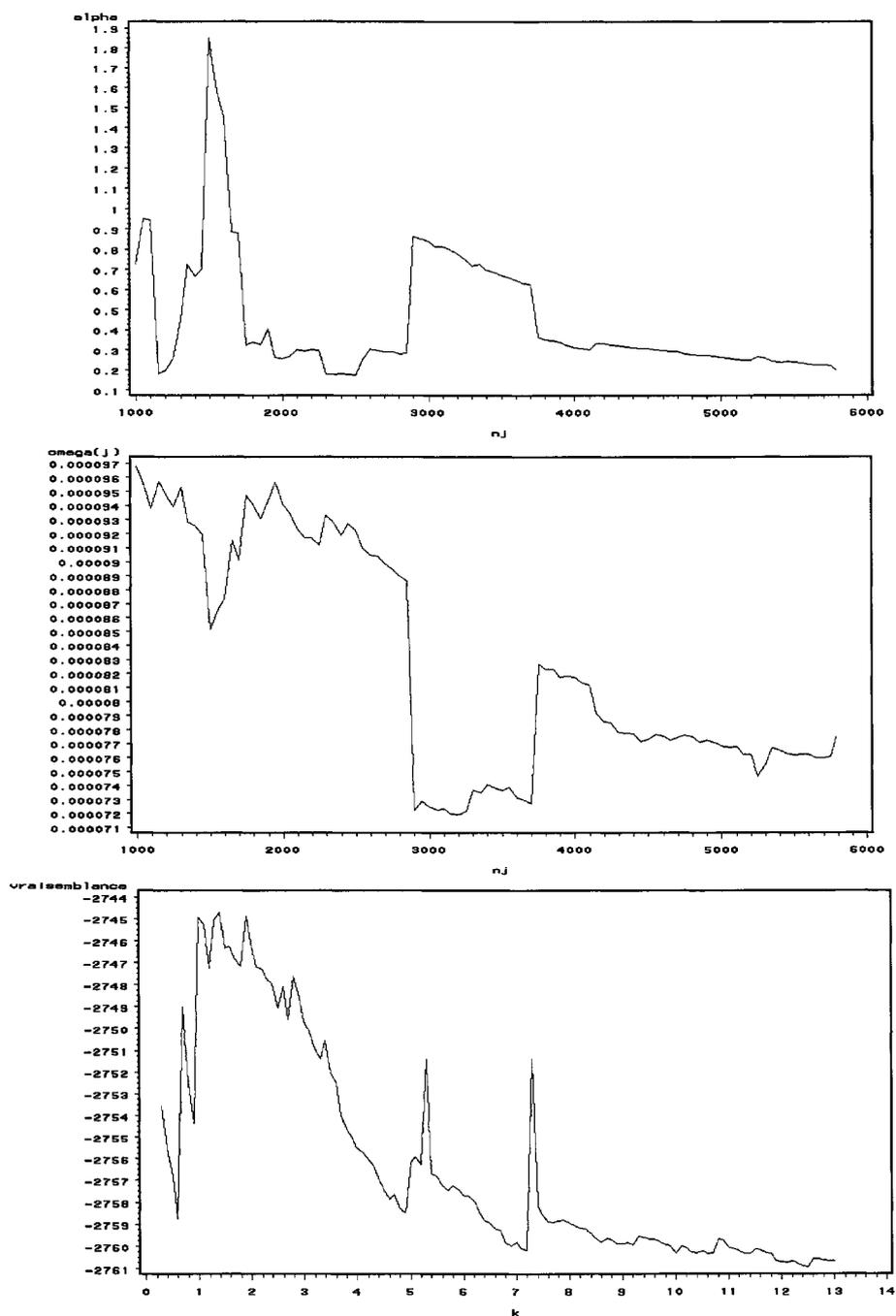


FIG. 4.7 – Rendements du Standard & Poor's 500 après la suppression des valeurs aberrantes : localisation graphique de  $k$ . Graphique du haut : en abscisses : taille  $m$  de l'échantillon ; en ordonnées : valeurs de  $\hat{\omega}^{(m)}$ . Graphique du milieu : en abscisses : taille  $m$  de l'échantillon ; en ordonnées : valeurs de  $\hat{\alpha}^{(m)}$ . Graphique du bas : vraisemblance en fonction de  $k$ .

TAB. 4.14 – Estimation par MV des paramètres du modèle (4.1) après la suppression des valeurs aberrantes des rendements du S&P 500

Paramètres	Estimation	Ecart-type [RATS]	T-stat	Signif
$k = 1.4$				
$\omega$	$7.78 * 10^{-5}$	$9.824 * 10^{-7}$	79.24	0.00
$\alpha_1$	0.568	0.039	14.36	0.00
$\alpha_2$	0.094	0.010	9.20	0.00

TAB. 4.15 – Caractéristiques empiriques de la série des rendements du S&P 500 et de celles des simulées des modèles estimés (4.1), ARCH(2) et GARCH(1,1) (taille  $n = 5788$ ).

Moyenne	E-type	Min	Q1	Méd	Q3	Max
<b>Série du S&amp;P 500</b>						
0.0004	0.0103	-0.2290	-0.0046	0.0005	0.0056	0.0871
Simulé (I) du modèle (4.1) avec [ $k = 5, \omega = 0.78167, \alpha_1 = 0.08548, \alpha_2 = 0.37695$ ]						
0.0032	1.009	-5.9216	-0.6500	0.0155	0.6575	4.4302
Simulé (II) du modèle (4.1) avec [ $k = 1.4, \omega = 7.78 * 10^{-5}, \alpha_1 = 0.56800, \alpha_2 = 0.09357$ ]						
$6.11 * 10^{-5}$	0.0099	-0.0370	-0.0064	$5.66 * 10^{-5}$	0.0065	0.0714
Simulé d'un ARCH(2) avec [ $\omega = 6.6 * 10^{-5}, \alpha_1 = 0.164237, \alpha_2 = 0.173759$ ]						
$1.5 * 10^{-5}$	0.0103	-0.0475	-0.0067	$3.2 * 10^{-5}$	0.0066	0.0463
Simulé d'un GARCH(1,1) avec [ $\omega = 2 * 10^{-6}, \alpha_1 = 0.08256, \beta_1 = 0.89978$ ]						
-0.0002	0.0108	-0.0500	-0.0068	-0.0001	0.0065	0.0658
<b>Carrés des valeurs de la série du S&amp;P 500</b>						
0.0001	0.0007	$8 * 10^{-11}$	$5.1 * 10^{-6}$	$2.6 * 10^{-5}$	$9 * 10^{-5}$	0.0524
Carrés des valeurs simulées (I) du modèle (4.1)						
1.0175	1.6341	$1.7 * 10^{-8}$	0.0960	0.4270	1.2678	35.0658
Carrés des valeurs simulées (II) du modèle (4.1)						
$9.7 * 10^{-5}$	0.0002	$9.76 * 10^{-15}$	$9.77 * 10^{-6}$	$4.19 * 10^{-5}$	0.0001	0.0051
Carrés des valeurs simulées d'un modèle ARCH(2)						
0.0001	0.0002	$2 * 10^{-12}$	$10^{-5}$	$4.4 * 10^{-5}$	0.0001	0.0023
Carrés des valeurs simulées d'un modèle GARCH(1,1)						
0.0001	0.0002	$3 * 10^{-16}$	$10^{-5}$	$4.5 * 10^{-5}$	0.0001	0.0043

TAB. 4.16 – Autocorrélations empiriques d'ordre 1, 2 et 3 de la série des rendements du S&P 500 et de celles des simulées des modèles estimés : (4.1)-(I) [ $k = 5, \omega = 0.78167, \alpha_1 = 0.08548, \alpha_2 = 0.37695$ ], (4.1)-(II) [ $k = 1.4, \omega = 7.78 * 10^{-5}, \alpha_1 = 0.56800, \alpha_2 = 0.09357$ ], ARCH(2) [ $\omega = 0.000066, \alpha_1 = 0.164237, \alpha_2 = 0.173759$ ] et GARCH(1,1) [ $\omega = 0.000002, \alpha_1 = 0.08256, \beta_1 = 0.89978$ ] ( $n = 5788$ ).

type	$\rho(1)$	$\rho(2)$	$\rho(3)$
<b>Valeurs de rendement du S&amp;P 500 (<math>\varepsilon_t</math>)</b>			
S & P 500	0.03770181	-0.03765795	-0.03514824
modèle (4.1)-(I)	0.03110298	0.00458911	0.02128153
modèle (4.1)-(II)	-0.02697171	-0.01010968	-0.00196357
modèle ARCH(2)	-0.01524802	0.00365087	-0.02230645
modèle GARCH(1,1)	0.02651659	0.00529818	-0.03642704
<b>Carrés des valeurs de rendement du S&amp;P 500 (<math>\varepsilon_t^2</math>)</b>			
S & P 500	0.11227068	0.14841353	0.07608516
modèle (4.1)-(I)	0.22775493	-0.00319569	-0.00427009
modèle (4.1)-(II)	0.151192	0.12090335	0.04730214
modèle ARCH(2)	0.18677668	0.22835592	0.05280784
modèle GARCH(1,1)	0.12565087	0.14601436	0.11467397

## 4.6 Annexe

TAB. 4.17 – Effectif en fonction de  $k$  des observations telles que  $\varepsilon_t^2 > k\varepsilon_{t-1}^2$  pour la série des rendements en logarithme de l'indice CAC 40 sur la période 1988-1998. (2385 observations)

$k$	Effectif (%)
0.1	78.94
0.5	59.86
1	50.92
2	40.10
3	33.77
4	30.41
5	27.56
10	21.22
15	18.25
20	15.48
30	13.17
40	11.58
50	10.11
100	7.47
500	2.89
1000	1.93
10000	0.59

TAB. 4.18 – Effectif en fonction de  $k$  des observations telles que  $\varepsilon_t^2 > k\varepsilon_{t-1}^2$  pour la série des rendements en logarithme de l'indice S&P500. (5804 observations)

$k$	Effectif (%)
0,1	78.46
0,5	59.74
1	49.81
2	40.88
3	35.25
4	31.66
5	29.08
10	21.70
15	18.18
20	15.94
30	13.05
40	11.34
50	10.10
100	7.15
500	3.24
1000	2.38
10000	0.93



# Bibliographie

- [1] **Chan, K. S., et Tong, H. (1990)**, *On Likelihood Ratio Tests for Threshold Autoregression*, Journal of the Royal Statistical Society, Ser. B. 52, 469-476.
- [2] **Hansen, B. E. (1996)**, *Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis*, Econometrica, 64, 413-430.
- [3] **Mak, T.K., Wong H. et Li W.K. (1997)**, *Estimation of Nonlinear Time Series With Conditional Heteroscedastic Variances By Iteratively Weighted Least Squares*, Computational Statistics & Data Analysis, Vol. 24, Issue 2, 169-178.
- [4] **Noh J., R.F. Engle et A. Kane (1994)**, *Forecasting Volatility and Option Prices of the S&P 500 Index*, Journal of Derivatives, Fall 1994, 17-30.
- [5] **Rockinger M. et E. Jondeau (2001)**, *Entropy Densities With an Application Conditional Skewness and Kurtosis*, Journal of Econometrics, Vol. 106/1, 119-142.
- [6] **Tsay R.S (1989)**, *Testing and Modeling Threshold Autoregressive Processes*, Journal of American Statistical Association, 84, no. 405, 231-240.
- [7] **Tsay R.S (1998)**, *Testing and Modeling Multivariate Threshold Models*, Journal of American Statistical Association, 93, no. 443, 1188-1202.

