RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITÉ A.M.O DE BOUIRA

FACULTÉ DES SCIENCES ET DES SCIENCES APPLIQUÉES

DÉPARTEMENT DE MATHÉMATIQUES

MÉMOIRE POUR L'OBTENTION DU DIPLOME
DE MASTER EN MATHÉMATIQUES

Spécialité :

Recherche Opérationnelle

THEME

Études probabilistes et statistiques des modèles ARCH et GARCH

Présenté par : OUKIL Nassima

Devant le jury composé de :

Président :	HAMID Karim	MAA	U.A/M/O Bouira
Encadreur :	DEMMOUCHE Nacer	MCB	U.A/M/O Bouira
Examinatreur :	BOUDREF Mohamed Ahmed	MCA	U.A/M/O Bouira
Examinatrice :	BOUDANE Khadidja	MAA	U.A/M/O Bouira

Année universitaire :2019/2020

Etudes probabilistes et statistiques des modèles ARCH et GARCH

Présenté par : OUKIL Nassima

Sous la direction de : DEMMOUCHE Nacer
Département de Mathématiques.
Faculté des Sciences et Sciences appliquées.
Université Akli Mohand Oulhadj de Bouira.

2020

Table des matières

Remerciements

Je remercie Dieu le tout Puissant de m'avoir donné la santé et la volonté d'entamer et de terminer ce mémoire.

Tout d'abord, ce travail ne serait pas aussi riche et n'aurait pas pu avoir le jour sans l'aide et l'encadrement de Mr Demmouche Nacer, je le remercie pour la qualité de son encadrement exceptionnel, pour sa patience, sa rigueur et sa disponibilité durant ma préparation de ce mémoire.

J'adresse aussi mes vifs remerciements aux membres de jurys pour avoir bien voulu examiner et juger ce travail.

Je remercie également tous mes enseignants de m'avoir inciter à travailler en mettant à notre disposition leurs expériences et leurs compétences

Je remercie infiniment ma famille et spécialement mes grand parents, qui m'ont encouragé et aidé à arriver à ce stade de ma formation.

Introduction générale

Une série temporelle, ou série chronologique, est une suite de valeurs numériques représentant l'évolution d'une quantité spécifique au cours du temps. De telles suites de variables aléatoires peuvent être exprimées mathématiquement afin d'en analyser le comportement, généralement pour comprendre son évolution passée et pour en prévoir le comportement futur. Une telle transposition mathématique utilise le plus souvent des concepts de probabilités et de statistique.

L'objet des séries temporelles est l'étude des variables au cours du temps. Même s'ils n'ont pas été à l'origine de cette discipline, ce sont les économètres qui ont assuré les grandes avancées qu'a connues cette discipline . Parmi ses principaux objectifs figurent la détermination de tendances au sein de ces séries ainsi que la stabilité des valeurs (et de leur variation) au cours du temps.

C'est de la déception des prévisions issues des modèles structurels d'inspiration keynésienne qu'est née la théorie des séries temporelles telle qu'on la connait aujourd'hui. Et sur ce point, c'est la publication de l'ouvrage de **Box et Jenkins** en 1970 qui a été décisive. En effet, dans l'ouvrage les deux auteurs développent le très populaire modèle $ARMA$ (Autoregressive Moving Average). Le modèle $ARMA$ est un cas particulier d'un modèle beaucoup plus général nommé $ARIMA$ (Autoregressive Integrated Moving Average). En effet, le modèle $ARMA$ ne permet de traiter que les séries dites stationnaires (des moments du premier ordre qui sont invariants au cours du temps). Les modèles $ARIMA$ permettent de traiter les séries non stationnaires après avoir déterminé le niveau d'intégration (le nombre de fois qu'il faut différencier la série avant de la rendre stationnaire).

Bien que possédant d'excellentes qualités prévisionnelles, le modèle ARIMA ou ARMA souffre d'une lacune majeure : il est incapable de traiter simultanément plus d'une variable (série). Pour contourner ce problème, il faut pouvoir généraliser le modèle $ARIMA$ dans le cas à plusieurs variables. C'est ce qu'a fait en partie **Christopher Sims** en proposant en 1980 le modèle VAR (Vector Autoregressive) qui permet de traiter concomitamment plusieurs variables. Mais, contrairement au modèle structurel à plusieurs variables, dans les modèles VAR, toutes les variables sont endogènes. Cette manière de modéliser en faisant abstraction d'une théorie économique a donné naissance à ce que l'on a appelé l'Économétrie sans théorie.

Ces modèles ($ARIMA$ et VAR) ne permettent de traiter que des phénomènes qui sont linéaires ou approximativement mais ne permettent pas de "capturer" les propriétés des phénomènes qui sont non linéaires (les variables financières par exemple, inflation, cours d'action etc.). Alors pour une modélisation plus réaliste de ces séries les modèles $ARCH$ (Autoregressive Conditionally Heteroscedastic) et les modèles $GARCH$ ($ARCH$ *Generalized*) sont les plus utiles.

Dans le premier chapitre, on donne quelques rappels théoriques tels que l'indépendance de variables aléatoires, les théorèmes limites pour les suites de variables aléatoires indépendantes identiquement distribuées, la stationnarité stricte et stationnarité au second ordre des processus aléatoires, l'ergodicité et le théoroème ergodique.

Dans le deuxième chapitre, on étudie deux types de modèles de séries chronologiques, à savoir le modèle $ARCH$ et le modèle $ARCH$ généralisé i.e. , $GARCH$. L'étude de ces deux modèles consiste à trouver des conditions suffisantes sur les coefficients qui assurent l'existence des solutions strictement stationnaires et ergodiques ayant des moments d'ordres supérieurs.

Dans le troisième chapitre, on istime les paramètres des deux modèles par la méthode de quasi-maximum de vraisemblance (QMV) et on étudie les propriétés asymptotiques de l'estimateur de QMV.

Chapitre 1

Outils préliminaires

1.1 Variables aléatoires

L'orsqu'on envisage d'observer une expérience aléatoire et lui associer un espace de probabilité $(\Omega, \mathcal{F}, P)$, il arrive que l'on s'intéresse plutôt à une fonction numérique du résultat attendu qu'au résultat lui-même, mais l'objectif reste de pouvoir assigner des probabilités à des événements concernant cette fonction du résultat. La fonction étudiée doit donc satisfaire certaines conditions de mesurabilité : Si $X(.)$ est une telle fonction définie sur Ω à valeurs réelles, des événements concernant $X(.)$, comme $\{\omega \in \Omega : X(\omega) = a\}, \{\omega \in \Omega : X(\omega) > a\}$, $a \in \mathbb{R}$, doivent être dans $\mathcal{F}$ pour qu'on puisse leur assigner des probabilités parce que P n'est pas définie pour des événements n'appartenant pas à $\mathcal{F}$. Cependant la quasi-totalité des événements concernant $X(.)$ que l'on puisse formuler peuvent s'exprimer par des opérations ensemblistes élémentaires des événements $\{\omega \in \Omega : X(\omega) \leq a\}, a \in \mathbb{R}$ et donc il suffit d'exiger que ceux-ci soient membres de $\mathcal{F}$ pour tout $a \in \mathbb{R}$. Une fonction vérifiant une telle condition est dite définie sur $(\Omega, \mathcal{F}, P)$.

Définition 1 *Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Une fonction $X(.)$ définie sur Ω à valeurs dans $\mathbb{R}$ est dite variable aléatoire sur $(\Omega, \mathcal{F}, P)$ si $\{\omega \in \Omega : X(\omega) \leq a\} \in \mathcal{F}$ pour tout $a \in \mathbb{R}$.*

Définition 2 ***(Tribu engendrée par une variable aléatoire)*** *La tribu $\mathcal{F}(X)$ engendrée par une variable aléatoire X définie sur un espace de probabilité $(\Omega, \mathcal{F}, P)$ est la plus petite tribu contenant les événements de la forme $\{\omega \in \Omega : X(\omega) \leq a\}, a \in \mathbb{R}$.*

Définition 3 ***(Distribution de probabilité)*** *Soit X une variable aléatoire réelle définie sur $(\Omega, \mathcal{F}, P)$. On appelle distribution de probabilité de X la fonction ensembliste $P_X(.)$ définie sur $\mathcal{B}(\mathbb{R})$ à valeurs dans $[0,1]$ par*

$$\begin{aligned} P_X(.) \quad &: \quad \mathcal{B}(\mathbb{R}) \to [0,1] \\ B \quad &\mapsto \quad P_X(x) = P\{\omega \in \Omega : X(\omega) \in B\} = P(X^{-1}(B)) \end{aligned}$$

Définition 4 ***(Fonction de répartition)*** *Soit X une variable aléatoire réelle définie sur $(\Omega, \mathcal{F}, P)$. On appelle fonction de répartition de X la fonction $F_X(.)$ définie sur $\mathbb{R}$ à valeurs dans $[0,1]$ par*

$$\begin{aligned} F_X(.) \quad &: \quad \mathbb{R} \to [0,1] \\ x \quad &\mapsto \quad F_X(x) = P\{\omega \in \Omega : X(\omega) \leq x\} = P_X(]-\infty, x]) = P(X^{-1}(]-\infty, x])). \end{aligned}$$

La fonction de répartition $F_X(.)$ est aussi dite distribution de probabilité de X.

Théorème 5 ***(Propriétés de la fonction de répartition)*** *Soit $F_X(.)$ la fonction de répartition d'une variable aléatoire X, alors*

1. *F_X est une fonction croissante.*
2. *F_X est continue à droite.*
3. $\lim_{x\to-\infty} F_X(x) = 0$.
4. $lim_{x\to\infty} F_X(x) = 1$.

Définition 6 ***(Variable aléatoire discrète)*** *Une variable aléatoire X définie sur $(\Omega, \mathcal{F}, P)$ à valeurs dans $\mathbb{R}$ est dite discrète si l'ensemble de valeurs de X est fini ou infini dénombrable.*

Définition 7 ***(Fonction de masse d'une variable aléatoire discrète)*** *Pour une variable aléatoire discrète X définie sur $(\Omega, \mathcal{F}, P)$ à valeurs dans $I = X(\Omega)$, on définit la fonction de masse $p_X(.)$ par*

$$\begin{aligned} p_X(.) \quad &: \quad I \to [0,1] \\ x \quad &\mapsto \quad p_X(x) = P\{\omega : X(\omega) = x\} = P(X^{-1}(\{x\})). \end{aligned}$$

Définition 8 ***(Variable aléatoire continue)*** *Une variable aléatoire X définie sur $(\Omega, \mathcal{F}, P)$ à valeurs dans $\mathbb{R}$ est dite continue si et seulement si sa fonction de répartition F_X est continue sur $\mathbb{R}$. De manière équivalente, X est continue si et seulement si $P(X = x) = 0$ pour tout $x \in \mathbb{R}$.*

Définition 9 ***(Variable aléatoire absolument continue)*** *Une variable aléatoire X définie sur $(\Omega, \mathcal{F}, P)$ à valeurs dans $\mathbb{R}$ est dite absolument continue si sa fonction de répartition F_X est absolument continue au sens de définition suivante :*

Définition 10 ***(Continuité absolue d'une fonction)*** *Une fonction réelle f, définie sur un intervalle I, est dite absolument continue sur I si pour tout $\epsilon > 0$, il existe $\delta > 0$ tel que pour toute suite finie d'intervalles ouverts de I 2-à-2 disjoints $]x_1, y_1[, ...,]x_n, y_n[$, $n \in \mathbb{N}^*$ vérifiant $\sum_{i=1}^{n} |y_i - x_i| \leq \delta$, on a $\sum_{i=1}^{n} |f(y_i) - f(x_i)| < \varepsilon$.*

Théorème 11 ***(Densité de probabilité)*** *Si X est absolument continue, alors il existe une fonction positive Borel mesurable $f_X(.)$ définie sur $\mathbb{R}$, dite densité de probabilité de X, telle que $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$. La fonction $f_X(.)$ est aussi dite dérivée de Radon-Nikodym de la mesure P_X par rapport à la mesure de Lebesgue.*

1.1.1 Caractéristiques d'une variable aléatoire

Dans cette partie on introduit les caractéristiques des variables aléatoires à savoir l'espérance mathématique, la variance, la covariance et le coefficient de corrélation. Soit X une variable aléatoire définie sur $(\Omega, \mathcal{F}, P)$ à valeurs dans $\mathbb{R}$.

Définition 12 *L'espérance mathématique de X, notée $E(X)$, est l'intégrale de Lebesgue de X sur Ω par rapport à la mesure $P(.)$. Autrement dit*

$$E(X) = \int_{\Omega} X(\omega) dP(\omega).$$

Ceci est bien défini si $E|X| < \infty$.

Remarque 13 *L'espérance $E(X)$ peut être également définie en terme de :*

1. *La distribution de probabilité $P_X(.)$ de X, $E(X) = \int_{\mathbb{R}} x P_X(dx)$.*
2. *La fonction de répartition F_X de X, $E(X) = \int_{\mathbb{R}} x dF_X(x)$.*
3. *La fonction de masse $p_X(.)$ si X est discrète, $E(X) = \sum_{x \in X(\Omega)} x p_X(x)$.*
4. *La densité de probabilité $f_X(.)$ de X si X est absolument continue, $E(X) = \int_{\mathbb{R}} x f_X(x) dx$.*

L'espérance mathématique d'une variable aléatoire peut être finie, infini ou peut ne pas exister. Les moments d'ordre supérieurs d'une variable aléatoires sont données comme suit.

Définition 14 ***(Moments)*** *Soit X une variable aléatoire et $p \in \mathbb{N}^*$ tels que $E(|X|^p) < \infty$.*

1. *Le moment d'ordre p de X est $E(X^p) < \infty$.*
2. *Le moment factoriel d'ordre p de X est $E(X(X-1)...(X-p+1))$.*
3. *Le moment centré d'ordre p de X est le moment d'ordre p de $X - E(X)$.*

Les inégalités pour les espérances qu'on développe ici sont régulièrement utilisées à la fois en théorie des probabilités et en analyse.

Théorème 15 ***(Inégalité de Markov)*** *Soit X une variable aléatoire sur $(\Omega, \mathcal{F}, P)$ et $\phi : \mathbb{R} \to \mathbb{R}^+$ une fonction Borel mesurable, alors pour tout $\epsilon > 0$*

$$P(\phi(X) \geq \epsilon) \leq \frac{E(\phi(X))}{\epsilon}.$$

Théorème 16 ***(Inégalité de Chebyshev)*** *Si X est une variable aléatoire sur $(\Omega, \mathcal{F}, P)$ telle que $E(X^2) < \infty$, alors pour tout $\epsilon > 0$*

$$P(|X - E(X)| \geq \epsilon) \leq \frac{E(X - E(X))^2}{\epsilon^2}.$$

Théorème 17 ***(Inégalité de Cauchy-Buniakowski-Schwarz)*** *Soit X et Y deux variables aléatoires sur $(\Omega, \mathcal{F}, P)$ telles que $E(X^2) < \infty$ et $E(Y^2) < \infty$, alors $E|XY| < \infty$ et*

$$(E|XY|)^2 \leq E(X^2).E(Y^2).$$

Théorème 18 ***(Inégalité Jensen)*** *Si $\phi : \mathbb{R} \to \mathbb{R}$ est une fonction convexe et X une variable aléatoire sur $(\Omega, \mathcal{F}, P)$ telle que $E(X)$ et $E[\phi(X)]$ existent, alors*

$$\phi[E(X)] \leq E[\phi(X)].$$

Théorème 19 ***(Inégalité de Holder)*** *Soit $1 < p < \infty$ et $1 < q < \infty$ tels que $\frac{1}{p} + \frac{1}{q} = 1$. Si $E|X|^p < \infty$ et $E|Y|^q < \infty$, alors $E|XY| < \infty$ et*

$$E|XY| \leq [E|X|^p]^{\frac{1}{p}}[E|Y|^q]^{\frac{1}{q}}.$$

Théorème 20 ***(Inégalité de Minkowski)*** *Si $E|X|^p < \infty$ et $E|Y|^q < \infty, 1 \leq p < \infty$, alors $E|X+Y|^p < \infty$ et*

$$(E|X+Y|^p)^{\frac{1}{p}} \leq (E|X|^p)^{\frac{1}{p}} + (E||Y|^p)^{\frac{1}{p}}.$$

Définition 21 ***(Variance, covariance)*** *Soit X et Y deux variables aléatoires sur $(\Omega, \mathcal{F}, P)$ telle que $E(X^2) < \infty$ et $E(Y^2) < \infty$.*

1) Le moment centré d'ordre 2 de X est dit variance de X, noté par $Var(X)$,

$$Var(X) = \sigma_X^2 = E(X - E(X))^2 = E(X^2) - (E(X))^2 \geq 0,$$

qui existe toujours parce que $\sigma_X^2 \leq E(X^2) < \infty$.

2) La covariance de X et Y, notée $cov(X, Y)$, est définie par

$$cov(X, Y) = E(X - E(X))(Y - E(Y)) = E(XY) - E(X)E(Y),$$

qui existe toujours puisque par l'inégalité de CBS

$$|cov(X, Y)|^2 \leq \sigma_X^2.\sigma_Y^2 < \infty.$$

3) Si $Var(X) > 0$ et $Var(Y) > 0$, la covariance normalisée est dite coefficient de corrélation entre X et Y, défini par

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X.\sigma_Y},$$

où par l'inégalité de CBS on a $|\rho(X,Y)| \leq 1$.

1.2 Variables aléatoires indépendantes

Définition 22 *Deux variables aléatoires X et Y définies sur $(\Omega, \mathcal{F}, P)$ à valeurs dans $\mathbb{R}$ sont dites indépendantes, si pour tous $B_1, B_2 \in B(\mathbb{R})$,*

$$P(X \in B_1, Y \in B_2) = P(X \in B_1)P(Y \in B_2).$$

L'indépendance d'une suite finie de variables aléatoires est donnée comme suit.

Définition 23 *Les variables aléatoires $X_1, X_2, ..., X_n$ définies sur $(\Omega, \mathcal{F}, P)$ à valeurs dans $\mathbb{R}$ sont dites indépendantes si pour tous boréliens $B_1, B_2, ..., B_n \in B(\mathbb{R})$,*

$$P(X_1 \in B_1, X_2 \in B_2, ..., X_n \in B_n) = \prod_{i=1}^{n} P(X_i \in B_i).$$

L'indépendance de variables aléatoires est préservée sous des transformations mesurables.

Théorème 24 *Si $X_1, ..., X_n$ sont des variables aléatoires indépendantes, alors pour toutes fonctions Borel mesurables $f_1, ..., f_n$ de $\mathbb{R}$ dans $\mathbb{R}$, les variables aléatoires $f_1(X_1), ..., f_n(X_n)$ sont également indépendantes.*

L'indépendance d'une suite infinie de variables aléatoires est donnée comme suit.

Définition 25 *Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires définies sur $(\Omega, \mathcal{F}, P)$ à valeurs dans $\mathbb{R}$. On dit que $(X_n)_{n\in\mathbb{N}^*}$ est une suite de variables aléatoires indépendantes, si pour tout $n \geq 2$, les variables aléatoires $X_1, X_2, ..., X_n$ sont indépendantes.*

1.3 Convergence stochastique et théorèmes limites

Rappelons qu'une suite de variables aléatoires $(X_n)_{n\in\mathbb{N}^*}$ (notée aussi $X = (X_1, X_2, ...)$) définies sur $(\Omega, \mathcal{F}, P)$ peut être définie comme une application mesurable de Ω dans $\mathbb{R}^\infty$ qui pour tout issu ω associe une suite de nombres réels $X(\omega) = (X_1(\omega), X_2(\omega), ...)$. Comme pour le cas de suites réelles, il est donc très important de connaitre le comportement des suites $X(\omega) = (X_1(\omega), X_2(\omega), ...)$ lorsque n croit indéfiniment. Cependant la question n'est pas aussi simple que pour les suites réelles, puisque dans le cas de suite de variables aléatoires il existe plusieurs façons (ou modes) selon lesquelles une suite de variables aléatoires peut s'approcher ou tendre vers une limite X. Un premier critère est celui de convergence d'une suite $X(\omega) = (X_1(\omega), X_2(\omega), ...)$ pour tout $\omega \in \Omega$.

Définition 26 ***(Convergence en une issue)*** *Une suite de variables aléatoires* $(X_n)_{n\in\mathbb{N}^*}$ *définies sur* $(\Omega, \mathcal{F}, P)$ *est dite avoir une limite (ou convergente) en un point* $\omega \in \Omega$ *si la suite de nombres réels* $(X_n(\omega))_{n\in\mathbb{N}^*}$ *converge vers une limite* $X(\omega)$.

Cette définition n'est pas utile en soi puisqu'on ne s'intéresse pas au comportement d'une suite en un résultat spécifique $\omega \in \Omega$ de l'expérience. Cependant, elle peut servir pour d'autres concepts de limite plus importants. Il est utile de noter qu'une suite de variable aléatoire étant une suite de fonctions définies sur Ω, on peut en définir le concept de convergence simple.

Définition 27 ***(Convergence simple)*** *Une suite de variables aléatoires* $(X_n)_{n\in\mathbb{N}^*}$ *définies sur* $(\Omega, \mathcal{F}, P)$ *converge simplement vers la variable aléatoire* X *définie sur* $(\Omega, \mathcal{F}, P)$ *si pour tout* $\omega \in \Omega$,

$$\lim_{n\to\infty} X_n(\omega) = X(\omega).$$

La définition 29 est trop forte puisqu'elle exige la convergence pour tout $\omega \in \Omega$. En pratique, il existe rarement des suites de variables aléatoires convergentes simplement vers une variable aléatoire. C'est pourquoi, une telle définition est allégée en exigeant la convergence seulement sur un sous-ensemble particulier de Ω.

Définition 28 *Soit une suite de variables aléatoires* $(X_n)_{n\in\mathbb{N}^*}$ *et une variable aléatoire* X *définies sur* $(\Omega, \mathcal{F}, P)$. *L'ensemble* $\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\} \in \mathcal{F}$ *est dit ensemble de convergence de la suite* $(X_n)_{n\in\mathbb{N}^*}$.

On peut parler de la probabilité de convergence d'une suite $(X_n)_{n\in\mathbb{N}^*}$ en faisant référence à la probabilité de son ensemble de convergence. Une définition de convergence moins forte que la convergence simple mais qui reste quand même assez forte est le concept de convergence presque sûre (p.s) (dite aussi convergence avec probabilité un).

Définition 29 ***(Convergence p.s)*** *Une suite de variables aléatoires* $(X_n)_{n\in\mathbb{N}^*}$ *définies sur* $(\Omega, \mathcal{F}, P)$ *converge presque sûrement vers la variable aléatoire* X *définie sur* $(\Omega, \mathcal{F}, P)$ *et on écrit* $X_n \to X$ *p.s., s'il existe un ensemble nul* A *dans* $(\Omega, \mathcal{F}, P)$ *tel que la suite* $(X_n(\omega))_{n\in\mathbb{N}^*}$ *converge vers* $X(\omega)$ *pour tout* $\omega \in \Omega - A$. *Autrement dit, si son ensemble de convergence est de probabilité* 1, *i.e.,*

$$P\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\} = 1.$$

On peut exhiber des exemples dans lesquels pour tout $\omega \in \Omega$, la suite $(X_n(\omega))_{n\in\mathbb{N}^*}$ ne converge pas vers $X(\omega)$, mais $\lim_{n\to\infty} P\{\omega \in \Omega : |X_n(\omega) = X(\omega)| \geq \epsilon\} = 0$ pour tout $\epsilon > 0$.

Définition 30 ***(Convergence en probabilité)*** *Une suite de variables aléatoires* $(X_n)_{n\in\mathbb{N}^*}$ *définies sur* $(\Omega, \mathcal{F}, P)$ *converge en probabilité vers la variable aléatoire* X *définie sur* $(\Omega, \mathcal{F}, P)$ *et on écrit* $X_n \to^P X$, *si pour tout* $\epsilon > 0$

$$\lim_{n\to\infty} P(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \epsilon\}) = 0.$$

On note par $L^1 = L^1(\Omega, \mathcal{F}, P)$ l'espace vectoriel de variables aléatoires intégrables, i.e., l'espace des variables aléatoires X définies sur $(\Omega, \mathcal{F}, P)$ telles que $E|X| < \infty$. De même pour $p \in]0, \infty[$, $L^p = L^p(\Omega, \mathcal{F}, P)$ est la classe de toutes les variables aléatoires X définies sur $(\Omega, \mathcal{F}, P)$ pour lequelles $E|X|^p < \infty$.

Définition 31 *Soit* $(X_n)_{n\in\mathbb{N}^*}$ *une suite de variables aléatoires sur* $(\Omega, \mathcal{F}, P)$ *et* X *une variable sur* $(\Omega, \mathcal{F}, P)$. *Supposons qu'il existe* $p > 0$ *tel que* $E|X_n|^p < \infty$ *pour tout* $n \geq 1$. *On* *dit que la suite* $(X_n)_{n\in\mathbb{N}^*}$ *converge en moyenne d'ordre* p *(ou converge dans* L^p*) vers* X *et on écrit* $X_n \to^p X$ *ou* $X_n \to^{L^p} X$*, si* $\lim_{n\to\infty} E|X_n - X|^p = 0$.

Remarque 32 *Lorsque* $p = 1$*, on parle de la convergence en moyenne et lorsque* $p = 2$*, on parle de convergence en moyenne quadratique qu'on note par* $X_n \to^{m.q} X$.

Un autre mode de convergence très répandu et qui n'implique pas directement les valeurs de X_n, mais plutôt leurs distributions est connu sous le nom de convergence en distribution.

Définition 33 ***(Convergence en distribution)*** *On dit que la suite de variables aléatoires* $(X_n)_{n\in\mathbb{N}^*}$ *définie sur* $(\Omega, \mathcal{F}, P)$ *converge en distribution (ou en loi) vers la variable aléatoire* X *définie sur* $(\Omega, \mathcal{F}, P)$ *et on écrit* $X_n \to^d X$*, si* $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ *pour tout* x *où* F_X *est continue.*

Notons que la propriété :$X_n \to X \Rightarrow E(X_n) \to E(X)$ n'est pas valable sans restriction, mais elle est vraie sous deux hypothèses largement applicables, comme le montre le théorème de convergence monotone et le théorème de convergence dominé.

Théorème 34 ***(Théorème de convergence monotone)*** *Soit* $X, X_1, X_2, ...$ *des variables aléatoires positives sur* $(\Omega, \mathcal{F}, P)$ *avec* $X_n \uparrow X$*, alors* $E(X_n) \uparrow E(X)$.

Théorème 35 ***(Théorème de convergence dominée)*** *Soit* $(X_n)_{n\in\mathbb{N}^*}$ *une suite de variables aléatoires sur* $(\Omega, \mathcal{F}, P)$ *telle que* $X_n \to^{p.s} X$ *et* $|X_n| < Y$*,* $n \in \mathbb{N}^*$*, avec* $E(Y) < \infty$*, alors* $\lim_{n\to\infty} E(X_n) = E(X)$.

Le résultat suivant illuste les implications qui sont toujours valables.

Théorème 36 *On a les implication suivantes :*

1. $X_n \to^{p.s} X \Rightarrow X_n \to^P X \Rightarrow X_n \to^d X$.
2. $X_n \to^{m.q} X \Rightarrow X_n \to^{L^1} X \Rightarrow X_n \to^P X$.
3. $p > 0, X_n \to^{L^p} X \Rightarrow X_n \to^P X$.

La convergence en loi vers une constante implique une convergence en probabilité.

Théorème 37 *Si* $X_n \rightarrow^d c$, *alors* $X_n \rightarrow^P c$.

La convergence presque sûre, la convergence en probabilité, la convergence en moyenne quadratique et la convergence en moyenne sont conservées sous l'addition.

Théorème 38 *Soit* $X, Y, X_n, Y_n, n \in \mathbb{N}$ *des variables aléatoires.*

1. *Si* $X_n \rightarrow^{p.s} X$ *et* $Y_n \rightarrow^{p.s} Y$, *alors* $X_n + Y_n \rightarrow^{p.s} X + Y$.
2. *Si* $X_n \rightarrow^{P} X$ *et* $Y_n \rightarrow^{P} Y$, *alors* $X_n + Y_n \rightarrow^{P} X + Y$.
3. *Si* $X_n \rightarrow^{m.q} X$ *et* $Y_n \rightarrow^{m.q} Y$, *alors* $X_n + Y_n \rightarrow^{m.q} X + Y$.
4. *Si* $X_n \rightarrow^{L^1} X$ *et* $Y_n \rightarrow^{L^1} Y$, *alors* $X_n + Y_n \rightarrow^{L^1} X + Y$.

Soit $X, Y, X_n, Y_n, n \in \mathbb{N}$ des variables aléatoires. Si $X_n \rightarrow^d X$ et $Y_n \rightarrow^d Y$, cela n'implique pas que $X_n + Y_n \rightarrow^d X + Y$, mais ceci est vrai lorsque une des limites X ou Y est constante.

Théorème 39 ***(Théorème de Slutsky)*** *Soit* $X, X_n, Y_n, n \in \mathbb{N}$ *des variables aléatoires. Si* $X_n \rightarrow^d X$ *et* $Y_n \rightarrow^d c \in \mathbb{R}$, *alors* $X_n + Y_n \rightarrow^d X + c$.

La convergence presque sûrement et en probabilité sont conservées sous la multiplication. La convergence en moyenne quadratique des produits ne tient pas en générale parce que XY *peut ne pas être dans* L^2 *quand* X *et* Y *sont dans* L^2. *Cependant, le produit des variables aléatoires dans* L^2 *appartient à* L^1 *et la convergence des facteurs dans* L^2 *implique la convergence des produits dans* L^1.

Théorème 40 *Soit* $X, Y, X_n, Y_n, n \in \mathbb{N}$ *des variables aléatoires.*

1. *Si* $X_n \rightarrow^{p.s} X$ *et* $Y_n \rightarrow^{p.s} Y$, *alors* $X_n Y_n \rightarrow^{p.s} XY$.
2. *Si* $X_n \rightarrow^{p} X$ *et* $Y_n \rightarrow^{p} Y$, *alors* $X_n Y_n \rightarrow^{p} XY$.
3. *Si* $X_n \rightarrow^{m.q} X$ *et* $Y_n \rightarrow^{m.q} Y$, *alors* $X_n Y_n \rightarrow^{L^1} XY$.

Le théorème de Slutsky pour la multiplication est donnée comme suit.

Théorème 41 ***(Théorème de Slutsky)*** *Soit* $X, X_n, Y_n, n \in \mathbb{N}$ *des variables aléatoires. Si* $X_n \rightarrow^d X$ *et* $Y_n \rightarrow^d c \in \mathbb{R}$, *alors* $X_n Y_n \rightarrow^d cX$.

1.4 Théorèmes limites

Dans cette partie on donne quelques théorèmes limites pour les suites de variables aléatoires indépendantes.

1.4.1 Loi des grands nombres

On commence par rappeler l'énoncé de la loi des grands nombres pour le schéma de Benoulli. Rappelons qu'une épreuve de Beroulli de paramètre p compris entre 0 et 1 est une expérience aléatoire comportant deux issues, le succés ou l'échec. On appelle schéma de Benoulli de paramètre n et p toute expérience aléatoire consistant à répéter n fois de façon indépendante une épreuve de Bernoulli de paramètre p. Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables indépendantes identiquement distribuées avec $P(X_1 = 1) = p$ et $P(X_1 = 0) = 1 - p$. En terme de concept de convergence en probabilité, la loi des grands nombres de Bernoulli peut être énoncée comme suit :

$$\frac{S_n}{n} \to^P p, n \to \infty, \text{ où } S_n = X_1 + ... + X_n$$

Maintenant, soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables indépendantes avec $E|X_n| < \infty$, $n \in \mathbb{N}^*$ et les variaces $Var(X_n), n \in \mathbb{N}^*$ sont telles que $Var(X_n) \leq \alpha < \infty$, alors par l'inégalité de Chebyshev on a la loi faible des grands nombres

$$\frac{S_n - E(S_n)}{n} \to^P 0, n \to \infty \tag{1.1}$$

Une loi forte des grands nombres est une proposition dans laquelle la convergence en probabilité est remplacée par la convergence presque sûrement. L'un des premiers résultats dans cette direction est le théorème suivant.

Théorème 42 ***(Kolmogorov)*** *Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes avec $E(X_n^2) < \infty, n \in \mathbb{N}^*$ et $b_n > 0, n \in \mathbb{N}^*$ tels que $b_n \uparrow \infty$ et $\sum_{n-1}^{\infty} Var(X_n)/b_n^2 < \infty$, alors*

$$\frac{S_n - E(S_n)}{b_n} \to^{p.s} 0, n \to \infty.$$

Dans le cas où $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes et identiquement distribuées, on peut obtenir une loi forte des grands nombres sans supposer l'existence de second moment, on suppose juste l'existence de premier moment absolu.

Théorème 43 ***(Kolmogorov)*** *Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes et identiquement distribuées avec $E|X_1| < \infty$, alors*

$$\frac{S_n}{n} \to^{p.s} E(X_1), n \to \infty.$$

1.4.2 Théorème Central Limite

Les preuves des premiers théorèmes limites de la théorie des probabilité, la loi des grands nombres, le théorème de Moivre-Laplace et le théorème limite de Poisson pour le shéma de Bernoulli, reposent sur une analyse directe des fonctions de distribution F_n qui s'expriment assez simplement en termes des probabilités binomiales. (Dans le shéma de Bernoulli on a ajouté des variables aléatoires qui prennent que deux valeurs de sorte qu'on puisse trouver F_n explicitement). Chebyshev a fait la première étape dans la démonstration des théorèmes limites pour les sommes de variables aléatoires arbitrairement distribuées. L'inégalité qu'il a découverte permet non seulement de donner une preuve élémentaire de la loi des grands nombre de Bernoulli, mais aussi d'établir des conditions trés générales pour la validité de cette loi lorsqu'elle est énoncée dans la forme (1.1). De plus, Chebyshev créa la méthode des moments qui permet de montrer que la conclusion de théorème de Moivre-Laplace écrite sous la forme

$$P\{(S_n - ES_n)/\sqrt{VarS_n} \leq x\} \to \frac{1}{\sqrt{2\Pi}} \int_{-\infty}^{x} e^{\frac{-u^2}{2}} du (= \phi(x)), n \to \infty.$$

est universelle dans le sens qu'elle est valable sous des hypothèses trés générales concernant la nature des variables aléatoires. Pour cette raison, il est connu sous le nom de théorème central limite de la théorie de probabilité. La méthode des fonctions charactéristiques est appliquée pour montrer ces théorèmes limites suivants.

Théorème 44 ***(Loi faible des grands nombre de Khinchin)*** *Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de* *variables aléatoires indépendantes et identiquement distribuées avec* $E|X_1| < \infty$ *et* $E(X_1) = m$*, alors* $S_n/n \to^P E(X_1)$.

Théorème 45 ***(Théorème Central Limite pour une suite iid)*** *Soit* $(X_n)_{n\in\mathbb{N}^*}$ *une suite de variables aléatoires indépendantes et identiquement distribuées non dégénérés avec* $E(X_1^2) < \infty$*, alors*

$$P\{P\{(S_n - ES_n)/\sqrt{VarS_n} \leq x\} \leq x\} \to \phi(x), x \in \mathbb{R},$$

où $\phi(x)$ *est la fonction de distribution de la loi normale centrée réduite* $N(0,1)$*, i.e.,*

$$\phi(x) = \frac{1}{\sqrt{2\Pi}} \int_{-\infty}^{x} e^{\frac{-u^2}{2}} du.$$

1.5 Espace de probabilité d'un phénomène évolutif

Pour représenter mathématiquement un phénomène aléatoire évolutif donné, la première étape consiste à lui associer un espace de probabilités $(\Omega, \mathcal{F}, P)$ où Ω désigne l'ensemble des réalisations possibles du phénomène, $\mathcal{F}$ est une tribu de parties de Ω représentant les évènements que l'on puisse formuler, à priori, à propos des réalisations possibles de ce phénomène et P est une mesure de probabilité sur $\mathcal{F}$, mesurant les chances de réalisation des évènements de $\mathcal{F}$associés. L'espace fondamental Ω est supposé dépendre implicitement d'un ensemble T représentant le domaine d'évolution du phénomène. A chaque moment $t \in T$ de l'évolution, on peut associer un ensemble Ω_t représentant les résultats possibles du phénomène à l'instant t. Sur tout le domaine T, l'ensemble des réalisations possibles Ω sera le produit cartésien des ensemble Ω_t, i.e., $\Omega = \prod\limits_{i\in T} \Omega_t$. La tribu $\mathcal{F}$, quant elle, dépend également du domaine d'évolution T. A tout moment $t \in T$ de l'évolution, on peut associer également une tribu élémentaire $\mathcal{F}_t$ représentant les évènements liés au phénomène à l'instant t. Sur tout le domaine T, la tribu $\mathcal{F}$des évènements liés au phénomène est le produit $\mathcal{F} = \bigotimes\limits_{t\in T} \mathcal{F}_t$ qui est la plus petite tribu contenant tous les ensembles $\prod\limits_{t\in T} A_t$ avec $A_t \in \mathcal{F}_t$ et pour éviter des situations de dégénérescence, cette tribu est telle qu'elle doit renfermer les évènements élémentaires associés à tout moment d'évolution (e.g. $\{\omega_t\} \in \mathcal{F}$ pour tout t). De tels évènements sont dits projections des évènements élémentaires $\{\omega = (\omega_t, t \in T)\}$. Puisque Ω représente un produit d'espaces élémentaires Ω_t, les évènements élémentaires projections peuvent s'exprimer à travers des ensembles cylindriques que la tribu $\mathcal{F}$doit contenir. Pa exemple, dans le cas où $T = \mathbb{N}$, les ensembles cylindriques sont de la forme $\prod_{k=0}^{j-1} \Omega_k \times A_j \times \prod_{k=j+1}^{\infty} \Omega_k$, $A_j \in \mathcal{F}_j, j \in \mathbb{N}$. L'espace probabilisable $(\Omega, \mathcal{F})$ est le produit des espaces élémentaires $(\Omega_t, \mathcal{F}_t)$, i.e., $(\Omega, \mathcal{F}) = \prod_{t \in T} (\Omega_t, \mathcal{F}_t)$. La mesure de probabilité P définie sur $\mathcal{F}$ dépend également de T. Pour tout $t \in T$, soit P_t une mesure de probabilité sur l'espace élémentaire $(\Omega_t, \mathcal{F}_t)$, alors sur tout le domaine d'évolution T, la mesure de probabilité P est fonction des $(P_t \in T)$. Dans le cas où le phénomène est soumis à l'indépendance mutuelle, $P = \prod_{t \in T} P_t$. L'espace de probabilité $(\Omega, \mathcal{F}, P)$ est donc le produit d'espaces élémentaires $(\Omega_t, \mathcal{F}_t, P_t)$, i.e., $(\Omega, \mathcal{F}, P) = \prod_{t \in T} (\Omega_t, \mathcal{F}_t, P_t)$.

1.5.1 Processus aléatoire

L'espace de probabilité $(\Omega, \mathcal{F}, P)$ étant complètement spécifié, on peut en associer une application qui permet de numériser l'ensemble Ω des réalisation possibles, lequel peut être quelconque. Par numériser on entend associer à chaque réalisation du phénomène, à tout moment $t \in T$ de l'évolution, un nombre ou un vecteur de nombres réels. Lorsque Ω est d'emblée numérique dans le sens où les composantes de $\omega \in \Omega$ sont des nombres ou des vecteurs de réels, on peut prendre pour application l'identité et l'espace $(\Omega, \mathcal{F}, P)$ est alors dit canonique. Une telle application numérisant Ω, dite processus aléatoire, et qui doit vérifier certaines conditions de probabilisabilité (mesurabilité), peut être définie de plusieurs manières. On réservera ici trois définitions dont chacune se base sur un angle de vue différent.

1. vision verticale : regarde un processus d'un point de vue de ses membres, c'est-à-dire de variables aléatoires le constituant.
2. vision horizontale : définit un processus par rapport à ses réalisations.
3. vision croisée : définit le processus du point de vu du croisement d'un membre du

processus et d'une réalisation possible.

Définition 46 *Un processus aléatoire de domaine d'évolution* T*, défini sur un espace de probabilité* $(\Omega, \mathcal{F}, P)$ *et à valeurs dans* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *est une famille de variables aléatoires* $(X_t, t \in T)$ *chacune définie sur* $(\Omega, \mathcal{F})$ *à valeurs dans* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$*. Autrement dit,* $(X_t, t \in T)$ *est une application qui associe pour tout* t *dans* T *une variable aléatoire* X_t *dans* $D((\Omega, \mathcal{F}), (\mathbb{R}, \mathcal{B}(\mathbb{R})))$*, ensemble des fonctions* $\mathcal{F}$*-mesurables, définies de* Ω *à valeurs dans* $\mathbb{R}$*, i.e.,*

$$\begin{aligned} (X_t, t \in T) : T &\to D((\Omega, \mathcal{F}), (\mathbb{R}, \mathcal{B}(\mathbb{R}))) \\ t &\longmapsto X_t \end{aligned}$$

Remarque 47 *Cette définition, d'essence récursive, ne se préoccupe pas à priori de la condition de mesurabilité qui est déja garantie par le fait d'avoir défini un processus en tant que famille de variables aléatoires, donc l'application déja mesurable. Les variables aléatoires* $X_t, t \in T$ *peuvent être discrètes ou continues.*

La définition suivante définit un processus aléatoire en terme de ses réalisations (dites aussi trajectoires), i.e., définit un processus aléatoire comme fonction aléatoire (élément aléatoire à valeurs dans $\mathbb{R}^T$). Soit $\mathcal{B}(\mathbb{R}^T)$ la pribu de Borel de sous ensembles de $\mathbb{R}^T$, i.e., la plus petite tribu de sous ensembles de $\mathbb{R}^T$ contenant tous les ensembles cylindriques de la forme

$$C = \{x = (x_t, t \in T) \in \mathbb{R}^T : x_{t_1} \in B_1, ..., x_{t_n} \in B_n\}, \text{ où } B_k \in \mathcal{B}(\mathbb{R}), k = 1, ..., n, n \in \mathbb{N}^*.$$

Définition 48 *Un processus aléatoire* $X = (X_t, t \in T)$ *de domaine d'évolution* T *défini sur un espace de probabilité* $(\Omega, \mathcal{F}, P)$ *et à valeurs dans* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *est une application mesurable de* $(\Omega, \mathcal{F})$ *dans* $(\mathbb{R}^T, \mathcal{B}(\mathbb{R}^T))$*. Autrement dit, c'est une application qui pour toute réalisation possible* ω *du phénomène, associe une famille de nombres* $\mathbf{X}(\omega) = (X_t(\omega), t \in T)$ *et est telle que l'image réciproque de tout Borélien* B *de* $\mathcal{B}(\mathbb{R}^T)$ *est un membre de* $\mathcal{F}$*, i.e.,*

$$\begin{aligned} \mathbf{X} \quad &: \quad \Omega \to \mathbb{R}^T \\ \omega &\longmapsto \mathbf{X}(\omega) = (X_t(\omega), t \in T) \end{aligned}$$

où pour tout $B \in \mathcal{B}(\mathbb{R}^T), \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$.

Une troisième définition mais qui n'est pas tout à fait équivalente aux deux premières dans le sens où elle exige des conditions supplémentaires, regarde un processus du point de vue et de ses membres et de ses réalisations.

Définition 49 *Un processus aléatoire $X = (X_t, t \in T)$ de domaine d'évolution T défini sur un espace de probabilité $(\Omega, \mathcal{F}, P)$ et à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est une application mesurable de $(\Omega \times T, \mathcal{F} \bigotimes \mathcal{T})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ où $\mathcal{T}$ est une tribu associée à T et $\mathcal{F} \bigotimes \mathcal{T}$ la tribu engendrée par tous les ensembles $A \times B$, $A \in \mathcal{F}$, $B \in \mathcal{T}$. Ceci se traduit par*

$$\begin{aligned} \mathbf{X} \quad &: \quad \Omega \times T \to \mathbb{R} \\ (\omega, t) \quad &\mapsto \quad \mathbf{X}(\omega, t) = X_t(\omega) \end{aligned}$$

où pour tout $C \in \mathcal{B}(\mathbb{R})$, $\{(\omega, t) \in \Omega \times T : X_t(\omega) \in C\} \in \mathcal{F} \bigotimes \mathcal{T}$.

1.5.2 Distribution de probabilité d'un processus aléatoire

Pour un processus aléatoire $\mathbf{X} = (X_t, t \in T)$ défini sur un espace de probabilité $(\Omega, \mathcal{F}, P)$ à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et de domaine d'évolution T dénombrable, sa structure de probabilité est caractérisée par la distribution infini-dimensionnelle $P_{\mathbf{X}}(.)$ définie sur $B(\mathbb{R}^T)$ à valeurs dans [0,1] comme suit :

$$\begin{aligned} P_{\mathbf{X}}(.) \quad &: \quad \mathcal{B}(\mathbb{R}^T) \longrightarrow [0,1] \\ B \quad &\longmapsto \quad P_{\mathbf{X}}(B) = P\{\omega \in \Omega : \mathbf{X}(\omega) \in B\}. \end{aligned}$$

Cette distribution infini-dimensionnelle est difficile à manipuler et qu'il existe un outil simple, la distribution fini-dimensionnelle, permettant de simplifier son analyse. L'introduuction de cette distribution est justifiée par le théorème d'extension de Kolmogorov qui stipule que la distribution infini-dimensionnelle est uniquement déterminée par les probabilités fini-dimensionnelles. La définition suivante est valable pour tout type de processus.

Définition 50 *Pour $n \in \mathbb{N}^*$ et pour tous $t_1, t_2, ..., t_n \in T$, la distribution fini-dimensionnelle de $\mathbf{X} = (X_t, t \in T)$ est la distribution de toute sous-suite finie $(X_{t_1}, X_{t_2}, ..., X_{t_n})$ de $\mathbf{X}$.*

Autrement dit, c'est la fonction ensembliste $P_{X_{t_1},X_{t_2},\ldots,X_{t_n}}(.)$ définie sur $\mathcal{B}(\mathbb{R}^n)$ par

$$\begin{aligned} P_{X_{t_1},X_{t_2},\ldots,X_{t_n}}(.) \quad &: \quad \mathcal{B}(\mathbb{R}^n) \to [0,1] \\ B \quad &\longmapsto \quad P_{X_{t_1},X_{t_2},\ldots,X_{t_n}}(B) = P\{\omega \in \Omega : (X_{t_1}(\omega), X_{t_2}(\omega), \ldots, X_{t_n}(\omega)) \in B\}. \end{aligned}$$

Afin d'éviter la manipulation de Boréliens, on peut de manière équivalente caractériser la structure probabiliste du processus aléatoire au moyen de la fonction de répartition (de distribution) fini-dimensionnelle.

Définition 51 *Pour $n \in \mathbb{N}^*$ et pour tout $t_1, t_2, \ldots, t_n \in T$, la fonction de répartition fini-dimensionnelle de $\mathbf{X} = (X_t, t \in T)$ est la fonction de répartition de toute suite finie $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$ de $\mathbf{X}$. Autrement dit, c'est la fonction $F_{X_{t_1},X_{t_2},\ldots,X_{t_n}}(.)$ définie sur $\mathbb{R}^n$ à valeurs dans $[0,1]$ par*

$$\begin{aligned} F_{X_{t_1},X_{t_2},\ldots,X_{t_n}}(.) \quad &: \quad \mathbb{R}^n \to [0,1] \\ (x_1, x_2, \ldots, x_n) \quad &\longmapsto \quad F_{X_{t_1},X_{t_2},\ldots,X_{t_n}}(x_1, x_2, \ldots, x_n) = P(\bigcap_{j=1}^{n}\{\omega : X_{t_j}(\omega) \leq x_j\}). \end{aligned}$$

Pour tout processus aléatoire défini sur un espace de probabilité $(\Omega, \mathcal{F}, P)$ à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, il est clair que l'ensemble de ses fonctions de répartition fini-dimensionnelles $\{F_{X_{t_1},X_{t_2},\ldots,X_{t_n}}, n \in \mathbb{N}^*, t_1, \ldots t_n \in T\}$ vérifient :

1. $F_{X_{t_1},X_{t_2},\ldots,X_{t_n}}(x_1, x_2, \ldots, x_n) = F_{X_{t_{i_1}},X_{t_{i_2}},\ldots,X_{t_{i_n}}}(x_{i_1}, x_{i_2}, \ldots, x_{i_n})$ pour toute permutation $i_1, i_2, \ldots, i_n$ de $1, 2, \ldots, n$ et tout $x_i \in \mathbb{R}$.
2. L'ensemble des fonctions de répartition fini-dimensionnelles a la propriété de consistance suivante

$$F_{X_{t_1},X_{t_2},\ldots,X_{t_n}}(x_1, x_2, \ldots, x_{n-1}, \infty) = F_{X_{t_1},X_{t_2},\ldots,X_{t_{n-1}}}(x_1, x_2, \ldots, x_{n-1}). \qquad (1.2)$$

Maintenant, il est naturel de poser le problème inverse suivant : sous quelles conditions qu'une famille $\{F_{t_1,\ldots t_n}, n \in \mathbb{N}^*, t_1, \ldots t_n \in T\}$ de fonctions de distributions soit une famille de fonctions de répartitions fini-dimensionnelles d'un processus aléatoire ? Ce problème a été résolu par Kolmogorov en 1933.

Théorème 52 *(Théorème de Kolmogorov) Soit* $\{F_{t_1,t_2,...,t_n}, n \in \mathbb{N}^*, t_1, t_2, ..., t_n \in T\}$ *une famille de fonctions de distribution fini-dimensionnelles qui satisfait la condition de consistance précédente (1.2), alors il existe un espace de probabilité* $(\Omega, \mathcal{F}, P)$ *et un processus aléatoire* $\mathbf{X} = (X_t, t \in T)$ *tel que*

$$P\{\omega \in \Omega : X_{t_1}(\omega) \leq x_1, ..., X_{t_n}(\omega) \leq x_n\} = F_{t_1,t_2,...,t_n}(x_1, ..., x_n)$$

1.5.3 Caractéristiques de la distribution d'un processus aléatoire

Soit $X = (X_t, t \in T)$ un processus aléatoire défini sur un espace de probabilité $(\Omega, \mathcal{F}, P)$ à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et de domaine d'évolution T. Comme pour les variables aléatoires, la distribution infini-dimensionnelle d'un processus aléatoire est aussi caractérisée par certaines familles particulières définies sur T à valeurs dans $\mathbb{R}$, à savoir : la fonction moyenne, la fonction variance et la fonction d'autocovariance.

Définition 53 *La fonction moyenne* $\mu(.)$ *est une fonction de* T *dans* $\mathbb{R}$ *qui pour tout* $t \in T$ *associe l'éspérence mathématique du membre* X_t *et ayant comme domaine de définition l'ensemble* $D_\mu = \{t \in T : E(\ X_t) \text{ existe}\}$.

Définition 54 *La fonction variance* $\sigma^2(.)$ *est une fonction de* T *dans* $\mathbb{R}^+$ *qui pour tout* $t \in T$ *associe la variance du membre* X_t *et ayant comme domaine de définition l'ensmble* $D_{\sigma^2} = \{t \in T : E(\ X_t^2) < \infty\}$.

Définition 55 *La fonction d'autocovariancec* $\gamma(.,.)$ *est une fonction de* $T \times T$ *dans* $\mathbb{R}$ *qui pour tout couple* $(t, s) \in T \times T$ *associe la covariance entre les membres* X_t *et* X_s *et ayant comme domaine de définition l'ensemble* $D_\gamma = \{(t, s) \in T \times T : E(X_t, X_s) < \infty\}$.

Définition 56 *La fonction d'autocorrélation* $\rho(.,.)$ *est une fonction de* $T \times T$ *dans* $[-1, 1]$ *qui pour tout couple* $(t, s) \in T \times T$ *associe la corrélation entre les membres* X_t *et* X_s *et ayant comme domaine de définition* $D_\rho = D_\gamma$.

1.6 Stationnarité des processus aléatoires

Soit $\mathbf{X} = (X_t, t \in T)$ un processus aléatoire de domaine d'évolution T défini sur $(\Omega, \mathcal{F}, P)$ et à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ avec le domaine d'évolution T a la propriété que la somme de deux points de T est également dans T. Souvent on prend $T = \mathbb{N}$ mais peut être $\mathbb{Z}, \mathbb{R}^+$ou $\mathbb{R}$.

Définition 57 *Le processus $\mathbf{X} = (X_t, t \in T)$ est strictement stationnaire si pour tout $n \geq 1$ et tous points $t_1, ..., t_n, h$ dans T, la distribution de $(X_{t_1}, ..., X_{t_n})$ est la même que la distribution de $(X_{t_1+h}, ..., X_{t_n+h})$, i.e.,*

$$P(X_{t_1}, ..., X_{t_n}) \in B) = P((X_{t_1+h}, ..., X_{t_n+h}) \in B), B \in \mathcal{B}(\mathbb{R}^n).$$

Soit $\mathbf{X} = (X_t, t \in T)$ un processus aléatoire aléatoire strictement stationnaire. Si $m_t = E(X_t)$ existe et finie, il s'en suit que m_t est constante pour tout $t \in T$, i.e., $m_t = m$. De même si $E(X_t^2) < \infty$, alors la variance $\sigma_t^2 = E[X_t - E(X_t)^2]$ est constante et indépendante de t. Soit $t, s \in T$ et supposons que $t > s$. En utilisant la propriété de stationnarité stricte, on a

$$cov(X_t, X_s) = E[(X_t - m)(X_s - m)] = E[(X_{t-s} - m)(X_0 - m)],$$

i.e., $cov(X_t, X_s)$ ne dépend que de différence $t - s$. Si on définit la fonction de covariance

$$\gamma_{\mathbf{X}}(h) = cov(X_h, X_0) = E[(X_h - m)(X_0 - m)], \ h \in T,$$

alors pour $t, s \in T, cov(X_t, X_s) = E[(X_t - m)(X_s - m)] = \gamma_X(|t - s|)$. Notons que $\sigma^2 = \gamma_X(0)$. Parfois il est pratique de standariser la fonction covariance en produisant ce qu'on appelle fonction d'autocorrélation ou bien fonction de corrélation de processus $X = (X_t, t \in T)$ définie par

$$\rho_{\mathbf{X}}(t) = \frac{\gamma_{\mathbf{X}}(t)}{\gamma_{\mathbf{X}}(0)}$$

Notons que $\rho_{\mathbf{X}}(0) = 1$ et par l'inégalité de CBS, $-1 \leq \rho_{\mathbf{X}}(t) \leq 1$ pour tout $t \in T$.

Il est important de noter que tous les moments d'un processus strictement stationnaire, lorsqu'ils existent, sont invariants dans le temps. Le processus est ainsi dit stationnaire en tous les moments ou à l'ordre ∞. Ainsi, la définition de stationnarité stricte semble contraignante puisque lorsque tous les moments existent, elle exige l'invariance de tous ces moments par rapport au temps. De plus, elle repose sur la connaissance des distributions fini-dimensionnelles du processus qui ne peuvent être connues en pratique, sauf dans des cas trés spéciaux. Cependant, plusieurs propriétés probabilistes essentielles des processus aléatoires peuvent être obtenues juste à partir des deux premiers moments (lorsqu'il existent) et pour les moments restants, la distinction est souvent négligeable. La stationnarité de ces deux premiers moments peut donc être suffisante pour expliquer, du moins avec bonne précision, la stationnarité dans la distribution du processus. C'est pourquoi, on a souvent besoin d'un concept de stationnarité moins fort mais qui peut être rencontré en pratique.

Définition 58 *Le processus aléatoire* $\mathbf{X} = (X_t, t \in T)$ *est dit faiblement stationnaire si*

1. $E(X_t^2) < \infty, t \in T,$
2. $E(X_t) = m, t \in T,$
3. $cov(X_t, X_s) = \gamma_{\mathbf{X}}(|t-s|), t, s \in T.$

D'autre termes utilisés dans la littérature comme synonyme de faiblement stationnaire sont stationnaire au second-ordre ou covariance stationnaire et le processus strictement stationnaire est souvent appelé stationnaire. Un processus strictement stationnaire qui a des moments d'ordre deux finis est faiblement stationnaire (mais, bien sur, un processus strictement stationnaire peut ne pas avoir aucun moment fini). Il est fort possible qu'un processus faiblement stationnaire ne soit pas strictement stationnaire, mais il existe une exception importante à cette règle générale. Un processus aléatoire $\mathbf{X} = (X_t, t \in T)$ pour lequelle pour tout $n \geq 1$ et tous $t_1, ..., t_n$, le vecteur aléatoire $(X_{t_1}, ..., X_{t_n})$ a une distribution Normale est dit processus Gaussien. Puisque la distribution normale multivariétée est déterminée par ses deux premiers moments, le vecteur moyenne et la matrice de covariance, le processus Gaussien qui est stationnaire au second-ordre sera aussi strictement stationnaire.

1.7 Processus stationnaire et théorème ergodique

Soit $\mathbf{X} = (X_n, n \in \mathbb{N}^*)$ un processus aléatoire défini sur $(\Omega, \mathcal{F}, P)$ et à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, i.e., une suite de variables aléatoires. Si $\mathbf{X} = (X_n, n \in \mathbb{N}^*)$ est une suite de variables aléatoires indépendante identiquement distribuées avec $E|X_1| < \infty$, alors par la loi forte des grands nombres

$$\frac{X_1 + ... + X_n}{n} \rightarrow^{p.s} E|X_1|. \tag{1.3}$$

Au lieu d'exiger que les variables $(X_n, n \in \mathbb{N}^*)$ sont indépendantes identiquement distribuées, Brkhoff a prouvé la convergence presque sûre de la moyenne empirique dans (1.3) en exigent seulemet que la distribution du processus $(X_n, n \in \mathbb{N}^*)$ ne dépend pas du placement de l'origine. Rappelons qu'on note par $\mathbb{R}^\infty$ est l'ensemble de toutes les suites infinies de nombres réels $x = (x_{1,}x_2, ...)$. Un rectangle de dimension n dans $\mathbb{R}^\infty$ est l'ensemble $\{x \in \mathbb{R}^\infty : x_1 \in I_1, ..., x_n \in I_n\}$, où $I_1, ...I_n$ sont des intervalles finis ou infinis. Un cylindre de dimension n dans $\mathbb{R}^\infty$ est tout ensemble de la forme $\{x \in \mathbb{R}^\infty : (x_1, ..., x_n) \in B_n\}$, où $B_n \in B(\mathbb{R}^n)$. La σ-algèbre de Borel $\mathcal{B}(\mathbb{R}^\infty)$ est la plus petite σ-algèbre des sous-ensembles $\mathbb{R}^\infty$ contenant tous les rectangles de dimensions finis.

Définition 59 *Un processus aléatoire $(X_n, n \in \mathbb{N}^*)$ est strictement stationnaire si pour tout $k \in \mathbb{N}^*$, le processus aléatoire $(X_{n+k}, n \in \mathbb{N}^*)$ a la même distribution que $(X_n, n \in \mathbb{N}^*)$, i.e., pour tout $B \in \mathcal{B}(\mathbb{R}^\infty)$*

$$P((X_1, X_2, ...) \in B) = P((X_{k+1}, X_{k+2}, ...) \in B). \tag{1.4}$$

Notons que la distribution d'un processus contient toute l'information pertinente pour la théorie des probabilités. Tous les théorèmes qu'on démontrera ne dépent que de la distribution du processus et par conséquent il sont valables pour tous les processus ayant cette distribution. Parmi tous les processus aléatoires ayant la même distribution donnée $\widetilde{P}$ sur $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$, il y'en a un qui est le plus simple.

Définition 60 *Pour toute distribution $\widetilde{P}$ sur $\mathcal{B}(\mathbb{R}^\infty)$, on définit un processus $(\widetilde{X_n}, n \in \mathbb{N}^*)$*

sur $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty),\ \widetilde{P})$ *par*

$$\widetilde{X_n}(x_1, x_2, ...) = x_n.$$

Ce processus est dit processus de représentation en cordonnées et il a la même distribution que le processus original parce que pour tout $B \in B(\mathbb{R}^\infty)$,

$$\widetilde{P}\{x \in \mathbb{R}^\infty : (\widetilde{X_1}, \widetilde{X_2}, ...) \in B\} = \widetilde{P}\{x \in \mathbb{R}^\infty : x \in B\} = \widetilde{P}(B).$$

À partir de tout processus strictement stationnaire, on peut construire une infinité de processus strictement stationnaires.

Théorème 61 *Soit* $(X_n, n \in \mathbb{N}^*)$ *un processus aléatoire strictement stationnaire et* $f : \mathbb{R}^\infty \to \mathbb{R}$ *une fonction Borel mesurable, alors le processus aléatoire* $(Y_n, n \in \mathbb{N}^*)$ *défini par* $Y_n = f(X_n, X_{n+1}, ...)$ *est strictement stationnaire.*

Un cas particulier des processus strictement stationnaire est la suite de variables aléatoires indépendantes et identiquement distribuées.

Corollaire 62 *Soit* $(X_n)_{n\in\mathbb{N}^*}$ *une suite indépendantes et identiquement distribuée et* $f : \mathbb{R}^\infty \to \mathbb{R}$ *une fonction Borel mesurable, alors le processus aléatoire* $(Y_n)_{n\in\mathbb{N}^*}$ *défini par* $Y_n = f(X_n, X_{n+1}, ...)$ *est strictement stationnaire.*

1.7.1 Transformations préservant la mesure

Soit un espace de probabilité $(\Omega, \mathcal{F}, P)$ et $T : \Omega \to \Omega$. Rappelons que T est une application mesurable de $(\Omega, \mathcal{F})$ dans $(\Omega, \mathcal{F})$ si $T^{-1}(A) = \{\omega \in \Omega : T(\omega) \in A\} \in \mathcal{F}$ pour tout $A \in \mathcal{F}$.

Définition 63 *Une transformation mesurable* $T : \Omega \to \Omega$ *est dite préservant la mesure* P, *si* $P(T^{-1}A) = P(A),\ A \in \mathcal{F}$.

À partir des transformations préservant la mesure P, un grand nombres de processus stationnaires peuvent être générés.

Proposition 64 *Soit* $T : \Omega \to \Omega$ *une transformation préservant la mesure sur* $(\Omega, \mathcal{F}, P)$ *et* $X(\omega)$ *une variable aléatoire définie sur* $(\Omega, \mathcal{F}, P)$, *alors la suite* $X(T^{n-1}\omega), n = 1, 2, ...$ *est strictement stationnaire.*

Le processus stationnaire construit dans la proposition 64 est dit processus généré par la transformation préservant la mesure T. En termes de distribution, tout processus strictement stationnaire peut être générée par une transformation préservant la mesure. Considérons un processus aléatoire strictement stationnaire $\mathbf{X} = (X_n, n \in \mathbb{N}^*)$ défini sur $(\Omega, \mathcal{F}, P)$ et $\widetilde{\mathbf{X}} = (\widetilde{X_n}, n \in \mathbb{N}^*)$ le processus de représentation des coordonnées sur $(\mathbb{R}^\infty, B(\mathbb{R}^\infty), P_X)$. Par définition, $\widetilde{X_n}(x) = x_n$.

Définition 65 *Sur l'espace mesurable* $(\mathbb{R}^\infty, B(\mathbb{R}^\infty))$, *on définit la transformation du décalage* $S : \mathbb{R}^\infty \to \mathbb{R}^\infty$ *par*

$$Sx = S(x_1, x_2, ...) = (x_2, x_3, ...).$$

On remarque immédiatement que $\widetilde{X_n}(x) = \widetilde{X_1}(S^{n-1}x), n \geq 2$.

Proposition 66 *La transformation du décalage* $S : \mathbb{R}^\infty \to \mathbb{R}^\infty$ *est Borel mesurable et si* $\mathbf{X} = (X_n, n \in \mathbb{N}^*)$ *est strictement stationnaire, alors* S *préserve la mesure* P_X *.*

Définition 67 *Soit* T *une transformation préservant la mesure sur* $(\Omega, \mathcal{F}, P)$. *Un évènement* $A \in \mathcal{F}$ *est dit invariant si* $A = T^{-1}A$.

Proposition 68 *La classe* $\mathcal{E}$ *des évènements invariants est une tribu.*

Définition 69 *Une transformation préservant la mesure sur* $(\Omega, \mathcal{F}, P)$ *est dite ergodique si pour tout* $A \in \mathcal{E}$, $P(A) = 0$ *ou* 1.

1.7.2 Théorème ergodique

L'un des résultats des théorèmes limites est le théorème ergodique qui stipule que la moyenne empirique d'une suite de variables aléatoires strictement stationnaire de moyenne finie converge presque sûrement vers une variable aléatoire.

Théorème 70 ***(Birkhoff and Khinchin)*** *Soit T une transformation préservant la mesure sur $(\Omega, \mathcal{F}, P)$ et X une variable aléatoire telle que $E|X| < \infty$, alors*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} X(T^k\omega) = E(X|\mathcal{E}) \quad p.s.$$

Une conséquence directe du théorème 70 est que si T est ergodique, alors la variable limite est constante presque sûrement et egale a la moyenne théorique.

Corollaire 71 *Soit T une transformation préservant la mesure sur $(\Omega, \mathcal{F}, P)$ et ergodique, alors pour toute variable aléatoire X telle que $E|X| < \infty$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} X(T^k\omega) = E(X) \quad p.s.$$

1.7.3 Théorème ergodique pour les processus stationnaires

Soit $X = (X_n, n \in \mathbb{N}^*)$ un processus strictement stationnaire et $S : \mathbb{R}^\infty \to \mathbb{R}^\infty$ la transformation de décalage préservant la mesure P_X. Un ensemble $B \in \mathcal{B}(\mathbb{R}^\infty)$ est invariant sous S si $S^{-1}B = B$. La transformation S est ergodique si pour tout $B \in \mathcal{B}(\mathbb{R}^\infty)$ invariant, $P_{\mathbf{X}}(B) = 0$ ou 1. Par le théorème ergodique 70 et le corollaire 71, si S est ergodique, alors

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \widetilde{X_k} = E(X_1) \; p.s.$$

Si S est ergodique, on obtient les mêmes conclusions pour le processus original parce que la convergence presque sûrement dépend seulement de la distribution du processus. Presque toutes les définitions concernant l'invariance et l'ergodicité peuvent être formulées en termes du processus original au lieu du processus $\widetilde{\mathbf{X}} = (\widetilde{X_n}, n \in \mathbb{N}^*)$ défini sur $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty), P_X)$.

Définition 72 *Un événement $A \in \mathcal{F}$ est invariant s'il existe $B \in \mathcal{B}(\mathbb{R}^\infty)$ tel que pour chaque $n \geq 1$,*

$$A = \{(X_n, X_{n+1}, ...) \in B\}.$$

Le théorème ergodique 70 se traduit comme suit.

Théorème 73 *Soit* $\mathbf{X} = (X_n, n \in \mathbb{N}^*)$ *un processus strictement stationnaire,* $E|X_1| < \infty$ *et* $\mathcal{E}$ *la tribu des événements invariants, alors*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k = E(X|\mathcal{E}) \ \ p.s.$$

Définition 74 *Un processus strictement stationnaire* $\mathbf{X} = (X_n, n \in \mathbb{N}^*)$ *est ergodique si tout événement invariant est de probabilité égale à zéro ou un.*

Théorème 75 *Toute suite de variables aléatoires indépendantes et identiquement distribuées* $(X_n, n \in \mathbb{N}^*)$ *est un processus ergodique.*

Une conséquence directe du théorème 73 est que si le processus aléatoire est ergodique, alors la variable limite est constante presque sûrement et égale à la moyenne théorique.

Corollaire 76 *Si* $\mathbf{X} = (X_n, n \in \mathbb{N}^*)$ *est un processus aléatoire strictement stationnaire et ergodique avec* $E|X| < \infty$*, alors*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k = E(X_1) \ \ p.s.$$

L'ergodicité est préservée sous des fonctions mesurables.

Théorème 77 *Soit* $X = (X_n, n \in \mathbb{N}^*)$ *un processus aléatoire ergodique et* $f : \mathbb{R}^\infty \to \mathbb{R}$ *une fonction Borel mesurable, alors le processus aléatoire* $\mathbf{Y} = (Y_n, n \in \mathbb{N}^*)$ *défini par*

$$Y_n = f(X_n, X_{n+1}, ...)$$

est ergodique.

Chapitre 2

Modèles ARCH et GARCH

2.1 Introduction

Les modèles autorégressifs conditionnellement hétérocédastiques ($ARCH$) ont été proposés par Engle [9] pour capturer la volatilité instantanée caractérisant les séries financières. Leurs extension aux modèles Autorégressifs conditionnellement hétérocédastiques généralisés ($GARCH$) est due à Bollerslev [1]. Dans ces modèles, le concept principal est la variance conditionnelle, c'est-à-dire la variance conditionnée par le passé du processus (conditionnellement à l'ensemble d'informations fournies par le passé de la série chronologique). Dans les modèles classiques $GARCH$, la variance est exprimée comme une fonction linéaire du carré des valeurs passées de la série chronologique. Cette spécification particulière peut capturer les faits stylisés principaux caractérisant la série financière. La structure linéaire de ces modèles est exposée par plusieurs représentations qu'on étudiera dans cette partie. On présente d'abord des définitions et des représentations des modèles $GARCH(p,q)$, puis on étudie les conditions de stationnarité forte ainsi que les conditions de stationnarité au second ordre.

Nous donnons une première définition d'un processus $GARCH$ fondée sur les deux premiers moments de ε_t conditionnels à son passé, i.e., conditionnellement sur la σ-algèbre $\mathcal{F}_{t-1} = \sigma(\varepsilon_u, u < t)$.

Définition 78 ***(Processus GARCH(p,q))*** *On dit que* $(\varepsilon_t)_{t\in\mathbb{Z}}$ *est un processus* $GARCH(p,q)$ *si ses deux premiers moments conditionnels existent et vérifient :*

1. $E(\varepsilon_t/\mathcal{F}_{t-1}) = 0, t \in \mathbb{Z}$.

2. Il existe des constantes $\omega, \alpha_i, i = 1, ..., q$ et β_j, $j = 1, ..., p$ telle que

$$\sigma_t^2 = var(\varepsilon_t/\mathcal{F}_{t-1}) = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2, t \in \mathbb{Z}. \tag{2.1}$$

L' équation (2.1) peut être écrite sous la forme symbolique suivante :

$$\mathcal{B}(B)\sigma_t^2 = \omega + \mathcal{A}(B)\varepsilon_t^2, t \in \mathbb{Z}, \tag{2.2}$$

où B est l'opérateur de retard ($B^i\varepsilon_t^2 = \varepsilon_{t-i}^2$ et $B^i\sigma_t^2 = \sigma_{t-i}^2, \forall i \in \mathbb{N}$), $\mathcal{A}$ et $\mathcal{B}$ sont deux polynômes de degrés q et p, respectivement :

$$\mathcal{A}(B) = \sum_{i=1}^{q} \alpha_i B^i \,; \,\mathcal{B}(B) = 1 - \sum_{j=1}^{p} \beta_j B^j.$$

Si $\mathcal{B}(z) = 1$ on a

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2$$

et le processus (ε_t) est appelé un processus $ARCH(q)$.

Par définition, les innovations relatives au processus $(\varepsilon_t^2)_{t\in\mathbb{Z}}$ sont données par le processus $(\xi_t)_{t\in\mathbb{Z}}$, défini par $\xi_t = \varepsilon_t^2 - \sigma_t^2, t \in \mathbb{Z}$.

En remplaçant la variable σ_{t-j}^2 par $\varepsilon_{t-j}^2 - \xi_{t-j}$ dans l'équation (2.1), on obtient la représentation suivante

$$\varepsilon_t^2 = \omega + \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i)\varepsilon_{t-i}^2 + \xi_t - \sum_{j=1}^{p} \beta_j \xi_{t-j}; \quad t \in \mathbb{Z}, \tag{2.3}$$

avec la convention $\alpha_i = 0$ ($\beta_j = 0$) si $i > q$ ($j > p$). Cette équation a la structure linéaire d'un modèle $ARMA(\max(p, q), p)$. Sous des conditions supplémentaires, incluant la stationnarité au second ordre du processus $(\varepsilon_t^2)_{t\in\mathbb{Z}}$, on peut dire que si $(\varepsilon_t)_{t\in\mathbb{Z}}$ est un processus $GARCH(p, q)$, alors (ε_t^2) est un processus $ARMA(\max(p, q), p)$. En particulier, le carré d'un processus $ARCH(q)$ satisfait, s'il est stationnaire au second ordre, un modèle autorégressif d'ordre q, $AR(q)$. La représentation $ARMA$ est utile pour l'estimation et l'identification des processus $GARCH$.

Remarque 79 *L'innovation ξ_t est conditionnellement hétérocédastique. En effet, si la loi conditionnelle de ε_t sachant $\mathcal{F}_{t-1}$ l'ensemble d'informations disponibles à l'instant $t-1$, suit une loi normale $N(0, \sigma_t^2)$, alors*

$$\begin{aligned}
E(\xi_t^2/\mathcal{F}_{t-1}) &= E((\varepsilon_t^2 - E(\varepsilon_t^2/\mathcal{F}_{t-1}))^2/\mathcal{F}_{t-1}) \\
&= E((\varepsilon_t^4 - 2\varepsilon_t^2 E(\varepsilon_t^2/\mathcal{F}_{t-1}) + (E(\varepsilon_t^2/\mathcal{F}_{t-1}))^2/\mathcal{F}_{t-1}) \\
&= E(\varepsilon_t^4/\mathcal{F}_{t-1}) - 2(E(\varepsilon_t^2/\mathcal{F}_{t-1}))^2 + (E(\varepsilon_t^2/\mathcal{F}_{t-1}))^2 \\
&= E(\varepsilon_t^4/\mathcal{F}_{t-1}) - (E(\varepsilon_t^2/\mathcal{F}_{t-1}))^2 \\
&= 3\sigma_t^4 - \sigma_t^4 = 2\sigma_t^4.
\end{aligned}$$

La définition (78) ne fournit pas directement de processus la vérifiant. La définition plus restrictive suivante permettra d'obtenir explicitement des processus solutions. Soit η une variable aléatoire d'espérance nulle et de variance unité.

Définition 80 ***(Processus GARCH(p,q) fort)*** *Soit (η_t) une suite de variables aléatoires i.i.d. On dit que $(\varepsilon_t)_{t\in\mathbb{Z}}$ est un processus $GARCH(p,q)$ au sens fort s'il vérifie :*

$$\left\{ \begin{array}{c} \varepsilon_t = \sigma_t \eta_t, \\ \sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2, \end{array} \right. \tag{2.4}$$

où $\omega > 0$, $\alpha_i \geq 0, i = 1, ..., q$, $\beta_j \geq 0, j = 1, ..., p$.

Il est clair qu'un processus $GARCH$ fort tel que σ_t^2 est mesurable par rapport à la σ-algèbre $\mathcal{F}(\varepsilon_u, u < t)$ est un processus $GARCH$ au sens de définition (78). La réciproque n'est cependant pas vraie. Les processus $GARCH$ au sens de la définition (78) sont souvent qualifiés de semi-forts. En remplaçant ε_{t-i} par $\sigma_{t-i}\eta_{t-i}$ dans l'équation (2.1) on obtient :

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \sigma_{t-i}^2 \eta_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2, \tag{2.5}$$

qu'on peut écrire sous la forme

$$\sigma_t^2 = \omega + \sum_{i=1}^{\max\{p,q\}} (\alpha_i \eta_{t-i}^2 + \beta_i)\sigma_{t-i}^2. \tag{2.6}$$

2.2 Solutions stationnaires

Dans cette section, on s'intéresse à l'existence de solutions stationnaires (au sens strict et du second ordre) de modèle(2.4). On s'intéresse principalement aux solutions non anticipatives, c'est-à-dire aux processus $(\varepsilon_t)_{t\in\mathbb{Z}}$ tels que ε_t soit une fonction mesurable des variables $\eta_{t-i}, i > 0$. Pour de tels processus, σ_t est indépendante de $\sigma(\eta_t, \eta_{t+1}, ...)$, la σ-algèbre générée par les variables $\eta_{t+i}, i \geq 0$ et ε_t est indépendante de $\sigma(\eta_{t+1}, \eta_{t+2}, ...)$, la σ-algèbre générée par les variables $\eta_{t+i}, i \geq 0$. On montrera que de telles solutions sont également ergodiques. On considère d'abord le modèle $GARCH(1,1)$ qui peut être étudié d'une manière explicite que le cas général.

2.2.1 Etude d'un modèle GARCH(1,1)

Soit un modèle $GARCH(1,1)$ défini par :

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t, \\ \sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2, \end{cases} \tag{2.7}$$

avec $\omega \geq 0, \alpha_1 \geq 0, \beta_1 \geq 0$, et $(\eta_t)_{t\in\mathbb{Z}}$ est une suite de variables aléatoires i.i.d telle que $E(\eta_t) = 0$ et $var(\eta_t^2) = 1$.

Le modèle $GARCH(1,1)$ peut être écrit comme équation aux récurrences stochastique multivariée suivante

$$X_t = A_t X_{t-1} + B_t \ , \ \ t \in \mathbb{Z}, \tag{2.8}$$

où $(A_t, B_t)_{t\in\mathbb{Z}}$ est une suite iid, $(X_t)_{t\in\mathbb{Z}}$ est une suite de vecteurs aléatoires, $(A_t)_{t\in\mathbb{Z}}$ est une suite de matrices aléatoires iid et $(B_t)_{t\in\mathbb{Z}}$ est une suite de vecteurs aléatoires iid définis par

$$A_t = \begin{pmatrix} \alpha\eta_t^2 & \beta\eta_t^2 \\ \alpha & \beta \end{pmatrix}, X_t = \begin{pmatrix} \varepsilon_t^2 \\ \sigma_t^2 \end{pmatrix}, B_t = \begin{pmatrix} \omega\eta_t^2 \\ \omega \end{pmatrix}. \tag{2.9}$$

Le processus de volatilité $(\sigma_t^2)_{t\in\mathbb{Z}}$ satisfait l'équation aux récurrences stochastique univariée de type (2.8) où $A_t = \alpha\eta_{t-1}^2 + \beta$, $B_t = \omega$ est la suite $(A_t)_{t\in\mathbb{Z}}$ est iid. Rappelons que l'unique solution strictement stationnaire et ergodique de l'équation aux différences stochastique univariée de type (2.8) est donnée par le théorème suivant de Vervaat [16].

Théorème 81 *Supposons que $E\log^+|A_1| < \infty$ et $E\log^+|B_1| < \infty$. Si $\gamma = E\log|A_1| < 0$, alors pour tout $t \in \mathbb{Z}$ la série*

$$X_t = \sum_{k=0}^{\infty}\prod_{j=0}^{k-1} A_{t-j}B_{t-k}, \tag{2.10}$$

converge presque sûrement et le processus $(X_t)_{t\in\mathbb{Z}}$ est l'unique solution non-anticipative strictement stationnaire et ergodique de l'équation aux différence stochastique univariée (2.8). De plus, s'il existe $p \in [1,\infty[$ tel que $E|B_1|^p < \infty$ et $E|A_1|^p < 1$, alors la série (2.10) converge en moyenne d'ordre p pour tout $t \in \mathbb{Z}$ et les moments $E(X_t^m)$ sont determinés par

$$E(X_t^m) = \sum_{k=0}^{m}\binom{m}{k} E(A_1^k B_1^{m-k})E(X_t^k), m = 1, ..., [p], \tag{2.11}$$

où $[p]$ est la partie entière de p.

Nelson [15] a démontré une condition suffisante pour que le modèle $GARCH(1,1)$ défini par (2.7) admet une unique solution non-anticipative strictement stationnaire et ergodique en utilisant l'équation aux différences stochastiques univarié de type (2.8).

Théorème 82 ***(Stationnarité stricte du processus GARCH(1,1))*** *Soit un modèle $GARCH(1,1)$ défini par (2.7) avec $\omega > 0$. Si*

$$-\infty \leq E\log(\alpha_1\eta_t^2 + \beta_1) < 0, \tag{2.12}$$

alors la série

$$X_t = \left\{1 + \sum_{k=1}^{\infty} \prod_{j=1}^{k} (\alpha \eta_{t-1-j}^2 + \beta)\right\} \omega, \tag{2.13}$$

converge presque sûrement pour tout $t \in \mathbb{Z}$ *et le processus* $(\varepsilon_t)_{t\in\mathbb{Z}}$ *défini par* $\varepsilon = \sqrt{X_t}\eta_t$ *est l'unique solution non-anticipative strictement stationnaire et ergodique du modèle (2.7). Si* $\gamma \geq 0$ *et* $\omega > 0$*, il n'existe pas de solution strictement stationnaire.*

Soit $p \in [1, +\infty[$ *tel que*

$$E(\alpha \eta_t^2 + \beta)^p < 1, \tag{2.14}$$

alors $E(\sigma_t^{2p}) < \infty$ *et la série en (2.13) converge en moyenne d'ordre* p*. Les moments* $E(\sigma_t^{2m})$ *sont déterminés par l'équation*

$$E(\sigma_t^{2m}) = [1 - E(\alpha \eta_t^2 + \beta)^m]^{-1} \sum_{k=0}^{m-1} \binom{m}{k} E(\alpha_1 \eta_t^2 + \beta)^k \omega^{m-k} E(\sigma_t^{2k}), m = 1, ..., [p], \tag{2.15}$$

où $[p]$ *est la partie entière de* p*.*

Remarque 83 *1. Le coefficient* $\gamma = E \log(\alpha \eta_t^2 + \beta)$ *existe toujours dans* $[-\infty, +\infty]$ *car* $E \log^+(\alpha \eta_t^2 + \beta) \leq E(\alpha \eta_t^2 + \beta) = \alpha + \beta$*.*

2. *Dans le cas où* $\omega = 0$ *et* $\gamma < 0$*, D'aprés (2.13) l'unique solution strictement stationnaire de modèle (2.7) est* $\varepsilon_t = 0$*. Par conséquent, c'est l'origine d'imposer la condition* $\omega > 0$ *dans le modèle (2.7) .*

3. *C'est clair que la condition (2.12) dépend de la distribution de* η_t *et qu'elle n'est pas symétrique en* α *et* β*.*

4. *La condition (2.12) implique que* $\beta < 1$*. Inversement, si* $\alpha + \beta < 1$*, alors la condition (2.12) est vérifiée. En effet, par l''inégalité de Jensen on a*

$$E \log(\alpha \eta_t^2 + \beta) \leq \log E(\alpha \eta_t^2 + \beta) = \log(\alpha + \beta) < 0.$$

5. *Si la condition (2.12) est vérifiée, alors elle est vérifiée pour tout* (α_1, β_1) *tel que* $\alpha_1 \leq \alpha$ *et* $\beta_1 \leq \beta$*. En particulier, la stationnarité stricte du modèle* $GARCH(1,1)$ *implique*

la stationnarité stricte du modèle $ARCH(1)$ qui correspond à $\beta = 0$ dans le modèle $GARCH(1,1)$.

Démonstration. En utilisant de manière itérative la deuxième équation du modèle (2.7), on obtient

$$\begin{aligned}
\delta_t^2 &= \omega + (\alpha\eta_{t-1}^2 + \beta)\sigma_{t-1}^2 \\
&= \omega + \left\{1 + \sum_{k=1}^{n} (\alpha\eta_{t-k}^2 + \beta)\right\} + (\alpha\eta_{t-1}^2 + \beta)...(\alpha\eta_{t-n-1}^2 + \beta)\sigma_{t-n-1}^2 \\
&= X_t(n) + (\alpha\eta_{t-1}^2 + \beta)...(\alpha\eta_{t-n-1}^2 + \beta)\sigma_{t-n-1}^2. \qquad (2.16)
\end{aligned}$$

Le processus limite $X_t = \lim_{N\to\infty} X_t(n)$ existe dans $\overline{\mathbb{R}}^+$ parce que les termes de la somme sont positifs.

De plus, en faisant tendre n vers l'infini dans la relation $X_t(n) = \omega + (\alpha\eta_{t-1}^2 + \beta)X_{t-1}(n-1)$ on obtient

$$X_t = \omega + (\alpha\eta_{t-1}^2 + \beta)X_{t-1}.$$

On montre que X_t est finie presque sûrement si $\gamma < 0$. Supposons que $\gamma < 0$. Par la loi forte des grands nombres on a

$$\frac{1}{n}\sum_{i=1}^{n} \log(\alpha\eta_{t-i}^2 + \beta) \to^{p.s} E\log(\alpha\eta_t^2 + \beta) = \gamma < 0, n \to \infty.$$

Par la règle de Cauchy pour les séries à termes positifs, on a

$$\left|\prod_{j=1}^{n} (\alpha\eta_{t-j}^2 + \beta)\right|^{\frac{1}{n}} = \exp\left\{\frac{1}{n}\prod_{j=1}^{n} \log(\alpha\eta_{t-j}^2 + \beta)\right\} \to^{p.s} e^{\gamma} < 1, n \to \infty. \qquad (2.17)$$

Par conséquent, la série (2.13) converge p.s et le processus $(X_t)_{t\in\mathbb{Z}}$ à valeurs positives et strictement stationnaire et ergodique parce que X_t est fonction mesurable de processus ergodique $(\eta_t^2)_{t\in\mathbb{Z}}$. De même $(X_t)_{t\in\mathbb{Z}}$ est non anticipatif parce que X_t est une fonction mesurable de $\eta_{t-i}^2, i \geq 1$. Le processus $(\varepsilon_t)_{t\in\mathbb{Z}}$ défini par

$$\varepsilon_t = \sqrt{X_t}\eta_t = \left\{\omega + \sum_{k=1}^{\infty}\prod_{j=1}^{k}(\alpha\eta_{t-j}^2 + \beta)\omega\right\}^{1/2}\eta_t \tag{2.18}$$

est strictement stationnaire et ergodique car ε_t est fonction mesurable des processus ergodiques $(\eta_t^2)_{t\in\mathbb{Z}}$ et $(X_t)_{t\in\mathbb{Z}}$. De même, $(\varepsilon_t)_{t\in\mathbb{Z}}$ est non anticipatif parce que ε_t est fonction mesurable de $\eta_{t-i}, i \geq 0$. De plus, $(\varepsilon_t)_{t\in\mathbb{Z}}$ satisfait le modèle (2.7).

On montre maintenant que $(\varepsilon_t)_{t\in\mathbb{Z}}$ défini par (2.18) est l'unique solution de modèle (2.7). Soit $\overline{\varepsilon}_t = \sigma_t\eta_t$ une autre solution strictement stationnaire du modèle (2.7). D'aprés l'équation (2.16) on a :

$$\sigma_t^2 = X_t(n) + (\alpha\eta_{t-1}^2 + \beta)...(\alpha\eta_{t-n-1}^2 + \beta)\sigma_{t-n-1}^2.$$

Par conséquent

$$\sigma_t^2 - X_t = \{X_t(n) - X_t\} + (\alpha\eta_{t-1}^2 + \beta)...(\alpha\eta_{t-n-1}^2 + \beta)\sigma_{t-n-1}^2.$$

Le terme $\{X_t(n) - X_t\}$ tend vers 0 p.s lorsque $n \to \infty$. D'autres parts, puisque la série qui définit X_t converge presque sûrement, on a

$$\prod_{j=1}^{n}(\alpha\eta_{t-j}^2 + \beta) \to^{p.s} 0, n \to \infty.$$

De plus, par stationnarité, la loi de σ_{t-n-1}^2 est indépendante de n. Par conséquent,

$$\prod_{j=1}^{n-1}(\alpha\eta_{t-j}^2 + \beta)\sigma_{t-n-1}^2 \to^{p} 0, n \to \infty.$$

On a prouvé que $\sigma_t^2 - X_t \to^P 0$, $n \to \infty$. Puisque $\sigma_t^2 - X_t$ ne dépend pas de n, on a nécessairement $X_t = \sigma_t^2$ pour tout $t \in \mathbb{Z}$ presque sûrement .

Si $\gamma > 0$, d'aprés (2.17) et la critère de Cauchy

$$\sum_{k=1}^{n}\prod_{j=1}^{k}(\alpha\eta_{t-j}^2 + \beta) \to^{p.s} \infty, n \to \infty.$$

Donc, si $\omega > 0, X_t = \infty$ p.s. D'aprés (2.16), il est clair que $\sigma_t^2 = +\infty$, p.s. Il s'ensuit qu' il n'existe pas de solution finie presque sûrement de modèle (2.7). Dans le cas où $\gamma = 0$, on montre le résultat par absurde. Supposons qu'il existe une solution strictement stationnaire $(\varepsilon_t)_{t\in\mathbb{Z}}$ du modèle (2.7). On a, pour chaque $n > 0$, on a

$$\sigma_0^2 \geq \omega\{1 + \sum_{i=1}^{n}\prod_{j=1}^{k}(\alpha\eta_{-j}^2 + \beta)\}$$

et on déduit que $\prod_{j=1}^{n}(\alpha\eta_{-j}^2 + \beta)\omega \rightarrow^{p.s} 0$, $n \rightarrow \infty$, de manière équivalente :

$$\sum_{i=1}^{n}\log(\alpha\eta_i^2 + \beta) + \log\omega \rightarrow^{p.s} -\infty,\ n \rightarrow \infty. \tag{2.19}$$

Le théorème de Chung-Fuchs stipule que si $X_1, X_2, ...$ est une suite iid telle que $E(X_1) = 0$ et $E|X_1| > 0$, alors $\limsup_n \sum_{i=1}^{n} X_i = \infty$ et $\liminf_n \sum_{i=1}^{n} X_i = -\infty$. Par conséquent $\limsup_n \log(\alpha\eta_i^2 + \beta) = \infty$, ceci contredit (2.19).

Pour une variable aléatoire X , soit $||X||_p = (E|X|^p)^{\frac{1}{p}}$. Par l'inégalité de Jensen, la condition $E(\alpha\eta_t^2+\beta)^p < 1$ implique que la condition (2.12) est vérifiée. De plus, par l'inégalité de Minkowski on a

$$||\sigma_t^2||_p \leq \omega + \sum_{k=1}^{\infty}\left\|\prod_{j=1}^{k}(\alpha\eta_{t-j}^2 + \beta)\omega\right\|_p = \omega\left\{1 + \sum_{k=1}^{\infty}||\alpha\eta_t^2 + \beta||_p^k\right\} < \infty. \tag{2.20}$$

Par conséquent, $E(\sigma_t^{2p}) < \infty$ la série en (2.13) converge en moyenne d'ordre p. De l'équation aux différences stochastiques univariée de type (2.8) , on déduit que

$$\begin{aligned}
E(X_t^m) &= \sum_{k=0}^{m}\binom{m}{k}E(A_t^k B_t^{m-k})E(X_t^k) \\
&= \sum_{k=0}^{m}\binom{m}{k}E[(\alpha\eta_t^2 + \beta)^k]\omega^{m-k}E(X_t^k) \\
&= \sum_{k=0}^{m-1}\binom{m}{k}E[(\alpha\eta_t^2 + \beta)^k]\omega^{m-k}E(X_t^k) + E[(\alpha\eta_t^2 + \beta)^m]E(X_t^k).
\end{aligned}$$

Par conséquent (2.15) est vérifiée pour chaque $m = 1, ..., [p]$. ■

Le résultat suivant de Franq et Zakoïan [12] montre la non-stationnarité du processus $GARCH(1,1)$.

Corollaire 84 *Soit un modèle $GARCH(1,1)$ défini dans (2.7) pour $t \geq 1$. Si $E(\log(\alpha_1\eta_0^2 + \beta_1)) > 0$, alors*

$$\lim_{t\to\infty} \sigma_t^2 = +\infty \ p.s. \ (n \to \infty). \tag{2.21}$$

Si, en plus, $E(|\log(\eta_t^2)|) < \infty$, alors

$$\lim_{t\to\infty} \varepsilon_t^2 = +\infty \ p.s. \ (n \to \infty). \tag{2.22}$$

Démonstration. On a :

$$\sigma_t^2 \geq \alpha_0 \sum_{m=1}^{t} \prod_{j=1}^{m-1} (\alpha_1\eta_{t-j}^2 + \beta_1) \geq \alpha_0(\alpha_1\eta_{t-1}^2 + \beta_1)...(\alpha_1\eta_1^2 + \beta_1),$$

alors

$$\lim_{t\to\infty} \inf \frac{1}{t} \log \sigma_t^2 \geq \lim_{t\to\infty} \inf \frac{1}{t} \sum_{j=1}^{t-1} \log(\alpha_1\eta_{t-j}^2 + \beta_1) = E(\log(\alpha_1\eta_0^2 + \beta_1)).$$

Par conséquent $\log \sigma_t^2 \to +\infty$ et $\sigma_t^2 \to +\infty$, p.s.

De la même manière, on a :

$$\begin{aligned} \lim_{t\to\infty} \inf \frac{1}{t} \log \varepsilon_t^2 &= \lim_{t\to\infty} \inf \frac{1}{t} (\log \sigma_t^2 + \log \eta_t^2) \\ &\geq E(\log(\alpha_1\eta_0^2 + \beta_1)) + \lim_{t\to\infty} \inf \frac{1}{t} \log \eta_t^2 = E(\log(\alpha_1\eta_0^2 + \beta_1)), \end{aligned}$$

alors $\log \varepsilon_t^2 \to +\infty$ et $\varepsilon_t^2 \to +\infty$, p.s.

Bollerslev [1] a établit une condition nécessaire et suffisante pour l'existence d'une solution stationnaire au second ordre du modèle $GARCH(1,1)$. ■

Théorème 85 ***(Stationnarité au second-ordre du processus*** $GARCH(1,1)$*)*

Lorsque $\omega > 0$*, le modèle* $GARCH(1,1)$ *défini par (2.7) admet une unique solution non-anticipative stationnaire au second ordre* $(\varepsilon_t)_{t\in\mathbb{Z}}$ *donnée par (2.18), avec* $E(\varepsilon_t) = 0, var(\varepsilon_t) = \alpha_0(1-\alpha_1-\beta_1)^{-1}$ *et* $cov(\varepsilon_t, \varepsilon_s) = 0$ *pour* $t \neq s$*, i.e.,* $(\varepsilon_t)_{t\in\mathbb{Z}}$ *est un bruit blanc si et seulement si*

$$\alpha_1 + \beta_1 < 1. \tag{2.23}$$

Démonstration. *Si* $(\varepsilon_t)_{t\in\mathbb{Z}}$ *est un processus* $GARCH(1,1)$ *au sens de définition **(78)** qui est stationnaire au second ordre et non anticipatif, alors on a*

$$E(\varepsilon_t^2) = E(E(\varepsilon_t^2/\mathcal{F}_{t-1})) = E(\sigma_t^2) = \omega + (\alpha+\beta)E(\varepsilon_{t-1}^2),$$

c'est-à-dire

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

Par conséquent, il faut que $1-\alpha-\beta > 0$.

Inversement, supposons que $\alpha+\beta < 1$*. D'aprés la remarque, la condition de stationnarité stricte (2.12) est vérifiée. Il suffit donc de montrer que la solution non-anticipative strictement stationnaire définie en **(2.18)** admet une variance finie. La variable* X_t *définie en (2.13) étant une limite croissante de variables aléatoires positives, d'aprés le théorème de Beppo Levi, on peut invertir l'éspérance et la somme infinie et on obtient*

$$\begin{aligned} E(\varepsilon_t^2) &= E(\sigma_t^2)E(\eta_t^2) = E(X_t) = \left\{1 + \sum_{k=1}^{\infty} E(\prod_{j=1}^{k}(\alpha\eta_{t-j}^2+\beta)\right\}\omega \\ &= \left\{1 + \sum_{k=1}^{\infty}\{E(\alpha\eta_t^2+\beta)\}^k\right\}\omega \\ &= \left\{1 + \sum_{k=1}^{\infty}(\alpha+\beta)^k\right\}\omega = \frac{\omega}{1-(\alpha+\beta)}. \end{aligned}$$

Ceci montre que la solution non-anticipative strictement stationnaire et ergodique définie en (2.18) est stationnaire au second-ordre. De plus, cette solution est un bruit blanc parce *que* $E(\varepsilon_t) = E(E(\varepsilon_t/\mathcal{F}_{t-1})) = 0$ *et pour tout* $h > 0$,

$$cov(\varepsilon_t, \varepsilon_{t-h}) = E(\varepsilon_t\varepsilon_{t-h}) = E(E(\varepsilon_t\varepsilon_{t-h}/\mathcal{F}_{t-1})) = E(\varepsilon_{t-h}E(\varepsilon_t/\mathcal{F}_{t-1})) = 0.$$

Il reste à montrer que $(\varepsilon_t)_{t\in\mathbb{Z}}$ *défini en (2.18) est l'unique solution non-anticipative stationnaire au second-ordre du modèle (2.7). Soit* $\widetilde{\varepsilon}_t = \sqrt{\widetilde{X}_t}\eta_t$ *une autre solution non-anticipative stationnaire au second ordre du modèle (2.7), alors*

$$\left|X_t - \widetilde{X}_t\right| = \prod_{j=1}^{n}(\alpha\eta_{t-j}^2 + \beta)\left|X_{t-n-1} - \widetilde{X}_{t-n-1}\right|$$

et par conséquent,

$$\begin{aligned} E\left|X_t - \widetilde{X}_t\right| &= E\left\{\prod_{j=1}^{n}(\alpha\eta_{t-j}^2 + \beta)\right\} E|X_{t-n-1} - \widetilde{X}_{t-n-1}| \\ &= (\alpha+\beta)^n E|X_{t-n-1} - \widetilde{X}_{t-n-1}|. \end{aligned}$$

Pour tout $t \in \mathbb{Z}$ *et tout* $n > 0, E|X_{t-n-1} - \widetilde{X}_{t-n-1}| \leq E|X_{t-n-1}| + E|\widetilde{X}_{t-n-1}|$ *où* $E|X_{t-n-1}|$ *et* $E|\widetilde{X}_{t-n-1}|$ *sont finies et indépendantes de* n *par la stationnarité au second-ordre de* $(X_t)_{t\in\mathbb{Z}}$ *et* $(\widetilde{X}_t)_{t\in\mathbb{Z}}$*. Maintenant,* $(\alpha+\beta)^n$ *tend vers* 0 *quand* $n \to \infty$ *. On déduit que pour tout* $t \in \mathbb{Z}$, $E|X_t - \widetilde{X}_t| = 0$ *et donc* $X_t = \widetilde{X}_t$ *pour tout* $t \in \mathbb{Z}$ *p.s.* ∎

2.2.2 Modèle GARCH(p,q)

Dans le cas général, le modèle $GARCH(p,q)$ défini par (2.4) peut être écrit comme équation aux récurrence stochastique multivariée (2.8) où

$$\begin{aligned}
A_t &= \begin{pmatrix}
\alpha_1\eta_t^2 & \alpha_2\eta_t^2 & ... & \alpha_{q-1}\eta_t^2 & \alpha_q\eta_t^2 & \beta_1\eta_t^2 & \beta_2\eta_t^2 & ... & \beta_{p-1}\eta_t^2 & \beta_p\eta_t^2 \\
1 & 0 & ... & 0 & 0 & 0 & 0 & ... & 0 & 0 \\
. & . & ... & . & . & . & . & . & . & . \\
. & . & ... & . & . & . & . & ... & . & . \\
. & . & ... & . & . & . & . & ... & . & . \\
0 & 0 & ... & 1 & 0 & 0 & 0 & ... & 0 & 0 \\
\alpha_1 & \alpha_2 & ... & \alpha_{q-1} & \alpha_q & \beta_1 & \beta_2 & ... & \beta_{p-1} & \beta_p \\
0 & 0 & ... & 0 & 0 & 1 & 0 & ... & 0 & 0 \\
0 & 0 & ... & 0 & 0 & 0 & 1 & ... & 0 & 0 \\
. & . & ... & . & . & . & . & ... & . & . \\
. & . & ... & . & . & . & . & ... & . & . \\
. & . & ... & . & . & . & . & ... & . & . \\
0 & 0 & ... & 0 & 0 & 0 & 0 & ... & 1 & 0
\end{pmatrix}, && (2.24) \\
X_t &= (\varepsilon_t^2, ..., \varepsilon_{t-q+1}^2, \sigma_t^2, ..., \sigma_{t-p+1}^2)', && (2.25) \\
B_t &= (\alpha_0\eta_t^2, 0, ..., \alpha_0, 0, ..., 0)', && (2.26)
\end{aligned}$$

tels que A_t est une matrice aléatoire à valeurs dans $M^+(p+q)$, ensemble des matrices réelles carrées de tailles $(p+q)(p+q)$ à coefficients positifs, X_t et B_t sont des vecteurs aléatoires à valeurs dans $(\mathbb{R}^+)^{p+q}$. La suite $(A_t, B_t)_{t\in\mathbb{Z}}$ est iid à valeurs dans $M^+(p+q)\times(\mathbb{R}^+)^{p+q}$.

Rappelons d'abord la définition de l'exposant de Lyapunov. Soit $||.||$ une norme quelconque sur $\mathbb{R}^d$ et on définit l'opérateur de norme sur l'ensemble $M(d)$ des matrices réelles de tailles $d \times d$ par

$$||A|| = \sup\{||Ax||/||x|| : x \in \mathbb{R}^d, x \neq 0\}.$$

L'exposant de Lyapunov associé à la suite de matrices aléatoires $(A_t)_{t\in\mathbb{Z}}$ iid à valeurs dans $M(d)$ est défini lorsque $E\log^+||A_t|| < \infty$ par

$$\gamma = \inf_{t\in\mathbb{N}^*} \frac{1}{t} E(\log||A_t A_{t-1}...A_1||).$$

Par exemple, on a $\gamma \leq E \log ||A_t||$, avec égalité lorsque $d = 1$. Par le théorème ergodique sous-additif de Kingman (voir Kingman (1973), théorème 6), on a

$$\gamma = \lim_{t \to \infty} \frac{1}{t} \log ||A_t A_{t-1} ... A_1|| \quad \text{p.s.} \tag{2.27}$$

Puisque toutes les normes sont équivalentes sur $\mathbb{R}^d$, l'exposant de Lyapunov γ ne dépend pas de la norme choisie. Dans le cas d'un modèle $ARCH(q)$, X_t ne contient que ε_t^2 et ses $q-1$ premières valeurs passées et A_t se limite au bloc supérieur gauche de la matrice définie en (2.24). En itérant (2.8) on obtient

$$X_t = B_t + \sum_{k=1}^{\infty} \prod_{j=0}^{k-1} A_{t-j} B_{t-k}. \tag{2.28}$$

Pour vu que la série converge presque sûrement. Dans cequi suit on donne des conditions suffisantes assurant la convergence presque sûrement de la série en (2.28). Une condition suffisante pour que

$$B_t + \sum_{k=1}^{\infty} \prod_{j=0}^{k-1} A_{t-j} B_{t-k} > 0 \quad \text{p.s.} \tag{2.29}$$

dans le sens que toutes les composantes sont strictement positives presque sûrement est que

$$\omega > 0, \alpha_i \geq 0, i = 1, ..., q, j = 1, ..., p. \tag{2.30}$$

Cette condition n'est cependant pas toujours nécessaire comme on le verra plus loin.

Rappelons que pour toute matrice réelle A de taille $(p+q) \times (p+q)$, le rayon spectral qu'on note par $\rho(A)$ est le plus grand module des ses valeurs propres et un résultat d'algèbre dit que

$$\lim_{t \to \infty} \frac{1}{t} \log ||A_t|| = \log \rho(A). \tag{2.31}$$

La propriété (2.31) s'étend aux matrices aléatoires à travers le résultat suivant (voir Bougerol-Lacroix [2]).

Théorème 86 *Soit $(A_t)_{t\in\mathbb{Z}}$ une suite de matrices aléatoires strictement stationnaire et ergodique à valeurs dans $M(d)$, alors*

$$\lim_{t\to\infty}\frac{1}{t}E(\log||A_tA_{t-1}...A_1||)=\gamma=\inf_{t\in\mathbb{N}^*}\frac{1}{t}E(\log||A_tA_{t-1}...A_1||). \quad (2.32)$$

Le lemme général suivant est trés utile pour l'étude du produit de matrices aléatoires.

Lemme 87 *Soit $(A_t)_{t\in\mathbb{Z}}$ une suite de matrics aléatoires iid à valeurs dans $M(d)$ telle que $E\log^+||A_t||$ est finie. Si*

$$\lim_{t\to\infty}||A_0A_{-1}...A_{-t}||=0. \quad (2.33)$$

alors l'exposant de Lyapunov associé à la suite iid $(A_t)_{t\in\mathbb{Z}}$ est strictement négatif.

Le théorème suivant de Franq et Zakoian [11] donne une condition nécessaire et suffisante pour que le modèle $GARCH(p,q)$ admette une unique solution strictement stationnaire, non anticipative et ergodique.

Théorème 88 ***(Stationnarité stricte du processus** $GARCH(p,q)$)*

Le modèle $GARCH(p,q)$ définie par (2.4) a une solution strictement stationnaire si et seulement si l'exposant de Lyapunov associé à la suite iid $(A_t)_{t\in\mathbb{Z}}$ définie par (2.24) est strictement négatif. De plus, cette solution est unique, non-anticipative et ergodique.

Démonstration. *Nous allons utiliser la norme définie par $||A||=\sum|a_{ij}|$. Par commodité, La norme sera notée de manière identique quelle que soit la dimension de A. Avec cette convention, la norme est clairement multiplicative $||AB||\leq||A||\;||B||$ pour toutes les matrices A et B telles que AB existe. Puisque les variables η_t sont de variance finie, tous les termes de la matrice A_t définie en (2.24) sont intégrable. Par conséquent, on a donc*

$$E\log^+||A_t||\leq E||A_t||<\infty.$$

Supposons $\gamma<0$, alors l'égalité (2.27) implique que la série

$$\widetilde{X}_t=B_t+\sum_{k=0}^{\infty}\prod_{j=0}^{k}A_{t-j}B_{t-k-1}$$

converge presque sûrement pour tout $t \in \mathbb{Z}$. En effet, en utilisant la multiplicativité de la norme,

$$||\widetilde{X}_t|| \leq ||B_t|| + \sum_{k=0}^{\infty} \left\| \prod_{j=0}^{k} A_{t-j} \right\| ||B_{t-k-1}||. \tag{2.34}$$

Puisque $E|\log ||B_{t-k-1}||| \leq |\log \omega| + E\log^+ ||B_{t-k-1}|| \leq |\log \omega| + E||B_{t-k-1}|| < \infty, k^{-1}\log ||B_{t-k-1}|| \to 0$ *p.s.,*$k \to \infty$ *et*

$$||\prod_{j=0}^{k} A_{t-j}||^{\frac{1}{k}} ||B_{t-k-1}||^{\frac{1}{k}} = \exp\{\frac{1}{k}\log ||A_t...A_{t-k}|| + \frac{1}{k}\log ||B_{t-k-1}||\} \to^{p.s} e^{\gamma} < 1, k \to \infty. \tag{2.35}$$

Par conséquent, par la règle de Cauchy, $\widetilde{X}_t$ *est bien défini dans* $(\mathbb{R}^{*+})^{p+q}$. *Soit* $\widetilde{X}_{q+1,t}$ *désigne le* $q+1$*-ème composante de* $\widetilde{X}_t$, *mettons* $\varepsilon_t = \sqrt{\widetilde{X}_{q+1,t}}\eta_t$ *on définit une solution strictement stationnaire du modèle (2.4). D'aprés (2.28),* ε_t *peut être exprimé comme une fonction mesurable de* $\eta_t, \eta_{t-1}, ...$. *La solution est donc non-anticipative et ergodique puisque* $(\eta_t)_{t\in\mathbb{Z}}$ *est ergodique. L'unicité se démontre par le même raisonnement que dans le cas* $p = q = 1$. *Supposons qu'il existe une autre solution strictement stationnaire du modèle (2.4), ou de manière équivalente une autre solution strictement stationnaire positive* $(X_t^*)_{t\in\mathbb{Z}}$ *de (2.4). Pour tout* $n \geq 0$ *,*

$$\begin{aligned} X_t^* &= B_t + B_t + \sum_{k=0}^{n}\prod_{j=0}^{k} A_{t-j}B_{t-k-1} + A_t...A_{t-n}X_{t-n-1}^* \\ &= \widetilde{X}_t(n) + A_t...A_{t-n}X_{t-n-1}^*, \end{aligned}$$

alors

$$||X_t^* - \widetilde{X}_t|| \leq ||\widetilde{X}_t(n) - \widetilde{X}_t|| + ||A_t...A_{t-n}||||X_{t-n-1}^*||.$$

Puisque $\widetilde{X}_t(n) \to^{p.s} \widetilde{X}_t, n \to \infty$ *, on a* $||\widetilde{X}_t(n) - \widetilde{X}_t|| \to^{p.s} 0, n \to \infty$. *De plus, puisque la série (2.34) converge p.s., on a* $||A_t...A_{t-n}|| \to^{p.s} 0, n \to \infty$. *Maintenant, la distribution de* X_{t-n-1}^* *est indépendante de* n *par stationnarité stricte de processus* $(X_t^*)_{t\in\mathbb{Z}}$. *Par conséquent,* $||A_t...A_{t-n}||||X_{t-n-1}^*|| \to^p 0, n \to \infty$. *On a montré donc que* $X_t^* - \widetilde{X}_t \to^p 0, n \to \infty$. *Le terme* $X_t^* - \widetilde{X}_t$ *ne dépend pas de* n, *alors on a nécessairement* $X_t^* = \widetilde{X}_t$ *p.s pour tout* $t \in \mathbb{Z}$.

Ensuite on montre la condition nécessaire de théorème précédent. D'aprés le lemme 87, il suffit d'établir (2.33). On montre que pour tout $1 \leq i \leq p+q$

$$\lim_{t\to\infty} A_0 ... A_{-t} e_i = 0 \ p.s., \tag{2.36}$$

où e_i est le i^{eme} élément de base de $\mathbb{R}^{p+q}$. Soit $(\varepsilon_t)_{t\in\mathbb{Z}}$ une solution strictement stationnaire non anticipative et ergodique du modèle (2.4) et soit $(X_t)_{t\in\mathbb{Z}}$ un processus aléatoire défini par l'équation aux récurrences stochastique. Pour $t > 0$

$$\begin{aligned} X_0 &= B_0 + A_0 X_{-1} \\ &= B_0 + \sum_{k=0}^{t-1} \prod_{j=0}^{k} A_{-j} B_{-k-1} + A_0 ... A_{-t} X_{-t-1} \\ &\geq \sum_{k=0}^{t-1} \prod_{j=0}^{k} A_{-j} B_{-k-1}, \end{aligned}$$

parce que les coefficients de matrices A_t, B_0 et X_t sont non négative. Il s'ensuit que la série $\sum_{k=0}^{t-1} \prod_{j=0}^{k} A_{-k} B_{-k-1}$ converge p.s et donc $A_0 ... A_{-k} B_{-k-1} \to^P 0$, $k \to \infty$. Puisque $B_{-k-1} = \omega \eta_{-k-1}^2 e_1 + \omega e_{q+1}$, il suit que $A_0 ... A_{-k} B_{-k-1}$ peut être décomposé en deux termes positifs, et on a

$$\lim_{k\to\infty} A_0 ... A_{-k} \omega \eta_{-k-1}^2 e_1 = 0 \ p.s, \ \lim_{k\to\infty} A_0 ... A_{-k} \omega e_{q+1} = 0 \ p.s. \tag{2.37}$$

Puisque $\omega > 0$, la relation(2.36) est vérifiée pour $i = q+1$. En utilisant la relation

$$A_{-k} e_{q+i} = \beta_i \eta_{-k}^2 e_1 + \beta_i e_{q+1} + e_{q+i+1} \quad , i = 1, ..., p, \tag{2.38}$$

avec par convention $e_{p+q+1} = 0$. Pour $i = 1$, la relation (2.38) on obtient

$$0 = \lim_{t\to\infty} A_0 ... A_{-k} e_{q+1} \geq \lim_{k\to\infty} A_0 ... A_{-k+1} e_{q+2} \geq 0, \tag{2.39}$$

donc (2.36) est vérifiée pour $i = q+2$ et par récurrence, pour $i = q+j$, $j = 1, ..., p$ en utilisant (2.38). De plus, on remarque que $A_{-k} e_q = \alpha_q \eta_{-k}^2 e_1 + \alpha_q e_{q+1}$, ce qui nous permet de voir que, d'aprés (2.37), que (2.36) est vérifiée pour $i = q$. On conclut pour les autres valeurs de i en utilisant $A_{-k} e_i = \alpha_i \eta_{-k}^2 e_1 + \alpha_i e_{q+1} + e_{i+1}$, $i = 1, ..., q-1$ et une récurrence ascendante. ■

Remarque 89 *1. Notons que dans le cas d'un modèle $GARCH(p,q)$ on a utilisé l'équation aux récurrence stochastique multivarié (2.8) à coefficients $((A_t, B_t))_{t\in\mathbb{Z}}$ iid à valeurs dans $M^+(p+q) \times (\mathbb{R}^+)^{p+q}$ où la matrice A_t et les vecteurs X_t et B_t sont définies par (2.24), (2.25), (2.26), respectivement. Bougerol et Picard* [4] *ont montré la stationnarité stricte du processus $GARCH(p,q)$ en utilisant l'équation aux récurrences stochastique multivariée (2.18)à coefficients $((A_t, B_t))_{t\in\mathbb{Z}}$ iid à valeurs dans $M^+(p+q-1) \times (\mathbb{R}^+)^{p+q}$ où la matrice A_t et les vecteurs X_t et B_t sont définies par*

$$\begin{aligned}
A_t &= \begin{pmatrix} \delta_t & \beta_p & \alpha & \alpha_q \\ I_{p-1} & 0 & 0 & 0 \\ \xi_t & 0 & 0 & 0 \\ 0 & 0 & I_{q-2} & 0 \end{pmatrix} \in M^+(p+q-1), \\
X_t &= (\sigma_t^2, ..., \sigma_{t-p+1}^2, \varepsilon_{t-1}^2, ..., \varepsilon_{t-q+1}^2)' \in (\mathbb{R}^+)^{p+q-1}, \\
B_t &= (\omega, 0, ..., 0)' \in (\mathbb{R}^+)^{p+q-1},
\end{aligned}$$

avec

$$\begin{aligned}
\delta_t &= (\alpha_1 \eta_{t-1}^2 + \beta_1, \beta_2, ..., \beta_{p-1})' \in (\mathbb{R}^+)^{p-1}, \\
\xi_t &= (\eta_{t-1}^2, 0, ..., 0)' \in (\mathbb{R}^+)^{p-1}, \\
\alpha &= (\alpha_2, ..., \alpha_{q-1})' \in (\mathbb{R}^+)^{q-2},
\end{aligned}$$

I_{p-1} et I_{q-2} sont des matrices identité de tailles $p-1$ et $q-2$, respectivement. Cette représentation de Bougerol et Picard n'est définie que pour $p \geq 1$ et $q \geq 2$.

2. Francq et Zakoïan [11] *ont montré la stationnarité stricte du processus $GARCH(p,q)$ en utilisant la représentation (2.6) pour le processus de volatilité $(\sigma_t^2)_{t\in\mathbb{Z}}$ et l'équation aux récurrence stochastique multivariée (2.8) à coefficients $((A_t, B_t))_{t\in\mathbb{Z}}$ iid à valeurs dans $M^+(r) \times (\mathbb{R}^+)^r$ où la matrice A_t et les vecteurs X_t et B_t sont définies par*

$$\begin{aligned}
A_t &= \begin{pmatrix} \alpha_1 \eta_{t-1}^2 + \beta_1 & ... & \alpha_r \eta_{t-r}^2 + \beta_r \\ I_{r-1} & ... & 0 \end{pmatrix} \in M^+(p+q-1), \\
X_t &= (\sigma_t^2, ..., \sigma_{t-r+1}^2)' \in (\mathbb{R}^+)^r, \\
B_t &= (\omega, 0, ..., 0)' \in (\mathbb{R}^+)^r.
\end{aligned}$$

3. *Si un modèle $GARCH(p,q)$ admet une solution strictement stationnaire, non anticipative et ergodique, alors tout modèle $GARCH(p,q)$ obtenu en remplaçant les α_i et les β_j par des coefficients plus petits admet également une solution strictement stationnaire, non anticipative et ergodique. En effet, l'exposant de Lyapunov associé à la suite $(A_t)_{t\in\mathbb{Z}}$ iid du modèle ainsi défini sera nécessairement inférieur à celui associé à la suite $(A_t)_{t\in\mathbb{Z}}$ iid du modèle initial car, avec la norme utilisé, $0 \leq A \leq B$ implique $||A|| \leq ||B||$. En particulier, la stationnarité stricte du modèle $GARCH(p,q)$ implique celle du modèle $ARCH(q)$ obtenu en annulant les coefficients β_j.*

Le résultat suivant de Bougerol et Picard [4] fournit une condition nécessaire de stationnarité stricte du processus $GARCH(p,q)$.

Corollaire 90 *Si l'exposant de Lyapunov γ associé à la suite iid $(A_t)_{t\in\mathbb{Z}}$ défini dans (2.24) est strictement négatif, alors les propriétés suivantes sont équivalentes :*

(a) $\sum_{i=1}^{p} \beta_i < 1$.

(b) *Les racines de*

$$1 - \beta_1 z - ... - \beta_p z^p = 0,$$

sont en dehors du disque unité.

(c) $\rho(B_0) < 1$, *avec $\rho(B_0)$ est le rayon spectral de la sous matrice B de la matrice A_t définie par*

$$\begin{pmatrix} \beta_1 & \beta_2 & ... & \beta_p \\ 1 & 0 & ... & 0 \\ 0 & 1 & ... & 0 \\ . & . & ... & .. \\ 0 & . & ... & 0 \end{pmatrix}.$$

Démonstration. Comme les éléments de la matrice a_t sont non négatifs, il est clair que l'exposant de Lyapunov associé à la suite $(A_t)_{t\in\mathbb{Z}}$ est plus grand que l'exposant de Lyapunov associé à la suite des matrices constantes obtenues en remplaçant par 0 les composantes des q premières lignes et les q premières colonne de la matrice α_t. La matrice obtenue a les mêmes valeurs propres non nulles que la matrice B et le même rayon spectral que B. Par la propriété (2.31) et le théorème (88), si $A_t = A$ pour tout $t \in \mathbb{Z}$, alors $\gamma = \log \rho(A)$. Par conséquent, $\log \rho(B) \leq \gamma$ et $\log \rho(B) < 0$ si $\gamma < 0$. On peut vérifier par récurrence sur p en calculant le determinant par rapport à la dernière colonne que pour $\lambda \neq 0$,

$$\det(\lambda I_p - B) = \lambda^p - \lambda^{p-1}\beta_1 - ... - \lambda\beta_{p-1} - \beta_p = \lambda^p \mathcal{B}(\frac{1}{\lambda}), \tag{2.40}$$

où $\mathcal{B}(z) = 1 - \beta_1 z - ... - \beta_p z^p$. On en déduit que entre $(b) \Leftrightarrow (c)$. On montre maintenant que $(a) \Leftrightarrow (b)$. On a $\mathcal{B}(0) = 1$ et $\mathcal{B}(1) = 1 - \sum_{i=1}^p \beta_i$. Donc, si $\sum_{i=1}^p \beta_i \geq 1$ alors $\mathcal{B}(1) \leq 0$ et par continuité de $\mathcal{B}(z)$ il existe $z_0 \in]0, 1]$ tel que $\mathcal{B}(z_0) = 0$. Donc on a montré $(b) \Rightarrow (a)$.

Inversement, si $\sum_{i=1}^p \beta_i \leq 1$ et si $\mathcal{B}(z_0) = 0$ pour un certain z_0 de module inférieur ou égal à 1, alors

$$1 = \sum_{i=1}^p \beta_i z_0^i = |\sum_{i=1}^p \beta_i z_0^i| \leq \sum_{i=1}^p \beta_i |z_0|^i \leq \sum_{i=1}^p \beta_i < 1,$$

ce qui est impossible. On a montré alors $(a) \Rightarrow (b)$.

Maintenant on peut donner la condition nécessaire et suffisante de stationnarité stricte pour le processus $GARCH(1,1)$ en utilisant l'équation aux récurrences stochastique multivariée (2.8). Dans ce cas, la matrice A_t définie en (2.24) se réduit à la matrice A_t définie en (2.9), le vecteur X_t défini en (2.25) se réduit au vecteur X_t défini en (2.9) et le vecteur B_t défini en (2.26) se réduit au vecteur B_t défini en (2.9). Donc on a

$$A_t A_{t-1} ... A_1 = \prod_{k=1}^{t-1} (\alpha_1 \eta_{t-k}^2 + \beta) A_t,$$

et

$$\log ||A_t A_{t-1} ... A_1|| = \sum_{k=1}^{t-1} \log(\alpha_1 \eta_{t-k}^2 + \beta) + \log ||A_t||.$$

Par la propriété (2.27) et la loi forte des grands nombres, on a $\gamma = E \log(\alpha \eta_t^2 + \beta)$. Par conséquent, la condition nécessaire et suffisante de la stationnarité stricte de théorème (2.4) se réduit à $E \log(\alpha_1 \eta_t^2 + \beta) < 0$, comme on l'a déja vu dans le théorème 2.2.

Le théorème suivant de Francq et Zakoïan [11] donne les moments d'un certain ordre de la solution strictement stationnaire, non anticipative et ergodique du modèle $GARCH(p,q)$.

■

Théorème 91 *Supposons que l'exposant de Lyapunov* γ *associé à la suite* $(A_t)_{t\in\mathbb{Z}}$ *définie en (2.9) est strictement négatif. Soit* $(\varepsilon_t)_{t\in\mathbb{Z}}$ *la solution strictement stationnaire, non anticipative et ergodique du modèle* $GARCH(p,q)$ *défini par (2.4), alors il existe* $s>0$ *tel que*

$$E(\sigma_t^{2s})<\infty \text{ et } E(\varepsilon_t^{2s})<\infty.$$

Démonstration. Supposons que $\gamma<0$. Par définition, $\gamma=\inf_k \frac{1}{k}E(\log||A_kA_{k-1}...A_1||)$, alors il existe $k_0\geq 1$ tel que

$$E(\log||A_{k_0}A_{k_0-1}...A_1||)<0.$$

De plus, en utilisant la norme $||A||=\sum_{i,j}|A_{ij}|$, la positivité des éléments des A_t, l'indépendance et l'équidistribution des A_t, on obtient

$$\begin{aligned}
E||A_{k_0}A_{k_0-1}...A_1|| &= ||E(A_{k_0}A_{k_0-1}...A_1)|| \\
&= ||(EA_1)^{k_0}|| \\
&\leq (E||A_1||)^{k_0}<\infty.
\end{aligned}$$

Rappelons que si X est une variable aléatoire positive $E(X^r)<\infty$ pour un certain $r>0$ et $E\log X<0$, alors il existe $s>0$ tel que $E(X^s)<1$. Ce dernier résultat entraîne donc l'existence d'un $s\in]0,1[$ tel que

$$\delta=E||A_{k_0}A_{k_0-1}...A_1||^s<1.$$

La solution strictement stationnaire de l'équation aux récurrences stochastique multivariée

(2.8) est donnée par (2.28) et satisfait

$$\begin{aligned}
E||X_t||^s &\leq E||B_t||^s + \sum_{k=1}^{\infty} E||A_t...A_{t-k+1}||^s E||B_{t-k}||^s \\
&= E||B_1||^s \{1 + \sum_{k=1}^{k_0} E||A_t...A_{t-k+1}||^s + \sum_{k=k_0+1}^{\infty} E||A_t...A_{t-k+1}||^s\} \\
&\leq E||B_1||^s \{1 + \sum_{k=1}^{k_0} (E||A_1||^s)^k + \sum_{k=k_0+1}^{2k_0} E||A_t...A_{t-k+1}||^s + \sum_{k=2k_0+1}^{\infty} E||A_t...A_{t-k+1}||^s\} \\
&\leq E||B_1||^s \{1 + \sum_{k=1}^{k_0} (E||A_1||^s)^k + \delta \sum_{k=1}^{k_0} (E||A_1||^s)^k + \sum_{k=2k_0+1}^{\infty} E||A_t...A_{t-k+1}||^s\} \\
&\quad . \\
&\quad . \\
&\quad . \\
&\leq E||B_1||^s \{1 + \sum_{j=0}^{\infty} \delta^j \sum_{k=1}^{k_0} (E||A_1||^s)^k\} < \infty.
\end{aligned}$$

Pour l'inégalité précédente nous avons, en plus des arguments déja donnés, utilisé la relation $(\sum_i u_i)^s \leq \sum_i u_i^s$ pour toute suite de nombres positifs (u_i). On conclut que $E(\sigma_t^{2s}) < \infty$ et $E(\varepsilon_t^{2s}) < \infty$ en remarquant $\sigma_t^s \leq ||X_t||^s$ et $\varepsilon_t^s \leq ||X_t||^s$.

Le théorème suivant de Bollerslev [1] donne une condition nécessaire et suffisante pour que le modèle $GARCH(p,q)$ défini en (2.4) admette une unique solution stationnaire au second-ordre et non anticipative. ■

Théorème 92 ***(Stationnarité au second-ordre du processus GARCH(p,q))***

Le modèle $GARCH(p,q)$ défini par (2.4) admet une unique solution stationnaire au second-ordre et non anticipative avec $E(\varepsilon_t) = 0, var(\varepsilon_t) = \omega(\mathcal{B}(1) - \mathcal{A}(1))^{-1}$ et $cov(\varepsilon_t, \varepsilon_s) = 0$ pour $t \neq s$ si et seulement si $\mathcal{A}(1) - \mathcal{B}(1) < 0$.

Démonstration. Soit $(\varepsilon_t)_{t \in \mathbb{Z}}$ un processus $GARCH(p,q)$ au sens de la définition (78) stationnaire au second-ordre et non anticipatif, alors

$$E(\varepsilon_t^2) = E(E(\varepsilon_t^2 / \mathcal{F}_{t-1})) = E(\sigma_t^2) < \infty$$

et ne dépend pas de t. En prennant l'éspérance mathématique des deux membres de (2.1), on obtient

$$E(\varepsilon_t^2) = \omega + \sum_{i=1}^{q} \alpha_i E(\varepsilon_t^2) + \sum_{j=1}^{p} \beta_j E(\varepsilon_t^2).$$

qui implique que

$$E(\varepsilon_t^2) = \omega(1 - \sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j)^{-1},$$

prouvu que $(1 - \sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j)^{-1} = (\mathcal{B}(1) - \mathcal{A}(1))^{-1} \neq 0$.

Supposons maintenant que $\mathcal{A}(1) - \mathcal{B}(1) < 0$. On montre qu'il existe une solution stationnaire au second-ordre pour le modèle $GARCH(p,q)$ défini en (2.4). Pour $t, k \in \mathbb{Z}$, soit

$$X_t(k) = \begin{cases} 0 & \text{si } k < 0, \\ A_t X_{t-1}(k-1) + B_t & \text{si } k \geq 0. \end{cases}$$

Donc, on a

$$X_t(k) - X_t(k-1) = \begin{cases} 0 & \text{si } k < 0, \\ B_t & \text{si } k = 0, \\ A_t(X_{t-1}(k-1) + X_{t-1}(k-2)) + B_t & \text{si } k > 0. \end{cases}.$$

Par itération sur (2.41), on obtient pour $k > 0$,

$$X_t(k) - X_t(k-1) = A_t A_{t-1} ... A_{t-k+1} B_{t-k}.$$

Par conséquent,

On sait que les matrices aléatoires A_t définie en (2.24) sont à coefficients positifs, alors en utilisant la norme matricielle $||A|| = \sum_{i,j} |a_{ij}|$, on obtient

$$E||X_t(k) - X_t(k-1)|| = ||E(A_t A_{t-1} ... A_{t-k+1} B_{t-k})||.$$

puisque $(\eta_t)_{t \in \mathbb{Z}}$ est iid, les termes $A_t A_{t-1} ... A_{t-k+1}$ et B_{t-k} sont indépendants et

$$E(A_t) = A \ et \ E(B_t) = B, t \in \mathbb{Z}.$$

Par conséquent

$$E||X_t(k) - X_t(k-1)|| = ||A^k B|| = (1, ..., 1)A^k B. \tag{2.41}$$

La condition $\mathcal{A}(1) - \mathcal{B}(1) < 0$ implique que les valeurs propres de A sont de module strictement inférieur à 1. En effet, on peut vérifier que

$$\det(\lambda I_{p+q} - A) = \lambda^{p+q}(1 - \sum_{i=1}^{q} \alpha_i \lambda^{-i} + \sum_{j=1}^{p} \beta_j \lambda^{-j}),$$

par exemple en retranchant la $q+1$-ème ligne de $\lambda I_{p+q} - A$ à la première, puis en développant le déterminant par rapport à la première ligne. Donc si $|\lambda| \geq 1$, en utilisant l'inégalité $|a-b| \geq |a| - |b|$, on obtient

$$|\det(\lambda I_{p+q} - A)| \geq |1 - \sum_{i=1}^{q} \alpha_i \lambda^{-i} + \sum_{j=1}^{p} \beta_j \lambda^{-j}| \geq 1 - \sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j = \mathcal{B}(1) - \mathcal{A}(1) > 0,$$

et donc $\rho(\mathcal{A}) < 1$. En utilisant le résultat (2.31), on déduit que $A^k \to 0, k \to \infty$. Par conséquent, pour t fixé, $X_t(k)$ converge dans L^1 et presque sûrement lorsque $k \to \infty$. Soit X_t la limite de $(X_t(k))_{k\in\mathbb{N}}$. Pour k fixé, le processus $(X_t(k))_{k\in\mathbb{Z}}$ est strictement stationnaire. Le processus limite $(X_t)_{t\in\mathbb{Z}}$ est aussi strictement stationnaire et est une solution de l'équation aux récurrences stochastique multivariée (2.8). Le processus $(\varepsilon_t)_{t\in\mathbb{Z}}$ défini par $\varepsilon_t = \eta_t \sqrt{X_{q+1,t}}$ est la solution strictement stationnaire, non anticipative et ergodique de modèle $GARCH(p,q)$ défini par (2.4). Puisque

$$E(\varepsilon_t^2) = \omega(1 - \sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j)^{-1} < \infty.$$

indépendament de t, la solution strictement stationnaire $(\varepsilon_t)_{t\in\mathbb{Z}}$ est aussi stationnaire au second-ordre. De plus, cette solution est un bruit blanc parce que $E(\varepsilon_t) = E(E(\varepsilon_t/\mathcal{F}_{t-1})) = 0$ pour tout $t \in \mathbb{Z}$ et pour tout $h > 0$, on a

$$cov(\varepsilon_t, \varepsilon_{t-h}) = E(\varepsilon_{t-h} E(\varepsilon_t/\mathcal{F}_{t-1})) = 0.$$

L'unicité de la solution stationnaire au second-ordre et non anticipative du modèle $GARCH(p,q)$ (2.4) se démontre comme dans le cas d'un modèle $GARCH(1,1)$. $\blacksquare$

Chapitre 3

Estimation des paramètres

3.1 Introduction

L'utilisation de la méthode du quasi-maximum de vraisemblance est particulièrement intéressante pour les modèles GARCH car elle est valide, asymptotiquement, pour tout processus GARCH strictement stationnaire, sans hypothèse de moments sur le processus observé.

Dans ce chapitre, nous étudions en détail la méthode du quasi-maximum de vraisemblance conditionnelle (à des valeurs initiales). Dans un premier temps, nous nous limiterons aux GARCH purs (le processus observé est un GARCH). Cette vraisemblance est écrite comme si la loi des variables η_t était normale centrée réduite (on parle de pseudo ou quasi-vraisemblance), mais cette hypothèse n'est pas nécessaire pour la convergence forte de l'estimateur. Elle a évidemment un effet sur la variance de la loi normale asymptotique de l'estimateur.

3.2 Quasi-vraisemblance conditionnelle

On supposera que les observations $\varepsilon_1, ..., \varepsilon_n$ constituent une réalisation (de longueur n) d'un processus $GARCH(p, q)$, solution strictement stationnaire non anticipative du modèle

$$\left\{ \begin{array}{c} \varepsilon_t = \sqrt{h_t}\eta_t, \\ h_t = \omega + \sum_{i=1}^{q} \alpha_{0i}\varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_{0j} h_{t-j}, \forall t \in \mathbb{Z}, \end{array} \right. \quad (3.1)$$

où $(\eta_t)_{t\in\mathbb{Z}}$ est une suite de variables iid de variance unité, $\omega_0 > 0, \alpha_{0i} \geq 0$ $(i = 1, ..., q), \beta_{0j} \geq 0$ $(j = 1, ..., p)$.

Les ordres p et q sont supposés connus. Le vecteur des paramètres

$$\theta = (\theta_1, ..., \theta_{p+q+1})' := (\omega, \alpha_1, ..., \alpha_q, \beta_1, ..., \beta_p)', \tag{3.2}$$

appartient à un espace de paramètres

$$\Theta \subset]0, +\infty[\times[0, \infty[^{p+q}. \tag{3.3}$$

La vraie valeur du paramètre est inconnue. Elle est notée

$$\theta_0 = (\omega_0, \alpha_{01}, ..., \alpha_{0q}, \beta_{01}, ..., \beta_{0p})'.$$

Pour écrire la vraisemblance du modèle, il faut spécifier une distribution particulière pour les variables iid η_t. On considère généralement la quasi-vraisemblance gaussienne, i.e, la vraisemblance obtenue à partir d'une loi normale centrée réduite pour les η_t. Nous ne ferons cependant pas l'hypothèse que cette loi constitue la vraie distribution du processus iid.

La spécification d'une distribution gaussienne pour les variables η_t ne permet pas d'en déduire simplement la loi de l'échantillon. On travaille avec la vraisemblance de $\varepsilon_1, ..., \varepsilon_n$ conditionnellement à certaines valeurs initiales.

Étant données des valeurs initiales $\varepsilon_0, ..., \varepsilon_{1-q}, \widetilde{\sigma}_0^2, ..., \widetilde{\sigma}_{1-p}^2$ que nous allons préciser, la vraisemblance conditionnelle gaussienne $L_n(\theta)$ s'écrit

$$L_n(\theta) = L_n(\theta; \varepsilon_1, ..., \varepsilon_n) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\widetilde{\sigma}_t^2}} \exp(-\frac{\varepsilon_t^2}{2\widetilde{\sigma}_t^2}),$$

où les $\widetilde{\sigma}_t^2$ sont définis récursivement, pour $t \geq 1$, par

$$\widetilde{\sigma}_t^2 = \widetilde{\sigma}_t^2(\theta) = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \widetilde{\sigma}_{t-j}^2. \tag{3.4}$$

Pour une valeur donnée de θ, sous l'hypothèse de stationnarité au second ordre, la variance non conditionnelle (correspondant à cette valeur de θ) est un choix raisonnable pour les valeurs initiales inconnues :

$$\varepsilon_0^2 = ... = \varepsilon_{1-q}^2 = \sigma_0^2 = ... = \sigma_{1-p}^2 = \frac{\omega}{1 - \sum_{i=1}^q \alpha_i - \sum_{j=1}^p \beta_j}. \tag{3.5}$$

On peut prendre comme valeurs initiales

$$\varepsilon_0^2 = ... = \varepsilon_{1-q}^2 = \widetilde{\sigma}_0^2 = ... = \widetilde{\sigma}_{1-p}^2 = \omega. \tag{3.6}$$

ou encore

$$\varepsilon_0^2 = ... = \varepsilon_{1-q}^2 = \widetilde{\sigma}_0^2 = ... = \widetilde{\sigma}_{1-p}^2 = \varepsilon_1^2. \tag{3.7}$$

Un estimateur du QMV de θ est défini comme toute solution mesurable $\widehat{\theta}_n$ de

$$\widehat{\theta}_n = \arg_{\theta\in\Theta} \max L_n(\theta).$$

On voit, en prenant le logarithme, que maximiser la vraisemblance revient à minimiser par rapport à θ

$$\widehat{I}_n(\theta) = n^{-1} \sum_{t=1}^n \widetilde{l}_t, \quad \text{où } \widetilde{l}_t = \widetilde{l}_t(\theta) = \frac{\varepsilon_t^2}{\widetilde{\sigma}_t^2} + \log\ \widetilde{\sigma}_t^2, \tag{3.8}$$

et $\widetilde{\sigma}_t^2$ est définie en (3.4). Un estimateur du quasi-maximum de vraisemblance est donc une solution mesurable de l'équation

$$\widehat{\theta}_n = \arg_{\theta\in\Theta} \min \widetilde{l}_t(\theta). \tag{3.9}$$

Équation de vraisemblance

On obtient les équations de vraisemblance en annulant la dérivée par rapport à θ du critère $\widetilde{I}_n(\theta)$, ce qui donne

$$\frac{1}{n} \sum_{t=1}^n \{\varepsilon_t^2 - \widetilde{\sigma}_t^2\} \frac{1}{\widetilde{\sigma}_t^4} \frac{\partial \widetilde{\sigma}_t^2}{\partial \theta} = 0. \tag{3.10}$$

Ces équations s'interprètent, pour n grand, comme des relations d'orthogonalité. En effet, comme nous le verrons plus précisément dans la partie suivante, le terme de gauche de l'égalité précédente se comporte asymptotiquement comme

$$\frac{1}{n}\sum_{t=1}^{n}\{\varepsilon_t^2-\sigma_t^2\}\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\theta}, \tag{3.11}$$

l'influence des valeurs initiales étant nulle lorsque$n \to \infty$. Or, pour la vraie valeur du paramètre, l'innovation de ε_t^2 est $\nu_t = \varepsilon_t^2 - \sigma_t^2$. Donc sous réserve que l'espérance existe, on a

$$E_{\theta_0}(\nu_t\frac{1}{\sigma_t^4(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}) = 0,$$

car $\frac{1}{\sigma_t^4(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}$ est une fonction mesurable des $\epsilon_{t-i}, i > 0$. Ce résultat n'est autre que la version asymptotique de (3.10) en θ_0, en utilisant le théorème ergodique.

3.2.1 Propriétés asymptotiques de l'estimateur du QMV

Dans tout le chapitre, nous utiliserons comme norme d'une matrice $A = (a_{ij})$ quelconque la norme $||A|| = \sum |a_{ij}|$. Le rayon spectral d'une matrice carrée A sera noté $\rho(A)$. Le produit de Kronecker sera noté $\otimes$ et le symbole $\to^{\mathcal{L}}$ désignera la convergence en loi.

Convergence forte

Rappelons que le modèle (3.1) possède une solution strictement stationnaire si et seulement si la suite de matrices $A_0 = (A_{0t})$, où

$$A_{0t} = \begin{pmatrix} \alpha_{01}\eta_t^2 & & \dots & & \alpha_{0q}\eta_t^2 & \beta_{01}\eta_t^2 & & \dots & & \beta_{0p}\eta_t^2 \\ 1 & 0 & \dots & & 0 & 0 & & \dots & & 0 \\ 0 & 1 & \dots & & 0 & 0 & & \dots & & 0 \\ . & . & \dots & & . & . & & \dots & & . \\ 0 & & \dots & 1 & 0 & 0 & & \dots & 0 & 0 \\ \alpha_{01} & & \dots & & \alpha_{0q} & \beta_{01} & & \dots & & \beta_{0p} \\ 0 & & \dots & & 0 & 1 & 0 & \dots & & 0 \\ 0 & & \dots & & 0 & 0 & 1 & \dots & & 0 \\ . & . & \dots & & . & . & & \dots & & . \\ 0 & & \dots & 0 & 0 & 0 & & \dots & 1 & 0 \end{pmatrix},$$

admet un coefficient de Lyapounov strictement négatif, $\gamma(A_0) < 0$, où

$$\gamma(A_0) := \inf_{t\in\mathbb{N}^*} \frac{1}{t} E(\log ||A_t A_{t-1}...A_1||) = \lim_{t\to\infty} p.s. \frac{1}{t} \log ||A_t A_{t-1}...A_1||. \tag{3.12}$$

Notons

$$\mathcal{A}_\theta(z) = \sum_{i=1}^{q} \alpha_i z^i \quad \text{et} \quad \mathcal{B}_\theta(z) = 1 - \sum_{j=1}^{p} \beta_j z^j.$$

Par convention $\mathcal{A}_\theta(z) = 0$ si $q = 0$ et $\mathcal{B}_\theta(z) = 1$ si $p = 0$.

Pour montrer la convergence forte, les hypothèses suivantes seront faites.

A1 : $\theta \in \Theta$ et Θ est compact.

A2 : $\gamma(A_0) < 0$ et $\forall\theta \in \Theta, \sum_{j=1}^{p} \beta_j < 1$.

A3 : η_t^2 a une loi non dégénérée et $E\eta_t^2 = 1$.

A4 : si $p > 0, \mathcal{A}_{\theta_0}(z)$ et $\mathcal{B}_{\theta_0}(z)$ n'ont pas de racine commune, $\mathcal{A}_{\theta_0}(1) \neq 0$, et $\alpha_{0q} + \beta_{0p} \neq 0$.

Dans le théorème suivant de Francq et Zakoïan [11] donne la convergence forte de l'estimateur de quasi-maximum de vraisemblance (EQMV).

Théorème 93 ***(Convergence forte de l'estimateur du QMV)*** *Soit $(\widehat{\theta}_n)$ une suite d'estimateurs du QMV satisfaisant (3.9), avec les conditions initiales (3.6) ou (3.7). Sous les hypothèses* ***A1-A4****, presque sûrement*

$$\widehat{\theta}_n \to \theta_0, \quad \textit{quand} \;\; n \to \infty.$$

Démonstration. La démonstration repose sur une représentation vectorielle autorégressive d'ordre un du vecteur $\sigma_{-t}^2 = (\sigma_t^2, \sigma_{t-1}^2, ..., \sigma_{t-p+1}^2)$, analogue à celle utilisée pour l'étude de la stationnarité. L'hypothèse **A2** permet d'exprimer σ_{-t}^2 sous forme d'une série dépendant du passé infini de la variable ε_t^2. On montre que les valeurs initiales n'ont pas d'importance asymptotiquement en utilisant le fait que, sous l'hypothèse de stationnarité stricte, ε_t^2admet nécessairement un moment d'ordre s, avec $s > 0$. Cette propriété permet également de vérifier que l'espérance de $l_t(\theta_0)$ est bien définie dans $\mathbb{R}$ et que $E_{\theta_0}(l_t(\theta)) - E_{\theta_0}(l_t(\theta_0)) \geq 0$, ce qui assure que le critère limite est minimisé en la vraie valeur. La difficulté provient du fait que $E_{\theta_0}(l_t^+(\theta))$ peut être égal à $+\infty$.Les hypothèses **A3** et **A4** sont cruciales pour établir l'identifiabilité : la première exclut l'existence d'une combinaison linéaire constante entre les $\varepsilon_{t-j}^2, j \geq 0$. On utilise égale-ment l'hypothèse d'absence de racines communes. L'ergodicité de $l_t(\theta)$ et un argument de compacité permettent de conclure.

On a

$$\sigma_{-t}^2 = c_{-t} + B\sigma_{-t-1}^2, \tag{3.13}$$

où

$$\sigma_{-t}^2 = \begin{pmatrix} \sigma_t^2 \\ \sigma_{t-1}^2 \\ . \\ . \\ . \\ \sigma_{t-p+1}^2 \end{pmatrix}, \quad c_{-t} = \begin{pmatrix} \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 \\ 0 \\ . \\ . \\ . \\ 0 \end{pmatrix}, \quad B = \begin{pmatrix} \beta_1 & \beta_2 & ... & \beta_p \\ 1 & 0 & ... & 0 \\ . & & & \\ . & & & \\ . & & & \\ 0 & ... & 1 & 0 \end{pmatrix}. \tag{3.14}$$

Nous allons établir les résultats intermédiaires suivants.

a) $\lim_{n\to\infty} \sup_{\theta\in\Theta} |I_n(\theta) - \widetilde{I}_n(\theta)| = 0$, p.s.

b) $(\exists t \in \mathbb{Z}$ tel que $\sigma_t^2(\theta) = \sigma_t^2(\theta_0) \quad P_{\theta_0}$ p.s $) \Rightarrow \theta = \theta_0$,

c) $E_{\theta_0}|l_t(\theta_0)| < \infty$, et si $\theta \neq \theta_0$, $E_{\theta_0} l_t(\theta) > E_{\theta_0} l_t(\theta_0)$,

d) pour tout $\theta \neq \theta_0$ il existe un voisinage $V(\theta)$ tel que $\liminf_{n\to\infty} \inf_{\theta^*\in V(\theta)} \widetilde{I}_n(\theta^*) > E_{\theta_0} l_1(\theta_0)$, p.s .

a) Oubli asymptotique des valeurs initiales. la condition $\sum_{j=1}^p \beta_j$ de l'hypothèse **A2** implique $\rho(B) < 1$. La compacité de Θ implique que

$$\sup_{\theta\in\Theta} \rho(B) < 1. \tag{3.15}$$

On obtient donc en itérant (3.13)

$$\sigma_{-t}^2 = c_{-t} + B^2 c_{-t-2} + ... + B^{t-1} c_{-1} + B^t \sigma_{-0}^2 = \sum_{k=0}^{\infty} B^k c_{-t-k}. \tag{3.16}$$

Soit $\widetilde{\sigma}^2_{-t}$ le vecteur obtenu en remplaçant σ^2_{t-i}par $\widetilde{\sigma}^2_{t-i}$ dans σ^2_{-t}, et soit $\widetilde{c}_{-t}$ le vecteur obtenu en remplaçant $\varepsilon^2_0, ..., \varepsilon^2_{1-q}$ par les valeurs initiales (3.6) ou (3.7). Nous avons

$$\widetilde{\sigma}^2_{-t} = c_{-t}+Bc_{-t-1}+...+B^{t-q-1}c_{-q+1}+B^{t-q}\widetilde{c}_{-q}+...+B^{t-q}\widetilde{c}_{-q}+...+B^{t-1}\widetilde{c}_{-1}+B^t\widetilde{\sigma}^2_{-0}. \quad (3.17)$$

On déduit de (3.15) que presque sûrement

$$\begin{aligned}\sup_{\theta\in\Theta}||\sigma^2_{-t} - \widetilde{\sigma}^2_{-t}|| &= \sup_{\theta\in\Theta}||\{\sum_{k=1}^{q} B^{t-k}(c_{-k} - \widetilde{c}_{-k}) + B^t(\sigma^2_{-0} - \widetilde{\sigma}^2_{-0})\}|| \\ &\leq K\rho^t, \forall t.\end{aligned} \quad (3.18)$$

Pour $x > 0$ on a $\log x \leq x - 1$. Par suite, pour $x, y > 0, |\log \frac{x}{y}| \leq \frac{|x-y|}{\min(x,y)}$. On a donc presque sûrement, en utilisant (3.18),

$$\begin{aligned}\sup_{\theta\in\Theta}|I_n(\theta) - \widetilde{I}_n(\theta)| &\leq n^{-1}\sum_{t=1}^{n}\sup_{\theta\in\Theta}\{|\frac{\widetilde{\sigma}^2_t - \sigma^2_t}{\widetilde{\sigma}^2_t\sigma^2_t}|\varepsilon^2_t + |\log(\frac{\sigma^2_t}{\widetilde{\sigma}^2_t})|\} \\ &\leq \{\sup_{\theta\in\Theta}\frac{1}{\omega^2}\}Kn^{-1}\sum_{t=1}^{n}\rho^t\varepsilon^2_t + \{\sup_{\theta\in\Theta}\frac{1}{\omega}\}Kn^{-1}\sum_{t=1}^{n}\rho^t.\end{aligned} \quad (3.19)$$

L'existence d'un moment d'ordre $s > 0$ pour ε^2_t permet d'affirmer que $\rho^t\varepsilon^2_t \to 0$ p.s.

b)Identifiabilité du paramètre. Supposons que $\sigma^2_t(\theta) = \sigma^2_t(\theta_0)P_{\theta_0}$ p.s. Le polynôme $\mathcal{B}_\theta(B)$ est inversible sous l'hypothèse **A2**. On obtient

$$\{\frac{\mathcal{A}_\theta(B)}{\mathcal{B}_\theta(B)} - \frac{\mathcal{A}_{\theta_0}(B)}{\mathcal{B}_{\theta_0}(B)}\}\varepsilon^2_t = \frac{\omega_0}{\mathcal{B}_{\theta_0}(1)} - \frac{\omega}{\mathcal{B}_\theta(1)} \text{ p.s. } \forall t.$$

Si la série en B entre accolades était non nulle, cela signifierait qu'il existerait une combinaison linéaire des ε^2_{t-j} , $j \geq 0$, égale à une constante. Donc l'innovation linéaire du processus (ε^2_t) serait nulle. Or, la loi de η_t étant non dégénérée d'après **A3**,

$$\varepsilon^2_t - E_{\theta_0}(\varepsilon^2_t/\varepsilon^2_{t-1}, ...) = \sigma^2_t(\theta_0)(\eta^2_t - 1) \neq 0, \quad \text{avec probabilité positive.}$$

On a donc

$$\frac{\mathcal{A}_\theta(z)}{\mathcal{B}_\theta(z)} = \frac{\mathcal{A}_{\theta_0}(z)}{\mathcal{B}_{\theta_0}(z)}, \forall|z| \leq 1 \text{ et } \frac{\omega_0}{\mathcal{B}_{\theta_0}(1)} = \frac{\omega}{\mathcal{B}_\theta(1)}. \quad (3.20)$$

Sous l'hypothèse d'absence de racines communes **A4**, ceci entraîne $\mathcal{A}_\theta(z) = \mathcal{A}_{\theta_0}(z), \mathcal{B}_\theta(z) = \mathcal{B}_{\theta_0}(z)$ et $\omega = \omega_0$. On a montré b).

c) Le critère limite est minimisé en la vraie valeur. Le critère que l'on minimise n'est pas intégrable en tout point, mais remarquons que $E_{\theta_0} I_n(\theta) = E_{\theta_0} l_t(\theta)$ est bien défini dans $\mathbb{R} \cup \{+\infty\}$ car, notant $x^- = \min(x, 0)$ et $x^+ = \max(x, 0)$,

$$E_{\theta_0} l_t^-(\theta) \leq E_{\theta_0} \log^- \sigma_t^2 \leq \max\{0, -\log \omega\} < \infty.$$

Il reste à montrer que $E_{\theta_0} l_t^+(\theta_0) < \infty$. Utilisant l'inégalité de Jensen et, à nouveau, l'existence d'un moment d'ordre $s > 0$ pour ε_t^2, il vient

$$E_{\theta_0} \log^+ \sigma_t^2(\theta_0) < \infty. \tag{3.21}$$

car

$$E_{\theta_0} \log \sigma_t^2(\theta_0) = E_{\theta_0} \frac{1}{s} \log\{\sigma_t^2(\theta_0)\}^s \leq \frac{1}{s} \log E_{\theta_0} \log\{\sigma_t^2(\theta_0)\}^s < \infty.$$

Donc

$$E_{\theta_0} l_t(\theta_0) = E_{\theta_0}\{\frac{\sigma_t^2(\theta_0)\eta_t^2}{\sigma_t^2(\theta_0)} + \log \sigma_t^2(\theta_0)\} = 1 + E_{\theta_0} \log \sigma_t^2(\theta_0) < \infty.$$

Ayant déjà établi que $E_{\theta_0} l_t^-(\theta_0) < \infty$, on en déduit que $E_{\theta_0} l_t(\theta_0)$ est bien dans $\mathbb{R}$.

Puisque pour tout $x > 0, \log x \leq x - 1$, avec égalité si et seulement si $x = 1$, on a

$$\begin{aligned} E_{\theta_0} l_t(\theta) - E_{\theta_0} l_t(\theta_0) &= E_{\theta_0} \log \frac{\sigma_t^2(\theta)}{\sigma_t^2(\theta_0)} + E_{\theta_0} \frac{\sigma_t^2(\theta_0)\eta_t^2}{\sigma_t^2(\theta)} - E_{\theta_0} \eta_t^2 \\ &= E_{\theta_0} \log \frac{\sigma_t^2(\theta)}{\sigma_t^2(\theta_0)} + E_{\theta_0} \frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)} - 1 \\ &\geq E_{\theta_0}\{\log \frac{\sigma_t^2(\theta)}{\sigma_t^2(\theta_0)} + \log \frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)}\} = 0, \end{aligned} \tag{3.22}$$

avec égalité si et seulement si $\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)} = 1 P_{\theta_0}$ p.s, c'est-à-dire, étant donné b), si et seulement si $\theta = \theta_0$.

d) Utilisation de la compacité de Θ et de l'ergodicité de $(l_t(\theta))$. Pour tout $\theta \in \Theta$ et tout entier positif k, soit $V_k(\theta)$ la boule ouverte de centre θ et de rayon $1/k$. En raison de a),

$$\begin{aligned}\liminf_{n\to\infty} \inf_{\theta^*\in V_k(\theta)\cap\Theta} \widetilde{I}_n(\theta^*) &\geq \liminf_{n\to\infty} \inf_{\theta^*\in V_k(\theta)\cap\Theta} I_n(\theta^*) - \limsup_{n\to\infty} \sup_{\theta\in\Theta} |I_n(\theta) - \widetilde{I}_n(\theta)| \\ &\geq \liminf_{n\to\infty} n^{-1} \sum_{t=1}^{n} \inf_{\theta^*\in V_k(\theta)\cap\Theta} l_t(\theta^*).\end{aligned}$$

Pour obtenir la convergence de cette moyenne empirique on ne peut utiliser le théorème ergodique standard car nous avons vu que $l_t(\theta^*)$ n'est pas nécessairement intégrable, sauf en θ_0. Une adaptation de ce théorème pour une suite strictement stationnaire et ergodique de variables admettant une espérance dans $\mathbb{R} \cup \{+\infty\}$, ce qui est le cas de $\{l_t(\theta^*)\}$ et par suite de $\{\inf_{\theta^*\in V_k(\theta)\cap\Theta} l_t(\theta^*)\}$, permet d'affirmer que

$$\liminf_{n\to\infty} n^{-1} \sum_{t=1}^{n} \inf_{\theta^*\in V_k(\theta)\cap\Theta} l_t(\theta^*) = E_{\theta_0} \inf_{\theta^*\in V_k(\theta)\cap\Theta} l_1(\theta^*).$$

D'après le théorème de Beppo-Levi, $E_{\theta_0} \inf_{\theta^*\in V_k(\theta)\cap\Theta} l_1(\theta^*)$ tend en croissant vers $E_{\theta_0} l_1(\theta)$ quand $k \to \infty$. Étant donné (3.22), nous avons montré d).

La fin de la preuve du théorème utilise un argument de compacité. Remarquons d'abord que pour tout voisinage $V(\theta_0)$ de θ_0

$$\limsup_{n\to\infty} \inf_{\theta^*\in V(\theta_0)} \widetilde{I}_n(\theta^*) \leq \lim_{n\to\infty} \widetilde{I}_n(\theta_0) = \lim_{n\to\infty} I_n(\theta_0) = E_{\theta_0} l_1(\theta_0). \tag{3.23}$$

Le compact Θ est recouvert par la réunion d'un voisinage quelconque $V(\theta_0)$ de θ_0 et de l'ensemble des $V(\theta), \theta \in \Theta \backslash V(\theta_0)$, où $V(\theta)$ vérifie d). Il existe donc un sous recouvrement fini de Θ par $V(\theta_0), V(\theta_1), ..., V(\theta_k)$, d'où l'on déduit que

$$\inf_{\theta\in\Theta} \widetilde{I}_n(\theta) = \min_{i=0,1,\ldots,k} \inf_{\theta\in\Theta\cap V(\theta_i)} \widetilde{I}_n(\theta).$$

Les relations d) et (3.23) montrent que, presque sûrement, $\widehat{\theta}_n$ appartient à $V(\theta_0)$ pour n assez grand. Ceci étant vrai pour tout voisinage $V(\theta_0)$, le résultat est montré. ■

Remarque 94 *1. On ne suppose pas que la vraie valeur θ_0 du paramètre appartient à l'intérieur de Θ. Le théorème permet donc de traiter les cas où certains coefficients, α_i ou β_j, sont nuls.*

2. *L'hypothèse **A4** disparaît dans le cas $ARCH$. Dans le cas général, elle permet de sur-identifier l'un des ordres, p ou q, mais pas les deux. On estime ainsi de manière convergente les paramètres d'un processus $GARCH(p-1,q)$ (ou d'un $GARCH(p,q-1)$) si on utilise un $GARCH(p,q)$.*

3. *Quand $p \neq 0$, l'hypothèse A4 exclut le cas où tous les α_{0i} sont nuls. Ceci est évidemment nécessaire, sinon le modèle a pour solution un bruit blanc fort qui peut s'écrire de multiples manières. Par exemple, un bruit blanc fort de variance 1 peut s'écrire sous la forme d'un $GARCH(1,1)$ avec $\sigma_t^2 = \sigma^2(1-\beta) + 0 \times \varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$.*

4. *L'hypothèse d'absence de racines communes, dans **A4**, n'est restrictive que si $p > 1$ et $q > 1$. En effet si $q = 1$, la seule racine de $\mathcal{A}_{\theta_0}(z)$ est 0 et $\mathcal{B}_{\theta_0}(0) \neq 0$.*

   Si $p = 1$ et $\beta_{01} \neq 0$, la seule racine de $\mathcal{B}_{\theta_0}(z)$ est $\frac{1}{\beta_{01}} > 0$ (si $\beta_{01} = 0$, le polynôme n'admet pas de racine). En raison de la positivité des coefficients α_{0i}, cette valeur ne peut annuler $\mathcal{A}_{\theta_0}(z)$.

5. *L'hypothèse que $E\eta_t = 0$ n'est pas nécessaire pour les propriétés asymptotiques de l'estimateur du QMV d'un $GARCH$. La variance conditionnelle de ε_t est donc, en général, seulement proportionnelle à h_t dans ce modèle :*

   $var(\varepsilon_t/\varepsilon_u, u < t) = \{1-(E\eta_t)^2\}h_t$. L'hypothèse $E\eta_t^2 = 1$ est faite pour des raisons d'identifiabilité et n'est pas restrictive dès que $E\eta_t^2 < \infty$.

Normalité asymptotique

On considère les hypothèses supplémentaires suivantes.

A5 : $\theta_0 \in \Theta^o$, où Θ^o est l'intérieur de Θ .

A6 : $\mathcal{K}_\eta = E\eta_t^4 < \infty$.

La loi limite de $\widehat{\theta}_n$ est donnée par le résultat suivant de Francq et Zakoïan [10].

Théorème 95 ***(Normalité asymptotique de l'estimateur du QMV)***

*Sous les hypothèses **A1-A6** ,*

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \rightarrow^{\mathcal{L}} \mathcal{N}(0, (\mathcal{K}_\eta - 1)J^{-1}),$$

où

$$J := E_{\theta_0}(\frac{\partial^2 l_t(\theta_0)}{\partial\theta\partial\theta'}) = E_{\theta_0}(\frac{1}{\sigma_t^4(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'} \qquad (3.24)$$

est une matrice définie positive.

Démonstration. La preuve de ce théorème repose classiquement sur un développement de Taylor du critère (3.8) en θ_0. On a

$$\begin{aligned} 0 &= n^{\frac{-1}{2}}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\widetilde{l}_t(\widehat{\theta}_n) \\ &= n^{\frac{-1}{2}}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\widetilde{l}_t(\theta_0) + (\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\widetilde{l}_t(\theta_{ij}^*))\sqrt{n}(\widehat{\theta}_n - \theta_0), \end{aligned} \qquad (3.25)$$

où les θ_{ij}^* sont entre $\widehat{\theta}_n$ et θ_0. Nous montrerons que

$$n^{\frac{-1}{2}}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\widetilde{l}_t(\theta_0) \rightarrow^{\mathcal{L}} \mathcal{N}(0, (\mathcal{K}_\eta - 1)J), \qquad (3.26)$$

et que

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\widetilde{l}_t(\theta_{ij}^*) \rightarrow J(i,j) \text{ en probabilité.} \qquad (3.27)$$

La preuve du théorème en découlera immédiatement. Nous allons à nouveau décomposer la démonstration en plusieurs points.

a) $E_{\theta_0}\left\|\frac{\partial l_t(\theta_0)}{\partial\theta}\frac{\partial l_t(\theta_0)}{\partial\theta'}\right\| < \infty, E_{\theta_0}\left\|\frac{\partial l_t(\theta_0)}{\partial\theta\partial\theta'}\right\| < \infty,$

b) J est inversible et $Var_{\theta_0}\{\frac{\partial l_t(\theta_0)}{\partial\theta}\} = \{\mathcal{K}_\eta - 1\}J,$

c) il existe un voisinage $\mathcal{V}(\theta_0)$ de θ_0 tel que, pour tous $i, j, k \in \{1, ..., p+q+1\}$,

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

d) $\left\| n^{\frac{-1}{2}} \sum_{t=1}^n \{ \frac{\partial l_t(\theta_0)}{\partial\theta} - \frac{\partial \widetilde{l}_t(\theta_0)}{\partial\theta} \} \right\|$ et $\sup_{\theta \in \mathcal{V}(\theta_0)} \left\| n^{-1} \sum_{t=1}^n \{ \frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 \widetilde{l}_t(\theta)}{\partial\theta\partial\theta'} \} \right\|$ tendent en probabilité vers 0 quand $n \to \infty$,

e) $n^{\frac{-1}{2}} \sum_{t=1}^n \frac{\partial}{\partial\theta} l_t(\theta_0) \to^{\mathcal{L}} \mathcal{N}(0, (\mathcal{K}_\eta - 1)J)$,

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

a) Intégrabilité des dérivées du critère en θ_0. Puisque $l_t(\theta) = \varepsilon_t^2/\sigma_t^2 + \log \sigma_t^2$, nous avons

$$\frac{\partial l_t(\theta)}{\partial\theta} = \left\{ 1 - \frac{\varepsilon_t^2}{\sigma_t^2} \right\} \left\{ \frac{1}{\sigma_t^2} \frac{\partial\sigma_t^2}{\partial\theta} \right\}. \tag{3.28}$$

$$\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'} = \left\{ 1 - \frac{\varepsilon_t^2}{\sigma_t^2} \right\} \left\{ \frac{1}{\sigma_t^2} \frac{\partial^2\sigma_t^2}{\partial\theta\partial\theta'} \right\} + \left\{ 2\frac{\varepsilon_t^2}{\sigma_t^2} - 1 \right\} \left\{ \frac{1}{\sigma_t^2} \frac{\partial\sigma_t^2}{\partial\theta} \right\} \left\{ \frac{1}{\sigma_t^2} \frac{\partial\sigma_t^2}{\partial\theta'} \right\}. \tag{3.29}$$

Pour $\theta = \theta_0, \varepsilon_t^2/\sigma_t^2 = \eta_t^2$ est indépendant des termes en σ_t^2 ou en ses dérivées. Pour montrer a) il suffira donc de montrer

$$E_{\theta_0} \left\| \frac{1}{\sigma_t^2} \frac{\partial\sigma_t^2}{\partial\theta}(\theta_0) \right\| < \infty, E_{\theta_0} \left\| \frac{1}{\sigma_t^2} \frac{\partial^2\sigma_t^2}{\partial\theta\partial\theta'}(\theta_0) \right\| < \infty, E_{\theta_0} \left\| \frac{1}{\sigma_t^4} \frac{\partial\sigma_t^2}{\partial\theta} \frac{\partial\sigma_t^2}{\partial\theta'}(\theta_0) \right\| < \infty. \tag{3.30}$$

D'après (3.16) nous avons

$$\frac{\partial\sigma_{-t}^2}{\partial\omega} = \sum_{k=0}^{\infty} B^k 1. \quad \frac{\partial\sigma_{-t}^2}{\partial\alpha_i} = \sum_{k=0}^{\infty} B^k \varepsilon_{-t-k-i}^2, \tag{3.31}$$

$$\frac{\partial\sigma_{-t}^2}{\partial\beta_j} = \sum_{k=1}^{\infty} \left\{ \sum_{i=1}^{k} B^{i-1} B^{(j)} B^{k-i} \right\} c_{-t-k}, \tag{3.32}$$

où $\underline{1}=(1,0,...,0)\prime, \varepsilon^2_{-t} = (\varepsilon^2_t, 0, ..., 0)', B^{(j)}$ est une matrice $p \times p$ qui possède un 1 en position $(1, j)$, et des 0 partout ailleurs. Remarquons que, d'après la positivité des coefficients et (3.31)-(3.32), les dérivées de σ^2_t sont positives ou nulles. D'après(3.31), il est clair que $\partial\sigma^2_t/\partial\omega$ est bornée. Puisque $\sigma^2_t \geq \omega > 0$, il en est de même pour $\{\partial\sigma^2_t/\partial\omega\}/\sigma^2_t$. Cette variable possède donc des moments de tous ordres. D'après la seconde égalité de (3.31) et la positivité de tous les termes considérés, nous avons

$$\alpha_i \frac{\partial\sigma^2_{-t}}{\partial\alpha_i} = \sum_{k=0}^{\infty} B^k \alpha_i \varepsilon_{-t-k-i} \leq \sum_{k=0}^{\infty} B^k c_{-t-k} = \sigma^2_{-t}.$$

On en déduit que

$$\frac{1}{\sigma^2_t}\frac{\partial\sigma^2_t}{\partial\alpha_i}. \tag{3.33}$$

La variable $\sigma_t^{-2}(\partial\sigma^2_t/\partial\alpha_i)$ possède donc des moments de tous ordres en $\theta = \theta_0$. D'après (3.32) et $\beta_j B^{(j)} \leq B$, nous avons

$$\beta_j \frac{\partial\sigma^2_{-t}}{\partial\beta_j} \leq \sum_{k=1}^{\infty} \left\{ \sum_{i=1}^{k} B^{i-1} B B^{k-i} \right\} c_{-t-k} = \sum_{k=1}^{\infty} k B^k c_{-t-k}. \tag{3.34}$$

En utilisant (3.15), nous avons $\left\|B^k\right\| \leq K\rho^k$ pour tout k. De plus ε^2_t possèdant un moment d'ordre $s \in]0,1[$, il en est donc de même pour $c_{-t}(1) = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon^2_{t-i}$. En utilisant de plus (3.34), les minorations $\sigma^2_t \geq \omega + B^k(1,1)c_{-t-k}(1)$ et la relation $x/(1+x) \leq x^s$ pour tout $x \geq 0$, on obtient

$$\begin{aligned}
E_{\theta_0} \frac{1}{\sigma^2_t}\frac{\partial\sigma^2_t}{\partial\beta_j} &\leq E_{\theta_0} \frac{1}{\beta_j} \sum_{k=1}^{\infty} \frac{k B^k(1,1) c_{-t-k}(1)}{\omega + B^k(1,1)c_{-t-k}(1)} \\
&\leq \frac{1}{\beta_j} \sum_{k=1}^{\infty} k E_{\theta_0} \left\{ \frac{B^k(1,1)c_{-t-k}(1)}{\omega} \right\} \\
&\leq \frac{K^s}{\omega^s \beta_j} E_{\theta_0} \left\{c_{-t-k}(1)\right\}^s \sum_{k=1}^{\infty} k\rho^{sk} \leq \frac{K}{\beta_j}.
\end{aligned} \tag{3.35}$$

Sous l'hypothèse **A5** on a $\beta_{0j} > 0$ pour tout j, ce qui permet de conclure que la première espérance dans (3.30) existe.

Regardons maintenant les dérivées d'ordre supérieur de σ_t^2. D'après la première égalité de (3.31), on a

$$\frac{\partial \sigma_{-t}^2}{\partial \omega^2} = \frac{\partial^2 \sigma_{-t}^2}{\partial \omega \partial \alpha_i} = 0 \quad \text{et} \quad \frac{\partial^2 \sigma_{-t}^2}{\partial \omega \partial \beta_j} = \sum_{k=1}^{\infty} \left\{ \sum_{i=1}^{k} B^{i-1} B^{(j)} B^{(k-i)} \right\} \underline{1}. \tag{3.36}$$

On a donc

$$\beta_i \frac{\partial^2 \sigma_{-t}^2}{\partial \omega \partial \beta_j} \leq \sum_{k=1}^{\infty} k B^k \underline{1},$$

qui est un vecteur de constantes finies (puisque $\rho(B) < 1$). On en déduit que $\partial^2 \sigma_t^2(\theta_0)/\partial \omega \partial \theta_i$ est bornée et admet donc des moments de tous ordres. Il en est bien sûr de même pour $\left\{\partial^2 \sigma_t^2(\theta_0)/\partial \omega \partial \theta_i\right\}/\sigma_t^2(\theta_0)$. La seconde égalité de (3.31) donne

$$\frac{\partial^2 \sigma_{-t}^2}{\partial \alpha_i \partial \alpha_j} = 0 \quad \text{et} \quad \frac{\partial^2 \sigma_{-t}^2}{\partial \alpha_i \partial \beta_j} = \sum_{k=1}^{\infty} \left\{ \sum_{i=1}^{k} B^{i-1} B^{(j)} B^{k-i} \right\} \varepsilon_{-t-k-i}^2. \tag{3.37}$$

Les arguments utilisés pour montrer (3.35) donnent alors

$$E_{\theta_0} \frac{\partial^2 \sigma_t^2 / \partial \alpha_i \partial \beta_j}{\sigma_t^2} \leq \frac{K^*}{\beta_j}.$$

Ceci montre que $\left\{\partial^2 \sigma_t^2(\theta_0)/\partial \alpha_i \partial \theta\right\}/\sigma_t^2(\theta_0)$ est intégrable. La dérivation par rapport à $\beta_{j'}$ de la relation (3.32) donne

$$\begin{aligned}
\beta_j \beta_{j'} \frac{\partial^2 \sigma_{-t}^2}{\partial \beta_j \partial \beta_{j'}} &= \beta_j \beta_{j'} \sum_{k=2}^{\infty} \left[\begin{array}{c} \sum_{i=2}^{k} \left\{ \left(\sum_{l=1}^{i-1} B^{l-1} B^{(j')} B^{i-1-l} \right) B^{(j)} B^{k-i} \right\} \\ + \sum_{i=1}^{k-1} \left\{ B^{i-1} B^{(j)} \left(\sum_{l=1}^{k-i} B^{l-1} B^{(j')} B^{k-i-l} \right) \right\} \end{array} \right] c_{-t-k} \\
&\leq \sum_{k=2}^{\infty} \left[\sum_{i=2}^{k} (i-1) B^k + \sum_{i=1}^{k-1} (k-i) B^k \right] c_{-t-k} = \sum_{k=2}^{\infty} k(k-1) B^k c_{-t-k}, \qquad (3.38)
\end{aligned}$$

car $\beta_j B^{(j)} \leq B$. Comme pour (3.35), on en déduit

$$E_{\theta_0} \frac{\partial^2 \sigma_t^2 / \left\{\partial \beta_j \partial \beta_j'\right\}}{\sigma_t^2} \leq \frac{K^*}{\beta_j \beta_{j'}},$$

et l'existence de la deuxième espérance dans (3.30) est prouvée. Par ailleurs, puisque $\left\{\partial \sigma_t^2 / \partial \omega\right\} / \sigma_t^2$ est bornée, et puisque par (3.33), les variables $\left\{\partial \sigma_t^2 / \partial \alpha_i\right\} / \sigma_t^2$ sont bornées en θ_0, il est clair que

$$E_{\theta_0} \left\| \frac{1}{\sigma_t^4(\theta_0)} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta_i} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta} \right\| < \infty,$$

pour $i = 1, ..., q+1$. Avec des notations et arguments déjà utilisés pour montrer (3.35), et en utilisant l'inégalié élémentaire $x/(1+x) \leq x^{s/2}$pour tout $x \geq 0$, l'inégalité de Minkowski donne

$$\left\{ E_{\theta_0} \left(\frac{1}{\sigma_t^2(\theta_0)} \frac{\partial \sigma_t^2(\theta_0)}{\partial \beta_j} \right)^2 \right\}^{1/2} \leq \frac{1}{\beta_{0j}} \sum_{k=1}^{\infty} k \left\{ E_{\theta_0} \left(\frac{B^k(1,1) c_{-t-k}(1)}{\omega_0} \right)^s \right\}^{1/2} < \infty.$$

Finalement l'inégalité de Cauchy-Schwarz permet de conclure que le troisième espérance de (3.30) existe.

b) Inversibilité de J et lien avec la variance de la dérivée du critère. En utilisant a) et à nouveau l'indépendance entre $\eta_t^2 = \varepsilon_t^2 / \sigma_t^2(\theta_0)$ et σ_t^2 ainsi que ses dérivées, nous avons d'après (3.28)

$$E_{\theta_0} \left\{ \frac{\partial l_t(\theta_0)}{\partial \theta} \right\} = E_{\theta_0}(1 - \eta_t^2) E_{\theta_0} \left\{ \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta} \right\} = 0.$$

De plus, étant donné (3.30), J existe et vérifie bien (3.24). Nous avons également

$$\begin{aligned} Var_{\theta_0} \left\{ \frac{\partial l_t(\theta_0)}{\partial \theta} \right\} &= E_{\theta_0} \left\{ \frac{\partial l_t(\theta_0)}{\partial \theta} \frac{\partial l_t(\theta_0)}{\partial \theta'} \right\} \qquad (3.39) \\ &= E\left\{(1 - \eta_t^2)^2\right\} E_{\theta_0} \left\{ \frac{\partial \sigma_t^2(\theta_0)/\partial \theta}{\sigma_t^2(\theta_0)} \frac{\partial \sigma_t^2(\theta_0)/\partial \theta'}{\sigma_t^2(\theta_0)} \right\} \\ &= \{\mathcal{K}_\eta - 1\} J. \end{aligned}$$

Supposons maintenant que J soit non inversible. Alors il existe un vecteur non nul λ de $\mathbb{R}^{p+q+1}$ tel que $\lambda' \{\partial\sigma_t^2(\theta_0)/\partial\theta\} = 0$ p.s. D'après la stationnarité de $\{\partial\sigma_t^2(\theta_0)/\partial\theta\}_t$, on a

$$0 = \lambda' \frac{\partial\sigma_t^2(\theta_0)}{\partial\theta} = \lambda' \begin{pmatrix} 1 \\ \varepsilon_{t-1}^2 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_{t-q}^2 \\ \sigma_{t-1}^2(\theta_0) \\ \cdot \\ \cdot \\ \cdot \\ \sigma_{t-p}^2(\theta_0) \end{pmatrix} + \sum_{j=1}^{p} \beta_j \lambda' \frac{\partial\sigma_{t-j}^2(\theta_0)}{\partial\theta} = \lambda' \begin{pmatrix} 1 \\ \varepsilon_{t-1}^2 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_{t-q}^2 \\ \sigma_{t-1}^2(\theta_0) \\ \cdot \\ \cdot \\ \cdot \\ \sigma_{t-p}^2(\theta_0) \end{pmatrix}.$$

Posons $\lambda = (\lambda_0, \lambda_1, ..., \lambda_{q+p})'$. Il est clair que $\lambda_1 = 0$, sinon ε_{t-1}^2 serait mesurable par rapport à la tribu engendrée par $\{\eta_u, u < t-1\}$. Pour la même raison on a $\lambda_2 = ... = \lambda_{2+i} = 0$ si $\lambda_{q+1} = ... = \lambda_{q+i} = 0$. Par conséquent, $\lambda \neq 0$ implique une représentation $GARCH(p-1, q-1)$. Ceci est impossible en raison de **A4** en argumentant comme pour établir (3.20). Par suite $\lambda' J \lambda = 0$ implique $\lambda = 0$, ce qui termine la preuve de b).

c) Intégrabilité uniforme des dérivées d'ordre 3 du critère. En dérivant (3.29), nous obtenons

$$\begin{aligned} \frac{\partial^3 l_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k} \quad = \quad & \left\{1 - \frac{\varepsilon_t^2}{\sigma_t^2}\right\} \left\{\frac{1}{\sigma_t^2} \frac{\partial^3\sigma_t^2}{\partial\theta_i\partial\theta_j\partial\theta_k}\right\} \qquad (3.40) \\ & + \left\{2\frac{\varepsilon_t^2}{\sigma_t^2} - 1\right\} \left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_i}\right\} \left\{\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta_i\partial\theta_k}\right\} \\ & + \left\{2\frac{\varepsilon_t^2}{\sigma_t^2} - 1\right\} \left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_j}\right\} \left\{\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta_i\partial\theta_k}\right\} \\ & + \left\{2\frac{\varepsilon_t^2}{\sigma_t^2} - 1\right\} \left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_k}\right\} \left\{\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta_i\partial j}\right\} \\ & + \left\{2 - 6\frac{\varepsilon_t^2}{\sigma_t^2}\right\} \left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_i}\right\} \left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta j}\right\} \left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_k}\right\}. \end{aligned}$$

Commençons par étudier l'intégrabilité de $\{1 - \varepsilon_t^2/\sigma_t^2\}$. C'est le terme le plus délicat à traiter. En effet nous n'avons pas intégrabilité $\varepsilon_t^2/\sigma_t^2$ de uniformément sur Θ : en $\theta =$

$(\omega, 0')$, le rapport $\varepsilon_t^2/\sigma_t^2$ n'est intégrable que si $E\ \varepsilon_t^2$ existe. Nous allons cependant montrer l'intégrabilité de $\{1 - \varepsilon_t^2/\sigma_t^2\}$ uniformément en θ au voisinage de θ_0. Soit Θ^*un compact contenant θ_0 et contenu dans l'intérieur de Θ ($\forall \theta \in \Theta^*$, on a $\theta \geq \theta_* > 0$ composante par composante). Notons B_0 la matrice B (définie en (3.14)) évaluée au point $\theta = \theta_0$. Pour tout $\delta > 0$, il existe un voisinage $\mathcal{V}(\theta_0)$ de θ_0, entièrement contenu dans Θ^*, tel que pour tout $\theta \in \mathcal{V}(\theta_0)$,

$$B_0 \leq (1+\delta)B \text{ (i.e. } B_0(i,j) \leq (1+\delta)B(i,j) \text{ pour tout } i \text{ et tout } j).$$

Notons que, puisque $\mathcal{V}(\theta_0) \subset \Theta^*$, on a $\sup_{\theta \in \mathcal{V}(\theta_0)} 1/\alpha_i < \infty$. De (3.16), on tire

$$\sigma_t^2 = \omega \sum_{k=0}^{\infty} B^k(1,1) + \sum_{i=1}^{q} \alpha_i \left\{ \sum_{k=0}^{\infty} B^k(1,1)\varepsilon_{t-k-i}^2 \right\},$$

et, en utilisant encore $x/(1+x) \leq x^s$ pour tout $x \geq 0$ et tout $s \in]0,1[$,

$$\begin{aligned}
& \sup_{\theta \in \mathcal{V}(\theta_0)} \frac{\sigma_t^2(\theta_0)}{\sigma_t^2} \\
\leq\ & \sup_{\theta \in \mathcal{V}(\theta_0)} \left\{ \frac{\omega_0 \sum_{k=0}^{\infty} B_0^k(1,1)}{\omega} + \sum_{i=1}^{q} \alpha_{0i} \left(\sum_{k=0}^{\infty} \frac{B_0^k(1,1)\varepsilon_{t-k-i}^2}{\omega + \alpha_i B^k(1,1)\varepsilon_{t-k-i}^2} \right) \right\} \\
\leq\ & K + \sum_{i=1}^{q} \sup_{\theta \in \mathcal{V}(\theta_0)} \left\{ \frac{\alpha_{0i}}{\alpha_i} \sum_{k=0}^{\infty} \frac{B_0^k(1,1)}{B^k(1,1)} \left(\frac{\alpha_i B^k(1,1)\varepsilon_{t-k-i}^2}{\omega} \right)^s \right\} \\
\leq\ & K + K \sum_{i=1}^{q} \sum_{k=0}^{\infty} (1+\delta)^k \rho^{ks} \varepsilon_{t-k-i}^{2s}. \qquad (3.41)
\end{aligned}$$

Si on choisit s tel que $E\varepsilon_t^{2s} < \infty$ et, par exemple, $\delta = (1-\rho^s)/(2\rho^s)$ alors l'espérance de la série précédente est finie. On en déduit qu'il existe un voisinage $\mathcal{V}(\theta_0)$ de θ_0 tel que

$$E_{\theta_0} \sup_{\theta \in \mathcal{V}(\theta_0)} \frac{\varepsilon_t^2}{\sigma_t^2} = E_{\theta_0} \sup_{\theta \in \mathcal{V}(\theta_0)} \frac{\sigma_t^2(\theta_0)}{\sigma_t^2} < \infty.$$

En utilisant (3.41), en conservant le même choix de δ mais en choisissant s tel que $E\varepsilon_t^{4s} < \infty$, l'inégalité triangulaire donne

$$\begin{aligned} \left\| \sup_{\theta\in\mathcal{V}(\theta_0)} \frac{\varepsilon_t^2}{\sigma_t^2} \right\| &= \mathcal{K}_\eta^{1/2} \left\| \sup_{\theta\in\mathcal{V}(\theta_0)} \frac{\sigma_t^2(\theta_0)}{\sigma_t^2} \right\| \\ &\leq \mathcal{K}_\eta^{1/2} K + \mathcal{K}_\eta^{1/2} K q \sum_{k=0}^{\infty} (1+\delta)^k \rho^{ks} \left\| \varepsilon_t^{2s} \right\|_2 < \infty. \end{aligned} \tag{3.42}$$

Étudions maintenant le deuxième terme entre accolades dans (3.40). En dérivant (3.36), (3.37) et (3.38), à l'aide des arguments utilisés pour montrer (3.33), on obtient

$$\sup_{\theta\in\Theta^*} \frac{1}{\sigma_t^2} \frac{\partial^3 \sigma_t^2}{\partial\theta_{i_1}\partial\theta_{i_2}\partial\theta_{i_3}} \leq K.$$

quand les indices i_1, i_2 et i_3 ne sont pas tous dans $\{q+1, q+2, ..., q+1+p\}$ (i.e., quand on dérive par au moins un paramètre autre qu'un des β_j). En reprenant les arguments utilisés pour montrer (3.34) et (3.38), puis (3.35), on obtient

$$\beta_i\beta_j\beta_k \frac{\partial^3 \sigma_t^2}{\partial\beta_i\partial\beta_j\partial\beta_k} \leq \sum_{k=3}^{\infty} k(k-1)(k-2) B^k(1,1) c_{-t-k}(1),$$

$$\sup_{\theta\in\Theta^*} \frac{1}{\sigma_t^2} \frac{\partial^3 \sigma_t^2}{\partial\beta_i\partial\beta_j\partial\beta_k} \leq K \left\{ \sup_{\theta\in\Theta^*} \frac{1}{\omega^s \partial\beta_i\partial\beta_j\partial\beta_k} \right\} \sum_{k=3}^{\infty} k(k-1)(k-2)\rho^{ks} \left\{ \sup_{\theta\in\Theta^*} c_{-t-k}(1) \right\}^s$$

pour tout $s \in]0,1[$. Comme $E_{\theta_0} \{\sup_{\theta\in\Theta^*} c_{-t-k}(1)\}^{2s} < \infty$ pour un $s > 0$, on en déduit

$$E_{\theta_0} \sup_{\theta\in\Theta^*} \left| \frac{1}{\sigma_t^2} \frac{\partial^3 \sigma_t^2}{\partial\theta_i\partial\theta_i\partial\theta_i} \right|^2 < \infty. \tag{3.43}$$

Remarquons que la relation (3.43) la puissance 2 peut être remplacée par une puissance d arbitrairement grande :

$$E_{\theta_0} \sup_{\theta\in\Theta^*} \left| \frac{1}{\sigma_t^2} \frac{\partial^3 \sigma_t^2}{\partial\theta_i\partial\theta_j\partial\theta_k} \right|^d < \infty. \tag{3.44}$$

En utilisant l'inégalité de Cauchy-Schwarz, (3.42) et (3.43), on obtient

$$E_{\theta_0} \sup_{\theta\in\mathcal{V}(\theta_0} \left| \left\{ 1 - \frac{\varepsilon_t^2}{\sigma_t^2} \right\} \left\{ \frac{1}{\sigma_t^2} \frac{\partial^3 \sigma_t^2}{\partial\theta_i \partial\theta_j \partial\theta_k} \right\} \right| < \infty,$$

Les autres termes entre parenthèses dans (3.40) se traitent de la même manière. On montre en particulier que

$$E_{\theta_0} \sup_{\theta\in\Theta^*} \left| \frac{1}{\sigma_t^2} \frac{\partial^2 \sigma_t^2}{\partial\theta_i \partial\theta_j} \right|^d < \infty, \quad E_{\theta_0} \sup_{\theta\in\Theta^*} \left| \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial\theta_i} \right|^d < \infty, \tag{3.45}$$

pour tout entier d. Ceci permet par exemple d'établir à l'aide de Hölder que

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

Les autres termes de la somme dans (3.40) se traitent de la même manière. Ainsi on obtient c).

d) Oubli asymptotique des valeurs initiales. En utilisant (3.17), on obtient les équations analogues à (3.31)-(3.32) pour les dérivées de $\widetilde{\sigma}^2_{-t}$:

$$\frac{\partial \widetilde{\sigma}^2_{-t}}{\partial\omega} = \sum_{k=0}^{t-1-q} B^k 1 + \sum_{k=1}^{q} B^{t-k} \frac{\partial \widetilde{c}_{-k}}{\partial\alpha_i} + B^t \frac{\partial \widetilde{\sigma}^2_{-0}}{\partial\omega}, \tag{3.46}$$

$$\frac{\partial \widetilde{\sigma}^2_{-t}}{\partial\alpha_i} = \sum_{k=0}^{t-1-q} B^k \varepsilon^2_{-t-k-i} + \sum_{k=1}^{q} B^{t-k} \frac{\partial \widetilde{c}_{-k}}{\partial\alpha_i} + B^t \frac{\partial \widetilde{\sigma}^2_{-0}}{\partial\alpha_i}, \tag{3.47}$$

$$\frac{\partial \widetilde{\sigma}^2_{-t}}{\partial\beta_j} = \sum_{k=1}^{t-1-q} \left\{ \sum_{i=1}^{k} B^{i-1} B^{(j)} B^{k-i} \right\} c_{-t-k} + \sum_{k=1}^{q} \left\{ \sum_{i=1}^{t-k} B^{i-1} B^{(j)} B^{t-k-i} \right\} \widetilde{c}_{-k}, \tag{3.48}$$

où $\frac{\partial \widetilde{\sigma}^2_{-0}}{\partial \omega}$ vaut $(0, ..., 0)'$ quand les conditions initiales sont données par (3.7) et vaut $(1, ..., 1)'$ quand les conditions initiales sont données par (3.6). Les dérivées secondes ont des expressions similaires. La compacité de Θ, et le fait que $\rho(B) < 1$ permettent d'affirmer que, presque sûrement

$$\sup_{\theta \in \Theta} \left\| \frac{\partial \sigma_t^2}{\partial \theta} - \frac{\partial \widetilde{\sigma}_t^2}{\partial \theta} \right\| < K\rho^t, \quad \sup_{\theta \in \Theta} \left\| \frac{\partial^2 \sigma_t^2}{\partial \theta \partial \theta'} - \frac{\partial^2 \widetilde{\sigma}_t^2}{\partial \theta \partial \theta'} \right\| < K\rho^t, \forall t \in \mathbb{Z}. \tag{3.49}$$

En utilisant (3.18) on obtient

$$\left| \frac{1}{\sigma_t^2} - \frac{1}{\widetilde{\sigma}_t^2} \right| = \left| \frac{\widetilde{\sigma}_t^2 - \sigma_t^2}{\sigma_t^2 \widetilde{\sigma}_t^2} \right| \leq \frac{K\rho^t}{\sigma_t^2}, \qquad \frac{\sigma_t^2}{\widetilde{\sigma}_t^2} \leq 1 + K\rho^t. \tag{3.50}$$

Puisque

$$\frac{\partial \widetilde{l}_t(\theta)}{\partial \theta} = \left\{ 1 - \frac{\varepsilon_t^2}{\widetilde{\sigma}_t^2} \right\} \left\{ \frac{1}{\widetilde{\sigma}_t^2} \frac{\partial \widetilde{\sigma}_t^2}{\partial \theta} \right\} \quad \text{et} \quad \frac{\partial l_t(\theta)}{\partial \theta} = \left\{ 1 - \frac{\varepsilon_t^2}{\sigma_t^2} \right\} \left\{ \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta} \right\},$$

on a, en utilisant (3.50) et la première inégalité dans (3.49),

$$\begin{aligned} \left| \frac{\partial l_t(\theta_0)}{\partial \theta_i} - \frac{\partial \widetilde{l}_t(\theta_0)}{\partial \theta_i} \right| &= \left| \begin{array}{c} \left\{ \frac{\varepsilon_t^2}{\widetilde{\sigma}_t^2} - \frac{\varepsilon_t^2}{\sigma_t^2} \right\} \left\{ \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta_i} \right\} + \left\{ 1 - \frac{\varepsilon_t^2}{\widetilde{\sigma}_t^2} \right\} \left\{ \frac{1}{\sigma_t^2} - \frac{1}{\widetilde{\sigma}_t^2} \right\} \left\{ \frac{\partial \sigma_t^2}{\partial \theta_i} \right\} \\ + \left\{ 1 - \frac{\varepsilon_t^2}{\widetilde{\sigma}_t^2} \right\} \left\{ \frac{1}{\widetilde{\sigma}_t^2} \right\} \left\{ \frac{\partial \sigma_t^2}{\partial \theta_i} - \frac{\partial \widetilde{\sigma}_t^2}{\partial \theta i} \right\} \end{array} \right| (\theta_0) \\ &\leq K\rho^t (1 + \eta_t^2) \left| 1 + \left\{ \frac{1}{\sigma_t^2(\theta_0)} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta_i} \right\} \right|. \end{aligned}$$

Par suite

$$\left| n^{-1/2} \sum_{t=1}^n \left\{ \frac{\partial l_t(\theta_0)}{\partial \theta_i} - \frac{\partial \widetilde{l}_t(\theta_0)}{\partial \theta_i} \right\} \right| \leq K^* n^{-1/2} \sum_{t=1}^n \rho^t (1 + \eta_t^2) \left| 1 + \frac{1}{\sigma_t^2(\theta_0)} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta_i} \right|. \tag{3.51}$$

L'inégalité de Markov, (3.30), et l'indépendance entre η_t et $\sigma_t^2(\theta_0)$ implique que, pour tout $\varepsilon > 0$,

$$\begin{aligned} &P\left(n^{-1/2} \sum_{t=1}^n \rho^t (1 + \eta_t^2) \left| 1 + \frac{1}{\sigma_t^2(\theta_0)} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta_i} \right| > \varepsilon \right) \\ \leq\ & \frac{2}{\varepsilon} \left(1 + E_{\theta_0} \left| \frac{1}{\sigma_t^2(\theta_0)} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta_i} \right| \right) n^{-1/2} \sum_{t=1}^n \rho^t \to 0, \end{aligned}$$

ce qui, par (3.51), montre la première partie de d).

Regardons maintenant l'influence asymptotique des valeurs initiales sur les dérivées secondes du critère en un voisinage de θ_0. D'après (3.29) et les majorations précédentes, nous avons

$$\begin{aligned}
&\sup_{\theta\in\mathcal{V}(\theta_0}\left|n^{-1}\sum_{t=1}^{n}\left\{\frac{\partial^2 l_t(\theta_0)}{\partial\theta_i\partial\theta_j}-\frac{\partial^2\widetilde{l}_t(\theta_0)}{\partial\theta_i\partial\theta_j}\right\}\right| \\
\leq\;& n^{-1}\sum_{t=1}^{n}\sup_{\theta\in\mathcal{V}(\theta_0}\left|\begin{array}{c}
\left\{\frac{\varepsilon_t^2}{\widetilde{\sigma}_t^2}-\frac{\varepsilon_t^2}{\sigma_t^2}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta_i\partial\theta_j}\right\} \\
+\left\{1-\frac{\varepsilon_t^2}{\widetilde{\sigma}_t^2}\right\}\left\{\left(\frac{1}{\sigma_t^2}-\frac{1}{\widetilde{\sigma}_t^2}\right)\frac{\partial^2\sigma_t^2}{\partial\theta_i\partial\theta_j}+\frac{1}{\widetilde{\sigma}_t^2}\left(\frac{\partial^2\sigma_t^2}{\partial\theta_i\partial\theta_j}-\frac{\partial^2\widetilde{\sigma}_t^2}{\partial\theta_i\partial\theta_j}\right)\right\} \\
+\left\{2\frac{\varepsilon_t^2}{\sigma_t^2}-2\frac{\varepsilon_t^2}{\widetilde{\sigma}_t^2}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_i}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_j}\right\} \\
+\left\{2\frac{\varepsilon_t^2}{\sigma_t^2}-1\right\}\left\{\left(\frac{1}{\sigma_t^2}-\frac{1}{\widetilde{\sigma}_t^2}\right)\frac{\partial\sigma_t^2}{\partial\theta_i}+\frac{1}{\widetilde{\sigma}_t^2}\left(\frac{\partial\sigma_t^2}{\partial\theta_i}-\frac{\partial\widetilde{\sigma}_t^2}{\partial\theta_i}\right)\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_j}\right\} \\
+\left\{2\frac{\varepsilon_t^2}{\sigma_t^2}-1\right\}\left\{\frac{1}{\widetilde{\sigma}_t^2}\frac{\partial\widetilde{\sigma}_t^2}{\partial\theta_i}\right\}\left\{\left(\frac{1}{\sigma_t^2}-\frac{1}{\widetilde{\sigma}_t^2}\right)\frac{\partial\sigma_t^2}{\partial\theta_j}+\frac{1}{\widetilde{\sigma}_t^2}\left(\frac{\partial\sigma_t^2}{\partial\theta_j}-\frac{\partial\widetilde{\sigma}_t^2}{\partial\theta j}\right)\right\}
\end{array}\right| \\
\leq\;& Kn^{-1}\sum_{t=1}^{n}\rho^t\Upsilon_t,
\end{aligned}$$

où

$$\Upsilon_t=\sup_{\theta\in\mathcal{V}(\theta_0}\left\{1+\frac{\varepsilon_t^2}{\sigma_t^2}\right\}\left\{1+\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta_i\partial\theta_j}+\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_i}\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_j}\right\}.$$

D'après (3.42), (3.45) et l'inégalité de Hölder, on voit que, pour un certain voisinage $\mathcal{V}(\theta_0)$, l'espérance de Υ_t est une constante finie. En utilisant à nouveau l'inégalité de Markov on montre alors la seconde convergence de d).

e) Utilisation d'un TCL pour accroissements de martingale. Rappelons que $\varepsilon_u, u < t$ désigne la tribu engendrée par les variables $\varepsilon_{t-i}, i \geq 0$. Le vecteur des scores conditionnels est évidemment centré, ce qui peut se retrouver directement à partir de (3.28), en utilisant le fait que $\sigma_t^2(\theta_0) \in \varepsilon_u, u < t$ et $E_{\theta_0}(\varepsilon_t^2/\varepsilon_u, u < t) = \sigma_t^2(\theta_0)$:

$$E_{\theta_0}\left(\frac{\partial}{\partial\theta}l_t(\theta_0)\backslash\varepsilon_u, u<t\right)=\frac{1}{\sigma_t^4(\theta_0)}\left(\frac{\partial}{\partial\theta}\sigma_t^2(\theta_0)\right)E_{\theta_0}\left(\sigma_t^2(\theta_0)-\varepsilon_t^2\backslash\varepsilon_u, u<t\right)=0.$$

Notons également que, d'après (3.39), $Var_{\theta_0}(\partial l_t(\theta_0)/\partial\theta$ est finie. D'après l'inversibilité de J et les hypothèses sur la loi de η_t (qui entraînent $0 < \mathcal{K}_\eta - 1 < \infty$), cette matrice de variance est non dégénérée. Nous en déduisons que $\forall\lambda \in \mathbb{R}^{p+q+1}$ la suite $\left\{\lambda' \frac{\partial}{\partial\theta} l_t(\theta_0), \varepsilon_t\right\}_t$ est une différence de martingale stationnaire ergodique de carré intégrable.

f) Utilisation d'un second développement limité et du théorème ergodique. Reprenons le développement de Taylor (3.25) du critère en θ_0. On a, pour tous i et j

$$n^{-1}\sum_{t=1}^{n} \frac{\partial^2}{\partial\theta_i\partial\theta_j} l_t(\theta^*_{ij}) = n^{-1}\sum_{t=1}^{n} \frac{\partial^2}{\partial\theta_i\partial\theta_j} l_t(\theta_0) + n^{-1}\sum_{t=1}^{n} \frac{\partial}{\partial\theta'}\left\{\frac{\partial^2}{\partial\theta_i\partial\theta_j} l_t(\widetilde{\theta}_{ij})\right\}(\theta^*_{ij} - \theta_0),$$

où $\widetilde{\theta}_{ij}$ est entre θ^*_{ij} et θ_0. La convergence presque sûre de $\widetilde{\theta}_{ij}$ vers θ_0, le théorème ergodique et iii) impliquent que p.s.

$$\begin{aligned}
\lim_{n\to\infty}\sup \left\| n^{-1}\sum_{t=1}^{n} \frac{\partial}{\partial\theta'}\left\{\frac{\partial^2}{\partial\theta_i\partial\theta_j} l_t(\widetilde{\theta}_{ij})\right\}\right\| &\leq \lim_{n\to\infty}\sup n^{-1}\sum_{t=1}^{n} \sup_{\theta\in\mathcal{V}(\theta_0}\left\| \frac{\partial}{\partial\theta'}\left\{\frac{\partial^2}{\partial\theta_i\partial\theta_j} l_t(\theta)\right\}\right\| \\
&= E_{\theta_0} \sup_{\theta\in\mathcal{V}(\theta_0}\left\| \frac{\partial}{\partial\theta'}\left\{\frac{\partial^2}{\partial\theta_i\partial\theta_j} l_t(\theta)\right\}\right\| < \infty.
\end{aligned}$$

Puisque $\left\|\theta^*_{ij} - \theta_0\right\| \to 0$ presque sûrement, le second terme du membre de droite de (3.52) converge vers 0 avec probabilité 1. Pour achever la preuve de ce théorème il suffit d'appliquer le lemme de Slutsky. Par d), e) et f) nous obtenons (3.26) et (3.27). ■

Remarque 96 1. *L'hypothèse **A5** est classique car elle permet d'utiliser le fait que les conditions du premier ordre sont valides, au moins asymptotiquement. En effet si $\widehat{\theta}_n$ est convergent, il appartient également à l'intérieur de Θ pour n grand. En tant que maximum, il doit donc annuler la dérivée de la fonction critère. Cette hypothèse est cependant restrictive car elle exclut par exemple le cas $\alpha_{01} = 0$.*

2. *Lorsque certaines composantes de θ_0 sont nulles, **A5** n'est pas vérifiée et le théorème ne permet pas de conclure. Il est clair que, dans ce cas, la distribution asymptotique de $\sqrt{n}(\widehat{\theta}_n - \theta_0)\sqrt{n}(\widehat{\theta}_n - \theta_0)$ ne peut être normale car l'estimateur est contraint. Si par exemple $\alpha_{01} = 0$, la loi de $\sqrt{n}(\widehat{\alpha}_1 - \alpha_{01})$ est, pour tout n, concentrée sur $[0, \infty[$ et ne peut donc être asymptotiquement normale. Ce type de problèmes, dits "de bord".*

3.2.2 Le cas $ARCH(1)$: évaluations numérique de la variance asymptotique

Considérons le modèle $ARCH(1)$

$$\varepsilon_t = \{\omega_0 + \alpha_0 \varepsilon_{t-1}^2\}^{\frac{1}{2}} \eta_t,$$

avec $\omega_0 > 0, \alpha_0 > 0$ et supposons les variables η_t iid de variance 1 et vérifiant **A3**. Le paramètre est ici $\theta = (\omega, \alpha)'$. La contrainte de stationnarité stricte, **A2**, s'écrit

$$\alpha_0 < \exp\{-E(\log \eta_t^2)\}.$$

Afin que l'hypothèse **A1** soit vérifiée, on peut prendre un espace des paramètres de la forme $\Theta = [\delta, \frac{1}{\delta}] \times [0, \frac{1}{\delta}]$ où $\delta > 0$ est une constante suffisamment petite pour que la vraie valeur $\theta_0 = (\omega_0, \alpha_0)'$appartienne à Θ. L'estimateur du QMV de θ est alors fortement convergent d'après le théorème 93.

Puisque $\partial \widetilde{\sigma}_t^2 / \partial \theta = (1, \varepsilon_{t-1}^2)'$, l'estimateur du QMV $\widehat{\theta}_n = (\widehat{\omega}_n, \widehat{\alpha}_n)'$ est caractérisé par les équations normales

$$\frac{1}{n} \sum_{t=1}^{n} \frac{\varepsilon_t^2 - \widehat{\omega}_n - \widehat{\alpha}_n \varepsilon_{t-1}^2}{(\widehat{\omega}_n + \widehat{\alpha}_n \varepsilon_{t-1}^2)^2} \begin{pmatrix} 1 \\ \varepsilon_{t-1}^2 \end{pmatrix} = 0, \tag{3.52}$$

en prenant par exemple $\varepsilon_0^2 = \varepsilon_1^2$. Cet estimateur n'a pas de forme explicite et doit être obtenu numériquement pour un échantillon donné. L'application du théorème 95 donnant la loi asymptotique de l'estimateur ne nécessite comme seule hypothèse supplémentaire que θ_0 appartienne à $\Theta^\circ =]\delta, 1/\delta[\times]0, 1/\delta[$. Ainsi, si $\alpha_0 = 0$ (modèle conditionnellement homoscédastique), l'estimateur reste convergent mais il ne peut être asymptotiquement normal. La matrice J prend la forme suivante

$$J = E_{\theta_0} \begin{bmatrix} \frac{1}{(\omega_0 + \alpha_0 \varepsilon_{t-1}^2)^2} & \frac{\varepsilon_{t-1}^2}{(\omega_0 + \alpha_0 \varepsilon_{t-1}^2)^2} \\ \frac{\varepsilon_{t-1}^2}{(\omega_0 + \alpha_0 \varepsilon_{t-1}^2)^2} & \frac{\varepsilon_{t-1}^4}{(\omega_0 + \alpha_0 \varepsilon_{t-1}^2)^2} \end{bmatrix}, \tag{3.53}$$

et la variance asymptotique de $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ est

$$Var_{as}\{\sqrt{n}(\widehat{\theta}_n - \theta_0)\} = \mathcal{K}_\eta - 1)J^{-1}.$$

3.2.3 Le cas non stationnaire : exemple de l'$ARCH(1)$

Lorsque la contrainte de stationnarité stricte n'est pas vérifiée, c'est-à dire dans le cas $ARCH(1)$ lorsque

$$\alpha_0 \geq \exp\{-E\log\eta_t^2\}, \tag{3.54}$$

on peut définir un processus $ARCH(1)$ en l'initialisant. Pour une valeur 0 donnée, on définit

$$\varepsilon_t = h_t^{\frac{1}{2}}\eta_t, h_t = \omega_0 + \alpha_0\varepsilon_{t-1}^2, \quad t = 1, 2, ... \tag{3.55}$$

avec $\omega_0 > 0, \alpha_0 > 0$ avec les hypothèses usuelles sur la suite (η_t). Comme nous l'avons déjà remarqué, σ_t^2 converge vers l'infini, presque sûrement lorsque

$$\alpha_0 \geq \exp\{-E\log\eta_t^2\}, \tag{3.56}$$

et seulement en probabilité lorsque l'inégalité (3.55) est une égalité.

Est-il possible d'estimer les coefficients d'un tel modèle ? La réponse est seulement partiellement positive : il est possible d'estimer de manière convergente le coefficient α_0 mais sans doute pas le coefficient ω_0. La portée pratique de ce résultat est donc limitée mais l'intérêt théorique qu'il comporte mérite qu'on s'y attarde.

Considérons l'estimateur QMV d'un $ARCH(1)$, c'est à dire une solution mesurable de

$$(\widehat{\omega}_n, \widehat{\alpha}_n) = \arg\min_{\theta\in\Theta} \frac{1}{n}\sum_{t=1}^{n} l_t(\theta), \; l_t(\theta) = \frac{\varepsilon_t^2}{\sigma_t^2(\theta)} + \log\sigma_t^2(\theta), \tag{3.57}$$

où $\theta = (\omega, \alpha), \Theta$ est un compact de $(0, \infty)^2$, est $\sigma_t^2(\theta) = \omega + \alpha\varepsilon_{t-1}^2$ pour tout $t = 1, ..., n$ (en fixant une valeur initiale pour ε_0^2). La convergence presque sûre de ε_t^2vers l'infini va nous permettre de montrer la convergence forte de l'estimateur QMV.

Conclusion

Dans ce mémoire, on a étudié deux modèles de séries chronologiques, le modèle $ARCH(p)$ et le modèle $GARCH(p,q)$. On a montré que les deux modèles peuvent s'écrire sous forme d'équation aux différences stochastique bilatérale et multi-dimensionnelle de type $X_n = A_n X_{n-1} + B_n, n \in \mathbb{Z}$ où $(A_n)_{n\in\mathbb{Z}}$ est une suite de matrices aléatoires iid et $(B_n)_{n\in\mathbb{Z}}$ est une suite de vecteurs aléatoires iid. On a montré que la condition nécessaire et suffisante pour que les deux modèles admettent des solutions strictement stationnaires et ergodiques est que l'exposant de Lyapunov associé à la suite $(A_n)_{n\in\mathbb{Z}}$ soit strictement inférieur à zéro. Par la suite, on a estimé les paramètres des deux modèles par la méthode de quasi-maximum de vraisemblance (QMV) et on a montré la normalité asymptotique de l'éstimateur de QMV. On a montré aussi la convergence presque sûrement de l'estimateur de QMV vers le vecteur des vraies valeurs des paramètres.

Bibliographie

[1] Bollersev, T. (1986) Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econotrics, 31, 307-327.

[2] Bougerol, P., and Lacroix, J. (1985). Products of Random Matrices with applications to Schrodinger Operators. Progress in Probility and Statistics Vol. 8., Boston.

[3] Bougerol, P., and Picard, N. (1992a). Strict stationnarity of genelized autoregressive processes. The Annals of Probability, 20, 1717-1730.

[4] Bougerol, P., and Picard, N. (1992b). Stationnarity of GARCH processes and some nonegative times series. Journal of Econometrics, 52, 115-127.

[5] Brieman, L. (1968) . Probability. Addison-Welsley Publishing Company, Reading, Massachusetts.

[6] Davis, R. A., and Mikodch, T. (1998) . The sample autocorrelations of heavy-tailed processes with applications to ARCH. Annals of statistics, 26, 2049-2080.

[7] de Saporta, B. (2005) . Etude de la solution de l'équation $Y_{n+1} = A_n Y_n + B_n$ à n coefficients aléatoires. PhD thesis, Université de Rennes 1.

[8] Doob, J.L. (1953) . Stochastic Processes. John Wiley, New York.

[9] Engle, R. F. (1982) . Autoregressive conditional heteroscedasticity with estimates of variance of United Kingdom inflation. Econometrica, 50, 987-1008.

[10] Feller, W . (1950) . Intoduction to Probability Theory, vol. I. John Wiley, New York.

[11] Francq, C., and Zakoïan, J, M. (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. Bernoulli, 10, 605-637.

[12] Francq, C., and Zakoïan, J, M. (2010) GARCH Models, Structure, Statistical inference and financial applications. John Wiley and Sons, Ltd, Publication

[13] Karlin, S. (1975) . A First Course in Stochastic Processes, 2nd ed. Academic Press, New York.

[14] Mikosch, T., and Starica, C. (2000) Limit theory for the sample autocorrelations and extrems of a GARCH(1,1) process. Annals of statistics, 28, 1427-1451.

[15] Nelson, D. B. (1990) Stationnarity and Presistence in the GARCH(1,1) Model. Econometric Theory, 6, 318-334.

[16] Vervaat, W. (1979). On a stochastic difference equation and representation of non-negative infinitly divisible random variables. *Advanced probability*, 11, 750-783.