

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministre de L'Enseignement Supérieur et de La Recherche Scientifique

Université Mohamed Khider - Biskra
Faculté des Sciences et Sciences de l'Ingénieur
Département de Mathématiques



MEMOIRE

Présenté pour Obtenir le Diplôme de Magister en Mathématiques

Option: *Analyse & Modèles Aléatoires*

THEME

SERIES TEMPORELLES ET TEST D'ADEQUATION POUR UN MODELE GARCH (1, 1)

Présenté par : **YAHIA DJABRANE**

Devant le Jury :

<i>Président :</i>	K. BOUKHETALA	PR	U. S. T. H. B. Alger
<i>Rapporteur :</i>	A. NECIR	MC	U. BISKRA
<i>Examineur :</i>	B. MEZERDI	PR	U. BISKRA
<i>Examineur :</i>	Z. MOHDEB	PR	U. M. Constantine

Soutenu le : 28/02/2005

Remerciements



En cette belle et heureuse occasion.

*Je tiens à remercier chaleureusement mon directeur de recherche Monsieur **Necir Abdelhakim**, Maître de conférence à l'université de Biskra, pour ses conseils bénéfiques et son apport précieux tout au long de la réalisation de cette thèse. Il n'a ménagé aucun effort pour m'accompagner dans ma recherche et m'a permis, par sa rigueur et ses connaissances, d'améliorer mes aptitudes dans ce domaine...*

Comme je suis également très reconnaissant aux membres du Jury pour avoir accepté de participer à ma soutenance...

*Je remercie infiniment Monsieur **Boukhetela Kamel**, Professeur à l'U.S.T.H.B. d'Alger de me faire l'honneur de présider le Jury de mon mémoire de magistère.*

*Mes remerciements vont également à Monsieur **Mohdeb Zaher**, Professeur à l'université de Constantine, pour avoir accepté d'examiner ce travail et à mon enseignant Monsieur **Mezerdi Brahim**, Professeur à l'université de Biskra qui m'a honoré par sa présence dans ce Jury.*

Mes remerciements sont adressées également aux membres du département de Mathématiques : Enseignants, Etudiants, Collègues, ...

Enfin, je tiens à exprimer toute ma gratitude à mes parents et à toute ma famille pour leur confiance sans faille en mes capacités et leur soutien, durant mon cursus d'étude.

A mon meilleur ami Guettouchi Nadjib.

Résumé

Nous présentons dans ce mémoire un aperçu sur la théorie des séries temporelles. Nos objectifs sont : Premièrement, l'étude du comportement asymptotique des valeurs extrêmes dont l'hypothèse d'indépendance et d'équidistribuité est remplacé par des conditions sur la stationnarité, la dépendance et l'hétéroscédasticité. En particulier, ces conditions sont satisfaisantes par les séries temporelles financières: taux de change, rendements d'indice boursier,... etc.

Deuxièmement, résoudre le problème d'adéquation de ces séries pour les modèles non linéaires de type GARCH. Les techniques utilisées pour aborder ce genre de problème se basent sur la théorie spectrale, dont le principe consiste à vérifier l'ajustement de la fonction de distribution spectrale empirique à celle du modèle.

Abstract

We present in this memory an outline on the theory of time series. Our objectives are: Firstly, the study of asymptotic behavior of extreme values, of which the independence and identically distributed hypothesis are replaced by conditions on the stationnarity, dependence and heteroskedasticity. In particular, these conditions are satisfactory by some financial time series: exchange rates, returns of stock indices... etc.

Secondly, is to solve the goodness-of-fit problems for this time series for non linear GARCH-type models. The techniques used to approach this kind of problem are based on a spectral theory, whose principle consists to verify the ajustement between the empirical spectral distribution function and this of the model.

.GARCH

Table des matières

0.1	Introduction	03
0.2	Notations	07
Chapitre 1 : Séries temporelles		08
1.1	Analyse des séries temporelles	08
1.1.1	Modélisation d'une série temporelle.....	09
1.1.2	Modèle stationnaire.....	10
1.2	Introduction à la théorie spectrale	13
1.3	Moments empiriques	16
1.4	Représentation des séries temporelles	19
1.4.1	Modèles : AR, MA et ARMA.....	19
1.4.2	Modèles : ARCH, GARCH.....	21
Chapitre 2 : Séries temporelles et valeurs extrêmes		25
2.1	Valeurs extrêmes et indépendance	26
2.2	Valeurs extrêmes et dépendance	31
2.2.1	Mesure de dépendance.....	31
2.3	Etude d'index des queues de distribution	33
2.3.1	Consistance de l'estimateur de l'index.....	34
2.4	Valeurs extrêmes et hétéroscédasticité	40
2.5	Normalité asymptotique et mélangeance	44

Chapitre 3 : Adéquation au modèle GARCH(1, 1)	52
3.1 Construction du modèle	53
3.1.1 Relations entre : GARCH, AR et ARMA.....	56
3.1.2 Propriétés des distributions du modèle.....	58
3.1.3 Estimation des paramètres du modèle.....	60
3.2 Test d'adéquation au modèle GARCH	64
3.2.1 Notion d'adéquation.....	64
3.2.2 Construction du test d'adéquation.....	67
 Chapitre 4 : Etude empirique d'une série financière	 75
4.1 Propriétés statistiques de la série	76
4.1.1 Séries des prix et des rendements journaliers.....	76
4.1.2 Moments empiriques.....	78
4.1.3 Hétéroscédastécité et autocorrélation.....	80
4.2.4 Test d'hypothèse de normalité.....	84
4.2 Modélisation GARCH du S&P 500	85
4.2.1 Identification <i>a priori</i> du modèle.....	85
4.2.2 Forme explicite du modèle.....	86
 Bibliographie	 89

0.1 Introduction

L'analyse des séries temporelles, ou séries chronologiques, se réfère à la branche de statistique où les observations sont régulièrement espacées dans le temps. Leur domaine d'application est très vaste et s'étend de l'astronomie à l'économie et finance en passant par la biologie, psychologie, géophysique ou la théorie du signal...etc. Elles ont donc suscité un très vif intérêt, ce qui a eu par conséquence le développement de nombreux modèles : AR, MA, ARMA, ARCH et GARCH....

Historiquement, c'est dans les XVI^e - XVII^e siècles qu'apparaissent les premières séries temporelles (ou chronologiques) dans les écrits scientifiques. Mais le plus ancien graphique connu remonte à l'antiquité (en terme de représentation de l'inclinaison des orbites des planètes en fonction du temps), les Romains savaient déjà que l'année dure 365 et 1/4 jours.

L'astronome *T. Brahe* (1546- 1601) enregistre des données nombreuses et précises sur les orbites des planètes (1572). Son assistant *J. Kupler* (astronome et mathématicien Allemand, 1571-1630) a utilisé ces données pour formuler les lois du mouvement des planètes qui portent, ainsi son nom (1609).

Ces travaux sont restés dans le cadre d'analyse astronomique, jusqu'au XVIII^e siècle. En 1786, l'économiste anglais *W. Playfair*, publie "*The Commercial & Political Atlas*", qui contient plus de 40 graphiques de données économiques.

En 1807, *Joseph Fourier* (Mathématicien français, 1768-1830), propose que toute fonction périodique peut s'écrire comme une série de fonctions sinusoïdales et cosinusoïdales (i.e. sin et cos). Ce qui a fait ensuite, de la transformation de Fourier la base de la théorie spectrale.

Au XIX^e siècle, les premières études empiriques des données économiques sont parues dans l'analyse des séries de prix, de taux d'intérêt,... En 1919, *W. Person*, propose une approche globale de décomposition d'une série temporelle en quatre composantes (i.e. Tendances, Saisonnière, Cycle et Résidu), connue sous le nom de (Baromètre de Harvard).

Dans les années 20, *Slytsky* (1927 - 1933) montre qu'en calculant une moyenne mobile (MA) à partir d'un (bruit blanc), on obtient une série dont les observations ne sont pas indépendantes et qui présentent des cycles apparents "*Effet Slytsky*". Et dans le même ordre d'idées, en (1926-27) *Yule* propose le modèle autorégressif (AR), et montre que celui-ci peut conduire à l'apparition des fluctuations cycliques.

Suite aux recherches sur les fluctuations économiques, et au développement de la statistique mathématique, les chercheurs développent une approche combinant plusieurs éléments :

- L'estimation basée sur la théorie économique,
- L'utilisation de la statistique dans l'analyse des données,
- L'utilisation des modèles économétriques pour analyser les politiques.

Après le développement de la théorie des processus stationnaires et l'introduction de la notion de spectre d'une série stationnaire durant les années (1925-45), l'analyse des séries temporelles se développe rapidement et les applications se multiplient, notamment en : Physique, Génie électrique, Médecine, Biologie, Géophysique et principalement en Economie et Finance...

Jusqu'à la fin des années 60, cette recherche sera dominée par l'étude du modèle de régression. Ce n'est qu'en 1970 qu'apparaissent en effet, l'adéquation au modèle (ARIMA) et la prévision (Approche de *Box et Jenkins*).

Suite aux travaux pionniers de *Engle* en 1982, il est devenu courant dans la littérature financière de retenir comme mesure de volatilité (i.e. variance conditionnelle) la variance d'un modèle "Autorégressive Conditionnellement Hétéroscidastique" : ARCH, qui est une généralisation de l'approche de *Box et Jenkins*. En outre, c'est le passage des modèles linéaires vers les modèles non linéaires. Ce qui permet d'intégrer des propriétés observées empiriquement, telles que la dépendance entre les observations ou la distribution à queue lourde...

Au milieu des années quatre vingt *Bollerslev*, alors étudiant de *Engle*, généralise la famille des modèles ARCH en permettant à la variance conditionnelle du modèle de dépendre aussi de son propre passé. *Bollerslev* définit donc en 1986 le modèle Autorégressif Conditionnellement Hétéroscédastique Généralisés, c'est-à-dire GARCH, qui fera l'objet principal de cette thèse.

La disponibilité d'un grand nombre de données temporelles, dans ces dernières années a permis non seulement de rechercher des dynamiques non linéaires, mais aussi de s'intéresser aux queues de distribution, c'est-à-dire aux risques extrêmes. En effet, la théorie des valeurs extrêmes est l'une des avancées majeures dans la modélisation des phénomènes scientifiques dépendant du temps, si on prend en compte les diverses caractéristiques des données observées telles que : l'invariance par translation du temps (Stationnarité), la dépendance et la linéarité ...

Le présent mémoire, est alors une synthèse des travaux de recherches concernant la théorie des séries temporelles, ainsi que des différentes propriétés probabilistes et statistiques du modèle GARCH(1,1), en apportant des réponses à certaines questions notamment :

- Quels sont les objectifs et comment se fait la modélisation des séries temporelles ?

- Que deviendra la théorie des risques où celle de l'assurance dans l'étude des valeurs extrêmes en rejetant les hypothèses classiques d'indépendance, d'iqui-distribueté (i.i.d) et de gaussienté ?

- Comment choisir un modèle qui presente le mieux les données (i.e. problème d'adéquation) ?

L'organisation de la thèse est la suivante :

Chapitre 1 : " Séries temporelles "

Tout d'abord, nous rappelons les principaux définitions, notations et résultats que nous verrons souvent dans la théorie des séries temporelles, et que l'on va utiliser par la suite, notamment : les modèles AR, MA, ARMA et les modèles hétéroscédastiques ARCH et GARCH. Ainsi aurons-nous à étudier la théorie spectrale, les fonctions d'autocovariance et d'autocorrélation, la stationnarité et les moments d'échantillonnage,...

Chapitre 2 : " Séries temporelles et valeurs extrêmes "

Le deuxième chapitre est consacré à l'étude des valeurs extrêmes, dont l'objectif est de mesurer en terme de probabilité l'intervention des événements rares, ou l'apparition des catastrophes : grands sinistres, orages, et crises économiques, . . . etc.

Ce qui rend la connaissance des caractéristiques des distributions et propriétés (comportements) des estimateurs de ces événements, en particulier la probabilité de dépasser un certain seuil (i.e.valeurs extrêmes) une préoccupation majeure des assureurs dans les compagnies d'assurance et un excellent outil d'aide à la décision, pour éviter les faillites retentissantes ou les situations des ruines...

Presque tous les résultats sur le comportement asymptotique des quantiles extrêmes disponibles maintenant étant restreint à des observations indépendantes dans l'étude des séries temporelles financières en particulier, ce chapitre a donc pour but d'enquêter sur le comportement asymptotique : la consistance et la normalité asymptotique de l'estimateur des quantiles extrêmes, basées sur les grandes observations.

Après avoir rappelé certains résultats établis dans le cas d'indépendance, nous allons surtout nous intéresser au cas où se pose les problèmes de dépendance, de stationnarité et d'hétérosédasticité dans l'étude des séries temporelles.

Chapitre 3 : " Test d'adéquation au modèle GARCH(1, 1)"

L'un des problème qui se pose, lorsque on veut appliquer les divers résultats établis dans le chapitre précédent à des séries temporelles réelles (séries financier en particulier), est comment assurer l'ajustement de ces données au modèle ?

Ce chapitre donc a pour but de résoudre ce problème, du moins pour le modèle GARCH (1,1). Mais, le principe reste neanmoins le même soit pour le cas général GARCH(p, q), soit pour un ARCH(p) ou ARCH(1) qui sont des cas particulier des premiers... Nous commençons par la construction d'un modèle GARCH(1,1), puis nous établissons ses relations avec les autres modèles et en particulier les modèles AR et ARMA. En suite on étudiera les propriétés des distributions conditionnelles et non conditionnelles du modèle, l'estimation de ces paramètres en nous rappelant quelques notions d'adéquations. Ce qui permet d'aborder le problème d'adéquation dans le domaine des fréquences pour un modèle GARCH(1. 1).

Chapitre 4 : "Etude empirique d'une série temporelle financière"

Enfin, dans le dernier chapitre nous proposons une étude empirique sur la dynamique d'une série temporelle financière (série des rendements journaliers de l'indice boursier S&P500). Dans une première section ,nous présenterons les diverses propriétés statistiques, caractéristiques et comportements de cette série, on les comparant avec celles d'un modèle de type GARCH.

Ce qui nous permet d'établir un modèle d'ajustement pour cette série, dont nous retiendrons la mielleure spécification.

Les données représentent les prix de fermeture quotidiens d'indice S&P 500, de la période allant du 03 Janvier 1994 au 13 Mai 2004, soit peut plus de 2610 observations. Il sont extraites du cite Internet :

"[http :// www.chart2.finance.dcn.yahoo.com](http://www.chart2.finance.dcn.yahoo.com)".

0.2 Notations

(Ω, F, P)	Espace de probabilité.
$\mathbb{E}[\cdot]$	Espérance mathématique (Moyenne).
$var(\cdot)$	Variance.
$F(\cdot)$	Fonction de distribution.
$F_n(\cdot)$	Fonction de distribution empirique.
$\bar{F}(\cdot)$	Fonction des queues de distribution.
$Q(\cdot)$	Fonction des quantiles.
$Q_n(\cdot)$	Fonction des quantiles empiriques.
$\gamma_X(\cdot)$	Fonction d'autocovariance de X .
$\rho_X(\cdot)$	Fonction d'autocorrélation de X .
$\stackrel{\mathcal{D}}{=}$	Egalité en distribution.
$\stackrel{\mathcal{D}}{\rightarrow}$	Convergence en distribution.
$W_t(\cdot)$	Mouvement brownien (processus de Weiner).
$B_t(\cdot)$	Pont brownien.
$[\cdot]$	Partie entière.
$\mathcal{N}(0, 1)$	Loi normale centré et réduite.
$\mathbf{1}_A(\cdot)$	Fonction indicatrice de l'événement A .
$\delta_a(\cdot)$	Mass de Dirac au point a .
$\mathbb{D}[0, 1]$	Espace de Skorohod.
$(X_t)_t$	Série temporelle.
\mathcal{F}_t	Filtration.
$\mathbb{E}[\cdot \mathcal{F}_t]$	Espérance conditionnelle par rapport à \mathcal{F}_t .

Chapitre 1

Séries temporelles

La théorie des séries temporelles est une combinaison de deux concepts, probabiliste et statistique, le probabiliste dont on étudie les caractéristiques des variables aléatoires X_t . Le problème statistique est de donner les caractéristiques des distributions de la série temporelle X_t , pour les observations X_1, X_2, \dots, X_n au temps $t = 1, 2, \dots, n$. Le modèle statistique résultant sert à la compréhension du système stochastique d'une part et la prédiction du future (i.e. X_{n+1}, X_{n+2}, \dots) d'autre part.

1.1 Analyse des séries temporelles

Le terme "*série temporelle*" désigne à la fois les séries chronologiques réelles et une suite théorique des variables aléatoires indécis par le temps ($t \in \mathbb{T}$), qui va servir à modéliser ces premières.

Définition 1.1.1 : "*Série temporelle*"

Une série temporelle est une suite d'observations répétées, correspondant à des dates différentes, où encore à un ensemble de valeurs représentant l'évolution d'un phénomène au cours du temps.

Généralement, les observations d'un phénomène sont équidistantes, les unes des autres (temps discret; $t \in \mathbb{N}, \mathbb{Z}, \dots$), le temps correspondant à un jour, un mois, une année.... Si par exemple on travaille dans des domaines tels que les finances, on peut citer entre autres : la valeur journalière du Dollar (\$) en Euro (€), à l'ouverture du marché boursière, les données mensuelles du chômage, les prix d'action, relevé annuel de la production d'acier sur 30 ans,...etc. Mais il existe d'autres domaines (comme en Physique), où les observations sont élevées de façon continue, l'indice t prend des valeurs dans une intervalle de \mathbb{R} .

1.1.1 Modélisation d'une série temporelle

La modélisation des séries temporelles se fait à partir de la décomposition classique, "*décomposition de Persons*" en fonction des quatre éléments suivants :

1. Tendance (T_t) : mouvement à long terme (longue période).
2. Saisonnière (S_t) : fonction périodique du temps (période courte).
3. Cycle (C_t) : cycle d'affaires, fluctuation périodique (moyenne terme).
4. Résidu (R_t) : partie irrégulière, correspondante à la notion d'écart au modèle ou encore bruit.

D'une manière générale, on peut proposer un modèle qui représente la série temporelle étudiée en combinaison des quatre éléments précédents. Pour cela, on a trois types de modèles : le premier est le modèle d'ajustement de forme additive ou multiplicative comme suit :

$$X_t := T_t + S_t + C_t + R_t \quad \text{ou} \quad X_t := T_t S_t C_t + R_t.$$

Le deuxième type est le modèle autoprojectif, dont on suppose que X_t est une fonction de ces valeurs passées et d'une perturbation aléatoire R_t :

$$X_t := f(X_{t-1}, X_{t-2}, \dots; R_t),$$

dans cette classe, on peut citer les modèles *ARIMA*, *ARMA*,...

Le troisième type est le plus important, c'est le modèle explicatif. Dans cette catégorie de modèle, la variable aléatoire X_t est exprimée en fonction d'un vecteur aléatoire Y_t et d'une perturbation aléatoire R_t :

$$X_t := f(Y_t, R_t),$$

où Y_t est soit déterministe ou aléatoire, dans ce dernier cas les processus $(Y_t)_t$ et $(R_t)_t$ ont certaines propriétés d'indépendance ou du non-corrélation.

Ces modèles sont les modèles de base de l'économétrie et nous les considérons essentiellement pour faire le lien entre eux et les modèles autoprojectifs. On a ainsi deux cas particuliers de modèle explicatif, modèle explicatif statique où les variables Y_t ne contiennent pas de valeurs passées de X_t et les R_t sont indépendantes entre eux. Le deuxième cas est le modèle explicatif dynamique, où les R_t sont autocorrélés et les Y_t contiennent des valeurs passées de X_t .

1.1.2 Modèle stationnaire

La représentation des phénomènes aléatoires dépendants du temps est faite de la manière suivante :

Soient (Ω, \mathcal{F}, P) un espace de probabilité, \mathbb{T} un ensemble non vide d'indice (par exemple : $\mathbb{N}, \mathbb{Z}, \mathbb{R}^+, \dots$), X_t une fonction de $\mathbb{T} \times \Omega$ à valeurs dans \mathcal{E} qui associé à tout couple (t, ω) le processus $X_t(\omega)$, avec \mathcal{E} désignant l'espace d'états du processus. D'où :

- i- pour $t \in \mathbb{T}$, fixé : $X_t(\omega)$ est une variable aléatoire.
- ii- pour $\omega \in \Omega$, fixé : $X_t(\omega)$ est une trajectoire.

Définition 1.1.2 : "Processus stochastique "

Un processus stochastique défini sur \mathbb{T} noté $(X_t(\omega))_{t \in \mathbb{T}}$ ou simplement $(X_t)_t$, est une collection de variable aléatoire X_t de Ω à valeurs dans \mathbb{R} , de telle manière qu'à chaque élément $t \in \mathbb{T}$ est associée une variable aléatoire X_t . On a ainsi deux cas :

- i- un processus en temps discrète si \mathbb{T} est discrète ; ($\mathbb{T} \subseteq \mathbb{Z}$).
- ii- un processus en temps continue si \mathbb{T} est continue ; ($\mathbb{T} \subseteq \mathbb{R}$).

Par conséquence, on s'intéresse aux modèles stochastiques, dont les éléments X_t de la série temporelle $(X_t)_t$ sont considérés comme des variables aléatoires. Par la suite on désigne par un modèle, le processus stochastique qui modélise la série temporelle.

Généralement, les variables d'une série $(X_t)_t$ ne sont ni indépendantes, ni identiquement distribuées. Les moyennes, variances et covariances de ces variables dépendent de leurs positions dans la série. En particulier, si on suppose que $(X_t)_t$ est de carré intégrable (i.e. $\mathbb{E}[X_t^2] < \infty, \forall t \in \mathbb{T}$), alors :

$$\begin{aligned} \mathbb{E}[X_t] &= \mu_t, \quad \text{var}(X_t) = \sigma_t^2, \\ \text{Cov}(X_t, X_{t+k}) &= \mathbb{E}[(X_t - \mu_t)(X_{t+k} - \mu_{t+k})], \quad k \in \mathbb{Z}. \end{aligned}$$

Définition 1.1.3 : "Fonction d'autocovariance "

La fonction d'autocovariance d'une série temporelle $(X_t)_t$ est une suite $(\gamma(h))_{h \in \mathbb{Z}}$, avec

$$\gamma(h) := \text{Cov}(X_t, X_{t+h}),$$

c'est une fonction paire, semi-définie positive, i.e.

$$\begin{aligned} \sum_{i=1}^N \sum_{j=1}^N a_i a_j \gamma(t_i - t_j) &\geq 0, \\ |\gamma(h)| &\leq \gamma(0) = \text{var}(X_t), \quad h \in \mathbb{Z}. \end{aligned}$$

Définition 1.1.4 : "*Fonction d'autocorrélation*"

De même, on définit une suite $(\rho(k))_{k \in \mathbb{Z}}$ qu'on l'appelle fonction d'autocorrélation de la série $(X_t)_t$:

$$\rho_k := \frac{Cov(X_t, X_{t+k})}{\sqrt{var(X_t)var(X_{t+k})}} = \frac{\gamma(k)}{\gamma(0)},$$

c'est une fonction paire, semi-définie positive, i.e.

$$\begin{aligned} \sum_{i=1}^N \sum_{j=1}^N a_i a_j \rho(t_i - t_j) &\geq 0, \\ |\rho(k)| &\leq \rho(0) = 1, \quad k \in \mathbb{Z}. \end{aligned}$$

L'une des propriétés importantes dans l'étude des séries temporelles est la stationnarité, et il existe deux types de stationnarité comme suit :

Définition 1.1.5 : "*Série fortement stationnaire*"

Une série temporelle $(X_t)_t$ est fortement (ou strictement) stationnaire, si et seulement si :

$$(X_{t_1}, X_{t_2}, \dots, X_{t_n}) \stackrel{\mathcal{D}}{=} (X_{t_1+k}, X_{t_2+k}, \dots, X_{t_n+k}). \quad (1.1.1)$$

On dit alors, que la loi de vecteur $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ est identique à celle de $(X_{t_1+k}, X_{t_2+k}, \dots, X_{t_n+k})$ pour tout sous-ensemble $\{t_1, t_2, \dots, t_n\} \subseteq \mathbb{T}, \forall k \in \mathbb{N}$.

Définition 1.1.6 : "*Série faiblement stationnaire*"

Une série temporelle $(X_t)_t$ est faiblement stationnaire (ou stationnaire du second ordre), si et seulement si :

1. $\mathbb{E}[X_t^2] < \infty, \quad \forall t \in \mathbb{T},$
2. $\mathbb{E}[X_t] = \mathbb{E}[X_s] = \mu, \quad \forall s, t \in \mathbb{T},$
3. $Cov(X_s, X_t) = Cov(X_{s+k}, X_{t+k}), \quad \forall s, t \in \mathbb{T}, \quad \forall k \in \mathbb{N}.$

L'existence de la fonction d'autocovariance d'une série temporelle stationnaire (on dit souvent stationnaire au lieu de faiblement stationnaire) est assurée par la proposition suivante :

Proposition 1.1.1 :

Si la série temporelle $(X_t)_t$ est stationnaire, alors il existe une fonction $\gamma_X : \mathbb{Z} \rightarrow \mathbb{R}$, telle que les autocovariances ne dépendent que de la différence entre les observations :

$$Cov(X_s, X_t) = \gamma_X(|t - s|), \quad \forall s, t \in \mathbb{Z}. \quad (1.1.2)$$

Preuve :

Soit $r \in \mathbb{Z}$, comme $(X_t)_t$ est stationnaire, donc pour tout $s, t \in \mathbb{Z}$:

1^{ère} cas : Si $s \leq t$

$$\begin{aligned} \text{Cov}(X_s, X_t) &= \text{Cov}(X_{s+r-r}, X_{t+s-s+r-r}) \\ &= \text{Cov}(X_r, X_{r+t-s}), \quad \text{si } s \geq r. \end{aligned}$$

De plus,

$$\begin{aligned} \text{Cov}(X_s, X_t) &= \text{Cov}(X_{s+r-s}, X_{t+r-s}) \\ &= \text{Cov}(X_r, X_{r+t-s}), \quad \text{si } s < r. \end{aligned}$$

2^{ème} cas : Si $s > t$

$$\text{Cov}(X_s, X_t) = \text{Cov}(X_t, X_s) = \text{Cov}(X_r, X_{r+s-t}),$$

d'où $\forall s, t \in \mathbb{Z}$:

$$\text{Cov}(X_s, X_t) = \text{Cov}(X_r, X_{r+|t-s|}) = \gamma_X(|t-s|). \quad \blacksquare$$

Remarques 1.1.1 :

1. Si $\forall t \in \mathbb{T}$, $(X_t)_t$ est de carré intégrable, la stationnarité forte implique alors la stationnarité faible.
2. Si la série $(X_t)_t$ est gaussienne¹, la stationnarité faible implique alors la stationnarité forte.
3. Par la suite, on désigne par un bruit blanc la suite $(\xi_t)_t$ de variables aléatoires i.i.d, centrée et de variance σ^2 ($\text{var}(\xi_t) = \mathbb{E}[\xi_t^2] = \sigma^2$). Dans ce cas, il est clair qu'un bruit blanc est un processus stationnaire du second ordre, avec :

$$\gamma_\xi(0) = \sigma^2 \quad \text{et pour tout } k \neq 0, \quad \gamma_\xi(k) = 0. \quad (1.1.3)$$

¹Un processus $(X_t)_t$ est gaussien, si toute combinaison linéaire finie de $(X_t)_t$, i.e : $\forall \lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{R}$ et $\forall t_1, t_2, \dots, t_n \in \mathbb{T}$: $\sum_{j=1}^n \lambda_j X_{t_j}$ suit une loi gaussienne.

1.2 Introduction à la théorie spectrale

Nous allons dans cette section introduire la notion de la densité spectrale, qui a *posteriori* contient la même information que la fonction d'autocovariance, mais elle a comme avantage les propriétés d'inversibilité et de périodicité, ce qui permet (en particulier dans le cas des tests) d'étudier et d'utiliser l'analyse de Fourier (transformation de Fourier)...

Soit $(X_t)_t$ une série temporelle stationnaire, de fonction d'autocovariance γ . On utilise les principes généraux de la transformation de Fourier dont nous verrons tout à l'heure l'intérêt. On démontrera qu'à la suite des autocovariances $(\gamma_h)_h$ on peut toujours associer une fonction réelle $F_X(\lambda)$ sur $]-\pi, \pi]$, de la variable numérique λ telle que :

$$\gamma(h) = \int_{-\pi}^{\pi} \exp(-ih\lambda) dF_X(\lambda), \quad h \in \mathbb{Z}, \quad (1.2.1)$$

i désignant le nombre imaginaire $i^2 = -1$.

Cette fonction est appelée "*distribution spectrale*" de la série $(X_t)_t$, elle est croissante, impaire et si elle admet une densité par rapport à la mesure de Lebesgue, i.e.

$$F_X(\lambda) = \int_{-\infty}^{\lambda} f_X(x) dx,$$

on substituera alors $f_X(\lambda)d\lambda$ dans la formule précédente, ce qui donne :

$$\gamma(h) = \int_{-\pi}^{\pi} \exp(-ih\lambda) f_X(\lambda) d\lambda, \quad h \in \mathbb{Z}. \quad (1.2.2)$$

Si la série de terme général $|\gamma(h)|$ est absolument convergente, ce qui permet de changer le rôle de la somme par une intégrale, dans ce cas on définit la densité spectrale comme suit :

Définition 1.2.1 : "*Densité spectrale*"

La densité spectrale notée f , d'une série temporelle $(X_t)_t$ est une fonction définie sur \mathbb{R} par :

$$f(\lambda) := \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma(h) \exp(ih\lambda), \quad \forall \lambda \in \mathbb{R}. \quad (1.2.3)$$

Propriété 1.2.1 :

La densité spectrale est une fonction paire, continue, périodique de période 2π et positive, de plus on a :

$$f(\lambda) := \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma(h) \cos(h\lambda), \quad \forall \lambda \in \mathbb{R}. \quad (1.2.4)$$

Preuve : Utilisant les propriétés (continuité, parité,...) des fonction ($\gamma(\cdot)$ et $\cos(\cdot)$) et le fait que : $\cos x = \frac{1}{2}(e^{ix} + e^{-ix})$, pour cela, **(1.2.3)** implique :

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

donc **(1.2.4)** est vérifiée. ■

Propriété 1.2.2 :

Il est équivalent de connaître la fonction d'autocovariance $\gamma(\cdot)$ ou la densité spectrale $f(\cdot)$, i.e.

$$\gamma(h) = \int_{-\pi}^{\pi} f(\lambda) \exp(-ih\lambda) d\lambda = \int_{-\pi}^{\pi} f(\lambda) \cos(\lambda h) d\lambda. \quad (1.2.5)$$

Preuve :

On a par définition, **(1.2.3)** implique que

$$\begin{aligned}
 \int_{-\pi}^{\pi} f(\lambda) \exp(-ih\lambda) d\lambda &= \int_{-\pi}^{\pi} \frac{\exp(-ih\lambda)}{2\pi} \sum_{k \in \mathbb{Z}} \gamma(k) \exp(i\lambda k) d\lambda \\
 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{k \in \mathbb{Z}} \gamma(k) \exp i\lambda(k-h) d\lambda.
 \end{aligned}$$

Donc par linéarité de l'intégrale et le fait que la série $\sum_{h \in \mathbb{Z}} |\gamma(h)|$ est convergente, alors :

$$\int_{-\pi}^{\pi} f(\lambda) \exp(-ih\lambda) d\lambda = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \gamma(k) \int_{-\pi}^{\pi} e^{-i\lambda(h-k)} d\lambda,$$

avec

$$\int_{-\pi}^{\pi} e^{-i\lambda(h-k)} d\lambda = \begin{cases} 0 & \text{si } h \neq k \\ 2\pi & \text{si } h = k, \end{cases}$$

d'où

$$\int_{-\pi}^{\pi} f(\lambda) \exp(-ih\lambda) d\lambda = \gamma(h). \quad \blacksquare$$

Propriété 1.2.3 : "Propriété d'un bruit blanc"

1. La densité spectrale d'un bruit blanc est constante en λ .
2. Tout processus stationnaire de densité spectrale constante est un bruit blanc.

Preuve :

Soit $(\xi_t)_t$ un bruit blanc i.i.d, centré et de variance $\sigma^2 < \infty$, de fonction d'autocovariance $\gamma_\xi(h)$ et de densité spectrale $f_\xi(\lambda)$.

1. Pour ce bruit blanc, on a

$$\gamma_\xi(h) = \text{cov}(\xi_t, \xi_{t+h}) = \mathbb{E} [\xi_t \xi_{t+h}] = \begin{cases} \sigma^2, & \text{si } h = 0, \\ 0, & \text{si non.} \end{cases}$$

donc

$$f_\xi(\lambda) := \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma_\xi(h) \exp(ih\lambda) = \frac{\sigma^2}{2\pi}, \quad (\text{constante en } \lambda).$$

2. Inversement, si $f_\xi \equiv C^{te}$, on a d'après la propriété (1.2.2),

$$\begin{aligned} \gamma_\xi(h) &= \int_{-\pi}^{\pi} \exp(-ih\lambda) f_\xi(h) d\lambda \\ &= \int_{-\pi}^{\pi} C^{te} \exp(-ih\lambda) d\lambda = \begin{cases} \gamma_\xi(h), & \text{si } h = 0, \\ 0, & \text{si non.} \end{cases} \end{aligned}$$

Pour que ce processus soit un bruit blanc, il suffit donc qu'il soit stationnaire; c'est le cas (par hypothèse). ■

1.3 Moments empiriques

En pratique, la moyenne, la fonction d'autocovariance, la fonction d'autocorrélation et la densité spectrale,... d'un modèle stationnaire sont inconnus et doivent être estimés à partir des observations. Cependant, il est naturel d'approcher les divers moments théoriques par les moments empiriques correspondants. Ainsi, il est important d'étudier les principales propriétés asymptotiques de ces moments...

Notons par : X_1, X_2, \dots, X_N l'ensemble d'observations disponibles (i.e. échantillon aléatoire), issu d'une série stationnaire $(X_t)_t$ de moyenne μ , l'estimateur naturel de μ est la moyenne empirique \bar{X} :

$$\hat{\mu} = \bar{X} = \frac{1}{N} \sum_{t=1}^N X_t,$$

avec

$$\mathbb{E}[\hat{\mu}] = \frac{1}{N} \sum_{t=1}^N \mathbb{E}[X_t] = \mu \text{ et } \text{var}(\bar{X}) = \mathbb{E}[(\bar{X} - \mu)^2] = \mathbb{E}[\bar{X}^2] - \mu^2.$$

Or

$$\mathbb{E}[\bar{X}^2] = \mathbb{E}\left[\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N X_i X_j\right] = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \mathbb{E}[(X_i - \mu)(X_j - \mu)] + \mu^2,$$

donc

$$\begin{aligned} \text{var}(\bar{X}) &= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \mathbb{E}[(X_i - \mu)(X_j - \mu)] \\ &= \frac{1}{N^2} \sum_{i=1}^N \mathbb{E}[(X_i - \mu)^2] + \frac{1}{N^2} \sum_{i=1}^N \sum_{j \neq i}^N \mathbb{E}[(X_i - \mu)(X_j - \mu)] \\ &= \frac{\gamma(0)}{N} + \frac{1}{N^2} \sum_{i=1}^N \sum_{j \neq i}^N \gamma(i - j) = \frac{\gamma(0)}{N} + \frac{2}{N^2} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \gamma(i - j) \\ &= \frac{\gamma(0)}{N} + \frac{2}{N^2} \sum_{i=1}^{N-1} \sum_{k=1}^{N-i} \gamma(k) = \frac{\gamma(0)}{N} + \frac{2}{N^2} \sum_{k=1}^{N-1} \sum_{i=1}^{N-k} \gamma(k) \\ &= \frac{\gamma(0)}{N} + \frac{2}{N^2} \sum_{k=1}^{N-1} (N - k) \gamma(k) = \frac{1}{N^2} \sum_{k=-N+1}^{N-1} \left(1 - \frac{|k|}{N}\right) \gamma(k), \end{aligned}$$

d'où

$$\text{var}(\bar{X}) = \sum_{k=-N+1}^{N-1} \left(\frac{N - |k|}{N^2}\right) \gamma(k).$$

Il est clair que si $\gamma(k) = 0$, $var(\bar{X}) = \frac{\gamma(0)}{N}$.

D'autre part, l'estimateur de la fonction d'autocovariance $\gamma(\cdot)$, est donné par :

$$\hat{\gamma}(k) = \frac{1}{N} \sum_{t=1}^{N-k} (X_t - \mu)(X_{t+k} - \mu), \quad (\mathbf{1.3.1})$$

ce qui permet de définir la fonction d'autocorrélation empirique comme suit :

$$\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)},$$

avec

$$\begin{aligned} \mathbb{E}[\hat{\gamma}(k)] &= \frac{1}{N} \sum_{t=1}^{N-k} \mathbb{E}[(X_t - \mu)(X_{t+k} - \mu)] \\ &= \left(\frac{N-k}{N}\right)\gamma(k) = \gamma(k) - \frac{k}{N}\gamma(k). \end{aligned}$$

Par conséquent, $\hat{\gamma}(\cdot)$ donnée par **(1.3.1)** est un estimateur biaisé mais asymptotiquement sans biais de la fonction d'autocovariance $\gamma(\cdot)$.

Définition 1.3.1 : " *Périodogramme* "

Le périodogramme des observations est défini comme le module au carré de la transformation de Fourier discrète de ces observations, on le note $I_N(\cdot)$:

$$I_N(\lambda) = \frac{1}{2\pi N} \left| \sum_{t=1}^N X_t \exp(-it\lambda) \right|^2.$$

Le plus souvent, on estime le périodogramme selon les fréquences de Fourier ; i.e. $\lambda_k = 2k\pi/N$ pour $k = 1, \dots, N$, il constitue donc un estimateur de la densité spectrale $f(\lambda)$. En effet,

$$\begin{aligned} I_N(\lambda) &= \frac{1}{2\pi N} \left(\sum_{t=1}^N X_t \exp(it\lambda) \right) \left(\sum_{t=1}^N X_t \exp(-it\lambda) \right) \\ &= \frac{1}{2\pi} \sum_{k=-N+1}^{N-1} \left(\frac{1}{N} \sum_{t=-\infty}^{+\infty} X_t X_{t-k} \right) \exp(ik\lambda) \\ &= \frac{1}{2\pi} \sum_{k=-N+1}^{N-1} \hat{\gamma}(k) \exp(ik\lambda). \end{aligned}$$

Propriété 1.3.1 :

Le périodogramme $I_N(\lambda)$ est un estimateur asymptotiquement sans biais de la densité spectrale $f(\lambda)$, mais il est non-convergent.

La non convergence du périodogramme n'interdit pas son utilisation. Même si l'estimateur est sans biais, on peut avoir une bonne idée de sa forme générale;

$$\lim_{N \rightarrow \infty} \text{var}(I_N(\lambda)) = \begin{cases} 2f^2(\lambda), & \text{si } \lambda = 0, \pm \pi, \\ f^2(\lambda), & \text{si } \lambda \neq 0, \pm \pi. \end{cases}$$

De plus, les valeurs associées à des fréquences différentes ($\lambda_1 \neq \lambda_2$) du périodogramme sont non corrélées i.e.

$$\lim_{N \rightarrow \infty} \text{cov}(I_N(\lambda_1), I_N(\lambda_2)) = 0, \quad \text{si } \lambda_1 \neq \lambda_2.$$

Maintenant, on effectue la transformation de Fourier inverse de la relation :

$$I_N(\lambda) = \frac{1}{2\pi} \sum_{k=-N+1}^{N-1} \hat{\gamma}(k) \exp(ik\lambda), \quad (\mathbf{1.3.2})$$

ce qui donne un estimateur de la fonction d'autocovariance (i.e. Autocovariance empirique);

$$\begin{aligned} \hat{\gamma}(k) &= \int_{-\pi}^{\pi} \cos(\lambda k) I_N(\lambda) d\lambda, \quad \text{avec} \\ \gamma(k) &= \int_{-\pi}^{\pi} \cos(\lambda k) f(\lambda) d\lambda. \end{aligned}$$

1.4 Représentation des séries temporelles

Généralement, les séries temporelles sont représentées par deux types de modèles, le premier est linéaire (modèles : AR, MA,...), le deuxième est non-linéaire (modèles : ARCH, GARCH,...). Ce dernier type fera l'objet d'un intérêt particulier dans notre mémoire.

Définition 1.4.1 : " Opérateurs retard "

L'opérateur retard noté par L , est l'opérateur linéaire qui associe à un processus $(X_t)_t$ le processus $(Y_t)_t$ tel que :

$$Y_t = LX_t := X_{t-1}.$$

D'une façon générale, on a la relation de récurrence suivante :

$$\left(\sum_{i=0}^n a_i L^i\right)X_t = \sum_{i=0}^n a_i X_{t-i}, \quad \forall n \in \mathbb{N}. \quad (1.4.1)$$

De même, on définit les polynômes Φ et Ψ en L , de degré p (resp. q) par :

$$\begin{aligned} \Phi(L) &: = 1 + \alpha_1 L + \dots + \alpha_p L^p, \\ \Psi(L) &: = 1 + \beta_1 L + \dots + \beta_q L^q, \end{aligned}$$

avec $\alpha_1, \dots, \alpha_p$ et β_1, \dots, β_q sont des constantes.

1.4.1 Modèles : AR, MA et ARMA

Définition 1.4.2 : "Modèle Auto-Regressif d'ordre p "

Un modèle autoregressif d'ordre p noté $AR(p)$, est un processus stationnaire $(X_t)_t$ qui vérifie la relation de type :

$$X_t := \xi_t + \sum_{j=1}^p \alpha_j X_{t-j}, \quad t \in \mathbb{Z},$$

où les $\alpha_1, \dots, \alpha_p$ sont des réelles et $(\xi_t)_t$ un bruit blanc (suite de variables aléatoires (i.i.d), centrée et de variance finie).

En fonction d'opérateur L , la relation précédente prend la forme :

$$X_t := \xi_t + \Phi(L)X_t.$$

Définition 1.4.3 : "Modèle Moyenne Mobile d'ordre q "

Un modèle moyenne mobile (*Moving Average*) d'ordre q noté $MA(q)$, est un processus stationnaire $(X_t)_t$ qui vérifie la relation de type :

$$X_t := \xi_t - \sum_{j=1}^q \beta_j \xi_{t-j}, \quad t \in \mathbb{Z}.$$

Cette relation s'écrit en fonction d'opérateur retard L :

$$X_t := \xi_t - \Psi(L)\xi_t.$$

Définition 1.4.4 : "Modèle Auto-Regressif Moyenne Mobile"

Une série temporelle stationnaire $(X_t)_t$, admet une représentation d'un modèle autoregressif moyenne mobile noté $ARMA(p, q)$, si elle satisfait :

$$X_t := \xi_t + \sum_{j=1}^p \alpha_j X_{t-j} - \sum_{k=0}^q \beta_k \xi_{t-k},$$

ou ce qui est équivalent à :

$$(1 - \Phi(L))X_t := (1 - \Psi(L))\xi_t,$$

avec ;

- i - $\alpha_p \neq 0, \beta_q \neq 0$.
- ii - Les polynômes $\Phi(L)$ et $\Psi(L)$ ont leurs racines de modules strictement supérieurs à 1.
- iii - $\Phi(L)$ et $\Psi(L)$ n'ont pas de racines communes.
- iv - $(\xi_t)_t$ un bruit blanc : suite de variables aléatoires (i.i.d), centrée et de variance finie.

Remarque 1.4.1 :

Remarquons que les modèles $AR(p)$ et $MA(q)$ sont des cas particuliers de la représentation $ARMA(p, q)$. Ainsi, un modèle $AR(p)$ (resp. $MA(q)$) est un modèle $ARMA(p, 0)$ (resp. $ARMA(0, q)$).

De plus, dans le cas où $p = \infty$ (resp. $q = \infty$), on dit qu'une série temporelle $(X_t)_t$ admet la représentation d'un modèle autoregressif infini (resp. moyenne mobile infini) noté $AR(\infty)$ (resp. $MA(\infty)$).

1.4.2 Modèles : ARCH, GARCH

En 1982, Engle, R.F. [24] a proposé l'approche "Auto-Regressive Conditionally Heteroscedastic" ARCH, ce modèle non linéaire qui offre une représentation (modélisation) d'une grande famille de séries temporelles financières, comme par exemple : les taux d'intérêts, les rendements d'actions ou d'indices boursiers et les taux de change,...etc.

Pour aborder ce type de modèle, on introduit la notion de filtration qui représente l'information disponible comme suit :

Soit (Ω, F, P) un espace de probabilité, dont on considère une filtration \mathcal{F}_t (une collection croissante de sous-tribus de F , i.e., $\mathcal{F}_s \subset \mathcal{F}_t$, pour $t \geq s$). Un processus X_t est adapté à une filtration \mathcal{F}_t , si pour tout X_t est \mathcal{F}_t -mesurable.

De plus, la filtration naturelle associée à un processus X_t est par définition la famille de sous-tribus $\mathcal{F}_t = \sigma(X_s, s \leq t)$, où \mathcal{F}_t est la plus petite tribu rendant mesurable les applications $\omega \rightarrow X_s(\omega)$, pour $s \leq t$.

Avec ces notations, la série temporelle $(X_t)_t$, est une différence de martingale relativement à la filtration \mathcal{F}_t , telle que : X_t est \mathcal{F}_t -mesurable et X_t est orthogonale à tout variable aléatoire $Y_t \in \mathcal{F}_{t-1}$, i.e.

$$\mathbb{E}[|X_t|] < \infty \quad \text{et} \quad \mathbb{E}[X_t | \mathcal{F}_{t-1}] = 0, \quad \text{P. presque sûrement. } \forall t,$$

cela contient les variables : $X_s \in \mathcal{F}_s \subset \mathcal{F}_{t-1}$, $s < t$, de sorte que :

$$\mathbb{E}[X_t X_s] = 0, \quad \forall s < t.$$

Définition 1.4.5 : "Modèle ARCH(1)"

Un modèle X_t est dit autoregressif conditionnellement hétéroscédastique d'ordre 1 qu'on note par ARCH(1), s'il admet l'écriture suivante :

$$X_t := \sigma_t \xi_t,$$

dont la variance conditionnelle $\sigma_t^2 = \mathbb{E}[X_t^2 | \mathcal{F}_{t-1}]$ satisfaite pour tout $t \in \mathbb{Z}$ et $\alpha_0 > 0$, $\alpha_1 \geq 0$, des constantes données :

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2,$$

où $(\xi_t)_t$ un bruit blanc : suite de variables aléatoires (i.i.d), centrée et réduite.

On suppose souvent que les variables ξ_t sont indépendantes de la filtration \mathcal{F}_{t-1} et que σ_t dépend de X_t .

Propriété 1.4.1 :

Le modèle $ARCH(1)$ admet les propriétés suivantes :

1. $var(X_t|\mathcal{F}_{t-1}) = \alpha_0 + \alpha_1 X_{t-1}^2 = \sigma_t^2$.

Dans ce cas, on dit que X_t est hétéroscédastique conditionnellement à son passé.

2. $var(X_t) = \alpha_0 + \alpha_1 var(X_{t-1})$.

Preuve : Il suffit d'appliquer la décomposition de la variance et la variance conditionnelle comme suit : pour tout $X_t \in \mathbb{L}^2$:

$$\begin{aligned} var(X_t|\mathcal{F}_{t-1}) &= \mathbb{E}[X_t^2|\mathcal{F}_{t-1}] - \mathbb{E}[X_t|\mathcal{F}_{t-1}]^2. \\ var(X_t) &= var(\mathbb{E}[X_t|\mathcal{F}_{t-1}]) + \mathbb{E}[var(X_t|\mathcal{F}_{t-1})]. \end{aligned}$$

- 1.

$$\begin{aligned} var(X_t|\mathcal{F}_{t-1}) &= \mathbb{E}[\xi_t^2(\alpha_0 + \alpha_1 X_{t-1}^2)|\mathcal{F}_{t-1}] - \mathbb{E}[\xi_t]^2 \mathbb{E}[\sigma_t|\mathcal{F}_{t-1}]^2 \\ &= \mathbb{E}[\xi_t^2] \mathbb{E}[\alpha_0 + \alpha_1 X_{t-1}^2|\mathcal{F}_{t-1}] - 0 \\ &= \alpha_0 + \alpha_1 X_{t-1}^2 = \sigma_t^2. \end{aligned}$$

- 2.

$$\begin{aligned} var(X_t) &= var(\mathbb{E}[X_t|\mathcal{F}_{t-1}]) + \mathbb{E}[var(X_t|\mathcal{F}_{t-1})] \\ &= var(0) + \mathbb{E}[\alpha_0 + \alpha_1 X_{t-1}^2] \\ &= \alpha_0 + \alpha_1 var(X_{t-1}). \end{aligned}$$

Sous-hypothèse de stationnarité (i.e. $var(X_t) = var(X_{t-1})$), ce qui implique que ;

$$var(X_t) = \frac{\alpha_0}{1 - \alpha_1},$$

pour que cette variance non-conditionnelle existe, il suffit donc que : $\alpha_1 \in]0, 1[$. ■

Définition 1.4.6 : "Modèle $ARCH(p)$ "

L'extension du modèle précédent $ARCH(1)$, donne un modèle autoregressif conditionnellement hétéroscédastique d'ordre p qu'on note $ARCH(p)$, qui prend la forme suivante :

$$X_t := \sigma_t \xi_t,$$

dont la variance conditionnelle $\sigma_t^2 = \mathbb{E}[X_t^2|\mathcal{F}_{t-1}]$ satisfaite pour tout $t \in \mathbb{Z}$ et $\alpha_0 > 0$, $\alpha_1, \alpha_2, \dots, \alpha_p$ des constantes positives données :

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2.$$

Sous-hypothèse de stationnarité de (X_t) , la variance non-conditionnelle existe pour $\alpha_1 + \alpha_2 + \dots + \alpha_p < 1$. et on a :

$$\text{var}(X_t) = \frac{\alpha_0}{1 - \alpha_1 + \alpha_2 + \dots + \alpha_p}.$$

Au milieu des années quatre vingt *Bollersllev*, *T.* alors étudiant de *Engle*, généralise la famille des modèles ARCH en permettant à la variance conditionnelle σ_t^2 du modèle de dépendre aussi de son propre passé σ_{t-1}^2 . *Bollerslev* définit donc en 1986 ([04]) le modèle Autorégressif Conditionnellement Hétéroscédastique Généralisés, "*Generalized-ARCH*".

Définition 1.4.7 : "*Modèle GARCH(p, q)*"

Le modèle *GARCH(p, q)* est une suite de différence de martingale X_t , relativement à la filtration \mathcal{F}_t :

$$X_t := \sigma_t \xi_t, \tag{1.4.2}$$

sa variance conditionnelle $\sigma_t^2 = \mathbb{E}[X_t^2 | \mathcal{F}_{t-1}]$ satisfaite pour tout $t \in \mathbb{Z}$, $\alpha_0 > 0$, $\alpha_1, \alpha_2, \dots, \alpha_p$ et $\beta_1, \beta_2, \dots, \beta_q$ des constantes positifs données :

$$\sigma_t^2 := \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{k=1}^q \beta_k \sigma_{t-k}^2. \tag{1.4.3}$$

Pour $\sigma_t > 0$, $X_t := \sigma_t \xi_t$, on a

$$\mathbb{E}[\xi_t | \mathcal{F}_{t-1}] = 0, \quad \mathbb{E}[\xi_t^2 | \mathcal{F}_{t-1}] = 1,$$

et les variables ξ_t sont indépendantes et identiquement distribuées, de plus ξ_t est indépendant de \mathcal{F}_{t-1} , ce qui est équivalent à supposer que la loi conditionnelle de la variable $\xi_t = X_t/\sigma_t$ par rapport a \mathcal{F}_{t-1} , est approximativement d'une distribution normale.

En utilisant les polynômes Φ et Ψ en L de degré p (resp. q), (1.4.3) prend la forme :

$$\sigma_t^2 := \alpha_0 + \Phi(L)X_t^2 + \Psi(L)\sigma_t^2. \tag{1.4.4}$$

Pour les conditions de stationnarité, on peut citer à titre d'exemple : *Bougerol et Picard* [5] pour un modèle *GARCH(p, q)*, *Nelson* [48], pour le modèle *GARCH(1, 1)*.

En effet, la série donnée par les équations (1.4.2) et (1.4.3) est strictement stationnaire si : $\alpha_0 > 0$ et :

$$\sum_{i=1}^p \alpha_i + \sum_{k=1}^q \beta_k < 1.$$

Remarques 1.4.2 :

1. Le modèle $ARCH(p)$ est une forme restreinte du modèle $GARCH(p, q)$, avec comme restriction $q = 0$.

2. On réécrit l'équation (1.4.4) avec les notations :

$\varphi(L) = \Phi(L) + \Psi(L)$ et $Z_t = X_t^2 - \sigma_t^2$, on trouve

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \Phi(L)X_t^2 + \Psi(L)\sigma_t^2 \\ &= \alpha_0 + \Phi(L)X_t^2 - \Psi(L)Z_t, \end{aligned}$$

d'où :

$$(1 - \varphi(L))X_t^2 = \alpha + (1 - \Psi(L))Z_t,$$

ce qui implique que X_t^2 admet la représentation d'un modèle ARMA (définition 1.4.4).

3. *Engle et Bollerslev* [25], ont interprété le nom :

"Integrated Genarilized Autoregressive Conditionaly Heteroscedastic".
 $IGARCH(p, q)$, dans le cas où on a ;

$$\alpha_0 > 0 \quad \text{et} \quad \sum_{i=1}^p \alpha_i + \sum_{k=1}^q \beta_k < 1.$$

Chapitre 2

Séries temporelles et valeurs extrêmes

L'objectif de la théorie des valeurs extrêmes est de mesurer en terme de probabilité l'apparition des événements rares et autres catastrophes : grands sinistres, orages, et crises économiques, . . . etc.

Ce qui rend la connaissance des caractéristiques des distributions et propriétés (comportement) des estimateurs de ces évènements, en particulier la probabilité de dépasser un certain seuil (i.e. valeurs extrêmes) une préoccupation majeure des assureurs dans les compagnies d'assurance, et un excellent outil d'aide à la décision, pour éviter les faillites retentissantes ou les situations de ruines...

La théorie -classique- des valeurs extrêmes, concerne initialement le comportement des sommes. Elle aura à triter par la suite la distribution des maxima ou minima d'une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d). Mais l'un des récents développements de cette théorie a consisté à élever ces hypothèses, en particulier l'hypothèse d'indépendance des observations.

Intuitivement, l'indépendance des observations mènera à une augmentation de l'erreur de l'estimation, donc il est importe de ne pas utiliser des intervalles de confiance classique développée par des suites indépendantes (si la dépendance entre les variables n'est pas négligeable). Si par exemple, la valeur de risque d'un investissement financier sera estimée, alors une borne supérieure de confiance à obtenir avec la théorie classique, indiquera souvent un risque inférieur à celle de réel.

Malheureusement, presque tous les résultats sur le comportement asymptotique des quantiles extrêmes disponibles maintenant sont restreints à des observations indépendantes dans l'étude des séries temporelles financiers, ce qui est cependant rarement réaliste.

Ce chapitre donc comme but d'enquêter sur le comportement asymptotique et la consistance de l'estimateur des quantiles extrêmes, basées sur les grandes observations et sous-conditions (hypothèses) sur la dépendance et la stationnarité d'une série temporelle. Après avoir rappelé certains résultats établis dans le cas d'indépendance, nous allons surtout nous intéresser au cas où se pose le problème de dépendance et de stationnarité dans l'étude des séries temporelles.

2.1 Valeurs extrêmes et indépendance

Comme on l'a signalé au début de ce chapitre, on ne s'intéresse qu'à la partie jugée importante dans la représentation des observations d'un phénomène par des suites de variables aléatoires, en l'occurrence la partie des queues de distribution. Il est donc important de comprendre le comportement asymptotique du maxima ou minima de ces observations....

Soit (Ω, F, P) un espace de probabilité. Dans tout ce qui suit, on suppose que l'on dispose d'un échantillon X_1, X_2, \dots, X_n où l'ensemble de n observations d'une suite de variables aléatoires $(X_n)_n$ extraite d'une série temporelle $(X_t)_t$, pour le moment indépendantes et identiquement distribuées (i.i.d), de même fonction de distribution (répartition) F :

$$F(x) := P\{\omega \in \Omega / X(\omega) \leq x\} = P\{X \leq x\}.$$

On définit d'abord ce que nous appelons la statistique d'ordre.

Définition 2.1.1 : " *Statistique d'ordre* "

On appelle statistique d'ordre, l'application mesurable S de $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})^n$ à valeur dans $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})^n$ classant les observations de la suite $(X_n)_n$ par ordre croissant ;

$$(X_1, X_2, \dots, X_n) \xrightarrow{S} (X_{1,n}, X_{2,n}, \dots, X_{n,n}), \text{ avec } X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n},$$

de sorte que :

$$X_{1,n} := \min(X_1, X_2, \dots, X_n) \text{ et } X_{n,n} := \max(X_1, X_2, \dots, X_n).$$

Le vecteur $(X_{1,n}, X_{2,n}, \dots, X_{n,n})$ est appelé l'échantillon ordonné, et $X_{k,n}$ étant la $k^{\text{ième}}$ statistique d'ordre.

On notera dans la suite respectivement par :

La fonction de distribution empirique basée sur les n premières variables aléatoires, est donnée par :

$$\begin{aligned} F_n(x) &= \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i < x\}}, \quad x \in \mathbb{R} \\ &= \frac{i}{n} \text{ pour } x \in]X_{i,n}, X_{i+1,n}]. \end{aligned}$$

La fonction des quantiles où l'inverse généralisé de la fonction de distribution F noté par Q : Pour chaque entier $n \geq 1$

$$Q(s) := F^{-1}(s) = \inf \{x : F(x) \geq s\}, \quad \text{pour } 0 < s < 1.$$

De même, la fonction des quantiles empiriques Q_n est définie pour $n \geq 1$ par :

$$Q_n(s) = \inf \{x : F_n(x) \geq s\} = \begin{cases} X_{1,n} & \text{pour } s < 0, \\ X_{i,n} & \text{pour } \frac{i-1}{n} < s \leq \frac{i}{n}, \quad 1 \leq i \leq n-1, \\ X_{n,n} & \text{pour } s > 1. \end{cases}$$

Les lois de maxima $X_{n,n}$ (resp. de minima $X_{1,n}$) de la statistique d'ordre associées à la suite $(X_n)_n$ de variables aléatoires (i.i.d), sont données par :

1. **Loi de maxima** $X_{n,n}$:

$$\begin{aligned} F_{n,n}(x) &= P\{X_{n,n} \leq x\} = P\{\max(X_1, X_2, \dots, X_n) \leq x\} \\ &= \prod_{k=1}^n P\{X_k \leq x\} = [F(x)]^n. \end{aligned}$$

2. **Loi de minima** $X_{1,n}$:

$$\begin{aligned} F_{1,n}(x) &= P\{X_{1,n} \leq x\} = 1 - P\{X_{1,n} > x\} \\ &= 1 - P\{\min(X_1, X_2, \dots, X_n) > x\} \\ &= 1 - \prod_{k=1}^n P\{X_k > x\} = 1 - \prod_{k=1}^n (1 - P\{X_k \leq x\}) \\ &= 1 - [1 - F(x)]^n. \end{aligned}$$

Définition 2.1.2 : "*Point extrême*"

On note par x^+ (resp. x^-) le point extrême supérieur (resp. inférieur) de la distribution F (i.e. la plus grande valeur possible pour X_k peut prendre la valeur $+\infty$ (resp. $-\infty$)) au sens où

$$x^+ := \sup \{x : F(x) < 1\}, \text{ respectivement } x^- := \inf \{x : F(x) > 0\}.$$

Définition 2.1.3 : "*Loi Asymptotique des valeurs extrêmes*"

La loi asymptotique des valeurs extrêmes, ou simplement (loi des extrêmes) est la loi de la variable aléatoire Y , telle que :

$$X_{n,n} \xrightarrow{\mathcal{D}} Y, \quad \text{quand } n \text{ tend vers } +\infty.$$

Concernant le comportement du maxima d'une suite de v.a (i.i.d), le résultat le plus important est fourni par le théorème de *Fischer et Tippet* (1928) [26].

Théorème 2.1.1 :

Soit $(X_n)_n$ une suite de variables aléatoires (i.i.d), s'il existe des constantes de normalisation (a_n) et (b_n) , $n \in \mathbb{N}$, avec $a_n > 0$ et $b_n \in \mathbb{R}$, telles que :

$$\left(\frac{X_{n,n} - b_n}{a_n}\right) \xrightarrow{\mathcal{D}} G_\gamma, \quad \text{quand } n \rightarrow \infty. \quad (2.1.1)$$

Alors G_γ a la forme suivante :

$$G_\gamma(x) = \begin{cases} \exp\left(- (1 + \gamma x)^{-1/\gamma}\right) & \text{pour } \gamma \neq 0, 1 + \gamma x > 0, \\ \exp(-e^{-x}) & \text{pour } \gamma = 0, x \in \mathbb{R}. \end{cases}$$

On déduit les trois familles de distribution (dite loi standard ou traditionnelle) des valeurs extrêmes suivantes :

Gumbel (typeI) : $\Lambda(x) = \exp(-e^{-x}), \quad x \in \mathbb{R}.$

Fréchet (typeII) : $\Phi_\gamma(x) = \begin{cases} 0, & x \leq 0, \\ \exp(-x^{-\gamma}), & x > 0, \text{ avec } \gamma > 0. \end{cases}$

Weibull(typeIII) : $\Psi_\gamma(x) = \begin{cases} \exp(-(-x)^{-\gamma}), & x < 0, \text{ avec } \gamma < 0, \\ 1, & x \geq 0. \end{cases}$

En plus, notons par \bar{F} la fonction de survie (ou la fonction des queues) :

$$\bar{F}(x) := P\{\omega \in \Omega / X(\omega) > x\} = 1 - F(x).$$

Ainsi, pour tout $p_n \in]0, 1[$, on note par $x_{p_n} = Q(p_n)$ la quantile d'ordre p_n de la distribution F .

Compte tenu du fait que l'on peut renormaliser les suites (a_n) et (b_n) , il n'existe d'après le théorème précédent que trois lois possibles pour G_γ selon la forme de la queue de la distribution F des X_i :

- **Loi de Gumbel** $\Lambda(x)$: avec,

$$a_n = Q(1 - 1/n) \text{ et } b_n = \bar{F}(a_n)^{-1} \int_{a_n}^{\infty} \bar{F}(s) ds.$$

- **Loi de Fréchet** $\Phi_\gamma(x)$: pour $\gamma > 0$ et

$$a_n = 0 \text{ et } b_n = a_n - F^{-1}(1 - 1/n).$$

- **Loi de Weibull** $\Psi_\gamma(x)$: pour $\gamma < 0$ et,

$$a_n = x^+ \text{ et } b_n = a_n - F^{-1}(1 - 1/n).$$

Ces lois limites que l'on peut représenter au passage à la limite de γ en 0 et à une normalisation près, sous la forme suivante connue sous le nom de représentation de *Jenkinson* :

$$G_\gamma(x) = \exp \left\{ - (1 + \gamma x)^{-1/\gamma} \right\}, \text{ pour } 1 + \gamma x > 0.$$

Ainsi, le cas limite $\gamma \rightarrow \infty$ correspond à la loi de Gumbel, le cas $\gamma > 0$ à la loi de Fréchet et enfin le cas $\gamma < 0$ à celle de Weibull.

Le comportement asymptotique de la loi de $X_{n,n}$ dépend de la distribution initiale F de X . C'est à *van Mises*(1923), à *Gnedenko*(1943) et à *de Haan*(1971) que l'on doit les conditions nécessaires, ou nécessaires et suffisantes sur la distribution F , pour obtenir l'une des trois limites définies précédemment.

Définition 2.1.4 : " *Domaine d'attraction* "

Soit G_γ une fonction de distribution sur \mathbb{R} , une fonction de distribution F sera dite appartenant au domaine d'attraction de G_γ ($F \in D(G)$), s'il existe

deux suites de nombres réels (a_n) et (b_n) , $n \in \mathbb{N}$, avec $a_n > 0$ et $b_n \in \mathbb{R}$, telles que :

$$\lim_{n \rightarrow \infty} F(a_n x + b_n) = \lim_{n \rightarrow \infty} P \left\{ \frac{X_{n,n} - b_n}{a_n} \leq x \right\} = G_\gamma(x), \quad (2.1.2)$$

pour tout x point de continuité de G_γ .

Les caractéristiques des domaines d'attraction sont établies dans les trois théorèmes suivantes, l'ensemble des démonstrations peut être trouvé dans : *B. Gnendenko* (1943), *L. de Haan*(1976) ou *R. von Mises*(1936).

Théorème 2.1.2 :

Soit F une fonction de distribution de densité f positive qui admet une dérivée f' négative sur $]x^-, x^+[$, nulle sur $[x^-, x^+[$, (x^+ fini où non) ;

$$Si \lim_{x \uparrow x^+} \frac{f'(x)(1 - F(x))}{f^2(x)} = -1, \quad \text{alors } F \in D(\Lambda).$$

Théorème 2.1.3 :

Soit F une fonction de distribution de densité f positive sur $[x^+, +\infty[$;

$$Si \lim_{x \rightarrow \infty} \frac{x f(x)}{1 - F(x)} = \gamma, \quad \text{alors } F \in D(\Phi_\gamma).$$

Théorème 2.1.4 :

Soit F une fonction de distribution de densité f positive sur $]x^{(-)}, x^+[$, nulle sur $]x^+, +\infty[$;

$$Si \lim_{x \uparrow x^+} \frac{(x^+ - x)f(x)}{f^2(x)} = -1, \quad \text{alors } F \in D(\Psi_\gamma).$$

Signallons à la fin de cette section qu'en terme de risque, l'obtention des lois précédent et en particulier l'estimation du paramètre γ de la distribution limite des valeurs extrêmes, est important pour évaluer la probabilité que l'ensemble de la population est au-dessus d'un certain seuil s . Entre autres il s'agit d'évaluer la quantité :

$$P \{X_{n,n} < s\} \approx \exp \left\{ - (1 + \gamma(s - a_n) / b_n)^{-1/\gamma} \right\}.$$

2.2 Valeurs extrêmes et dépendance

Comme on l'a signalé, l'une des pistes naturelles de généralisation de la théorie -classique- des valeurs extrêmes, repose sur la remise en cause de l'hypothèse de dépendance.

Il est clair que la théorie classique sur les variables aléatoires indépendantes mène à des intervalles de confiance qui sont plus courtes que celles qui prennent en considération la dépendance de la série étudiée.

Dans ce qui suit, nous rappelons les principales mesures de dépendance, en particulier la dépendance finie, le α -mélangeance et le β -mélangeance. Pour cela, on peut se référer au livre de *Doukhan* [20] qui fournit une étude détaillée des différentes sortes de mélangeance.

2.2.1 Mesure de dépendance

On peut mesurer la dépendance entre le passé des variables aléatoires (X_k ; avant $k = m$) et le futur (X_k ; après $k = m$), suivant la vitesse avec laquelle leur séparation en $n = m$ augmente.

Définition 2.2.1 : "*m*-dépendance "

La série temporelle $(X_t)_t$ est dite *m*-dépendante (dépendance finie), si les vecteurs aléatoires (\dots, X_{t-1}, X_t) et $(X_{t+m+1}, X_{t+m+2}, \dots)$ sont indépendants, $\forall t \in \mathbb{Z}$.

En outre, le "passé" et le "futur" sont indépendants si l'on fait abstraction des *m* "présent" variables, i.e.

- Les variables X_t et X_s sont dépendantes quand $|t - s| \leq m$.
- Les variables X_t et X_s sont indépendantes quand $|t - s| > m$.

Précisément, pour t_1, t_2, \dots, t_n on note :

$$F(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = P(X_{t_1} \leq x_1, X_{t_2} \leq x_2, \dots, X_{t_n} \leq x_n).$$

La *m*-dépendance correspond donc à :

Si : $s_1 < s_2 < \dots < s_p < t_1 < t_2 < \dots < t_q$, avec $t_1 - s_p > m$, alors :

$$\begin{aligned} F(x_1, \dots, x_p; s_1, \dots, s_p, y_1, \dots, y_q; t_1, \dots, t_q) \\ = F(x_1, \dots, x_p; s_1, \dots, s_p)F(y_1, \dots, y_q; t_1, \dots, t_q). \end{aligned}$$

L'une des mesure de dépendance importante que l'on va l'utiliser par la suite, est le β -mixing :

Définition 2.2.2 : " *β -mélangeance*"

Une série temporelle $(X_t)_t$ est dite *β -mélangeante* (*β -mixing*) ou absolument régulière, s'il existe une suite $\beta(\ell)$ qui tend vers 0 quand ℓ tend vers ∞ telle que :

$$\beta(\ell) := \sup_{m \in \mathbb{N}} \mathbb{E} \left[\sup_{A \in \mathcal{B}_{m+\ell+1}^\infty} |P(A/\mathcal{B}_1^m) - P(A)| \right] \rightarrow 0, \quad \forall m, \forall \ell, \quad (2.2.1)$$

avec :

$\mathcal{B}_{m+\ell+1}^\infty$ est le σ -Algèbre engendré par $(X_{m+\ell+1}, X_{m+\ell+2}, \dots)$,

\mathcal{B}_1^m est le σ -Algèbre engendré par (X_1, X_2, \dots, X_m) .

La suite $\beta(\ell)$ est appelé le coefficient de mélangeance.

Définition 2.2.3 : " *α -mélangeance*"

De même, une série temporelle $(X_t)_t$ est dite *α -mélangeante* (*strong mixing*), s'il existe une suite $\alpha(\ell)$ coefficient de mélangeance, qui tend vers 0 quand ℓ tend vers ∞ , telle que :

$$\alpha(\ell) := \sup_{A \in \mathcal{B}_{m+\ell+1}^\infty, B \in \mathcal{B}_1^m} |P(A \cap B) - P(A)P(B)| \rightarrow 0, \quad \forall m, \forall \ell. \quad (2.2.2)$$

2.3 L'index des queues de distributions

Cette section est consacrée à l'étude de quelques estimateurs de l'index des valeurs extrêmes, intervenant dans l'estimation des quantiles extrêmes. Dans le cas où l'index des valeurs extrême γ est positif, l'un des meilleurs estimateurs est celui de Hill (*Hill* (1975)) qui a été généralisé en 1989, par *Dekkers, Einmahl et de Haan* [12] sur tous les domaines d'attraction.

Aussi, *Csörgrö, Deheuvels et Mason* (1985) l'on régularisé par un noyau. Mais malheureusement, cet estimateur de type noyau n'est valable que pour $\gamma > 0$. Et c'est ainsi que *Groeneboom, Lopuha et de Wolf* (2003) ont introduit une nouvelle classe d'estimateurs à noyau dont la consistance forte de cette classe d'estimateurs a été établie par *Necir, A.*(2004) [44].

Dans cette section, on va étudier les comportements asymptotiques des estimateurs de l'index des queues de distribution de type Pareto ou l'index d'une série temporelle à variations régulières, sous des hypothèses sur la stationnarité et la dépendance... On se base seulement sur le cas d'un index positif (i.e. $\gamma > 0$).

Considérons la suite $(X_n)_{n \geq 1}$ extraite d'une série temporelle stationnaire $(X_t)_t$ et soit $X_{1,n} \leq \dots \leq X_{n,n}$ la statistique d'ordre associée à X_1, X_2, \dots, X_n et dont la fonction de distribution est à queue lourde, i.e. $\exists \alpha > 0, \exists c > 0$ tel que quand x tend vers ∞ :

$$\bar{F}(x) := P\{X_t > x\} \sim x^{-\alpha} L(x),$$

où $L(\cdot)$ est une fonction à variations lentes et α l'index extrême à estimer.

Ce qui permet de déterminer les quantiles extrêmes de la distribution F , typiquement $x_{p_n} = F^{-1}(1 - p_n) = Q(1 - p_n)$ est la quantile extrême à estimer, et nous nous intéressons souvent au cas où p_n soit très petit, parfois inférieur à $1/n$ (i.e. $p_n = O(1/n)$), à celle des queues de distribution.

Dans le cas d'indépendance, on dit que la de distribution F appartient au domaine d'attraction d'une fonction G_γ , s'il existe deux suites réelles b_n et $a_n > 0$ telles que :

$$\lim_{n \rightarrow \infty} F(a_n x + b_n) = \lim_{n \rightarrow \infty} P \left\{ \frac{X_{n,n} - b_n}{a_n} \leq x \right\} = G_\gamma(x).$$

Un des cas importante en applications financières, est celui où l'index γ de la distribution limite des valeurs extrêmes est positif ($\gamma > 0$). Dans ce cas, l'estimateur de Hill, est sans doute le plus utilisé dans la théorie des valeurs extrêmes, et la condition qu'une fonction de distribution F appartenant au domain d'attraction des lois des extrêmes G_γ , pour $\gamma > 0$ est équivalente à :

$$\Psi(\lambda, t) := \frac{Q(1 - \lambda t)}{Q(1 - \lambda)} - t^{-\gamma} \rightarrow 0, \quad t > 0 \quad \text{quand } \lambda \downarrow 0, \quad (2.3.1)$$

une convergence que l'on peut lire comme une approximation pour λ assez petite (i.e. $\lambda = \frac{k_n}{n}$, pour $t = \frac{np_n}{k_n}$) :

$$\begin{aligned} x_{p_n} & : = Q(1 - p_n) \approx Q\left(1 - \frac{k_n}{n}\right) \left(\frac{np_n}{k_n}\right)^{-\gamma} \\ & \approx X_{n-k_n, n} \left(\frac{np_n}{k_n}\right)^{-\hat{\gamma}} =: \hat{x}_{p_n}, \end{aligned} \quad (2.3.2)$$

où $\hat{\gamma}$ est l'estimateur souhaitable pour γ , dépend de la $(k_n + 1)^{i\text{ème}}$ statistique d'ordre $X_{k_n+1, n}$ et la suite k_n vérifiée :

$$k_n \rightarrow \infty \quad \text{et} \quad \frac{k_n}{n} \rightarrow 0, \quad \text{quand } n \rightarrow \infty. \quad (2.3.3)$$

En utilisant la méthode de maximum de vraisemblance, *B. M. Hill* (1975) a proposé un estimateur de l'index des valeurs extrêmes ($\gamma > 0$) comme suit :

$$\hat{\gamma}^H = \frac{1}{k_n} \sum_{i=1}^{k_n} \log\left(\frac{X_{n-i+1, n}}{X_{n-k_n, n}}\right), \quad (2.3.4)$$

cet estimateur est largement étudié dans le cas i.i.d. A titre d'exemple, on peut citer en substance les travaux de : *Hall* (1982), *Mason* (1982, 88), *de Haan et Resnick* (1996), *Davis et Resnick* (1984)...

2.3.1 Consistance de l'estimateur de l'index

Notre but maintenant, est de caractériser les divers comportements asymptotiques de l'estimateur des quantiles extrêmes ($\gamma > 0$), en particulier la consistance de l'estimateur de *Hill* (2.3.4) au cas où les séries temporelles dont les hypothèse d'indépendance et d'iquidistributé sont fortement rejetées... Dont les premiers travaux concernant ce point de vue sont ceux de : *Hsing* (1991), *Rootzen* (1995).

Commençant par l'étude de la consistance de l'estimateur de *Hill* dans le cas d'une série temporelle stationnaire dont la distribution est à queue lourde. Pour cela on admet les notations suivantes :

Soient $\mathcal{E} =]0, \infty]$, \mathcal{B} une σ -algèbre de Borel dans \mathcal{E} et $M_+(\mathcal{E})$ l'espace des mesures positives de Radon dans \mathcal{E} muni d'une topologie engendrée par une famille dénombrable de semi-normes \mathcal{S} (*Kallenberg* (1983) [35], *Resnick* (1987) [48]) :

$$\mathcal{S} = \{p_f : M_+(\mathcal{E}) \rightarrow \mathbb{R}^+ : p_f(\mu) = \mu(f), \quad |f| \leq 1 \text{ et } f \in \mathbb{C}_c^+(\mathcal{E})\},$$

où $\mathbb{C}_c^+(\mathcal{E})$ est l'espace des fonctions positives, continues dans \mathcal{E} à support compact.

Pour tout $x \in \mathcal{E}$ et $A \in \mathcal{B}$, on définit la masse de Dirac par :

$$\delta_x(A) = \begin{cases} 1 & \text{si } x \in A, \\ 0 & \text{si } x \in \bar{A}. \end{cases}$$

Enfin, à l'échantillon X_1, X_2, \dots, X_n on associe la mesure empirique :

$$\mu_n(\omega) := \sum_{t=1}^n \delta_{X_t}(\omega), \quad \omega \in \Omega.$$

La proposition suivante exprime la convergence de la mesure empirique des queues de distribution, ainsi que la consistance de l'estimateur de Hill pour une série temporelle stationnaire :

Proposition 2.3.1 : (*Resnick et Stărică*) [50]

Soient pour tout $n \geq 1$ $(X_{n,i})_{i \geq 1}$ une suite de variables aléatoires stationnaire d'éléments de \mathcal{E} , $(k = k_n)_n$ une suite vérifiant la condition (2.3.3). On suppose que $(X_{n,i})_{i \geq 1}$ satisfait les deux conditions :

(1) Pour tout $f \in \mathbb{C}_c^+(\mathcal{E})$

$$\lim_{n \rightarrow \infty} \frac{n}{k^2} \sum_{j=2}^k \mathbb{E} [f(X_{n,1})f(X_{n,j})] = 0.$$

(2) Pour toute suite $(\ell_n)_n$ telle que :

$$\ell_n \rightarrow \infty \text{ et } \frac{\ell_n}{k} \rightarrow 0, \quad \text{quand } n \rightarrow \infty. \quad (2.3.5)$$

Et considérons les intervalles :

$$I_1 = [1, k - \ell_n], \quad I_2 = [k + 1, 2k - \ell_n], \dots, \quad I_{\lfloor n/k \rfloor} = [(\lfloor n/k \rfloor - 1)k + 1, \lfloor n/k \rfloor k - \ell_n],$$

d'une sorte que pour toute fonction $f \in \mathbb{C}_c^+(\mathcal{E})$:

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\prod_{j=1}^{\lfloor n/k \rfloor} \exp\left(\frac{-1}{k} \sum_{i \in I_j} f(X_{n,i})\right) \right] - \prod_{j=1}^{\lfloor n/k \rfloor} \mathbb{E} \left[\exp\left(\frac{-1}{k} \sum_{i \in I_j} f(X_{n,i})\right) \right] = 0.$$

Supposons de plus que pour tout $x \in \mathcal{E}$,

$$\frac{n}{k} P \{X_{n,1} \in \cdot\} \rightarrow \mu, \quad \text{où } \mu(\{x\}) = 0, \quad (2.3.6)$$

alors

$$\mu_n := \frac{1}{k} \sum_{i=1}^n \delta_{X_{n,i}} \Longrightarrow \mu \quad \text{dans } M_+(\mathcal{E}). \quad (2.3.7)$$

D'ailleurs, si $X_{n,i} = \frac{X_i}{b_n}$, $i = 1, 2, \dots, n$, où $(X_n)_{n \geq 1}$ est une suite de variables aléatoires stationnaire et b_n tend vers l'infini, et si μ satisfait : $\int_1^\infty \log(s) \mu(ds) < \infty$. Alors,

$$\hat{\gamma}^H = \frac{1}{k} \sum_{i=1}^k \log\left(\frac{X_{i,n}}{X_{k+1,n}}\right) \xrightarrow{\mathbf{P}} \int_1^\infty \log(s) \mu(ds).$$

Preuve :

Supposons que $f \in \mathbb{C}_c^+(\mathcal{E})$. Pour montrer (2.3.7), il suffit (d'après *Koltenberg* (1983) ou *Resnick* (1987)) de montrer que :

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\exp\left\{\frac{-1}{k} \sum_{i=1}^n f(X_{n,i})\right\} \right] = \exp\{-\mu(f)\}. \quad (2.3.8)$$

Pour cela, admettons les notations suivantes : $f_i = f(X_{n,i})$ et $p = \lfloor n/k \rfloor$. Alors les intervalles précédents redeviennent

$$I_j = \{(j-1)k_n + 1, \dots, jk_n - \ell_n\}, \quad I_j^* = \{jk_n - \ell_n + 1, \dots, jk_n\},$$

$$I_p = \{(p-1)k_n + 1, \dots, pk_n - \ell_n\} \text{ et } I_p^* = \{pk_n - \ell_n + 1, \dots, pk_n\}.$$

On a donc,

$$\begin{aligned}
& \left| \mathbb{E}[\exp\{\frac{-1}{k} \sum_{i=1}^n f_i\}] - \exp\{-\mu(f)\} \right| \\
\leq & \left| \mathbb{E}[\exp\{\frac{-1}{k} \sum_{i=1}^n f_i\}] - \mathbb{E}[\exp\{\frac{-1}{k} \sum_{j=1}^p \sum_{i \in I_j} f_i\}] \right| \\
& + \left| \mathbb{E}[\exp\{\frac{-1}{k} \sum_{j=1}^p \sum_{i \in I_j} f_i\}] - \mathbb{E}[\exp\{\frac{-1}{k} \sum_{i \in I_j} f_i\}]^p \right| \\
& + \left| \mathbb{E}[\exp\{\frac{-1}{k} \sum_{i \in I_j} f_i\}]^p - \mathbb{E}[\exp\{\frac{-1}{k} \sum_{i=1}^k f_i\}]^p \right| \\
& + \left| \mathbb{E}[\exp\{\frac{-1}{k} \sum_{i=1}^k f_i\}]^p - \exp\{-\mu(f)\} \right| \\
= & : I_n + II_n + III_n + IV_n.
\end{aligned}$$

Maintenant, pour le terme I_n on a

$$\begin{aligned}
I_n & \leq \left| \mathbb{E}[\exp\{\frac{-1}{k} \sum_{j=1}^p (\sum_{i \in I_j} f_i + \sum_{i \in I_j^*} f_i)\}] \right| \\
& \leq \mathbb{E} \left| 1 - \exp\{\frac{-1}{k} \sum_{j=1}^p \sum_{i \in I_j^*} f_i\} \right| \\
& \leq \sum_{j=1}^p \sum_{i \in I_j^*} \frac{1}{k} \mathbb{E}[f_i] \\
& \leq p \frac{\ell_n}{k} \mathbb{E}[f_i] \sim \frac{\ell_n}{k} \frac{n}{k} \mathbb{E}[f_1] \sim \frac{\ell_n}{k} \mu(f),
\end{aligned}$$

le dernier terme tend vers 0 quand $n \rightarrow \infty$, cela d'après **(2.3.5)** et **(2.3.8)**.

De même pour III_n ;

$$\begin{aligned}
III_n &\leq p \left| \mathbb{E}[\exp\{\frac{-1}{k} \sum_{i \in I_j} f_i\}] - \mathbb{E}[\exp\{\frac{-1}{k} \sum_{i=1}^k f_i\}] \right| \\
&\leq p \mathbb{E} \left| 1 - \exp\{\frac{-1}{k} \sum_{i \in I_j^*} f_i\} \right| \\
&\leq p \frac{\ell_n}{k} \mathbb{E}[f_1] \rightarrow 0 \quad \text{quand } n \rightarrow \infty.
\end{aligned}$$

De plus, de la condition **(2)** dans la proposition, il découle que $II_n \rightarrow 0$.

Finalement, pour IV_n on pose $y_i = 1 - \exp\{\frac{-1}{k} f_i\}$,

$$\begin{aligned}
\mathbb{E}[\exp\{\frac{-1}{k} \sum_{i=1}^k f_i\}] &= \mathbb{E}[\prod_{i=1}^k (1 - y_i)] \\
&\leq 1 - \mathbb{E}[\sum_{i=1}^k y_i] + \mathbb{E}(\sum_{1 < i < j \leq k} y_i y_j) \\
&\leq 1 - \mathbb{E}[y_1] + k \sum_{\ell=2}^k \mathbb{E}(y_1 y_\ell).
\end{aligned}$$

D'autre part,

$$\mathbb{E}[\exp\{\frac{-1}{k} \sum_{i=1}^k f_i\}]^p \leq (1 - \frac{1}{p} k p (\mathbb{E}[y_1] k \sum_{\ell=2}^k \mathbb{E}(y_1 y_\ell)))^p,$$

$$\begin{aligned}
k p \sum_{\ell=2}^k \mathbb{E}(y_1 y_\ell) &\sim n \sum_{\ell=2}^k \mathbb{E}[1 - \exp\{\frac{-1}{k} f_1\}] \mathbb{E}[1 - \exp\{\frac{-1}{k} f_\ell\}] \\
&\leq \frac{n}{k^2} \sum_{\ell=2}^k \mathbb{E}[f_1 f_\ell] \rightarrow 0 \quad \text{quand } n \rightarrow \infty.
\end{aligned}$$

Identiquement, par la condition **(1)** ;

$$\begin{aligned}
k p \mathbb{E}(y_1) &\sim n \mathbb{E}[1 - \exp\{\frac{-1}{k} f_1\}] \leq \frac{n}{k} \mathbb{E}[f_1] \rightarrow \mu(f) \quad \text{et} \\
n \mathbb{E}[1 - \exp\{\frac{-1}{k} f_1\}] &\geq n \mathbb{E}[\frac{f_1}{k} - \frac{f_1^2}{2k^2}] \sim \mu(f) - \frac{1}{2k^2} \mu(f^2) \rightarrow \mu(f).
\end{aligned}$$

On conclut que,

$$\begin{aligned}\limsup_{n \rightarrow \infty} \mathbb{E}[\exp\{\frac{-1}{k} \sum_{i=1}^k f_i\}]^p &\leq \exp\{-\mu(f)\}, \\ \liminf_{n \rightarrow \infty} \mathbb{E}[\exp\{\frac{-1}{k} \sum_{i=1}^k f_i\}]^p &\leq \exp\{-\mu(f)\},\end{aligned}$$

cela confirme la preuve de **(2.3.7)** et par conséquent celle de **(2.3.8)**.

Finalement, pour démontrer la consistance de l'estimateur de Hill **(2.3.4)** sous les conditions de cette proposition, on utilise la *proposition 2.4* de *Resnick et Stărică* (1995)[49], qui confirme que la convergence de la mesure empirique de la queue de distribution implique la consistance de l'estimateur de Hill. ■

Proposition 2.3.2 :

Sous les conditions de la propriété précédente, on remplace uniquement celle concernant les fonctions $f \in \mathbb{C}_c^+(\mathcal{E})$ par la condition suivante :

$$\lim_{n \rightarrow \infty} \left| \mathbb{E} \left[\prod_{j=1}^{\lfloor n/k \rfloor} \left(1 - \frac{1}{k} \sum_{i \in I_j} f(X_{n,i}) \right) \right] - \prod_{j=1}^{\lfloor n/k \rfloor} \mathbb{E} \left[1 - \frac{1}{k} \sum_{i \in I_j} f(X_{n,i}) \right] \right| = 0, \tag{2.3.9}$$

pour tout fonction f de la forme :

$$f = \sum_{h=2}^s \beta_h \mathbf{1}_{]x_h, \infty]}, \text{ avec } \beta_h > 0, \quad h = 1, \dots, s \quad \text{et} \quad x_h > 0, \quad h \geq 1.$$

Alors, les assertions de la proposition 2.3.1 sont vérifiées.

Preuve : La preuve de cette proposition est semblable à la proposition précédente. ■

2.4 Valeurs extrêmes et hétéroscédasticité

Dans l'étude des séries temporelles présentant certains phénomènes économétriques tel que : les rendements d'indice boursier, actifs financiers et taux de change,...etc, un problème important se pose. C'est celui de mesurer les risques intervenant après l'apparition de certains événements rares.

Pour résoudre ce type de problème, on doit étudier ce risque comme étant un indice des queues de distribution. Comme on le verra plus loin (voir le prochain chapitre), les modèles qui peuvent représenter (modéliser) ce genre de séries temporelles financières sont les modèles hétéroscédastiques : en particulier le modèle ARCH, GARCH,...

Concernant le modèle ARCH(1), C'est à *de Haan, Resnick, Rootzèn et de Vriès* (1989) [15] que l'on doit les résultats portant sur le comportement des extrema triés de ce modèle. Leur démarche s'inscrit dans la lignée des travaux de *Kesten* (1973) et *Vervaat* (1979), portant sur les équations stochastiques aux différences (ou récurrentes). La même approche a ensuite été appliquée à des modèles plus généraux, le modèle GARCH(1,1), GARCH(p, q),...

Par la suite, on va étudier la consistance de l'estimateur de Hill ($\gamma > 0$), qui est basé sur un échantillon trié d'un modèle vérifiant l'équation stochastique aux différences (ou récurrentes) de la forme suivante :

$$Y_t = A_t Y_{t-1} + B_t, \quad t \geq 1, \quad Y \geq 0, \quad (2.4.1)$$

où (A_t, B_t) est un couple de variable aléatoire i.i.d. à valeurs dans \mathbb{R}_+^2 .

Puis, on suppose que A_1, B_1 ont des fonctions de distribution absolument continues. D'après *Kesten* (1973) les conditions d'existence d'une solution stationnaire de l'équation (2.4.1) et dont la distribution marginale est à queue lourde sont : $A_1, B_1 > 0$ et il existe $\kappa > 0$, telle que

$$\mathbb{E}[A_1^\kappa] = 1, \quad \mathbb{E}[A_1^\kappa \max(0, \ln A_1)] < \infty \quad \text{et} \quad 0 < \mathbb{E}[B_1^\kappa] < \infty. \quad (2.4.2)$$

Dans ce cas, la fonction de distribution F de X_1 satisfait pour une constante $c > 0$:

$$1 - F(x) \sim cx^{-\kappa}, \quad \text{quand } x \rightarrow \infty,$$

et elle appartient au domaine d'attraction de G_γ , avec l'index des valeurs extrêmes $\gamma = 1/\kappa$.

D'autre part, *Davis et al.* (1999)[11] ont donné les conditions nécessaires et suffisantes pour l'existence d'une solution de l'équation (2.4.1). Si on note par $|\cdot|$ la norme euclidienne dans \mathbb{R}^n et pour toute matrice de dimension $d \times d$, $\|A\| = \sup_{|x|=1} |Ax|$ l'opérateur normé correspondant.

Théorème 2.4.1 : (*Davis et al. 1999*) [11]

Si $\mathbb{E}[A_1^\kappa \max(0, \ln \|A_1\|)] < \infty$, $\mathbb{E}[\max(0, \ln |B_1|)] < \infty$ et si l'exposant de Lyapunov λ définie par :

$$\lambda := \inf \left\{ \frac{1}{n} \mathbb{E} [\ln \|A_1 A_2 \dots A_n\|] \right\},$$

est strictement négatif, alors (2.4.1) n'admet qu'une unique solution fortement stationnaire et causale.

De plus, les conditions

$$\mathbb{E}[\|A_1\|^\epsilon] < 1, \quad \mathbb{E}[|B_1|^\epsilon] < \infty, \quad \text{pour tout } \epsilon \in]0, 1],$$

sont nécessaires et suffisantes pour l'existence et l'unicité d'une solution stationnaire de l'équation (2.4.1).

Une classe importante des modèles satisfaisant cette équation stochastique aux différences, est celle des modèles hétéroscédastique, en particulier les modèles ARCH(1) donnée sous l'équation

$$X_t = \xi_t(\alpha_0 + \alpha_1 X_{t-1}^2)^{1/2}, \quad t \in \mathbb{Z},$$

avec $\alpha_0 > 0$ et $\alpha_1 \geq 0$. Il est clair que le processus X_t^2 est la solution de l'équation (2.4.1) avec, respectivement $A_t = \alpha_0 \xi_t^2$ et $B_t = \alpha_1 \xi_t^2$

En plus, (*de Haan, Resnick, Rootzèn et de Vries* (1989) [15]) ont montré que sous la condition (2.4.2), il existe un indice γ appartenant à l'intervalle $]0, \kappa[$, et une constante c_0 telle que

$$0 < c_0 < 1 \quad \text{et} \quad \mathbb{E}[A_1^\gamma] = c_0 < 1. \quad (2.4.3)$$

Revenant maintenant à l'équation (2.4.1), pour tout $t \geq 1$

$$\begin{aligned} Y_t &= A_t Y_{t-1} + B_t \\ &= A_t (A_{t-1} Y_{t-1} + B_{t-1}) + B_t, \end{aligned}$$

par itération, on trouve

$$Y_t = \sum_{j=0}^{\infty} \left(\prod_{i=t-j+1}^t A_i \right) B_{t-j} := \sum_{j=0}^{\infty} Y_t^{(j)} \quad \text{avec} \quad \prod_{i=t+1}^t A_i = 1. \quad (2.4.4)$$

Proposition 2.4.1 :

Sous la condition **(2.4.2)** et supposons de plus que $(Y_t)_t$ est une solution de l'équation **(2.4.1)**, choisissons une suite $k := k_n$ telle que $n = o(k^{3/2})$. Alors l'estimateur de Hill basé sur la suite $(Y_t)_t$ est consistant, i.e.

$$\frac{1}{k} \sum_{i=1}^k \log \frac{Y_{i,n}}{Y_{i+k,n}} \xrightarrow{P} \frac{1}{\alpha}, \quad \text{quand } n \text{ tend vers } \infty.$$

Preuve :

De la condition posée sur la suite k_n (i.e. $n = o(k^{3/2})$), il est possible de prendre une autre suite $\ell_n = o(1)$ de sorte que $\frac{n}{k} < \ell_n < \frac{k^2}{n}$. En effet, ce choix permet d'obtenir respectivement : la condition **(2.3.5)** dans la proposition 2.3.1 et la condition $\frac{n}{k} = o(\ell_n)$, en concluant par la proposition 2.3.1. ■

Identiquement, pour un modèle atoregressif conditionnellement hétéroscé-
dastique généralisé GARCH(1,1), définie par le système d'équations suivant :

$$\begin{cases} X_t := \sigma_t \xi_t & t \in \mathbb{Z}, \\ \sigma_t^2 := \alpha_0 + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2, \end{cases}$$

avec α_0, α et β des constantes positives données et $(\xi_t)_t$ est un bruit blanc (i.e. une suite de variables aléatoires, indépendantes et identiquement distribuées centrée réduite).

En transformant ce système sous forme matricielle, ce qui permet de remarquer que le modèle GARCH(1,1) vérifie l'équation stochastique aux différences **(2.4.1)**, avec respectivement Y_t, A_t et B_t sont données par :

$$Y_t = \begin{pmatrix} X_t \\ \sigma_t^2 \end{pmatrix}, \quad A_t = \begin{pmatrix} \alpha \xi_{t-1}^2 & \beta \xi_{t-1}^2 \\ \alpha & \beta \end{pmatrix} \text{ et } B_t = \begin{pmatrix} \alpha \xi_{t-1}^2 \\ \alpha_0 \end{pmatrix}.$$

De même, pour la variance conditionnelle σ_t^2

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2 \\ &= (\alpha \xi_{t-1}^2 + \beta) \sigma_{t-1}^2 + \alpha_0, \end{aligned}$$

qui est une simple forme de l'équation **(2.4.1)** avec :

$$Y_t = \sigma_{t-1}^2, \quad A_t = (\alpha \xi_{t-1}^2 + \beta) \quad \text{et} \quad B_t = \alpha_0.$$

Suite à l'étude des valeurs extrêmes, en particulier le comportement et l'estimation de l'index de la queue de distribution des modèles hétéroscédastiques vérifiant une équation de type **(2.4.1)**, un résultat remarquable concernant le modèle GARCH(1, 1) est celui obtenu par *T. Mikosch* et *C. Stărică* (1998) [39] sous des conditions comparables à celles du théorème établi pour les modèles ARCH.

Ces auteurs ont montré dans leur article que : les suites des maxima des variances conditionnelles ($M_{t,\sigma} = \max(\sigma_\tau^2, 0 < \tau < t)$), celles des valeurs absolues du modèle ($M_{t,|X|} = \max(|X_\tau|, 0 < \tau < t)$) et enfin celles des maxima du modèle ($M_{t,X} = \max(X_\tau, 0 < \tau < t)$) appartenant tous au domaine d'attraction d'une loi de Fréchet (voir aussi le type II dans le théorème 2.1.1), dont la valeur de l'index extrême pour chacune des suites précédentes sont données respectivement par :

$$\begin{aligned}\gamma_\sigma &= \int_1^\infty P \left\{ \sup_{t \geq 1} \prod_{j=1}^t A_j \leq y^{-1} \right\} \kappa y^{-\kappa-1} dy, \\ \gamma_{|X|} &= \lim_{\kappa \rightarrow \infty} \mathbb{E} \left[|\xi_1|^\kappa - \max_{j=1,2,\dots,\kappa+1} (0 \vee \left| \xi_j^2 \prod_{j=1}^t A_j \right|) / \mathbb{E}[|\xi_1|^\kappa] \right], \\ \gamma_X &= 2\gamma_{|X|} \left(1 - \tilde{\pi}\left(\frac{1}{2}\right) \right).\end{aligned}$$

Remarque 2.4.1 :

Dans l'expression de γ_X , la quantité $\tilde{\pi}(\cdot)$ correspond à une fonction qui génère des probabilités du processus de poisson composées (*Hsing* [34]), la constante κ est celle donnée dans **(2.4.2)**. D'autre part, à cause de la complexité de ces expressions analytiques (γ_σ , $\gamma_{|X|}$ et γ_X), on peut les obtenir par simulation (technique de Monte-Carlo). Pour une étude détaillée, on peut citer à ce propos l'article des auteurs [39].

2.5 Normalité asymptotique et mélangeance

Dans cette section, nous allons étudier le comportement asymptotique de l'estimateur des quantiles de type **(2.3.2)**, et surtout nous nous intéressons au cas où se pose le problème de dépendance et stationnarité dans l'étude des séries temporelles, en particulier quand la série étudiée possède un certain type de mélangeance et précisément le β -mélangeance...

Comme exemple des modèles de séries temporelles satisfaisant ces conditions, une classe particulière de modèle non-linéaire, tels que les modèles : ARCH, GARCH,...

Soit $(X_t)_t$ une série temporelle strictement stationnaire **(1.1.1)**, absolument régulière (où β -mélangeante **(2.2.1)**).

Précisément, supposons qu'il existe une suite $(\ell_n)_{n \in \mathbb{N}}$, telle que :

(C1)

$$\lim_{n \rightarrow \infty} n \frac{\beta(\ell_n)}{\ell_n} + \ell_n k_n^{-1/2} \log^2 k_n = 0,$$

cette condition est vérifiée avec $\ell_n = \lceil c \log n \rceil$, pour une constante $c > 0$ et k_n satisfait :

$$\log^2 n \log^4(\log n) = o(k_n). \quad \textbf{(2.5.1)}$$

De plus, on suppose une condition de régularité de la queue jointe de variables aléatoires (X_1, X_{m+1}) ;

(C2)

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{n}{k_n} P \left\{ X_1 > Q \left(1 - \frac{k_n}{n} x \right); X_{m+1} > Q \left(1 - \frac{k_n}{n} y \right) \right\} \\ \rightarrow C_m(x, y), \quad \forall m \in \mathbb{N}, 0 < x, y \leq 1 + \varepsilon. \end{aligned}$$

Ajoutons une troisième condition : pour que la probabilité qu'a à la fois X_1, X_{m+1} appartenant à une intervalle extrême soit bornée ;

(C3) Il existe $\delta > 0$ et une suite $\tilde{\rho}(m)$, $m \in \mathbb{N}$ avec $\sum_{m=1}^{\infty} \tilde{\rho}(m) < \infty$, telle que :

$$\begin{aligned} \frac{n}{k_n} P \{ X_1 \in I_n(x, y); X_{m+1} \in I_n(x, y) \} \leq (y - x)(\tilde{\rho}(m) + \delta \frac{k_n}{n}), \\ \forall m \in \mathbb{N}, 0 < x, y \leq 1 + \varepsilon, \end{aligned}$$

où $I_n(x, y) := [Q(1 - \frac{k_n}{n}y), Q(1 - \frac{k_n}{n}x)]$.

Définition 2.5.1 : " ρ -mélangeance"

Une série $(X_t)_t$ est dite ρ -mélangeante, s'il existe une suite $\rho(\ell)$ qui tend vers 0 quand ℓ tend vers ∞ , telle que :

$$\rho(\ell) := \sup_{m \in \mathbb{N}} \sup_{A \in L^2(\mathcal{B}_1^m), B \in L^2(\mathcal{B}_{m+\ell+1}^\infty)} \frac{|cov(A, B)|}{\sqrt{var(A)var(B)}}, \quad \forall m, \quad (2.5.2)$$

$L^2(\mathcal{B}_{m+\ell+1}^\infty)$ (resp. $L^2(\mathcal{B}_1^m)$) est l'espace des fonctions de carré intégrable, $\mathcal{B}_{m+\ell+1}^\infty$ -mesurable (resp. \mathcal{B}_1^m -mesurable).

Proposition 2.5.1 :

Supposons que $\ell_n = o(\frac{n}{k_n})$ et que les conditions (C2), (C3) sont vérifiées. Soit $\varepsilon > 0$, alors pour tout $x, y; \quad 0 < x, y \leq 1 + \varepsilon$:

$$\lim_{n \rightarrow \infty} \frac{n}{\ell_n k_n} cov \left(\sum_{i=1}^{\ell_n} \mathbf{1}_{\{X_i > Q(1 - \frac{k_n}{n}x)\}}, \sum_{i=1}^{\ell_n} \mathbf{1}_{\{X_i > Q(1 - \frac{k_n}{n}y)\}} \right) = C(x, y), \quad (2.5.3)$$

avec

$$C(x, y) := x \wedge y + \sum_{m=1}^{\infty} [C_m(x, y) + C_m(x, y)]. \quad (2.5.4)$$

De plus, il existe $d > 0$, tel que pour tout x et $y; \quad 0 < x, y \leq 1 + \varepsilon$ et pour tout $n \in \mathbb{N}$:

$$\frac{n}{\ell_n} \mathbb{E} \left[\sum_{i=1}^{\ell_n} \mathbf{1}_{\{X_i \in I_n(x, y)\}} \right]^2 \leq d(y - x). \quad (2.5.5)$$

Preuve :

Dans (C3), choisissant $y = 1 + \varepsilon$ et faisant tendre x vers 0, on obtient :

$$\begin{aligned} \frac{n}{k_n} P \left\{ X_1 > Q \left(1 - \frac{k_n}{n}(1 + \varepsilon) \right); X_{m+1} > Q \left(1 - \frac{k_n}{n}(1 + \varepsilon) \right) \right\} \\ \leq (1 + \varepsilon)(\tilde{\rho}(m) + \delta \frac{k_n}{n}). \end{aligned}$$

D'autre part,

$$\lim_{n \rightarrow \infty} \sum_{m=1}^{\ell_n} (\tilde{\rho}(m) + (\delta + (1 + \varepsilon)^2) \frac{k_n}{n}) = \sum_{i=1}^{\ell_n} \tilde{\rho}(m) < \infty,$$

avec $\ell_n = o\left(\frac{n}{k_n}\right)$, donc $\frac{\ell_n k_n}{n} = o(1)$, ou encore $\left(\frac{\ell_n k_n}{n} \rightarrow 0, \text{ quand } n \rightarrow \infty\right)$.

Il s'ensuit que :

$$\lim_{n \rightarrow \infty} \frac{n}{k_n} \sum_{m=1}^{\ell_n} \text{cov} \left(\mathbf{1}_{\{X_1 > Q(1 - \frac{k_n}{n}x)\}}, \mathbf{1}_{\{X_{m+1} > Q(1 - \frac{k_n}{n}y)\}} \right) = \sum_{m=1}^{\infty} C_m(x, y) \in \mathbb{R}.$$

Comme la série $(X_t)_t$ est stationnaire (par hypothèse), alors :

$$\begin{aligned} & \frac{n}{k_n \ell_n} \sum_{1 \leq i < j \leq \ell_n} \text{cov} \left(\mathbf{1}_{\{X_i > Q(1 - \frac{k_n}{n}x)\}}, \mathbf{1}_{\{X_j > Q(1 - \frac{k_n}{n}y)\}} \right) \\ &= \frac{1}{\ell_n} \sum_{i=1}^{\ell_n} \frac{n}{k_n} \sum_{j=i+1}^{i+\ell_n-1} \text{cov} \left(\mathbf{1}_{\{X_i > Q(1 - \frac{k_n}{n}x)\}}, \mathbf{1}_{\{X_j > Q(1 - \frac{k_n}{n}y)\}} \right) \\ & \quad - \frac{1}{\ell_n} \sum_{i=2}^{\ell_n} \frac{n}{k_n} \sum_{j=\ell_n+1}^{i+\ell_n-1} \text{cov} \left(\mathbf{1}_{\{X_i > Q(1 - \frac{k_n}{n}x)\}}, \mathbf{1}_{\{X_j > Q(1 - \frac{k_n}{n}y)\}} \right), \end{aligned}$$

or, le second terme de cette égalité est borné par

$$\sum_{m=1}^{\ell_n} \frac{m}{\ell_n} (\tilde{\rho}(m) + (\delta + (1 + \varepsilon)^2) \frac{k_n}{n}) \rightarrow 0,$$

et le premier terme tend vers :

$$\sum_{m=1}^{\infty} C_m(x, y), \quad \text{quand } n \rightarrow \infty,$$

ce qui implique **(2.5.3)**.

D'autre part on a ;

$$\begin{aligned} \frac{n}{\ell_n k_n} \mathbb{E} \left[\sum_{i=1}^{\ell_n} \mathbf{1}_{\{X_i \in I_n(x, y)\}} \right]^2 &= \frac{n}{\ell_n k_n} \sum_{i=1}^{\ell_n} \mathbb{E} \left[\left(\mathbf{1}_{\{X_i \in I_n(x, y)\}} \right)^2 \right] \\ & \quad + \frac{n}{k_n \ell_n} \sum_{i \neq j}^{\ell_n} \mathbb{E} \left[\mathbf{1}_{\{X_i \in I_n(x, y)\}}, \mathbf{1}_{\{X_j \in I_n(x, y)\}} \right] \\ &= \frac{n}{k_n \ell_n} \sum_{i=1}^{\ell_n} P \{X_i \in I_n(x, y)\} \\ & \quad + \frac{n}{k_n \ell_n} \sum_{i \neq j}^{\ell_n} P \{X_i \in I_n(x, y), X_{m+1} \in I_n(x, y)\} \\ &\leq (y - x) \left[1 + 2 \sum_{m=1}^{\ell_n} (\tilde{\rho}(m) + \delta \frac{k_n}{n}) \right] \\ &\leq (y - x) (\tilde{\rho}(m) + \delta \frac{k_n}{n}), \end{aligned}$$

car la série $\sum_{m=1}^{\infty} \tilde{\rho}(m) < \infty$ et $\frac{\ell_n k_n}{n} = o(1)$. C.Q.D. ■

Introduisons maintenant des conditions sur la vitesse de convergence et la suite intermédiaire (k_n) comme suit :

(C4) Supposons que la fonction des quantiles $Q(\cdot)$ admet la représentation suivante :

$$Q(1-t) = dt^{-\gamma}(1+r(t)), \quad \text{avec } |r(t)| \leq \Phi(t),$$

- pour tout $d > 0$ et tout fonction Φ à τ -variation au voisinage de 0^1 .

- pour tout $\tau \geq 0$ et Φ croissante avec : $\lim \Phi(t) = 0$, quand $t \downarrow 0$.

(C5) La suite intermédiaire (k_n) vérifiée :

$$\lim_{n \rightarrow \infty} k_n^{1/2} \Phi\left(\frac{k_n}{n}\right) = 0.$$

Théorème 2.5.1 :

Sous les conditions (C1), (C2), (C3), (C4) et (C5) avec $\frac{\ell_n k_n}{n} = o(1)$, alors il existe un processus gaussien $e(\cdot)$ centré d'une fonction de covariance C , tel que quand n tend vers ∞ :

$$\sup_{t \in]0,1]} t^{\gamma+1/2} (1 + |\log t|)^{-1/2} \left| k_n^{1/2} \left(\frac{Q_n(t)}{Q(1 - k_n/n)} - t^{-\gamma} \right) - \gamma t^{-(\gamma+1)} e(t) \right| \xrightarrow{P} 0, \quad (2.5.6)$$

où $Q_n(t) = Q_{n, k_n}(t) := X_{n - [k_n t]}$, $t \in]0, 1]$.

Démonstration :

Etant donnée **(C4)**, le terme $\Psi(\lambda, t)$ donné par **(2.3.1)** ;

$$\Psi(\lambda, t) = t^{-\gamma} O(\Phi(\lambda t) + \Phi(\lambda)),$$

uniformément pour t et ce, puisque Φ est croissante à τ -variation au voisinage de 0 ($\tau > 0$).

¹Une fonction Φ est à τ -variation au voisinage de 0, si et seulement si pour une fonction auxiliaire $a(\cdot)$:

$$\lim_{t \rightarrow \infty} \frac{\Phi\left(\frac{1}{tx}\right) - \Phi\left(\frac{1}{t}\right)}{a(t)} = \log(x).$$

En plus, la condition **(C5)** et le fait que Φ croissante avec : $\lim \Phi(t) = 0$, quand $t \downarrow 0$, implique que :

$$\sup_{t \in]0,1]} t^{\gamma+1/2} (1 + |\log t|)^{-1/2} \left| \Psi\left(\frac{k_n}{n}, t\right) \right| \xrightarrow{P} 0,$$

en combinaison avec la proposition **2.5.1**, les conditions du théorème **3.1** dans Drees, H. (2000) [21] établissent pour le processus empirique uniforme de la queue de distribution sont vérifiées, ce qui implique l'assertion. Remarquons seulement que dans le cas des observations indépendantes, le processus gaussien $e(\cdot)$ est un mouvement brownien standard. ■

Pour établir la normalité asymptotique pour la classe des estimateurs de l'index des valeurs extrêmes $\hat{\gamma}$, qui s'écrivent en fonction d'un fonctionnel statistique des queues de distribution (i.e., $\hat{\gamma} := T(Q_n)$), on a besoin des conditions de régularité de cet fonctionnel statistique T , ces conditions sont :

[T0] T est un fonctionnel à valeur réelle, mesurable par rapport à la tribu Borélien dans l'ensemble des fonctions $z \in \mathbb{D}[0, 1]$:

$$\lim_{t \downarrow 0} t^{\gamma+1/2} |\log t|^{-1/2} z(t) \rightarrow 0.$$

[T1] $T(az) = T(z)$, $\forall a > 0$.

[T2] $T((t^{-\gamma})_{1 < t \leq 1}) = \gamma$.

[T3] Il existe une mesure signée² $\vartheta_{T,\gamma}$ dans $]0, 1]$ avec

$$\int_{]0,1]} t^{-\gamma-1/2} (1 + |\log t|)^{-1/2} |\vartheta_{T,\gamma}|(dt) < \infty,$$

telle que :

$$\lim_{\varepsilon_n \downarrow 0} \varepsilon_n^{-1} [T((t^{-\gamma} + \varepsilon_n z_n(t))_{1 < t \leq 1}) - T((t^{-\gamma})_{1 < t \leq 1})] \rightarrow \int_{]0,1]} z(t) \vartheta_{T,\gamma}(dt),$$

et z_n satisfait :

$$\sup_{t \in]0,1]} t^{\gamma+1/2} (1 + |\log t|)^{-1/2} |z_n(t) - z(t)| \rightarrow 0,$$

pour toute fonction z définie en **[T0]**.

² *Mesure signée* : ou mesure réelle généralisée : est une fonction d'ensembles σ -additive, définie sur un clan \mathfrak{S} à valeurs dans \mathbb{R} et prend au plus l'une des deux valeurs $+\infty$ et $-\infty$.

Avec la condition **(C5)** précédente, on ajoute la condition suivante :

$$\lim_{n \rightarrow \infty} \frac{1}{\sqrt{k_n}} \log(np_n) = 0 \quad \text{et} \quad \lim_{n \rightarrow \infty} \frac{np_n}{k_n} = 0. \quad (2.5.7)$$

Théorème 2.5.2 : (Drees, H. (2002)) [22]

Sous les conditions de théorème 2.5.1, la condition (2.5.7) et si $\hat{\gamma}_n = T(Q_n)$, avec T satisfait les conditions de régularité [T]. Alors

$$\begin{aligned} \frac{\sqrt{k_n}}{\log(k_n/(np_n))} \log\left(\frac{\hat{x}_{p_n}}{x_{p_n}}\right) &\sim \frac{\sqrt{k_n}}{\log(k_n/(np_n))} \left(\frac{\hat{x}_{p_n}}{x_{p_n}} - 1\right) \\ &\sim \sqrt{k_n}(\hat{\gamma}_n - \gamma) \xrightarrow{\mathcal{D}} \mathcal{N}(o, \sigma_{T,\gamma}^2), \end{aligned} \quad (2.5.8)$$

faiblement avec :

$$\sigma_{T,\gamma}^2 := \gamma^2 \iint_{]0,1]^2} (st)^{-(\gamma+1)} C(s,t) \vartheta_{T,\gamma}(ds) \vartheta_{T,\gamma}(dt),$$

où $N(o, \sigma^2)$: désigne la loi normale centrée de variance σ^2 .

Démonstration :

Premièrement, pour montrer la convergence faible de $\sqrt{k_n}(\hat{\gamma}_n - \gamma)$ vers $\mathcal{N}(o, \sigma_{T,\gamma}^2)$, on a d'après la condition **[T1]** ;

$$\begin{aligned} \hat{\gamma}_n &= T(Q_n) = T\left(\frac{Q_n}{Q(1 - k_n/n)}\right) \\ &\stackrel{\mathcal{D}}{=} T\left[(t^{-\gamma} + k_n^{-1/2} \gamma t^{-(\gamma+1)} e(t) + o(k_n^{-1/2}))_{1 < t \leq 1}\right], \end{aligned}$$

et d'après **[T3]**

$$\hat{\gamma}_n = T((t^{-\gamma})_{1 < t \leq 1}) + \frac{\gamma}{\sqrt{k_n}} \int_{]0,1]} t^{-\gamma-1/2} e(t) \vartheta_{T,\gamma}(dt) + o(\sqrt{k_n}),$$

finalement, par **[T1]** on trouve que :

$$\sqrt{k_n}(\hat{\gamma}_n - \gamma) \rightarrow \gamma \int_{]0,1]} t^{-\gamma-1/2} e(t) \vartheta_{T,\gamma}(dt).$$

ce qui prouve l'assertion.

Deuxièmement, comme on a : $\log(1+x) \sim x$, quand $x \downarrow 0$. Alors

$$\log\left(\frac{\hat{x}_{p_n}}{x_{p_n}}\right) \sim \left(\frac{\hat{x}_{p_n}}{x_{p_n}} - 1\right),$$

donc pour montrer **(2.5.8)**, ils suffit de montrer que

$$\frac{\sqrt{k_n}}{\log(k_n/(np_n))} \left(\frac{\hat{x}_{p_n}}{x_{p_n}} - 1 \right) \sim \sqrt{k_n} (\hat{\gamma}_n - \gamma),$$

où ce qui revient à dire :

$$\frac{1}{\log(k_n/(np_n))} \left(\frac{\hat{x}_{p_n}}{x_{p_n}} - 1 \right) = \hat{\gamma}_n - \gamma + o(\sqrt{k_n}). \quad (2.5.9)$$

Écrivent

$$\frac{\hat{x}_{p_n}}{x_{p_n}} - 1 = \frac{\hat{x}_{p_n} - x_{p_n}}{x_{p_n}},$$

avec,

$$\begin{aligned} \hat{x}_{p_n} - x_{p_n} &= Q_n(1) - Q(1 - p_n) \\ &= (Q_n(1) - Q(1 - k_n/n)) \left(\frac{np_n}{k_n} \right)^{-\hat{\gamma}} \\ &\quad + Q(1 - k_n/n) \left(\left(\frac{np_n}{k_n} \right)^{-\hat{\gamma}} - \left(\frac{np_n}{k_n} \right)^{-\gamma} \right) \\ &\quad + Q(1 - k_n/n) \left(\left(\frac{np_n}{k_n} \right)^{-\gamma} - Q(1 - p_n) \right) \\ &: = I_n + II_n + III_n. \end{aligned}$$

Le théorème 2.5.1, et les conditions **(C4)** et **(C5)** impliquent :

$$\begin{aligned} \left| \frac{Q(1 - p_n)}{Q(1 - k_n/n)} \left(\frac{np_n}{k_n} \right)^{\gamma} - 1 \right| &= \left| \Psi\left(\frac{k_n}{n}, \frac{np_n}{k_n} \right) \right| \left(\frac{np_n}{k_n} \right)^{\gamma} \\ &= O\left(\Phi\left(\frac{k_n}{n} \right) \right) = o(k_n^{-1/2}). \end{aligned}$$

Toujours d'après le théorème 2.5.1, la condition **(2.5.7)** et le fait que :

$$\hat{\gamma}_n - \gamma = O_p(k_n^{-1/2}),$$

alors

$$\begin{aligned} \frac{I_n}{x_{p_n} \log(k_n/np_n)} &= k_n^{-1/2} (e(1) + o(1)) \frac{Q(1 - k_n/n)}{Q(1 - p_n) \log(k_n/np_n)} \left(\frac{np_n}{k_n} \right)^{-\hat{\gamma}} \\ &= o(k_n^{-1/2}). \end{aligned}$$

D'autre part, puisque $\frac{\partial}{\partial t} x^t = x^t \log x$, utilisons le théorème de la valeur moyenne et **(2.5.7)**,

$$\begin{aligned} \frac{II_n}{x_{p_n} \log(k_n/np_n)} &= \frac{Q(1 - k_n/n)}{Q(1 - p_n)} \left(\frac{np_n}{k_n}\right)^{-\gamma} \left(\frac{np_n}{k_n}\right)^{\theta(\gamma - \hat{\gamma})} (\hat{\gamma} - \gamma) \\ &= (\hat{\gamma} - \gamma)(1 + o(1)), \quad \forall \theta \in]0, 1]. \end{aligned}$$

Finalement,

$$\begin{aligned} \frac{III_n}{x_{p_n} \log(k_n/np_n)} &= \left(\frac{Q(1 - k_n/n)}{Q(1 - p_n)} \left(\frac{np_n}{k_n}\right)^{-\gamma} \right) / \log\left(\frac{k_n}{np_n}\right) \\ &= o(k_n^{-1/2}), \end{aligned}$$

ce qui prouve **(2.5.9)**, et le théorème est démontré. ■

Remarques 2.5.1 :

1. Si $(F \in D(G_\gamma))$, la condition suivante, qui remplace **(C4)** et **(C5)** est toujours satisfaite $\forall k_n \rightarrow \infty$:

$$k_n^{1/2} \sup_{t \in]0, 1 + \varepsilon]} t^{\gamma+1/2} (1 + |\log t|)^{-1/2} \left| \Psi\left(\frac{k_n}{n}, t\right) \right| \rightarrow 0, \text{ quand } n \rightarrow \infty, \quad \textbf{(2.5.10)}$$

sous les conditions **(C1)**, **(C3)** et **(2.5.10)**, l'assertion du théorème 2.5.1 est donc vérifiée.

2. Pour démontrer la normalité asymptotique de l'estimateur des quantiles extrêmes, on ajoute à la condition **(2.5.7)** la condition suivante :

$$\frac{\sqrt{k_n}}{\log(k_n/(np_n))} \left(\frac{np_n}{k_n}\right)^\gamma \Psi\left(\frac{k_n}{n}, \frac{np_n}{k_n}\right) \rightarrow 0, \quad \text{quand } n \rightarrow \infty,$$

cette convergence implique que :

$$\begin{aligned} \frac{Q(1 - p_n)}{Q(1 - k_n/n)} \left(\frac{np_n}{k_n}\right)^\gamma - 1 &= \left(\frac{np_n}{k_n}\right)^\gamma \Psi\left(\frac{k_n}{n}, \frac{np_n}{k_n}\right) \\ &= o\left(k_n^{-1/2} \log \frac{k_n}{np_n}\right), \end{aligned}$$

ce qui permet d'établir la convergence **(2.5.8)** dans le théorème précédent.

Chapitre 3

Adéquation au modèle GARCH(1, 1)

Le problème qui se pose lorsque on veut appliquer les divers résultats établis dans le chapitre précédente, à des séries temporelles réelles (séries financières en particulier), est comment assurer l'ajustement de ces données à l'un des modèles vérifiant les divers conditions de : stationnarité, hétéroscédasticité et mélangeance... ?.

Ce chapitre, donc, a pour but de résoudre ce problème, du moins pour le modèle GARCH(1, 1), mais comme nous le verrons ci-après, le principe reste le même soit pour le cas général GARCH(p, q), soit pour un ARCH(p) ou ARCH(1) qui sont des cas particuliers du premier...

La construction d'un test d'adéquation au modèle GARCH(1, 1) est similaire aux tests classiques : *Kolmogorov* (1933), *Smirnov* (1939, 1941), *Cramér* (1928) et *von Mises* (1931)... à la différence que celle-ci s'est penchée sur le domaine des fréquences, et ce, en se basant sur des transformations des distributions et densités spectrales empiriques.

Pour aborder ce sujet, les étapes suivantes sont inéluctables :

- Construction du modèle GARCH(1, 1).
- Relations entre les modèles GARCH, AR et ARMA.
- Propriétés des distributions du modèle.
- Estimation des paramètres du modèle.

3.1 Construction du modèle

L'un des modèles classiques qui fournit une bonne représentation des phénomènes financiers comme : les taux de change, prix d'actifs ou d'option, rendements d'indice boursier, . . . etc, en terme d'un processus $(P_t)_t$ qui désigne l'évolution de ces phénomènes est celui de *Black* et *Scholes* :

$$P_t := P_0 \exp(\sigma_t W_t + \mu_t t),$$

où W_t est un mouvement brownien, σ_t et μ_t sont des processus stochastiques. Ce modèle, qui est basé sur l'équation de diffusion de type :

$$dP_t = \mu_t P_t dt + \sigma_t P_t dW_t,$$

ou encore sous forme d'intégrale :

$$\log \frac{P_t}{P_0} = \int_0^t (\mu_s - \frac{1}{2} \sigma_s^2) ds + \int_0^t \sigma_s dW_s.$$

A l'origine de ce modèle, le processus σ_t est supposé constant et on dit qu'il représente la volatilité de P_t . Et on a deux cas : le premier est que σ_t et μ_t sont des processus déterministes ; les log rendements $(\log(\frac{P_t}{P_{t-1}}), t = 1, 2, \dots)$ dans une intervalle $]t-1, t]$ sont des variables aléatoires indépendantes d'une distribution normale, de moyenne et variance données respectivement par :

$$\bar{\mu}_t = \int_{t-1}^t (\mu_s - \frac{1}{2} \sigma_s^2) ds \quad \text{et} \quad \bar{\sigma}_t^2 = \int_{t-1}^t \sigma_s^2 dW_s,$$

en d'autre terme, les variables :

$$\xi_t = \frac{1}{\bar{\sigma}_t} (\log(\frac{P_t}{P_{t-1}}) - \bar{\mu}_t),$$

forment un échantillon i.i.d, d'une loi normale standard centrée et réduite. Dans ce cas $\bar{\sigma}_t$ est appelée la "*volatilité moyenne*" sur l'intervalle $]t-1, t]$.

Le deuxième cas, si σ_t et μ_t sont indéterministes, alors le processus $(\xi_t)_t$ n'est pas nécessairement gaussien. De plus, si l'unité du temps est petite, alors on peut dire que $(\xi_t)_t$ est d'une distribution approximativement normale, de même que les processus σ_t et μ_t sont approximativement constants, on peut les remplacer donc par $\bar{\sigma}_t$ et $\bar{\mu}_t$ respectivement.

Pour simplifier les notations, on prend $\bar{\mu}_t = 0$, ce qui revient à dire que :

$$\log\left(\frac{P_t}{P_{t-1}}\right) = \sigma_t \xi_t =: X_t,$$

cela définit un modèle de type (ARCH, GARCH,...), d'un processus à volatilité stochastique σ_t et un processus centré réduit $(\xi_t)_t$.

Un des apports de ces modèles hétéroscédastiques, était donc de mieux s'ajuster aux données, en particulier aux données financières.

Pour cela, on pose

$$X_t := \log\left(\frac{P_t}{P_{t-1}}\right), t = 1, 2, \dots$$

P_t désigne : les rendements d'indice, prix d'actions,... etc.

Notre étude sera donc basée sur les caractéristiques de la série temporelle $(X_t)_t$, dont le modèle souhaitable pour la représenter est de type hétéroscédastique : ARCH, GARCH,... Comme cas important de ce type, on a choisi le modèle GARCH(1, 1), qui est une restriction du modèle GARCH(p, q) d'une part, et d'autre part est une généralisation du modèle ARCH(1) .

Définition 3.1.1 : "*Modèle GARCH(1, 1)*"

On dit qu'un modèle X_t est autorégressif conditionnellement hétéroscédastique généralisé (GARCH(1, 1)), s'il admet l'écriture suivante :

$$X_t := \sigma_t \xi_t, \tag{3.1.1}$$

et dont la variance conditionnelle satisfaite pour tout $t \in \mathbb{Z}$ et α_0, α et β des constantes positives données :

$$\sigma_t^2 := \alpha_0 + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{3.1.2}$$

avec σ_t est \mathcal{F}_t -mesurable, $(\xi_t)_t$ est un bruit blanc (i.e. une suite de variables aléatoires, indépendantes et identiquement distribuées), telle que

$$\mathbb{E} [\xi_t | \mathcal{F}_{t-1}] = 0, \quad \mathbb{E} [\xi_t^2 | \mathcal{F}_{t-1}] = 1.$$

Propriété 3.1.1 : *Le modèle GARCH(1, 1) à les propriétés suivantes :*

1. $var(X_t|\mathcal{F}_{t-1}) = \alpha_0 + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2 = \sigma_t^2,$

on dit donc que X_t est hétéroscédastique conditionnellement à X_{t-1} .

2. $var(X_t) = \alpha_0 + (\alpha + \beta)var(X_{t-1}).$

Preuve : (Semble à celle de la propriété (1.4.1))

1.

$$\begin{aligned} var(X_t|\mathcal{F}_{t-1}) &= \mathbb{E}[X_t^2|\mathcal{F}_{t-1}] - \mathbb{E}[X_t|\mathcal{F}_{t-1}]^2 \\ &= \mathbb{E}[\sigma_t^2 \xi_t^2|\mathcal{F}_{t-1}] - \mathbb{E}[\sigma_t \xi_t|\mathcal{F}_{t-1}]^2 \\ &= \alpha_0 + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2 = \sigma_t^2. \end{aligned}$$

2.

$$\begin{aligned} var(X_t) &= var(\mathbb{E}[X_t|\mathcal{F}_{t-1}]) + \mathbb{E}[var(X_t|\mathcal{F}_{t-1})] \\ &= var(0) + \mathbb{E}[\alpha_0 + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2|\mathcal{F}_{t-1}]. \end{aligned}$$

Comme,

$$\mathbb{E}[X_{t-1}^2|\mathcal{F}_{t-1}] = \mathbb{E}[\sigma_{t-1}^2 \xi_{t-1}^2|\mathcal{F}_{t-1}] = \mathbb{E}[\sigma_{t-1}^2|\mathcal{F}_{t-1}] = var(X_{t-1}),$$

donc,

$$var(X_t) = \alpha_0 + (\alpha + \beta)var(X_{t-1}).$$

Sous hypothèse de stationnarité, (i.e. $var(X_t) = var(X_{t-1})$), ce qui implique que :

$$var(X_t) = \frac{\alpha_0}{1 - (\alpha + \beta)}, \quad (3.1.3)$$

pour que cette variance existe, il faut donc que : $\alpha + \beta < 1$. ■

3.1.1 Relations entre : *GARCH*, *AR* et *ARMA*

C'est à partir des relations qu'entretient les modèles GARCH et les modèles autorégressif et autorégressif moyenne mobile (*AR*, *ARMA*, ...) qu'on ait arriver à la construction des premiers modèles.

Conciderant la variance conditionnelle σ_t^2 d'un modèle GARCH(1, 1), et on réécrit l'équation (3.1.2), en utilisant le fait de récurrence ($X_t = \sigma_t \xi_t$, $X_{t-1} = \sigma_{t-1} \xi_{t-1}$, ...), donc

$$\sigma_t^2 := \alpha_0 + (\beta + \alpha \xi_{t-1}^2) \sigma_{t-1}^2, \quad (3.1.4)$$

ce qui définit un modèle $Y_t = \sigma_t^2$ autorégressif d'ordre 1 (*AR*(1)), dont le coefficient est une variable aléatoire ($\beta + \alpha \xi_{t-1}^2$) et le bruit est une fonction déterministe (définition 1.4.2).

De même, si on suppose que la série $(X_t)_t$ est stationnaire, i.e.

$$\mathbb{E}[X_t^2] = \mathbb{E}[\sigma_t^2] = \frac{\alpha_0}{1 - \alpha - \beta},$$

ne dépend pas du temps t , et si on pose $(U_t)_t := (X_t^2 - \sigma_t^2)_t$, alors :

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2 \\ &= \alpha_0 + \alpha X_{t-1}^2 + \beta X_{t-1}^2 - \beta X_{t-1}^2 + \beta \sigma_{t-1}^2, \end{aligned}$$

donc

$$X_t^2 - U_t = \alpha_0 + (\alpha + \beta) X_{t-1}^2 - \beta U_{t-1}.$$

D'où

$$(1 - (\alpha + \beta)L)X_t^2 = \alpha_0 + (1 - \beta L)U_t,$$

L : désigne l'opérateur retard (définition 1.3.1).

On conclut que le processus X_t^2 admet l'écriture d'un modèle linéaire ARMA(1, 1) centré, d'un bruit blanc $U_t = X_t^2 - \sigma_t^2$, polynôme autorégressif $1 - (\alpha + \beta)L$ et moyenne mobile $1 - \beta L$, avec

$$\mathbb{E}[U_t] = \mathbb{E}[X_t^2] - \mathbb{E}[\mathbb{E}[X_{t-1}^2 | \mathcal{F}_{t-1}]] = 0,$$

une relation qu'on va l'utiliser pour calculer la variance de la variable aléatoire $(X_t X_{t+h})$, (i.e. $\text{var}(X_t X_{t+h})$), cette quantité notée par $v_X(h)$, et que l'on va utiliser dans la construction du test d'adéquation.

En effet,

$$\begin{aligned}
v_X(h) &= \text{var}(X_t X_{t+h}) = \mathbb{E}[(U_t + \sigma_t^2)(U_{t+h} + \sigma_{t+h}^2)] \\
&= \mathbb{E}[U_t U_{t+h}] + \mathbb{E}[X^2]^2 \\
&= \gamma_U(h) + \text{var}^2(X).
\end{aligned}$$

Or, la valeur de la fonction d'autocovariance $\gamma_U(h)$ du processus U_t (modèle ARMA(1, 1)) est donnée dans *Brockwell et Davis* [7] comme suit :

$$\begin{aligned}
\gamma_U(0) &= \text{var}(U) \left[1 - \frac{\alpha^2}{1 - (\alpha + \beta)^2} \right], \\
\gamma_U(1) &= \text{var}(U) \left[\alpha + \frac{\alpha^2(\alpha + \beta)}{1 - (\alpha + \beta)^2} \right],
\end{aligned}$$

et pour tout $h \geq 2$,

$$\gamma_U(h) = (\alpha + \beta)^{h-1} \gamma_U(1).$$

De plus, puisque U est centré, donc

$$\begin{aligned}
\text{var}(U) &= \mathbb{E}[U_t^2] - \mathbb{E}[U_t]^2 = \mathbb{E}[(X_t^2 - \sigma_t^2)^2] \\
&= (\mathbb{E}[\xi^4] - 1) \text{var}^2(X),
\end{aligned}$$

ce qui donne, (en remplaçant $\text{var}(X)$ par sa valeur donnée par (3.1.4)) :

$$\text{var}(U) = \frac{1 + \alpha + \beta}{1 - (\alpha + \beta)} \frac{\alpha_0^2 (\mathbb{E}[\xi^4] - 1)}{1 - ((\alpha + \beta)^2 + \alpha^2 (\mathbb{E}[\xi^4] - 1))}.$$

Enfin, après simplification $v_X(h)$ vaut, pour $h \geq 2$:

$$v_X(h) = \text{var}^2(X) \left[\frac{\alpha(1 - \alpha\beta^2) \mathbb{E}[\xi^4] - 1}{1 - ((\alpha + \beta)^2 + \alpha^2 (\mathbb{E}[\xi^4] - 1))} (\alpha + \beta)^{h-1} + 1 \right]. \quad (3.1.5)$$

Il est clair que l'existence de cette quantité, est assurée par celle de moment d'ordre quatre du modèle (i.e. $\mathbb{E}[X^4] < \infty$). Or, cela d'après (3.1.5) est satisfaite si la condition suivante est vérifiée :

$$(\alpha + \beta)^2 + \alpha^2 (\mathbb{E}[\xi^4] - 1) < 1. \quad (3.1.6)$$

3.1.2 Propriétés des distributions du modèle

Malheureusement, la connaissance de la loi conditionnelle du modèle (et cela, même avec des hypothèses sur le bruit) est insuffisante. On doit chercher alors, si nous ne pouvons pas avoir la loi non conditionnelle de ce modèle, au moins ces caractéristiques intéressantes (en terme des moments d'ordre supérieur), en les comparant avec les autres lois analogues (à loi normale en générale). Pour cela on définit respectivement le coefficient d'aplatissement appelé *Kurtosis* (κ) et le coefficient d'asymétrie appelé *Skewness* (s), d'une série temporelle $(X_t)_t$ comme suit :

$$\kappa := \frac{\mathbb{E} [(X - \mathbb{E}(X))^4]}{\mathbb{E} [(X - \mathbb{E}(X))^2]^2}, \quad s := \frac{\mathbb{E} [(X - \mathbb{E}(X))^3]}{\mathbb{E} [(X - \mathbb{E}(X))^2]^{3/2}}.$$

L'une des propriétés importantes de la distribution d'un modèle hétéroscédastique est qu'elle est à queue lourde (i.e. distribution *Liptokrotique*¹). Il est évident que pour une variable aléatoire centrée réduite $\mathcal{N}(0, 1)$, la kurtosis κ (resp. le skewness s) vaut 3 (resp. 0).

On a vu (propriété 3.1.1), qu'un modèle GARCH(1, 1) suit conditionnellement à son passé (sous-hypothèse de normalité du bruit ξ_t) une loi normale centrée d'une variance σ_t^2 :

$$X_t | \mathcal{F}_{t-1} \sim \mathcal{N}(0, \sigma_t^2).$$

En particulier,

$$\begin{aligned} \mathbb{E}[X_t] &= \mathbb{E}[\mathbb{E}(X_t | \mathcal{F}_{t-1})] = 0 \\ \text{et } \mathbb{E}[X_t^2] &= \alpha_0 + (\alpha + \beta)\mathbb{E}[X_{t-1}^2]. \end{aligned}$$

De plus, sous hypothèse de stationarité, la variance non conditionnelle existe pour $\alpha + \beta < 1$, et vaut

$$\text{var}(X_t) = \mathbb{E}[X_t^2] = \frac{\alpha_0}{1 - \alpha - \beta}.$$

¹Une distribution est liptokrotique si le coefficient d'aplatissement (**kurtosis**) est plus grand que 3, celle de la distribution normale.

Maintenant, pour calculer la kurtosis d'un modèle $GARCH(1, 1)$ centré et du bruit blanc indépendant de X_t et de σ_t^2 , il suffit de calculer le moment d'ordre quatre, on le divise en suite par le carré du moment d'ordre deux ;

$$\begin{aligned}\mathbb{E}[X_t^4] &= \mathbb{E}[\mathbb{E}(X_t^4|\mathcal{F}_{t-1})] = \mathbb{E}[\mathbb{E}(\xi_t^4\sigma_t^4|\mathcal{F}_{t-1})] \\ &= 3\mathbb{E}[\mathbb{E}[(\alpha_0 + \alpha X_{t-1}^2 + \beta\sigma_{t-1}^2)^2|\mathcal{F}_{t-1}]] \\ &= 3\mathbb{E}[(\alpha_0 + \alpha X_{t-1}^2 + \beta\sigma_{t-1}^2)^2] \\ &= 3(\alpha_0^2 + 2\alpha_0(\alpha + \beta)\mathbb{E}[X_{t-1}^2] + (\alpha + \beta)^2\mathbb{E}[X_{t-1}^4]),\end{aligned}$$

sous l'hypothèse de stationnarité, on trouve que

$$\mathbb{E}[X_t^4] = 3(\alpha_0^2 + \frac{2\alpha_0^2(\alpha + \beta)}{1 - \alpha - \beta} + (\alpha + \beta)^2\mathbb{E}[X_{t-1}^4]),$$

d'où :

$$\mathbb{E}[X_t^4] = \frac{3\alpha_0^2(1 + \alpha + \beta)}{(1 - \alpha - \beta)(1 - 3\alpha^2 - \beta^2 - 2\alpha\beta)}.$$

Alors, le moment d'ordre quatre d'un modèle stationnaire $GARCH(1, 1)$ existe à condition que :

$$3\alpha^2 + \beta^2 + 2\alpha\beta < 1, \quad (\mathbf{3.1.7})$$

dans ce cas, le coefficient d'aplatissement (**kurtosis**) est donné par :

$$\begin{aligned}\kappa &= \mathbb{E}[X_t^4] / \mathbb{E}[X_t^2]^2 \\ &= 3 \frac{1 - (\alpha + \beta)^2}{(1 - 3\alpha^2 - \beta^2 - 2\alpha\beta)} \\ &= 3 \frac{1 - (\alpha + \beta)^2}{1 - (\alpha + \beta)^2 - 2\alpha^2} > 3.\end{aligned}$$

Remarque 3.1.1 :

Il est clair que cette quantité supérieure à 3 (la kurtosis de la distribution normale), mais il faut que la condition **(3.1.7)** soit vérifiée pour assurer l'existence du moment d'ordre 4, ainsi que la kurtosis κ de la distribution non conditionnelle du modèle $GARCH(1, 1)$. Dans ce cas, on dit que la distribution d'un modèle $GARCH$ stationnaire est leptokurtique, ce qui signifie qu'elle a une queue lourde par rapport à celle d'une distribution normale.

3.1.3 Estimation des paramètres du modèle

Quand on est on face du problème d'ajustement d'un modèle hétéroscédastiques aux données, la première étape est de choisir l'ordre (p, q) et le type - classe - du modèle à utiliser. Généralement, pour le choix du modèle, l'approche la plus utile est d'estimer un modèle, faire une validation et passer à d'autres modèles,... ainsi de suite jusqu'à ce que nous arrivions à la classe préférable.

Une démarche inspirée de celle de *Box et Jenkins*, pour les modèles ARIMA (*Gourieroux, C et Monfort, A* [28]), qui s'appuie sur l'identification *a priori* des paramètres p et q (degré de retard) en utilisant des méthodes qui ne nécessitent pas l'estimation des autres paramètres $(\alpha_0, \alpha$ et $\beta)$.

L'estimation des paramètres des modèles hétéroscédastiques, est généralement fondée sur la méthode du maximum vraisemblance. L'optimisation est effectuée en supposant que la suite des variables aléatoires $(\xi_t)_t$ soit indépendante et identiquement distribuée, d'espérance nulle et de variance unité.

Notons par $\theta = (\alpha_0, \alpha, \beta)^\top$, le vecteur des paramètres du modèle à estimer, et par $\hat{\theta}$ sont estimateur (i.e. $\hat{\theta} = (\hat{\alpha}_0, \hat{\alpha}, \hat{\beta})^\top$).

Rappelons tout d'abord, le principe de la méthode du maximum vraisemblance. Considérons une variable aléatoire X , d'une loi de probabilité P_θ et densité f_θ qui dépendent du paramètre θ (θ appartient à une famille paramétrique $\Theta \subset \mathbb{R}^n$), à estimer en fonction des observations disponibles de la variable X (i.e. l'échantillon (x_1, x_2, \dots, x_n)). On appelle *fonction de vraisemblance* du paramètre θ , l'application L de Θ à valeur dans \mathbb{R}^+ , définie pour tout $\theta \in \Theta$ par :

$$L(\theta) = L(x_1, x_2, \dots, x_n; \theta) := \prod_{i=1}^n f_\theta(x_i),$$

avec $f_\theta(x_i)$: désigne la densité marginale de la densité jointe $f_\theta(x_1, x_2, \dots, x_n)$.

Le principe de la méthode de maximum vraisemblance, consiste à déterminer la valeur du paramètre θ en fonction des observations (x_1, x_2, \dots, x_n) , et qui assure la plus grande probabilité d'apparition de ces observations.

Définition 3.2.2 : "*Estimateur du maximum vraisemblance*"

On appelle estimateur du maximum vraisemblance d'un paramètre $\theta \in \Theta$, la statistique $\hat{\theta}$ qui fait correspondre au vecteur (x_1, x_2, \dots, x_n) la quantité $\hat{\theta}(x_1, x_2, \dots, x_n)$, telle que :

$$\forall \theta \in \Theta, \quad L(x_1, x_2, \dots, x_n; \hat{\theta}) \geq L(x_1, x_2, \dots, x_n; \theta).$$

Lorsque on suppose que cette fonction de vraisemblance $L(\cdot)$ est deux fois continûment différentiable par rapport à θ , $\forall \theta \in \Theta$, alors l'estimateur du maximum vraisemblance $\hat{\theta}$ constituer une solution du système suivant :

$$\left(\frac{\partial L(\theta)}{\partial \theta} \right)_{\theta=\hat{\theta}} = 0, \quad \text{et} \quad \left(\frac{\partial^2 L(\theta)}{\partial \theta^2} \right)_{\theta=\hat{\theta}} < 0. \quad (\mathbf{3.1.8})$$

Revenant maintenant au modèle GARCH (1, 1), comme nous ne connaissons pas la vraie distribution de ce modèle, on doit utiliser les faits que la distribution conditionnelle est normale et que le bruit ξ_t suit une loi gaussienne centré réduite, à savoir :

$$X_t | \mathcal{F}_{t-1} \sim \mathcal{N}(0, \sigma_t^2), \quad \frac{X_t}{\sigma_t} = \xi_t \sim \mathcal{N}(0, 1).$$

L'estimateur du maximum vraisemblance des paramètres d'un modèle GARCH(1, 1), est non linéaire et peut-être obtenu en résolvant les conditions de premier ordre **(3.1.8)**. Le log-vraisemblance du modèle se calcule généralement à partir de la distribution conditionnelle.

Pour T observations (X_1, X_2, \dots, X_T) , le log vraisemblance d'un modèle GARCH(1, 1) du paramètre $\theta = (\alpha_0, \alpha, \beta)^\top$ se calcule donc, à partir du produit de toutes les densités conditionnelles de X_t ;

$$L_T(\theta) = \sum_{t=1}^T \ell_t(\theta), \quad \text{et} \quad \ell_t(\theta) = \log f_\xi\left(\frac{X_t}{\sigma_t}\right),$$

f_ξ : désigne la densité de ξ_t , qui vaut dans le cas gaussien $\mathcal{N}(0, 1)$:

$$f_\xi\left(\frac{X_t}{\sigma_t}\right) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left\{-\frac{1}{2}\left(\frac{X_t}{\sigma_t}\right)^2\right\}.$$

D'où :

$$\ell_t(\theta) = \frac{-1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma_t^2) - \frac{1}{2} \left(\frac{X_t}{\sigma_t}\right)^2.$$

Le log-vraisemblance par rapport à θ vaut donc :

$$\begin{aligned}
L_T(\theta) &= \sum_{t=1}^T \left[\frac{-1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma_t^2) - \frac{1}{2} \left(\frac{X_t}{\sigma_t} \right)^2 \right] \\
&= \frac{-T}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T \log(\alpha_0 + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2) \\
&\quad - \frac{1}{2} \sum_{t=1}^T X_t^2 (\alpha_0 + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2)^{-1}.
\end{aligned}$$

Ce qu'on peut remarquer, c'est que la méthode de maximum vraisemblance pose un problème concernant la valeur initiale du processus de volatilité σ_t , quand on veut maximiser $L_T(\theta)$. Le processus σ_t dépend à la fois de X_{t-1} et de σ_{t-1} . Or, si la première observation a lieu au temps $t = 1$, les valeurs X_0 et σ_0 vont être spécifiées, donc par récurrence σ_t est calculée.

L'avantage de cette méthode est que les estimateurs sont consistants, la matrice des variances covariances égale à l'inverse de la matrice d'information. Par définition, la matrice d'information notée $I_T(\theta)$ est donnée par :

$$I_T(\theta) = \mathbb{E} \left[\frac{\partial L_T(\theta)}{\partial \theta} \frac{\partial L_T(\theta)}{\partial \theta^\top} \right] = -\mathbb{E} \left[\frac{\partial^2 L_T(\theta)}{\partial \theta \partial \theta^\top} \right]. \quad (3.1.9)$$

Dans le cas d'un modèle GARCH(1, 1)-gaussien,

$$\frac{\partial L_T(\theta)}{\partial \theta} = \sum_{t=1}^T \frac{\partial \ell_t(\theta)}{\partial \theta},$$

avec

$$\begin{aligned}
\frac{\partial \ell_t(\theta)}{\partial \theta} &= \frac{\partial}{\partial \theta} \left(\frac{-1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma_t^2) - \frac{1}{2} \left(\frac{X_t}{\sigma_t} \right)^2 \right) \\
&= \frac{-1}{2\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta} + \frac{1}{2\sigma_t^2} \frac{X_t^2}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta} \\
&= \frac{1}{2\sigma_t^2} \left(\frac{X_t^2}{\sigma_t^2} - 1 \right) \frac{\partial \sigma_t^2}{\partial \theta}.
\end{aligned}$$

Pour calculer la matrice d'information sous **(3.1.9)**, les valeurs passées de σ_t^2 et celles de $\frac{\partial \sigma_t^2}{\partial \theta}$ doivent être spécifiées (i.e. $\frac{\partial \sigma_{t-1}^2}{\partial \theta}, \frac{\partial \sigma_{t-2}^2}{\partial \theta}, \dots, \frac{\partial \sigma_0^2}{\partial \theta}$).

Une approche qui est largement utilisé est d'estimer la matrice d'information $I_T(\theta)$ par le *Hessian*, à savoir :

$$I_T(\theta) \simeq H := -\frac{\partial^2 L_T(\theta)}{\partial \theta \partial \theta^\top}, \quad (\mathbf{3.1.10})$$

la valeur de $\frac{\partial L_T(\theta)}{\partial \theta}$ est nulle quand θ maximise $L_T(\cdot)$, ce qui indique que

l'approximation de l'espérance dans **(3.1.9)** par les valeurs observées des dérivées d'ordre deux est toujours vérifiée, si aucune des contraintes sur les paramètres ne paraît :

$$\alpha_0 > 0, \alpha \geq 0, \beta \geq 0 \quad \text{et} \quad \alpha + \beta < 1. \quad (\mathbf{3.1.11})$$

Dans notre cas, la matrice des covariances est estimée par l'inverse du Hessian observé, que l'on calcule en appliquant des approximations numériques des dérivées du second ordre.

3.2 Test d'adéquation au modèle GARCH

3.2.1 Notion d'adéquation

Le problème d'adéquation en une simple forme consiste à tester l'ajustement des données à une fonction de distribution fixée. À savoir, pour un échantillon de taille n , de même fonction de distribution inconnue F , on veut tester l'hypothèse H_0 (que l'on appelle souvent l'hypothèse nulle) :

$$H_0 : F = F_0, \text{ pour } F_0 \text{ une distribution fixé.}$$

La première approche de ce genre de problèmes d'adéquation est celui proposé en 1900 par *Karl Pearson* (Le test de χ^2). Dans les années (1920-1930), une autre méthode a été proposée en se basant sur l'utilisation d'une distance fonctionnelle, qui mesure l'écart entre la distribution F_0 et la fonction de distribution empirique F_n ; *Cramér* (1928), *von Mises* (1931) ont proposer la mesure :

$$\omega_n^2 := \int_{\mathbb{R}} (F_n(x) - F_0(x))^2 \mu(x) dx,$$

$\mu(\cdot)$ désigne une fonction de poids.

Statistiquement, $F_n(\cdot)$ est un estimateur fortement consistant, sans biais de la distribution $F(\cdot)$. Il est clair, que pour x fixé, $F_n(x)$ suit une loi de *Bernoulli* de moyenne F_n et de variance $\frac{1}{n}F(x)(1 - F(x))$. Par conséquence, d'après la loi des grands nombres, F_n converge presque sûrement vers F quand n tend vers l'infini.

Théorème 3.2.1 : (Contelli 1917 - Glivenko 1930)

Soit $(X_n, n \in \mathbb{N}^*)$ une suite de variables aléatoires *i.i.d.* de fonction de distribution F_0 , alors

$$\lim_{n \rightarrow \infty} \sup_{-\infty < x < \infty} |F_n(x) - F_0(x)| \xrightarrow{P.S.} 0.$$

Ce théorème est celui le plus fondamental de la statistique mathématique, puisque *Kolmogrov* (1933), en ajoutant la vitesse de convergence a ce théorème, a réussi à étudier la statistique :

$$K_n := \sqrt{n} \sup_{-\infty < x < \infty} |F_n(x) - F_0(x)|,$$

Par la suite, *Smirnov*(1939, 1941) a proposé les deux statistiques suivantes :

$$K_n^+ := \sqrt{n} \sup_{-\infty < x < \infty} (F_n(x) - F_0(x)),$$

$$K_n^- := \sqrt{n} \sup_{-\infty < x < \infty} (F_0(x) - F_n(x)).$$

Les statistique K_n , K_n^+ et K_n^- sont appelées les statistique de *Kolmogrov-Smirnov*. Le célèbre résultat asymptotique concernant ces statistiques est donné sous le théorème suivant :

Théorème 3.2.2 :

Si F est une fonction de distribution continue, alors :

i) - (Kolmogrov 1933)

$$\lim_{n \rightarrow \infty} P \{K_n \leq x\} = K(x) := \sum_{k \in \mathbb{Z}} (-1)^k \exp(-2k^2 x^2), \quad x > 0.$$

ii)- (Smirnov 1941)

$$\lim_{n \rightarrow \infty} P \{K_n^+ \leq x\} = S(x) := 1 - \exp(-2x^2), \quad x > 0.$$

Définition 3.2.1 : "Pont brownien"

On appelle pont brownien sur $[0, 1]$, le processus stochastique $(B_t)_t$ tel que : $(B_t, t \in [0, 1]) \stackrel{\mathcal{D}}{=} (W_t - tW_1, t \in [0, 1])$, avec $(W_t)_t$ est un mouvement brownien standard. Donc il verifie les conditions suivantes :

1. La distribution jointe de $B_{t_1}, B_{t_2}, \dots, B_{t_n}$ ($0 \leq t_1 < t_2 < \dots < t_n \leq 1$; $n \geq 1$) est gaussienne, avec $\mathbb{E}[B_t] = 0$.
2. La fonction de covariance de B_t vaut,

$$cov(B_t, B_s) = \mathbb{E}[B_t B_s] = t \wedge s - ts,$$

cela implique que $B_0 = B_1 = 0$, *P.S.*

3. B_t est à trajectoires presque sûremnet continues.

L'existence d'un tel processus stochastique est une simple conséquence du lemme suivant :

Lemme 3.2.1 :

Pour tout processus de Weiner (ou mouvement brownien) $(W_t)_t$, le processus défini par :

$$W_t - tW_1, \quad t \in [0, 1], \quad (3.2.1)$$

est un pont brownien.

Preuve :

Les conditions (1, 2 et 3) précédentes, sont clairement vérifiées par la représentation (3.2.1). ■

En 1949, *Doob* a considéré une distribution finidémontionnelle, pour montrer la convergence du processus empirique $\sqrt{n}(F_n(x) - F_0(x))$ vers le suprémum d'un pont brownien $(B_t)_t$. En particulier, *Doob* a montré que :

$$P \left\{ \sup_{0 \leq t \leq 1} |B_t| \leq x \right\} = K(x) \quad \text{et} \quad P \left\{ \sup_{0 \leq t \leq 1} B_t \leq x \right\} = S(x),$$

où $K(x)$ (resp. $S(x)$) est la distribution de *Kolmogrov* (resp. *Smirnov*), donnée par le théorème précédent, ce qui prouve du nouveau les résultats établis par *Kolmogrov-Smirnov*.

3.2.2 Construction du test d'adéquation

Pour tester l'adéquation d'une série temporelle au modèle GARCH(1, 1), l'approche est la même décrite au paragraphe précédent, sauf que l'écart que l'on doit mesurer est celui entre la distribution spectrale empirique (appelée Périodogramme intégré $J_n(\lambda)$) et la distribution spectrale $F(\lambda)$ (i.e. $\sup(J_n(\lambda) - F(\lambda))$). Ce type de test joue un rôle important dans le problème de modélisation des séries temporelles.

Historiquement, ce sont *Bartlett* (1955), *Grenander et Rosenblatt* (1957) qui ont commencé par l'utilisation du périodogramme pour tester une hypothèse simple d'un bruit blanc. Cette idée a été généralisée selon le modèle sous hypothèse de test, linéaire ou non.

Considérons un modèle X_t atoregressif conditionnellement hétéroscédastique généralisé GARCH(1, 1), qui admet l'écriture suivante :

$$X_t := \sigma_t \xi_t \quad \text{avec} \quad \sigma_t^2 := \alpha_0 + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2,$$

pour tout $t \in \mathbb{Z}$, $\alpha_0 > 0$, α et β des constantes positives données, σ_t est \mathcal{F}_t -mesurable et $(\xi_t)_t$ est un bruit blanc (i.e. une suite de variables aléatoires i.i.d, centré et réduite), et qu'elle admet une densité de probabilité par rapport à la mesure de Lebesgue.

Comme on la dit dans la section 3.1.2, la distribution non conditionnelle de ce type des modèles est inconnue, donc la construction d'un test d'adéquation doit être faite en se basant sur la distribution spectrale (dans le domaine des fréquences).

Pour une série temporelle stationnaire $(X_t)_t$, d'une moyenne μ , d'une densité spectrale $f_X(\cdot)$ et d'une fonction d'autocovariance $\gamma_X(\cdot)$, la fonction d'autocovariance empirique $\gamma_{n,X}(\cdot)$ est donnée par :

$$\gamma_{n,X}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (X_t - \mu)(X_{t-h} - \mu), \quad h = 0, \dots, n-1.$$

De même en définit le périodogramme (voir la définition **1.3.1**), qui est l'estimateur naturel par la méthode des moments de la densité spectrale $f(\lambda)$ estimée selon les fréquences de Fourier ; i.e. $\lambda_k = 2k\pi/n$ pour $k = 1, \dots, n$ (*Brockwell et Davis* [6]) :

$$I_{n,X}(\lambda) := \frac{1}{2\pi n} \left| \sum_{t=1}^n e^{-i\lambda t} X_t \right|^2, \quad \lambda \in [0, \pi].$$

En plus, notons par $J_{n,X}(\cdot)$ la fonction de distribution spectrale empirique qui l'on appelle *périodogramme intégré* (i.e. un estimateur consistant de la distribution spectrale) telle que :

$$J_{n,X}(\lambda) := \int_0^\lambda I_{n,X}(y)dy, \quad \lambda \in [0, \pi], \quad (3.2.2)$$

c'est un estimateur consistant de la distribution spectrale

$$F_X(\lambda) := \int_0^\lambda f_X(x)dx, \quad \lambda \in [0, \pi],$$

a condition que la densité spectrale $f_X(\cdot)$ soit bien définie. Il est clair que cette distribution dans le cas d'un modèle GARCH(1. 1) vaut

$$F_X(\lambda) = \frac{\lambda}{2\pi} \gamma_X(0), \quad \lambda \in [0, \pi], \quad (3.2.3)$$

car pour tout $k \geq 0$,

$$\gamma_X(k) = cov(X_t, X_{t+k}) = \begin{cases} \mathbb{E}[X_t X_{t+k}] = 0, & \text{si } k \neq 0, \\ \gamma_X(0), & \text{si } k = 0, \end{cases}$$

avec

$$F_X(\lambda) := \frac{1}{2\pi} \int_0^\lambda \sum_{k=1}^n \gamma_X(k) e^{-ixk} dx = \frac{\lambda}{2\pi} \gamma_X(0).$$

Maintenant, pour avoir un processus centré tel que le modèle GARCH, on considère une variable aléatoire centrée X ($X = X_t - \bar{X}$), et on définit respectivement pour tout $h \in \mathbb{Z}$, la covariance entre X_0 (à la date $t = 0$) et X_h (à la date $t = h$), ainsi que la variance de la variable aléatoire $X_0 X_h$ par :

$$\gamma_X(h) = cov(X_0, X_h) \quad \text{et} \quad v_X(h) := var(X_0 X_h) = \mathbb{E} [(X_0 X_h)^2].$$

Prenons de plus, pour tout $x \in [0, 1]$

$$\gamma_{n, [nx], X}(h) = \frac{1}{n} \sum_{t=1}^{[nx]-h} X_t X_{t-h}, \quad h = 0, 1, \dots, [nx],$$

$[x]$ désigne la partie entière de x , de sorte que : $\gamma_{n,n,X}(h) := \gamma_{n,X}(h)$.

Pour cela, **(3.2.2)** est équivalente a

$$\begin{aligned}
J_{n,X}(x, \lambda) &= \frac{1}{2\pi} \int_0^\lambda \sum_{h=-\lfloor nx \rfloor + 1}^{\lfloor nx \rfloor - 1} \gamma_{n, \lfloor nx \rfloor, X}(h) e^{-iyh} dy, \quad \lambda \in [0, \pi] \\
&= \frac{1}{2\pi} \int_0^\lambda \left\{ \gamma_{n, \lfloor nx \rfloor, X}(0) + \sum_{h=-\lfloor nx \rfloor + 1}^{-1} \gamma_{n, \lfloor nx \rfloor, X}(h) e^{-iyh} \right. \\
&\quad \left. + \sum_{h=1}^{\lfloor nx \rfloor - 1} \gamma_{n, \lfloor nx \rfloor, X}(h) e^{-iyh} \right\} dy \\
&= \frac{1}{2\pi} \int_0^\lambda \left\{ \gamma_{n, \lfloor nx \rfloor, X}(0) + \sum_{h=1}^{\lfloor nx \rfloor - 1} \gamma_{n, \lfloor nx \rfloor, X}(h) (e^{-iyh} + e^{+iyh}) \right\} dy \\
&= \frac{1}{2\pi} \int_0^\lambda \left\{ \gamma_{n, \lfloor nx \rfloor, X}(0) + 2 \sum_{h=1}^{\lfloor nx \rfloor - 1} \gamma_{n, \lfloor nx \rfloor, X}(h) \cos(yh) \right\} dy,
\end{aligned}$$

donc,

$$J_{n,X}(x, \lambda) = \frac{1}{2\pi} \left(\lambda \gamma_{n, \lfloor nx \rfloor, X}(0) + 2 \sum_{h=1}^{\lfloor nx \rfloor - 1} \gamma_{n, \lfloor nx \rfloor, X}(h) \frac{\sin(\lambda h)}{h} \right). \quad (3.2.4)$$

Maintenant, on suppose que les moments d'ordre supérieur à quatre du processus X sont finis. Cela assure à son tour l'existence du $v_X(h)$, pour un modèle GARCH(1, 1) (voir la condition **(3.1.7)**), i.e.

$$\mathbb{E} [|X_t|^{4+\epsilon}] < \infty, \quad \text{pour tout } \epsilon > 0, \quad (3.2.5)$$

ce qui permet d'enoncer le resultat suivant (*T. Mikosch et C. Stărică (2002)*), dont le principe est qu'une modification du processus $J_{n,X}(x, \lambda)$ satisfait le théorème central limite fonctionnel :

Théorème 3.2.3 : (*T. Mikosch et C. Stărică (2002)*) [40]

Considerons le modèle stationnaire GARCH(p, q), s'il vérifie la condition (3.2.5). Alors, quand $n \rightarrow \infty$:

$$\sqrt{n} \left(\sum_{h=1}^{\lfloor nx \rfloor - 1} \frac{\gamma_{n, \lfloor nx \rfloor, X}(h) \sin(\lambda h)}{v_X^{1/2}(h) h} \right)_{\lambda, x} \xrightarrow{\mathcal{D}} \left(\sum_{h=1}^{+\infty} W_h(x) \frac{\sin(\lambda h)}{h} \right)_{\lambda, x},$$

dans $\mathbb{D}([0, 1] \times [0, \pi])$ l'espace de Skorohod (i.e. l'espace des fonctions continues à droite avec une limite à gauche : cadlag), avec $(W_h(\cdot))_{h \geq 1}$ est une suite indépendante et identiquement distribuée de mouvements browniens standards.

Remarques 3.2.1 :

1. Pour $x = 1$, la convergence en distribution dans le théorème précédent peut prendre la forme suivante : quand $n \rightarrow \infty$

$$\sqrt{n} \sum_{h=1}^{n-1} \frac{\gamma_{n,X}(h) \sin(\lambda h)}{v_X^{1/2}(h) h} \xrightarrow{\mathcal{D}} B(\cdot) := \sum_{h=1}^{+\infty} W_h(1) \frac{\sin(\cdot h)}{h},$$

dans l'espace des fonctions continues dans $[0, \pi]$ (i.e. $\mathbb{C}[0, \pi]$), la série à droite est appelée la représentation de *Paley-Wiener* d'un pont brownien dans $[0, \pi]$.

2. L'apparition de la quantité $v_X(h)$ dans le théorème est motivée par le fait que pour chaque $m \geq 1$, quand $n \rightarrow \infty$

$$\sqrt{n} (\gamma_{n, \lfloor nx \rfloor, X}(h), h = 1, \dots, m)_x \xrightarrow{\mathcal{D}} \left(v_X^{1/2}(h) W_h(x), h = 1, \dots, m \right)_x, \quad (3.2.6)$$

dans $\mathbb{D}([0, 1], \mathbb{R}^m)$, (Lemme 2.1. [40]). Il est donc naturel de diviser l'autocovariance empirique par $v_X^{1/2}(h)$, ce qui permet d'avoir une distribution asymptotique qui ne dépend pas du processus sous hypothèse.

Avec ces remarques et sous conditions du théorème précédent, on a le corollaire suivant, qui est à la base de la construction du test d'adéquation.

Corollaire 3.2.1 :

sous condition du théorème précédent, on a quand $n \rightarrow \infty$

$$\sqrt{n} \sup_{\lambda \in [0, \pi]} \left| \sum_{h=1}^{n-1} \frac{\gamma_{n,X}(h) \sin(\lambda h)}{v_X^{1/2}(h) h} \right| \xrightarrow{\mathcal{D}} \sup_{\lambda \in [0, \pi]} |B(\lambda)|,$$

$$n \int_0^\pi \left(\sum_{h=1}^{n-1} \frac{\gamma_{n,X}(h) \sin(\lambda h)}{v_X^{1/2}(h) h} \right)^2 d\lambda \xrightarrow{\mathcal{D}} \int_0^\pi B^2(\lambda) d\lambda.$$

Les résultats précédents sont valables pour les modèles $GARCH(p, q)$, mais dans ce qui suit nous ne étudierons que le cas particulier d'un modèle $GARCH(1, 1)$. D'après **(3.1.5)**, la quantité $v_X(h)$ dépend à la fois des paramètres du modèle (i.e. α_0, α et β) et du moment d'ordre quatre du bruit (i.e. $\mathbb{E}[\xi^4]$);

$$v_X(h) = \text{var}^2(X) \left[\frac{\alpha(1 - \alpha\beta^2)\mathbb{E}[\xi^4] - 1}{1 - ((\alpha + \beta)^2 + \alpha^2(\mathbb{E}[\xi^4] - 1))} (\alpha + \beta)^{h-1} + 1 \right], \quad h \geq 2.$$

Supposons maintenant qu'on possède un échantillon (X_1, X_2, \dots, X_n) extrait d'une série temporelle $(X_t)_t$, et que nous voulons tester l'hypothèse nulle suivante :

$$\mathbb{H}_0 : (X_1, X_2, \dots, X_n) \text{ est adéquate au modèle } GARCH(1, 1).$$

Pour l'application des résultats précédents à cette série temporelle, il est donc naturel de remplacer les quantités inconnues (i.e. α_0, α et β) du modèle par leurs estimateurs correspondants (i.e. $\hat{\alpha}_0, \hat{\alpha}$ et $\hat{\beta}$), d'un part et le moment d'ordre 4 du bruit (i.e. $\mathbb{E}[\xi^4]$) par le moment empirique $\mathbb{E}[\hat{\xi}^4]$ d'autre part, à savoir :

$$\mathbb{E}[\hat{\xi}^4] = \frac{1}{n} \sum_{t=1}^n \xi_t^4, \quad \xi_t = \frac{X_t}{\hat{\sigma}_t},$$

avec

$$\hat{\sigma}_t^2 := \hat{\alpha}_0 + \hat{\alpha}X_{t-1}^2 + \hat{\beta}\hat{\sigma}_{t-1}^2,$$

dans ce cas, la variance $v_X(h)$ va être remplacée par sa version empirique qu'on note par $\hat{v}_X(h)$.

Maintenant, pour montrer l'adéquation de la série temporelle $(X_t)_t$ au modèle $GARCH(1, 1)$ (sous l'hypothèse \mathbb{H}_0), ou ce qui revient à d'une certaine distance entre la distribution spectrale $F_X(\cdot)$ du modèle et le périodogramme intégré $J_{n,X}(\cdot)$ vers un fonctionnel d'un pont brownien. En effet, cela nous ramène à étudier le comportement de la statistique :

$$\sqrt{n} \sup_{\lambda \in [0, \pi]} |J_{n,X}(\lambda) - F_X(\lambda)|, \quad \lambda \in [0, \pi],$$

ou après pondération,

$$\sqrt{n} \sup_{\lambda \in [0, \pi]} \left| \frac{2\pi}{v_X^{1/2}(h)} (J_{n,X}(\lambda) - F_X(\lambda)) \right|, \quad \lambda \in [0, \pi].$$

D'après **(3.2.3)** et pour $x = 1$ dans **(3.2.4)**, cette quantité vaut

$$\sqrt{n} \sup_{\lambda \in [0, \pi]} \left| 2 \sum_{h=1}^{n-1} \frac{\gamma_{n,X}(h) \sin(\lambda h)}{v_X^{1/2}(h) h} + \frac{\lambda}{v_X^{1/2}(h)} (\gamma_{n,X}(0) - \gamma_X(0)) \right|.$$

La consistance de la variance empirique $\gamma_{n,X}(0)$ (i.e. $|\gamma_{n,X}(0) - \gamma_X(0)|$ est stochastiquement bornée), nous permet donc de construire la statistique :

$$\sqrt{n} \sup_{\lambda \in [0, \pi]} \left| \sum_{h=1}^{n-1} \frac{\gamma_{n,X}(h) \sin(\lambda h)}{v_X^{1/2}(h) h} \right|, \quad \lambda \in [0, \pi],$$

dont le comportement asymptotique -après remplacement de $v_X(h)$ par sa version empirique $\hat{v}_X(h)$ - est donnée par le théorème suivant :

Théorème 3.2.4 :

Supposons que les estimateurs $\hat{\alpha}_0$, $\hat{\alpha}$ et $\hat{\beta}$ des paramètres α_0 , α et β basent sur l'échantillon (X_1, X_2, \dots, X_n) sont indépendantes de $(\text{sign}(X_t))$, et qu'ils sont consistents au sens où :

$$\hat{\alpha}_0 \xrightarrow{\mathbb{P}} \alpha_0, \quad \hat{\alpha} \xrightarrow{\mathbb{P}} \alpha \text{ et } \hat{\beta} \xrightarrow{\mathbb{P}} \beta. \quad (3.2.7)$$

Alors le théorème et le corollaire précédents restent valable pour un modèle GARCH(1, 1), même si $v_X(\cdot)$ est remplacée par sa version empirique $\hat{v}_X(\cdot)$.

En particulier, quand $n \rightarrow \infty$

$$\sqrt{n} \sup_{\lambda \in [0, \pi]} \left| \sum_{h=1}^{n-1} \frac{\gamma_{n,X}(h) \sin(\lambda h)}{\hat{v}_X^{1/2}(h) h} \right| \xrightarrow{\mathcal{D}} \sup_{\lambda \in [0, \pi]} |B(\lambda)|,$$

$$n \int_0^\pi \left(\sum_{h=1}^{n-1} \frac{\gamma_{n,X}(h) \sin(\lambda h)}{\hat{v}_X^{1/2}(h) h} \right)^2 d\lambda \xrightarrow{\mathcal{D}} \int_0^\pi B^2(\lambda) d\lambda.$$

Démonstration :

La démonstration de ce théorème est basée sur le théorème précédent, il suffit donc de démontrer la consistance de processus $\hat{v}_X(\cdot)$, et d'appliquer le théorème **3.2.3**, cela comme suit :

Par construction, les estimateurs du modèle obtenus par la méthode de maximum du vraisemblance, sont définis à partir des processus X_t^2 et σ_t^2 , donc ils sont indépendants de $(\text{sign}(X_t))$, d'un part.

D'autre part, la consistance de $\hat{\alpha}_0$, $\hat{\alpha}$ et $\hat{\beta}$ implique la consistance de $\mathbb{E}[\hat{\xi}^4]$, a savoir :

$$\begin{aligned}
\mathbb{E}[\hat{\xi}^4] - \frac{1}{n} \sum_{t=1}^n \xi_t^4 &= \frac{1}{n} \sum_{t=1}^n \frac{X_t^4}{\hat{\sigma}_t^4} - \frac{1}{n} \sum_{t=1}^n \frac{X_t^4}{\sigma_t^4} \\
&= \frac{1}{n} \sum_{t=1}^n X_t^4 \left(\frac{\sigma_t^4 - \hat{\sigma}_t^4}{\hat{\sigma}_t^4 \sigma_t^4} \right) \\
&= \frac{1}{n} \sum_{t=1}^n X_t^4 \frac{(\sigma_t^2 - \hat{\sigma}_t^2)(\sigma_t^2 + \hat{\sigma}_t^2)}{\hat{\sigma}_t^4 \sigma_t^4} \\
&= \frac{1}{n} \sum_{t=1}^n \frac{X_t^4}{\hat{\sigma}_t^4 \sigma_t^4} [(\alpha_0 - \hat{\alpha}_0) + (\alpha - \hat{\alpha})X_{t-1}^2 + (\beta - \hat{\beta})\sigma_{t-1}^2 + \hat{\beta}(\sigma_t^2 - \hat{\sigma}_t^2)] \\
&= \frac{(\alpha_0 - \hat{\alpha}_0)}{n} \sum_{t=1}^n X_t^4 (1 + \hat{\beta} + \hat{\beta}^2 + \dots + \hat{\beta}^t) \frac{\sigma_t^2 + \hat{\sigma}_t^2}{\hat{\sigma}_t^4 \sigma_t^4} \\
&\quad + \frac{(\alpha - \hat{\alpha})}{n} \sum_{t=1}^n X_t^4 (X_{t-1}^2 + \hat{\beta}X_{t-2}^2 + \dots + \hat{\beta}^t X_0^2) \frac{\sigma_t^2 + \hat{\sigma}_t^2}{\hat{\sigma}_t^4 \sigma_t^4} \\
&\quad + \frac{(\beta - \hat{\beta})}{n} \sum_{t=1}^n X_t^4 (\sigma_{t-1}^2 + \hat{\beta}\sigma_{t-2}^2 + \dots + \hat{\beta}^t X_0^2) \frac{\sigma_t^2 + \hat{\sigma}_t^2}{\hat{\sigma}_t^4 \sigma_t^4} \\
&\quad + \frac{(\sigma_0^2 - \hat{\sigma}_0^2)}{n} \sum_{t=1}^n X_t^4 \hat{\beta}^t \frac{\sigma_t^2 + \hat{\sigma}_t^2}{\hat{\sigma}_t^4 \sigma_t^4} \\
&= (\alpha_0 - \hat{\alpha}_0)I_n + (\alpha - \hat{\alpha})II_n + (\beta - \hat{\beta})III_n + (\sigma_0^2 - \hat{\sigma}_0^2)IV_n.
\end{aligned}$$

Premièrement, par consistance des paramètres $\hat{\alpha}_0$, $\hat{\alpha}$, $\hat{\beta}$ et comme on a $X_t := \sigma_t \xi_t$, donc :

$$\begin{aligned}
I_n &\leq \frac{1}{n(1 - \hat{\beta})} \sum_{t=1}^n \xi_t^4 \frac{\sigma_t^2 + \hat{\sigma}_t^2}{\hat{\sigma}_t^4} \\
&\leq \frac{1}{n(1 - \hat{\beta})} \sum_{t=1}^n \xi_t^4 (\sigma_t^2 \hat{\alpha}_0^{-2} + \hat{\alpha}_0^{-1}) \\
&\xrightarrow{P.S} \frac{1}{(1 - \hat{\beta})} \mathbb{E}[\xi^4] (\hat{\alpha}_0^{-1} + \hat{\alpha}_0^{-2} \mathbb{E}[\sigma_t^2]).
\end{aligned}$$

Par le même argument, pour tout $\delta > 0$ et $\epsilon > 0$ tel que $\beta + \epsilon < 1$,

$$\begin{aligned}
& P(II_n > \delta) \\
\leq & P\left(\frac{1}{n} \sum_{t=1}^n X_t^4 (X_{t-1}^2 + (\beta + \epsilon)X_{t-2}^2 + \dots + (\beta + \epsilon)^t X_0^2) \frac{\sigma_t^2 + \hat{\sigma}_t^2}{\hat{\sigma}_t^4 \sigma_t^4} > \delta\right) \\
& + P(\hat{\beta} - \beta > \epsilon) \\
\leq & P\left(\frac{1}{n} \sum_{t=1}^n \xi_t^4 (X_{t-1}^2 + (\beta + \epsilon)X_{t-2}^2 + \dots + (\beta + \epsilon)^t X_0^2) (\sigma_t^2 \hat{\alpha}_0^{-2} + \hat{\alpha}_0^{-1}) > \delta\right)
\end{aligned}$$

Appliquant l'inégalité de *Cauchy-Schwarz*², aux deux termes suivants :

$$\begin{aligned}
& \frac{1}{n} \sum_{t=1}^n \xi_t^4 (X_{t-1}^2 + (\beta + \epsilon)X_{t-2}^2 + \dots + (\beta + \epsilon)^t X_0^2) \text{ et} \\
& \frac{1}{n} \sum_{t=1}^n \xi_t^4 (X_{t-1}^2 + (\beta + \epsilon)X_{t-2}^2 + \dots + (\beta + \epsilon)^t X_0^2) \sigma_t^2,
\end{aligned}$$

leurs premiers moments sont uniformément bornés pour n , donc II_n est stochastiquement borné, idem pour III_n . Finalement,

$$P\left(\frac{1}{n} IV_n > \delta\right) \leq P\left(\sum_{t=1}^n \xi_t^4 (\beta + \epsilon)^t (\sigma_t^2 \hat{\alpha}_0^{-2} + \hat{\alpha}_0^{-1}) > \delta\right) + P(\hat{\beta} - \beta > \epsilon),$$

les séries $\sum_{t=1}^n \xi_t^4 (\beta + \epsilon)^t$ et $\sum_{t=1}^n \xi_t^4 (\beta + \epsilon)^t \sigma_t^2$ ont des moments bornés, et par consistance de $\hat{\beta}$ (i.e. $P(\hat{\beta} - \beta > \epsilon) = o(1)$). Ce qui implique que IV_n est stochastiquement bornée.

On conclut par la loi des grands nombres, la consistance de $\hat{\alpha}_0$, $\hat{\alpha}$, $\hat{\beta}$ et que les termes I_n , II_n , III_n et IV_n sont stochastiquement bornées, le processus $\hat{v}_X(\cdot)$ est donc consistant. ■

²Inégalité de *Cauchy-Schwarz* : Soient $X, Y \in \mathbf{L}^2$, alors

$$XY \in \mathbf{L}^1 \text{ et } \mathbb{E}|XY| \leq \mathbb{E}|X^2|^{1/2} \mathbb{E}|Y^2|^{1/2}.$$

Chapitre 4

Problème d'adéquation pour la classe des distributions de type Pareto

Au vu des résultats établis par *Mikosch, T. et Stărică, C.* (1998), les suites des maximas d'un modèle GARCH(1, 1), celles des variances conditionnelles et des valeurs absolues du modèle, appartenant tous au domaine d'attraction de *Fréchet* (Type II). De plus, d'après *Kesten* (1973), la distribution marginale d'une solution de l'équation stochastique aux différences de la forme (2.4.1) est à queue lourde (section 2.4).

Il est donc nécessaire de ne pas se limiter à la construction d'un test d'adéquation pour le modèle, il est temps de penser de faire prolonger ce problème d'adéquation pour la classe des distributions de type Pareto, coïncidant avec celles appartenant au domaine d'attraction de *Fréchet*, et permettant de modéliser les queues de distribution. Les données extrêmes sont les seules utilisées dans l'estimation de ces paramètres, ce qui assure le meilleur ajustement du modèle à la queue.

Ce problème est récemment étudié par *Beirlant, de Wet et Goegebeur* (2004) dans le cadre empirique des suites de variables aléatoires i.i.d. D'autre part, *Necir, A. et Boukhetala, K.* (2004)[46] ont proposé une nouvelle distance pondérée comme modification de la distance \mathbb{L}^2 -Wasserstein, et qui sert à vérifier l'ajustement dans la classe des distributions de type Pareto.

Cela est l'objet du présent chapitre, on envisage au futur, l'application de ces résultats en particulier au domaine d'assurance non vie, en présence des hypothèses sur la dépendance, la stationnarité ou l'hétéroscédasticité...

4.1 Classe des distributions de type Pareto

Etant donné une suite de variables aléatoires (i.i.d) $(X_n)_n$, de même fonction de distribution F , et soit $(X_{1,n}, \dots, X_{n,n})$, avec $X_{1,n} \leq \dots \leq X_{n,n}$ la statistique d'ordre associée. Nous voulons vérifier si un modèle statistique permet d'obtenir une bonne approximation de la distribution des extrimas triés de ce modèle. Typiquement, la méthode du quantile quantile -plot (QQ-plot) est une méthode graphique, empirique utilisée pour tester l'adéquation d'une distribution empirique F_n à une loi F donnée, se basant simplement sur le fait que les quantiles empiriques $Q(X_{i,n})$ suivant la même loi que n variables aléatoires uniformément ordonnées $(U_{i,n})$, d'espérances respectives $\frac{i}{n+1}$, de sorte que les points $(X_{i,n}, Q(\frac{i}{n+1}))$ pour i grand doivent être quasiment alignées sur une droite.

Le comportement des valeurs extrêmes est gouverner par le paramètre γ (i.e. indice des valeurs extrême) qui est celui de la forme de la distribution généralisée des valeurs extrêmes G_γ :

$$G_\gamma(x) = \begin{cases} \exp\left(- (1 + \gamma x)^{-1/\gamma}\right), & \text{pour } \gamma \neq 0, 1 + \gamma x > 0, \\ \exp(-e^{-x}), & \text{pour } \gamma = 0, x \in \mathbb{R}. \end{cases} \quad (4.1.1)$$

S'il existe des constantes de normalisation $(a_n > 0)_n$ et $(b_n)_n \in \mathbb{R}$, telles que :

$$\lim_{n \rightarrow \infty} P\left(\frac{X_{n,n} - b_n}{a_n}\right) = \lim_{n \rightarrow \infty} F^n(b_n + a_n x) = G_\gamma, \quad (4.1.2)$$

pour tout point de continuité de G_γ , alors G_γ est de type (4.1.1) (Fisher et Tippett, 1982, Gnedenko, 1943). De plus, si (4.1.2) est satisfait pour la fonction de distribution cumulative F , alors elle appartient au domaine d'attraction de G_γ ($F \in \mathcal{D}(G_\gamma)$).

Définition 4.1.1 : " *Distribution de Pareto généralisée* "

La distribution de Pareto généralisée ou simplement distribution de Pareto notée F_{GPD} est employée pour modéliser la queue d'une distribution présentant des valeurs extrêmes. Elle est donnée par :

$$F_{GPD}(x) = 1 - \left(1 + \frac{x\gamma}{\rho}\right)^{-1/\gamma}, \quad \rho > 0, \quad 1 + \frac{x\gamma}{\rho} > 0.$$

La fonction des quantiles associées à F_{GPD} notée Q_{GPD} est définie par :

$$Q_{GPD}(t) = \begin{cases} \rho \frac{(1-t)^{-\gamma} - 1}{\gamma}, & \text{si } \gamma \neq 0, \rho > 0, \\ -\rho \log(1-t), & \text{si } \gamma = 0. \end{cases}$$

Dans le cas $\gamma > 0$, $\mathcal{D}(G_\gamma)$ coïncide avec la classe des distributions de type Pareto avec l'indice $1/\gamma$. C'est une classe de distribution à queue lourde, qui possèdent un point limite supérieur $x_F^+ = +\infty$, i.e.

$$1 - F(x) = x^{-1/\gamma} L_F(x), \quad x > 0,$$

avec $L_F(\cdot)$ désigne une fonction à variation lente à l'infini :

$$\lim_{x \rightarrow \infty} \frac{L_F(tx)}{L_F(x)} = 1, \quad \text{pour tout } t > 0,$$

le paramètre γ mesure dans ce cas l'alourdissement de la queue de la distribution F , et comme exemple de modèle de type Pareto on a notamment : la distribution de Pareto généralisée, la distribution de Burr¹, et entre autres particulièrement la distribution marginale du modèle stationnaire GARCH,...

La plupart des récentes recherches se concentrent sur la distribution à queue lourde, dont l'index des valeurs extrêmes γ est strictement positif. D'après *Csörgő et Viharos (1998)*, quand $\gamma > 0$, la distribution comulative F du modèle X vérifiant (4.1.2) est de type Pareto i.e.

$$\lim_{x \rightarrow \infty} \frac{1 - F(tx)}{1 - F(x)} = t^{-1/\gamma}, \quad \text{pour tout } t > 0,$$

et comme estimateur de l'index $\gamma > 0$, le plus célèbre est celui de Hill (*Hill, 1975*) donné par :

$$\gamma_{k_n}^H = \frac{1}{k_n} \sum_{i=1}^{k_n} \log X_{n-i+1,n} - \log X_{n-k_n,n},$$

¹La distribution de Burr (*Burr, 1942*) :

$$F(x) = 1 - \left(\frac{\eta}{\eta + x^\tau} \right)^\lambda, \quad x > 0, \lambda, \eta, \tau > 0.$$

Il est clair que cette distribution est de type Pareto avec $\gamma = 1/(\lambda\tau)$ et $\rho = -1/\lambda$.

avec k_n est une suite d'entier positif ($1 \leq k_n < n$), satisfait souvent la condition :

$$k_n \rightarrow \infty \quad \text{et} \quad \frac{k_n}{n} \rightarrow 0, \quad \text{quand } n \text{ tend vers l'infini.} \quad (4.1.3)$$

L'adéquation d'un modèle statistique à l'échantillon X_1, \dots, X_n est visuellement confirmé par l'interprétation du QQ-plot, et du fait que le log-transformation de la variable X suivant une distribution de Pareto est exponentiellement distribué (dans ce cas $L_F(x) = 1$). Pour cette raison *Beirlant, de Wet et Goegebeur* (2004), ont interprété le Pareto quantile plot (PQ-plot) comme une méthode empirique qui sert à vérifier si les données sont générées par une distribution de la classe de type Pareto ou non.

Définition 4.1.2 : " Pareto quantile plot "

De même que la méthode de QQ-plot, le Pareto quantile plot (PQ-plot) consiste à dessiner les points de coordonnées $(\log \frac{n+1}{i}, \log X_{n-i+1,n})$, $i = 1, 2, \dots, n$. Pour avoir une bonne adéquation, le Pareto quantile plot doit être alors approximativement linéaire, et la pente ajustée aux points de PQ-plot coïncide avec l'estimateur de l'index γ .

En utilisant les liens entre la distribution de Paréto et la distribution exponentielle, *Beirlant et al.* (2004) ont modifié la statistique de *Jackson* pour tester l'exponentialité appliqué à la $i^{\text{ème}}$ statistique d'ordre dans le PQ-plot, ce qui permet de faire des comparaisons entre des distribution appartenant à la classe de type Pareto et autres qui n'appartenant pas.

Soient X_1, \dots, X_n des variables aléatoires i.i.d, d'une distribution exponentielle notée $Exp(\lambda)$ ², la statistique de Jackson est donnée par :

$$J_n = \frac{\sum_{j=1}^n t_{j,n} X_{j,n}}{\sum_{j=1}^n X_j},$$

où $t_{j,n} = \lambda E[X_{j,n}] = \sum_{k=1}^j (n - k + 1)^{-1}$.

²La distribution exponentielle $Exp(\lambda)$ est donnée par :

$$F(x) = 1 - \exp\{-\lambda x\}, \quad x \in \mathbb{R}.$$

On pose $V_j = (n - j + 1)(X_{j,n} - X_{j-1,n})$, $j = 1, \dots, n$, la statistique J_n vaut dans ce cas :

$$J_n = \frac{\sum_{j=1}^n C_{j,n} V_j}{\sum_{j=1}^n V_j},$$

où $C_{1,n} = 1$ et $C_{j,n} = 1 + t_{j-1,n}$, $j = 2, \dots, n$, et la distribution limite de cette statistique est donnée par le théorème suivant :

Théorème 4.1.1 : (Jackson, 1967)

Supposons que X_1, \dots, X_n sont des variables aléatoires i.i.d, d'une distribution exponentielle $Exp(\lambda)$, alors quand $n \rightarrow \infty$,

$$\sqrt{n}(J_n - 2) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Considérons maintenant n variables aléatoires i.i.d X_1, \dots, X_n d'une distribution de Pareto stricte d'index α notée $Pa(\alpha)$ ³, alors $Y_{j,k} = X_{n-k+j,n}/X_{n-k,n}$, $j = 1, \dots, k$ sont des $j^{\text{ème}}$ statistiques d'ordre Pareto distribuées. Par conséquence, $Y_{j,k}^* = \log Y_{j,k}$, $j = 1, \dots, k$, suivent une distribution exponentielle de paramètre $\lambda = 1/\gamma$. Dans ce cas, la statistique de Jackson appliquée à $Y_{j,k}^*$, $j = 1, \dots, k$, vaut :

$$J_k^* = \frac{\frac{1}{k} \sum_{j=1}^k C_{k-j+1,k} Z_j}{\gamma_{k,n}^H},$$

avec $Z_j = j(\log X_{n-j+1,n} - \log X_{n-j,n})$, $j = 1, \dots, k$ et $\gamma_{k,n}^H = \frac{1}{k} \sum_{j=1}^k Z_j$.

La distribution limite de cette statistique pour la suite intermédiaire $k = k_n$ satisfait la condition (4.1.3), et sous condition sur la fonction à variation lente de la fonction des quantiles de la queue $U := Q(1 - \frac{1}{x})$, comme suit :

(R_L) Il existe une constante $\varrho < 0$ et une fonction $b(\cdot)$ satisfait $b(x) \rightarrow 0$, quand $x \rightarrow \infty$, telle que pour tout $\lambda \geq 1$, quand $x \rightarrow \infty$,

$$\frac{L_U(\lambda x)}{L_U(x)} - 1 \sim b(x) \frac{\lambda^\varrho - 1}{\varrho}.$$

³La distribution de Pareto stricte $Pa(\alpha)$ est donnée par :

$$F(x) = 1 - x^{-\alpha}, \quad x > 1, \quad \alpha > 0.$$

Théorème 4.1.2 (*Beirlant et al., 2004*)

Supposons que X_1, \dots, X_n sont des variables aléatoires i.i.d, d'une distribution F , $F \in D(G_\gamma)$ pour tout $\gamma > 0$ et L_U vérifie la condition (R_L) , alors quand n tend vers l'infini, $k = k_n$ satisfait la condition (4.1.3), et $\sqrt{k}b(\frac{n}{k}) \rightarrow c$,

$$\sqrt{k}(J_k^* - 2) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{c\varrho}{\gamma(1-\varrho)^2}, 1\right),$$

$\mathcal{N}(\mu, 1)$, $\xrightarrow{\mathcal{D}}$ désignent respectivement la loi normale réduite de moyenne μ et la convergence en distribution.

4.2 Test d'adéquation des indices des queues via la distance \mathbb{L}^2 -Wasserstein pondérée

On désigne par \mathfrak{S} l'espace des fonctions de distribution G , telle que :

$$\int_0^1 (G^{-1}(s))^2 ds < \infty. \quad (4.2.1)$$

Définition 4.2.1 : "*Distance \mathbb{L}^2 - Wasserstein*"

On définit la distance \mathbb{L}^2 - Wasserstein entre deux fonctions de distribution G_1 et $G_2 \in \mathbb{L}^2(0, 1)$, notée $d(G_1, G_2)$ par

$$d(G_1, G_2) := \left(\int_0^1 (G_1^{-1}(s) - G_2^{-1}(s))^2 ds \right)^{1/2},$$

avec $G^{-1}(t) = \inf \{x : G(x) \geq t\}$, $0 < t < 1$, est l'inverse généralisé de G ou la fonction des quantiles.

Dans la classe des distributions de type Pareto (Pareto, log-normale, log gamma,...) la condition (4.2.1) n'est pas vérifiée. Pour résoudre ce problème, *Necir, A. et Boukhetala, K.* (2004)[46] ont proposé une autre version de la distance d précédente; on pondérant la queue de distribution, du sorte que :

$$d_\omega^2(G_1, G_2) := \int_0^1 \omega(s) (G_1^{-1}(1-s) - G_2^{-1}(1-s))^2 ds,$$

pour tous fonctions G_1 et $G_2 \in \mathfrak{S}_\omega$, où \mathfrak{S}_ω est l'ensemble des distributions G tels que :

$$\int_0^1 \omega(s) (G^{-1}(s))^2 ds < \infty,$$

avec $\omega(s)$ est une fonction de poids dans $]0, 1[$, telle que : $\int_0^1 \omega(s) ds = 1$.

Considérons la suite X_1, \dots, X_n des variables aléatoires indépendantes et identiquement distribuées, d'une fonction de distribution $F(x) = P\{X \leq x\}$, $x \in \mathbb{R}$. On note par $\bar{F}(x) = 1 - F(x)$ sa fonction de survie et par $X_{1,n}, \dots, X_{n,n}$ la statistique d'ordre correspondante, et on désigne par F_n la fonction de distribution empirique.

Supposons maintenant que la fonction de distribution F est à queue lourde, autrement dit, la queue de F est de type Pareto ; il existe une constante $\gamma > 0$, telle que

$$\lim_{x \rightarrow \infty} 1 - F(x) = x^{-1/\gamma} L_F(x), \quad \text{pour tout } x > 0, \quad (\mathbf{4.2.2})$$

L_F est une fonction à variation lente, et le paramètre γ est appelé l'index des queues de la distribution F .

Pour tester l'ajustement entre la distribution F sous la condition **(4.2.2)** et la distribution empirique F_n de l'échantillon, *Necir, A. et Boukhetala, K.* (2004)[46] ont proposé la distance $d_\omega^2(F_n, F)$, pour une fonction de poids

$$\omega_{\delta,k}(s) := \begin{cases} (1 + 2\delta)(k/n)^{-(2\delta+1)} s^{2\delta}, & 0 \leq s \leq k/n, \\ 0, & \text{ailleurs,} \end{cases}$$

avec δ est une constante souhaitable strictement positive ($\delta > 0$), et $k := k_n$,

$n \geq 1$ est une suite d'entier positif satisfait la condition

$$1 \leq k_n < n, \quad k_n \rightarrow \infty \quad \text{et} \quad \frac{k_n}{n} \rightarrow 0, \quad \text{quand } n \rightarrow \infty. \quad (\mathbf{4.2.3})$$

Il est clair que $\int_0^1 \omega_{\delta,k}(s) ds = 1$, pour tout $\delta > 0$. La distance correspondante est donc :

$$\begin{aligned} d_{\delta,k}^2(F_n, F) &:= \int_0^1 \omega_{\delta,k}(s) \{F_n^{-1}(1-s) - F^{-1}(1-s)\}^2 ds \\ &= \int_0^{k/n} \frac{(1+2\delta)s^{2\delta}}{(k/n)^{(2\delta+1)}} \{F_n^{-1}(1-s) - F^{-1}(1-s)\}^2 ds. \end{aligned}$$

Effectuons le changement de variable ($s = kt/n$), $d_{\delta,k}^2(F_n, F)$ se réécrit pour tout $\delta > 0$, comme suit

$$d_{\delta,k}^2(F_n, F) = (1 + 2\delta) \int_0^1 s^{2\delta} \{F_n^{-1}(1 - ks/n) - F^{-1}(1 - ks/n)\}^2 ds.$$

Finalement, on suppose que la fonction de la queue $\bar{F} = 1 - F$ est à variation régulière au second ordre dont le premier indice est $-1/\gamma$ et le second étant $\rho \leq 0$. Par conséquent, il existe une fonction $A^*(t) \rightarrow 0$ quand $t \rightarrow 0$ et qui a un signe constant, de sorte que la modification suivante de (4.2.2) soit vérifiée :

$$\lim_{t \rightarrow 0} A^*(t)^{-1} \left\{ \frac{1 - F(tx)}{1 - F(t)} - x^{-1/\gamma} \right\} = x^{-1/\gamma} \frac{x^\rho - 1}{\rho}, \quad \text{pour tout } x > 0. \quad (4.2.4)$$

Si $\rho = 0$, on remplace $\frac{x^\rho - 1}{\rho}$ par $\log x$. Par simplicité, on pose

$$A(t) := A^* [F^{-1}(1 - 1/t)], \quad t > 1,$$

$F^{-1}(1 - 1/t)$ est la fonction des quantiles de la queue. Notons que $|A|$ est à variation régulière à l'infini d'indice $\rho\gamma$, avec $A(t) \rightarrow 0$, quand $n \rightarrow \infty$.⁴

Théorème 4.2.1 : (Necir, A. et Boukhetala, K. (2004) [46])

Supposons que (4.2.4) soit vérifiée avec $\gamma > 0$. Soit $k = k_n$ satisfait la condition (4.2.3), et $kA(k/n) \rightarrow 0$ quand n tend vers l'infini. Alors, pour tout $\delta > \gamma$ nous avons

$$\frac{k}{\gamma^2(2\delta + 1) [F^{-1}(1 - k/n)]^2} d_{\delta,k}^2(F_n, F) \stackrel{\mathcal{D}}{=} \int_0^1 s^{2(\delta - \gamma - 1)} W^2(s) ds + o_p(1),$$

quand $n \rightarrow \infty$, avec W le mouvement brownien standard et $\stackrel{\mathcal{D}}{=}$ désigne l'égalité en distribution.

⁴Pour une étude détaillée des propriétés de la fonction $A(\cdot)$ et la variation régulière au second ordre, citons : de Haan et Stadtmüller (1996).

L'une des fonctions spéciales ; solutions d'équations différentielles particulières est la fonction de Bessel. Dans ce paragraphe nous nous intéressons à quelques propriétés de cette fonction, en particulier le développement de Karhunen- Loeve pour un processus de Weiner pondéré via la fonction de Bessel.

Définition 4.2.2 : "*Fonction de Bessel*"

La fonction de Bessel notée $J_v(z)$ de première espèce d'ordre v , est définie par

$$J_v(z) = \left(\frac{z}{2}\right)^v \sum_{\ell=0}^{\infty} \frac{(-1)^\ell}{\ell! \Gamma(v + \ell + 1)} \left(\frac{z}{2}\right)^{2\ell},$$

c'est une solution particulière de l'équation différentielle suivante :

$$\frac{\partial^2 y}{\partial z^2} + \frac{1}{z} \frac{\partial y}{\partial z} + \left(1 + \frac{v^2}{z^2}\right) y = 0.$$

Considérons la fonction de Bessel $J_v(\cdot)$ d'indice v (citons : *Deheuvels et Martynov (2004)*, section 2.1 pour une étude détaillée des propriétés de la fonction $J_v(\cdot)$). Notons les zéros de $J_v(\cdot)$ (i.e. solutions de l'équation $J_v(z) = 0$) par $0 < z_{v,1} < z_{v,2} < \dots$

Le corollaire suivant a pour objectif de préciser les liens entre la fonction de Bessel et le processus de Weiner $W(\cdot)$ (resp. pont brownien $B(\cdot)$) :

Corollaire 4.2.1 : (*Deheuvels et Martynov (2004)*)

Pour tout $\beta > -1$, ou par équivalence pour tout $v = 1 / (2(1 + \beta)) > 0$,

$$\int_0^1 s^{2\beta} W^2(s) ds \stackrel{\mathcal{D}}{=} 4v^2 \sum_{\ell=1}^{\infty} \frac{\omega_\ell^2}{z_{v-1,\ell}^2},$$

$$\int_0^1 s^{2\beta} B^2(s) ds \stackrel{\mathcal{D}}{=} 4v^2 \sum_{\ell=1}^{\infty} \frac{\omega_\ell^2}{z_{v,\ell}^2},$$

avec $\{\omega_\ell : \ell \geq 1\}$ est une suite de variables aléatoires i.i.d. d'une distribution normale centré et réduite $\mathcal{N}(0, 1)$.

Preuve : Voir le Corollaire 1.1 de *Deheuvels et Martynov (2004)*.

Posons maintenant $v := 1/(2(\delta - \gamma))$, pour $\delta > \gamma > 0$. Dans ce cas on a $\beta = \delta - \gamma - 1 > -1$, la condition du corollaire précédent est donc vérifiée. La combinaison de celui-ci avec le théorème 4.2.1 permet d'énoncer le résultat suivant :

Théorème 4.2.2 : (*Necir, A. et Boukhetala, K. (2004) [46]*)

Supposons que (4.2.4) soit vérifiée avec $\gamma > 0$. Soit $k = k_n$ satisfait la condition (4.2.3), et $kA(k/n) \rightarrow 0$ quand n tend vers l'infini. Alors, pour tout $\delta > \gamma$ nous avons

$$\frac{(\delta - \gamma)^2 k}{\gamma^2 (2\delta + 1) [F^{-1}(1 - k/n)]^2} d_{\delta, k}^2(F_n, F) \stackrel{\mathcal{D}}{=} \sum_{\ell=1}^{\infty} \frac{\omega_{\ell}^2}{z_{v-1, \ell}^2} + o_p(1),$$

quand $n \rightarrow \infty$, avec $\{\omega_{\ell} : \ell \geq 1\}$ est une suite de variables aléatoires i.i.d. d'une distribution normale centré et réduite $\mathcal{N}(0, 1)$ et $\stackrel{\mathcal{D}}{=}$ désigne l'égalité en distribution.

Chapitre 5

Etude empirique d'une série temporelle financière

Du point de vue de l'investisseur, le rendement de l'investissement est beaucoup plus intéressant que le prix de celui-ci, car il insiste plus sur le gain relatif réalisable, et puisque c'est le rendement qui détermine les profits que peut tirer l'actionnaire de son projet. De plus, les séries temporelles des rendements représentent le mieux les changements des prix, ce qui permet de faire des comparaisons entre compagnies et titres boursiers et d'avoir une prévision améliorée... Ces séries qui ont beaucoup de propriétés ne possèdent pas, en revanche, celles des prix, telles que la stationnarité, la volatilité,...

Ce chapitre sera donc consacré à l'étude des diverses propriétés statistiques, aux caractéristiques et comportements d'une série temporelle des rendements de l'indice boursier *S&P500* (*Standard and Poor's 500*)¹.

Cet indice conçu par *Standard and Poor's*, s'appuie sur 500 (blue chips) américaines. Les actions étant choisies pour leur taille et leur liquidité et l'indice est pondéré par les capitalisations boirsières.

¹Fin 2001, le S&P 500 comprenait 421 sociétés du **Nyse**, 77 du **Nasdaq** et 2 de l'**Amex**. Il représentait ainsi une capitalisation boursière totale de plus de 10.000 milliards de Dollars fin novembre 2001.

5.1 Propriétés statistiques de la série

5.1.1 Séries des prix et des rendements journaliers

Les premières études sur la loi de probabilité des prix boursiers ont été basées sur la loi normale, ou loi de Gauss. L'hypothèse normale sous-entend donc que la loi gaussienne modélise le mieux les variations relatives des rendements d'indice boursier, prix des actions ou des taux de change... En vue de tester cette hypothèse de normalité, le skewness et la kurtosis ont été mis à contribution. Les résultats obtenus doivent en principe se rapprocher des hypothèses couramment émises dans la théorie financière. Les données de notre étude sont issues du cite internet :

"www.chart2.finance.dcn.yahoo.com".

La figure 5.1 représente les prix de fermeture quotidienne de l'indice boursier S&P 500, de la période allant du 03/01/1994 au 13/05/2004 (i.e. 10 ans), soit 2610 observations. Du fait que les marchés boursiers sont fermés pendant les fins de la semaine (i.e. Dimanche et Lindi) et les jours fériés, les jours non ouvrables ne sont pas donc pris en compte. Ainsi une année commerciale contient presque 250 jours ouvrables en moyenne.

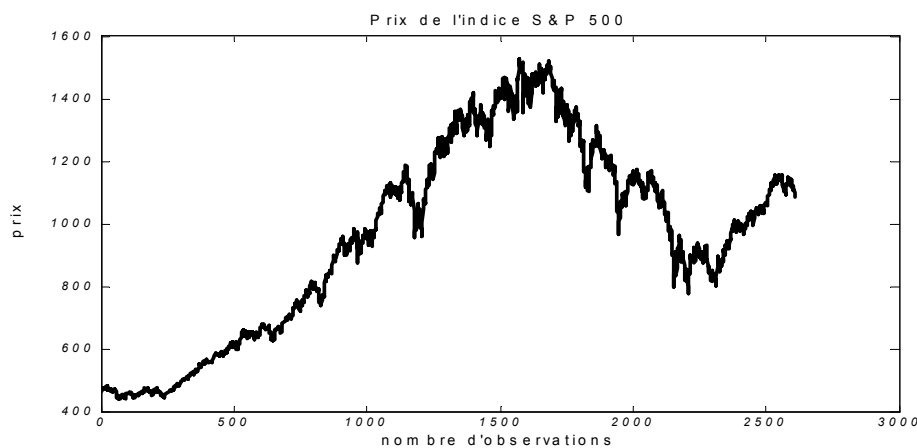


FIG. 5.1 – Série des prix de fermeture quotidienne de l'indice boursier S&P 500, de la période allant du 03/01/1994 au 13/05/2004.

Ces données (i.e. les prix de fermeture journalier) ont été transformées, ce qui nous permet d'étudier la série des rendements réalisables.

Il existe beaucoup de mesures de rendement d'un indice boursier, celle fréquemment utilisée est le rendement géométrique ou le log-rendement, qui consiste à calculer le logarithme du différentiel des valeurs en t et $t - 1$:

Définition 5.1.1 : " *Log-rendement* "

Soit P_t le prix d'un titre au temps t , on définit donc le rendement de ce titre à l'instant t par :

$$R_t := \log\left(\frac{P_t}{P_{t-1}}\right), \quad t = 1, 2, \dots$$

La figure 5.2 représente l'évolution de la série des rendements de l'indice S&P 500, de la période allant du 03/01/1994 au 13/05/2004, l'axe des abscisses présente le nombre d'observations (i.e. jours ouvrables) et l'axe des ordonnées présente le rendement de l'indice journalier. La première constatation que l'on peut faire est que la série est stationnaire, mais que l'on observe quand même des périodes de forte variation. La seconde est l'existence de quelques valeurs extrêmes.

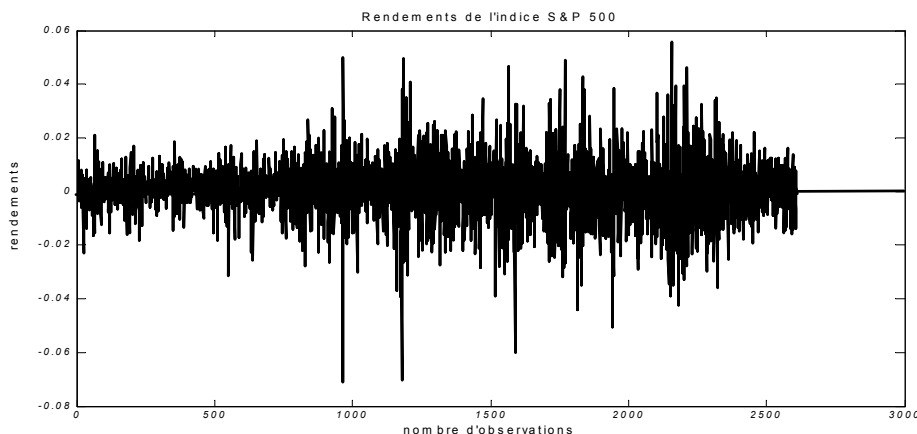


FIG. 5.2 – Série des rendements de l'indice boursier S&P500, de la période allant du 03/01/1994 au 13/05/2004.

5.1.2 Moments empiriques

L'une des propriétés importantes de la distribution d'une série du rendement est qu'elle est à queue lourde. Pour étudier cette propriété, on donne les versions empiriques de la moyenne, variance, coefficient d'aplatissement et coefficient d'asymétrie de cette série, comme suit :

Etant donné N observations d'une série temporelle $(X_t)_t : t = 1, 2, \dots, N$, l'estimateur naturel de la moyenne notée \bar{X} et de la variance de X_t notée S^2 sont données respectivement par :

$$\bar{X} := \frac{1}{N} \sum_{t=1}^N X_t, \quad S^2 := \frac{1}{N} \sum_{t=1}^N (X_t - \bar{X})^2.$$

Une série de rendement $(X_t)_t$ possède un coefficient d'aplatissement (*kurtosis*) plus grand que celui d'un bruit blanc gaussien. Ce coefficient peut être calculé empiriquement. On a ainsi :

$$\kappa := \frac{\mathbb{E} [(X - \mathbb{E}(X))^4]}{\mathbb{E} [(X - \mathbb{E}(X))^2]^2} \simeq \frac{\frac{1}{T} \sum_{t=1}^N (X_t - \bar{X})^4}{\left(\frac{1}{T} \sum_{t=1}^N (X_t - \bar{X})^2\right)^2}.$$

De même, le coefficient d'asymétrie (*skewness*), qui vaut :

$$s := \frac{\mathbb{E} [(X - \mathbb{E}(X))^3]}{\mathbb{E} [(X - \mathbb{E}(X))^2]^{3/2}} \simeq \frac{\frac{1}{T} \sum_{t=1}^N (X_t - \bar{X})^3}{\left(\frac{1}{T} \sum_{t=1}^N (X_t - \bar{X})^2\right)^{3/2}}.$$

Statistique	Valeur
Minimum	-0.0711
Maximum	0.0557
Moyenne	3.28 10^{-4}
Variance	1.28 10^{-4}
Kurtosis (κ)	6.2306 > 3
Skewness (s)	-0.1102 $\neq 0$

Tableau 5.1 : *Moments empiriques de la série des rendements.*

Dans le tableau 5.1 sont représentés les différentes caractéristiques statistiques de la série des rendements de l'indice S&P 500, sur la période couvrant les dix dernières années (1994-2004). Le premier constat dans ce tableau, est que la série des rendements de l'indice S&P 500 a une distribution leptokurtique : le coefficient d'aplatissement (kurtosis $\kappa = 6.2306$) est plus grand que 3, celle de la distribution normale et le coefficient d'asymétrie (skewness $s = -0.1102$) est différent de zéro.

L'analyse du skewness et de la kurtosis conduit aux conclusions usuelles dans les études des cours boursiers. Elles sont différentes de 0 et 3, ce qui signifie que la distribution n'est pas normale mais plutôt asymétrique avec des queues épaisses caractérisant une distribution leptokurtique. Ce qui nous conduit au rejet de l'hypothèse de normalité.

De plus, on peut résumer ces propriétés dans la figure 5.3 qui représente l'histogramme de la série des rendements, et cela en partageant les rendements journaliers de l'indice S&P 500 en 60 classes. On remarque bien que la distribution empirique des rendements n'est pas normalement aplatie. Elle a en particulier des queues de distribution trop épaisses (kurtosis supérieure à 3).

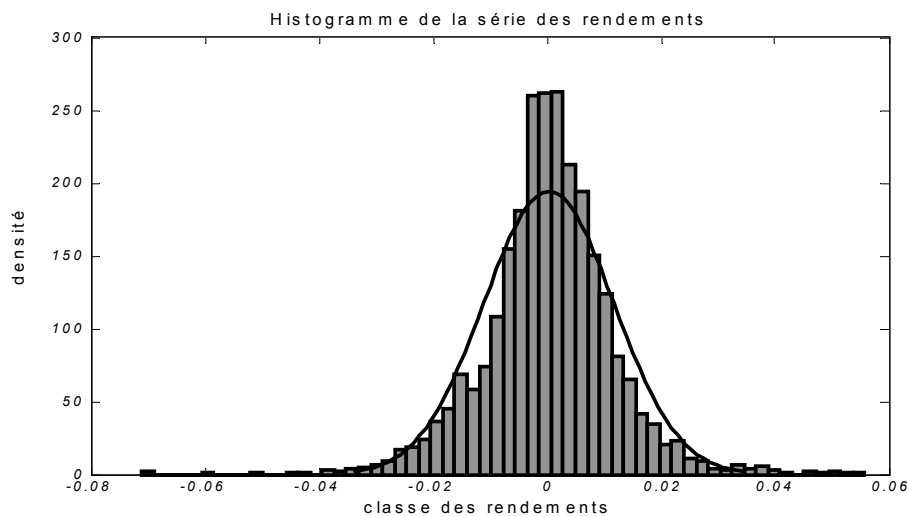


FIG. 5.3 – Héstogramme de la série des rendements de l'indice S&P500, du 03/01/1994 au 13/05/2004.

5.1.3 Hétéroscédasticité et Autocorrélation

La variance de la série du rendement change avec le temps, pour démontrer ce caractère important des séries des rendements, il est intéressant de subdiviser la série des données et de comparer ensuite l'écart type (i.e. $S = \sqrt{\text{var}(X)}$) de chacune des sous-séries obtenues...

sous-série	écart type	sous-série	écart type
1-250	2.4324	1506-1756	13.429
251-501	2.9131	1757-2007	6.7115
502-752	3.4637	2008-2258	35.019
753-1003	2.2736	2259-2509	27.512
1004-1254	47.472	2510-2610	0.1173
1255-1505	2.9734	/	/

Tableau 5.2 : *Ecart types $\times 10^{-8}$ des différentes sous-séries du rendement.*

Le tableau 5.2 représente les écarts types des différentes sous-séries (i.e. 250 jours \approx 1 année). En analysant ce tableau, il paraît évident que la série des rendements de l'indice S&P 500 de la période allant du : 03/01/1994 au 13/05/2004, ne possède pas une variance (écart type) constante. Cette série démontre donc une certaine hétéroscédasticité.

Afin de vérifier ce résultat, il est nécessaire de s'intéresser aux fonctions d'autocorrélation. Pour étudier la présence d'autocorrélation dans la série du rendement, on utilise le test de *Ljung-Box* ou celui de *Portmanteau* connu encore sous le nom de *Q-test*.

Pour tester l'autocorrélation d'une série d'observation $(X_t, t = 1, \dots, N)$, d'une variance finie ($\sigma^2 < \infty$) et sous-l'hypothèse qu'il n'existe pas d'autocorrélation, la statistique de *Ljung-Box* est donnée par :

$$Q(h) := N(N+2) \sum_{k=1}^h \frac{\hat{\rho}^2(k)}{N-k},$$

avec $\hat{\rho}$: désigne la fonction d'autocorrélation empirique ;

$$\hat{\rho}_X^2(k) = \frac{\sum_{t=k+1}^N (X_t - \bar{X})(X_{t-k} - \bar{X})}{\sum_{i=1}^N (X_t - \bar{X})^2},$$

$$\hat{\rho}_{X^2}^2(k) = \frac{\sum_{t=k+1}^N (X_t^2 - \hat{\sigma}^2)(X_{t-k}^2 - \hat{\sigma}^2)}{\sum_{i=1}^N (X_t^2 - \hat{\sigma}^2)^2} \quad \text{et} \quad \hat{\sigma}^2 = \frac{1}{N} \sum_{t=1}^N X_t^2.$$

Cette statistique suit asymptotiquement une distribution de χ^2 à h degré de liberté. L'hypothèse que les coefficients d'autocorrélations sont nuls est rejetés au seuil α , lorsque la statistique Q est supérieure au χ^2 au seuil $(1 - \alpha)$ avec h degré de liberté (i.e. $\chi^2_{1-\alpha}(h)$).

Les rendements ne démontrent qu'une faible autocorrélation (Fig 5.4), tandis que le carré et la valeur absolue des rendements sont autocorrélés à un niveau significatif. Il convient le plus souvent que le corrélogramme estimé à partir d'une série de N observations est nul, au taux 95%, si chacune des valeurs calculées appartient à l'intervalle $[-\frac{2}{\sqrt{N}}, \frac{2}{\sqrt{N}}]$, et si l'ensemble de ces valeurs est distribué au hasard autour de zéro.

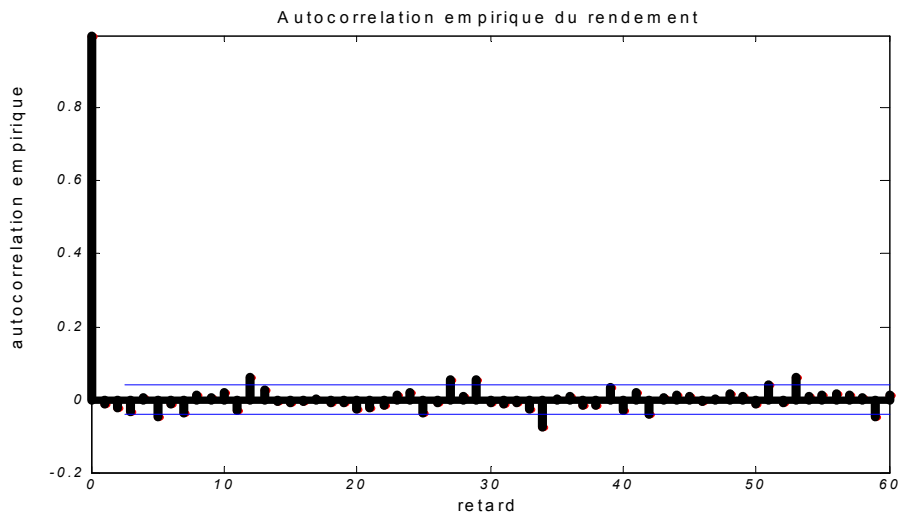


FIG. 5.4 – Autocorrélogramme de la série du rendement de l'indice S&P500.

Si une série est bruit blanc strict (i.e. il n'existe aucune autocorrélation significative), alors le sont également les séries déduites du carré et de la valeur absolue de celle-ci. Or nous remarquons clairement la présence d'une dépendance de ces variables entre elles, traduites par des autocorrélations significatives durables pour les deux séries (Fig 5.5 et Fig 5.6). Ceci nous conduit au rejet de l'hypothèse d'absence d'autocorrélation des cours et met en évidence la présence d'une hétéroscédasticité.

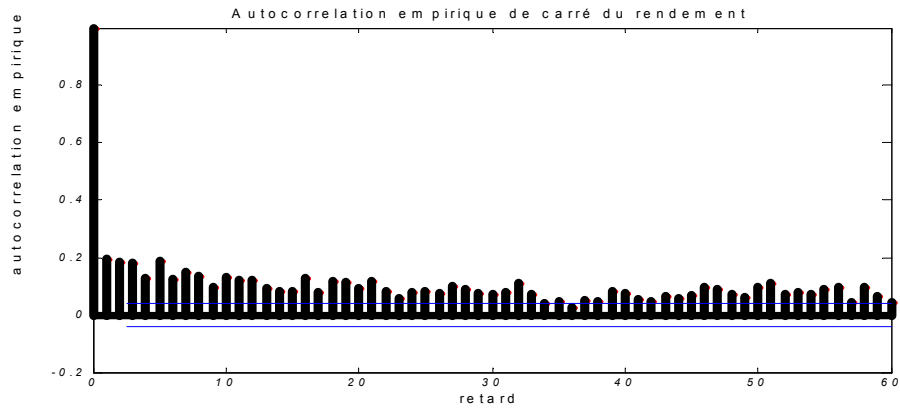


FIG. 5.5 – Autocorrélogramme de la série de carré du rendement de l'indice S&P500.

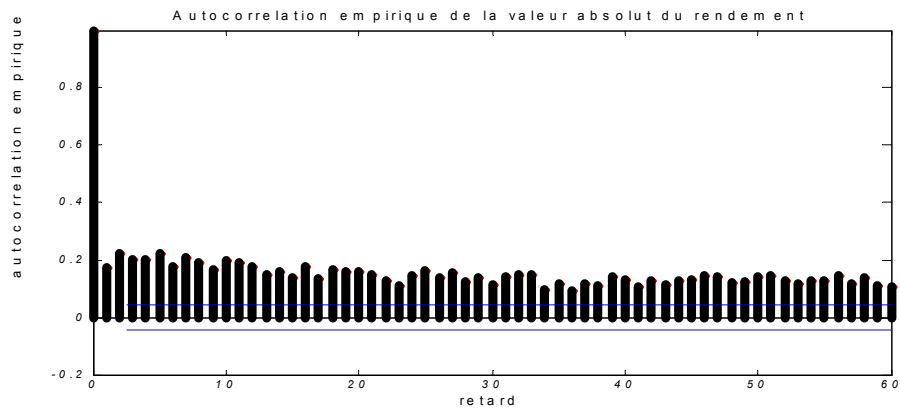


FIG. 5.6 – Autocorrélogramme de la série des valeurs absolues du rendement de l'indice S&P500.

A cause de la stationnarité, $\rho_X(k) = \text{corr}(X_t, X_{t-k}), \forall t$. en particulier, $\rho_X(k) = \text{corr}(X_n, X_{n-k})$, et quand n est la dernière valeur observée, associée au temps présent, nous dirons souvent que X_{n-k} est la variable associée au passé d'ordre k . Ainsi $\rho_X(k)$ représente l'expression du lien linéaire entre le présent et le passé d'ordre k .

Nous allons introduire ici une notion faisant intervenir les variables intermédiaires dans l'expression de la fonction d'autocorrélation de la série $(X_t)_t$. Puisque, les variables du passé sont elles-mêmes liées entre elles, les valeurs $\rho_X(k)$, $k = 1, 2, \dots$, sont donc insuffisantes pour expliquer comment le présent est relié au passé. La fonction d'autocorrélation intervient donc comme une mesure de dépendance entre les variables X_t et X_{t-k} , mais elle ne fait pas intervenir uniquement ces variables.

Définition 5.1.2 : "*Autocorrélation partielle*"

Pour une série stationnaire $(X_t)_t$, on définit la fonction d'autocorrélation partielle notée $\phi_X(k)$, par :

$$k \mapsto \phi_X(k) = \text{corr}(\tilde{X}_t, \tilde{X}_{t-k}),$$

où

$$\begin{cases} \tilde{X}_{t-k} = X_{t-k} - \mathbb{E}[X_{t-k} | X_{t-1}, \dots, X_{t-k+1}] \\ \tilde{X}_t = X_t - \mathbb{E}[X_t | X_{t-1}, \dots, X_{t-k+1}]. \end{cases}$$

Par analogie avec le corrélogramme, on trace souvent le graphe de la fonction d'autocorrélation partielle. Ce graphe constitue le corrélogramme partiel (figure 5.7), et ils jouent un rôle important dans l'estimation des ordres p et q du modèle d'ajustement. En particulier, pour le processus autoregressif, l'autocorrélogramme s'annule à partir de p et pour le processus moyenne mobile, l'autocorrélogramme partiel s'annule à partir de q .

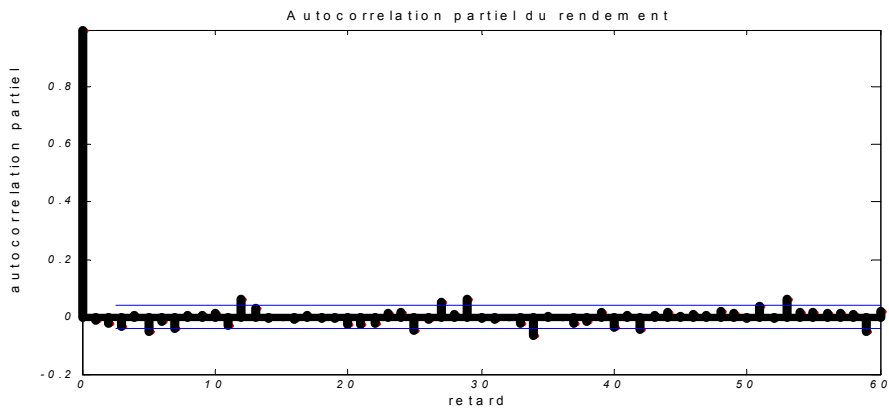


FIG. 5.7 – Autocorrélogramme partiel de la série du rendement de l'indice S&P500.

5.1.4 Test de l'hypothèse de normalité

Les premières études sur la loi de probabilité des prix boursiers ont été basées sur la loi normale, ou loi de Gauss. L'hypothèse normale sous-entend donc que la loi gaussienne modélise le mieux les variations relatives des prix d'indice boursier. L'objet de cette section est donc de tester cette hypothèse de normalité soit qu'on l'accepte ou on la rejette.

Pour tester l'hypothèse de normalité de la série $(X_t)_t$, on utilise le test de *Jarque-Bera* (1984). Ce test statistique permet de vérifier la normalité de la distribution d'un échantillon $(X_t, t = 1, \dots, N)$, il est basé sur les estimateurs empiriques des coefficients d'aplatissement et d'asymétrie. En effet, sous-hypothèse de normalité des observations, on a normalité des estimateurs de la kurtosis κ et du skewness s . Alors, quand N tend vers l'infini

$$\kappa \xrightarrow{\mathcal{D}} \mathcal{N}(3, \sqrt{24/T}) \quad \text{et} \quad s \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sqrt{6/T}).$$

Le test de Jarque-Bera confirme donc, que si la distribution de l'échantillon est normale, alors la statistique JB suivante :

$$JB = \frac{N}{24}(\kappa - 3)^2 + \frac{N}{6}s^2,$$

suit asymptotiquement une loi de khi deux à 2 degré de liberté. Ainsi, si $JB \geq \chi_{1-\alpha}^2(2)$, on rejette l'hypothèse H_0 de normalité au seuil α .

Dans notre cas, la kurtosis vaut **6.2306** et le skewness **-0.1102**. La statistique de Jarque & Bera vaut alors **1140.3**, qui est supérieur au quantile du χ^2 à deux degré de liberté au seuil $\alpha = 0.05$, soit **5.9991** (i.e. $\chi_{(95\%)}^2 = 5.99$, $\chi_{(99\%)}^2 = 1.96, \dots$). L'hypothèse de normalité est donc fortement rejetée au seuil $\alpha = 0.05$.

Remarque 5.1.1 :

Utilisant l'un des tests d'adéquation à une distribution fixée, par exemple, le test de Kolmogorov - Smirnov (§ 3.2.1) ou le test de khi deux, on arrive à rejeter une fois de plus toute hypothèse d'adéquation de la série des rendements aux lois classiques : Normales, Lognormales, Weibull et Student,...

5.2 Modélisation GARCH du S&P 500

5.2.1 Identification *a priori* du modèle

Considérons la série des prix de l'indice S&P500, observée du 03 Janvier 1994 au 13 Mai 2004 (soit 2610 observations), cette série n'est pas stationnaire (figure 5.1).

Pour stationnariser la série, considérons la différenciée correspondant aux rendements quotidiens de l'indice S&P500 sur la période 03 janvier 1994 au 13 mai 2004. Au vu de l'allure de cette série (figure 5.2), cette transformation semble stationnariser la série au second ordre.

Afin de déterminer le type de modélisation à adopter, l'analyse des moments empiriques de la série des rendements (tableau 5.1) conduit aux conclusions usuelles dans les études des cours boursiers, dont le skewness et la kurtosis sont différentes de 0 et 3, ce qui signifie que la distribution du modèle n'est pas normale mais plutôt asymétrique avec des queues épaisses caractérisant une distribution leptokurtique. Ce qui nous conduit au rejet de l'hypothèse de normalité, et de penser à une modélisation GARCH, dont les propriétés précédentes sont bien respectées (section 3.1.2).

L'étude des autocorrélogrammes de la série des rendements et de la série au carré semblent confirmer le phénomène de non linéarité. En effet, l'autocorrélogramme de la série des rendements n'indiquera pas de corrélations importantes (figure 5.4), ce qui nous conduit à penser que la série pourrait être un bruit blanc, alors que des autocorrélations significativement non-nulles apparaissent en considérant la série au carré des rendements (figure 5.5). Ces résultats incitent à adapter une modélisation de la série des rendements de type hétéroscédastique (ARCH ou GARCH) et non pas ARMA.

Au regard de l'autocorrélogramme partiel (figure 5.7), une modélisation à l'aide d'un modèle GARCH(1, 1) semblerait possible. En effet, l'autocorrélogramme et l'autocorrélogramme partiel sont significativement nuls à partir des premiers retards (i.e. $p = q = 1$).

5.2.2 Forme explicite du modèle

De l'identification *a priori*, on conclut que la forme GARCH(1, 1) est retenue comme un modèle d'ajustement adéquat du rendement de l'indice S&P500 qu'on note $(R_t)_t$, sur la période considérée.

Comme nous l'avions vu (définition 3.1.1), le modèle autoregressif conditionnellement hétéroscédastique généralisé (GARCH(1, 1)) admet l'écriture suivante :

$$X_t := \sigma_t \xi_t, \quad \sigma_t^2 := \alpha_0 + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2,$$

pour tout $t \in \mathbb{Z}$ et α_0 , α et β , des constantes vérifiant les contraintes suivantes :

$$\alpha_0 > 0, \quad \alpha \geq 0, \quad \beta \geq 0 \quad \text{et} \quad \alpha + \beta < 1, \quad (5.2.1)$$

avec σ_t est \mathcal{F}_t -mesurable, $(\xi_t)_t$ est un bruit blanc (i.e. une suite de variables aléatoires, indépendantes et identiquement distribuées), telle que

$$\mathbb{E}[\xi_t | \mathcal{F}_{t-1}] = 0, \quad \mathbb{E}[\xi_t^2 | \mathcal{F}_{t-1}] = 1.$$

Supposons maintenant que l'on cherche le rendement à la date t . Deux informations sont particulièrement intéressantes : la moyenne et la variance. Plus particulièrement, on souhaiterait modéliser la moyenne et la variance conditionnelle de R_t , conditionnellement à l'information connue (\mathcal{F}_{t-1}). La modélisation GARCH du $(R_t)_t$ revient donc à écrire

$$R_t = \mu + X_t, \quad X_t = \sigma_t \xi_t,$$

où μ et σ_t sont respectivement la moyenne et la variance conditionnelle.

L'estimation des paramètres des modèles hétéroscédastiques, est généralement fondée sur la méthode du maximum vraisemblance. L'optimisation est effectuée en supposant que la suite des variables aléatoires $(\xi_t)_t$ soit indépendante et identiquement distribuée, d'une distribution normale d'espérance nulle et de variance unité.

Notons par $\theta = (\alpha_0, \alpha, \beta)^\top$, le vecteur des paramètres du modèle à estimer, et par $\hat{\theta}$ sont estimateur (i.e. $\hat{\theta} = (\hat{\alpha}_0, \hat{\alpha}, \hat{\beta})^\top$).

L'expression du log-vraisemblance à minimiser (voir la section 3.1.3) est donc :

$$\begin{aligned}
L_N(\theta) &= \sum_{t=1}^N \left[\frac{-1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma_t^2) - \frac{1}{2} \left(\frac{X_t}{\sigma_t} \right)^2 \right] \\
&= \frac{-N}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^N \log(\alpha_0 + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2) \\
&\quad - \frac{1}{2} \sum_{t=1}^N X_t^2 (\alpha_0 + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2)^{-1},
\end{aligned}$$

N : désigne le nombre d'observations.

Grâce à la puissance de calcul des ordinateurs actuels, il est possible de minimiser effectivement le log-vraisemblance de façon numérique. Pour l'instant, le calcul des estimateurs d'un modèle *GARCH(1, 1)* sous les contraintes (5.2.1) sera effectué en utilisant le logiciel *MatLab 6p5*, qui réalise ce type de calcul de manière automatique.

Le modèle retenu est donc un *GARCH(1, 1)* de paramètres :

$$\alpha_0 = 7.6386 \cdot 10^{-7}, \quad \alpha = 0.0773 \quad \text{et} \quad \beta = 0.9200, \quad (5.2.2)$$

et de moyenne conditionnelle μ , d'estimateur $\hat{\mu}$ qui vaut $6.5860 \cdot 10^{-4}$.

On remarque que la somme $\alpha + \beta = 0.9973 < 1$, elle vérifie donc la condition (3.1.3) de la stationnarité.

D'où, finalement, la série des rendements $(R_t)_t$ admet la représentation suivante :

$$R_t = 6.586010^{-4} + \sigma_t \xi_t,$$

la variance conditionnelle σ_t^2 est donnée par :

$$\sigma_t^2 = 7.6386 \cdot 10^{-7} + 0.0773 X_{t-1}^2 + 0.9200 \sigma_{t-1}^2.$$

La sortie graphique suivante (figure 5.8), correspond à la variance conditionnelle ou volatilité du modèle (i.e. σ_t^2) avec les paramètres (5.2.2) en bas, et la série $X_t = \sigma_t \xi_t$ appelée souvent série des innovations, en haut.

La première constatation que l'on peut faire est la présence des périodes de forte volatilité sur la série de la variance conditionnelle, indiquant parallèlement l'existence des valeurs extrêmes sur la série X_t . La seconde, est que l'effet d'hétéroscédasticité est nettement clair ; la variance conditionnelle change fortement au cours du temps.

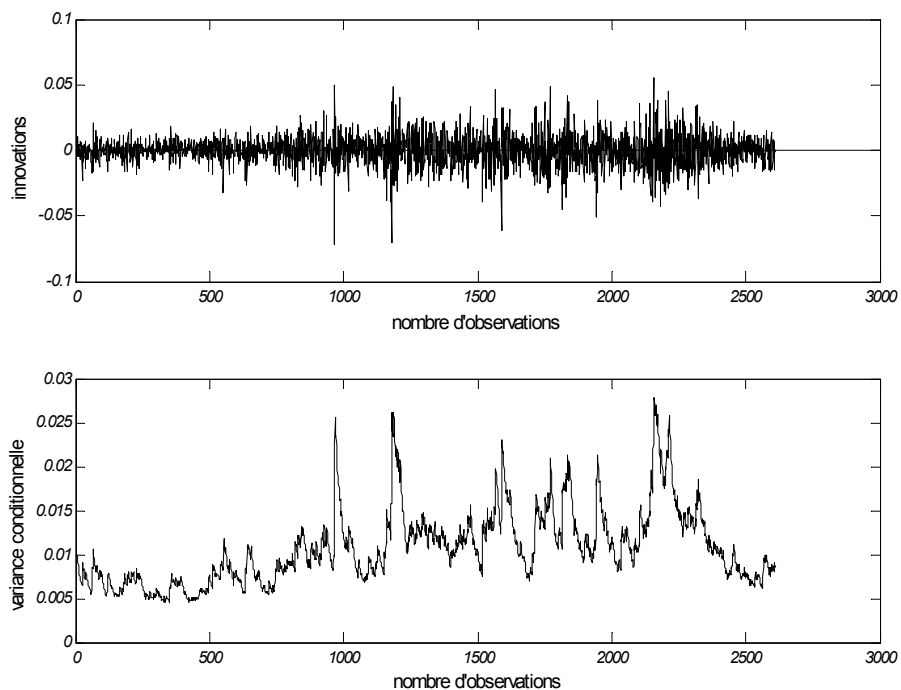


FIG. 5.8 – Variance conditionnelle en bas, et série d'innovations en haut de l'indice S&P500.

Bibliographie

- [1] Anderson, T.W. (1993). Goodness of fit tests for spectral distributions. *Ann. Statist.* **21**, 830-847.
- [2] Bartlett, M.S. (1954). Problemes de l'analyse spectrale des séries temporelles stationnaires. *Publ. Inst. Statist. Université Paris III.* **3**, 119-134.
- [3] Beirlant, J., de Wet, T. et Goegebeur, Y. (2004). A goodness-of-fit statistic for Pareto-type behaviour. Technical report, University Centre for Statistics, K.U.Leuven, (submitted : *Canadian Journal of Statistics*).
- [4] Bollerslev, T. (1986). Generalized autoregressive conditional hetereskedasticity. *J. Econometrics.* **31**, 307-327.
- [5] Bougerol, P. et Picard, N. (1992b). Stationarity of GARCH processes and of some nonnegative time series. *J. Econometrics.* **52**, 115-127.
- [6] Brockwell, P.J., et Davis, R.A. (1991). *Time Series : Theory and Methods*, 2nd edition. Springer, New York.
- [7] Brockwell, P.J., et Davis, R.A. (1996). *Introduction to Time Series and Forecasting*. Springer, New York.
- [8] Csörgő, M. et Révész, P. (1981). *Strong Approximations in Probability and Statistics*. Academic Press, New York.
- [9] Csörgő, S. et Viharos, L. (1998). Esimating the tail index. In '*Asymptotic Methods in Probability and Statistics*', ed. Szyszkowicz, B., North- Hollet, 833-881.
- [10] Davis, R.A. et Resnick, S.L. (1984). Tail estimates motivated by extreme value theory. *Ann. Statist.* **12**, 1467-1487.
- [11] Davis, R.A., Mikosch, T. et Basrak, B. (1999). Sample ACF of multivariate stochastic recurrence equations with application to GARCH. Technical Report, University of Groningen. disponible dans : *www.math.rug.nl/~mikosch*.

- [12] Dekkers, A.L.M., Einmahl, J.H.J., et de Haan, L. (1989). a moment estimator for the index of an extreme-value distribution. *Annals of Statistics*. **17**, 1833-1855.
- [13] Dekkers, A.L.M. et de Haan, L. (1989). On the estimation of the extreme-value index and large quantile estimation. *Annals of Statistics*. **17**, 1795-1832.
- [14] de Haan, L. (1976). Sample extrêmes : an elementary introduction. *Statistica Neerlandica*. **30**, 161-172.
- [15] de Haan, L., Resnick, S., Rootzén, H. et de Vries, C. (1989). Extremal behavior of solutions of a stochastic difference equation with applications to ARCH processes. *Stochastic Processes and their Applications*. **32**, 213-224.
- [16] de Haan, L. et Resnick, S. (1996). On asymptotic normality of the Hill estimator. Preprint, disponible au TR1155. ps.Z, [http : //www.orie.cornell.edu/trilst/trilst.html](http://www.orie.cornell.edu/trilst/trilst.html).
- [17] de Haan, L. et Stadtmüller, U. (1996). Generalized regular variation of second order. *J. Australian Math. Soc.*, (serie A), **61**, 381-395.
- [18] Deheuvels, P. et Martynov, G. (2004). Karhunen-Loève Expansion for weighted Weiner processes and Brownian Bridges via Bessel functions. *Progress in Probability*, **55**, 57-93. Birkhäuser Verlag, Basel.
- [19] Del Barrio, E., Custa-Albertos, J.A. et Matrán, C. (2000). Contributions of empirical and quantile to the asymptotic theory of goodness-of-fit tests. *Test*. **9**, 1-96.
- [20] Doukhan, P. (1994). *Mixing Properties and Examples* Springer Verlag.
- [21] Drees, H. (2000). Weighted approximations of tail processes for β -mixing random variables. *Annals of Applied Probability*. **10**, 1274-1301.
- [22] Drees, H. (2002). Extreme quantile estimation for dependent data with applications to finance. Preprint N° **68**. University of Saarland. disponible dans : [www. math.uni-sb.de](http://www.math.uni-sb.de).
- [23] Embrechts, P., Klüppelberg, C. et Mikosch, T.(1996). *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin.
- [24] Engle, R.F. (1982). Autoregressive conditional heteroskedastic models with estimates of the variance of the United Kingdom inflation. *Econometrica*. **50**, 987-1007.

- [25] Engle, R.F et Bollerslev, T. (1986). Modelling the presistence of conditional variances. *Econometric Rev.* **5**, 1-87.
- [26] Fisher, R.A. et Tippett, L.H.C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society.* **24**, 180-190.
- [27] Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics.* **44**, 423-453.
- [28] Gouriéroux, C. et Monfort, A. (1989). *Statistique et modèles économétriques*, Economica, Paris.
- [29] Gouriéroux, C. (1997). *ARCH Models and Financial Applications*. Springer-Verlag. New York.
- [30] Grenander, U. et Rosenblatt, M. (1984). *Statistical Analysis of Stationary Time Series*, 2nd edition. Chelsea Publishing Co. New York.
- [31] Groeneboom, P., Lopuhaä, H.P. et de Wolf, P.P. (2003). Kernel-type estimators for the extreme value index. *Annals of statistics* (à paraître).
- [32] Hall, P. (1982). On some simple estimates of an exponent of regular variation. *J. Roy. Statist. Soc. Ser. B.* **44**, 37-42.
- [33] Hill, B.M (1975). A sample approach to inference about the tail of a distribution. *Ann. Statist.* **3**, 1163-1174.
- [34] Hsing, T. (1995). On tail estimation using dependent data. *Ann. Statist.* **19**, 1547-1569.
- [35] Kallenberg, O. (1983). *Random Measures*. Third edition. Akademie-Verlag, Berlin.
- [36] Kesten, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Math.* **131**. 207-248.
- [37] Mason, D. (1982). Lows of large numbers for sums of extreme values. *Ann. Probability.* **10**, 754-764.
- [38] Mason, D. (1988). A strong invariance theorem for the tail empirical process. *Ann. Inst. Henri Poincaré.* **24**, 491-509.
- [39] Mikosch, T. et Stărică, C. (1998). Limite Theory for the sample auto-correlations and extremes of a GARCH(1,1) proses. Technical Raport. University of Groningen et Chalmers University of Gothenburg. disponible dans : *www.math.rug.nl/~mikosch*.

- [40] Mikosch, T. et Stărică, C. (2002). Changes of structure in financial time series and the GARCH model *Perprint*. disponible dans : *www.math.chalmers.se/~starica*.
- [41] Mikosch, T. et Stărică, C. (2003). Non-stationarities in financial time series, the long range dependence and the IGARCH effects. *The Review of Economics et Statistics*. (à paraître). disponible dans : *www.math.chalmers.se/~starica*.
- [42] von Mises, R.(1936). La distribution de la plus grande n valeurs. Selected Papers II (*American Mathematical Society*). 271-294.
- [43] Necir, A. (2003). On the kernel estimators of the tail index. *Bulltin of International Institut*. Berlin.
- [44] Necir, A. (2004). A functional law of the iterated logarithm for the kernel-type estimators of the tail index. *Jornal of Statitdical Planning and Inference*. (à paraître).
- [45] Necir, A. et Boukhetala, K. (2004). Estimating the risk adjusted premium for the largest reinsurance covers. *Compstat'2004, Physica-Verlag* (à paraître).
- [46] Necir, A. et Boukhetala, K. (2004). A goodness-of-fit test of tail indices via a weighted \mathbb{L}^2 -Wasserstein distance. (à paraître).
- [47] Nelson, D.B. (1990). Stationarity and presistence in the GARCH(1,1) model. *Econometric Theory*. **6**, 318-334.
- [48] Resnick, S. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer, New York.
- [49] Resnick, S. et Stărică, C.(1995). Consistency of Hill's estimator for dependent data. *J Applied probability*. **32**. 139-167.
- [50] Resnick, S. et Stărică, C.(1998). Tail index estimation for dependent data. *Ann. Applied probability*. **29**. 271-293.
- [51] Rootzén, H.(1995). The tail empirical process for stationary sequences. *Perprint*. Chalmers University of Technology.
- [52] Shorack, G. R. et Wellner, J. A. (1986). *Empirical processes with applications to statistics*. Wiley, New York.
- [53] Vervaat, W. (1979). On a stochastic difference equation and a representation of non-negative infinitely divisible random variables. *Applied probability*. **11**, 750-783.