



Remerciement

Au terme de ce travail, je souhaite remercier Madame Professeure Nadia IDRISSE FATMI, ma directrice de thèse et mon co-encadrant de thèse, Monsieur le Professeur Mohammed BADAoui pour leurs conseils avisés, leur attention de tout instant sur mes travaux, leur disponibilité, leurs nombreux encouragements et surtout à la confiance qu'ils m'ont accordé. Je suis éternellement reconnaissant pour votre soutien inconditionnel et votre écoute prépondérante pour la bonne réussite de cette thèse. J'ai beaucoup appris à vos côtés et je vous adresse ma gratitude pour tout cela.

Je tiens à remercier les membres du jury qui ont accepté de lire et évaluer ce travail. Aux rapporteurs : Professeur Driss DRIOUCHI, Professeur El Hachmi HAMMOU et Professeur Aziz IFZARNE et à l'examineur Professeur Imad HAFIDI à qui j'adresse mes sincères remerciements pour son aide considérable dans l'élaboration de ce travail.

Je saisis aussi l'occasion pour remercier les membres de notre Laboratoire d'Ingénierie des Procédés, Informatique et Mathématiques. J'associe à ces remerciements tous mes amis chercheurs. Je pense entre autres à Saida, Abdessamad, Yassin, Ghasane et Hammadi. Je ne saurais terminer sans remercier Messieurs Abdeljabbar SADIK, Abderrahim ELBOURAIMI, Khaled ELAARARI et à tous mes collègues au ministère de la justice.

Enfin, je remercie infiniment toute personne ayant participé de près ou de loin à la réalisation de ce travail.

Résumé

Cette thèse présente une contribution par un ensemble d'algorithmes numériques pour l'estimation par quasi-maximum de vraisemblance des modèles GARCH(p, q), CGARCH(N) et log-GARCH(p, q), pour tous les ordres $p \geq 1$, $q \geq 1$ et $N \geq 1$. Notre approche est basée sur une combinaison des méthodes de filtrage optimal (le filtre de Kalman sans et sous contrainte, et le filtre d'information) destinées à l'estimation de la variance conditionnelle de ces modèles, et les méthodes d'approximation stochastique (l'algorithme des perturbations simultanées (SPSA) et l'algorithme fonctionnel lissé) pour l'estimation des paramètres. Ces algorithmes permettent de remédier à quatre problèmes restrictifs de ces modèles, à savoir, la positivité de la variance conditionnelle des modèles GARCH et CGARCH, l'existence des moments du log-GARCH et la robustesse de son kurtosis, ainsi que l'existence des valeurs nulles dans les séries des rendements sujets d'ajustement par le modèle log-GARCH.

Ce travail porte sur la pré-estimation robuste de la variance conditionnelle générée par le modèle GARCH vis à vis des contraintes de positivité des paramètres. Pour cela, le filtre de Kalman sous contrainte est utilisé d'une manière permettant au modèle de tenir compte de toute information de bornitude de la volatilité lors de l'estimation des paramètres. Il s'agit d'une relaxation du modèle GARCH qui permet une estimation flexible du modèle CGARCH par rapport aux ordres élevés en fournissant un espace des paramètres adéquat pour l'estimation par quasi-maximum de vraisemblance au sens de la méthode SPSA. Nous proposons également un algorithme intégrant le filtre d'information et l'algorithme fonctionnel lissé dans l'estimation par quasi-maximum de vraisemblance des paramètres du modèle log-GARCH en relaxant cette fois-ci les conditions d'existence des moments du modèle. Ainsi, des estimations des mesures robustes du kurtosis sont fournies suite à cette relaxation. Par la suite, cet algorithme est combiné avec l'algorithme EM pour traiter par imputation les valeurs nulles dans les séries des rendements sujets d'ajustement par le modèle log-GARCH. Les résultats sont validés théoriquement via des simulations Monte Carlo, ainsi que par des applications empiriques sur des séries financières réelles.

Abstract

This thesis presents a contribution by a set of numerical algorithms for the quasi-maximum likelihood estimation of GARCH(p, q), CGARCH(N) and log-GARCH(p, q) models, for all orders $p \geq 1$, $q \geq 1$ and $N \geq 1$. Our approach is based on a combination of optimal filtering methods (unconstrained and constrained Kalman filter, and the information filter) for the pre-estimation of the conditional variance of these models, and stochastic approximation methods (the simultaneous perturbation algorithm (SPSA) and the smoothed functional algorithm) for the estimation of the parameters. These algorithms address four restrictive limits of these models, namely, the non-negativity of the conditional variance of the GARCH and CGARCH models, the existence of the moments of the log-GARCH and the robustness of its kurtosis, and the existence of zeros in the return series fitted by the log-GARCH model.

This work focuses on the robust pre-estimation of the conditional variance generated by the GARCH model with respect to the non-negativity constraints of the parameters. To this aim, a constrained Kalman filter is used in order to allow the model to take into account any volatility boundedness information when estimating the parameters. Such a relaxation of the GARCH model that allows a flexible estimation of the CGARCH model with respect to high orders by providing an adequate parameter space for the quasi-maximum likelihood estimation in the sense of the SPSA method. We propose also an algorithm incorporating the information filter and the smoothed functional algorithm in the quasi-maximum likelihood estimation of the log-GARCH parameters by relaxing this time the conditions of existence of the moments. Thus, estimates of robust measures of kurtosis are provided following this relaxation. Subsequently, this algorithm is combined with the EM algorithm to treat by imputation of zeros in the series of returns fitted by the log-GARCH model. The results are validated theoretically through Monte Carlo simulations, as well as through empirical applications to real financial series.

ملخص

تقدم هذه الأطروحة مجموعة من الخوارزميات العددية التي تهدف إلى تقدير شبه الاحتمال الأرجح لنماذج $GARCH(p,q)$ ، $CGARCH(N)$ و $\log-GARCH(p,q)$ من أجل كل الرتب $p \geq 1$ ، $q \geq 1$ و $N \geq 1$. نعتد في ذلك على مزيج من طرق الترشيح المثلى (مرشح كالمان المقيد ومرشح المعلومات) قصد تقدير المغايرة المشروطة لهذه النماذج، و أساليب التقريب العشوائي (خوارزمية الاضطراب المتزامن (SPSA) و الخوارزمية الدالية السلسلة) لتقدير البارامترات. تسمح هذه الخوارزميات بمعالجة أربع مشاكل مقيدة لهذه النماذج، ألا وهي موجبية المغايرة المشروطة لنماذج $GARCH$ و $CGARCH$ ، وجود عزوم نموذج $\log-GARCH$ ، متانة التفرطح، فضلاً عن وجود القيم المنعدمة في سلاسل المردودات المراد تقديرها بنموذج $\log-GARCH$.

يركز هذا العمل على التقدير المسبق المتين للمغايرة المشروطة لنموذج $GARCH$ دون اعتبار شروط موجبية البارامترات. لهذا، يتم استخدام مرشح كالمان المقيد بطريقة تسمح للنموذج باستغلال أي معلومات عن محدودية المغايرة المشروطة عند تقدير البارامترات. يتعلق الأمر بتخفيف لنموذج $GARCH$ يسمح بتقدير مرن لنموذج $CGARCH$ فيما يتعلق بالرتب العالية عن طريق اقتراح فضاء للبارامترات يسمح بتقديرها بطريقة شبه الاحتمال الأرجح. كما نقترح خوارزمية تدمج مرشح المعلومات والخوارزمية الدالية السلسلة لتقدير شبه الاحتمال الأرجح لبارامترات نموذج $\log-GARCH$ عن طريق تخفيف شروط وجود عزومات النموذج. ومن ثم، نقدم تقديرات للقياسات المتينة للتفرطح بعد هذا التخفيف. بعد ذلك، يتم الجمع بين هذه الخوارزمية وخوارزمية EM لمعالجة القيم المنعدمة في سلاسل المردودات المراد تقديرها بنموذج $\log-GARCH$. تم التحقق من صحة النتائج نظرياً عن طريق محاكاة Monte Carlo، فضلاً عن تطبيقات تجريبية على سلاسل مالية حقيقية.

Table des matières

Table des figures	viii
Liste des tableaux	x
Liste des algorithmes	x
Notations et Abréviations	xii
Introduction Générale	1
1 Généralités	7
1.1 Notions d'espérance conditionnelle	7
1.2 Modèles autorégressifs conditionnellement hétéroscédastiques	8
1.2.1 Modèle ARCH	8
1.2.2 Modèle GARCH	9
1.2.3 Modèle CGARCH (Component GARCH)	15
1.2.4 Modèle log-GARCH	17
1.3 Filtrage optimal	19
1.3.1 Système dynamique	19
1.3.2 Filtre de Kalman	20
1.3.3 Filtre de Kalman sous contrainte	23
1.3.4 Filtre d'information	26
1.4 Optimisation par approximation stochastique	27
1.4.1 Algorithme d'approximation stochastique à perturbations simultanées (SPSA)	27
1.4.2 Algorithme fonctionnel lissé (Smoothed Functional)	28
2 Pré-estimation robuste de la variance conditionnelle générée par un GARCH	31
2.1 Modèle GARCH relaxé	34
2.2 Représentation espace d'état du modèle r -GARCH et structure de covariance	37

2.2.1	Représentation espace d'état du modèle r -GARCH	37
2.2.2	Structure d'autocovariance	38
2.3	Estimation robuste de la variance conditionnelle	43
2.3.1	Prédiction	43
2.3.2	Robustification	44
2.4	Illustration numérique	45
2.4.1	Estimation des paramètres du GARCH relaxé	46
2.4.2	Redressement des estimations en présence d'information sur la volatilité	53
2.5	Conclusion	55
3	Estimation des paramètres des modèles r-GARCH et CGARCH : Étude de convergence	56
3.1	Estimation par QCKSA du modèle GARCH relaxé	58
3.1.1	Maximisation de la vraisemblance par SPSA	58
3.1.2	Illustration numérique	60
3.2	Estimation du modèle CGARCH	63
3.2.1	Analyse de convergence	65
3.2.2	Illustration numérique	70
3.3	Conclusion	74
4	Estimation du modèle log-GARCH basée sur le filtre d'information	75
4.1	Représentation espace d'état du modèle log-GARCH	77
4.2	Filtre d'information	79
4.3	Estimation par quasi-maximum de vraisemblance du log-GARCH	83
4.3.1	Estimation du kurtosis droit et gauche	84
4.3.2	Estimation des paramètres du log-GARCH en présence d'ob- servations nulles	86
4.4	Illustration numérique	88
4.4.1	Estimation par QIFSF des paramètres du log-GARCH	89
4.4.2	Estimation du kurtosis droit et gauche	89
4.4.3	Estimation par QIFSF du log-GARCH en présence d'observa- tions nulles	91
4.5	Conclusion	95
	Conclusions et Perspectives	96
	Bibliographie	107

Table des figures

2.1	Volatilités estimées par QCK (à gauche) et par QML (à droite) pour une simulation de taille $n=500, 1000$ et 5000 du GARCH(1,1). Le graphe en noire indique la volatilité originale simulée.	48
2.2	Volatilités estimées par QCK (à gauche) et par QML (à droite) pour une simulation de taille $n=500, 1000$ et 5000 du GARCH(3,1). Le graphe en noire indique la volatilité originale simulée.	50
2.3	Volatilités estimées par QCK (à gauche) et par QML (à droite) pour une simulation de taille $n=500, 1000$ et 5000 du GARCH(2,3). Le graphe en noir indique la volatilité originale simulée.	52
2.4	L'erreur absolue δ des estimations par QCK en fonction de la précision i . La ligne pointillée indique $\delta(\tilde{\omega})$, les lignes continues grises et noires représentent respectivement $\delta(\tilde{\alpha}_1)$ et $\delta(\tilde{\beta}_1)$	54
3.1	Taux de change journalier du mark/dollar (à gauche) et la série des rendements correspondante (à droite) du 1 ^{er} Juin 1973 au 28 Janvier 1985.	62
3.2	Taux de change journalier du franc/dollar (à gauche) et la série des rendements correspondante (à droite) du 1 ^{er} Juin 1973 au 28 Janvier 1985.	62
3.3	Taux de change journalier du yen/dollar (à gauche) et la série des rendements correspondante (à droite) du 1 ^{er} Juin 1973 au 28 Janvier 1985.	62
3.4	L'indice S&P500 (à gauche) et la série des rendements correspondante (à droite) du 28 Octobre 2010 au 27 Novembre 2020.	71
3.5	La volatilité de la composante à long terme $\sigma_{1,t}^2$ (à gauche) et la composante à court terme $\sigma_{2,t}^2$ (à droite) estimées par QML.	73
3.6	La volatilité de la composante à long terme $\sigma_{1,t}^2$ (à gauche) et la composante à court terme $\sigma_{2,t}^2$ (à droite) estimées par QCKSA.	73

4.1	Biais des estimations par QML (ligne noire continue), par ECQ (ligne noire pointillée) et par EM-QIFS (ligne grise pointillée) pour différentes proportions d'observations nulles.	92
4.2	Rendements journaliers en pourcentage.	93

Liste des tableaux

2.1	Propriétés en échantillons finis des modèles r -GARCH(1,1) et GARCH(1,1) estimés respectivement par QCK et QML.	47
2.2	Propriétés en échantillons finis des modèles r -GARCH(3,1) et GARCH(3,1) estimés respectivement par QCK et QML.	49
2.3	Propriétés en échantillons finis des modèles r -GARCH(2,3) et GARCH(2,3) estimés respectivement par QCK et QML.	51
2.4	Précision des volatilités estimées par QCK et QML pour les processus GARCH(1,1), GARCH(3,1) et GARCH(2,3).	51
2.5	Estimations des paramètres par QCK en fonction du niveau d'information i	54
3.1	Modèles GARCH estimés pour les taux de change journaliers du 1 ^{er} Juin 1973 au 28 Janvier 1985.	61
3.2	Statistiques descriptives des rendements du S&P500.	71
3.3	Estimation des paramètres du CGARCH(2) par QML et QCKSA.	72
4.1	Propriétés en échantillons finis du log-GARCH(1,1) estimé par QML, ECQ et QIFSF.	90
4.2	Propriétés en échantillons finis des mesures du kurtosis droit et gauche calculées par les estimation QML, ECQ et QIFSF.	91
4.3	Propriétés en échantillons finis du log-GARCH(1,1) estimé par QML, ECQ et EM-QIFSF pour différentes proportions d'observations nulles.	92
4.4	Statistiques descriptives des rendements.	94
4.5	Estimations empiriques des spécifications log-GARCH(1,1) des trois séries de rendements.	95

Liste des algorithmes

2.1	Prédiction de la variance conditionnelle du GARCH par le filtre de Kalman.	44
2.2	Robustification de la variance conditionnelle prédite par troncature de la densité de probabilité.	45
3.3	Minimisation par SPSA de la log-vraisemblance du GARCH.	60
3.4	Minimisation par SPSA de la log-vraisemblance du CGARCH.	70
4.5	Prédiction de la variance conditionnelle du log-GARCH par le filtre d'information.	82
4.6	Minimisation par SF de la log-vraisemblance du log-GARCH.	84
4.7	Estimation du log-GARCH par EM-QIFS.	88

Notations

$\mathbb{N}, \mathbb{Z}, \mathbb{R}$	L'ensemble des entiers naturels, entiers relatifs, réels.
$\mathcal{M}_{(k,l)}$	L'ensemble des matrices réelles de taille (k, l) .
$0_{(k,l)}$	La matrice nulle de $\mathcal{M}_{(k,l)}$.
I_k	La matrice identité de $\mathcal{M}_{(k,k)}$.
${}^t M$	La matrice transposée d'une matrice $A \in \mathcal{M}_{(k,l)}$.
$Abs(M)$	La matrice de même taille que A formée des éléments de A en valeur absolue.
$Sp(M)$	Spectre d'une matrice A constitué de toute ses valeurs propres.
$\ M\ $	La norme d'une matrice $M = (a_{ij})_{1 \leq i, j \leq k} \in \mathcal{M}_{(k,k)}$ définie par $\sum_{i,j} a_{ij} $.
$M_{i\bullet}$	La i^{eme} ligne de M .
$M_{\bullet j}$	La j^{eme} colonne de M .
$M_{\times \bullet}$	${}^t ({}^t M_{1\bullet} \dots {}^t M_{k\bullet}) \in \mathbb{R}^{kl}$
$M_{\bullet \times}$	${}^t (M_{\bullet 1} \dots M_{\bullet l}) \in \mathbb{R}^{k^2}$
$M_{\bullet l}^{(s)}$	${}^t (a_{sl} \dots a_{kl})$
$M_{\bullet \times}^{(s)}$	${}^t (M_{\bullet 1}^{(s)} \dots M_{\bullet k}^{(s)}) \in \mathbb{R}^{(k-s+1)^2}$
\otimes	Produit de Kronecker.
ρ	Rayon spectral.
ρ_k	Fonction d'autocorrélation d'ordre k .
iid	Indépendant et identiquement distribué.
$iid(.,.)$	Indépendant et identiquement distribué selon une loi paramétrée par sa moyenne et sa variance.
$M^{(m)}$	$\mathbb{E} (Abs(M_t))^{\otimes m}$ pour toute suite iid de matrices aléatoires (M_t) et $m \in \mathbb{N}$.
${}^t e_k$	Le premier vecteur de la base canonique de \mathbb{R}^k .
$\ \cdot\ $	Toute norme usuelle de \mathbb{R}^k
$\ X\ _m$	La m -norme d'une variable aléatoire X , définie par $\{\mathbb{E}\ X\ ^m\}^{1/m}$.

L	Opérateur de retard polynomial tel que $LX_k = X_{k-1}$.
$[\cdot]$	La partie entière.
\underline{X}_t	(X_t, X_{t-1}, \dots)
$p.s$	Presque sûrement.
$\log^+ x$	$\max(\log x, 0)$.
$N(\cdot, \cdot)$	La distribution normal ou gaussienne paramétrée par sa moyenne et sa variance.
$\mathbb{1}_E$	Fonction indicatrice valant 1 en tout point de E et 0 ailleurs.
∇	Opérateur du gradient.
\mathcal{C}^∞	La classe des fonctions indéfiniment dérivables.
$\mathcal{U}_{[a,b]}$	La loi uniforme sur $[a, b]$.

Abréviations

ARMA	Autorégressif moyenne mobile.
MLE	Maximum de vraisemblance.
QML	Quasi-maximum de vraisemblance.
QCK	Quasi-maximum de vraisemblance basé sur le filtre de Kalman sous contrainte.
QCKSA	Quasi-maximum de vraisemblance basé sur le filtre de Kalman sous contrainte et l'algorithme d'approximation stochastique.
QMLNC	Quasi-maximum de vraisemblance basé sur les contraintes relaxées de Nelson et Cao.
QIFSF	Quasi-maximum de vraisemblance basé sur le filtre d'information et l'algorithme fonctionnel lissé.
ECQ	<i>Exponential Chi-squared.</i>
EM	<i>Expectation-Maximization.</i>
EM-QIFSF	Quasi-maximum de vraisemblance basé sur le filtre d'information et l'algorithme fonctionnel lissé combinés avec l'algorithme EM.
MSE	Erreur quadratique moyenne donnée par $\mathbb{E}(\theta - \hat{\theta})^2$, avec $\hat{\theta}$ un estimateur de θ .
BIC	Critère d'information bayésien.
AIC	Critère d'information d'Akaike.

Introduction Générale

Il est incontestable que l'introduction des modèles autorégressifs conditionnellement hétéroscédastiques a instauré les piliers de l'économétrie financière en réalisant ainsi un progrès très marquant de l'analyse des séries temporelles issues des domaines financiers et monétaires. Par conséquent, la théorie financière s'est placée dans un cadre quantitatif qui permettait non seulement de valider statistiquement ses modèles fondamentaux, mais aussi d'intégrer entre autres, la notion du risque dans la prévision de la rentabilité des différents actifs financiers à différents termes.

Dans un contexte dominé par les modèles ARMA, ayant été largement popularisé grâce à la méthodologie d'estimation de Box-Jenkins, il s'est avéré que les séries financières et monétaires (prix d'action, taux d'intérêt, taux de change, etc.) s'échappent de l'éventail de validité de ce cadre en imposant une dynamique particulière loin d'être reflétée par les modèles ARMA. En effet, les séries des rendements¹ associées aux séries précédentes présentent de nombreuses régularités statistiques (ou faits stylisés) qui ont été exhibées et empiriquement vérifiées par MANDELROT (1963) et FAMA (1965). Parmi celles qui sont inappropriées aux modèles ARMA, on peut citer :

- (i) Les rendements ne démontrent qu'une faible autocorrélation, tandis que les carrés des rendements (ou en valeur absolue) sont souvent fortement autocorrélés. Cela ne contredit pas l'hypothèse supposant que la série des rendements est un bruit blanc mais démontre de plus que le bruit blanc n'est pas indépendant.
- (ii) Les carrés des rendements (ou en valeur absolue) connaissent un regroupement de fortes variations que l'on appelle *volatility clustering* en anglais. Autrement dit, une forte variation de la série est souvent suivie par une autre forte variation, croissante ou décroissante. En conséquence, puisqu'une forte valeur du carré du rendement à la date $t - 1$ semble être suivie par une forte valeur à la date t , la variance de la série des rendements conditionnellement

1. Le rendement ε_t (ou log rendement) d'une série $(X_t)_t$ peut être défini par $\varepsilon_t = \log(X_t/X_{t-1})$ pour $t \geq 1$.

à ses valeurs passées (appelée aussi volatilité) ne semble pas constante (hétéroscédasticité conditionnelle).

Face à ces caractéristiques, on aperçoit l'insuffisance des modèles ARMA étant limités à l'exploitation des moments d'ordre 2 du processus (structure d'autocovariance). En conséquence, ces modèles sont inopérants à la fois, vis à vis de la propriété (i), et en présence du phénomène d'hétéroscédasticité conditionnelle abordé dans (ii), car dans ces cas, l'ARMA traiterait sans distinction toutes les séries des rendements comme des bruits blancs (faibles ou forts).

Robert F ENGLE (1982) propose initialement les modèles ARCH (Autorégressifs Conditionnellement hétéroscédastiques) où l'aspect hétéroscédastique est introduit via une dynamique autorégressive sur les carrés des rendements, en exprimant la variance conditionnelle (volatilité) des rendements comme fonction déterministe de leur passé. Selon une telle spécification, les faits stylisés précités sont justement vérifiés. Cependant, on observe sur les séries financières une décroissance assez lente des autocorrélations. Une modélisation ARCH réaliste nécessiterait ainsi un ordre très élevé, et donc un grand nombre de paramètres à estimer. En remédiant à cela, BOLLERSLEV (1986) généralise parcimonieusement les modèles ARCH en introduisant les modèles GARCH où des termes autorégressifs de la variance conditionnelle interviennent dans l'expression de cette dernière. Ainsi, les spécifications de type GARCH ne cessaient de se diversifier en constituant une famille de modèles GARCH recouvrant les différents aspects des séries financières et monétaires qu'un modèle omet par rapport à l'autre (e.g., contraintes de positivité des paramètres, asymétrie, changement de régimes, *etc.*). Nous nous focalisons dans la présente thèse sur les modèles : GARCH standard, CGARCH (Component GARCH) et log-GARCH.

Bien que la notion de variance conditionnelle représente l'apport fondamental des modèles GARCH, le GARCH standard ne maintient sa cohérence qu'en contraignant ses paramètres afin d'assurer la positivité de sa variance conditionnelle. Cela dit, des contraintes de positivité (BOLLERSLEV (1986)), où au plus sous forme d'inégalités plus relaxées (NELSON et CAO (1992)), doivent être identifiées à priori et imposées sur les paramètres lors de l'estimation. Cette cohérence dans la construction du GARCH standard est obtenue au détriment de sa flexibilité théorique (procédure d'estimation) et de sa performance empirique. En effet, sur le plan de l'estimation, la positivité des paramètres est simple à implémenter dans un programme d'optimisation sous contraintes. Mais cela contredit de nombreux résultats empiriques portant d'une part sur le comportement oscillatoire du processus de la variance conditionnelle (NELSON (1991)), et d'autre part, sur la flexibilité du processus GARCH quant aux formes d'autocorrélation du carré des rendements (HE et TERÄSVIRTA (1999b)). De plus, plusieurs études empiriques avaient montré des violations de la positivité des paramètres en fournissant des estimations significativement négatives (R. ENGLE, ITO et LIN (1990) ; FRENCH, SCHWERT et STAMBAUGH (1987)).

Certe, sous les contraintes relaxées de NELSON et CAO (1992), le GARCH devient empiriquement plus réaliste mais moins flexible au niveau d'estimation des paramètres en exigeant l'identification à priori de ces contraintes en fonctions des ordres du modèle. D'un point de vue numérique, cette approche n'est pas pratique car elle est difficilement convertible en un algorithme permettant l'estimation du GARCH

en fonction de ses ordres car en effet, les contraintes de Nelson et Cao sont différemment données pour chaque fourchette d'ordres en se limitant à l'ordre $q \geq 2$. C'est la raison pour laquelle, l'identification usuelle des ordres du GARCH selon l'approche de Box-Jenkins² ou par la méthode des coins (FRANCQ (2010); GOURIÉROUX et MONFORT (1995)) ne sont pas pratiques car elles nécessiteraient l'exécution de plusieurs algorithmes associés chacun à une certaine fourchette d'ordres à tester. De plus, ces approches ne cessent de se compliquer pour des GARCH d'ordres élevés. Par ailleurs, une limite majeure des modèles GARCH est qu'ils n'exploitent pas l'information sur la volatilité dans l'amélioration et le redressement des estimations des paramètres, notamment celle de bornitude. C'est le cas, par exemple, de la modélisation des prix sous un régime d'indexation où la volatilité reste plafonnée. Il s'agit d'une information plus fine qui généralise la positivité de la variance conditionnelle et qui doit être prise en considération lors de l'estimation des paramètres, chose que les GARCH ignorent totalement.

Ces considérations s'étendent au modèle CGARCH (Component GARCH) (DING, GRANGER et Robert F ENGLE (1993); Robert F ENGLE et G. LEE (1999)) en héritant les lacunes du GARCH standard. D'un côté, du fait de sa structure de variance conditionnelle décomposée en une somme de variances conditionnelles issues de spécifications généralement de type GARCH et d'un autre côté, parce qu'un CGARCH est en définitif un GARCH restreint. Cependant, si les contraintes relaxées de positivité de la variance conditionnelle au sens de Nelson et Cao sont prédéfinies et applicables directement pour certaines fourchettes d'ordres du GARCH standard, l'estimation des paramètres du CGARCH pose de plus une difficulté d'identification de ces contraintes. Cela requiert l'identification ex-post des paramètres du GARCH restreint résultant du CGARCH en fonction des paramètres de ce dernier. En ce sens, toute amélioration des méthodes d'estimation du GARCH se reflète sur celles du CGARCH en accord avec le contexte d'introduction de ce modèle, à savoir, mieux capter la dynamique à long et à court terme du processus de volatilité qu'il génère.

Par ailleurs, la famille des GARCH tendait à englober de nouvelles spécifications plus flexibles qui se libèrent des contraintes de positivité de la variance conditionnelle, e.g., EGARCH (NELSON (1991)) et GJR-GARCH (GLOSTEN, JAGANNATHAN et RUNKLE (1993)). À ces modèles s'ajoute le modèle log-GARCH (symétrique) (GEWEKE (1986); MILHØJ (1987); PANTULA (1986)) dont la variance conditionnelle est donnée en logarithme, en fonction du logarithme du passé des carrés des rendements. Toutefois, le log-GARCH est par construction inapplicable pour des séries de rendements contenant des valeurs nulles. Il s'agit d'une lacune que la littérature des log-GARCH traite soit par élimination des valeurs nulles (Christian FRANCQ, WINTENBERGER et ZAKOIAN (2013a)), soit par des algorithmes d'imputation de type EM (Expectation-Maximization) (SUCARRAT et ESCRIBANO (2013)).

Un autre fait stylisé des séries des rendements financiers concerne leurs distributions de probabilité qui sont généralement différentes d'une distribution normale. En effet, les densités de probabilité de ces séries présentent des queues épaisses et des pics en zéro que l'on appelle leptokurtiques. Une mesure de cet effet est obtenue

2. Cette approche est basée sur la sélection du modèle minimisant les critères AIC ou BIC après l'estimation du modèle pour différents ordres.

à partir du coefficient de kurtosis comme le rapport du moment empirique centré d'ordre 4 et du carré de la variance empirique. Or, en présence de leptokurticité, les moments supérieurs de ces séries peuvent bien ne pas exister, y compris celui d'ordre 4.

D'autre part, de nombreux travaux recouraient aux méthodes de filtrage optimal via des représentations espace d'état pour l'estimation de la variance conditionnelle générée par des modèles de la famille GARCH, notamment le filtre de Kalman (ALLAL et BENMOUMEN (2011); FERREIRA et al. (2017); KULIKOVA, TSYGANOVA et KULIKOV (2019); OSSANDÓN et BAHAMONDE (2013)). Dans le cadre d'estimation par maximum de vraisemblance, l'intérêt de tel outil est de construire la fonction de vraisemblance du modèle à travers la pré-estimation optimale de sa variance conditionnelle (FERREIRA et al. (2017); KULIKOVA, TSYGANOVA et KULIKOV (2019); OSSANDÓN et BAHAMONDE (2013)) en évitant d'imposer des valeurs initiales pour la récursion de l'équation de la variance conditionnelle (ALLAL et BENMOUMEN (2011)) contrairement aux méthodes d'estimations standards (QML et MCO) où ses valeurs sont fixées à partir des observations de l'échantillon.

Dans la présente thèse, nous intégrerons les méthodes de filtrage optimale, précisément, le filtre de Kalman et le filtre d'information dans l'estimation par QML des modèles GARCH, CGARCH et log-GARCH afin de la rendre plus relaxée vis à vis des contraintes de positivité de la variance conditionnelle (GARCH et CGARCH), des conditions d'existence des moments d'ordre élevé et l'existence des rendements nuls (log-GARCH). En effet, nous utiliserons le filtre de Kalman (sans et sous contrainte) pour :

- (i) Pré-estimer d'une manière robuste³ la variance conditionnelle générée par le GARCH standard en satisfaisant d'une part les contraintes relaxées de Nelson et Cao sans qu'elles soient ni identifiées ni appliquées et en permettant d'autre part d'intégrer toute information de bornitude de la volatilité dans la procédure d'estimation.
- (ii) Améliorer l'estimation par QML des modèles GARCH et CGARCH en utilisant la pré-estimation robuste de leurs variances conditionnelles.

Nous combinerons aussi le filtre d'information avec l'estimation par QML afin de :

- (iii) Estimer les paramètres du modèle log-GARCH en relaxant les conditions d'existence de ses moments.
- (iv) Estimer d'une manière robuste le kurtosis du processus log-GARCH en absence du moment d'ordre 4.
- (v) Estimer les paramètres du modèle log-GARCH en présence des rendements nuls.

Évidemment, la performance de ces estimations est impérativement liée au choix des méthodes d'optimisation utilisées. La littérature des GARCH n'accordait pas suffisamment d'intérêt à ce volet en se contentant de faire appel aux méthodes

3. La robustesse vis à vis des contraintes de positivité des paramètres du GARCH.

itératives numériques (e.g., Newton-Raphson, Berndt–Hall–Hall–Hausman (BHHH) (BERNDT et al. (1974); BOLLERSLEV (1986))). Dans la présente thèse, nous nous baserons sur les méthodes d’approximation stochastique, précisément l’algorithme des perturbation simultanées (SPSA) et l’algorithme fonctionnel lissé (Smoothed Functional), afin de les intégrer dans l’estimation par QML des modèles GARCH standard, CGARCH et log-GARCH. Nous notons que le choix de cette classe des méthodes d’optimisation n’est pas arbitraire, nous justifierons effectivement ce choix lié à l’effet de la pré-estimation robuste de la variance conditionnelle sur la construction de la fonction de vraisemblance.

Organisation du mémoire

Après avoir présenté le cadre général de ce travail de recherche qui concerne l’estimation des modèles GARCH, CGARCH et log-GARCH, et après avoir cerné les éléments problématiques et les objectifs du mémoire, la suite de ce document est composée de quatre chapitres, organisés comme suit :

- **Le Chapitre 1** Nous présentons un état de l’art des trois modèles ciblés dans ce travail, en rappelant leurs propriétés principales auxquelles nous ferons appel dans les chapitres suivants. Ce chapitre décrit également les méthodes de filtrage optimal, précisément le filtre de Kalman (sans et sous contrainte) et le filtre d’information et finit par donner un aperçu sur les méthodes d’approximation stochastique, en particulier l’algorithme de perturbation simultanée et l’algorithme fonctionnel lissé. Il s’agit donc d’une présentation globale de la boîte à outils utilisée dans ce travail.
- **Le Chapitre 2** Nous proposons une pré-estimation robuste de la variance conditionnelle générée par le modèle GARCH d’ordres $p, q \geq 1$, tout en restant positive en relaxant les contraintes de positivité des paramètres de BOLLERSLEV (1986) au sens de NELSON et CAO (1992). Suite à cette relaxation basée sur le filtre de Kalman sous contrainte, de nouvelles conditions de stationnarité et d’existence des moments du GARCH relaxé sont établies en déterminant sa structure d’autocovariance. Une illustration numérique par la méthode Monte Carlo met en évidence les résultats de ce chapitre. Ce chapitre fait l’objet d’une publication :
 - Settar, A., Fatmi, N. I., and Badaoui, M. (2021). New Approach in Dealing with the Non-Negativity of the Conditional Variance in the Estimation of GARCH Model. Central European Journal of Economic Modelling and Econometrics, 55-74.
- **Le Chapitre 3** Nous nous focalisons d’une part sur l’estimation par QML des paramètres du GARCH standard en utilisant l’algorithme SPSA. D’autre part, cette estimation est employée pour construire à priori un espace des paramètres dans lequel l’algorithme SPSA appliqué à l’estimation par QML des paramètres du CGARCH converge, sans l’identification ex-post de la forme

GARCH correspondante. Une illustration empirique examine la performance des estimations proposées.

- **Le Chapitre 4** Nous nous intéressons à l'estimation par QML des paramètres du modèle log-GARCH d'ordres $p, q \geq 1$. Cette fois-ci, l'estimation est combinée avec l'algorithme fonctionnel lissé sous des conditions réduites d'existence des moments. Une telle relaxation est établie par le biais du filtre d'information pour l'estimation de la variance conditionnelle du modèle. L'algorithme proposé est utilisé d'une part pour estimer les mesures du kurtosis droit et gauche du log-GARCH. D'autre part, la présence des rendements nuls est traitée par imputation comme un problème de valeurs manquantes, permettant l'estimation du log-GARCH à l'aide d'un algorithme EM. L'ensemble de ces apports est mis en exergue via des illustrations numériques et empiriques.

Ce chapitre a fait l'objet de deux publications :

- Settar, A., Fatmi, N. I., and Badaoui, M. (2021). Numerical estimation of log-GARCH model with relaxed moment assumptions. In *Journal of Physics : Conference Series* (Vol. 1743, No. 1, p. 012036). IOP Publishing.
- Settar, A., and Badaoui, M. (2021, March). On the treatment of zero returns in the estimation of log-GARCH model : Empirical study. In *IOP Conference Series : Materials Science and Engineering* (Vol. 1099, No. 1, p. 012029). IOP Publishing.

Enfin, nous présentons dans la conclusion les apports des travaux de recherche décrits dans ce document ainsi que les résultats des différentes illustrations numériques et empiriques. Nous décrivons quelques limites de ce travail, ainsi que quelques pistes pouvant inspirer des futurs travaux de recherche.

Notons que dans toutes les illustrations numériques présentées dans ce travail, portant sur l'estimation des paramètres des modèles GARCH, CGARCH et log-GARCH, le vecteur des paramètres dans les algorithmes proposés est toujours initialisé par son estimation par la méthode du QML rapportée dans la littérature relativement à chaque modèle étudié.

Dans ce chapitre, nous commençons par donner un aperçu sur la notion d'espérance conditionnelle grandement utilisée dans les processus GARCH ainsi que le filtre de Kalman. Ensuite, nous présentons les définitions et les propriétés fondamentales des modèles GARCH, CGARCH et log-GARCH sujets de la présente thèse. Les méthodes de filtrage optimal sont également discutées, à savoir, le filtre de Kalman et le filtre d'information. Finalement, nous entamons les algorithmes d'approximation stochastique, notamment l'algorithme de perturbation simultanée et l'algorithme fonctionnel lissé.

1.1 Notions d'espérance conditionnelle

Définition 1.1. Soient X une variable aléatoire réelle définie sur un espace probabilisé (Ω, \mathcal{A}) telle que $\mathbb{E}|X| < \infty$ et $\mathcal{B} \subset \mathcal{A}$ une sous-tribu de \mathcal{A} . L'unique variable aléatoire W \mathcal{B} -mesurable vérifiant les conditions :

- $\mathbb{E}|W| < \infty$
- $\mathbb{E}(XZ) = \mathbb{E}(WZ)$, pour toute variable aléatoire Z \mathcal{B} -mesurable et bornée, est appelée l'espérance conditionnelle de X sachant \mathcal{B} et est noté $\mathbb{E}(X|\mathcal{B})$.

Remarque 1.1. Lorsque la tribu \mathcal{B} est engendrée par une variable aléatoire Y , on note $\mathbb{E}(X|\mathcal{B})$ par $\mathbb{E}(X|Y)$.

Proposition 1.1. L'espérance conditionnelle a les propriétés suivantes :

- *Linéarité* : $\mathbb{E}(\lambda X + \mu Y|\mathcal{B}) = \lambda\mathbb{E}(X|\mathcal{B}) + \mu\mathbb{E}(Y|\mathcal{B})$ pour tout $\lambda, \mu \in \mathbb{R}$
- *Positivité* : $X \geq 0$ implique $\mathbb{E}(X|\mathcal{B}) \geq 0$ p.s
- Si X est \mathcal{B} -mesurable alors $\mathbb{E}(X|\mathcal{B}) = X$ p.s

- Si X est indépendante de \mathcal{B} alors $\mathbb{E}(X|\mathcal{B}) = \mathbb{E}X$ p.s
- Si Z est une variable aléatoire \mathcal{B} -mesurable et bornée alors :

$$\mathbb{E}(ZX|\mathcal{B}) = Z\mathbb{E}(X|\mathcal{B}) \text{ p.s}$$
- Soit \mathcal{C} une sous-tribu de \mathcal{B} alors $\mathbb{E}(\mathbb{E}(X|\mathcal{B})|\mathcal{C}) = \mathbb{E}(X|\mathcal{C})$ p.s
- $\mathbb{E}(\mathbb{E}(X|\mathcal{B})) = \mathbb{E}X$ p.s

Définition 1.2. *Étant donnés deux vecteurs aléatoires X et Y dont les espérances existent, on dit que X et Y sont corrélées si et seulement si $\text{Cov}(X, Y) \neq 0$.*

Proposition 1.2. *Soient X et Y deux variables aléatoires dont les espérances existent. Alors, les variables aléatoires $X - \mathbb{E}(X|Y)$ et Y ne sont pas corrélées.*

Définition 1.3 (Estimateur linéaire optimal). *Soient X et Y deux vecteurs aléatoires liés. On appelle l'estimateur linéaire optimal de X en fonction de Y , noté \mathbb{E}^* :*

$$\mathbb{E}^*(X|Y) = A_0Y + B_0$$

minimisant la variance de l'erreur d'estimation :

$$\mathbb{E}(\|X - AY - B\|^2)$$

avec A_0 et B_0 sont des matrices réelles (déterministes) de tailles convenables.

1.2 Modèles autorégressifs conditionnellement hétéroscédastiques

1.2.1 Modèle ARCH

L'économétrie traditionnelle basée sur les modèles ARIMA et la méthodologie de Box-Jenkins, repose principalement sur l'hypothèse d'homoscédasticité des séries chronologiques. Afin de généraliser de telle hypothèse restrictive, Robert F ENGLE (1982) a introduit le processus ARCH (Autoregressive Conditional Heteroskedasticity) permettant à sa variance conditionnelle de varier dans le temps en fonction du passé des observations (hétéroscédasticité conditionnelle) en gardant sa variance inconditionnelle constante. Nous présentons brièvement un aperçu de ce modèle ainsi que ses propriétés de stationnarité et d'existence des moments. Nous développons plus la structure du modèle dans sa forme généralisée GARCH.

Définition 1.4. *Soit ε_t un processus stochastique discret. Soit \mathcal{I}_t l'ensemble d'information jusqu'à la date t , i.e., $\mathcal{I}_{t-1} = \{\varepsilon_s; s < t\}$. ε_t est un ARCH(p) si et seulement si*

$$\varepsilon_t | \mathcal{F}_{t-1} \sim iid(0, \sigma_t^2) \quad (1.1)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 \quad (1.2)$$

avec :

$$\alpha_0 > 0, \quad \alpha_i \geq 0, \quad i = 1, \dots, p$$

Remarque 1.2.

- Si $p = 0$, ε_t est réduit à un bruit blanc.
- Les conditions de positivité des paramètres $\alpha_0, \alpha_1, \dots, \alpha_p$ sont suffisantes pour la positivité de σ_t^2 .

Théorème 1.1 (Stationnarité au second ordre, Robert F Engle (1982)). Soit ε_t un ARCH(p) avec $\alpha_0 > 0$ et $\alpha_i \geq 0$, $i = 1, \dots, p$. ε_t est stationnaire au second ordre si et seulement si

$$\sum_{i=1}^p \alpha_i < 1$$

Comme la distribution de ε_t est symétrique autour de 0, il s'ensuit que $\mathbb{E}\varepsilon_t^{2m+1} = 0$, pour tout $m \in \mathbb{N}$. Les conditions d'existence des moments d'ordre pair sont données par Robert F ENGLE (1982) pour le cas du ARCH(1) et généralisées par MILHØJ (1985) pour tout ordre $p \geq 1$.

Théorème 1.2 (Existence des moments, Robert F Engle (1982)). Soit ε_t un ARCH(1) avec $\alpha_0 > 0$ et $\alpha_1 \geq 0$. Pour tout $m \in \mathbb{N}^*$, $\mathbb{E}\varepsilon_t^{2m}$ existe si et seulement si

$$\alpha_1^m \prod_{j=1}^m (2j - 1) < 1$$

1.2.2 Modèle GARCH

Afin de permettre au modèle ARCH de capter la longue mémoire du processus, des termes autorégressifs de ε_t^2 doivent être ajoutés dans l'équation de la variance conditionnelle. Or, cela conduit à la violation des conditions de positivité des paramètres. Par le modèle GARCH (*Generalized Autoregressive Conditional Heteroskedastic*), BOLLERSLEV (1986) généralise parcimonieusement la spécification ARCH en ajoutant des termes autorégressifs de σ_t^2 dans l'équation de la variance conditionnelle (1.2).

Définition 1.5. ε_t est un GARCH(p, q) si et seulement si

$$\varepsilon_t = \sigma_t \eta_t, \quad \eta_t \sim iid(0, 1) \quad (1.3)$$

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 \quad (1.4)$$

où :

$$\omega > 0, \quad \alpha_i \geq 0, \quad i = 1, \dots, p, \quad \beta_j \geq 0, \quad j = 1, \dots, q$$

Remarque 1.3.

- La condition $\omega > 0$ est nécessaire pour que la variance inconditionnelle de ε_t soit positive, tandis que les conditions de positivité de $\alpha_1, \dots, \alpha_p$ et β_1, \dots, β_q sont suffisantes pour la positivité de σ_t^2 .
- Dans la définition 1.5, l'équation (1.1) a été remplacée par l'équation (1.3) vu que cette dernière permet de générer un processus GARCH à partir du processus $\{\eta_t; t \in \mathbb{Z}\}$ étant une suite de variables aléatoires iid, centrées, réduites, et indépendantes de \mathfrak{J}_{t-1} pour tout $t \in \mathbb{Z}$.

Proposition 1.3 (Représentation ARMA du GARCH(p, q)). Soit ε_t un GARCH(p, q), alors ε_t^2 est un ARMA($\max(p, q), q$) d'innovation linéaire $\nu_t = \varepsilon_t^2 - \sigma_t^2$, soit :

$$\varepsilon_t^2 = \omega + \sum_{i=1}^{\max(p, q)} (\alpha_i + \beta_i) \varepsilon_{t-i}^2 + \nu_t - \sum_{j=1}^q \beta_j \nu_{t-j}. \quad (1.5)$$

avec la convention : $\alpha_i = 0$ (resp. $\beta_j = 0$) si $i > p$ (resp. $j > q$).

De plus ν_t est par définition non autocorrélé et de moyenne nulle.

Stationnarité

La stationnarité au second ordre du processus GARCH a été abordée par BOLLERSLEV (1986) dont la condition est donnée comme suit :

Théorème 1.3. Soit ε_t un GARCH(p, q), alors ε_t est stationnaire au second ordre si et seulement si

$$\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$$

Cependant, il a été empiriquement démontré que certaines séries financières présentent généralement des paramètres qui ne se situent pas dans la région de stationnarité au second ordre (voir Robert F ENGLE et BOLLERSLEV (1986)). Bien que ces séries ne soient pas de carrés intégrables, elles sont strictement stationnaires. C'est bien le cas du modèle IGARCH (*Integrated GARCH*) qui correspond au cas $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j = 1$. En effet, le modèle est strictement stationnaire si la distribution de η_t est à support non borné et sans probabilité en 0 (Corollaire 2.5, FRANCO (2010)). Ainsi, BOUGEROL et PICARD (1992) établissent la condition de stationnarité stricte d'un GARCH(p, q).

Théorème 1.4. Soit ε_t un GARCH(p, q). ε_t est strictement stationnaire si et seulement si $\rho(A_t) < 1$, où :

$$A_t = \begin{pmatrix} \alpha_1 \eta_t^2 & \cdots & \alpha_p \eta_t^2 & \beta_1 \eta_t^2 & \cdots & \beta_q \eta_t^2 \\ & I_{p-1} & 0_{(p-1,1)} & & 0_{(p-1,q)} & \\ \alpha_1 & \cdots & \alpha_p & \beta_1 & \cdots & \beta_q \\ & 0_{(q-1,p)} & & & I_{q-1} & 0_{(q-1,1)} \end{pmatrix}$$

De plus, ε_t est la seule solution strictement stationnaire du modèle 1.5.

Remarque 1.4.

- La stationnarité au second ordre implique la stationnarité stricte mais le contraire n'est pas toujours vrai (pour plus de détails voir Chapitre 2 dans FRANCQ (2010)). En conséquence, la condition de stationnarité au second ordre reste pratiquement plus facile à vérifier.
- La condition suffisante du théorème précédant reste valable même en relaxant les conditions de positivité de $\alpha_1, \dots, \alpha_p$ et β_1, \dots, β_q . Ce qui n'est pas le cas pour la stationnarité au second ordre. (voir la démonstration du théorème 1 dans BOLLERSLEV (1986))

Théorème 1.5 (Représentation ARCH(∞) du GARCH(p, q)). Soit ε_t un GARCH(p, q) avec $\omega > 0$ et à priori sans aucune contrainte de positivité sur les paramètres $\alpha_1, \dots, \alpha_p$ et β_1, \dots, β_q . Si toutes les racines du polynôme $\beta(z) = 1 - \beta_1 z - \dots - \beta_q z^q$ sont à l'extérieur du cercle unitaire, alors ε_t admet une représentation ARCH(∞) donnée par :

$$\sigma_t^2 = \omega^* + \Psi(L)\varepsilon_t^2 \tag{1.6}$$

où :

$$\omega^* = \frac{\omega}{\beta(1)} \text{ et } \Psi(L) = \frac{\alpha(L)}{\beta(L)}$$

avec :

$$\alpha(L) = \sum_{i=1}^p \alpha_i L^i$$

Par le biais de la représentation ARCH(∞), NELSON et CAO (1992) sont parvenus à relaxer les conditions de positivité des paramètres $\alpha_1, \dots, \alpha_p$ et β_1, \dots, β_p assurant la positivité de σ_t^2 , notamment pour les GARCH($p, 1$) et GARCH($p, 2$). En effet, réécrivons (1.6) comme suit :

$$\sigma_t^2 = \omega^* + \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k-1}^2$$

Étant donné que $\omega^* > 0$ car sinon $\beta(1) \leq 0$, ce qui impliquerait l'existence d'une racine à l'intérieur du cercle unitaire puisque $\beta(0) = 1$, il ne reste qu'à dériver des contraintes sur les coefficients $\psi_0, \psi_2, \dots, \psi_k, \dots$ pour que σ_t^2 reste strictement positive, ce qui est donné par les théorèmes suivants :

Théorème 1.6 (GARCH($p, 1$), Nelson et Cao (1992)). Supposant que les polynômes $\alpha(z)$ et $\beta(z)$ n'ont pas de racines communes et que les racines du dernier sont à l'extérieur du cercle unitaire. $\omega^* \geq 0$ et $\psi_k \geq 0$ pour tout entier k , si et seulement si $\omega \geq 0$, $\beta_1 \geq 0$ et $\psi_k = \sum_{j=0}^k \alpha_{j+1} \beta_1^{k-j} \geq 0$.

Théorème 1.7 (GARCH($p, 2$), Nelson et Cao (1992)). *Supposant que les polynômes $\alpha(z)$ et $\beta(z)$ n'ont pas de racines communes et que les racines du dernier sont à l'extérieur du cercle unitaire. $\omega^* \geq 0$ et $\psi_k \geq 0$ pour tout $k \geq 0$, si et seulement si*

$$\begin{aligned} \omega/(1 - \Delta_1 - \Delta_2 + \Delta_1\Delta_2) &\geq 0 \\ (\Delta_1, \Delta_2) &\in \mathbb{R}^{+*} \times \mathbb{R} \\ \sum_{j=0}^{p-1} \alpha_{j+1}/\Delta_1^j &> 0 \\ \psi_k &\geq 0, k = 0, \dots, p \end{aligned}$$

où Δ_1 et Δ_2 sont les racines du polynôme $(1 - \beta_1/z - \beta_2/z^2)$.

Théorème 1.8 (GARCH(p, q), $q \geq 3$ Tsai et Chan (2008)). *Supposons que les racines des polynômes $\alpha(z) = \sum_{i=1}^p \alpha_i z^i$ et $\beta(z) = 1 - \sum_{i=1}^q \beta_i z^i$ n'ont pas de racines communes et les racine de β sont à l'extérieur du cercle unitaire. $\omega^* \geq 0$ et $\psi_k \geq 0$ pour tout $k \geq 0$, si et seulement si*

$$\begin{aligned} \Delta_1 &> 1 \\ \alpha(\Delta_1) &> 0 \\ \psi_k &\geq 0, k = 1, \dots, k^* \end{aligned}$$

où

$$k^* = \lfloor \max(0, \frac{\log r_1 - \log((q-1)r^*)}{\log|\Delta_1| - \log|\Delta_2|}) \rfloor$$

$$\text{avec } r^* = \max_{2 \leq j \leq q} |r_j| \text{ et } r_j = \frac{\alpha(|\Delta_j|)}{\frac{d}{dz}(1 - \beta(z))(|\Delta_j|)}, 1 \leq j \leq q$$

Remarque 1.5.

Outre que la représentation (1.6), une autre représentation autorégressive infinie de σ_t^2 , supposée strictement stationnaire, est donnée par l'existence de ω' et des réels $\phi_1, \dots, \phi_k, \dots$ tel que

$$\sigma_t^2 = \omega' + \sum_{k=1}^{\infty} \phi_k \eta_{t-k}^2 \sigma_{t-k}^2 \tag{1.7}$$

Existence des moments

La condition nécessaire et suffisante d'existence du moment d'ordre 4 du GARCH(p, q) a été formulée par KARANASOS (1999) et HE et TERÄSVIRTA (1999a). LING (1999) et LING et MCALEER (2002) généralisent la condition d'existence des moments du GARCH(p, q) pour tout ordre pair, comme le montre le théorème suivant :

Théorème 1.9 (Ling (1999)). *Soit ε_t un GARCH(p, q). Pour tout $m \in \mathbb{N}^*$, $\mathbb{E}\varepsilon_t^{2m}$ existe si et seulement si $\rho(\mathbb{E}A_t^{\otimes m}) < 1$.*

Fonction d'autocorrélation du carré du GARCH

Un des faits stylisés qui caractérisent les séries financières est le fait que les carrés des rendements sont autocorrélés, tandis que les rendements ne le sont pas (MANDELBROT (1963)). Grâce à sa représentation ARMA (1.5), le modèle GARCH est bien adapté à la prise en compte de cette propriété empirique.

La fonction d'autocorrélation du carré d'un GARCH(p, q) a été donnée par BOLLERSLEV (1986) par analogie aux équations de Yule-Walker sous l'existence du moment d'ordre 4, soit :

$$\rho_k = \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) \rho_{k-i}, \quad k \geq q + 1$$

Dans le cas du GARCH(1,1), BOLLERSLEV (1988) dérive la fonction d'autocorrélation des carrés comme suit :

$$\rho_k = \left(\alpha_1 + \frac{\alpha_1^2 \beta_1}{1 - 2\alpha_1 \beta_1 - \beta_1^2} \right) (\alpha_1 + \beta_1)^{k-1} \quad (1.8)$$

DING et GRANGER (1996) propose une approximation de ρ_k telle que :

$$\rho_k \approx \left(\alpha_1 + \frac{1}{3} \beta_1 \right) (\alpha_1 + \beta_1)^{k-1}$$

On trouve par ailleurs dans KARANASOS (1999) et HE et TERÄSVIRTA (1999a) des formulations récursives permettant de calculer la fonction d'autocorrélation du carré du GARCH(p, q), qui a été quant à elle sujet d'analyse faite par HE et TERÄSVIRTA (1999b) sous les conditions relaxées de NELSON et CAO (1992).

Kurtosis

Les distributions de probabilité empiriques des séries des rendements financiers ne correspondent pas généralement à une distribution gaussienne, elles sont plutôt leptokurtiques. Une mesure standard de la leptokurticité est le kurtosis donné par le rapport du moment d'ordre 4 et le carré du moment d'ordre 2, soit :

$$K = \frac{\mathbb{E}\varepsilon_t^4}{(\mathbb{E}\varepsilon_t^2)^2}$$

Une valeur de référence de K est asymptotiquement égale à 3 dans le cas d'une distribution normale.

En particulier, pour un GARCH normal ($\eta_t \sim iidN(0, 1)$), étant donnée que η_t est indépendante de σ_t , et est de variance 3, on trouve que :

$$\mathbb{E}(\varepsilon_t^4 / \mathcal{I}_{t-1}) = 3\mathbb{E}(\varepsilon_t^2 / \mathcal{I}_{t-1})^2$$

En prenant l'espérance mathématique de chaque coté, on obtient :

$$\begin{aligned}\mathbb{E}(\varepsilon_t^4) &= \mathbb{E}\mathbb{E}(\varepsilon_t^4/\mathfrak{I}_{t-1}) \\ &= 3\mathbb{E}[\mathbb{E}(\varepsilon_t^2/\mathfrak{I}_{t-1})]^2 \\ &\geq 3[\mathbb{E}\mathbb{E}(\varepsilon_t^2/\mathfrak{I}_{t-1})]^2 \\ &= 3[\mathbb{E}(\varepsilon_t^2)]^2\end{aligned}$$

D'où, un processus GARCH présente toujours un excé de kurtosis. On en déduit alors que le kurtosis d'un GARCH est, par construction de ce dernier, une mesure naturelle, voir adéquate de l'hétéroscédasticité conditionnelle.

Estimation des paramètres du GARCH par QML

Nous présentons dans ce qui suit la procédure d'estimation des paramètres du GARCH(p, q) par quasi-maximum de vraisemblance dans sa forme analytique (basée sur le calcul des dérivés) telle qu'elle est rapportée dans la littérature. En effet, dans son article fondateur, BOLLERSLEV (1986) utilise la méthode du maximum de vraisemblance (MLE) pour estimer les paramètres du GARCH(p, q) gaussien. Cependant, de nombreux exercices empiriques ont montré que les séries étudiées ont des distributions asymétriques de courbes leptokurtiques (plus lourde que la loi normale) (DIEBOLD (2012)), ce qui a remis en question la justification de l'hypothèse de la normalité du GARCH. Par conséquent, la MLE utilisant la distribution de Student ou des fonctions de vraisemblance gaussienne généralisée a été introduite dans les travaux de Robert F ENGLE et BOLLERSLEV (1986), BOLLERSLEV (1987), HSIEH (1989) et NELSON (1991). Toutefois, ces méthodes peuvent conduire à des estimations inconsistantes des paramètres du modèle si la distribution de l'innovation est mal spécifiée. Cependant, même si la vraie distribution de l'innovation du GARCH est loin d'être normale, la méthode MLE considérée comme un quasi-maximum de vraisemblance (QML) gaussien peut garantir la consistance et la normalité asymptotique des estimations des paramètres du GARCH dont l'innovation est de moment d'ordre 4 fini (Christian FRANCO et ZAKOIAN (2004); HALL et YAO (2003); HORV et KOKOSZKA (2003). Pourtant, il convient de noter que la théorie asymptotique de l'estimation des paramètres du GARCH avait été abordée par WEISS (1986) pour le modèle ARCH et par S.-W. LEE et HANSEN (1994) et LUMSDAINE (1996) pour le GARCH(1,1) et généralisée pour le GARCH(p, q) sous des conditions plus fortes par BOLLERSLEV et WOOLDRIDGE (1992). Un cadre plus robuste de l'estimation du modèle GARCH dans le cas où la distribution de l'innovation η_t est inconnue est élaboré par Christian FRANCO, LEPAGE et ZAKOIAN (2011), FAN, QI et XIU (2014) et ZHU et XIE (2016).

Quasi-vraisemblance conditionnelle

Soient $\varepsilon_1, \dots, \varepsilon_n$ n réalisations d'un processus GARCH(p, q) strictement stationnaire, d'innovation η_t , *iid*, centrée et réduite. Les paramètres du modèle sont supposés positifs constituant un vecteur θ , soit :

$$\theta = {}^t(\omega, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q)$$

On considère la quasi-vraisemblance gaussienne évaluée à partir d'une loi normale centrée réduite pour η_t , soit :

$$L_n(\theta) = \prod_{t=1}^n \frac{1}{\sqrt{2\pi\widehat{\sigma}_t^2}} \exp\left(-\frac{\varepsilon_t^2}{2\widehat{\sigma}_t^2}\right)$$

Par un choix convenable des valeurs initiales $\varepsilon_0, \dots, \varepsilon_{1-p}, \sigma_0^2, \dots, \sigma_{1-q}^2$, les valeurs de $\widehat{\sigma}_t^2$ sont données récursivement par :

$$\widehat{\sigma}_t^2 = \omega + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \widehat{\sigma}_{t-j}^2, \quad t = 1, \dots, n$$

Un estimateur du QML de θ est défini comme toute quantité $\widehat{\theta}_n$ vérifiant presque sûrement

$$L_n(\widehat{\theta}_n) = \sup_{\theta} L_n(\theta)$$

En prenant le logarithme, maximiser la vraisemblance revient à minimiser par rapport à θ le critère suivant :

$$\widehat{l}_n(\theta) = \frac{1}{n} \sum_{t=1}^n \frac{\varepsilon_t^2}{2\widehat{\sigma}_t^2} + \log(\widehat{\sigma}_t^2)$$

Équations de vraisemblance

On obtient les équations de vraisemblance en annulant la dérivée par rapport à θ du critère $\widehat{l}_n(\theta)$. Ce qui donne :

$$\frac{1}{n} \sum_{t=1}^n (\varepsilon_t^2 - \widehat{\sigma}_t^2) \frac{1}{\widehat{\sigma}_t^4} \frac{\partial \widehat{\sigma}_t^2}{\partial \theta} = 0$$

À ce niveau, un algorithme itératif (e.g., Newton-Raphson, Descente du gradient) peut être utilisé pour trouver une estimation $\widehat{\theta}_n$.

Notons également que l'étude asymptotique de cet estimateur concernant sa consistance et sa normalité asymptotique, est abordée par FRANCO (2010) à travers les théorèmes 3.3 et 3.4.

1.2.3 Modèle CGARCH (Component GARCH)

Processus à mémoire courte - à mémoire longue

Selon la théorie de l'analyse des séries chronologiques, la suite des autocorrélations (ou des autocovariances) d'un processus stochastique contient toute l'information sur sa mémoire et la résume entièrement. Nous distinguons trois situations :

- Si $\rho_k = 0$, pour tout k , cela signifie qu’il n’existe aucun lien entre les valeurs du processus à deux dates distantes de k . Le processus est dit **sans mémoire** (e.g., bruit blanc).
- Si ρ_k a tendance à décroître d’une façon exponentielle (donc rapidement) ou qu’elle tend vers 0 à vitesse géométrique, le processus est dit à **mémoire courte** (e.g., ARIMA).
- Si ρ_k a tendance à décroître d’une façon hyperbolique (donc lentement) ou qu’elle tend vers 0 lentement, le processus est dit à **mémoire longue**.

Ainsi, la présence de longue mémoire entraîne une corrélation à long terme entre les observations actuelles du processus et ses valeurs futures, et donc chaque observation porte la mémoire des observations qui l’ont précédé, ce qui crée plusieurs tendances dans l’évolution du processus. Tel est le phénomène de persistance.

Comme le montre sa fonction d’autocorrélation (1.8), le carré du GARCH(1,1) décroît d’une façon exponentielle, ce qui rend le modèle incapable de capter la propriété de mémoire longue éventuellement présente dans les séries étudiées. Effectivement, DING et GRANGER (1996) montrent la présence d’une telle propriété dans la série des rendements de l’indice S&P500 en valeur absolue, vu qu’elle décroît très lentement et reste significative même avec de longs retards. Les travaux de ANDERSEN et BOLLERSLEV (1997); ANDERSEN, BOLLERSLEV et al. (2001); BOLLERSLEV et WRIGHT (2000); HWANG (2001) sont parmi d’autres qui mettaient empiriquement en évidence la présence de la propriété de mémoire longue. DING et GRANGER (1996) considèrent que ce comportement d’autocorrélation empirique révèle l’existence de différentes composantes de la volatilité dominant différentes périodes. Certaines composantes de la volatilité peuvent avoir un effet très significatif à court terme, mais chutent très rapidement. D’autres peuvent avoir un effet à court terme relativement plus faible, mais elles durent pendant une longue période. Dans ce sens, DING et GRANGER (1996) introduit le modèle CGARCH (*Component GARCH*) comme une nouvelle spécification de la volatilité en la décomposant à plusieurs spécifications-composantes permettant à certaines de capter la dynamique à long terme de la volatilité, et aux autres de capter ses fluctuations à court terme.

Définition 1.6. Soit η_t une suite de variables aléatoires iid, centrées et de variance unité et $N \geq 1$. On dit que ε_t est un processus CGARCH(N) au sens de Ding et Granger s’il vérifie :

$$\begin{aligned}\varepsilon_t &= \sigma_t \eta_t \\ \sigma_t^2 &= \sum_{i=1}^N w_i \sigma_{i,t}^2, \quad \sum_{i=1}^N w_i = 1 \\ \sigma_{i,t}^2 &= \omega_i + \alpha_i \varepsilon_{t-1}^2 + \beta_i \sigma_{i,t-1}^2, \quad i = 1, \dots, N\end{aligned}$$

où $\omega_i \geq 0$, $\alpha_i \geq 0$ et $\beta_i \geq 0$, $i = 1, \dots, N$.

Remarque 1.6.

Une autre spécification du CGARCH a été proposée par Robert F ENGLE et G. LEE (1999) où la variance conditionnelle est décomposée en une composante permanente q_t et une composante transitoire $\sigma_t^2 - q_t$ afin de capter les mouvements de la volatilité à court et à long terme, soit :

$$\begin{aligned}\sigma_t^2 &= q_t + \sum_{i=1}^p \alpha_i (\varepsilon_{t-i}^2 - q_{t-i}) + \sum_{j=1}^q (\sigma_{t-j}^2 - q_{t-j}) \\ q_t &= \omega + \rho q_{t-1} + \phi (\varepsilon_{t-1}^2 - q_{t-1})\end{aligned}$$

Toutes les propriétés statistiques du GARCH discutées précédemment restent valables pour le CGARCH car en effet, tout CGARCH(N) composé de N GARCH(1,1), s'exprime comme un GARCH(N, N) restreint¹. Dans ce sens, KARANASOS (1999) est parvenu à expliciter la spécification GARCH(N, N) d'un CGARCH(N) composé d'un IGARCH(1,1) et ($N - 1$) GARCH(1,1).

Lemme 1.1 (Karanasos (1999)). Soit ε_t un CGARCH(N) avec $\omega_i = 0$ pour $i \geq 2$. Le GARCH(N, N) correspondant est de variance conditionnelle donnée par :

$$\begin{aligned}\sigma_t^2 &= \sum_{i=1}^N w_i \omega_i \varepsilon_{t-1}^2 + \sum_{i=1}^N \sum_{l=1}^N \left[\prod_{j=1}^{l-1} \left(\sum_{i_j=i_{j-1}+1, i_j \neq i}^{N-l+1+j} \right) \prod_{j=1}^{l-1} \beta_{i_j} (-1)^{l+1} w_i \alpha_i \varepsilon_{t-l}^2 \right] \\ &\quad + \sum_{l=1}^N \left[\prod_{j=1}^l \left(\sum_{i_j=i_{j-1}+1}^{N-l+j} \right) \prod_{j=1}^l \beta_{i_j} (-1)^{l+1} \sigma_{t-1}^2 \right] + w_1 \omega_1 \prod_{i=2}^N (1 - \beta_i)\end{aligned}$$

Pour des ordres modérés du CGARCH ($N \leq 3$), la relaxation des contraintes de positivité de σ_t^2 peut se faire en relaxant les contraintes de positivité des coefficients de l'équation de la variance conditionnelle donnée par le lemme 1.1 au sens de Nelson et Cao (voir Robert F ENGLE et G. LEE (1999) pour $N = 2$).

1.2.4 Modèle log-GARCH

Le modèle log-GARCH est l'une des extensions du GARCH, introduit par GEWEKE (1986), PANTULA (1986) et MILHØJ (1987) qui remédie au problème de positivité de la variance conditionnelle en modélisant son logarithme en fonction du passé des logarithmes des carrés des observations de sorte que les paramètres ne soient a priori soumis à aucune condition de positivité.

Définition 1.7. On dit que ε_t est un processus log-GARCH(p, q) (symétrique) s'il vérifie :

1. Bien que les paramètres du GARCH(N, N) soient négatifs, sa variance conditionnelle reste positif à condition que les paramètres de chaque composante GARCH(1,1) soient positifs.

$$\begin{aligned}\varepsilon_t &= \sigma_t \eta_t, \quad \eta_t \sim iid(0, 1) \\ \log \sigma_t^2 &= \omega + \sum_{i=1}^p \alpha_i \log \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \log \sigma_{t-j}^2\end{aligned}$$

où η_t est une suite de variables aléatoires iid, centrées, de variance unité et indépendantes de σ_t et $\omega, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q$ sont des réels.

Par construction, le modèle log-GARCH présente plusieurs avantages. En effet, grâce à l'absence des conditions de positivité de sa variance conditionnelle, la volatilité du modèle n'est pas minorée et par conséquent, elle maintient la flexibilité de sa dynamique contrairement au modèle GARCH où la valeur minimale de la volatilité est déterminée par le paramètres ω donné dans (1.4). À cela s'ajoute la stabilité du modèle relativement aux transformations en puissances de σ_t , à savoir celles de la forme $|\sigma_t|^d$ qui généralisent la spécification GARCH ($d = 2$). Les modèles associés à ces transformations sont incompatibles (e.g., le GARCH et le TGARCH ($d = 1$)). Au contraire, toute transformation de puissance de σ_t suivant un log-GARCH suit également un log-GARCH de mêmes paramètres sauf ω que l'on multiplie par $d/2$. De plus, le log-GARCH est inversible sous la condition d'inversibilité du polynôme $\beta(z) = 1 - \sum_{j=1}^q \beta_j z^j$ tel que :

$$\log \sigma_t^2 = \frac{\omega}{\beta(1)} + \sum_{i=1}^{\infty} c_i \log \varepsilon_{t-i}^2$$

avec :

$$\sum_{k=1}^{\infty} c_k z^k = \frac{\alpha(z)}{\beta(z)}, \quad |z| \leq 1$$

où :

$$\alpha(z) = \sum_{i=1}^p \alpha_i z^i$$

Stationnarité

Théorème 1.10 (Christian Francq, Wintenberger et Zakoian (2013a)). *Supposons que $\mathbb{E} \log^+ |\log \eta_0^2| < \infty$. Le log-GARCH(p, q) est strictement stationnaire si et seulement si toutes les racines du polynôme $1 - \sum_{i=1}^{\max(p, q)} (\alpha_i + \beta_i) z^i$ sont à l'extérieur du cercle unitaire.*

En particulier, dans le cas $p = q = 1$, le log-GARCH est strictement stationnaire si

$$|\alpha_1 + \beta_1| < 1 \tag{1.9}$$

De plus :

$$\mathbb{E} \log \sigma_t^2 = \frac{\omega + \alpha_1 \mathbb{E} \log \eta_t^2}{1 - \alpha_1 - \beta_1}$$

Il est à noter que la condition (1.9) est suffisante mais non nécessaire, car en effet, si $\alpha_1 + \beta_1 \neq 1$ et $\eta_0^2 = 1$ p.s, alors $\varepsilon_t = \exp\left(\frac{\omega}{2(1 - \alpha_1 - \beta_1)}\right) \eta_t$ est une solution strictement stationnaire du log-GARCH.

Existence des moments

Réécrivons le log-GARCH sous forme polynomiale telle que :

$$\beta(L)\log \sigma_t^2 = \omega + \alpha(L)\log \eta_t^2$$

où :

$$\beta(z) = 1 - \sum_{i=1}^r (\alpha_i + \beta_i) z^i \text{ et } \alpha(z) = \sum_{i=1}^r \alpha_i z^i$$

avec $r = \max(p, q)$, $\alpha_i = 0$ pour $i > p$ et $\beta_i = 0$ pour $i > q$.

En supposant que toutes les racines de β sont à l'extérieur du cercle unitaire, on obtient :

$$\log \sigma_t^2 = c + \sum_{i=1}^{\infty} \phi_i \log \eta_{t-i}^2$$

avec $c = \frac{\omega}{\beta(1)}$ et $\sum_{i=1}^{\infty} \phi_i z^i = \beta^{-1}(z)\alpha(z)$.

Théorème 1.11 (Christian Francq et Zakoian (2019)). Soient ε_t un log-GARCH et $\underline{\phi} = \min_i \phi_i$.

Supposons que $\eta_t \sim iidN(0, 1)$, alors :

- Si $\underline{\phi} < 0$, alors $\mathbb{E}|\varepsilon_t|^d < \infty$ si et seulement si $d < -\frac{1}{2\underline{\phi}}$
- Si $\underline{\phi} \geq 0$, alors $\mathbb{E}|\varepsilon_t|^d < \infty$ pour tout d .

Notons que le théorème ci-dessus est valable pour toute distribution de η_t admettant un moment fini d'ordre d . Cependant, quant au calcul explicite des moments de ε_t , le théorème est restreint aux distributions de η_t dont les moments sont explicitement connus. Un cadre plus général des log-GARCH symétriques et asymétriques est proposé par Christian FRANCQ, WINTENBERGER et ZAKOIAN (2013a) (Proposition 2.2 et 2.3) fournissant des conditions d'existence des log-moments $\mathbb{E}|\log \varepsilon_t^2|^m$, $m \in \mathbb{N}^*$ sous des conditions moins faibles sur η_t .

1.3 Filtrage optimal

1.3.1 Système dynamique

Définition 1.8. Un système dynamique est un modèle permettant de décrire l'évolution au cours du temps d'un ensemble d'objets en interaction dont l'état courant

dépend juste de l'état précédent, des commandes appliquées et d'un processus aléatoire.

Les équations décrivant un système dynamique constituent un modèle, dit espace d'état, associé à une suite d'observations bruitées $(Y_t)_{t \geq 0}$.

Définition 1.9. Un modèle (représentation) espace d'état est défini par le système des équations suivant :

$$\begin{cases} X_{t+1} = A_t X_t + W_t \\ Y_t = H_t X_t + V_t \end{cases} \quad (t \geq 0) \quad (1.10)$$

où :

- $X_t \in \mathbb{R}^r$ est le vecteur d'état du processus tel que l'état initial X_0 du processus est d'espérance et de variance connues, et décorrélé avec W_t et V_t pour $t \geq 0$;
- $A_t \in \mathcal{M}_{(r,r)}$ est la matrice déterministe de transition de l'état qui décrit la dynamique de X_t ;
- $W_t \in \mathbb{R}^r$ est un bruit blanc qui modélise l'erreur du processus, de moyenne connue et de matrice de covariance Q connue (semi-définie positive) ;
- $Y_t \in \mathbb{R}^m$ est le vecteur d'observation à l'instant t ;
- $H_t \in \mathcal{M}_{(m,r)}$ est la matrice déterministe de mesure ;
- $V_t \in \mathbb{R}^m$ est un bruit blanc qui modélise l'erreur d'observation, de moyenne connue et de matrice de covariance R connue (définie positive) et non corrélée avec W_t pour $t \geq 0$.

L'objectif sera de déterminer une estimation optimale et récursive des états X_t à partir des mesures $(Y_t)_{t \geq 1}$. Une telle estimation peut être classée selon la quantité d'information disponible $\{Y_1, \dots, Y_\tau\}$. En effet, on peut chercher à estimer l'état du système dynamique X_t à un instant donné t suivant la valeur de τ , soit :

- Si $t < \tau$, il s'agit d'un problème de lissage (*Smoothing*) ;
- Si $t = \tau$, il s'agit d'un problème de filtrage (*Filtering*) ;
- Si $t > \tau$, il s'agit d'un problème de prédiction (*Prediction*).

1.3.2 Filtre de Kalman

Le filtre de Kalman (KALMAN (1960)) est un filtre optimal pour l'estimation de l'état du système. Il garantit que l'erreur de l'estimation est en moyenne nulle et de variance minimale. Il existe une littérature grandement abondante traitant la théorie du filtre de Kalman, on en cite en particulier ANDERSON et MOORE (2012), D. SIMON (2006) et A. C. HARVEY (1990). Ainsi, le filtre de Kalman fournit, après initialisation, des estimations prédites et filtrées en conjuguant récursivement les étapes suivantes :

Initialisation : À l'instant $t = 0$, nous n'avons aucune mesure disponible. La première mesure arrive à l'instant $t = 1$. X_0 , l'état du système à l'instant $t = 0$ est une variable aléatoire, il paraît logique en l'absence de mesure de l'estimer par son espérance. Nous avons alors :

$$\widehat{X}_0 = \mathbb{E}X_0$$

La matrice de covariance de l'erreur est donnée comme :

$$P_0 = \mathbb{E}(X_0 - \widehat{X}_0)^t (X_0 - \widehat{X}_0)$$

Notez que si nous connaissons parfaitement l'état de départ du système, nous prenons $\widehat{X}_0 = X_0$ et dans ce cas P_0 est nulle.

Étape de prédiction : On se place à l'instant t . À cet instant, on dispose d'une estimation initiale fondée sur la connaissance du processus et des observations jusqu'à l'instant précédent $t - 1$. On cherche un estimateur $\widehat{X}_{t|t-1}$ de X_t sans tenir compte de la mesure Y_t , mais en utilisant uniquement l'équation d'état du système (1.10). On obtient donc :

$$\widehat{X}_{t|t-1} = A_{t-1} \widehat{X}_{t-1|t-1}$$

On définit l'erreur de cette estimation par :

$$e_{t|t-1} = X_t - \widehat{X}_{t|t-1}$$

ainsi que la matrice de covariance de l'erreur :

$$P_{t|t-1} = \mathbb{E}(e_{t|t-1}^t e_{t|t-1}) = A_t P_{t-1|t-1}^t A_t + Q$$

Étape de filtrage : Nous allons maintenant utiliser l'observation Y_t pour améliorer l'estimation $\widehat{X}_{t|t-1}$ et obtenir un nouvel estimateur $\widehat{X}_{t|t}$ de X_t , soit :

$$\widehat{X}_{t|t} = \widehat{X}_{t|t-1} + K_t(Y_t - \widehat{Y}_{t|t-1})$$

où $K_t = P_{t|t-1}^t H_t (H_t P_{t|t-1}^t H_t + R)^{-1}$ est appelé le gain du filtre de Kalman.

Comme les observations Y_0, \dots, Y_{t-1} ne dépendent linéairement que de X_0 et V_0, \dots, V_{t-1} , on a donc :

$$\widehat{Y}_{t|t-1} = \mathbb{E}^*(Y_t | Y_0, \dots, Y_{t-1}) = \mathbb{E}^*(H_t X_t + V_t | Y_0, \dots, Y_{t-1}) = H_t \widehat{X}_{t|t-1}$$

Ce qui implique :

$$\widehat{X}_{t|t} = \widehat{X}_{t|t-1} + K_t(Y_t - H_t \widehat{X}_{t|t-1})$$

D'où se définit l'erreur de cette estimation comme :

$$e_{t|t} = X_t - \widehat{X}_{t|t}$$

ainsi que la matrice de covariance de l'erreur :

$$P_{t|t} = \mathbb{E}(e_{t|t} {}^t e_{t|t})$$

Remarque 1.7.

Le gain du filtre de Kalman K_t tient compte des caractéristiques statistiques du bruit des mesures mais ne dépend pas des observations, donc il peut être calculé avant l'étape du filtrage. De plus, K_t existe toujours car la matrice R est définie positive.

Résumons l'algorithme du filtre de Kalman comme suit :

$$\begin{aligned} \widehat{X}_{0|0} &= \mathbb{E}X_0 \\ P_{0|0} &= P_0 \\ \text{Pour } t &= 1, 2, \dots, n \text{ faire} \\ \widehat{X}_{t|t-1} &= A_{t-1} \widehat{X}_{t-1|t-1} \\ P_{t|t-1} &= A_t P_{t-1|t-1} {}^t A_t + Q \\ K_t &= P_{t|t-1} {}^t H_t (H_t P_{t|t-1} {}^t H_t + R)^{-1} \\ \widehat{X}_{t|t} &= \widehat{X}_{t|t-1} + K_t (Y_t - H_t \widehat{X}_{t|t-1}) \\ P_{t|t} &= (I - K_t H_t) P_{t|t-1} \end{aligned} \tag{1.11}$$

Fin pour

Remarque 1.8.

- Sous l'hypothèse de normalité des bruits du modèle (1.10), la distribution conditionnelle de $(X_t | Y_1, \dots, Y_t)$ est normale et sa moyenne fournit donc un estimateur optimal (au sens L^2) de X_t . Or, lorsque l'hypothèse de normalité est en défaut, l'estimateur obtenu n'est plus optimal mais seulement optimal parmi les estimateurs linéaires.
- Lorsque le modèle (1.10) est à bruits non centrés mais dont les moyennes sont exactement connues, le filtre de Kalman maintient sa bonne performance en intégrant les moyennes des bruits dans les équations de mise à jour du filtre.

La connaissance des propriétés de stabilité du filtre de Kalman associées à un modèle espace d'état permet de les intégrer dans les algorithmes d'estimation comme des conditions contrôlant la convergence des équations du filtre. Ces propriétés ont été étudiées par ANDERSON et MOORE (2012) et ont été rediscutées par A. C. HARVEY (1990). Dans la suite, nous nous intéressons au modèle (1.10) dans sa forme invariante dans le temps, i.e., les matrices A_t et H_t sont indépendantes du temps. Ainsi on note pour tout $t \geq 0$, $A_t = A$ et $H_t = H$.

Définition 1.10 (Stabilité). *Le modèle (1.10) est dit stable si pour toute valeur initiale de X_t , X_0 , X_t converge vers une solution d'équilibre \bar{X} quand W_t est constante.*

Proposition 1.4. *Le modèle (1.10) est stable si et seulement si toutes les valeurs propres de A sont en module strictement inférieurs à 1.*

Définition 1.11 (État stationnaire). *Le filtre de Kalman est dit en état stationnaire si $P_{t|t-1}$ est invariante dans le temps, i.e., $P_{t|t-1} = \bar{P}$.*

Proposition 1.5. *Si le modèle (1.10) est stable, et $P_{0|-1}$ est semi-définie positive, alors :*

$$P_{t|t-1} \rightarrow \bar{P} \quad \text{quant } t \rightarrow \infty$$

Corollaire 1.1. *La variable d'état X_t est stationnaire au second ordre si et seulement si le modèle (1.10) est stable.*

1.3.3 Filtre de Kalman sous contrainte

Le filtre de Kalman décrit par le système (1.11) n'intègre pas l'information connue sur la variable d'état. Souvent, cette information prend via une matrice connue D et un vecteur connu d , la forme d'une égalité $DX = d$ ou d'une inégalité $DX \leq d$. Ces contraintes peuvent être incorporées de plusieurs façons dans les équations du filtre.

Une première approche est celle de la réduction du modèle (*Model reduction*). Elle consiste à réduire le modèle espace d'état en reformulant ses équations en fonction de la contrainte d'égalité sur X_t (WEN et DURRANT-WHYTE (1992)). Cependant, cette approche n'est pas applicable pour les contraintes d'inégalité et ne maintient pas la forme du modèle espace d'état, notamment l'interprétation de ses équations.

Une deuxième approche est celle des mesures parfaites (*Perfect measurements*). Elle consiste à considérer les valeurs de X_t vérifiant une contrainte d'égalité comme des observations sans bruits qui s'ajoutent au vecteur des mesures Y_t dans le modèle espace d'état (HAYWARD (1998); PORRILL (1988)). L'utilisation de cette méthode n'est directement applicable qu'aux contraintes d'égalité et indirectement aux contraintes d'inégalité en ajoutant des bruits de mesure non nuls aux mesures de X_t , ce qui rend difficile le contrôle de la convergence de l'estimation de la variable d'état (MAHATA et SÖDERSTRÖM (2004)).

Une troisième approche est celle de projection (*Projection*). Elle consiste à maximiser par rapport aux valeurs de X_t , la densité de probabilité de ce dernier conditionnellement aux mesures Y_1, \dots, Y_t sous la contrainte d'égalité ou d'inégalité sur X_t (T.-L. CHIA (1985); D. SIMON et T. L. CHIA (2002)). Ainsi, la méthode est convertie en un problème de maximisation sous contrainte. Bien qu'elle soit applicable aussi pour les contraintes d'égalité que d'inégalité, cette approche peut augmenter la complexité des algorithmes d'estimation où le filtre de Kalman intervient avec d'autres processus d'optimisation, e.g., la maximisation de la vraisemblance évaluée en fonction d'un processus estimé par le filtre de Kalman sous contrainte.

Méthode de troncature de la densité de probabilité

Cette méthode consiste à appliquer la contrainte d'inégalité sur X_t en tronquant sur l'intervalle de la contrainte, la densité de probabilité (supposée normale) de son estimation (filtrée ou prédite) notée \widehat{X}_t , de matrice d'erreur P_t , obtenues par le filtre de Kalman, de manière à ce que sa moyenne, notée \widetilde{X}_t , devient la nouvelle valeur contrainte de X_t (SHIMADA et al. (1998); D. SIMON (2006); D. SIMON et D. L. SIMON (2010)).

Supposons qu'à la date t , la valeur de X_t , x_t , est soumise à s contraintes d'inégalité :

$$a_{ti} \leq {}^t\phi_{ti}x_t \leq b_{ti}, \quad i = 1, \dots, s \quad (1.12)$$

où $(a_{ti}, b_{ti}) \in [-\infty, \infty[\times]-\infty, \infty]$.

Le but est de tronquer la distribution $N(\widehat{x}_t, P_t)$ sur les s contraintes (1.12) et en déduire la moyenne \widetilde{x}_t et sa covariance \widetilde{P}_t .

D'abord, initialisons le problème comme suit :

$$\begin{aligned} i &= 0; \\ \widetilde{x}_{ti} &= \widehat{x}_t; \\ \widetilde{P}_{ti} &= P_t. \end{aligned}$$

Posons ensuite :

$$z_{ti} = \varrho S^{-1/2} {}^tT(x_t - \widetilde{x}_{ti}) \quad (1.13)$$

où S est une matrice diagonale et T est une matrice orthogonale, obtenues par la décomposition de \widetilde{P}_{ti} telle que :

$$TS {}^tT = \widetilde{P}_{ti}$$

et ϱ est une matrice orthogonale de $\mathcal{M}_{(n,n)}$ qui satisfait :

$$\varrho S^{1/2} {}^tT\phi_{ti} = {}^t \left[({}^t\phi_{ti}\widetilde{P}_{ti}\phi_{ti})^{1/2} 0 \dots 0 \right] \quad (1.14)$$

Il découle de (1.13) et (1.14) que z_{ti} est un vecteur aléatoire centré réduit.

Ensuite, (1.12) est transformée comme suit :

$$c_{ti} \leq e_n z_{ti} \leq d_{ti}$$

où :

$$c_{ti} = \frac{a_{ti} - {}^t\phi_{ti}\widetilde{x}_{ti}}{({}^t\phi_{ti}\widetilde{P}_{ti}\phi_{ti})^{1/2}} \quad \text{et} \quad d_{ti} = \frac{b_{ti} - {}^t\phi_{ti}\widetilde{x}_{ti}}{({}^t\phi_{ti}\widetilde{P}_{ti}\phi_{ti})^{1/2}}$$

Remarquons que les éléments du vecteur z_{ti} sont indépendants comme sa covariance est égale à I_n . Seulement le premier élément de z_{ti} est contraint ayant la

distribution $N(0, 1)$. La troncature est donc unidimensionnelle que l'on applique à la loi $N(0, 1)$ entre c_{ti} et d_{ti} et on obtient la distribution normale tronquée comme suit :

$$\int_{c_{ti}}^{d_{ti}} \frac{1}{\sqrt{2\pi}} \exp(-u^2/2) du = \frac{1}{2} \left[\operatorname{erf}(d_{ti}/\sqrt{2}) - \operatorname{erf}(c_{ti}/\sqrt{2}) \right]$$

où $\operatorname{erf}(\cdot)$ est la fonction d'erreur de troncature définie comme :

$$\operatorname{erf}(v) = \frac{2}{\sqrt{\pi}} \int_0^v \exp(-u^2) du$$

Normalisons maintenant la densité tronquée que l'on note \tilde{f}_z telle que :

$$\tilde{f}_z(u) = \gamma \exp(-u^2/2) \text{ si } u \in [c_{ti}, d_{ti}], \text{ et } 0 \text{ sinon.}$$

où :

$$\gamma = \frac{\sqrt{2}}{\sqrt{\pi} [\operatorname{erf}(d_{ti}/\sqrt{2}) - \operatorname{erf}(c_{ti}/\sqrt{2})]}$$

On définit $z_{t,i+1}$ comme la variable aléatoire de densité \tilde{f}_z et dont le premier élément est de moyenne μ et de variance σ^2 , données par :

$$\mu = \mathbb{E}z_{t,i+1} = \gamma [\exp(-c_{ti}^2/2) - \exp(-d_{ti}^2/2)]$$

et

$$\sigma^2 = V(z_{t,i+1}) = \gamma [\exp(-c_{ti}^2/2)(c_{ti} - 2\mu) - \exp(-d_{ti}^2/2)(d_{ti} - 2\mu)] + \mu^2 + 1$$

Par conséquent, la moyenne et la variance de $z_{t,i+1}$ après l'application de la contrainte (1.12) pour $i = 1$, sont données par :

$$\tilde{z}_{t,i+1} = {}^t(\mu \ 0 \ \dots \ 0)$$

et

$$\operatorname{Cov}(\tilde{z}_{t,i+1}) = \operatorname{diag}(\sigma^2, 1, \dots, 1)$$

Il ne reste donc qu'inverser la transformation (1.13) pour obtenir la moyenne et la variance de x_t après l'application de (1.12), soient :

$$\tilde{x}_{t,i+1} = TS^{1/2} {}^t \varrho \tilde{z}_{t,i+1} + \tilde{x}_t i$$

et

$$\tilde{P}_{t,i+1} = TS^{1/2} {}^t \varrho \operatorname{Cov}(\tilde{z}_{t,i+1}) \varrho S^{1/2} {}^t T$$

On poursuit le même processus jusqu'à l'application du s^{eme} contrainte en obtenant ainsi l'état contrainte \tilde{x}_t et sa covariance \tilde{P}_t comme :

$$\tilde{x}_t = \tilde{x}_{ts} \text{ et } \tilde{P}_t = \tilde{P}_{ts}$$

La méthode de troncature s'avère la plus adéquate pour contraindre le filtre de Kalman dans le cas des contraintes d'inégalité sachant qu'elle est également applicable dans le cas des contraintes d'égalité (D. SIMON (2006)). En effet, cette méthode ne nécessite ni l'augmentation du modèle espace d'état ni le recours à des processus d'optimisation préalables.

1.3.4 Filtre d'information

Les équations du filtre de Kalman peuvent présenter dans certains cas une forme plus simple si on considère l'évolution des inverses des matrices de covariance des erreurs plutôt que ces matrices elles-mêmes (voir l'exemple 15.10 dans GOURIÉROUX et MONFORT (1995)). Dans ce cas, les matrices d'intérêt s'interprètent comme des matrices d'information, d'où le nom du filtre associé (FRASER (1967)).

D'ailleurs, l'utilité du filtre d'information est au-delà de cela, en fournissant une approche duale quand à la limite, l'information sur l'état initiale du système est inconnue. En effet, lorsqu'on a peu d'information sur l'état initiale X_0 , on est conduit à prendre une matrice de covariance P_0 très grande, voir infinie en absence de telle information tandis que P_0^{-1} devient très petite, voir nulle à la limite. Ainsi, la structure initiale de covariance peut être exprimée en fonction de la matrice d'information d'une manière permettant l'initialisation du filtre et par la suite, l'application des équations du filtre d'information. En introduisant les variables $\mathcal{I}_t = P_t^{-1}$ et $\hat{U}_t = \mathcal{I}_t \hat{X}_t$ en supposant que Q est inversible, on présente le filtre d'information comme suit (pour le détail du calcul voir GOURIÉROUX et MONFORT (1995), p.583) :

$$\begin{aligned} \hat{X}_{0|-1} &= \mathbb{E}X_0 \\ \mathcal{I}_{0|-1} &= 0 \\ \text{Pour } t &= 0, 1, \dots, n \text{ faire} \\ \hat{U}_{t|t} &= \hat{U}_{t|t-1} + {}^tHR^{-1}Y_t \\ \mathcal{I}_{t|t} &= \mathcal{I}_{t|t-1} + {}^tHR^{-1}H \\ \hat{U}_{t+1|t} &= [I_n - {}^tA^{-1}\mathcal{I}_{t|t}A^{-1}({}^tA^{-1}\mathcal{I}_{t|t}A^{-1} + Q^{-1})^{-1}] {}^tA^{-1}\hat{U}_{t|t} \\ \mathcal{I}_{t+1|t} &= [I_n - {}^tA^{-1}\mathcal{I}_{t|t}A^{-1}({}^tA^{-1}\mathcal{I}_{t|t}A^{-1} + Q^{-1})^{-1}] {}^tA^{-1}\mathcal{I}_{t|t}A^{-1} \\ \text{Finpour} \end{aligned}$$

Remarque 1.9.

- La notation P_t^{-1} est trompeuse car elle semble indiquer que P_t est inversible, ce qui peut ne pas être le cas, au moins pour les premières valeurs de t . L'équivalence avec le filtre de Kalman n'est effective qu'à partir du rang où $P_{t|t-1}$ et $P_{t|t}$ deviennent inversibles.
- D'après la proposition 1.4, si le filtre est stable, alors A est inversible et par conséquent, les équations du filtre d'information sont bien définies.
- Comme le filtre de Kalman nécessite l'inversion de la matrice $(HP_{t|t-1} {}^tH + R) \in \mathcal{M}_{(m,m)}$ contrairement au filtre d'information, ce dernier peut être numériquement

plus performant si le nombre des mesures dépasse largement la dimension du vecteur d'état ($m \gg r$).

1.4 Optimisation par approximation stochastique

L'approximation stochastique est un cadre général introduit par ROBBINS et MONRO (1951) qui englobe un ensemble d'algorithmes récursifs pour rechercher un minimum, généralement local, d'une fonction objectif à partir de ses mesures bruitées. Une classe de ces algorithmes dite à perturbations simultanées, est présentée par SPALL (1992) et s'intéresse aux algorithmes d'approximation stochastique à direction descente du gradient estimé par perturbation simultanée des paramètres de la fonction objectif.

1.4.1 Algorithme d'approximation stochastique à perturbations simultanées (SPSA)

On considère le problème de minimisation d'une fonction $L : \Theta \subset \mathbb{R}^p \rightarrow \mathbb{R}$. On cherche donc un certain $\theta^* \in \Theta$ qui vérifie :

$$\theta^* = \underset{\theta \in \Theta}{\operatorname{arg\,min}} L(\theta) \quad (1.15)$$

θ^* correspond également à une racine de l'équation du gradient de L , soit :

$$g(\theta^*) = \frac{\partial L}{\partial \theta}(\theta^*) = 0$$

Quand L et g sont parfaitement observées, les algorithmes récursifs dites déterministes peuvent être utilisés pour trouver θ^* (e.g., descente du gradient, Newton-Raphson). Dans le cadre de l'approximation stochastique, les mesures de L sont bruitées par un processus δ , et sont données par :

$$y(\theta) = L(\theta) + \delta(\theta)$$

Soit $\hat{\theta}_k$ l'estimation de θ à l'itération k , l'algorithme d'approximation stochastique a la forme :

$$\hat{\theta}_{k+1} = \hat{\theta}_k - a_k \hat{g}(\hat{\theta}_k) \quad (1.16)$$

où a_k est une suite de gain vérifiant certaines conditions assurant la convergence de (1.16). La perturbation simultanée consiste à estimer g en fonction de $y_k^\pm(\hat{\theta}_k) = L(\hat{\theta}_k \pm c_k \Delta_k) + \delta_k^\pm$, comme suit :

$$\hat{g}(\hat{\theta}_k) = \frac{y_k^+(\hat{\theta}_k) - y_k^-(\hat{\theta}_k)}{2c_k} \quad {}^t(\Delta_{k1}^{-1}, \dots, \Delta_{kp}^{-1})$$

où $\delta_{k \geq 0}^+$ et $\delta_{k \geq 0}^-$ sont deux vecteurs aléatoires *iid* et $\Delta_k = {}^t(\Delta_{k1}, \dots, \Delta_{kp})$ est un vecteur de p variables aléatoires centrées et mutuellement indépendantes satisfaisant certaines conditions assurant la convergence de (1.16).

Analyse de convergence

On considère l'algorithme SPSA à paramètre de perturbation fixe $c_k = c > 0$. Ainsi, la convergence de l'équation (1.16) dépend d'une part de la régularité de L , et d'autre part du choix des suites a_k et Δ_k . Les hypothèses suivantes (voir les hypothèses **5.1-5.6** dans BHATNAGAR, H. PRASHANTH et L. PRASHANTH (2013)) assurent l'existence d'un minimum local de L au sens du SPSA.

- (B1) L est lipschitzienne, continue et différentiable de dérivés secondes bornées. De plus, ∇L est lipschitzienne continue ;
- (B2) Pour tout $k \geq 0$, $a_k > 0$ et $a_k \rightarrow 0$ quand $k \rightarrow \infty$ de sorte que $\sum_k a_k = \infty$ et $\sum_k a_k^2 < \infty$;
- (B3) $\delta_{k \geq 0}^+$ et $\delta_{k \geq 0}^-$ sont des vecteurs aléatoires indépendants ayant une même distribution admettant des seconds moments finis.
- (B4) Les variables aléatoires Δ_{ki} , $k \geq 0$, $i = 1, \dots, p$ sont centrées et mutuellement indépendantes, de même distribution vérifiant pour tout $k \geq 0$,

$$\mathbb{E} [1/\Delta_{ki}^2] \leq \bar{K} \text{ pour un certain } \bar{K} < \infty;$$

- (B5) Les itérations suivant (1.16) restent uniformément stables presque sûrement, i.e.,

$$\sup_k \|\theta_k\| < \infty \text{ p.s}$$

- (B6) L'ensemble des minimums locaux de L est un compact de \mathbb{R}^p .

La proposition suivante assure, sous **(B1)-(B6)**, l'existence d'un minimum local de L dans un voisinage de l'ensemble de ses minimums locaux.

Proposition 1.6 (BHATNAGAR, H. PRASHANTH et L. PRASHANTH (2013)). *Soit H l'ensemble des minimums locaux de L . Supposons que **(B1)-(B6)** sont vérifiées. Alors pour tout $\eta > 0$, il existe $\bar{c} > 0$ tel que pour tout $c \in]0, \bar{c}]$, $\theta_k \rightarrow H_\eta$ p.s*

où H_η est un η -voisinage de H donné par :

$$H_\eta = \{\theta \in \Theta / \|\theta - \theta^*\| < \eta, \theta^* \in H\}$$

1.4.2 Algorithme fonctionnel lissé (Smoothed Functional)

L'algorithme fonctionnel lissé (Smoothed Functional algorithm) (CHIN (1997); KATKOVNIK et OY (1972); STYBLINSKI et TANG (1990)) appartient à la classe des méthodes de perturbation simultanée pour l'optimisation globale d'une fonction objectif. Cette fois-ci, le gradient de L est approximé par sa convolution avec une densité de probabilité convenable afin d'éviter les fluctuations et les extremums locaux de la fonction objectif, et assurer une convergence rapide de (1.16). Les conditions de convergence de cet algorithme ont été établies par BHATNAGAR, H. PRASHANTH et L. PRASHANTH (2013) dans le sens de la proposition 1.6, et elles ont été relaxées par CHEN et GAO (2015) en introduisant des troncatures dans l'équation récurrente

(1.16).

Considérons le problème d'optimisation (1.15) et introduisons la fonction de convolution suivante :

$$\tilde{L}(\theta, \beta) = \int_{\mathbb{R}^p} h(\eta, \beta) L(\theta - \eta) d\eta = \int_{\mathbb{R}^p} h(\theta - \eta, \beta) L(\eta) d\eta$$

où $\eta \in \mathbb{R}^p$ est le vecteur des perturbations aléatoires. D'où :

$$\tilde{L}(\theta, \beta) = \mathbb{E}_\eta [L(\theta - \eta)]$$

Ainsi, $\tilde{L}(\theta, \beta)$ est dite l'approximation lissée (smoothed approximation) de $L(\theta)$, et le noyau $h(\eta, \beta)$ est la fonction de densité de probabilité associée à η . Le paramètre β contrôle la dispersion de h , i.e., le degré de lissage de $L(\theta)$ (e.g., β peut contrôler l'écart-type de la suite η_1, \dots, η_p). De plus, quand β est suffisamment grand, le lissage devient capable d'éliminer les minimums locaux de $\tilde{L}(\theta, \beta)$ en fournissant son minimum global. De plus, pour que $\tilde{L}(\theta, \beta)$ ait de bonnes propriétés de lissage, RUBINSTEIN et KROESE (2016) proposent qu'elle vérifie les hypothèses suivantes :

- (a) $h(\eta, \beta) = \frac{1}{\beta^p} h(\eta/\beta)$ est différentiable par morceaux par rapport à η ;
- (b) $h(\eta, \beta) \rightarrow \delta(\eta)$ (mesure de Dirac) quand $\beta \rightarrow 0$;
- (c) $h(\theta, \beta)$ est une densité de probabilité.

Sous ces conditions, $\tilde{L}(\theta, \beta) \rightarrow L(\theta)$, quand $\beta \rightarrow 0$. Cela devrait se produire à la fin de l'optimisation pour assurer la convergence vers le véritable minimum de la fonction. Ainsi, ces conditions encadrent le choix des densités de probabilité candidates d'être utilisées comme des noyaux de convolution. La distribution normal, uniforme ou de Cauchy sont les plus utilisées.

L'idée est donc qu'en se donnant une fonction $L(\theta)$, on construit sa convolution lissée $\tilde{L}(\theta, \beta)$ et on l'utilise pour trouver l'extremum de $L(\theta)$ qui coïnciderait avec $\tilde{L}(\theta, \beta)$ quand $\beta \rightarrow 0$. Le problème maintenant est reformulé comme suit :

$$\min_{\theta \in \Theta} \tilde{L}(\theta, \beta)$$

avec $\beta \rightarrow 0$ quand $\theta \rightarrow \theta^*$.

En utilisant les propriétés de dérivation des convolutions, on parvient aisément à estimer le gradient lissé comme :

$$\nabla_\theta \tilde{L}(\theta, \beta) = \frac{1}{\beta} \int_{\mathbb{R}^p} \nabla_\eta h(\eta) L(\theta - \beta\eta) d\eta$$

que l'on peut estimer par :

$$\hat{\nabla}_\theta \tilde{L}(\theta, \beta) = \frac{-1}{\beta} \eta L(\theta - \beta\eta)$$

Une deuxième version de l'algorithme fonctionnel lissé consiste à utiliser une double perturbation de L autour de θ (*double-sided smoothed gradient*). Dans ce cas, on s'appuie sur la fonction lissée :

$$\tilde{L}(\theta, \beta) = \int_{\mathbb{R}^p} h(\eta, \beta) [L(\theta - \eta) + L(\theta + \eta)] d\eta$$

à laquelle correspond le gradient lissé :

$$\nabla_{\theta} \tilde{L}(\theta, \beta) = \frac{1}{\beta} \int_{\mathbb{R}^p} \nabla_{\eta} h(\eta) [L(\theta - \beta\eta) - L(\theta + \beta\eta)] d\eta$$

que l'on peut estimer par :

$$\widehat{\nabla}_{\theta} \tilde{L}(\theta, \beta) = \frac{1}{2\beta} \eta [L(\theta + \beta\eta) - L(\theta - \beta\eta)]$$

Remarque 1.10.

- Pour obtenir $\widehat{\nabla}_{\theta} \tilde{L}(\theta, \beta)$, on n'a pas besoin de connaître le vrai gradient $\nabla L(\theta)$. Cela est particulièrement intéressant si les formes analytiques de $L(\theta)$ ou son gradient ne sont pas connues et/ou ne peuvent être que simulées.
- L'algorithme fonctionnel lissé est indépendant de la convexité de $L(\theta)$ (STYBLINSKI et TANG (1990)).

Pré-estimation robuste de la variance conditionnelle générée par un GARCH

Conditions de positivité de Bollerslev

L'une des lacunes majeures dans l'estimation du modèle GARCH est qu'il est susceptible de générer une variance conditionnelle à valeurs négatives. Afin de remédier à cela, BOLLERSLEV (1986) a imposé la positivité des paramètres du modèle comme conditions suffisantes assurant la positivité de sa variance conditionnelle. Bien qu'elles sont simples à implémenter dans un processus d'estimation sous contrainte, ces conditions ne font que réduire la flexibilité du processus GARCH quant aux formes d'autocorrélation de son carrée, notamment pour des ordres $1 \leq p, q \leq 2$ (HE et TERÄSVIRTA (1999b)). De plus, la positivité des paramètres de la représentation (1.7) découlant de la positivité des paramètres du modèle, contredit le comportement cyclique de la volatilité, étant empiriquement vérifié, c'est-à-dire le fait qu'une période de forte volatilité peut entraîner une période de faible volatilité à un certain horizon (NELSON (1991)). Dans le même sens, plusieurs études empiriques ont montré des violations des contraintes de Bollerslev, citant à titre d'exemple, le GARCH(2,1) abordé par FRENCH, SCHWERT et STAMBAUGH (1987) et le GARCH(4,1) utilisé par R. ENGLE, ITO et LIN (1990). En effet, dans ces travaux, certains paramètres estimés étaient significativement négatifs, ce qui remettait en question la considération préalable des contraintes de Bollerslev dans l'estimation d'un GARCH.

Relaxation du modèle GARCH

Ainsi, l'accent a été mis sur la recherche des conditions moins restrictives de positivité de la variance conditionnelle. En ce sens, NELSON et CAO (1992) se sont appuyés sur la représentation ARCH(∞) du GARCH(p, q) (2.2) et en ont dérivé des contraintes nécessaires et suffisantes relaxant celles de Bollerslev pour $q \leq 2$

(voir les théorèmes 1.6 et 1.7), étant la fourchette d'ordre permettant de réduire ces contraintes à un nombre fini d'inégalités. Le même travail traite brièvement le cas $q \geq 3$ en fournissant des conditions suffisantes assurant la positivité presque sûre de la variance conditionnelle. Bien que TSAI et CHAN (2008) montrent que ces conditions sont également nécessaires (voir le théorème 1.8), elles ne semblent pas pratiques d'un point de vue numérique. Effectivement, NELSON et CAO (1992) ne préconisent pas de les imposer dans l'estimation des paramètres du modèle et favorisent en revanche l'utilisation d'autres routines plus pratiques pour les modèles GARCH d'ordres élevés.

Nous n'avons pas tort de dire que le large éventail d'utilisation des modèles GARCH(p, q) pour $1 \leq p, q \leq 2$ a freiné le développement des méthodes numériques flexibles permettant l'estimation des paramètres du GARCH d'ordres élevés, y compris le traitement du problème de la positivité de sa variance conditionnelle. Notons à titre d'exemple que sur le logiciel R, tous les packages d'estimation des modèles GARCH (e.g, fGarch (WUERTZ et al. (2013)), rugarch (A. GHALANOS, M. A. GHALANOS et RCPP (2019))) se restreignent à l'hypothèse de positivité des paramètres. Naturellement, cela ne devrait pas empêcher de disposer d'un cadre de calcul flexible pour les cas $p, q > 2$, d'autant plus que les approches précédentes (NELSON et CAO (1992); TSAI et CHAN (2008)) présentent quelques inconvénients majeurs. En effet, selon l'approche de Nelson et Cao, Tsai et Chan, l'extraction des contraintes sur les paramètres du modèle GARCH à partir de la représentation ARCH(∞), doit être effectuée à chaque fois que les ordres p et q sont modifiés. En outre, le nombre de ces contraintes devient de plus en plus important pour des GARCH d'ordres élevés, ce qui n'est pas pratique à mettre en œuvre dans une routine d'optimisation. Par exemple, pour le GARCH(6,1), les inégalités de Nelson et Cao selon le théorème (1.6) sont données comme suit :

$$\begin{aligned}
 \omega &> 0 \\
 \alpha_1 &\geq 0 \\
 \alpha_1\beta_1^2 + \alpha_2\beta_1 + \alpha_3 &\geq 0 \\
 \alpha_1\beta_1^3 + \alpha_2\beta_1^2 + \alpha_3\beta_1 + \alpha_4 &\geq 0 \\
 \alpha_1\beta_1^4 + \alpha_2\beta_1^3 + \alpha_3\beta_1^2 + \alpha_4\beta_1 + \alpha_5 &\geq 0 \\
 \alpha_1\beta_1^5 + \alpha_2\beta_1^4 + \alpha_3\beta_1^3 + \alpha_4\beta_1^2 + \alpha_5\beta_1 + \alpha_6 &\geq 0
 \end{aligned}$$

Une fois qu'on passe au GARCH(6,2), les inégalités de Nelson et Cao changent radicalement et deviennent selon le théorème (1.7) :

$$\begin{aligned}
 \omega/1 - \Delta_1 - \Delta_2 + \Delta_1\Delta_2 &\geq 0 \\
 \Delta_1, \Delta_2 &\in \mathbb{R} \\
 \Delta_1 &> 0 \\
 \alpha_1 + \Delta_1^{-1}\alpha_2 + \Delta_1^{-2}\alpha_3 + \Delta_1^{-3}\alpha_4 + \Delta_1^{-4}\alpha_5 &\geq 0
 \end{aligned}$$

où Δ_1 et Δ_2 sont les racines de $1 - \beta_1 z^{-1} - \beta_2 z^{-2}$. Le système des inégalités devient de plus en plus complexe pour des valeurs plus élevées de q .

Il est bien clair que l'inconvénient de telle approche ne réside pas dans le système des inégalités qu'elle génère, étant plus ou moins simple à implémenter, mais plutôt dans l'identification répétitive à priori de ces inégalités pour des ordres éventuellement choisis de p et q .

Pré-estimation robuste de la variance conditionnelle

Les limites de cette approche découlent de l'idée de chercher à contraindre les paramètres du modèle par rapport à la positivité de sa variance conditionnelle. Ce qui nécessite la modification de ces contraintes chaque fois que les ordres du GARCH changent, ou pour considérer une information plus fine sur la variance conditionnelle autre que la positivité. Notre approche consiste à contraindre directement la variance conditionnelle sans aucune identification à priori, ni application des contraintes sur les paramètres du modèle. Il s'agit d'une pré-estimation¹ robuste en deux étapes de la variance conditionnelle générée par le modèle GARCH tout en restant positive en relaxant les contraintes de positivité des paramètres de BOLLERSLEV (1986) au sens de NELSON et CAO (1992) et TSAI et CHAN (2008). Principalement, suite à cette relaxation, une nouvelle condition suffisante de stationnarité au second ordre est établie. Quant à l'existence des moments pour le nouveau GARCH relaxé, seul le moment d'ordre 4 est requis dont la condition suffisante d'existence est également établie. Sous ces conditions, la première étape de notre approche consiste à prédire la variance conditionnelle par le filtre de Kalman à travers une représentation espace d'état. Dans une deuxième étape, la variance conditionnelle prédite est robustifiée pour être indépendante des signes des paramètres du modèle en contraignant le filtre de Kalman utilisé par la méthode de troncature appliquée à la densité de probabilité de la série de la variance conditionnelle prédite.

Par ailleurs, notre approche généralise le concept de positivité de la variance conditionnelle considérée comme une information triviale sur la volatilité, à d'autres informations qui peuvent être disponibles et/ou imposées à priori sur la volatilité, notamment celle de bornitude. Bien que le GARCH standard n'exploite pas de telles informations dans l'amélioration et le redressement des estimations des paramètres, notre approche permet en revanche d'exprimer ces informations sous forme d'une contrainte sur la variance conditionnelle sans passer par l'identification à priori des contraintes sur les paramètres qui la vérifient.

En vu de mettre en exergue l'utilité de notre approche, nous présentons deux illustrations numériques par simulation Monte Carlo, dans lesquelles nous montrons d'une part, la capacité de l'estimation robuste de la variance conditionnelle d'améliorer la qualité des estimations de QML des paramètres du modèle GARCH relaxé pour des ordres élevés. D'autre part, nous préconisons l'utilité de notre approche à exploiter d'autres informations sur la volatilité dans le redressement des estimations des paramètres du modèle.

1. Nous utilisons ce terme en liaison avec l'estimation par QML pour désigner l'estimation de σ_t^2 permettant de construire la vraisemblance du modèle.

Dans la section §2.1, nous introduisons le modèle GARCH relaxé en dérivant également ses propriétés principales en termes de stationnarité au second ordre et d'existence des moments. Dans la section §2.2, nous représentons le GARCH relaxé sous forme d'espace d'état à partir duquel nous déterminons la structure d'auto-covariance du modèle. L'estimation robuste de la variance conditionnelle est présentée dans la sections §2.3. La performance de l'approche proposée est évaluée par simulation Monte Carlo dans la sections §2.4.

2.1 Modèle GARCH relaxé

La relaxation du modèle GARCH consiste au fait qu'à priori aucune contrainte n'est imposée sur les paramètres ARCH et GARCH du modèle 1.5. Seule la condition $\omega > 0$ est retenue étant nécessaire pour maintenir la positivité de la variance inconditionnelle. Suite à cette relaxation, le modèle 1.5 est désormais appelé GARCH relaxé et noté r -GARCH. Dans la suite, η_t est supposé indépendamment et identiquement distribué selon une loi normale centrée réduite.

Quant aux propriétés statistiques du modèle GARCH, à savoir la stationnarité au second ordre (Théorème 1.3) et l'existence des moments (Théorème 1.9), il convient de noter qu'elles sont intrinsèquement basées sur l'hypothèse de positivité des paramètres. Par conséquent, la relaxation de cette hypothèse dans le contexte d'un GARCH relaxé nécessite l'établissement de nouvelles conditions de stationnarité au second ordre et d'existence des moments, en particulier celui d'ordre 4. Dans ce sens, nous considérons les hypothèses suivantes :

$$\mathbf{(A1)} \quad \sum_{i=1}^p |\alpha_i| + \sum_{j=1}^q |\beta_j| < 1$$

$$\mathbf{(A2)} \quad \rho(A_1^{(2)}) < 1, \text{ où :}$$

$$A_t = \begin{pmatrix} \alpha_1 \eta_t^2 & \cdots & \alpha_p \eta_t^2 & \beta_1 \eta_t^2 & \cdots & \beta_q \eta_t^2 \\ & I_{p-1} & 0_{(p-1,1)} & & 0_{(p-1,q)} & \\ \alpha_1 & \cdots & \alpha_p & \beta_1 & \cdots & \beta_q \\ & 0_{(q-1,p)} & & & I_{q-1} & 0_{(q-1,1)} \end{pmatrix}$$

En se basant sur les hypothèses **(A1)** et **(A2)**, la condition de stationnarité au second ordre ainsi que celle d'existence du moment d'ordre 4 du GARCH relaxé sont données par les propositions 2.1 et 2.2 dont les démonstrations nécessitent l'utilisation du lemme suivant :

Lemme 2.1. *Soient a_1, a_2, \dots, a_s des réels tels que $a_s \neq 0$. Alors, toutes les racines du polynôme $P(z) = 1 - \sum_{i=1}^s a_i z^i$ sont à l'extérieur du cercle unité si et seulement si $\sum_{i=1}^s |a_i| < 1$.*

Démonstration. Supposons que $\sum_{i=1}^s |a_i| \geq 1$.

Puisque $P(0) = 1$ et $P(1) \leq 0$

Alors l'équation $P(z) = 0$ admet au moins une solution dans l'intervalle $[0, 1]$, ce qui est absurde.

Inversement, pour tout $|z| \leq 1$, on a :

$$\begin{aligned} |P(z)| &= \left| 1 - \sum_{i=1}^s a_i z^i \right| \geq \left| 1 - \left| \sum_{i=1}^s a_i z^i \right| \right| \geq 1 - \left| \sum_{i=1}^s a_i z^i \right| \geq 1 - \sum_{i=1}^s |a_i| |z^i| \\ &\geq 1 - \sum_{i=1}^s |a_i| > 0 \end{aligned}$$

D'où le résultat. □

Proposition 2.1. (Stationnarité au second ordre) Soit ε_t un r -GARCH(p, q) satisfaisant **(A1)**, alors ε_t est stationnaire au second ordre, et on a :

$$\mathbb{E}\varepsilon_t^2 = \frac{\omega}{1 - \sum_{i=1}^p \alpha_i - \sum_{j=1}^q \beta_j}$$

Démonstration. Rappelons d'une part que $\mathbb{E}\varepsilon_t = 0$ et que $Cov(\varepsilon_t, \varepsilon_{t-k}) = 0$ pour tout $k > 0$. D'autre part, partons de la représentation (1.5), étant donné que :

$$\sum_{i=1}^r |\alpha_i + \beta_i| \leq \sum_{i=1}^p |\alpha_i| + \sum_{j=1}^q |\beta_j| < 1$$

Nous concluons d'après le lemme 2.1 que ε_t^2 est inversible. Comme ν_t est un bruit blanc, on obtient :

$$\mathbb{E}\varepsilon_t^2 = \frac{\omega}{1 - \sum_{i=1}^p \alpha_i - \sum_{j=1}^q \beta_j} < \infty$$

D'où la proposition. □

Proposition 2.2. (Existence du moment d'ordre 4) Soit ε_t un r -GARCH(p, q), alors **(A1)** et **(A2)** sont suffisantes pour que $\mathbb{E}\varepsilon_t^4 < \infty$.

Démonstration. Nous nous inspirons de la démonstration de la proposition 2.2 dans Christian FRANCO, WINTENBERGER et ZAKOIAN (2013b). En effet, partons de la représentation Markovienne du modèle r -GARCH(p, q) étant la même que celle du GARCH(p, q) standard, soit :

$$Z_t = b_t + A_t Z_{t-1} \tag{2.1}$$

où :

$$Z_t = {}^t(\varepsilon_t^2, \dots, \varepsilon_{t-p+1}^2, \sigma_t^2, \dots, \sigma_{t-q+1}^2) \in \mathbb{R}^{p+q}$$

et

$$b_t = {}^t(\omega\eta_t^2, 0, \dots, \omega, 0, \dots, 0) \in \mathbb{R}^{p+q}$$

Réécrivons Z_t comme suit :

$$Z_t = \sum_{l=0}^{\infty} Z_{t,l}$$

où :

$$Z_{t,l} = A_{t,l}b_{t-l} \quad \text{et} \quad A_{t,l} = A_t A_{t-1} \dots A_{t-l+1}, \quad \text{pour tout } l > 0$$

Par convention $A_{t,0} = I_{p+q}$ et $Z_{t,0} = b_t$.

Notons que la positivité des paramètres n'est pas nécessaire pour que Z_t soit presque sûrement définie (voir FRANCO (2010), p.27).

Soit donc :

$$\tilde{A}_{t,l} = \prod_{j=0}^{l-1} Abs(A_{t-j})$$

Puisque A_{t-i} est une fonction de η_{t-i} en supposant à priori que σ_t^2 est positive, alors l'indépendance des matrices du produit $A_t A_{t-1} \dots A_{t-l+1} b_{t-l}$ implique que pour tout $m \in \mathbb{N}^*$, on a :

$$\begin{aligned} \|Z_t\|_m &= \|Abs(Z_t)\|_m \leq \sum_{l=0}^{\infty} \|\tilde{A}_{t,l} Abs(b_{t-l})\| \\ &= \sum_{l=0}^{\infty} \left\{ \mathbb{E} \|\tilde{A}_{t,l} Abs(b_{t-l})\|^m \right\}^{1/m} \\ &= \sum_{l=0}^{\infty} \left\{ \mathbb{E} \|\tilde{A}_{t,l}^{\otimes m} Abs(b_{t-l})^{\otimes m}\| \right\}^{1/m} \\ &= \sum_{l=0}^{\infty} \left\{ \|\mathbb{E} \tilde{A}_{t,l}^{\otimes m} Abs(b_{t-l})^{\otimes m}\| \right\}^{1/m} \end{aligned}$$

Comme $\tilde{A}_{t,l}^{\otimes m} Abs(b_{t-l})^{\otimes m}$ est positive, alors :

$$\begin{aligned} \|Z_t\|_m &\leq \sum_{l=0}^{\infty} \left\{ \|\mathbb{E} \prod_{j=0}^{l-1} Abs(A_{t-j})^{\otimes m} Abs(b_{t-l})^{\otimes m}\| \right\}^{1/m} \\ &= \sum_{l=0}^{\infty} \left\{ \left\| \prod_{j=0}^{l-1} \mathbb{E} Abs(A_{t-j})^{\otimes m} \mathbb{E} Abs(b_{t-l})^{\otimes m} \right\| \right\}^{1/m} \\ &= \|b_1^{(m)}\|^{1/m} \sum_{l=0}^{\infty} \|(A^{(m)})^l\|^{1/m} \end{aligned}$$

Par conséquent, $\rho(A^{(m)}) < 1$ entraîne $\|(A^{(m)})^l\| \rightarrow 0$ quand $l \rightarrow \infty$, et la somme converge. Ceci complète la preuve. □

Remarque 2.1.

- En relaxant les contraintes de positivité des paramètres, **(A1)** et **(A2)** relaxent respectivement les conditions de stationnarité au second ordre du théorème 1.3 et

d'existence du moment d'ordre quatre du théorème 1.9 associées au modèle GARCH standard. L'équivalence est obtenue quand les paramètres du modèle sont supposés à priori positifs.

- L'hypothèse **(A1)** relaxe l'hypothèse (8) de NELSON et CAO (1992)² dans le sens où elle assure à la fois la stationnarité au second ordre du r -GARCH et l'inversibilité de σ_t^2 telle qu'elle est donnée par le corollaire suivant :

Corollaire 2.1. Soit ε_t un r -GARCH(p, q) de variance conditionnelle $\sigma_t^2 > 0$. Sous l'hypothèse **(A1)**, σ_t^2 admet une représentation ARCH(∞) donnée par :

$$\sigma_t^2 = \omega^* + \Psi(L)\varepsilon_t^2 \quad (2.2)$$

où :

$$\omega^* = \frac{\omega}{\beta(1)} \quad \text{et} \quad \Psi(L) = \frac{\alpha(L)}{\beta(L)}$$

avec :

$$\beta(L) = 1 - \sum_{j=1}^q \beta_j L^j \quad \text{et} \quad \alpha(L) = \sum_{i=1}^p \alpha_i L^i$$

Démonstration. Partant de la représentation polynomiale de l'équation de la variance conditionnelle du GARCH(p, q), soit :

$$\beta(L)\sigma_t^2 = \omega + \alpha(L)\varepsilon_t^2$$

Sous **(A1)**, on a :

$$\sum_{j=1}^q |\beta_j| \leq 1 - \sum_{i=1}^p |\alpha_i| < 1$$

Ainsi, en vertu du lemme 2.1, on déduit que les racines de β sont à l'extérieur du cercle unité et par conséquent β est inversible. D'où la représentation (2.2). □

2.2 Représentation espace d'état du modèle r -GARCH et structure de covariance

2.2.1 Représentation espace d'état du modèle r -GARCH

Sous les hypothèses **(A1)** et **(A2)**, nous dérivons une représentation espace d'état du modèle r -GARCH(p, q) à partir de la représentation ARMA de ε_t^2 (1.5) en remplaçant ce dernier par $\nu_t + \sigma_t^2$, soit :

$$\sigma_t^2 = \omega + \sum_{i=1}^r \lambda_i \sigma_{t-i}^2 + \sum_{j=1}^r \alpha_j \nu_{t-j} \quad (t \in \mathbb{Z}) \quad (2.3)$$

2. Il s'agit de l'hypothèse que les racines du polynôme $\beta(L) = 1 - \sum_{j=1}^q \beta_j L^j$ sont à l'extérieur du cercle unitaire.

où $\lambda_i = \alpha_i + \beta_i$ tel que $\lambda_i = \beta_i$ pour $i > p$ et $\lambda_i = \alpha_i$ pour $i > q$.

Par conséquent, la représentation espace d'état est donnée sous forme d'innovation (voir Chapitre 9 dans ANDERSON et MOORE (2012)) basée sur le processus ν_t comme suit :

$$\begin{cases} h_t = \underline{\omega} + \Lambda h_{t-1} + \Phi \underline{\nu}_{t-1}, & \underline{\nu}_{t-1} = {}^t(\nu_{t-1} \dots \nu_{t-r}) \\ \varepsilon_t^2 = e_r h_t + \nu_t \end{cases} \quad (2.4)$$

où $h_t = {}^t(\sigma_t^2 \dots \sigma_{t-r+1}^2) \in \mathbb{R}^r$ représente la variable d'état, et $\underline{\omega}$, Λ et Φ sont les matrices de transition données par :

$$\begin{aligned} \underline{\omega} &= {}^t(\omega \ 0_{(r-1,1)}) \in \mathbb{R}^r \\ \Lambda &= \begin{pmatrix} \lambda_1 & \dots & \lambda_r \\ I_{r-1} & & 0_{(r-1,1)} \end{pmatrix} \in \mathcal{M}_{(r,r)} \\ \Phi &= \begin{pmatrix} \alpha_1 & \dots & \alpha_r \\ & 0_{(r-1,r-1)} & \end{pmatrix} \in \mathcal{M}_{(r,r)} \end{aligned}$$

Les bruits associés aux équations d'état et de mesure sont à la fois blancs, de moyenne zéro et non autocorrélés puisque ν_t est un bruit blanc. Comme condition initiale, nous supposons comme densité à priori, que h_0 est normal, ce qui équivaut à supposer que σ_t^2 a commencé dans le passé infini avec une distribution normale. Par voie de conséquence, le filtre de Kalman est applicable afin de prédire h_t étant données les mesures $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n\}$.

Notons que ν_t est loin d'être normal puisque η_t est supposé normal. En effet, $\nu_t = \sigma_t^2(\eta_t^2 - 1)$ est une forme à l'échelle d'un $\chi^2(1)$ non-centré. Néanmoins, même en l'absence de la normalité, le filtre de Kalman demeure linéairement optimal fournissant l'estimation linéaire optimale dans l'ensemble de toutes les estimations linéaires possibles (voir D. SIMON (2006), p.130).

Remarque 2.2.

Contrairement aux autres représentations espace d'état du modèle GARCH rapportées dans la littérature, la forme d'innovation de la représentation (2.4) possède certaines propriétés particulières. D'une part, tout modèle d'espace d'état peut être transformé en une forme d'innovation (voir la définition 3.2 dans ANDERSON et MOORE (2012)). D'autre part, cette forme est essentiellement unique (voir le théorème 3.2 dans ANDERSON et MOORE (2012)).

2.2.2 Structure d'autocovariance

L'application du filtre de Kalman nécessite la connaissance de la structure de covariance de h_t . Pour cela, nous proposons dans ce qui suit, une méthode récursive

basée sur le modèle espace d'état (2.4) permettant de calculer la fonction d'autocovariance de h_t .

Dans un premier temps, nous exprimons via le lemme suivant, $\mathbb{E}(\underline{\varepsilon}_{t-1}^2 \otimes \underline{\nu}_{t-1})$ et $\mathbb{E}(\underline{\nu}_{t-1} \otimes \underline{\varepsilon}_{t-1}^2)$ en fonction des paramètres du modèle et $\mathbb{E}\nu_t^2$ que l'on note ν . Nous en avons besoin pour le calcul de la fonction d'autocovariance de h_t donné par la proposition 2.3.

Lemme 2.2. *Soit ε_t un processus r -GARCH(p, q). Pour tout $k \in \llbracket 2, r \rrbracket$, soient $\underline{\alpha}_{k-1} = {}^t(\alpha_{k-1} \dots \alpha_1)$ et $\Gamma_{k-1}, M \in \mathcal{M}_{(k-1, k-1)}$ définies par :*

$$\Gamma_{k-1} = \begin{pmatrix} 1 & -\lambda_1 & -\lambda_2 & \cdot & \cdot & -\lambda_{k-1} \\ 0 & 1 & -\lambda_1 & \cdot & \cdot & -\lambda_{k-2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & -\lambda_1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \text{ et } M = \begin{pmatrix} 1 & m_1^{(1)} & m_1^{(2)} & \cdot & \cdot & m_1^{(k-1)} \\ 0 & 1 & m_2^{(2)} & \cdot & \cdot & m_2^{(k-1)} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & m_{k-1}^{(k-1)} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

où $m_i^{(k-1)}$ représente le i^{me} élément de $\Gamma_{k-1}^{-1} \underline{\alpha}_{k-1}$, $i = 1, \dots, k$. Alors :

$$\mathbb{E}(\underline{\varepsilon}_{t-1}^2 \otimes \underline{\nu}_{t-1}) = M_{\times \bullet} \nu \text{ et } \mathbb{E}(\underline{\nu}_{t-1} \otimes \underline{\varepsilon}_{t-1}^2) = M_{\bullet \times} \nu$$

Démonstration. Rappelons que le bruit blanc ν_t est l'innovation linéaire du processus ε_t^2 .

D'où pour tout $i > k > 0$:

$$\mathbb{E}(\varepsilon_{t-i}^2 \nu_{t-k}) = 0 \quad (2.5)$$

Il en découle :

$$\mathbb{E}(\varepsilon_{t-i}^2 \nu_{t-i}) = \mathbb{E} \left[\omega \nu_{t-i} + \sum_{j=1}^r \lambda_j \varepsilon_{t-i-j}^2 \nu_{t-i} + \nu_{t-i}^2 - \sum_{j=1}^q \beta_j \nu_{t-i-j} \nu_{t-i} \right] = \nu \quad (2.6)$$

Considérons maintenant le cas $0 < i < k \leq r$. En introduisant l'espérance mathématique de $\varepsilon_{t-i}^2 \nu_{t-k}$ dans (1.5), on obtient :

$$\mathbb{E}(\varepsilon_{t-i}^2 \nu_{t-k}) = \begin{cases} \alpha_{k-i} \nu + \sum_{j=1}^{k-i-1} \lambda_j \mathbb{E}(\varepsilon_{t-i-j}^2 \nu_{t-k}) & \text{si } 0 < i < k-1 \\ \alpha_1 \nu & \text{si } i = k-1 \end{cases} \quad (2.7)$$

On pose $X_{i,k} = \mathbb{E}(\varepsilon_{t-i}^2 \nu_{t-k})$. Alors, (2.5), (2.6) et (2.7) deviennent :

$$X_{i,k} = \begin{cases} \alpha_{k-i} \nu + \sum_{j=1}^{k-i-1} \lambda_j X_{i+j,k} & \text{si } 0 < i < k-1 \\ \alpha_1 \nu & \text{si } i = k-1 \\ \nu & \text{si } i = k \\ 0 & \text{si } i > k \end{cases} \quad (2.8)$$

Les deux premières équations du système (2.8) peuvent être exprimées sous forme matricielle comme :

$${}^t(X_{i,1} \dots X_{i,k-1}) = \Gamma_{k-1}^{-1} \underline{\alpha}_{k-1} \nu$$

où $\underline{\alpha}_{k-1} = {}^t(\alpha_{k-1} \dots \alpha_1)$ et

$$\Gamma_{k-1} = \begin{pmatrix} 1 & -\lambda_1 & -\lambda_2 & \cdot & \cdot & -\lambda_{k-1} \\ 0 & 1 & -\lambda_1 & \cdot & \cdot & -\lambda_{k-2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & -\lambda_1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Ce qui entraîne, en notant $m_i^{(k-1)}$ le i^{eme} élément de $\Gamma_{k-1}^{-1} \underline{\alpha}_{k-1}$, que :

$$\mathbb{E}(\varepsilon_{t-i}^2 \nu_{t-k}) = X_{i,k} = \begin{cases} m_i^{(k-1)} \nu & \text{si } 0 < i < k \\ \nu & \text{si } i = k \\ 0 & \text{si } i > k \end{cases} \quad (2.9)$$

Soit $\mathbf{X} = (X_{i,k})_{i,k=1,\dots,r} \in \mathcal{M}_{(r,r)}$ que l'on peut écrire comme $\mathbf{X} = M\nu$, où pour tout $0 < i < k \leq r$, on pose :

$$M = \begin{pmatrix} 1 & m_1^{(1)} & m_1^{(2)} & \cdot & \cdot & m_1^{(k-1)} \\ 0 & 1 & m_2^{(2)} & \cdot & \cdot & m_2^{(k-1)} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & m_{k-1}^{(k-1)} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

D'où, les deux espérances mathématiques $\mathbb{E}(\underline{\varepsilon}_{t-1}^2 \otimes \underline{\nu}_{t-1})$ et $\mathbb{E}(\underline{\nu}_{t-1} \otimes \underline{\varepsilon}_{t-1}^2)$ résultent. \square

Exemple 2.1. Nous considérons un modèle r -GARCH(2,2) et notons

$$\underline{\varepsilon}_{t-1}^2 = {}^t(\varepsilon_{t-1}^2 \ \varepsilon_{t-2}^2) \quad \text{et} \quad \underline{\nu}_{t-1} = {}^t(\nu_{t-1} \ \nu_{t-2})$$

alors :

$$\mathbb{E}(\underline{\varepsilon}_{t-1}^2 \otimes \underline{\nu}_{t-1}) = \mathbb{E} \begin{pmatrix} \varepsilon_{t-1}^2 \nu_{t-1} \\ \varepsilon_{t-1}^2 \nu_{t-2} \\ \varepsilon_{t-2}^2 \nu_{t-1} \\ \varepsilon_{t-2}^2 \nu_{t-2} \end{pmatrix} = \begin{pmatrix} 1 \\ m_1^{(1)} \\ 0 \\ 1 \end{pmatrix} \nu = \begin{pmatrix} {}^t M_{1\bullet} \\ {}^t M_{2\bullet} \end{pmatrix} \nu = M_{\times\bullet} \nu$$

et

$$\mathbb{E}(\underline{\nu}_{t-1} \otimes \underline{\varepsilon}_{t-1}^2) = \mathbb{E} \begin{pmatrix} \nu_{t-1} \varepsilon_{t-1}^2 \\ \nu_{t-1} \varepsilon_{t-2}^2 \\ \nu_{t-2} \varepsilon_{t-1}^2 \\ \nu_{t-2} \varepsilon_{t-2}^2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ m_1^{(1)} \\ 1 \end{pmatrix} \nu = \begin{pmatrix} M_{\bullet 1} \\ M_{\bullet 2} \end{pmatrix} \nu = M_{\bullet \times} \nu$$

La proposition suivante permet de déterminer la fonction d'autocovariance de h_t à partir de la représentation espace d'état (2.4).

Proposition 2.3. Soient Ω , P , Q , B et C les matrices définies par :

- $\Omega \in \mathcal{M}_{(r^2, r^2)}$ ayant comme éléments 1 aux positions $((k-1)r+k, (k-1)r+k)$ pour $k = 1, \dots, r$ et 0 ailleurs ;
- $P = 2(\Lambda \otimes \Phi)(M_{\times \bullet} e_{r^2} - \Omega) + \Lambda^{\otimes 2}$;
- $Q = \{(\Phi \otimes \Lambda)(M_{\bullet \times} e_{r^2} - \Omega) + \Phi^{\otimes 2}\} \Omega$;
- $B = (B_k)_k \in \mathbb{R}^{r^2}$ tel que $B_k = \omega^2 \left(1 - \sum_{i=1}^p \alpha_i - \sum_{j=1}^q \beta_j\right)^{-1}$ pour $k = 1, \dots, r$ et 0 ailleurs ;
- $C = (C_k)_k \in \mathbb{R}^{r^2}$ tel que $C_{(l-1)r+1} = \omega^2 \left(1 - \sum_{i=1}^p \alpha_i - \sum_{j=1}^q \beta_j\right)^{-1}$ pour $l = 1, \dots, r$ et 0 ailleurs.

Nous supposons que $I_{r^2} - P - 2Q$ est inversible. Alors, la fonction d'autocovariance de h_t est obtenue comme suit :

$$\text{Cov}(\sigma_t^2, \sigma_{t-s}^2) = e_{r^2} \mathbb{E}(h_t \otimes h_{t-s}) - (e_r E h_0)^2, \quad \text{pour tout } s \geq 0 \quad (2.10)$$

tel que :

$$\mathbb{E}(h_t^{\otimes 2}) = (I_{r^2} - P - 2Q)^{-1} \{B + (\Lambda \otimes I_r)C\} \quad (2.11)$$

et pour tout $s > 0$:

$$\mathbb{E}(h_t \otimes h_{t-s}) = B + (\Lambda \otimes I_r) \mathbb{E}(h_t \otimes h_{t-s+1}) + 2(\Phi \otimes I_r)(M_{\bullet \times}^{(s)} - g_k) e_{r^2} \Omega \mathbb{E}(h_t^{\otimes 2}) \quad (2.12)$$

où $g_k \in \mathbb{R}^{r^2}$ dont les éléments valent 1 aux positions $(s-1+k)(r-s+1)+k+1$ pour $k \geq 0$ et 0 ailleurs.

Démonstration. En appliquant les propriétés élémentaires du produit de Kronecker à l'équation d'état de la représentation (2.4), on obtient :

$$h_t^{\otimes 2} = \underline{\omega} \otimes h_t + \Lambda h_{t-1} \otimes (\underline{\omega} + \Lambda h_{t-1} + \Phi \underline{\nu}_{t-1}) + \Phi \underline{\nu}_{t-1} \otimes (\underline{\omega} + \Lambda h_{t-1} + \Phi \underline{\nu}_{t-1})$$

On pose :

$$\begin{aligned} T_1 &= \Lambda h_{t-1} \otimes (\underline{\omega} + \Lambda h_{t-1} + \Phi \underline{\nu}_{t-1}) \\ T_2 &= \Phi \underline{\nu}_{t-1} \otimes (\underline{\omega} + \Lambda h_{t-1} + \Phi \underline{\nu}_{t-1}) \end{aligned}$$

D'où, d'une part on a :

$$\begin{aligned} T_1 &= (\Lambda \otimes I_r) (h_{t-1} \otimes \underline{\omega}) + \Lambda^{\otimes 2} h_{t-1}^{\otimes 2} + (\Lambda \otimes \Phi) (h_{t-1} \otimes \underline{\nu}_{t-1}) \\ &= (\Lambda \otimes I_r) (h_{t-1} \otimes \underline{\omega}) + \Lambda^{\otimes 2} h_{t-1}^{\otimes 2} + (\Lambda \otimes \Phi) (\underline{\varepsilon}_{t-1}^2 \otimes \underline{\nu}_{t-1}) - (\Lambda \otimes \Phi) \underline{\nu}_{t-1}^{\otimes 2} \end{aligned}$$

Comme :

$$\nu = \mathbb{E} \nu_t^2 = \mathbb{E} (\varepsilon_t^2 - \sigma_t^2)^2 = \mathbb{E} (\eta_t^2 \sigma_t^2 - \sigma_t^2)^2 = \mathbb{E} (\eta_t^2 - 1)^2 \mathbb{E} \sigma_t^4 = 2e_{r,2} \mathbb{E} (h_t^{\otimes 2}) \quad (2.13)$$

et

$$\mathbb{E} \underline{\nu}_t^{\otimes 2} = 2\Omega \mathbb{E} (h_t^{\otimes 2})$$

Cela entraîne :

$$\begin{aligned} \mathbb{E} T_1 &= (\Lambda \otimes I_r) \mathbb{E} (h_{t-1} \otimes \underline{\omega}) + \{2(\Lambda \otimes \Phi) (M_{\bullet \times} e_{r,2} - I_{r,2}) \Omega + \Lambda^{\otimes 2}\} \mathbb{E} (h_t^{\otimes 2}) \\ &= (\Lambda \otimes I_r) C + P \mathbb{E} (h_t^{\otimes 2}) \end{aligned}$$

D'autre par, on a :

$$\begin{aligned} T_2 &= (\Phi \otimes I_r) (\underline{\nu}_{t-1} \otimes \underline{\omega}) + (\Phi \otimes \Lambda) (\underline{\nu}_{t-1} \otimes h_{t-1}) + \Phi^{\otimes 2} \underline{\nu}_{t-1}^{\otimes 2} \\ &= (\Phi \otimes I_r) (\underline{\nu}_{t-1} \otimes \underline{\omega}) + (\Phi \otimes \Lambda) (\underline{\nu}_{t-1} \otimes \underline{\varepsilon}_{t-1}^2) - \{(\Phi \otimes \Lambda) - \Phi^{\otimes 2}\} \underline{\nu}_{t-1}^{\otimes 2} \end{aligned}$$

Il en découle :

$$\begin{aligned} \mathbb{E} T_2 &= 2(\Phi \otimes \Lambda) M_{\bullet \times} e_{r,2} \Omega \mathbb{E} (h_t^{\otimes 2}) + 2(\Phi^{\otimes 2} - \Phi \otimes \Lambda) \Omega \mathbb{E} (h_t^{\otimes 2}) \\ &= 2\{(\Phi \otimes \Lambda) (M_{\bullet \times} e_{r,2} - I_{r,2}) + \Phi^{\otimes 2}\} \Omega \mathbb{E} (h_t^{\otimes 2}) \\ &= Q \mathbb{E} (h_t^{\otimes 2}) \end{aligned}$$

En conséquence :

$$\begin{aligned} \mathbb{E} (h_t^{\otimes 2}) &= \mathbb{E} (\underline{\omega} \otimes h_t) + (\Lambda \otimes I_r) C + P \mathbb{E} (h_t^{\otimes 2}) + Q \mathbb{E} (h_t^{\otimes 2}) \\ &= B + (\Lambda \otimes I_r) C + P \mathbb{E} (h_t^{\otimes 2}) + Q \mathbb{E} (h_t^{\otimes 2}) \end{aligned}$$

ce qui implique :

$$(I_{r,2} - P - Q) \mathbb{E} (h_t^{\otimes 2}) = B + (\Lambda \otimes I_r) C$$

Comme $(I_{r,2} - P - Q)$ est inversible, alors (2.11) résulte.

Par ailleurs, (2.12) est directement déduite en multipliant (2.4) par h_{t-s} et remarquant que $\mathbb{E} (\underline{\nu}_{t-1} \otimes \underline{\varepsilon}_{t-s}^2) = M_{\bullet \times}^{(s)} \nu$ et que $\mathbb{E} (\underline{\nu}_{t-1} \otimes \underline{\nu}_{t-s}) = g_k \nu$. \square

Remarque 2.3.

– FRANCQ (2010) utilise la représentation (2.1) pour calculer récursivement à la fois la fonction d'autocovariance de ε_t^2 et σ_t^2 . Le calcul de σ_t^2 apparait comme une étape intermédiaire dans des travaux antérieurs de HE et TERÄSVIRTA (1999a) afin de déterminer la fonction d'autocorrélation de ε_t^2 , et en particulier sous les conditions relaxées de Nelson et Cao (HE et TERÄSVIRTA (1999b)). Or, la proposition 2.3 utilise la même représentation espace d'état (2.4) si bien pour estimer σ_t^2 que pour calculer sa fonction d'autocovariance sans recourir à d'autres représentations supplémentaires. De plus, la proposition 2.3 est convertible en une forme algorithmique, ce qui s'avère délicat en passant par les expressions d'autocovariance développées par HE et TERÄSVIRTA (1999a) et HE et TERÄSVIRTA (1999b).

- D’après la proposition 2.3, le calcul de la fonction d’autocovariance de σ_t^2 ne nécessite pas la connaissance explicite de ν étant donnée (2.13), ce qui atténue éventuellement l’effet de la mal spécification de la distribution de ν_t sur les autocovariances calculées.

2.3 Estimation robuste de la variance conditionnelle

La méthode bien connue d’estimation de la variance conditionnelle générée par le modèle GARCH standard consiste à fixer une valeur initiale σ_0^2 et calculer récursivement les valeurs σ_t^2 en utilisant l’équation de la variance conditionnelle en se servant de la positivité des paramètres pour assurer la positivité de σ_t^2 , $t > 0$. Cependant, comme la valeur σ_0^2 n’est pas connue en pratique, le filtre de Kalman peut être utilisé afin de pré-estimer de manière optimale σ_t^2 à partir des mesures $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n\}$ (ALLAL et BENMOUMEN (2011); OSSANDÓN et BAHAMONDE (2013)).

Nous généralisons cette approche pour le modèle GARCH relaxé afin de fournir une estimation robuste de σ_t^2 dans le sens où à priori aucune restriction de positivité sur les paramètres ARCH et GARCH du modèle n’est imposée. Dans cette section, une estimation robuste en deux étapes est proposée selon laquelle nous commençons par prédire σ_t^2 par le filtre de Kalman en se basant sur la structure d’autocovariance établie dans §2.2.2. Ensuite, les valeurs prédites de σ_t^2 sont tronquées de manière à ce que la condition de positivité de la variance conditionnelle devienne une contrainte que l’on renforce indépendamment des signes des paramètres.

2.3.1 Prédiction

Soient $\hat{h}_{t|t-1}$, $\hat{h}_{t|t}$, $P_{t|t-1}$ et $P_{t|t}$ respectivement l’estimation prédite et filtré de h_t ainsi que leurs matrices d’erreur respectives. À partir des observations $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$, le filtre de Kalman est appliqué via l’algorithme 2.1.

Mentionnons que l’état initial de h_t est donné par sa moyenne :

$$\mathbb{E}h_0 = \frac{\omega}{1 - \sum_{i=1}^p \alpha_i - \sum_{j=1}^q \beta_j} {}^t(1 \dots 1) \in \mathbb{R}^r$$

et sa matrice d’autocovariance $\mathbb{E}\{(h_0 - \mathbb{E}h_0) {}^t(h_0 - \mathbb{E}h_0)\}$ déterminée par la proposition 2.3.

Par suite, l’exécution récursive des étapes de l’algorithme de Kalman permet d’extraire une estimation de σ_t^2 telle que pour tout $t = 1, \dots, n$:

$$\hat{\sigma}_t^2 = \hat{\sigma}_{t|t-1}^2 = e_r \hat{h}_{t|t-1}$$

Notons que cette extraction permet de récupérer la dimension univariée de σ_t^2 quel que soit l’ordre du GARCH. Par conséquent, les r contraintes de positivité des éléments de h_t sont réduites à celle de σ_t^2 .

Algorithme 2.1 Prédiction de la variance conditionnelle du GARCH par le filtre de Kalman.

Initialisation $\hat{h}_{0|0} = \mathbb{E}h_0$ et $P_{0|0} = \mathbb{E}\{(h_0 - \mathbb{E}h_0)^t(h_0 - \mathbb{E}h_0)\}$

Pour $t = 1, \dots, n$

Étape 1 : Équations de prédiction

$$\begin{aligned}\hat{h}_{t|t-1} &= \underline{\omega} + \Lambda \hat{h}_{t-1|t-1} \\ P_{t|t-1} &= \Lambda P_{t-1|t-1} {}^t\Lambda + \nu \Phi {}^t\Phi\end{aligned}$$

Étape 2 : Gain de Kalman

$$K_t = P_{t|t-1} {}^t e_r (e_r P_{t|t-1} {}^t e_r + \nu)^{-1}$$

Étape 3 : Équations de filtration

$$\begin{aligned}\hat{h}_{t|t} &= \hat{h}_{t|t-1} + K_t (\varepsilon_t^2 - e_r \hat{h}_{t|t-1}) \\ P_{t|t} &= P_{t|t-1} (I_r - {}^t e_r {}^t K_t)\end{aligned}$$

Fin Pour

Remarque 2.4.

La condition de stationnarité au second ordre de h_t (Proposition 2.1) assure la stabilité du filtre (2.1) vu que le système (2.4) est invariant dans le temps (voir le théorème 3.1, p.70 dans ANDERSON et MOORE (2012)).

2.3.2 Robustification

La robustification consiste à réestimer $\hat{\sigma}_{t|t-1}^2$ de manière à ce que la contrainte de positivité de la variance conditionnelle est directement renforcée indépendamment des signes des paramètres. Pour ce faire, un filtre de Kalman sous contrainte via la méthode de troncature de la fonction de densité (D. SIMON (2006)) est utilisée comme décrit par l'algorithme 2.2. Comme son nom l'indique, cette méthode consiste à tronquer la fonction de densité de probabilité (supposée normale) de $\{\hat{\sigma}_{t|t-1}^2\}_{t=1, \dots, n}$ sur un intervalle de positivité que l'on définit par une borne supérieure N choisie telle que pour tout $t = 1, \dots, n$, $\frac{1}{N} \leq \sigma_t^2 \leq N$. Cependant, à cause du caractère déterministe de telle contrainte, le choix empirique de N lié à la positivité de σ_t^2 reste arbitraire voir fallacieux en absence d'information de bornitude a priori sur la volatilité. Afin de contrôler un tel choix, nous explorons une contrainte de positivité à caractère aléatoire à travers une borne $N_{1-\tau}$ telle que pour un niveau de confiance $1-\tau$ et pour tout $t = 1, \dots, n$, on a :

$$1 - \tau = \mathbb{P} \left\{ \frac{1}{N_{1-\tau}} \leq \sigma_t^2 \leq N_{1-\tau} \right\} \quad (2.14)$$

Algorithme 2.2 Robustification de la variance conditionnelle prédite par troncature de la densité de probabilité.

Pour $t = 1, \dots, n$

Étape 1 : Générer une observation $\sigma_{t|t-1}^2$ suivant la loi $N(\widehat{\sigma}_{t|t-1}^2, p_{t|t-1})$, où $p_{t|t-1}$ représente le premier élément diagonale de $P_{t|t-1}$.

Étape 2 : Centrer-réduire $\sigma_{t|t-1}^2$ en effectuant la transformation :

$$\Sigma_t = \frac{\sigma_{t|t-1}^2 - \widehat{\sigma}_{t|t-1}^2}{\sqrt{p_{t|t-1}}}$$

Étape 3 : Fixer $\tau = 0.5\%$, d'où $N_{1-\tau} = Z_{0.995}\sqrt{p_{t|t-1}} + \widehat{\sigma}_{t|t-1}^2$. $Z_{0.995}$ représente le quantile d'ordre $1 - \tau$ de la loi normale centrée réduite.

Étape 4 : Calculer les deux bornes de la contrainte de positivité :

$$l_t = \frac{1 - N_{1-\tau}\widehat{\sigma}_{t|t-1}^2}{N_{1-\tau}\sqrt{p_{t|t-1}}} \quad \text{et} \quad u_t = \frac{N_{1-\tau} - \widehat{\sigma}_{t|t-1}^2}{\sqrt{p_{t|t-1}}}$$

Étape 5 : Calculer la densité normale tronquée de Σ_t , normalisée entre l_t et u_t , à savoir,

$$\tilde{f}(x) = c \exp(-x^2/2) \mathbf{1}_{[l_t, u_t]}(x)$$

où $c = \frac{\sqrt{2}}{\sqrt{\pi} [\text{erf}(u_t/\sqrt{2}) - \text{erf}(l_t/\sqrt{2})]}$ et $\text{erf}(\cdot)$ représente l'erreur de troncature.

Étape 6 : Calculer $\mu_\Sigma = \mathbb{E}_{\tilde{f}}(\Sigma_t)$ et déduire la variance conditionnelle obtenus après le renforcement de la contrainte (2.14) telle que :

$$\tilde{\sigma}_{t|t-1}^2 = \sqrt{p_{t|t-1}} \mu_\Sigma + \widehat{\sigma}_{t|t-1}^2 \quad (2.15)$$

Fin Pour

2.4 Illustration numérique

Dans cette section, nous menons par méthode Monte Carlo trois illustrations numériques d'estimation par quasi-maximum de vraisemblance des paramètres du GARCH relaxé dont la variance conditionnelle est estimée comme donnée dans 2.3. L'objectif est d'examiner la performance de l'estimation robuste de la variance conditionnelle relativement d'une part aux ordres du GARCH et d'autre part, à la présence de certaines informations sur la volatilité, notamment celles de bornitude.

Par ailleurs, l'estimation robuste de σ_t^2 basée sur la contrainte de positivité (2.14) est désormais utilisée pour la construction de la fonction de vraisemblance. Ainsi, à partir des observations $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n\}$, la fonction de la log-vraisemblance estimée,

notée $\widehat{\mathcal{L}}_n$, est définie pour tout $\theta \in \Theta$ par :

$$\widehat{\mathcal{L}}_n(\theta; \varepsilon_1, \varepsilon_2, \dots, \varepsilon_n) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \left(\frac{1}{n} \sum_{t=1}^n \frac{\varepsilon_t^2}{\widetilde{\sigma}_{t|t-1}^2(\theta)} + \log(\widetilde{\sigma}_{t|t-1}^2(\theta)) \right) \quad (2.16)$$

où $\theta = {}^t(\omega, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q)$ est le vecteur des paramètres et Θ est une partie de $\mathbb{R}^{*+} \times \mathbb{R}^{p+q}$ satisfaisant **(A1)** et **(A2)**.

Ainsi, maximiser $\widehat{\mathcal{L}}_n$ est équivalent à minimiser le critère \widehat{l}_n suivant :

$$\widehat{l}_n(\theta; \varepsilon_1, \varepsilon_2, \dots, \varepsilon_n) = \frac{1}{n} \sum_{t=1}^n \frac{\varepsilon_t^2}{\widetilde{\sigma}_{t|t-1}^2(\theta)} + \log(\widetilde{\sigma}_{t|t-1}^2(\theta))$$

Nous appelons l'estimation par QCK (**Q**uasi-maximum de vraisemblance par application du filtre de **K**alman sous **C**ontrainte), l'estimation via la minimisation de \widehat{l}_n évaluée en fonction de la variance conditionnelle robustifiée. Dans ce qui suit, nous utilisons la notation QML pour l'estimation standard du quasi-maximum de vraisemblance dont la moyenne Monte Carlo est notée $\widehat{\theta}$ et par $\widetilde{\theta}$ la moyenne Monte Carlo obtenue par QCK.

Dans toutes les simulations, on fixe le nombre des réplifications pour chaque taille d'échantillon à 1000 réplifications. La fonction *optim.* implémentée dans R (STEENBERGEN (2006)) est utilisée pour la minimisation de \widehat{l}_n . À des fins de comparaison, l'erreur quadratique moyenne (MSE) est utilisée pour mesurer l'écart entre les valeurs réelles des paramètres et ses valeurs estimées.

2.4.1 Estimation des paramètres du GARCH relaxé

Pour chaque taille d'échantillon $n \in \{500, 1000, 5000\}$, nous avons simulé trois processus GARCH(1,1), GARCH(3,1) et GARCH(2,3) respectivement de paramètres ${}^t(1.5, 0.3, 0.2)$, ${}^t(1.5, 0.1, 0.2, 0.1, 0.4)$ et ${}^t(1.5, 0.1, 0.2, 0.1, 0.4, 0.1)$. Le processus η_t est supposé normal centré réduit. Notons que le premier GARCH a une persistance de volatilité assez faible valant 0.5 relativement aux deux derniers GARCH étant à volatilité fortement persistante valant 0.8 pour le GARCH(3,1) et 0.9 pour le GARCH(2,3).

Les tableaux 2.1, 2.2 et 2.3 montrent que les paramètres sont mieux estimés par la méthode QCK basée sur le GARCH relaxé, comparativement à l'estimation par QML basée sur le modèle GARCH standard. En effet, pour tous les paramètres, l'estimation par QCK entraîne une baisse du MSE. Notons également que les faibles écarts types de $\widetilde{\theta}$ par rapport à celles de $\widehat{\theta}$ indiquent que les estimations sur l'ensemble des réplifications sont moins dispersées autour de $\widetilde{\theta}$ que autour de $\widehat{\theta}$. Autrement dit, la méthode proposée est plus stable autour des estimations QCK.

Par ailleurs, on voit bien que pour toutes les tailles n , l'estimation par QCK a capté le niveau de persistance de la volatilité caractérisant les séries GARCH simulées, en enregistrant les valeurs les plus proches des vraies valeurs de persistance fixées lors des simulations comparativement aux valeurs résultant de l'estimation par QML.

TABLE 2.1 – Propriétés en échantillons finis des modèles r -GARCH(1,1) et GARCH(1,1) estimés respectivement par QCK et QML.

			r -GARCH(1,1)		GARCH(1,1)	
			QCK		QML	
n			Moyenne	MSE	Moyenne	MSE
500	ω	1.5	1.5026 (0.0096)	< 0.0001*	1.5696 (0.4061)	0.1698
	α_1	0.3	0.3056 (0.0082)	< 0.0001*	0.3059 (0.0653)	0.0043
	β_1	0.2	0.1976 (0.0097)	< 0.0001*	0.1752 (0.1363)	0.0192
1000	ω	1.5	1.4999 (0.0099)	< 0.0001*	1.5135 (0.2789)	0.0780
	α_1	0.3	0.3050 (0.0086)	< 0.0001*	0.3022 (0.0556)	0.0031
	β_1	0.2	0.1995 (0.0099)	< 0.0001*	0.1852 (0.1126)	0.0129
5000	ω	1.5	1.5016 (0.0098)	< 0.0001*	1.5075 (0.1504)	0.0227
	α_1	0.3	0.3063 (0.0077)	< 0.0001*	0.3011 (0.0264)	0.0007
	β_1	0.2	0.2017 (0.0098)	< 0.0001*	0.1969 (0.0573)	0.0033

¹ Les écarts types estimés sont indiqués entre parenthèses. Les plus petites MSE sont marquées par (*).

Maintenant, nous complétons l'évaluation de la performance de l'estimation des paramètres par QCK en considérant trois trajectoires de taille $n \in \{500, 1000, 5000\}$, générées chacune par les GARCH utilisés précédemment. À partir des figures 2.1 et 2.2, on peut voir que pour toutes les séries simulées, la volatilité estimée par QCK est quasiment superposée sur la volatilité simulée. Par contre, l'écart important entre les volatilités simulées et estimées par QML est clairement visible. Ces résultats sont donc conformes aux résultats rapportés dans le tableau 2.3 où l'on observe que la volatilité estimée par QCK enregistre les valeurs les plus faibles du MSE relativement à la volatilité estimée par QML.

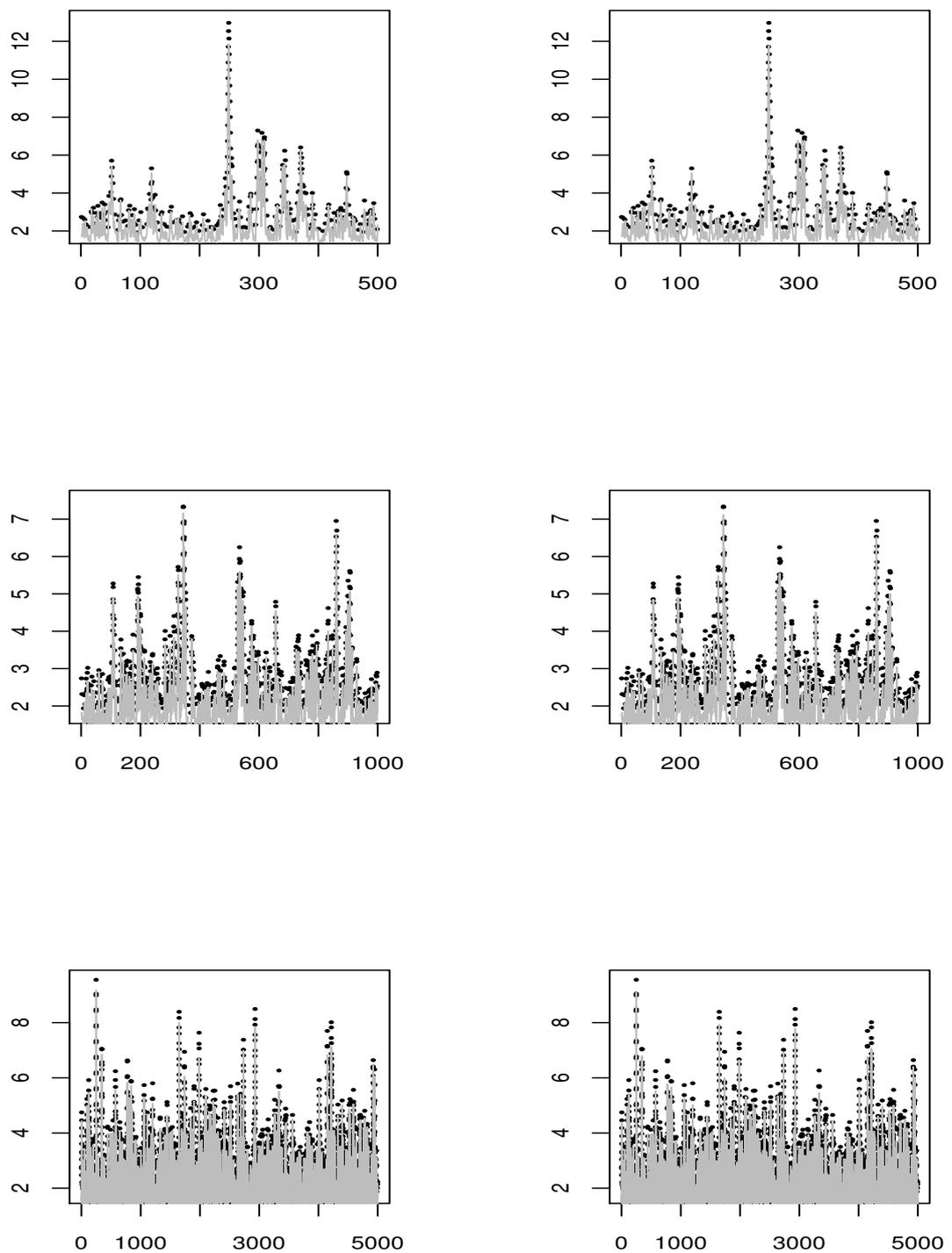


FIGURE 2.1 – Volatilités estimées par QCK (à gauche) et par QML (à droite) pour une simulation de taille $n=500$, 1000 et 5000 du GARCH(1,1). Le graphe en noire indique la volatilité originale simulée.

TABLE 2.2 – Propriétés en échantillons finis des modèles r -GARCH(3,1) et GARCH(3,1) estimés respectivement par QCK et QML.

		r -GARCH(3,1)			GARCH(3,1)	
		QCK		MSE	QML	
n			Moyenne	MSE	Moyenne	MSE
500	ω	1.5	1.5002 (0.0264)	0.0007*	1.4964 (0.4812)	0.2316
	α_1	0.1	0.0995 (0.0574)	0.0033*	0.1823 (0.2245)	0.0572
	α_2	0.2	0.1953 (0.0753)	0.0057*	0.1330 (0.2757)	0.0805
	α_3	0.1	0.0963 (0.0478)	0.0023*	0.0413 (0.1332)	0.0212
	β_1	0.4	0.4002 (0.0888)	0.0079*	0.5678 (0.3222)	0.1320
1000	ω	1.5	1.4972 (0.0932)	0.0087*	1.5478 (0.2573)	0.0685
	α_1	0.1	0.0968 (0.1009)	0.0102*	0.1034 (0.1956)	0.0383
	α_2	0.2	0.1950 (0.0942)	0.0089*	0.1955 (0.2887)	0.0834
	α_3	0.1	0.0921 (0.1026)	0.0106*	0.1416 (0.1251)	0.0174
	β_1	0.4	0.3974 (0.0989)	0.0098*	0.2654 (0.2865)	0.1002
5000	ω	1.5	1.5002 (0.0830)	0.0069*	1.6172 (0.1051)	0.0248
	α_1	0.1	0.1013 (0.0888)	0.0079*	0.0829 (0.1323)	0.0178
	α_2	0.2	0.1926 (0.0945)	0.0090*	0.2123 (0.2935)	0.0863
	α_3	0.1	0.0960 (0.0853)	0.0073*	0.1222 (0.1170)	0.0142
	β_1	0.4	0.3981 (0.0830)	0.0069*	0.3599 (0.2935)	0.0878

¹ Les écarts types estimés sont indiqués entre parenthèses. Les plus petites MSE sont marquées par (*).

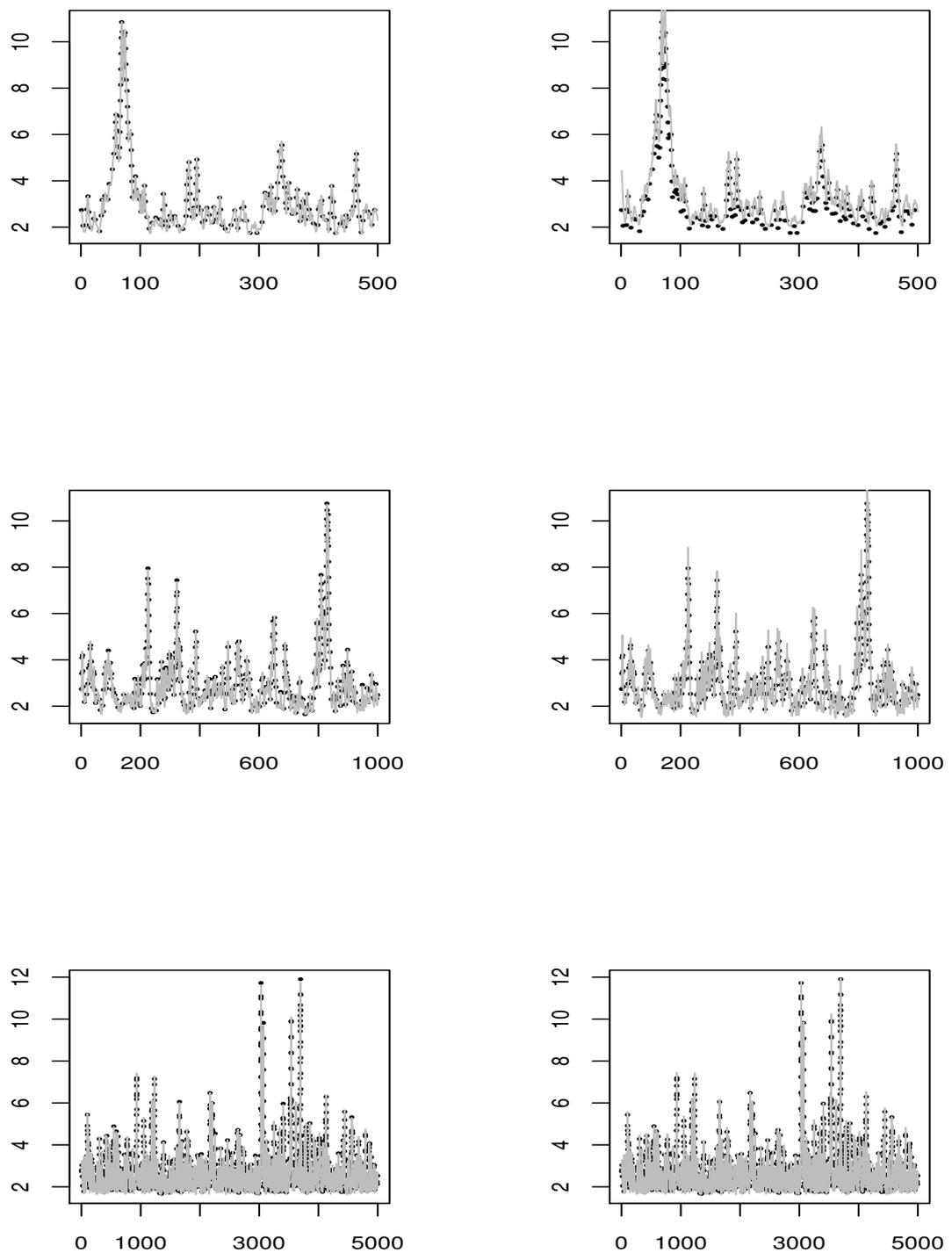


FIGURE 2.2 – Volatilités estimées par QCK (à gauche) et par QML (à droite) pour une simulation de taille $n=500$, 1000 et 5000 du GARCH(3,1). Le graphe en noire indique la volatilité originale simulée.

TABLE 2.3 – Propriétés en échantillons finis des modèles r -GARCH(2,3) et GARCH(2,3) estimés respectivement par QCK et QML.

			r -GARCH(2,3)		GARCH(2,3)	
			QCK		QML	
n			Moyenne	MSE	Moyenne	MSE
500	ω	1.5	1.5006 (0.0727)	0.0053*	1.5062 (0.0857)	0.0074
	α_1	0.1	0.0998 (0.0989)	0.0098*	0.0885 (0.2221)	0.0495
	α_2	0.2	0.1987 (0.0574)	0.0033*	0.0129 (0.0222)	0.0355
	β_1	0.1	0.1006 (0.0299)	0.0009*	0.1223 (0.2729)	0.0750
	β_2	0.4	0.4001 (0.0299)	0.0009*	0.1563 (0.2970)	0.1476
	β_3	0.1	0.1005 (0.0264)	0.0007*	0.1890 (0.3973)	0.1658
1000	ω	1.5	1.5019 (0.0754)	0.0057*	1.4814 (0.1164)	0.0139
	α_1	0.1	0.0989 (0.0632)	0.0040*	0.0901 (0.1212)	0.0148
	α_2	0.2	0.1993 (0.0678)	0.0046*	0.0149 (0.3882)	0.1850
	β_1	0.1	0.1009 (0.0519)	0.0027*	0.0757 (0.3667)	0.1351
	β_2	0.4	0.4303 (0.0942)	0.0098*	0.1430 (0.2373)	0.1224
	β_3	0.1	0.0991 (0.0399)	0.0016*	0.1643 (0.2320)	0.0580
5000	ω	1.5	1.5016 (0.0647)	0.0042	1.5280 (0.0389)	0.0023*
	α_1	0.1	0.0987 (0.0479)	0.0023*	0.0977 (0.0573)	0.0033
	α_2	0.2	0.1980 (0.0871)	0.0076*	0.0102 (0.3920)	0.1897
	β_1	0.1	0.0995 (0.0399)	0.0016*	0.0256 (0.4359)	0.1956
	β_2	0.4	0.3987 (0.0734)	0.0054	0.3990 (0.0479)	0.0023*
	β_3	0.1	0.0992 (0.0399)	0.0016*	0.1405 (0.0957)	0.0108

¹ Les écarts types estimés sont indiqués entre parenthèses. Les plus petites MSE sont marquées par (*).

TABLE 2.4 – Précision des volatilités estimées par QCK et QML pour les processus GARCH(1,1), GARCH(3,1) et GARCH(2,3).

n	GARCH(1,1)		GARCH(3,1)		GARCH(2,3)	
	MSE($\tilde{\sigma}_t$)	MSE($\hat{\sigma}_t$)	MSE($\tilde{\sigma}_t$)	MSE($\hat{\sigma}_t$)	MSE($\tilde{\sigma}_t$)	MSE($\hat{\sigma}_t$)
500	0.0010	0.0918	0.0090	0.1105	0.0025	0.2050
1000	0.0023	0.1004	0.0170	0.1050	0.0014	0.1979
5000	0.0031	0.1006	0.0105	0.1068	0.0048	0.2140

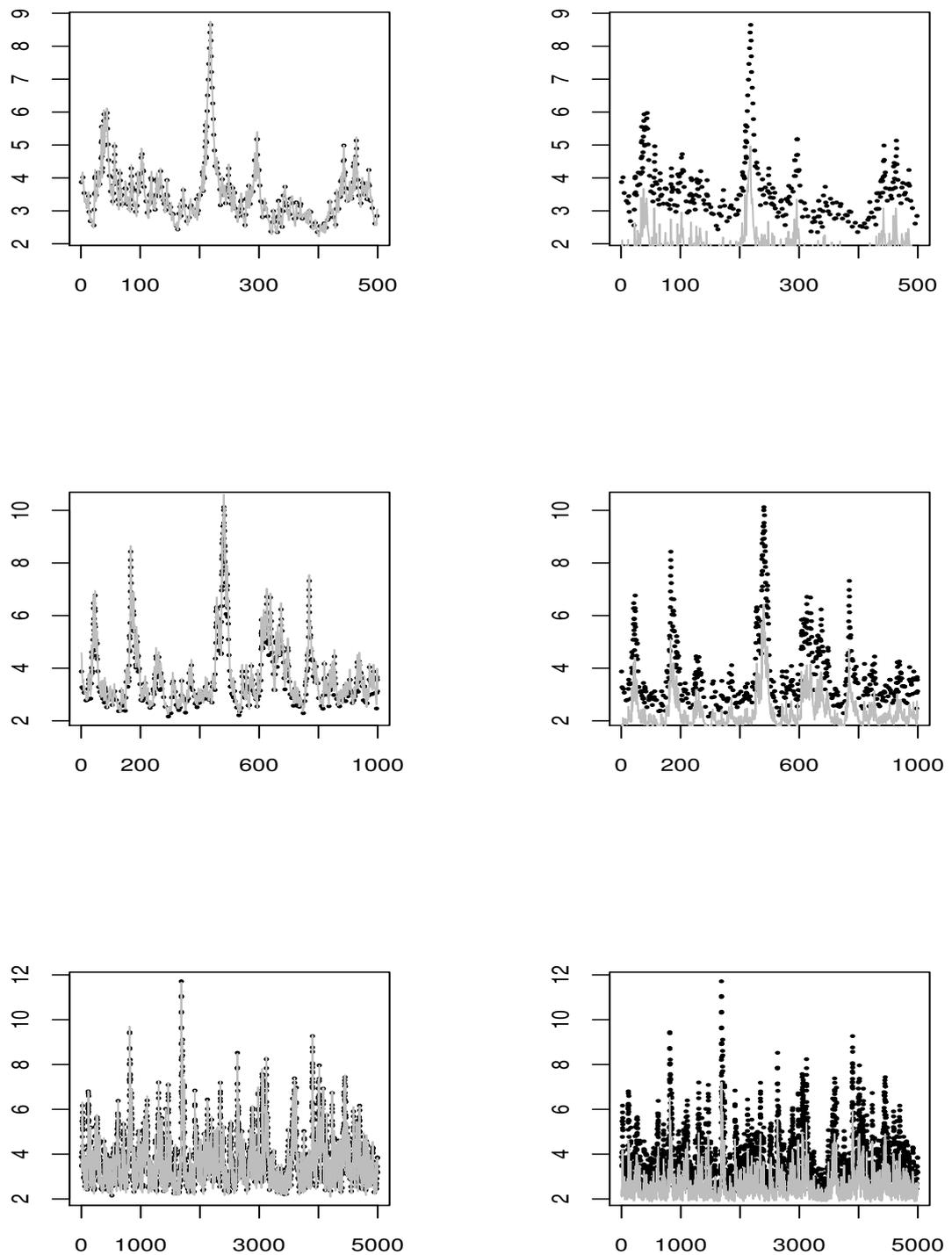


FIGURE 2.3 – Volatilités estimées par QCK (à gauche) et par QML (à droite) pour une simulation de taille $n=500$, 1000 et 5000 du GARCH(2,3). Le graphe en noir indique la volatilité originale simulée.

2.4.2 Redressement des estimations en présence d'information sur la volatilité

Nous nous intéressons à une situation particulière où la spécification de la variance conditionnelle est enrichie à priori par une information plus fine sur la bornitude de la volatilité que celle de positivité étant triviale. Notons que la contrainte (2.14) peut changer selon l'information disponible sur la volatilité vu que la méthode de troncature permet de prendre en considération toutes les formes possibles de bornitude. L'objectif est donc de mettre en évidence l'effet d'existence de telle information dans le redressement des estimations des paramètres en utilisant l'estimation QCK basée sur la pré-estimation robuste de la variance conditionnelle.

Dans ce sens, nous effectuons une étude par simulation de taille $n = 1000$ d'un processus GARCH(1,1) de paramètres $\omega_0 = 1.5$, $\alpha_0 = 0.4$ et $\beta_0 = 0.1$. Nous supposons qu'une information à priori sur la volatilité est disponible et donnée pour tout $t = 1, \dots, n$ comme suit :

$$\frac{1}{K} + \sigma_{t,0} - 1 \leq \sigma_t \leq K + \sigma_{t,0} - 1 \quad (2.17)$$

où $\sigma_{t,0}$ est la volatilité réelle³ du GARCH(1,1) généré et $K \geq 1$.

Remarquons que les bornes de (2.17) déterminent le niveau d'information sur σ_t relativement à $\sigma_{t,0}$. En effet, σ_t se rapproche de la valeur réelle $\sigma_{t,0}$ quand $K \rightarrow 1$. En d'autres termes, plus K se rapproche de 1, plus l'information sur σ_t est précise et inversement, plus la valeur de K est élevée, moins cette information est précise.

En posant $K = 1 + 10^{-i}$, $i = 1, \dots, 10$, nous mettons en évidence l'effet de la précision de l'information mesurée par i sur la qualité des estimations des paramètres en ajustant un GARCH(1,1) en utilisant l'algorithme QCK et nous observons l'erreur absolue $\delta(\cdot)$ des estimations obtenues. Le tableau 2.5 ainsi que la figure 2.4 montrent que la précision de l'estimation est améliorée en exploitant l'information disponible (2.17). En effet, l'erreur absolue des estimations des paramètres tend vers 0 dès que $i \geq 8$. En d'autres termes, plus l'information sur la volatilité est précise, mieux les paramètres sont estimés.

3. En pratique $\sigma_{t,0}$ peut être remplacée par la volatilité réalisée sous l'hypothèse de stabilité de la volatilité sur des périodes récentes.

TABLE 2.5 – Estimations des paramètres par QCK en fonction du niveau d'information i .

i	$\tilde{\omega}$	$\delta(\tilde{\omega})$	$\tilde{\alpha}_1$	$\delta(\tilde{\alpha}_1)$	$\tilde{\beta}_1$	$\delta(\tilde{\beta}_1)$
1	1.6251	0.1251	0.0315	0.3684	0.1630	0.0630
2	1.4564	0.0435	0.3215	0.0784	0.2064	0.1064
3	1.4897	0.0102	0.2395	0.1604	0.1091	0.0091
4	1.2805	0.2194	0.4003	0.0003	0.1003	0.0003
5	1.6512	0.1512	0.4198	0.0198	0.0870	0.0129
6	1.5391	0.0391	0.3755	0.0244	0.0056	0.0943
7	1.4031	0.0968	0.4968	0.0968	0.1123	0.0123
8	1.5000	0.0000	0.4000	0.0000	0.1000	0.0000
9	1.5000	0.0000	0.4000	0.0000	0.1000	0.0000
10	1.5000	0.0000	0.4000	0.0000	0.1000	0.0000

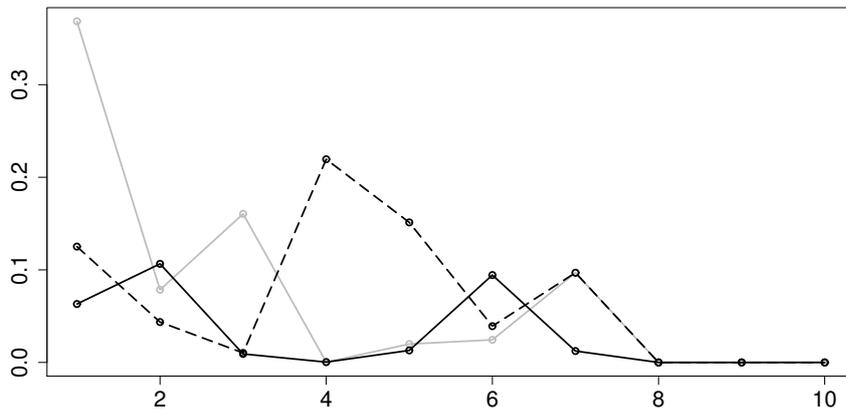


FIGURE 2.4 – L'erreur absolue δ des estimations par QCK en fonction de la précision i . La ligne pointillée indique $\delta(\tilde{\omega})$, les lignes continues grises et noires représentent respectivement $\delta(\tilde{\alpha}_1)$ et $\delta(\tilde{\beta}_1)$.

2.5 Conclusion

En se basant sur le filtre de Kalman sous contrainte, nous nous sommes focalisés sur la pré-estimation robuste de la variance conditionnelle générée par le modèle GARCH tout en restant positive, non seulement en relaxant les contraintes de positivité de Bollerslev, mais aussi sans la détermination à priori des contraintes moins restrictives sur les paramètres du modèle comme la préconise l'approche de Nelson et Cao. La méthode proposée a montré sa capacité d'améliorer les estimations de quasi-maximum de vraisemblance, y compris pour des ordres élevés. De plus, les résultats numériques ont montré l'effet de redressement que les informations additionnelles sur la volatilité peuvent avoir sur les estimations des paramètres en appliquant la pré-estimation proposée.

Estimation des paramètres des modèles r-GARCH et CGARCH : Étude de convergence

L'un des aspects de robustesse de la pré-estimation de la variance conditionnelle générée par le GARCH(p, q) relaxé vu au chapitre précédent est qu'elle est applicable quels que soient les ordres $p \geq 1$ et $q \geq 1$. En conséquence, l'estimation des paramètres du modèle par la méthode QCK telle qu'elle est présentée dans les illustrations numériques du chapitre précédent, est fortement généralisable quel que soit le nombre des paramètres. Cependant, pour des modèles GARCH d'ordres élevés, l'estimation standard du QML via la maximisation analytique basée sur le calcul des dérivées de la fonction de vraisemblance est moins pratique sur le plan du calcul. Nous transposons dans ce chapitre l'estimation QCK au contexte d'optimisation numérique stochastique en utilisant l'algorithme d'approximation stochastique à perturbation simultanée (SPSA) (SPALL (1992); SPALL (1997); SPALL (1998); SPALL (2005)). Il s'agit d'une classe d'algorithmes qui reposent sur une approximation du gradient de la fonction de vraisemblance indépendamment du nombre des paramètres en évitant ainsi les procédures de dérivation explicite.

Approximation Stochastique à Perturbation Simultanée

Le cadre fondamentale général de l'approximation stochastique a été introduit par ROBBINS et MONRO (1951) comme une méthode récursive de recherche des zéros d'une fonction objectif à partir de ses mesures bruitées. Dans la même veine, KIEFER et WOLFOWITZ (1952) ont proposé une estimation du gradient quand l'algorithme de Robbin et Monro est à direction descente du gradient. L'algorithme de Kiefer-Wolfowitz repose sur la génération d'un nombre suffisant de mesures de la fonction objectif en perturbant simultanément toutes les composantes du vecteur des paramètres, ce qui fait un nombre de mesures valant le double de la dimension du vecteur des paramètres. SPALL (1992) a introduit un estimateur remarquable du gradient qui nécessite seulement deux évaluations de la fonction objectif quel

que soit le nombre des paramètres. Cet estimateur est basé sur la perturbation simultanée de toutes les composantes du vecteur des paramètres à l'aide de variables aléatoires *iid* satisfaisant certaines propriétés qui sont le plus souvent satisfaites par des variables aléatoires uniforme ou de Bernoulli. L'algorithme de Robbins-Monro en conjonction avec cet estimateur est devenu célèbre dans la littérature sous le nom de l'algorithme d'approximation stochastique à perturbation simultanée (SPSA). L'application de l'approximation stochastique pour obtenir des estimations du QML est dû à SPALL (1987), en fournissant les conditions de convergence de tel algorithme appliqué à la fonction de vraisemblance. Quant au modèle GARCH, l'estimation des paramètres du modèle par QML combinée avec l'algorithme SPSA a été présentée par ALLAL et BENMOUMEN (2011) pour le cas $p = q = 1$, en se limitant au modèle standard, i.e., sous l'hypothèse de positivité des paramètres, sans aucun exercice d'analyse de convergence de l'algorithme utilisé. Ainsi, nous employons dans ce chapitre l'algorithme du SPSA dans l'estimation par QCK des paramètres du modèle GARCH afin d'améliorer l'optimalité des estimations relatives à l'estimation relaxée de NELSON et CAO (1992) §3.1.1. C'est ce qui a été effectivement confirmé par l'illustration numérique proposée §3.1.2.

Estimation du CGARCH (Component GARCH)

Dans la suite de ce chapitre, nous nous penchons sur l'estimation des paramètres du CGARCH(N) par la méthode QCKSA qui désigne désormais l'estimation par QCK basée sur l'algorithme SPSA. Premièrement, nous nous intéressons à prouver la convergence de l'algorithme du SPSA appliqué à la fonction de vraisemblance du CGARCH §3.2.1.

Dans quel contexte donc la méthode QCKSA intersecte-t-elle l'estimation du modèle CGARCH ? En effet, notons que :

- (i) Le CGARCH appartient en premier lieu à la famille des GARCH étant un GARCH(N, N) restreint dont les paramètres sont donnés en fonction des paramètres des N composantes GARCH(1,1) (DING et GRANGER (1996)). D'où, la méthode QCKSA est aussi valable pour l'estimation du CGARCH que le GARCH.
- (ii) Sachant que plus la variance conditionnelle du CGARCH est décomposée en plusieurs volatilités, plus le modèle est significatif en captant mieux la dynamique de la volatilité, le nombre des paramètres du modèle devient plus élevé. D'où l'intérêt de l'algorithme SPSA dans l'estimation du CGARCH indépendamment de sa dimension.
- (iii) Le CGARCH n'est pas seulement soumis aux contraintes de positivité des paramètres de Bollerslev assurant la positivité de sa variance conditionnelle, mais ces contraintes peuvent être ex-post relaxées au sens de NELSON et CAO (1992) après la dérivation de la spécification GARCH associée. D'où se voit l'utilité de l'estimation robuste de la variance conditionnelle vue au Chapitre 2 au sens d'imposer directement la positivité de la variance conditionnelle sans contraindre à priori les paramètres du modèle.

Ainsi, dans le cadre de l'estimation relaxée du GARCH (i), étant donné l'ordre élevé possible du CGARCH qui résulte de (ii) ainsi que l'identification des contraintes sur les paramètres relaxant le CGARCH comme mentionné dans (iii), il s'ensuit que la détermination à priori de l'espace des paramètres dans lequel l'algorithme SPSA converge doit être faisable d'une manière à éviter la déduction ex-post de l'espace des paramètres découlant de la spécification GARCH du modèle CGARCH. Nous montrons dans ce chapitre que la construction d'un tel espace est toujours faisable comme le produit des espaces des paramètres de chaque composante GARCH(1,1) relaxé §3.2.1. Ces résultats sont mis en évidence via une illustration empirique §3.2.2.

3.1 Estimation par QCKSA du modèle GARCH relaxé

3.1.1 Maximisation de la vraisemblance par SPSA

Rappelons qu'une estimation par QCK des paramètres du modèle GARCH(p, q) relaxé est obtenue en maximisant la fonction du log-vraisemblance $\widehat{\mathcal{L}}_n$ évaluée en fonction des estimations robuste de σ_t^2 étant données les observations $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n\}$. Cela revient à minimiser le critère \widehat{l}_n défini pour tout $\theta \in \Theta$ par :

$$\widehat{l}_n(\theta; \varepsilon_1, \varepsilon_2, \dots, \varepsilon_n) = \frac{1}{n} \sum_{t=1}^n \frac{\varepsilon_t^2}{\widetilde{\sigma}_{t|t-1}^2(\theta)} + \log(\widetilde{\sigma}_{t|t-1}^2(\theta)) \quad (3.1)$$

où $\theta = {}^t(\omega, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q)$ est le vecteur des paramètres et Θ est une partie de $\mathbb{R}^{*+} \times \mathbb{R}^{p+q}$ satisfaisant **(A1)** et **(A2)**.

Le choix de la routine de minimisation de \widehat{l}_n doit prendre en considération le processus de construction de la log-vraisemblance $\widehat{\mathcal{L}}_n$ basée sur la pré-estimation robuste de σ_t^2 . En particulier, ce choix doit être orienté par le cadre de relaxation de l'estimation du GARCH, notamment en termes de la dimension du modèle (le nombre des paramètres). Dans ce sens, notons que :

- (i) La pré-estimation robuste de σ_t^2 avant d'être insérée dans \widehat{l}_n , attribuée à la mesure de ce dernier un caractère aléatoire provenant du filtre de Kalman utilisé, ce qui diffère de l'évaluation traditionnelle de la vraisemblance associée au GARCH standard où σ_t^2 est calculée d'une manière récursive à travers l'équation de la variance conditionnelle relativement à une valeur initiale déterministe σ_0^2 (FRANCQ (2010)).
- (ii) Dans notre contexte consistant à généraliser l'estimation des paramètres du GARCH pour des ordres élevés, le recours aux méthodes d'optimisation analytiques basées sur le calculs des dérivées de \widehat{l}_n (gradient, Hessien) est remis en question relativement au nombre des dérivées demandé ainsi que le système des équations à résoudre qu'elles génèrent (Robert F ENGLE (1982), BOLLERSLEV (1986)).
- (iii) Une méthode d'optimisation faiblement dépendante de la forme fonctionnelle de la vraisemblance, notamment celle qui résulte de l'hypothèse de normalité

de η_t est bien souhaitable en vertu des travaux de Christian FRANCO, LEPAGE et ZAKOIAN (2011), FAN, QI et XIU (2014) et ZHU et XIE (2016) qui proposent l'estimation généralisée des paramètres du GARCH par QML pour toute densité de η_t généralisant en conséquence les formes fonctionnelles possibles de la vraisemblance.

L'algorithme SPSA (Simultaneous Perturbation Stochastic Approximation) a été considéré avec intérêt majeur dans le cadre des méthodes d'optimisation stochastiques fortement préconisées pour la minimisation des fonctions objectifs construites sur plusieurs étapes comme c'est le cas dans (i) (BHATNAGAR, H. PRASHANTH et L. PRASHANTH (2013)). De plus, en répondant à (ii) et (iii), l'algorithme SPSA se base sur le calcul de la fonction objectif bruitée en deux mesures uniquement quel que soit le nombre des paramètres, en approximant son gradient sans passer par des calculs des dérivées comme dans le cas des méthodes déterministes. Cela permet de considérer d'autres formes fonctionnelles de la vraisemblance (iii) sans tenir compte de la complexité analytique de ses dérivées.

Ainsi, nous généralisons l'utilisation de la méthode SPSA pour le modèle GARCH relaxé pour $p, q \geq 1$ à travers l'algorithme ci-dessous en exploitant son applicabilité indépendante du nombre des paramètres. De plus, nous employons une configuration de la méthode SPSA à paramètre de perturbation fixe, i.e., $c_k = c > 0$, où c est une constante choisie d'une manière à assurer la convergence de l'algorithme SPSA vers un voisinage de l'ensemble des minimums locaux de \hat{l}_n . Nous adoptons l'algorithme SPSA selon la calibration de SPALL (1998) comme suit :

Algorithme 3.3 Minimisation par SPSA de la log-vraisemblance du GARCH.

Calibration : $a = 0.16$, $c = 0.2$, $\lambda = 0.602$, $\gamma = 0.101$ et $A = 10\%$ du nombre des itérations.

Initialisation : ϵ , la précision de convergence. $iter$, le nombre des itérations. $\theta_0 \in \Theta$.

Pour $k = 1$ à $iter$, **faire**

Étape 1 : Calculer la suite du gain $a_k = a(A + k + 1)^{-\lambda}$.

Étape 2 : Générer le vecteur de perturbation aléatoire $\Delta_k \in \mathbb{R}^{p+q+1}$ distribué selon la loi de Bernoulli à valeurs de ± 1 et de paramètre $\frac{1}{2}$.

Étape 3 : Vérifier que $\omega > 0$ et respectivement à partir de **(A1)** et **(A2)**, la stationnarité au second ordre et l'existence de $\mathbb{E}\varepsilon_t^4$. Sinon, retourner à l'étape 2 et prendre $\hat{\theta}_{k-1}$ comme point initial.

Étape 4 : Obtenir deux mesures $y_k^\pm(\hat{\theta}_k)$ de \hat{l}_n basées sur la perturbation simultanée autour de $\hat{\theta}_k$, comme suit :

$$y_k^\pm(\hat{\theta}_k) = \hat{l}_n(\hat{\theta}_k \pm c\Delta_k) + \delta_k^\pm$$

où δ_k^- et δ_k^+ sont deux vecteurs aléatoires indépendants distribués selon la loi uniforme sur l'intervalle $[0, 1]$.

Étape 5 : Obtenir l'approximation du gradient de l_n , $\hat{g}(\hat{\theta}_k)$:

$$\hat{g}(\hat{\theta}_k) = \frac{y_k^+(\hat{\theta}_k) - y_k^-(\hat{\theta}_k)}{2c} \text{ }^t (\Delta_{k1}^{-1}, \dots, \Delta_{k,p+q+1}^{-1}) \text{ où } \Delta_{ki}, i = 1, \dots, p + q + 1 \text{ est le } i^{eme} \text{ élément de } \Delta_k.$$

Étape 6 : Itération $k + 1$:

$$\hat{\theta}_{k+1} = \hat{\theta}_k - a_k \hat{g}(\hat{\theta}_k) \tag{3.2}$$

Étape 7 : Retourner à l'étape 2 avec $k = k + 1$.

Étape 8 : Retourner $\hat{\theta}_k$ si $\|\hat{\theta}_{k+1} - \hat{\theta}_k\| < \epsilon$ ou $k = iter$. Sinon $k = k + 1$.

Fin Pour

3.1.2 Illustration numérique

Dans cette section, nous menons une application empirique de l'estimation par QCKSA dont l'objectif est de comparer sa performance par rapport à l'estimation relaxée établie par NELSON et CAO (1992) en termes des optimums obtenus par chaque approche. L'application porte sur trois séries réelles utilisées par BAILLIE et BOLLERSLEV (2002) et réutilisées par NELSON et CAO (1992), à savoir le taux de change du mark Allemand/dollar, du franc Français/dollar et du yen Japonais/dollar à partir du 1er Juin 1973 au 28 Janvier 1985, faisant chacune une taille de 2920 observations¹. Ces séries ainsi que les séries des rendements correspondantes sont présentées respectivement par les figures 3.1, 3.2 et 3.3. Nous désignons par (s_t) la série du taux de change et par (r_t) la série du rendement correspondant, soit :

1. Ces données sont disponibles sur le site web www.federalreserve.gov.

TABLE 3.1 – Modèles GARCH estimés pour les taux de change journaliers du 1^{er} Juin 1973 au 28 Janvier 1985.

	mark allemand		franc français		yen japonais	
	GARCH(2,2)		GARCH(2,1)		GARCH(3,1)	
	QCKSA	QMLNC	QCKSA	QMLNC	QCKSA	QMLNC
μ	0.0028	-0.0009	0.0116	-0.0002	-0.0006	0.0016
ω	0.0182	0.0186	0.0021	0.0079	0.0111	0.0002
α_1	0.0576	0.0573	0.1680	0.1024	0.0088	0.1888
α_2	0.2265	0.2262	0.2449	0.1444	0.1988	0.0752
α_3					-0.1011	-0.2344
β_1	0.3829	0.3833	0.5846	0.7735	0.4011	0.9730
β_2	0.3096	0.3100				
log-Lik	-1462.94	-2447.24	-1459.63	-2356.21	-1659.94	-2086.27

¹ log-Lik est la valeur de la log-vraisemblance.

$$r_t = 100 \log(s_t/s_{t-1}), \quad t = 1, \dots, 2920$$

Nous suivons l'ajustement de (r_t) adopté par NELSON et CAO (1992) comme un GARCH non centré dont les ordres sont sélectionnés par le critère AIC, soit :

$$r_t = \mu + \varepsilon_t, \quad (\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_1) \sim N(0, \sigma_t^2)$$

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

Le tableau 3.1 rapporte les estimations des paramètres du modèle obtenues par la méthode QCKSA et par l'approche de NELSON et CAO (1992) que l'on note QMLNC. Il est facile de vérifier que les estimations obtenus par QCKSA satisfont les inégalités de NELSON et CAO (1992) données par les théorèmes 1.6 et 1.7 et relaxent en conséquence les contraintes de positivité des paramètres de Bollerslev. C'est bien le cas de α_3 dans l'ajustement de la série des rendements du taux de change yen/dollar. Or le plus qu'apporte l'utilisation du SPSA apparait au niveau des valeurs obtenues de la log-vraisemblance. En effet, il est clair que l'estimation par QCKSA maximise la log-vraisemblance relativement à l'estimation obtenue par QMLNC, ce qui montre la capacité de l'algorithme SPSA d'améliorer l'optimalité des estimations obtenues.

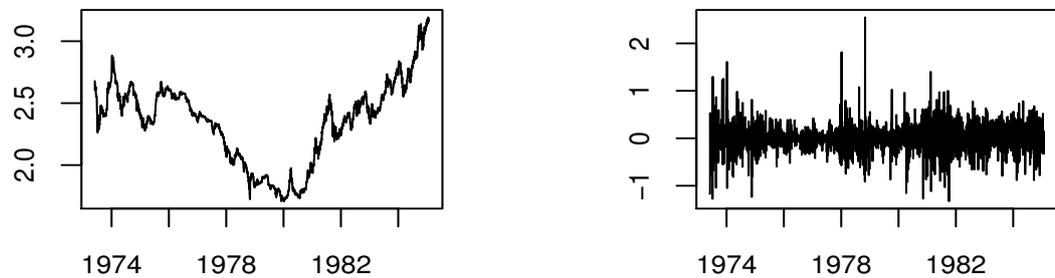


FIGURE 3.1 – Taux de change journalier du mark/dollar (à gauche) et la série des rendements correspondante (à droite) du 1^{er} Juin 1973 au 28 Janvier 1985.

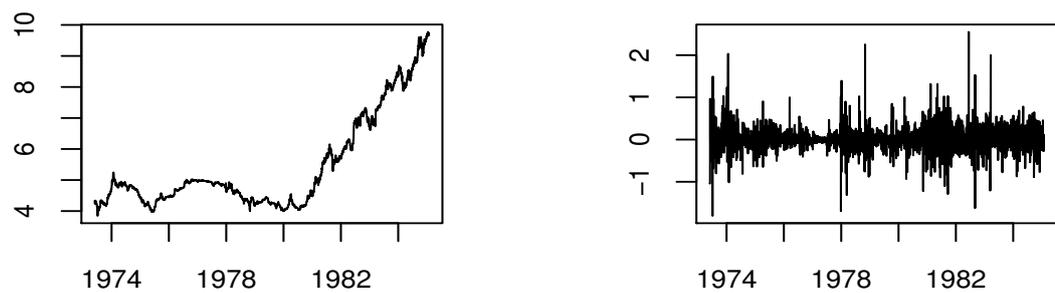


FIGURE 3.2 – Taux de change journalier du franc/dollar (à gauche) et la série des rendements correspondante (à droite) du 1^{er} Juin 1973 au 28 Janvier 1985.

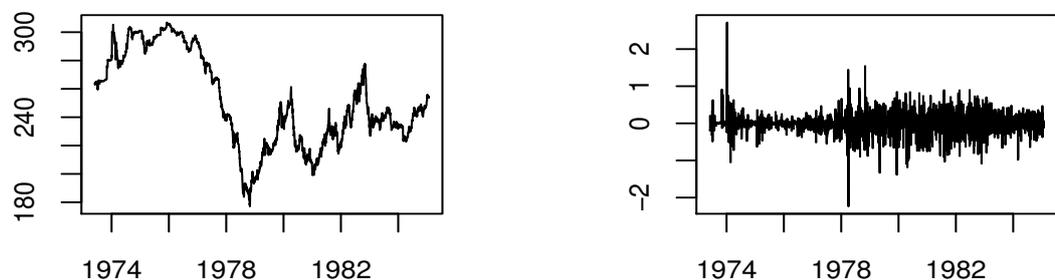


FIGURE 3.3 – Taux de change journalier du yen/dollar (à gauche) et la série des rendements correspondante (à droite) du 1^{er} Juin 1973 au 28 Janvier 1985.

3.2 Estimation du modèle CGARCH

Considérons le modèle CGARCH(N) sous la forme non pondérée présenté par MAHEU (2005), soit :

$$\begin{aligned}\varepsilon_t &= \left(\sum_{i=1}^N \sigma_{i,t}^2 \right)^{\frac{1}{2}} \eta_t = \sigma_t \eta_t, \quad \eta_t \sim iidN(0, 1) \\ \sigma_{i,t}^2 &= \omega_i + \alpha_i \varepsilon_{t-1}^2 + \beta_i \sigma_{i,t-1}^2, \quad i = 1, \dots, N\end{aligned}\quad (3.3)$$

Des contraintes de positivité de σ_t^2 sont nécessaires à imposer. Elles sont généralement identifiées à partir de la spécification GARCH(N, N) associée au CGARCH(N).

Cas du CGARCH(2) :

Pour $N = 2$, on a :

$$\begin{aligned}\sigma_t^2 &= \sigma_{1,t}^2 + \sigma_{2,t}^2 \\ &= \beta_1 \sigma_{1,t-1}^2 + \beta_2 \sigma_{2,t-1}^2 + \omega_1 + \omega_2 + (\alpha_1 + \alpha_2) \varepsilon_{t-1}^2\end{aligned}$$

D'une part, on a :

$$\begin{aligned}\beta_1 \sigma_{1,t-1}^2 &= (\beta_1 + \beta_2) \sigma_{1,t-1}^2 - \beta_2 \sigma_{1,t-1}^2 \\ &= (\beta_1 + \beta_2) \sigma_{1,t-1}^2 - \beta_2 (\omega_1 + \alpha_1 \varepsilon_{t-2}^2 + \beta_1 \sigma_{1,t-2}^2) \\ &= (\beta_1 + \beta_2) \sigma_{1,t-1}^2 - \beta_2 \omega_1 + \beta_2 \alpha_1 \varepsilon_{t-2}^2 + \beta_1 \beta_2 \sigma_{1,t-2}^2\end{aligned}$$

D'autre part, en développant de la même façon $\beta_2 \sigma_{2,t-1}^2$, on obtient :

$$\beta_2 \sigma_{2,t-1}^2 = (\beta_1 + \beta_2) \sigma_{2,t-1}^2 - \beta_1 \omega_2 + \beta_1 \alpha_2 \varepsilon_{t-2}^2 + \beta_1 \beta_2 \sigma_{2,t-2}^2$$

D'où la spécification GARCH(2,2) dont la variance conditionnelle est donnée par :

$$\sigma_t^2 = a_0 + a_1 \varepsilon_{t-1}^2 + a_2 \varepsilon_{t-2}^2 + b_1 \sigma_{t-1}^2 + b_2 \sigma_{t-2}^2$$

où :

$$\begin{aligned}a_0 &= \omega_1(1 - \beta_2) + \omega_2(1 - \beta_1) \\ a_1 &= (\alpha_1 + \alpha_2) \\ a_2 &= -(\alpha_1 \beta_2 + \alpha_2 \beta_1) \\ b_1 &= \beta_1 + \beta_2 \\ b_2 &= -\beta_1 \beta_2\end{aligned}$$

Par suite, les contraintes de positivité de σ_t^2 au sens de Bollerslev ainsi que les conditions de stationnarité et d'existence des moments du modèle GARCH sont

applicables sur les paramètres a_0, a_1, a_2, b_1 et b_2 afin de dériver des conditions sur les paramètres ω_i, α_i et $\beta_i, i = 1, 2$ pour que le CGARCH(2) soit bien défini. En particulier, comme noté par Robert F ENGLE et G. LEE (1999), les contraintes de positivité des paramètres du CGARCH(2) assurant la positivité de sa variance conditionnelle peuvent être relaxées telles que :

$$0 < \alpha_2 + \beta_2 < \alpha_1 + \beta_1 < 1, \alpha_1 < \beta_2, 0 < \alpha_1, 0 < \alpha_2, 0 < \omega_1 \text{ et } 0 < \omega_2$$

Cas du CGARCH(3) :

Pour $N = 3$, on a :

$$\begin{aligned} \sigma_t^2 &= \sigma_{1,t}^2 + \sigma_{2,t}^2 + \sigma_{3,t}^2 \\ &= \beta_1 \sigma_{1,t-1}^2 + \beta_2 \sigma_{2,t-1}^2 + \beta_3 \sigma_{3,t-1}^2 + \omega_1 + \omega_2 + \omega_3 + (\alpha_1 + \alpha_2 + \alpha_3) \varepsilon_{t-1}^2 \end{aligned}$$

On a :

$$\begin{aligned} \beta_1 \sigma_{1,t-1}^2 &= (\beta_1 + \beta_2 + \beta_3) \sigma_{1,t-1}^2 - (\beta_2 + \beta_3) \sigma_{1,t-1}^2 \\ &= (\beta_1 + \beta_2 + \beta_3) \sigma_{1,t-1}^2 - (\beta_2 + \beta_3) (\omega_1 + \alpha_1 \varepsilon_{t-2}^2 + \beta_1 \sigma_{1,t-2}^2) \\ &= (\beta_1 + \beta_2 + \beta_3) \sigma_{1,t-1}^2 - \omega_1 (\beta_2 + \beta_3) - \alpha_1 (\beta_2 + \beta_3) \varepsilon_{t-2}^2 + (\beta_1 \beta_2 + \beta_1 \beta_3) \sigma_{1,t-2}^2 \end{aligned}$$

De plus :

$$\begin{aligned} (\beta_1 \beta_2 + \beta_1 \beta_3) \sigma_{1,t-2}^2 &= (\beta_1 \beta_2 + \beta_1 \beta_3 + \beta_2 \beta_3) \sigma_{1,t-2}^2 - \beta_2 \beta_3 \sigma_{1,t-2}^2 \\ &= (\beta_1 \beta_2 + \beta_1 \beta_3 + \beta_2 \beta_3) \sigma_{1,t-2}^2 - \beta_2 \beta_3 (\omega_1 + \alpha_1 \varepsilon_{t-3}^2 + \beta_1 \sigma_{1,t-3}^2) \\ &= (\beta_1 \beta_2 + \beta_1 \beta_3 + \beta_2 \beta_3) \sigma_{1,t-2}^2 - \beta_2 \beta_3 \omega_1 - \beta_2 \beta_3 \alpha_1 \varepsilon_{t-3}^2 - \beta_2 \beta_3 \beta_1 \sigma_{1,t-3}^2 \end{aligned}$$

Nous procédons de la même façon pour $\beta_2 \sigma_{2,t-1}^2$ et $\beta_3 \sigma_{3,t-1}^2$, et nous en dérivons la spécification GARCH(3,3) dont la variance conditionnelle est donnée par :

$$\sigma_t^2 = a_0 + a_1 \varepsilon_{t-1}^2 + a_2 \varepsilon_{t-2}^2 + a_3 \varepsilon_{t-3}^2 + b_1 \sigma_{t-1}^2 + b_2 \sigma_{t-2}^2 + b_3 \sigma_{t-3}^2$$

où :

$$\begin{aligned} a_0 &= \omega_1 (1 - \beta_2 - \beta_3 - \beta_2 \beta_3) + \omega_2 (1 - \beta_1 - \beta_3 - \beta_1 \beta_3) + \omega_3 (1 - \beta_1 - \beta_2 - \beta_1 \beta_2) \\ a_1 &= (\alpha_1 + \alpha_2 + \alpha_3) \\ a_2 &= -\alpha_1 (\beta_2 + \beta_3) - \alpha_2 (\beta_1 + \beta_3) - \alpha_3 (\beta_1 + \beta_2) \\ a_3 &= -\alpha_1 \beta_2 \beta_3 - \alpha_2 \beta_1 \beta_3 - \alpha_3 \beta_1 \beta_2 \\ b_1 &= \beta_1 + \beta_2 + \beta_3 \\ b_2 &= -\beta_1 \beta_2 - \beta_2 \beta_3 - \beta_1 \beta_3 \\ b_3 &= -\beta_1 \beta_2 \beta_3 \end{aligned}$$

Contrairement au CGARCH(2), la relaxation des contraintes sur les paramètres du CGARCH(3) découlant de la spécification GARCH(3,3) n'est pas évidente même

au sens de TSAI et CHAN (2008) qui ne fournit pas une formulation explicite de ces contraintes pour des GARCH d'ordres $p, q > 3$. Bien que la forme générale du modèle GARCH associé à un modèle CGARCH donné avait été établie par KARANASOS (1999) (voir le lemme 1.1), la formulation relaxée des conditions de positivité de la variance conditionnelle, de stationnarité et d'existence des moments pour tout ordre N reste une question ouverte. En conséquence, la détermination de l'espace des paramètres du CGARCH dans lequel se fait la recherche d'un optimum au sens du QML, n'est possible que ex-post via sa représentation GARCH.

Nous attrapons alors le fil et nous proposons une estimation par QCKSA relaxant au sens du Chapitre 2 le CGARCH à travers la relaxation des N spécifications GARCH(1,1) associées aux $\sigma_{i,t}^2$ de paramètre $\theta_i = {}^t(\omega_i, \alpha_i, \beta_i)$, $i = 1, \dots, N$. Le but est de prouver l'existence d'un optimum du CGARCH(N) au sens du SPSA sur une partie déterminée de l'espace des paramètres sans recourir à la détermination ex-post des contraintes relaxées de positivité de σ_t^2 .

Dans ce qui suit, nous analysons la convergence de l'algorithme QCKSA appliqué au modèle CGARCH. Désormais, nous désignons respectivement par $\hat{\sigma}_{i,t}^2 : \Theta_i \rightarrow \mathbb{R}$ et $\tilde{\sigma}_{i,t}^2 : \Theta_i \rightarrow \mathbb{R}^{+*}$ la variance conditionnelle prédite par le filtre de Kalman et sa ré-estimation robuste, de paramètre inconnu θ_i estimant σ_t^2 pour une valeur donnée.

3.2.1 Analyse de convergence

GARCH(1,1) relaxé

Nous commençons par l'analyse de convergence de l'algorithme SPSA pour le GARCH(1,1) relaxé. En effet, pour tout $i = 1, \dots, N$, les hypothèses **(A1)** et **(A2)** appliquées à la i^{eme} composante GARCH(1,1) relaxé s'écrivent respectivement :

$$|\alpha_i| + |\beta_i| < 1 \quad (3.4)$$

$$\beta_i^2 + 2|\alpha_i\beta_i| + 3\alpha_i^2 < 1 \quad (3.5)$$

Sous ces hypothèses, nous pré-estimons $\sigma_{i,t}^2$ par le filtre de Kalman à partir de la représentation espace d'état que l'on dérive de (2.4), soit :

$$\begin{aligned} \sigma_{i,t}^2 &= \omega_i + (\alpha_i + \beta_i)\sigma_{i,t-1}^2 + \alpha_i\nu_{i,t-1} \\ \varepsilon_t^2 &= \sigma_{i,t}^2 + \nu_{i,t} \end{aligned}$$

Ensuite, sous les conditions initiales :

$$\hat{\sigma}_{i,0|0}^2 = \mathbb{E}(\sigma_{i,0}^2) = \frac{\omega_i}{1 - \alpha_i - \beta_i}$$

et

$$P_{0|0} = Var(\sigma_{i,0}^2) = \frac{2\omega_i^2\alpha_i^2}{(1 - \alpha_i - \beta_i)^2(1 - \beta_i^2 - 2\alpha_i\beta_i - 3\alpha_i^2)}$$

nous obtenons les équations du filtre de Kalman permettant de prédire dans une première étape $\sigma_{i,t}^2$ par $\widehat{\sigma}_{i,t|t-1}^2(\theta_i)$.

$$\widehat{\sigma}_{i,t|t-1}^2(\theta_i) = \omega_i + (\alpha_i + \beta_i)\widehat{\sigma}_{i,t-1|t-1}^2(\theta_i) \quad (3.6)$$

$$P_{t|t-1}(\theta_i) = (\alpha_i + \beta_i)^2 P_{t-1|t-1}(\theta_i) + \alpha_i^2 \nu(\theta_i) \quad (3.7)$$

$$K_t(\theta_i) = P_{t|t-1}(\theta_i) [P_{t|t-1}(\theta_i) + \nu(\theta_i)]^{-1} \quad (3.8)$$

$$\widehat{\sigma}_{i,t|t}^2(\theta_i) = \widehat{\sigma}_{i,t|t-1}^2(\theta_i) + K_t(\theta_i) [\varepsilon_t^2 - \widehat{\sigma}_{i,t|t-1}^2(\theta_i)] \quad (3.9)$$

$$P_{t|t}(\theta_i) = [1 - K_t(\theta_i)] P_{t|t-1}(\theta_i) \quad (3.10)$$

Rappelons que sous la condition de stationnarité au second ordre (3.4), le filtre (3.6)-(3.10) est stable au sens de la convergence (exponentielle) de la suite $(P_{t|t-1})$. En effet, on dérive à partir de (3.7), (3.8) et (3.10) l'expression récursive de $P_{t|t-1}$ comme suit :

$$P_{t|t-1} = \left[(\alpha_i + \beta_i)^2 \frac{P_{t-1|t-2}}{P_{t-1|t-2} + \nu} + \alpha_i^2 \right] \nu$$

La limite \bar{P} est la seule solution positive de l'équation de Riccati donnée par :

$$\bar{P} = \left[(\alpha_i + \beta_i)^2 \frac{\bar{P}}{\bar{P} + \nu} + \alpha_i^2 \right] \nu$$

Ainsi, par rapport à la norme $\|\cdot\|$ utilisée ici comme une norme usuelle de \mathbb{R} , on a :

$$\begin{aligned} \|P_{t|t-1} - \bar{P}\| &= (\alpha_i + \beta_i)^2 \|P_{t-1|t-2} - \bar{P}\| \left\| \frac{\nu}{P_{t-1|t-2} + \nu} \right\| \left\| \frac{\nu}{\bar{P} + \nu} \right\| \\ &\leq (\alpha_i + \beta_i)^2 \|P_{t-1|t-2} - \bar{P}\| \end{aligned}$$

$$\text{car} \quad \left\| \frac{\nu}{P_{t-1|t-2} + \nu} \right\| \leq 1 \text{ et } \left\| \frac{\nu}{\bar{P} + \nu} \right\| \leq 1.$$

Dans une deuxième étape, nous appliquons l'algorithme de la troncature de la densité de probabilité 2.2 sous la contrainte de positivité (2.14) et nous obtenons par suite une estimation robuste positive de $\sigma_{i,t}^2$, à savoir $\widetilde{\sigma}_{i,t}^2(\theta_i)$.

En utilisant l'algorithme SPSA, une estimation QCKSA des paramètres du i^{eme} GARCH(1,1) relaxé est donnée par :

$$\theta_i^* = \arg \min_{\theta_i \in \Theta_i} \widehat{l}_{i,n}(\theta_i)$$

où :

$$\widehat{l}_{i,n}(\theta_i) = \frac{1}{n} \sum_{t=1}^n \frac{\varepsilon_t^2}{\widetilde{\sigma}_{i,t}^2(\theta_i)} + \log(\widetilde{\sigma}_{i,t}^2(\theta_i)) \quad (3.11)$$

et

$$\Theta_i = \{ \theta_i \in \mathbb{R}^3 / \omega_i > 0, |\alpha_i| + |\beta_i| < 1, \beta_i^2 + 2|\alpha_i\beta_i| + 3\alpha_i^2 < 1 \}$$

Afin d'analyser la convergence de l'algorithme SPSA appliqué au GARCH(1,1) relaxé, nous nous référons aux hypothèses **(B1)**-**(B6)** §1.4.1.

D'abord, nous notons que les hypothèses **(B2)**-**(B4)** sont satisfaites par le choix de a_k , Δ_k et δ_k^\pm effectué respectivement aux étapes 1 et 2 et 4 dans l'algorithme 3.3 (BHATNAGAR, H. PRASHANTH et L. PRASHANTH (2013)). **(B5)** peut être pratiquement satisfaite à travers des vérifications ex-post faites à l'étape 3 pour éviter la divergence de la suite des itérations (θ_k) aux cas de non-stationnarité ou d'inexistence des moments. Revenant à **(B1)**, une idée qui permet de la satisfaire est d'extraire de Θ_i le plus petit compact (partie bornée et fermée de Θ_i) sur lequel $\widehat{l}_{i,n}$ est régulière au sens de **(B1)**. Ainsi, nous supposons que ω_i est borné, i.e., il existe $\bar{\omega}_i > 0$ tel que pour $\eta > 0$, $\eta \leq \omega_i \leq \bar{\omega}_i$.

Par suite, nous considérons la partie $\Theta_{i,\eta} \subset \Theta_i$ définie par :

$$\Theta_{i,\eta} = \{ \theta_i \in \mathbb{R}^3 / \eta \leq \omega_i \leq \bar{\omega}_i, |\alpha_i| + |\beta_i| \leq 1 - \eta, \beta_i^2 + 2|\alpha_i\beta_i| + 3\alpha_i^2 \leq 1 - \eta \}$$

Il en découle que $\Theta_{i,\eta}$ est effectivement à la fois un fermé comme union d'intervalles fermés de \mathbb{R}^3 , et est borné puisque pour tout $\theta_i \in \Theta_{i,\eta}$, $0 < \omega_i \leq \bar{\omega}_i$, $|\alpha_i| < 1$ et $|\beta_i| < 1$, constituant ainsi une partie compacte. De plus, $\Theta_{i,\eta} \rightarrow \Theta_i$ quand $\eta \rightarrow 0^+$.

Nous étudions dans ce qui suit la convergence de l'algorithme 3.3 en commençant par le lemme suivant montrant la régularité de $\widehat{l}_{i,n}$.

Lemme 3.1. *Pour tout $\eta > 0$, $\widehat{l}_{i,n}$ est régulière sur $\Theta_{i,\eta}$ au sens de **(B1)**.*

Démonstration. Soit $\eta > 0$ et notons respectivement par ϕ_t et ψ_t les fonctions définies sur $\Theta_{i,\eta}$ pour tout $t \in \mathbb{N}^*$ par :

$$\phi_t(\theta_i) = \widehat{\sigma}_{i,t|t-1}^2(\theta_i) \text{ et } \psi_t(\theta_i) = P_{t|t-1}(\theta_i)$$

À partir de (3.6), (3.8) et (3.9), on obtient facilement que :

$$\phi_t(\theta_i) = \omega_i + \frac{(\alpha_i + \beta_i) (\nu(\theta_i)\phi_{t-1}(\theta_i) + \psi_{t-1}(\theta_i)\varepsilon_{t-1}^2)}{\psi_{t-1}(\theta_i) + \nu(\theta_i)} \quad (3.12)$$

où ν est la fonction définie sur $\Theta_{i,\eta}$ par :

$$\nu(\theta_i) = \frac{2\omega_i^2(1 + \alpha_i + \beta_i)}{(1 - \alpha_i - \beta_i)(1 - 3\alpha_i^2 - \beta_i^2 - 2\alpha_i\beta_i)} \quad (3.13)$$

De même, on déduit de (3.7), (3.8) et (3.10) que :

$$\psi_t(\theta_i) = \left[(\alpha_i + \beta_i)^2 \frac{\psi_{t-1}(\theta_i)}{\psi_{t-1}(\theta_i) + \nu(\theta_i)} + \alpha_i^2 \right] \nu(\theta_i) \quad (3.14)$$

Dans un premier temps, nous montrons par récurrence que :

$$\psi_t \in \mathcal{C}^\infty(\Theta_{i,\eta}), \text{ pour tout } t \in \mathbb{N}^* \quad (3.15)$$

En effet, pour $t = 1$, on a $\psi_1 \in \mathcal{C}^\infty(\Theta_{i,\eta})$ comme fonction de $P_{1|0}$ qui est, d'après (3.7), fonction de $P_{0|0} \in \mathcal{C}^\infty(\Theta_{i,\eta})$ étant donnés (3.4) et (3.5).

Supposons maintenant que $\psi_t \in \mathcal{C}^\infty(\Theta_{i,\eta})$. En vertu de (3.4) et (3.5), on a :

$$\nu \in \mathcal{C}^\infty(\Theta_{i,\eta}) \quad (3.16)$$

Comme :

$$\psi_t(\theta_i) + \nu(\theta_i) > 0, \text{ pour tout } \theta_i \in \Theta_{i,\eta} \quad (3.17)$$

Alors :

$$\psi_{t+1} \in \mathcal{C}^\infty(\Theta_{i,\eta})$$

D'où le résultat (3.15).

Maintenant, nous montrons par récurrence que :

$$\phi_t \in \mathcal{C}^\infty(\Theta_{i,\eta}) \text{ pour tout } t \in \mathbb{N}^* \quad (3.18)$$

En effet, pour $t = 1$, on a $\phi_1 \in \mathcal{C}^\infty(\Theta_{i,\eta})$ comme fonction de $\widehat{\sigma}_{i,1|0}^2$ qui est d'après (3.6), fonction de $\widehat{\sigma}_{i,0|0}^2 \in \mathcal{C}^\infty(\Theta_{i,\eta})$ étant donnés (3.4).

En supposant que $\phi_t \in \mathcal{C}^\infty(\Theta_{i,\eta})$, étant donnés (3.17) et (3.16), on déduit que $\phi_{t+1} \in \mathcal{C}^\infty(\Theta_{i,\eta})$. D'où résulte (3.18).

En conséquence, d'après (2.15), on obtient :

$$\widetilde{\sigma}_{i,t|t-1}^2 \in \mathcal{C}^\infty(\Theta_{i,\eta}) \quad (3.19)$$

On en déduit, étant donnée (3.1) que :

$$\widehat{l}_{i,n} \in \mathcal{C}^\infty(\Theta_{i,\eta})$$

Puisque $\Theta_{i,\eta}$ est un compact, alors $\widehat{l}_{i,n}$ est bien régulière sur $\Theta_{i,\eta}$ au sens de **(B1)**. \square

D'après ce qui précède, nous énonçons la proposition suivante assurant la convergence presque sûre de l'équation récursive (3.2).

Proposition 3.1. *Soit H l'ensemble des minimums locaux de $\widehat{l}_{i,n}$ vérifiant **(B6)**, alors pour tout $\eta > 0$, il existe $\bar{c} > 0$ tel que pour tout $c \in]0, \bar{c}]$,*

$$\theta_i \rightarrow H_\eta \text{ p.s}$$

où H_η est un η -voisinage de H donné par :

$$H_\eta = \{\theta_i \in \Theta_{i,\eta} / \|\theta_i - \theta_i^*\| < \eta, \theta_i^* \in H\}$$

Démonstration. En admettant **(B6)**, nous appliquons le théorème 1.6 à $\widehat{l}_{i,n}$ vu que cette dernière vérifie **(B1)** d'après le lemme 3.1 et que **(B2)**-**(B5)** sont satisfaites par le choix des suites a_k , Δ_k et δ_k^\pm fait précédemment. D'où le résultat. \square

CGARCH(N), $N \geq 1$

À ce niveau, l'existence d'une estimation QCKSA des paramètres de chaque GARCH(1,1) relaxé composant le CGARCH(N) est prouvée sur une partie de l'espace des paramètres, à savoir $\Theta_{i,\eta}$ pour $\eta > 0$ et $i = 1, \dots, N$. Par suite, nous minimisons le critère associé à la fonction de log-vraisemblance du CGARCH(N) sur l'espace des paramètres :

$$\tilde{\Theta}_\eta = \prod_{i=1}^N \Theta_{i,\eta}, \text{ pour } \eta > 0$$

Ainsi, nous estimons le critère de vraisemblance pour tout $\theta = {}^t(\omega_1, \alpha_1, \beta_1, \dots, \omega_i, \alpha_i, \beta_i, \dots, \omega_N, \alpha_N, \beta_N) \in \tilde{\Theta}_\eta$ par :

$$\hat{l}_n(\theta) = \frac{1}{n} \sum_{t=1}^n \frac{\varepsilon_t^2}{\hat{\sigma}_t^2(\theta)} + \log(\hat{\sigma}_t^2(\theta)) \quad (3.20)$$

où $\hat{\sigma}_t^2$ est la variance conditionnelle estimée pour tout $\theta \in \tilde{\Theta}_\eta$ par :

$$\hat{\sigma}_t^2(\theta) = \sum_{i=1}^N \tilde{\sigma}_{i,t}^2(\theta_i) \quad (3.21)$$

Par construction $\hat{\sigma}_t^2$ est strictement positive de manière à ce que \hat{l}_n est désormais bien définie sur $\tilde{\Theta}_\eta$.

Remarque 3.1.

L'estimation (3.21) est en accord avec notre approche basée sur l'estimation du modèle CGARCH sans avoir besoin de sa représentation GARCH, car sinon la pré-estimation de $\sigma_{i,t}^2$ pour le GARCH(1,1) reste applicable à une telle représentation pour pré-estimer $\hat{\sigma}_t^2$ au sens de la section §3.1. Or, cela ne correspond pas vraiment à l'objectif de notre approche. Par conséquent, nous n'avons pas à appliquer les algorithmes de pré-estimation de la variance conditionnelle (2.1) et (2.2), mais nous employons directement l'algorithme 3.3 du SPSA pour minimiser \hat{l}_n par rapport à $\theta = {}^t(\omega_1, \alpha_1, \beta_1, \dots, \omega_i, \alpha_i, \beta_i, \dots, \omega_N, \alpha_N, \beta_N)$, en remplaçant Θ_η par $\tilde{\Theta}_\eta$ pour un certain $\eta > 0$ et en fixant $p = q = N$ comme décrit par l'algorithme suivant.

Nous étudions dans ce qui suit la convergence de l'algorithme 3.4 en commençant par le lemme suivant montrant la régularité de \hat{l}_n .

Lemme 3.2. *Pour tout $\eta > 0$, \hat{l}_n est régulière sur $\tilde{\Theta}_\eta$ au sens de (B1).*

Démonstration. Il découle de (3.19) et (3.21) que $\hat{\sigma}_t^2 \in \mathcal{C}^\infty(\tilde{\Theta}_\eta)$ comme une somme de fonctions de classe \mathcal{C}^∞ sur les parties formant $\tilde{\Theta}_\eta$.

Par conséquent, étant donnée (3.20), on a :

$$\hat{l}_n \in \mathcal{C}^\infty(\tilde{\Theta}_\eta)$$

Par ailleurs, $\tilde{\Theta}_\eta$ est un compact comme le produit cartésien des compacts $\Theta_{i,\eta}$ pour $i = 1, \dots, N$. D'où le résultat. \square

Algorithme 3.4 Minimisation par SPSA de la log-vraisemblance du CGARCH.

Pour $i = 1$ à N , **faire**

Appliquer les algorithmes 2.1 et 2.2 pour calculer $\tilde{\sigma}_{i,t}^2$

Appliquer l'algorithme 3.3 à $\hat{l}_{i,n}$ pour obtenir $\hat{\theta}_i = \arg \min_{\theta_i \in \Theta_{i,\eta}} \hat{l}_{i,n}(\theta_i)$

Fin Pour

Pour $t = 1$ à n , **faire**

$\hat{\sigma}_t^2(\theta) = \sum_{i=1}^N \tilde{\sigma}_{i,t}^2(\hat{\theta}_i)$

Fin Pour

Appliquer 3.3 à \hat{l}_n pour obtenir $\hat{\theta} = \arg \min_{\theta \in \tilde{\Theta}_\eta} \hat{l}_n(\theta)$.

La proposition suivante conduit à minimiser localement \hat{l}_n sur $\tilde{\Theta}_\eta$.

Proposition 3.2. *Soit L l'ensemble des minimums locaux de \hat{l}_n vérifiant (B6) alors, pour tout $\eta > 0$, il existe $\bar{c}' > 0$ tel que pour tout $c \in]0, \bar{c}'[$,*

$$\theta \rightarrow L_\eta \text{ p.s}$$

où L_η est un η -voisinage de L donné par :

$$L_\eta = \{\theta \in \Theta_\eta / \|\theta - \theta^*\| < \eta, \theta^* \in L\}$$

Démonstration. En admettant (B6), nous appliquons cette fois-ci le théorème 1.6 à \hat{l}_n vu que cette dernière vérifie (B1) d'après le lemme 3.2 et que (B2)-(B5) sont satisfaites par le choix des suites a_k , Δ_k et δ_k^\pm fait précédemment. D'où le résultat. \square

Nous parvenons donc à déterminer une partie Θ_η de l'espace des paramètres sur laquelle l'estimation des paramètres du CGARCH(N) par QCKSA est faisable quel que soit l'ordre N du modèle au lieu de déterminer cet espace ex-post à partir de sa spécification GARCH(N, N).

3.2.2 Illustration numérique

Nous présentons un exemple empirique d'application de la méthode QCKSA pour l'estimation des paramètres du CGARCH selon la spécification (3.3). Il s'agit de la série des rendements journaliers de l'indice S&P500 sur la période, du 28/10/2010 au 27/11/2020 constituant une taille $n = 2539$ observations² (figure 3.4). Les statistiques rapportées au tableau 3.2 révèlent la présence des caractéristiques usuelles d'excès de kurtosis et d'autocorrélation significative de la série des rendements et des carrés des rendements comme le montre les Q-statistiques de Ljung-Box (par exemple à l'ordre 10). Nous ajustons donc la série des rendements que l'on note r_t par un AR(1) à erreur de type CGARCH(2), soit :

2. Ces données sont disponibles sur le site www.investing.com

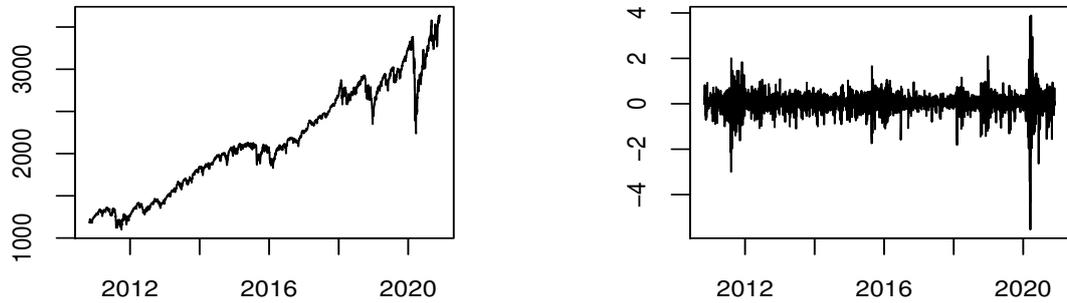


FIGURE 3.4 – L’indice S&P500 (à gauche) et la série des rendements correspondante (à droite) du 28 Octobre 2010 au 27 Novembre 2020.

$$\begin{aligned}
 r_t &= \phi_0 + \phi_1 r_{t-1} + \varepsilon_t, \quad (\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_1) \sim iidN(0, \sigma_t^2) \\
 \varepsilon_t &= (\sigma_{1,t}^2 + \sigma_{2,t}^2)^{\frac{1}{2}} \eta_t = \sigma_t \eta_t, \quad \eta_t \sim iidN(0, 1) \\
 \sigma_{i,t}^2 &= \omega_i + \alpha_i \varepsilon_{t-1}^2 + \beta_i \sigma_{i,t-1}^2, \quad i = 1, 2
 \end{aligned} \tag{3.22}$$

TABLE 3.2 – Statistiques descriptives des rendements du S&P500.

Min	Max	Moyenne	Variance	kurtosis	$Q_r(10)$	$Q_{r^2}(10)$
-5.5438	3.8948	0.0192	0.2289	20.8808	66.817	602.29

¹ $Q_r(10)$ et $Q_{r^2}(10)$ sont les Q-statistiques de Ljung-Box à l’ordre 10.

D’abord, nous estimons simultanément par QML les paramètres du CGARCH (3.22), à savoir ω_1 , α_1 , β_1 , ω_2 , α_2 et β_2 . Ensuite, nous estimons par QCKSA donnée par l’algorithme 3.4 les paramètres associés aux composantes $\sigma_{1,t}^2$ et $\sigma_{2,t}^2$. Nous présentons dans le tableau 3.3 l’ensemble des estimations obtenues outre que les valeurs de log-vraisemblance correspondantes.

Les résultats de l’estimation présentés dans le tableau 3.3 montrent que la composante $\sigma_{1,t}^2$ capture les fluctuations de volatilité à long terme tandis que la deuxième composante $\sigma_{2,t}^2$ modélise les mouvements de volatilité à court terme. En effet, par la méthode QML (resp. QCKSA), $\sigma_{1,t}^2$ commence faiblement en amplitude (0.0079 (resp. 0.0062)), mais décroît très lentement avec un taux de décroissance de 0.9944 (resp. 0.9913), tandis que $\sigma_{2,t}^2$ commence fortement en amplitude (0.0877 (resp. 0.0833)), mais décroît très rapidement avec un taux de décroissance de 0.7355 (resp.

TABLE 3.3 – Estimation des paramètres du CGARCH(2) par QML et QCKSA.

	$\hat{\phi}_0$	$\hat{\phi}_1$	$\hat{\omega}_1$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\omega}_2$	$\hat{\alpha}_2$	$\hat{\beta}_2$	log-Lik
QML	0.0191	-0.1621	0.0079	0.0007	0.9944	0.0877	0.2118	0.7355	-992.6334
QCKSA	0.0191	-0.1621	0.0062	0.0004	0.9913	0.0833	0.1821	0.7351	-985.7104

¹ log-Lik est la valeur de la log-vraisemblance.

0.7351) (les figures 3.5 et 3.6). Ces remarques sont également conformes aux persis- tances estimées par chaque composante, données par $\hat{\alpha}_i + \hat{\beta}_i$, $i = 1, 2$, selon lesquelles la composante $\sigma_{1,t}^2$ capture l'effet de persistance de σ_t^2 avec une valeur de 0.9973 (resp. 0.9917) dépassant celle de $\sigma_{2,t}^2$ valant 0.9473 (resp. 0.9172). De plus, $\sigma_{1,t}^2$ est largement moins sensible aux chocs ε_{t-1}^2 que $\sigma_{2,t}^2$ ($0.0007 < 0.2118$ et $0.0004 < 0.1821$). Notons en particulier que l'estimation par QCKSA distingue mieux le comportement de σ_t^2 en termes de persistance de la volatilité pour chaque composante vu qu'elle reflète mieux la persistance faible de $\sigma_{2,t}^2$ comparativement à celle obtenue par QML ($0.9172 < 0.9473$).

En guise de conclusion, l'estimation par QCKSA a fourni des résultats conformes à la structure du CGARCH reflétant à la fois le comportement à court et à long terme de la volatilité, en enregistrant de plus une valeur de log-vraisemblance relativement plus élevé ($-985.7104 > -992.6334$).

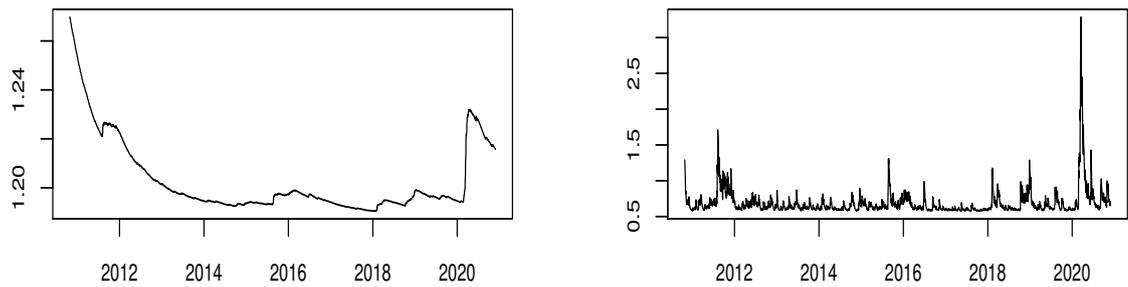


FIGURE 3.5 – La volatilité de la composante à long terme $\sigma_{1,t}^2$ (à gauche) et la composante à court terme $\sigma_{2,t}^2$ (à droite) estimées par QML.

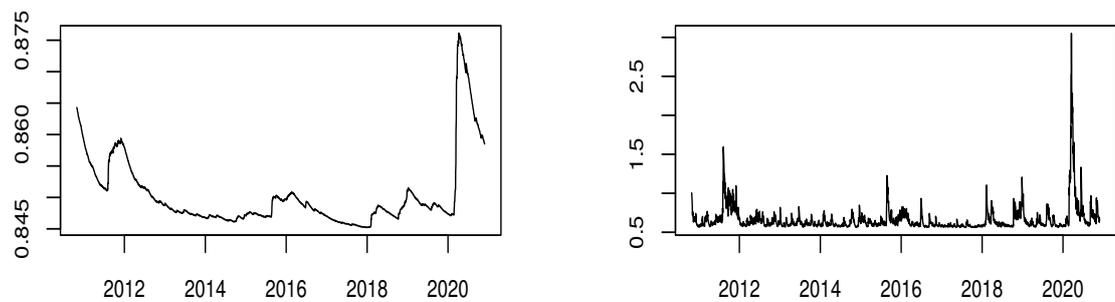


FIGURE 3.6 – La volatilité de la composante à long terme $\sigma_{1,t}^2$ (à gauche) et la composante à court terme $\sigma_{2,t}^2$ (à droite) estimées par QCKSA.

3.3 Conclusion

Dans l'objectif d'améliorer l'optimalité des estimations des paramètres des modèles GARCH et CGRACH, la méthode QCK a été combinée avec l'algorithme SPSA comme routine d'optimisation. L'illustration empirique appliquant le premier modèle a révélé l'effet de l'utilisation de l'algorithme du SPSA en obtenant un maximum de vraisemblance meilleur que celui du QML, en satisfaisant les conditions relaxées de Nelson et Cao. D'autre part, la convergence de l'algorithme SPSA appliqué au CGRACH a été prouvée dans un espace des paramètres défini à priori à partir des espaces de paramètres des composantes GARCH(1,1) relaxés. L'étude empirique a montré la performance de l'estimation QCKSA à capter la dynamique à court et à long terme de la volatilité du CGARCH, en maximisant sa vraisemblance relativement à l'estimation par QML.

Estimation du modèle log-GARCH basée sur le filtre d'information

Dans les chapitres précédents, un premier aspect d'utilisation du filtre de Kalman a été présenté à travers la robustification de l'estimation de la variance conditionnelle générée par le modèle GARCH. Dans ce chapitre, on poursuit l'intégration des variétés du filtre de Kalman dans l'estimation du modèle log-GARCH. Ce dernier est l'une des extensions du GARCH qui remédie au problème de positivité de la variance conditionnelle en modélisant son logarithme en fonction du passé des logarithmes des carrés des observations de sorte que les paramètres ne soient à priori soumis à aucune condition de positivité. Les premiers travaux qui ont introduit le modèle log-GARCH sont ceux de GEWEKE (1986), PANTULA (1986) et MILHØJ (1987), tandis que la première généralisation du log-GARCH est due à SUCARRAT et ESCRIBANO (2010), où le modèle dit *Power log-GARCH* en anglais, a été proposé comme une classe générale du modèle log-GARCH.

Les propriétés de consistance et de normalité asymptotique des estimateurs de QML des paramètres du log-GARCH ont été établies par Christian FRANCO, WINTENBERGER et ZAKOIAN (2013a) avec une extension au cas asymétrique. Dans la même veine, Christian FRANCO et SUCARRAT (2013) ont proposé une estimation du modèle log-GARCH via sa représentation ARMA, appelée *exponential Chi-squared*. Cette approche a conduit au développement d'une estimation plus robuste des modèles dits log-GARCH-X où des covariables ou d'autres variables conditionnelles sont ajoutées à l'équation de la volatilité avec une densité conditionnelle inconnue (SUCARRAT, GRØNNEBERG et ESCRIBANO (2016)). Une partie intéressante de la littérature qui coïncide avec les modèles log-GARCH est celle des modèles de volatilité stochastique. En effet, grâce à ces modèles, le log-GARCH peut être estimé d'une manière consistante via sa représentation ARMA (Christian FRANCO et ZAKOIAN (2006)) ou en utilisant des formes d'espace d'état combinées avec le filtre de Kalman (AKNOUCHE (2017)).

Mesure robuste du kurtosis

La structure des moments et des log-moments du processus log-GARCH a été étudiée en détail par Christian FRANCO, WINTENBERGER et ZAKOIAN (2013a); Christian FRANCO et ZAKOIAN (2019) dans un cadre qui comprend les cas symétriques et asymétriques. Cependant, l'existence des moments d'ordres élevés n'est pas empiriquement vérifiée, notamment lorsqu'on examine les rendements financiers réels qui présentent des queues épaisses, de sorte que les moments d'ordres supérieurs, notamment d'ordre 4, peuvent être infinis (ANATOLYEV (2019); BAI et NG (2005); CONT (2001)). D'où la nécessité de penser à d'autres mesures robustes du kurtosis qui reflètent significativement les propriétés empiriques de la queue de la distribution de probabilité des rendements financiers. Ainsi, une attention intéressante a été accordée à la non-robustesse de la mesure standard du kurtosis exprimée par le quatrième moment normalisé. Son inconvénient majeur est sa sensibilité à la présence des valeurs aberrantes, et encore moins que les rendements financiers ne sont pas compatibles avec une distribution normale. Parmi les travaux portant sur la construction d'alternatives à la mesure standard du kurtosis, nous citons SIDDIQUI et RAGHUNANDANAN (1967), HOGG (1972), HOGG (1974), MOORS (1988) et KIM et WHITE (2004). Dans cet esprit, FIORI (2008) a proposé un cadre pratique pour la construction d'estimateurs empiriques appelés kurtosis droit et gauche présentant des propriétés appropriées de consistance et de normalité asymptotique sous des conditions faibles d'existence des moments (FIORI et BELTRAMI (2012); FIORI et BELTRAMI (2014)).

Ajustement des rendements nuls par le log-GARCH

Par ailleurs, face à une série de rendements contenant des zéros, le modèle log-GARCH est par construction, inapproprié pour ajuster de telles données en raison de la spécification logarithmique de sa variance conditionnelle qui produit des termes indéfinis censés être des "log-rendements nuls".

En se référant à la littérature financière, on s'aperçoit que l'occurrence des rendements nuls est souvent considérée comme une mesure de liquidité dans le sens où plus ces rendements sont fréquents, plus la liquidité est faible (BEKAERT, C. R. HARVEY et LUNDBLAD (2007); LESMOND, OGDEN et TRZCINKA (1999); LEVINE et SCHMUKLER (2006)). Notons que la probabilité d'observer un rendement nul peut varier et dépendre des conditions du marché. HAUSMAN, LO et MACKINLAY (1992) proposent un cadre de modélisation probit permettant à cette probabilité de dépendre de certaines variables conditionnelles comme le volume des transactions des trader, leurs durées, la série des rendements passés, etc. Dans le même cadre s'inscrivent les travaux de Robert F ENGLE et RUSSELL (1998), RUSSELL et Robert F ENGLE (2005), BIEN, NOLTE et POHLMEIER (2011) et RYDBERG et SHEPHARD (2003). SUCARRAT et GRØNNEBERG (2016) établissent une nouvelle classe de modèle des rendements financiers avec une probabilité conditionnelle d'occurrence des valeurs nulles variable dans le temps, combinée avec les modèles GARCH et les modèles de volatilité stochastique (SV).

Bien que l'hypothèse d'une probabilité nulle d'occurrence des rendements nuls évite théoriquement la présence des rendements nuls, ils peuvent être empiriquement observés, à cause de certains problèmes au niveau des données, par exemple les valeurs manquantes associées aux jours fériés, la discrétisation des données cotées par une proportion fixe (la variation des rendement se fait à proportions fixe) ou même l'observation irrégulière de ces rendements, tous conduisent à des estimations de QML asymptotiquement biaisées (SUCARRAT et ESCRIBANO (2013)). Vis à vis de cela, Christian FRANCO, WINTENBERGER et ZAKOIAN (2013a) considère la spécification (1.7) sous l'hypothèse $Prob(\eta_t = 0) = 0$ et multiplie chaque paramètre α_i , pour $i = 1, \dots, p$, par $\mathbb{1}_{\{\eta_{t-i} \neq 0\}}$ afin d'éviter que les observations nulles se produisent dans la récursion de l'équation de la variance conditionnelle. Dans une autre contribution portant sur l'estimation par QML du log-GARCH, SUCARRAT et ESCRIBANO (2013) ont proposé une approche qui traite les observations nulles comme des valeurs manquantes en utilisant l'imputation par l'algorithme EM.

L'approche de SUCARRAT et ESCRIBANO (2013) élargit en effet l'éventail d'utilisation de l'algorithme EM pour l'estimation des paramètres du log-GARCH appliqué à des série de rendements contenant des observations nulles. Dans un tel contexte, les représentations espace d'état et le filtre de Kalman sont des outils adéquats pour une bonne performance de l'algorithme EM. Le cadre théorique d'une telle combinaison est rapporté dans BROCKWELL, DAVIS et FIENBERG (1991) avec une application au modèle ARCH dans BAHAMONDE (2012).

Ce chapitre est focalisé sur deux axes. Le premier porte sur l'estimation des paramètres du modèle log-GARCH (p, q) par quasi-maximum de vraisemblance sans aucune hypothèse sur l'existence des log-moments supérieurs à 1 §4.1. Pour cela, le modèle log-GARCH est en premier lieu représenté sous forme d'espace état §4.1. Ensuite, la log-variance conditionnelle est estimée à l'aide du filtre d'information sous la condition de stabilité assurant la stationnarité du modèle §4.2. Après avoir inséré la variance conditionnelle estimée dans la quasi-vraisemblance du modèle, cette dernière est évaluée et maximisée aléatoirement par l'algorithme fonctionnel lissé §4.3. La méthode d'estimation proposée est utilisée à deux fins. D'une part, l'estimation de la variance conditionnelle est utilisée pour normaliser les résidus du log-GARCH afin de calculer les mesures du kurtosis droit et gauche. D'autre part, la présence des observations nulles est traitée par imputation comme un problème de valeurs manquantes, permettant ainsi l'estimation du log-GARCH à l'aide d'un algorithme EM. L'ensemble de ces apports est mis en évidence via des illustrations numériques §4.4.

4.1 Représentation espace d'état du modèle log-GARCH

Rappelons la définition d'un processus $\varepsilon_t \sim \text{log-GARCH}(p, q)$, soit :

$$\varepsilon_t = \sigma_t \eta_t, \quad \eta_t \sim iid(0, 1) \quad (4.1)$$

$$\log \sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \log \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \log \sigma_{t-j}^2 \quad (4.2)$$

où $\sigma_t > 0$, ω , $\alpha_1, \dots, \alpha_p$ et β_1, \dots, β_q sont des réels.

D'abord nous transformons le modèle log-GARCH en une forme additive en log-transformant le carré de l'équation (4.1). Ainsi, le modèle log-GARCH(p, q) peut être reformulé sous forme d'espace d'état de variable d'état $H_t = {}^t(\log \sigma_t^2 \dots \log \sigma_{t-r+1}^2)$ représentant le vecteur des log-variances conditionnelles, avec $r = \max(p, q)$, et de variable de mesure $Y_t = \log \varepsilon_t^2$. Il en découle la représentation espace d'état suivante :

$$\begin{cases} H_t = \Omega + \Phi H_{t-1} + u_{t-1} \\ Y_t = e_r H_t + w_t \end{cases} \quad (4.3)$$

où Ω et Φ sont les matrices de transition données par :

$$\Omega = {}^t(\omega \ 0_{(r-1,1)}) \in \mathbb{R}^r \quad \text{et} \quad \Phi = \begin{pmatrix} \phi_1 & \dots & \phi_r \\ I_{r-1} & & 0_{(r-1,1)} \end{pmatrix} \in \mathcal{M}_{(r,r)}$$

avec $\phi_i = \alpha_i + \beta_i$ tel que $\phi_i = \beta_i$ pour $i > p$ et $\phi_i = \alpha_i$ pour $i > q$.

u_{t-1} et w_t sont des bruits blancs non corrélés (comme fonctions de η_t) associés respectivement à l'équation d'état et l'équation de mesure, et sont donnés par :

$$u_t = {}^t \left(\sum_{i=1}^p \alpha_i \log \eta_{t-i+1}^2 \quad 0_{(1,r-1)} \right) \sim iid(m_u, Q) \quad \text{et} \quad w_t = \log \eta_t^2 \sim iid(m_w, R)$$

où $m_u := \mathbb{E}u_t$, $Q := Cov(u_t)$, $m_w := \mathbb{E}w_t$ et $R := Cov(w_t)$.

Rappelons que la normalité du modèle espace d'état assure l'optimalité du filtre de Kalman en fournissant un estimateur efficace dans l'ensemble de tous les estimateurs linéaires et non linéaires possibles. Dans ce sens, nous normalisons le modèle (4.3) en conjonction avec la normalité du modèle (4.1)-(4.2). D'ailleurs, plusieurs voies ont été proposées pour dériver et/ou approcher la distribution de $\log \eta_t^2$ aussi bien pour le modèle log-GARCH (Christian FRANCO et SUCARRAT (2013)) que pour certaines représentations autorégressives périodiques de la volatilité stochastique (PAR-SV) (AKNOUCHE (2017)).

Dans notre contexte, nous procédons autrement en utilisant l'approximation log-normale de la distribution χ^2 afin d'approximer m_w et R de sorte que w_t soit normal et en déduire ensuite les paramètres m_u et Q de u_t héritant ainsi la normalité de w_t .

En effet, comme $\eta_t \sim iidN(0, 1)$, alors $\eta_t^2 \sim \chi_1^2$ de moyenne 1 et de variance 2. En même temps, supposer que $\log \eta_t^2 \sim iidN(m_w, R)$ équivaut $\eta_t^2 \sim iid \log-N(m_w, R)$. Nous déterminons donc m_w et R en liaison avec la moyenne et la variance de la loi χ_1^2 par la méthode des moments, soit :

$$e^{m_w + \frac{R}{2}} = 1 \quad \text{et} \quad (e^R - 1)e^{2m_w + R} = 2$$

d'où :

$$m_w = -\log \sqrt{3} \quad \text{et} \quad R = \log 3$$

Il s'ensuit que :

$$m_u = {}^t \left(m_w \sum_{i=1}^p \alpha_i \quad 0_{(1,r-1)} \right) \in \mathbb{R}^r \quad \text{et} \quad Q = \begin{pmatrix} R \sum_{i=1}^p \alpha_i^2 & 0 & \cdots & 0 \\ 0_{(r-1,r)} & & & \end{pmatrix} \in \mathcal{M}_{(r,r)}$$

Remarque 4.1.

Il est bien clair que le modèle (4.3) est à bruits non centrés dont les moyennes sont exactement connues comme fonction de $\eta_t \sim iidN(0,1)$. Ainsi, le filtre de Kalman maintient sa bonne performance en intégrant les moyennes m_u et m_w dans les équations de mise à jour du filtre.

4.2 Filtre d'information

Soit ε_t un log-GARCH(p, q) comme donné par (4.1)-(4.2). Nous réduisons les conditions d'existence des log-moments de ε_t à celle du premier ordre. Ainsi, nous supposons que $\mathbb{E}|\log \varepsilon_t^2| < \infty$.

Dans ce sens, nous pré-estimons la log-variance conditionnelle $\log \sigma_t^2$ en utilisant le filtre d'information permettant d'initialiser l'inverse de la matrice de covariance de l'erreur $P_t = \mathbb{E} \{ (H_t - \mathbb{E}H_t) {}^t (H_t - \mathbb{E}H_t) \}$ plutôt que P_t étant inconnue. Par suite, l'inverse de P_t représente la matrice d'information du système espace d'état (4.3) que l'on note \mathcal{I}_t . En pratique, nous considérons que P_0 est infinie et que $\mathbb{E}|\log \sigma_t^2| < \infty$ pour initialiser le filtre d'information comme suit :

$$\mathcal{I}_0 = 0 \quad \text{et} \quad H_0 = \mathbb{E}H_1 = \frac{\omega + m_w \sum_{i=1}^p \alpha_i} {1 - \sum_{i=1}^r \phi_i} {}^t (1 \dots 1) \in \mathbb{R}^r$$

Soient $\widehat{H}_{t|t-1}$ et $\widehat{H}_{t|t}$ respectivement les estimations prédites et filtrées de H_t . $\mathcal{I}_{t|t-1}$ et $\mathcal{I}_{t|t}$ sont leurs matrices d'information respectives.

À partir des observations Y_1, \dots, Y_n , les équations du filtre d'information appliquées au modèle (4.3) sont obtenues en inversant les équations du filtre de Kalman (4.4), (4.6), (4.11), (4.13) et (4.14) associées au même modèle via la variable d'état d'information prédite et filtrée notées respectivement $\widetilde{H}_{t|t-1}$ et $\widetilde{H}_{t|t}$.

En effet, l'équation de mise à jour de la covariance des erreurs filtrée est donnée par :

$$P_{t|t} = (I_r - K_t e_r) P_{t|t-1} \tag{4.4}$$

D'où :

$$K_t e_r P_{t|t-1} = P_{t|t-1} - P_{t|t} \tag{4.5}$$

Rappelons que le gain de Kamlan est donné par :

$$K_t = P_{t|t-1} {}^t e_r (e_r P_{t|t-1} {}^t e_r + R)^{-1} \tag{4.6}$$

Donc :

$$\begin{aligned} K_t(e_r P_{t|t-1} {}^t e_r + R) &= P_{t|t-1} {}^t e_r \\ K_t e_r P_{t|t-1} {}^t e_r + K_t R &= P_{t|t-1} {}^t e_r \end{aligned} \quad (4.7)$$

En substituant (4.5) à (4.7), on obtient le gain de Kalman comme :

$$K_t = P_{t|t} {}^t e_r R^{-1} \quad (4.8)$$

Quant à l'équation de mise à jour de l'information filtrée, l'équation (4.5) implique :

$$I_r - K_t e_r = P_{t|t} P_{t|t-1}^{-1} \quad (4.9)$$

En remplaçant (4.8) dans (4.9), on obtient :

$$P_{t|t}^{-1} = P_{t|t-1}^{-1} + {}^t e_r R^{-1} e_r$$

d'où :

$$\mathcal{I}_{t|t} = \mathcal{I}_{t|t-1} + {}^t e_r R^{-1} e_r \quad (4.10)$$

Concernant l'équation de mise à jour de l'information prédite, il suffit d'inverser la covariance de l'erreur :

$$P_{t|t-1} = \Phi P_{t|t} {}^t \Phi + Q \quad (4.11)$$

en utilisant la formule d'inversion des matrices suivante :

$$(A + B)^{-1} = A^{-1} - A^{-1} B (I + A^{-1} B)^{-1} A^{-1}$$

appliquée à :

$$A = \Phi P_{t|t} {}^t \Phi \quad \text{et} \quad B = Q$$

On obtient :

$$\begin{aligned} P_{t|t-1}^{-1} &= (\Phi P_{t|t} {}^t \Phi + Q)^{-1} \\ &= M_t - M_t Q (I_r + M_t Q)^{-1} M_t \\ &= M_t - M_t (Q^{-1} + M_t)^{-1} M_t \end{aligned}$$

où :

$$M_t = (\Phi P_{t-1/t-2} {}^t \Phi)^{-1}$$

Ainsi,

$$\mathcal{I}_{t|t-1} = M_t - M_t (Q^{-1} + M_t)^{-1} M_t \quad (4.12)$$

Pour les estimations filtrées et prédites de H_t , on dérive la première à partir de l'équation de mise à jour de l'état filtré :

$$\widehat{H}_{t|t} = \widehat{H}_{t|t-1} + K_t (Y_t - e_r \widehat{H}_{t|t-1} - m_w) \quad (4.13)$$

Ce qui entraîne :

$$\begin{aligned}\mathcal{I}_{t|t}\widehat{H}_{t|t} &= \mathcal{I}_{t|t}\widehat{H}_{t|t-1} + \mathcal{I}_{t|t}K_t(Y_t - e_r\widehat{H}_{t|t-1} - m_w) \\ \widetilde{H}_{t|t} &= \mathcal{I}_{t|t}\widehat{H}_{t|t-1} + {}^t e_r R^{-1}(Y_t - e_r\widehat{H}_{t|t-1} - m_w), \quad \text{D'après (4.8)} \\ \widetilde{H}_{t|t} &= \mathcal{I}_{t|t-1}\widehat{H}_{t|t-1} + {}^t e_r R^{-1}(Y_t - m_w), \quad \text{D'après (4.10)}\end{aligned}$$

D'où :

$$\widetilde{H}_{t|t} = \widetilde{H}_{t|t-1} + {}^t e_r R^{-1}(Y_t - m_w)$$

Pour la deuxième, nous partons de l'équation de mise à jour de l'état prédite :

$$\widehat{H}_{t|t-1} = \Omega + \Phi\widehat{H}_{t|t} + m_u$$

Il en découle :

$$\begin{aligned}\mathcal{I}_{t|t-1}\widehat{H}_{t|t-1} &= (M_t - M_t(Q^{-1} + M_t)^{-1}M_t)(\Omega + \Phi\widehat{H}_{t|t} + m_u), \quad \text{D'après (4.12)} \\ \widetilde{H}_{t|t-1} &= (I_r - M_t(Q^{-1} + M_t)^{-1}) {}^t \Phi^{-1}\mathcal{I}_{t|t}\Phi^{-1}(\Omega + \Phi\widehat{H}_{t|t} + m_u) \\ &= (I_r - M_t(Q^{-1} + M_t)^{-1}) {}^t \Phi^{-1}\mathcal{I}_{t|t}(\Phi^{-1}\Omega + \widehat{H}_{t|t} + \Phi^{-1}m_u)\end{aligned}$$

où :

$$N_t = (I_r - M_t(Q^{-1} + M_t)^{-1}) {}^t \Phi^{-1}$$

Donc :

$$\widetilde{H}_{t|t-1} = N_t\widetilde{H}_{t-1/t-1} + N_t {}^t \Phi^{-1}\mathcal{I}_{t-1/t-1}\Phi^{-1}(\Omega + m_u)$$

En somme, le système des équations du filtre d'information appliqué au modèle (4.3) est comme suit :

$$M_t = {}^t \Phi^{-1}\mathcal{I}_{t|t}\Phi^{-1} \quad (4.14)$$

$$\mathcal{I}_{t|t-1} = M_t - M_t(Q^{-1} + M_t)^{-1} {}^t M_t \quad (4.15)$$

$$\mathcal{I}_{t|t} = \mathcal{I}_{t|t-1} + {}^t e_r R^{-1}e_r \quad (4.16)$$

$$K_t = (\mathcal{I}_{t|t})^{-1} {}^t e_r R^{-1} \quad (4.17)$$

$$\widetilde{H}_{t|t} = \widetilde{H}_{t|t-1} + {}^t e_r R^{-1}(Y_t - m_w) \quad (4.18)$$

$$\widehat{H}_{t|t} = (\mathcal{I}_{t|t})^{-1} \widetilde{H}_{t|t} \quad (4.19)$$

$$N_t = [I_r - M_t(Q^{-1} + M_t)^{-1}] {}^t \Phi^{-1} \quad (4.20)$$

$$\widetilde{H}_{t|t-1} = N_t \left[\widetilde{H}_{t-1/t-1} + \mathcal{I}_{t-1/t-1}\Phi^{-1}(\Omega + m_u) \right] \quad (4.21)$$

$$\widehat{H}_{t|t-1} = (\mathcal{I}_{t|t-1})^{-1} \widetilde{H}_{t|t-1} \quad (4.22)$$

Cependant, l'implémentation directe des équations du filtre (4.14)-(4.22) pose deux problèmes principaux :

- (i) Q^{-1} apparaît dans les équations (4.15) et (4.20) sans que Q soit forcément inversible. Pour y faire face, nous procédons comme dans BROWN (1983) où l'équation (4.15) est remplacée par $\mathcal{I}_{t|t-1} = M_t$ chaque fois que Q soit singulière.

- (ii) La stabilité du filtre associé au modèle d'espace d'état (4.3) est requise afin d'éviter la divergence des équations de la matrice de covariance de l'erreur filtrée et prédite. En effet, le filtre associé au modèle (4.3) est stable si et seulement si toutes les valeurs propres de Φ soient de module inférieur à 1 (voir la proposition 1.4). Ce qui est équivalent à la condition :

$$\max |Sp(\Phi)| < 1$$

Remarque 4.2.

La condition de stabilité assure également la stationnarité au second ordre de H_t vu que le système (4.3) est invariant dans le temps (voir le corollaire 1.1). Ainsi, sous la condition de stabilité, les paramètres du log-GARCH satisfont :

$$\sum_{i=1}^r \phi_i < 1$$

En résumé, le filtre d'information final est présenté dans l'algorithme 4.5.

Algorithme 4.5 Prédiction de la variance conditionnelle du log-GARCH par le filtre d'information.

- **Initialisation** : $\widehat{H}_{0|-1} = \mathbb{E}H_0$ et $\mathcal{I}_{0|-1} = 0$
 - **Pour** $t = 0, \dots, n$, **faire** :
 - $\widetilde{H}_{t|t-1} = \mathcal{I}_{t|t-1} \widehat{H}_{t|t-1}$
 - Calculer $\mathcal{I}_{t|t}$ à partir de (4.16)
 - Calculer $\widetilde{H}_{t|t}$ à partir de (4.18)
 - Calculer $\widehat{H}_{t|t}$ à partir de (4.19)
 - Calculer M_t et N_t respectivement à partir de (4.14) et (4.20)
 - **Si** $\max |Sp(\Phi)| < 1$, **alors** :
 - **Si** $\det(Q) \neq 0$, **alors** :
 - Calculer $\mathcal{I}_{t+1|t}$ à partir de (4.15)
 - Calculer $\widetilde{H}_{t+1|t}$ à partir de (4.21)
 - **Si non**,
 - $\mathcal{I}_{t+1|t} = M_t$
 - $\widetilde{H}_{t+1|t} = {}^t\Phi^{-1} \widetilde{H}_{t|t} + M_t(\Omega + m_u)$
 - **Fin si**
 - **Si non**,
 - $\widetilde{H}_{t+1|t} = \widetilde{H}_{t|t-1}$
 - **Fin si**
- **Fin Pour**.
-

Étant donnée l'estimation prédite $\widehat{H}_{t|t-1}$, nous devrions extraire l'estimation de la variance conditionnelle, à savoir $\widehat{\sigma}_{t|t-1}^2 = \mathbb{E}(\sigma_t^2 / \underline{\varepsilon}_{t-1}^2)$, qui sera insérée dans la fonction de quasi-vraisemblance. Il est alors évident d'estimer $\log(\sigma_t^2)$ par :

$$\widehat{H}_{t|t-1}^{(1)} := e_r \widehat{H}_{t|t-1} = \mathbb{E}(\log \sigma_t^2 / \underline{\varepsilon}_{t-1}^2) \quad (4.23)$$

Cependant, l'extraction de la variance conditionnelle estimée comme $\exp \widehat{H}_{t|t-1}^{(1)}$ est incorrecte et conduit à un biais d'estimation, à moins qu'un terme multiplicatif C soit introduit pour le minimiser. Ainsi, et comme le préconise AKNOUCHE (2017), nous prenons C comme la variance prédite de $\log \sigma_t^2$, à savoir l'inverse du premier élément diagonale de la matrice d'information prédite, noté $(\mathcal{I}_{t|t-1}^{(1)})^{-1}$, soit :

$$\widehat{\sigma}_{t|t-1}^2 = \left(\mathcal{I}_{t|t-1}^{(1)} \right)^{-1} \exp \widehat{H}_{t|t-1}^{(1)} \quad (4.24)$$

4.3 Estimation par quasi-maximum de vraisemblance du log-GARCH

Soit $\varepsilon_1, \dots, \varepsilon_n$ les observations du processus log-GARCH(p, q) (4.1)-(4.2), de vecteur des paramètres inconnus $\theta = {}^t(\omega, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q) \in \Theta$.

Étant donnés (4.23) et (4.24), la log-vraisemblance est donnée par :

$$\widehat{\mathcal{L}}_n(\theta) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \left(\frac{1}{n} \sum_{t=1}^n \frac{\varepsilon_t^2}{\widehat{\sigma}_{t|t-1}^2(\theta)} + \widehat{H}_{t|t-1}^{(1)}(\theta) \right)$$

Par conséquent, une estimation de quasi-maximum de vraisemblance basée sur le filtre d'information, que l'on note $\widehat{\theta}_{QIF}$, est définie comme :

$$\widehat{\theta}_{QIF} = \arg \min_{\theta \in \Theta} \widehat{l}_n(\theta)$$

où \widehat{l}_n est le critère associé à la fonction de log-vraisemblance, et est donné par :

$$\widehat{l}_n(\theta) = \frac{1}{n} \sum_{i=1}^n \frac{\varepsilon_i^2}{\widehat{\sigma}_{i|i-1}^2(\theta)} + \widehat{H}_{i|i-1}^{(1)} \quad (4.25)$$

Pour les mêmes raisons discutées au chapitre précédant, nous optons encore une fois pour la minimisation stochastique de \widehat{l}_n . Plus précisément, nous employons l'algorithme fonctionnel lissé (*Smoothed Functional algorithm*) (CHIN (1997); KATKOVNIK et OY (1972); STYBLINSKI et TANG (1990)), désormais noté SF, qui appartient à la classe des méthodes à perturbation simultanée et consiste à approximer le gradient de \widehat{l}_n par sa convolution avec une densité uniforme qui assure une convergence rapide de l'algorithme. Ainsi, nous résumons l'algorithme SF appliqué à \widehat{l}_n comme suit :

Algorithme 4.6 Minimisation par SF de la log-vraisemblance du log-GARCH.

- **Calibration** : $(a, b, \lambda, \gamma) = (0.2, 1.5, 0.602, 0.101)$, $A = 10\%$ du nombre des itérations.
 - **Initialisation** : ϵ , la précision de convergence. iter, le nombre des itérations. $\theta_0 \in \Theta$.
 - **Pour** $k = 1$:iter, **faire** :
 - $a_k = a(A + k + 1)^{-\lambda}$ et $b_k = b(k + 1)^{-\gamma}$
 - Générer $\Delta_k \sim iid \mathcal{U}_{[-1,1]}$, $\Delta_k \in \mathbb{R}^d$
 - Générer $\delta_{k,1}, \delta_{k,2} \sim iid \mathcal{U}_{[0,1]}$
 - **Si** $max|Sp(\Phi(\theta_k))| < 1$
 - $y_{k,1}(\hat{\theta}_k) = \hat{l}_n(\hat{\theta}_k + b_k \Delta_k) + \delta_{k,1}$ et $y_{k,2}(\hat{\theta}_k) = \hat{l}_n(\hat{\theta}_k - b_k \Delta_k) + \delta_{k,2}$
 - $\hat{g}(\hat{\theta}_k) = \frac{y_{k,1}(\hat{\theta}_k) - y_{k,2}(\hat{\theta}_k)}{2b_k} \Delta_k$
 - $\hat{\theta}_{k+1} = \hat{\theta}_k - a_k \hat{g}(\hat{\theta}_k)$
 - **Si** $\|\hat{\theta}_k - \hat{\theta}_{k-1}\| < \epsilon$ ou $k = \text{iter}$
 - retourner $\hat{\theta}_k$
 - **Sinon** $k = k + 1$
 - **Fin si**
 - **Fin si**
 - **Fin Pour**.
-

Outre que l'estimation des paramètres du modèle log-GARCH, nous intégrons dans les parties suivantes l'estimation QIF basée sur l'algorithme SF, que l'on appelle désormais QIFSF, dans deux autres cadres enrichissant l'étude numérique du modèle log-GARCH, à savoir :

- (i) La mesure empirique du kurtosis sous les conditions réduites d'existence des moments considérées précédemment ;
- (ii) L'estimation par QIFSF des paramètres du log-GARCH en présence d'observations nulles.

4.3.1 Estimation du kurtosis droit et gauche

Dans cette partie, nous nous intéressons à l'estimation des mesures du kurtosis gauche et droit du processus log-GARCH en se basant sur une information très réduite sur l'existence de ses log-moments, à savoir, celui du premier ordre. Cela dit, l'utilisation de la mesure standard du kurtosis liée au moment d'ordre 4 s'avère inapproprié. D'où la nécessité de recourir à d'autres mesures, dites robustes, relativement aux contraintes d'existence des moments d'ordres élevés, chose qui n'est quasiment pas vérifiée dans le comportement des séries temporelles sujets d'ajustement par les modèles GARCH. Comme mesure alternative du kurtosis standard, ZENGA (2006) et FIORI (2008) avaient introduit de nouvelles mesures dites kurtosis droit et gauche décrites par une inégalité de part et d'autre de la médiane, basées sur le coefficient de concentration de Gini.

Dans ce qui suit, nous présentons brièvement les éléments pivots de la construction de telles mesures en montrant notre apport.

Soit ε_t un processus log-GARCH sous les conditions réduites d'existence des log-moments et $\widehat{\varepsilon}_t^*$ la série des résidus des rendements normalisés par l'écart-type conditionnel estimé $\widehat{\sigma}_t$, à savoir :

$$\widehat{\varepsilon}_t^* = \frac{\widehat{\varepsilon}_t}{\widehat{\sigma}_t}, \quad t = 1, 2, \dots, n$$

Selon ZENGA (2006) et FIORI (2008), le kurtosis de ε_t^* peut être interprété en termes d'inégalité de concentration des variables aléatoires conditionnelles X_R et X_L représentant respectivement les excès à droite et à gauche de ε_t^* par rapport à sa médiane notée M_ε , de moyennes respectives δ_R et δ_L , soit :

$$X_R = \varepsilon_t^* - M_\varepsilon \quad \text{si } \varepsilon_t^* > M_\varepsilon$$

$$X_L = M_\varepsilon - \varepsilon_t^* \quad \text{si } \varepsilon_t^* \leq M_\varepsilon$$

Ainsi, la mesure du kurtosis droit de ε_t^* , notée R , est donnée par le coefficient de concentration de Gini de X_R , soit :

$$R = \frac{\mathbb{E}|X_{R,1} - X_{R,2}|}{2\delta_R} \quad (4.26)$$

où $X_{R,1}$ et $X_{R,2}$ sont deux *iid* répliques de X_R .

De même, la mesure du kurtosis gauche de ε_t^* , notée L , est donnée par le coefficient de concentration de Gini de X_L , soit :

$$L = \frac{\mathbb{E}|X_{L,1} - X_{L,2}|}{2\delta_L} \quad (4.27)$$

où $X_{L,1}$ et $X_{L,2}$ sont deux *iid* répliques de X_L .

FIORI et BELTRAMI (2014) proposent des formes numériques estimant les deux mesures R et L , notées \widehat{R}_n et \widehat{L}_n telles que :

$$\widehat{R}_n = \frac{n^{-1} \sum_{i=k+1}^n \left(\frac{4i-2}{n} - 3 \right) \widehat{\varepsilon}_{(i)}^* - \left(\frac{k}{n} - 0.5 \right) (n-1) \widehat{\varepsilon}_{(k)}^*}{n^{-1} \sum_{i=k+1}^n \widehat{\varepsilon}_{(i)}^* + \left(\frac{k}{n} - 0.5 \right) \widehat{\varepsilon}_{(k)}^* - 0.5 F_n^{-1}(0.5)}$$

et

$$\widehat{L}_n = \frac{n^{-1} \sum_{i=k+1}^n \left(\frac{4i-2}{n} - 1 \right) \widehat{\varepsilon}_{(i)}^* - \left(\frac{k}{n} - 0.5 \right) (n-3) \widehat{\varepsilon}_{(k)}^*}{0.5 F_n^{-1}(0.5) - n^{-1} \sum_{i=1}^k \widehat{\varepsilon}_{(i)}^* + \left(\frac{k}{n} - 0.5 \right) \widehat{\varepsilon}_{(k)}^*}$$

où $k = \lfloor n/2 \rfloor$, $\widehat{\varepsilon}_{(\cdot)}^*$ est la statistique d'ordre liée aux résidus $\widehat{\varepsilon}_t^*$ et $F_n^{-1}(0.5)$ représente sa médiane empirique.

Nous proposons donc exploiter l'estimation de la variance conditionnelle abordée dans §4.2 dans l'amélioration des estimations \widehat{R}_n et \widehat{L}_n en les appliquant à la série

des résidus des rendements normalisés par l'écart-type conditionnel issus de (4.24), à savoir :

$$\widehat{\varepsilon}_t^* = \frac{\widehat{\varepsilon}_t}{\sqrt{\widehat{\sigma}_{t|t-1}^2}}, \quad t = 1, 2, \dots, n \quad (4.28)$$

Ces estimateurs sont fortement consistants pour toute variable aléatoire continue sous les conditions présentées dans le théorème suivant :

Théorème 4.1. (Fiori et Beltrami 2014) *Soit X une variable aléatoire continue de fonction quantile F^{-1} continue au voisinage de 0.5 et admettant une médiane unique. Si $\mathbb{E}|X|^s < \infty$ pour un certain $s > 0$, alors les estimateurs \widehat{R}_n et \widehat{L}_n sont fortement consistants respectivement pour les kurtosis droit (4.26) et gauche (4.27) de X .*

4.3.2 Estimation des paramètres du log-GARCH en présence d'observations nulles

Rappelons que par définition, le modèle log-GARCH n'est pas employable pour ajuster les données contenant des valeurs nulles. Bien que l'hypothèse $Prob(\eta_t = 0) = 0$ imposée par Christian FRANCO, WINTENBERGER et ZAKOIAN (2013a) rende théoriquement l'absence des valeurs nulles presque sûre, elle n'empêche pas en pratique l'occurrence de telles valeurs.

L'estimation basée sur le filtre d'information (QIFSF) en combinaison avec l'algorithme EM fournit une autre façon pour traiter par imputation le problème de la présence des rendements nuls comme valeurs manquantes afin d'obtenir une estimation des paramètres en se basant sur le même modèle (4.3).

Soient $\varepsilon_1, \dots, \varepsilon_n$ les observations issues d'un processus log-GARCH de paramètre inconnu $\theta = {}^t(\omega, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q) \in \Theta$, où :

$$\Theta = \{\theta \in \mathbb{R}^d / \max |Sp(\Phi(\theta))| < 1\} \quad \text{avec } d = p + q + 1.$$

Soit $\{\varepsilon_{i_1}, \varepsilon_{i_2}, \dots, \varepsilon_{i_m} / 1 \leq i_1 \leq \dots \leq i_m \leq n\}$ l'ensemble ordonné des observations nulles. Nous désignons par \mathcal{Y}_m la suite des "log-observations nulles" qui seront objets d'imputation, soit :

$$\mathcal{Y}_m = \{Y_{i_1}, \dots, Y_{i_m}\} \quad \text{avec } i_1 > 1$$

et

$$\mathcal{Y}_{n_1} = \{Y_1, Y_2, \dots, Y_{i_1-1}\} \quad \text{avec } n_1 = i_1 - 1$$

Ainsi, en se donnant une valeur initiale θ_0 , nous appliquons l'algorithme EM à partir de l'ensemble des observations \mathcal{Y}_{n_1} outre que les outputs du filtre d'information comme décrit ci-après :

- **Étape E (Expectation)** : Calculer à partir du filtre d'information (4.5) la prédiction de la mesure $\widehat{Y}_{i_1|i_1-1}(\widehat{\theta}_1)$ par laquelle Y_{i_1} est imputée, soit :

$$Y_{i_1}^* = \widehat{Y}_{i_1|i_1-1}(\widehat{\theta}_1) = \widehat{H}_{i_1|i_1-1}^{(1)} + m_w$$

- **Étape M (Maximization)** : Maximiser la fonction de log-vraisemblance $\widehat{\mathcal{L}}_{n_1+1}(\theta)$ évaluée relativement à $\mathcal{Y}_{n_1} \cup \{Y_{i_1}^*\}$, soit :

$$\widehat{\mathcal{L}}_{n_1+1}(\theta; \mathcal{Y}_{n_1}, Y_{i_1}^*) = -\frac{n_1+1}{2} \log(2\pi) - \frac{n_1+1}{2} \left(\frac{1}{n_1+1} \sum_{t=1}^{n_1+1} \frac{\varepsilon_t^2}{\widehat{\sigma}_{i_1|i_1-1}^2(\theta)} + \widehat{H}_{i_1|i_1-1}^{(1)}(\theta) \right)$$

Maximiser $\widehat{\mathcal{L}}_{n_1+1}$ revient à minimiser le critère $\widehat{l}_{n_1+1}(\theta)$ donné par :

$$\widehat{l}_{n_1+1}(\theta; \mathcal{Y}_{n_1}, Y_{i_1}^*) = \frac{1}{n_1+1} \sum_{t=1}^{n_1+1} \frac{\varepsilon_t^2}{\widehat{\sigma}_{i_1|i_1-1}^2(\theta)} + \widehat{H}_{i_1|i_1-1}^{(1)}(\theta)$$

Cela fournit dans une première itération de l'algorithme EM, une estimation de θ donnée par :

$$\widehat{\theta}_1 = \arg \min_{\theta \in \Theta} \widehat{l}_{n_1+1}(\theta)$$

Ensuite, nous posons $\mathcal{Y}_{n_2} = \{Y_1, Y_2, \dots, Y_{i_1}^*, \dots, Y_{i_2-1}\}$ avec $n_2 = i_2 - 1$, et nous poursuivons de manière ascendante l'imputation du reste des valeurs Y_{i_j} , $j = 2, \dots, m$ en exécutant récursivement les étapes E et M.

Finalement, nous obtenons une estimation de θ à partir des observations $\mathcal{Y}_{n_m+1} \cup \{Y_{i_m+1}, \dots, Y_n\}$ comme :

$$\widehat{\theta} = \arg \min_{\theta \in \Theta} \widehat{l}_n(\theta)$$

Dans le cas particulier où la première observation Y_1 correspond à un rendement nul ($i_1 = 1$), on ne dispose donc d'aucune observation antérieure à ε_1 . D'où, Y_1 est imputée par la moyenne théorique $\mathbb{E}Y_1$, soit :

$$Y_1^* = \frac{\omega + m_w \left(1 - \sum_{j=1}^q \beta_j\right)}{1 - \sum_{i=1}^r \phi_i} t(1, \dots, 1)$$

Nous résumons le traitement précédant dans l'algorithme suivant :

Algorithme 4.7 Estimation du log-GARCH par EM-QIFSFS.

- Fixer $\theta_0 \in \Theta$
- $\mathcal{Y}_n = \{Y_1, Y_2, \dots, Y_n\}$
- **Si** $i_1 = 1$, **alors**

$$Y_1 = Y_1^* = \frac{\omega + m_w \left(1 - \sum_{j=1}^q \beta_j\right)}{1 - \sum_{i=1}^r \phi_i} t(1, \dots, 1)$$

- **Sinon**,
 - **Pour** $j = 1 : m$, **faire** :
 - $\mathcal{Y}_{n_j} = \mathcal{Y}_n [1 : i_j - 1]$
 - Calculer $\widehat{H}_{i_j|i_{j-1}}^{(1)}(\mathcal{Y}_{n_j}; \theta_{j-1})$ et $\widehat{\sigma}_{i_j|i_{j-1}}^2(\mathcal{Y}_{n_j}; \theta_{j-1})$
 - Dédurre $\widehat{Y}_{i_j|i_{j-1}} = \widehat{H}_{i_1|i_{1-1}}^{(1)} + m_\xi$
 - $Y_{i_j}^* = \widehat{Y}_{i_j|i_{j-1}}$
 - Évaluer $\widehat{l}_{n_j+1}(\theta; \mathcal{Y}_{n_j}, Y_{i_j}^*)$
 - $\widehat{\theta}_j = \arg \min_{\theta \in \Theta} \widehat{l}_{n_j+1}(\theta)$
 - **Fin Pour**
 - $\widehat{\theta} = \arg \min_{\theta \in \Theta} \widehat{l}_n(\theta; \mathcal{Y}_{n_m+1} \cup \{Y_{i_m+1}, \dots, Y_n\})$
 - **Fin si**
-

4.4 Illustration numérique

Notations :

- $\widehat{\theta}_{QML}$: Vecteur des paramètres estimé par QML au sens de Christian FRANCO, WINTENBERGER et ZAKOIAN (2013a).
- $\widehat{\theta}_{ECQ}$: Vecteur des paramètres estimé par ECQ au sens de SUCARRAT et ESCRIBANO (2013).
- \widetilde{K}_{QIFSFS} : Vecteur des mesures estimées du kurtosis droit et gauche basées sur l'estimation par QIFSFS de la variance conditionnelle.
- \widetilde{K}_{QML} : Vecteur des mesures estimées du kurtosis droit et gauche basées sur l'estimation par QML de la variance conditionnelle.
- \widetilde{K}_{ECQ} : Vecteur des mesures estimées du kurtosis droit et gauche basées sur l'estimation par ECQ de la variance conditionnelle.

Dans cette partie, nous étudions la performance de l'estimation QIFSFS appliquée au modèle log-GARCH(1,1) avec une structure de log-moments réduite au premier ordre. Dans un premier temps, nous comparons en échantillons finis la performance des méthodes QIFSFS, QML et ECQ en termes de précision des estimations obtenues. Ensuite, nous évaluons la robustesse des mesures du kurtosis droit et gauche appliquées aux résidus du log-GARCH normalisés par l'écart-type conditionnel issu de (4.24).

Nous effectuons une simulation Monte Carlo pour $n \in \{500, 1000, 5000\}$ à partir

d'un modèle log-GARCH(1,1) de paramètre $t(1, 0.8, -0.5)$, avec $\eta_t \sim iid N(0, 1)$. Notons que par ce choix, le processus est stationnaire avec un second log-moment infini. Pour chaque n , nous avons généré 1000 répliquions issues du log-GARCH(1,1). Les moyennes des estimations $\hat{\theta}_{QML}$, $\hat{\theta}_{ECQ}$ et $\hat{\theta}_{QIFSF}$, leurs écarts types et leurs erreurs quadratiques moyennes (MSE) sur les 1000 répliquions sont présentées dans le tableau 4.1.

Par ailleurs, pour étudier les propriétés d'application de la méthode QIFSF pour l'estimation des paramètres en présences d'observations nulles comme décrit dans §4.3.2, nous réalisons dans un premier temps une expérience Monte Carlo similaire aux expériences précédentes avec $n = 10000$ réparti en 100 répliquions. Nous fixons π_0 à 5%, 10% et 20% de la taille de l'échantillon et nous choisissons d'une manière aléatoire πn observations et on leur affecte la valeur 0. Le tableau 4.3 contient les résultats des estimations et la figure 4.1 représente graphiquement les biais d'estimation en échantillons finis.

À titre d'application empirique, nous évaluons la performance de l'estimation EM-QIFSF comparativement aux méthodes existantes, appliquées à trois séries de rendements journaliers, à savoir, le prix du pétrole brut COWTI (West Texas Intermediate) (19/09/1983 – 23/08/1999), l'indice CAC40 (Cotation Assistée en Continu) (02/01/2005 – 31/12/2011) et l'indice FTSE100 (Indice footsie) (02/01/2005 – 31/12/2011) (Figure 4.2). Nous disposons de différentes proportions de zéros dans l'ensemble des trois séries, à savoir 1.47 %, 2.23 % et 5 % respectivement pour le COWTI, CAC40 et FTSE100¹. Il convient de noter que les rendements nuls incluent les rendements réellement observés et les jours fériés.

4.4.1 Estimation par QIFSF des paramètres du log-GARCH

Une première lecture du tableau 4.1 révèle une convergence des résultats de l'estimation par les trois méthodes avec une quasi surperformance de l'estimation QIFSF dans six estimations parmi les neufs réalisées pour chaque taille n . En particulier, pour toutes les tailles n , ω est le mieux estimé par QIFSF. Cela résulte du processus d'estimation à double étapes utilisée dans l'estimation ECQ où ω est estimé après l'estimation des autres paramètres (Christian FRANCQ et SUCARRAT (2013)). Cependant, pour un échantillon de grande taille, c'est-à-dire $n = 5000$, les trois estimations semblent relativement plus proches l'une de l'autre. En outre, parmi toutes les estimations, les petites valeurs des écarts types correspondent à $\hat{\theta}_{QIFSF}$ indiquant ainsi que les estimations sur l'ensemble des répliquions sont faiblement dispersées autour de $\hat{\theta}_{QIFSF}$ relativement aux dispersions de $\hat{\theta}_{QML}$ et $\hat{\theta}_{ECQ}$.

4.4.2 Estimation du kurtosis droit et gauche

De la même manière, nous nous appuyons sur la consistance des estimateurs droit et gauche du kurtosis (voir le théorème 4.1) pour les évaluer par simulation

1. Toutes ces séries sont disponibles sur le site www.investing.com

TABLE 4.1 – Propriétés en échantillons finis du log-GARCH(1,1) estimé par QML, ECQ et QIFSF.

n		QML		ECQ		QIFSF	
		$\hat{\theta}_{QML}$	MSE	$\hat{\theta}_{ECQ}$	MSE	$\hat{\theta}_{QIFSF}$	MSE
500	1	1.0068 (0.1106)	0.0123	0.9888 (0.1028)	0.0107	1.0635 (0.0312)	0.0050*
	0.8	0.8039 (0.0422)	0.0018	0.7971 (0.0298)	0.0009	0.8110 (0.0240)	0.0007*
	-0.5	-0.5025 (0.0537)	0.0029	-0.4961 (0.0329)	0.0011*	-0.5376 (0.0242)	0.0020
1000	1	1.0011 (0.0774)	0.0060	0.9990 (0.0714)	0.0051	1.0013 (0.0264)	0.0007*
	0.8	0.8004 (0.0316)	0.0010	0.8017 (0.0199)	0.0004*	0.7942 (0.0191)	0.0004*
	-0.5	-0.4945 (0.0420)	0.0018	-0.4961 (0.0241)	0.0006	-0.5013 (0.0199)	0.0004*
5000	1	1.0068 (0.0324)	0.0011	1.0069 (0.0291)	0.0009	0.9993 (0.0264)	0.0007*
	0.8	0.7984 (0.0098)	0.0001	0.7990 (0.0099)	0.0001*	0.7944 (0.0082)	0.0001
	-0.5	-0.5003 (0.0141)	0.0002*	-0.5003 (0.0244)	0.0006	-0.5013 (0.0244)	0.0006

¹ Les écarts types estimés sont affichés entre parenthèses.

² * représente la plus petite MSE.

Monte Carlo.

Concernant les propriétés en échantillons finies des estimations des mesures des kurtosis droit et gauche, les plus petites valeurs de la MSE pour tous les échantillons (tableau 4.2) confirment que le \tilde{K}_{QIFSF} surperforme les deux estimations \tilde{K}_{QML} et \tilde{K}_{ECQ} . En particulier, la performance de \tilde{K}_{QIFSF} s'améliore nettement pour un échantillon de grande taille ($n = 5000$) en vertu des valeurs du MSE qui atteint l'ordre de 10^{-3} . De plus, sachant que pour une distribution normale, les mesures du kurtosis droit et gauche sont toutes les deux proches de $\sqrt{2} - 1 \simeq 0,4142$, il est clair que les mesures estimées \tilde{K}_{QIFSF} sont robustes au comportement leptokurtique en restant les plus proches de \tilde{K}_0 . Au contraire, les valeurs de \tilde{K}_{QML} et \tilde{K}_{ECQ} restent proches de celles d'une distribution normale et significativement différentes des valeurs réelles correspondant aux séries simulées.

TABLE 4.2 – Propriétés en échantillons finis des mesures du kurtosis droit et gauche calculées par les estimation QML, ECQ et QIFSF.

n	\tilde{K}_0	QML		ECQ		QIFSF	
		\tilde{K}_{QML}	MSE	\tilde{K}_{ECQ}	MSfE	\tilde{K}_{QIFSF}	MSE
500	0.5242	0.4171 (0.0169)	0.1071	0.4159 (0.0170)	0.1083	0.5403 (0.0203)	0.0202*
	0.5367	0.4120 (0.0158)	0.1246	0.4108 (0.0157)	0.1259	0.5413 (0.0260)	0.0204*
1000	0.5116	0.4153 (0.0129)	0.0963	0.4145 (0.0128)	0.0971	0.5425 (0.0164)	0.0309*
	0.5484	0.4151 (0.0104)	0.1332	0.4146 (0.0103)	0.1337	0.5407 (0.0184)	0.0162*
5000	0.5467	0.4127 (0.0055)	0.1340	0.4126 (0.0055)	0.1341	0.5393 (0.0067)	0.0080*
	0.5424	0.4144 (0.0049)	0.1279	0.4143 (0.0049)	0.1281	0.5394 (0.0084)	0.0067*

¹ Les écarts types estimés sont mis entre parenthèses.

² * représente la plus petite MSE.

4.4.3 Estimation par QIFSF du log-GARCH en présence d'observations nulles

Étude Monte Carlo

L'examen visuel de la figure (4.1) montre qu'à l'exception de l'estimation de α_1 , les biais en échantillons finis des estimations de ω et β_1 augmentent faiblement à mesure que π_0 augmente, ce qui est attendu puisque les observations réelles sont perdues en traitant les zéros comme des valeurs manquantes. Plus précisément, outre que les valeurs proches des biais de $\hat{\omega}_{ECQ}$ et $\hat{\omega}_{QIFSF}$ l'un de l'autre, nous constatons que les deux approches ont conservé le même niveau de volatilité, contrairement à $\hat{\omega}_{QML}$, où l'effet des rendements nuls est plus significatif. Pour α_1 , les trois estimations ont fait diminuer le biais d'une manière plus importante pour les estimations ECQ et EM-QIFSF. Ainsi, la persistance du choc n'est pas fortement influencée. Quant à β_1 , le biais pour toutes les méthodes augmente légèrement par rapport aux autres paramètres tout en gardant le même niveau avec les pentes les plus faibles.

Étude Empirique

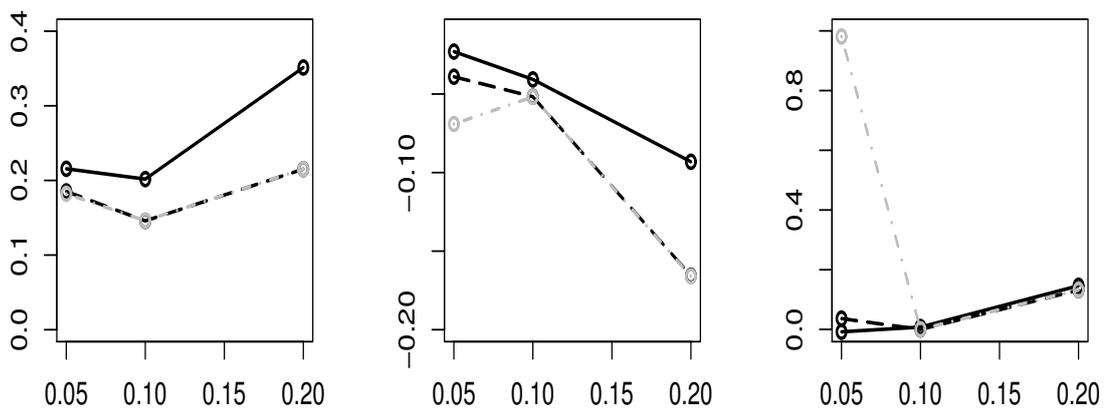
Comme l'indique le tableau 4.4, toutes les séries présentent les propriétés usuelles d'excès du kurtosis et de présence significative de l'effet ARCH. Pour chaque proportion π_0 , nous comparons notre estimation, à l'estimation ECQ (SUCARRAT et ESCRIBANO (2013)), ainsi qu'à l'estimation par QML (Christian FRANCO, WINTENBERGER et ZAKOIAN (2013a)) où les rendements nuls sont supprimés et les

TABLE 4.3 – Propriétés en échantillons finis du log-GARCH(1,1) estimé par QML, ECQ et EM-QIFSF pour différentes proportions d'observations nulles.

π_0		QML		ECQ		EM-QIFSF	
		$\hat{\theta}_{QML}$	Biais	$\hat{\theta}_{ECQ}$	Biais	$\hat{\theta}_{EM-QIFSF}$	Biais
0.05	1	1.2156	0.2156	1.1853	0.1853	1.1821	0.1821
	0.8	0.7770	-0.0230	0.7610	-0.0390	0.7309	-0.0691
	-0.5	-0.5081	-0.0081	-0.4637	0.0363	-0.4813	0.0187
0.1	1	1.2018	0.2018	1.1458	0.1458	1.1461	0.1461
	0.8	0.7593	-0.0407	0.7483	-0.0517	0.7482	-0.0518
	-0.5	-0.4922	0.0078	-0.5001	-0.0001	-0.4997	0.0003
0.2	1	1.3513	0.3513	1.2152	0.2152	1.2150	0.2152
	0.8	0.7069	-0.0931	0.6344	-0.1656	0.6340	-0.1660
	-0.5	-0.3538	0.1462	-0.3683	0.1317	-0.3680	0.1320

¹ π_0 : proportion des zéros injectés dans les données générées par le log-GARCH(1,1).

FIGURE 4.1 – Biais des estimations par QML (ligne noire continue), par ECQ (ligne noire pointillée) et par EM-QIFSF (ligne grise pointillée) pour différentes proportions d'observations nulles.



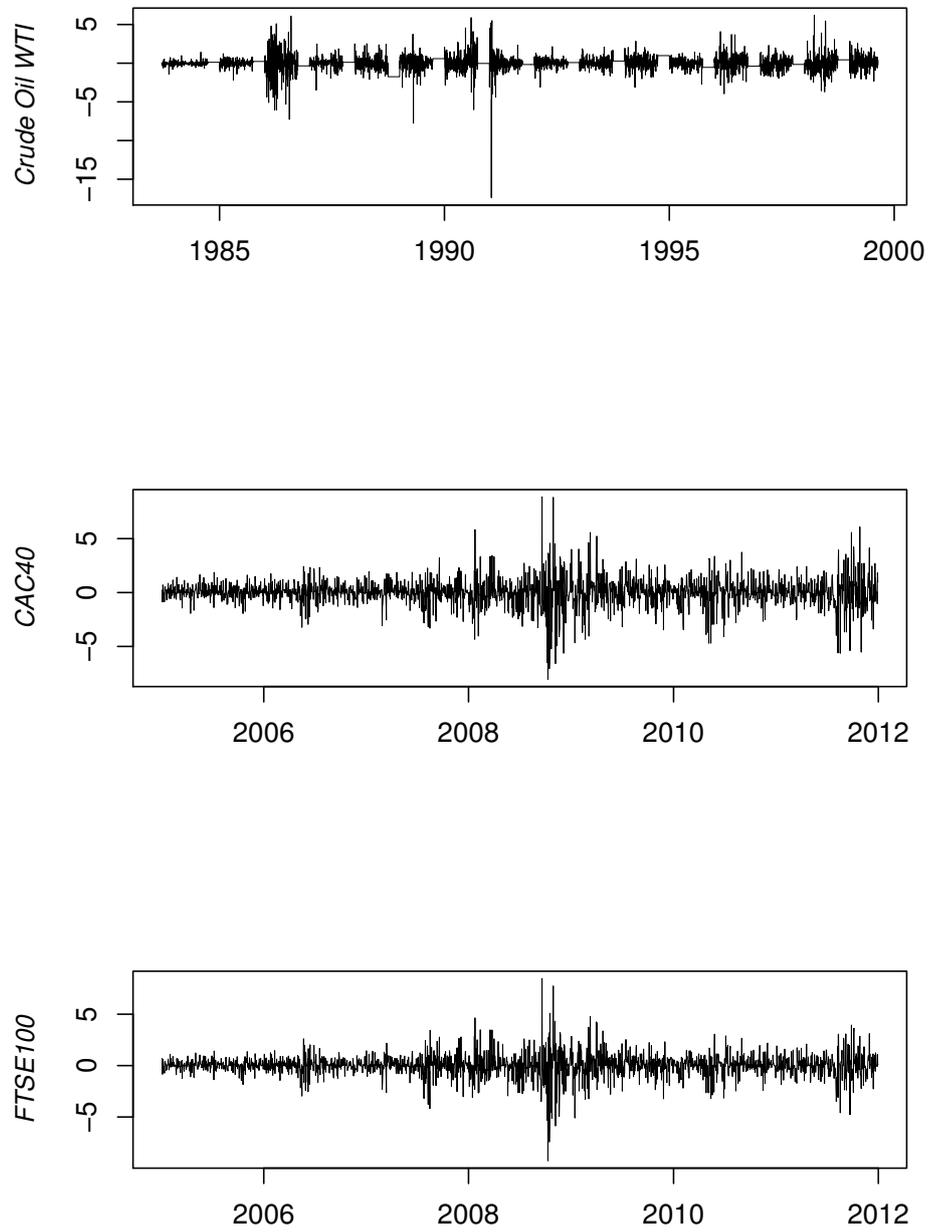


FIGURE 4.2 – Rendements journaliers en pourcentage.

TABLE 4.4 – Statistiques descriptives des rendements.

Série	n	π_0	s^2	s^4	test ARCH (p – value)
COWTI	5818	1.47%	0.8093	31.4116	0.0000
CAC40	2555	2.23%	1.4	21.1892	0.0000
FTSE100	2555	5%	1.0570	14.9132	0.0000

¹ s^2 , second moment empirique. s^4 , quatrième moment empirique. test ARCH, Test du multiplicateur de Lagrange au 1^{er} ordre. π_0 , proportion des observations nulles.

données sont reliées entre elles. Le tableau 4.5 présente le modèle log-GARCH(1,1) estimé pour chaque série, appliqué à la série des log-rendements en pourcentage r_t en utilisant les spécifications suivantes de la variance conditionnelle :

$$r_t = \varepsilon_t, \quad \varepsilon_t = \sigma_t \eta_t, \quad \eta_t \sim iidN(0, 1)$$

$$EM-QIFSF, ECQ: \quad \log \sigma_t^2 = \omega + \alpha_1 \log \varepsilon_{t-1}^2 + \beta_1 \log \sigma_{t-1}^2 \quad (4.29)$$

$$QML: \quad \log \sigma_t^2 = \omega + \alpha_1 \mathbb{1}_{\{\eta_{t-1}=0\}} \log \varepsilon_{t-1}^2 + \beta_1 \log \sigma_{t-1}^2 \quad (4.30)$$

Pour toutes les séries étudiées, la spécification EM-QIFSF correspond le mieux aux données selon le critère d'information bayésien (BIC) ayant également la plus grande log-vraisemblance. Ainsi, l'imputation par EM-QIFSF est plus performante que la méthode ECQ de sorte que les log-observations nulles imputées par notre méthode sont les plus compatibles avec les données auxquelles ils appartiennent. En outre, notre méthode semble plus adaptée que la méthode QML qui consiste à supprimer les zéros des données avant d'effectuer l'estimation des paramètres.

D'autre part, les plus petites différences d'estimation proviennent du COWTI (différence à la troisième décimale) ayant une proportion de zéros égale à 1,47% alors que les plus grandes différences proviennent du FTSE100 ayant une proportion de zéros égale à 5%. Dans ces cas, on peut voir que le nombre de zéros est l'une des causes produisant le biais d'estimation. Ce n'est pas une règle générale, car il suffit d'examiner les illustrations empiriques rapportées dans SUCARRAT et ESCRIBANO (2013) montrant un contre exemple.

TABLE 4.5 – Estimations empiriques des spécifications log-GARCH(1,1) des trois séries de rendements.

Série	Méthode	n	ω	α_1	β_1	log-Lik	BIC
COWTI	EM-QIFSF	5818	0.0802 (0.0006)	0.0546 (0.0004)	0.9373 (0.0009)	-4595.472	9216.495
	ECQ	5818	0.0832 (0.0040)	0.0576 (0.0050)	0.9333 (0.0050)	-6296.561	12618.67
	QML	5732	0.0852 (0.0062)	0.0580 (0.0076)	0.9336 (0.0076)	-4782.213	9589.244
CAC40	EM-QIFSF	2555	1.3471 (0.0059)	0.3568 (0.0075)	0.1405 (0.0054)	-2344.437	4712.412
	ECQ	2555	0.0701 (0.0030)	0.0345 (0.0052)	0.9487 (0.0052)	-3571.9	7167.337
	QML	2498	1.3321 (0.0190)	0.3524 (0.0422)	0.1506 (0.0422)	-4481.519	8986.508
FTSE100	EM-QIFSF	2555	1.0704 (0.0136)	0.3746 (0.0053)	0.0913 (0.0327)	-2349.862	4723.261
	ECQ	2555	0.0641 (0.0027)	0.0346 (0.0039)	0.9564 (0.0039)	-3099.961	6223.459
	QML	2427	1.0330 (0.0195)	0.3648 (0.0429)	0.1142 (0.0429)	-3768.716	7560.815

¹ Les écarts types estimés sont indiqués entre parenthèses.

² log-Lik est la valeur de log-vraisemblance.

4.5 Conclusion

Dans des conditions réduites d'existence des moments, nous avons proposé une estimation par quasi-maximum de vraisemblance des paramètres du modèle log-GARCH combinée avec le filtre d'information et l'algorithme fonctionnel lissé. Notre méthode a été utilisée pour l'imputation des observations nulles considérées comme des valeurs manquantes par un algorithme EM, tandis que les écarts types conditionnels estimés ont été utilisés pour normaliser les résidus du log-GARCH afin de calculer les mesures du kurtosis droit et gauche. En échantillons finis, l'estimation EM-QIFSF a le potentiel de mieux s'adapter au modèle log-GARCH et aussi de traiter avec précision le problème de l'estimation des paramètres en présence des observations nulles comparativement aux méthodes existantes. En particulier, l'algorithme surperforme largement les autres estimations en termes d'optimalité. En outre, notre approche fournit des écarts-types conditionnels estimés qui, combinés aux mesures du kurtosis droit et gauche, sont capables de mieux refléter le comportement leptokurtique des séries ajustées par le log-GARCH.

Conclusions et Perspectives

Dans cette thèse, nous nous sommes intéressés à l'estimation par quasi-maximum de vraisemblance des paramètres du modèle GARCH standard et de deux de ses extensions, à savoir, les modèles CGARCH et log-GARCH. En effet, nous sommes parvenus à établir un ensemble d'algorithmes permettant une telle estimation dans un cadre plus relaxé relativement aux approches rapportées dans la littérature.

Grâce au filtre de Kalman sous contrainte basé sur la méthode de troncature de la densité de probabilité, nous avons traité dans le chapitre 2 le problème de positivité de la variance conditionnelle générée par le GARCH standard. Les algorithmes proposés basés sur la représentation espace d'état et la structure de covariance qui en découle ont permis de la pré-estimer d'une manière robuste en satisfaisant automatiquement les contraintes relaxées de Nelson et Cao sans les identifier ni les imposer à priori. Outre qu'elle remédie aux contradictions empiriques avec les hypothèses de positivité des paramètres, notre approche a apporté deux généralisations dans l'estimation du GARCH(p, q) standard. La première concerne l'utilisation des contraintes de Nelson et Cao pour tous ordres $p, q \geq 1$. La deuxième a rendu le GARCH capable de tenir compte de l'information de bornitude de la volatilité lors de l'estimation des paramètres. Ainsi, des simulations de type Monte Carlo pour des GARCH de différents ordres, d'autres basées sur une contrainte de bornitude de la volatilité, ont montré la performance de notre approche dans l'amélioration de la précision des estimations du QML standard.

Le chapitre 3 s'est focalisé sur la routine de maximisation de la vraisemblance. Dans cette partie de la thèse, l'algorithme d'approximation stochastique est d'abord appliqué au GARCH(p, q) en prouvant sa convergence local. L'étude empirique de cette combinaison sur les séries de taux de change journalier du mark/dollar, du franc/dollar et du yen/dollar ont exhibé l'effet de l'utilisation de tel algorithme dans l'amélioration de l'optimalité de l'estimation proposée au chapitre 2. Nous avons également prouvé la convergence local de l'algorithme SPSA pour le modèle CGARCH à travers l'espace des paramètres que l'on a défini à l'aide des espaces des paramètres

associés aux composantes GARCH(1,1). Ainsi, nous avons révélé la performance du modèle à travers son application à la série de l'indice S&P500.

Dans le même cadre de relaxation, nous avons consacré le dernier chapitre à l'estimation des modèles log-GARCH. Cette fois-ci, la relaxation a porté sur l'existence des moments, notamment celui d'ordre 4. En effet, l'algorithme proposé a combiné le filtre d'information pour l'estimation de la variance conditionnelle et l'algorithme fonctionnel lissé pour l'estimation des paramètres. De plus, cet algorithme a été réadapté à deux fins. D'une part, nous avons fourni des estimations robustes et plus précises des mesures du kurtosis droit et gauche basées sur la variance conditionnelle estimée par le filtre d'information. Nous avons illustré ces résultats à travers des simulations Monte Carlo confirmant la performance des estimations proposées. D'autre part, nous avons rendu le modèle log-GARCH applicable aux séries contenant des observations nulles traitées comme des valeurs manquantes en intégrant dans l'algorithme initial l'algorithme EM. Des simulations Monte Carlo ainsi que l'application empirique de cette approche aux séries du prix du pétrole brut WTI (COWTI), l'indice CAC40 et l'indice FTSE100, ont tous confirmé la précision et l'optimalité des estimations obtenus comparativement aux méthodes rapportées dans la littérature.

Limites et Perspectives

Dans la perspective de ces travaux, il sera intéressant de développer l'algorithme QCK dans le sens de considérer d'autres contraintes/informations sur le comportement de la volatilité, notamment celui de changement de régime. Une formulation adéquate de la représentation espace d'état du GARCH ainsi que le système des contraintes reflétant un tel comportement rendra le nouveau modèle sujet de comparaison au modèle MS-GARCH (HAAS, MITNIK et PAOLELLA (2004) ; LIU (2006)).

Notons également que l'extension des algorithmes proposés dans ce travail appliqués au log-GARCH symétrique est faisable au cas asymétrique (Christian FRANCO, WINTENBERGER et ZAKOIAN (2013a)) en modifiant la représentation espace d'état associée.

Une autre extension de ce travail est l'amélioration des méthodes d'optimisation utilisées afin d'obtenir des optimums globaux comme des estimations de QML au lieu de se restreindre à des estimations locaux, sans oublier l'importance d'étudier les propriétés statistiques de ces estimations, à savoir, la consistance, la normalité asymptotique, l'efficacité, etc.

Il reste à mentionner quelques limites de ce travail. En effet, les algorithmes proposés sont fortement liés au choix des valeurs initiales du vecteur des paramètres. Une telle limite fait de ce travail un sujet d'amélioration en termes de robustesse et de stabilité des estimations obtenues relativement au choix des valeurs initiales. Une

deuxième limite de ce travail est de se restreindre aux spécifications GARCH dans la décomposition de la variance conditionnelle du CGARCH alors que la considération des autres spécifications (IGARCH, EGARCH, etc.) pourrait mieux capter la dynamique de la volatilité du modèle.

Bibliographie

- [1] Abdelhakim AKNOUCHE. « Periodic autoregressive stochastic volatility ». In : *Statistical Inference for Stochastic Processes* 20.2 (2017), pages 139-177 (cf. pages 75, 78, 83).
- [2] Jelloul ALLAL et Mohammed BENMOUMEN. « Parameter estimation for GARCH(1,1) models based on Kalman filter ». In : *Advances and Applications in Statistics* 25.2 (2011), pages 115-130 (cf. pages 4, 43, 57).
- [3] Stanislav ANATOLYEV. « Volatility filtering in estimation of kurtosis (and variance) ». In : *Dependence Modeling* 7.1 (2019), pages 1-23 (cf. page 76).
- [4] Torben G ANDERSEN et Tim BOLLERSLEV. « Heterogeneous information arrivals and return volatility dynamics : Uncovering the long-run in high frequency returns ». In : *The journal of Finance* 52.3 (1997), pages 975-1005 (cf. page 16).
- [5] Torben G ANDERSEN, Tim BOLLERSLEV, Francis X DIEBOLD et Paul LABYS. « The distribution of realized exchange rate volatility ». In : *Journal of the American statistical association* 96.453 (2001), pages 42-55 (cf. page 16).
- [6] Brian DO ANDERSON et John B MOORE. *Optimal filtering*. Courier Corporation, 2012 (cf. pages 20, 22, 38, 44).
- [7] NATALIA BAHAMONDE. « Nonlinear Parameter Estimation for State-Space ARCH Models with Missing Observations ». In : (2012) (cf. page 77).
- [8] Jushan BAI et Serena NG. « Tests for skewness, kurtosis, and normality for time series data ». In : *Journal of Business & Economic Statistics* 23.1 (2005), pages 49-60 (cf. page 76).
- [9] Richard T BAILLIE et Tim BOLLERSLEV. « The message in daily exchange rates : a conditional-variance tale ». In : *Journal of Business & Economic Statistics* 20.1 (2002), pages 60-68 (cf. page 60).
- [10] Geert BEKAERT, Campbell R HARVEY et Christian LUNDBLAD. « Liquidity and expected returns : Lessons from emerging markets ». In : *The review of financial studies* 20.6 (2007), pages 1783-1831 (cf. page 76).

- [11] Ernst R BERNDT, Bronwyn H HALL, Robert E HALL et Jerry A HAUSMAN. « Estimation and inference in nonlinear structural models ». In : *Annals of Economic and Social Measurement, Volume 3, number 4*. NBER, 1974, pages 653-665 (cf. page 5).
- [12] Shalabh BHATNAGAR, H.L PRASHANTH et L.A PRASHANTH. « Lecture Notes in Control and Information Sciences 434 ». In : (2013) (cf. pages 28, 59, 67).
- [13] Katarzyna BIEN, Ingmar NOLTE et Winfried POHLMEIER. « An inflated multivariate integer count hurdle model : an application to bid and ask quote dynamics ». In : *Journal of Applied Econometrics* 26.4 (2011), pages 669-707 (cf. page 76).
- [14] Tim BOLLERSLEV. « Generalized autoregressive conditional heteroskedasticity ». In : *Journal of econometrics* 31.3 (1986), pages 307-327 (cf. pages 2, 5, 9-11, 13, 14, 31, 33, 58).
- [15] Tim BOLLERSLEV. « A conditionally heteroskedastic time series model for speculative prices and rates of return ». In : *The review of economics and statistics* (1987), pages 542-547 (cf. page 14).
- [16] Tim BOLLERSLEV. « On the correlation structure for the generalized autoregressive conditional heteroskedastic process ». In : *Journal of time series analysis* 9.2 (1988), pages 121-131 (cf. page 13).
- [17] Tim BOLLERSLEV et Jeffrey M WOOLDRIDGE. « Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances ». In : *Econometric reviews* 11.2 (1992), pages 143-172 (cf. page 14).
- [18] Tim BOLLERSLEV et Jonathan H WRIGHT. « Semiparametric estimation of long-memory volatility dependencies : The role of high-frequency data ». In : *Journal of econometrics* 98.1 (2000), pages 81-106 (cf. page 16).
- [19] Philippe BOUGEROL et Nico PICARD. « Stationarity of GARCH processes and of some nonnegative time series ». In : *Journal of econometrics* 52.1-2 (1992), pages 115-127 (cf. page 10).
- [20] Peter J BROCKWELL, Richard A DAVIS et Stephen E FIENBERG. *Time series : theory and methods : theory and methods*. Springer Science & Business Media, 1991 (cf. page 77).
- [21] Robert Grover BROWN. *Introduction to random signal analysis and Kalman filtering*. Tome 8. Wiley New York, 1983 (cf. page 81).
- [22] Xing-Min CHEN et Chao GAO. « On the Convergence of Smoothed Functional Stochastic Optimization Algorithms ». In : *IFAC-PapersOnLine* 48.28 (2015), pages 229-233 (cf. page 28).
- [23] Tien-Li CHIA. « Parameter identification and state estimation of constrained systems ». Thèse de doctorat. Case Western Reserve University, 1985 (cf. page 23).

- [24] Daniel C CHIN. « Comparative study of stochastic algorithms for system optimization based on gradient approximations ». In : *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 27.2 (1997), pages 244-249 (cf. pages 28, 83).
- [25] Rama CONT. « Empirical properties of asset returns : stylized facts and statistical issues ». In : (2001) (cf. page 76).
- [26] Francis X DIEBOLD. *Empirical modeling of exchange rate dynamics*. Tome 303. Springer Science & Business Media, 2012 (cf. page 14).
- [27] Zhuanxin DING et Clive WJ GRANGER. « Modeling volatility persistence of speculative returns : a new approach ». In : *Journal of econometrics* 73.1 (1996), pages 185-215 (cf. pages 13, 16, 57).
- [28] Zhuanxin DING, Clive WJ GRANGER et Robert F ENGLE. « A long memory property of stock market returns and a new model ». In : *Journal of empirical finance* 1.1 (1993), pages 83-106 (cf. page 3).
- [29] RF ENGLE, T ITO et WL LIN. « Meteor Showers or Heat Waves ? Heteroskedastic Intra-Daily Volatility in the 20It must be noted that in last 20 years (1995-2014) tranquil phase occupies over 70% of the total period ». In : *Foreign Exchange Market. Econometrica* 58 (1990), pages 525-542 (cf. pages 2, 31).
- [30] Robert F ENGLE. « Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation ». In : *Econometrica : Journal of the Econometric Society* (1982), pages 987-1007 (cf. pages 2, 8, 9, 58).
- [31] Robert F ENGLE et Tim BOLLERSLEV. « Modelling the persistence of conditional variances ». In : *Econometric reviews* 5.1 (1986), pages 1-50 (cf. pages 10, 14).
- [32] Robert F ENGLE et Gary LEE. « A long-run and short-run component model of stock return volatility ». In : *Cointegration, Causality, and Forecasting : A Festschrift in Honour of Clive WJ Granger* (1999), pages 475-497 (cf. pages 3, 17, 64).
- [33] Robert F ENGLE et Jeffrey R RUSSELL. « Autoregressive conditional duration : a new model for irregularly spaced transaction data ». In : *Econometrica* (1998), pages 1127-1162 (cf. page 76).
- [34] Eugene F FAMA. « The behavior of stock-market prices ». In : *The journal of Business* 38.1 (1965), pages 34-105 (cf. page 1).
- [35] Jianqing FAN, Lei QI et Dacheng XIU. « Quasi-maximum likelihood estimation of GARCH models with heavy-tailed likelihoods ». In : *Journal of Business & Economic Statistics* 32.2 (2014), pages 178-191 (cf. pages 14, 59).

- [36] Guillermo FERREIRA, Jean P NAVARRETE, Francisco J RODRIGUEZ-CORTÉS et Jorge MATEU. « Estimation and prediction of time-varying GARCH models through a state-space representation : a computational approach ». In : *Journal of Statistical Computation and Simulation* 87.12 (2017), pages 2430-2449 (cf. page 4).
- [37] Anna Maria FIORI. « Measuring kurtosis by right and left inequality orders ». In : *Communications in Statistics—Theory and Methods* 37.17 (2008), pages 2665-2680 (cf. pages 76, 84, 85).
- [38] Anna Maria FIORI et Davide BELTRAMI. « Asymptotic estimation of right and left kurtosis measures, with applications to finance ». In : *Proceedings of the 46th Scientific Meeting of the Italian Statistical Society*. Citeseer. 2012 (cf. page 76).
- [39] Anna Maria FIORI et Davide BELTRAMI. « Right and left kurtosis measures : large sample estimation and an application to financial returns ». In : *Stat* 3.1 (2014), pages 95-108 (cf. pages 76, 85).
- [40] C FRANCO. « Jean-michel zakoian and ebrary inc garch models structure ». In : *Statistical Inference, and Financial Applications* (2010) (cf. pages 3, 10, 11, 15, 36, 42, 58).
- [41] Christian FRANCO, Guillaume LEPAGE et Jean-Michel ZAKOIAN. « Two-stage non Gaussian QML estimation of GARCH models and testing the efficiency of the Gaussian QMLE ». In : *Journal of Econometrics* 165.2 (2011), pages 246-257 (cf. pages 14, 59).
- [42] Christian FRANCO et Genaro SUCARRAT. « An exponential chi-squared QMLE for Log-GARCH models via the ARMA representation ». In : (2013) (cf. pages 75, 78, 89).
- [43] Christian FRANCO, Olivier WINTENBERGER et Jean-Michel ZAKOIAN. « GARCH models without positivity constraints : Exponential or Log GARCH ? » In : *Journal of Econometrics* 177.1 (2013), pages 34-46 (cf. pages 3, 18, 19, 75-77, 86, 88, 91, 97).
- [44] Christian FRANCO, Olivier WINTENBERGER et Jean-Michel ZAKOIAN. « GARCH models without positivity constraints : Exponential or Log GARCH ? » In : *Journal of Econometrics* 177.1 (2013), pages 34-46 (cf. page 35).
- [45] Christian FRANCO et Jean-Michel ZAKOIAN. « Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes ». In : *Bernoulli* 10.4 (2004), pages 605-637 (cf. page 14).
- [46] Christian FRANCO et Jean-Michel ZAKOIAN. *GARCH models : structure, statistical inference and financial applications*. John Wiley & Sons, 2019 (cf. pages 19, 76).
- [47] Christian FRANCO et Jean-Michel ZAKOIAN. « Linear-representation based estimation of stochastic volatility models ». In : *Scandinavian Journal of Statistics* 33.4 (2006), pages 785-806 (cf. page 75).

- [48] Donald Charles FRASER. « A new technique for the optimal smoothing of data ». Thèse de doctorat. Massachusetts Institute of Technology, 1967 (cf. page 26).
- [49] Kenneth R FRENCH, G William SCHWERT et Robert F STAMBAUGH. « Expected stock returns and volatility ». In : *Journal of financial Economics* 19.1 (1987), page 3 (cf. pages 2, 31).
- [50] J GEWEKE. « Comment on : modelling the persistence of conditional variances ». In : *Econometric Reviews* 5 (1986), pages 57-61 (cf. pages 3, 17, 75).
- [51] Alexios GHALANOS, Maintainer Alexios GHALANOS et LinkingTo RCPP. « Package 'rugarch' ». In : (2019) (cf. page 32).
- [52] Lawrence R GLOSTEN, Ravi JAGANNATHAN et David E RUNKLE. « On the relation between the expected value and the volatility of the nominal excess return on stocks ». In : *The journal of finance* 48.5 (1993), pages 1779-1801 (cf. page 3).
- [53] Christian GOURIÉROUX et Alain MONFORT. *Séries temporelles et modèles dynamiques*. Economica, 1995 (cf. pages 3, 26).
- [54] Markus HAAS, Stefan MITTNIK et Marc S PAOLELLA. « A new approach to Markov-switching GARCH models ». In : *Journal of financial econometrics* 2.4 (2004), pages 493-530 (cf. page 97).
- [55] Peter HALL et Qiwei YAO. « Inference in ARCH and GARCH models with heavy-tailed errors ». In : *Econometrica* 71.1 (2003), pages 285-317 (cf. page 14).
- [56] Andrew C HARVEY. *Forecasting, structural time series models and the Kalman filter*. Cambridge university press, 1990 (cf. pages 20, 22).
- [57] Jerry A HAUSMAN, Andrew W LO et A Craig MACKINLAY. « An ordered probit analysis of transaction stock prices ». In : *Journal of financial economics* 31.3 (1992), pages 319-379 (cf. page 76).
- [58] SD HAYWARD. « Constrained Kalman filter for least-squares estimation of time-varying beamforming weights ». In : *INSTITUTE OF MATHEMATICS AND ITS APPLICATIONS CONFERENCE SERIES*. Tome 67. 1. Oxford University Press. 1998, pages 113-126 (cf. page 23).
- [59] Changli HE et Timo TERÄSVIRTA. « Fourth moment structure of the GARCH (p, q) process ». In : *Econometric Theory* (1999), pages 824-846 (cf. pages 12, 13, 42).
- [60] Changli HE et Timo TERÄSVIRTA. « Properties of the Autocorrelation Function of Squared Observations for Second-order Garch Processes Under Two Sets of Parameter Constraints ». In : *Journal of Time Series Analysis* 20.1 (1999), pages 23-30 (cf. pages 2, 13, 31, 42).
- [61] Robert V HOGG. « More light on the kurtosis and related statistics ». In : *Journal of the American Statistical Association* 67.338 (1972), pages 422-424 (cf. page 76).

- [62] Robert V HOGG. « Adaptive robust procedures : A partial review and some suggestions for future applications and theory ». In : *Journal of the American Statistical Association* 69.348 (1974), pages 909-923 (cf. page 76).
- [63] Lajos HORV et Piotr KOKOSZKA. « GARCH processes : structure and estimation ». In : *Bernoulli* 9.2 (2003), pages 201-227 (cf. page 14).
- [64] David A HSIEH. « Modeling heteroscedasticity in daily foreign-exchange rates ». In : *Journal of Business & Economic Statistics* 7.3 (1989), pages 307-317 (cf. page 14).
- [65] Y HWANG. « Asymmetric long memory GARCH in exchange return ». In : *Economics Letters* 73.1 (2001), pages 1-5 (cf. page 16).
- [66] Rudolph Emil KALMAN. « A new approach to linear filtering and prediction problems ». In : (1960) (cf. page 20).
- [67] Menelaos KARANASOS. « The second moment and the autocovariance function of the squared errors of the GARCH model ». In : *Journal of Econometrics* 90.1 (1999), pages 63-76 (cf. pages 12, 13, 17, 65).
- [68] V Ya KATKOVNIK et KULCHITS. OY. « Convergence of a class of random search algorithms ». In : *Automation and Remote Control* 33.8 (1972), pages 1321-1326 (cf. pages 28, 83).
- [69] Jack KIEFER et Jacob WOLFOWITZ. « Stochastic estimation of the maximum of a regression function ». In : *The Annals of Mathematical Statistics* 23.3 (1952), pages 462-466 (cf. page 56).
- [70] Tae-Hwan KIM et Halbert WHITE. « On more robust estimation of skewness and kurtosis ». In : *Finance Research Letters* 1.1 (2004), pages 56-73 (cf. page 76).
- [71] MV KULIKOVA, JV TSYGANOVA et G Yu KULIKOV. « SVD-based state and parameter estimation approach for generalized Kalman filtering with application to GARCH-in-Mean estimation ». In : *Journal of Computational and Applied Mathematics* (2019), page 112487 (cf. page 4).
- [72] Sang-Won LEE et Bruce E HANSEN. « Asymptotic theory for the GARCH (1, 1) quasi-maximum likelihood estimator ». In : *Econometric theory* 10.1 (1994), pages 29-52 (cf. page 14).
- [73] David A LESMOND, Joseph P OGDEN et Charles A TRZCINKA. « A new estimate of transaction costs ». In : *The review of financial studies* 12.5 (1999), pages 1113-1141 (cf. page 76).
- [74] Ross LEVINE et Sergio L SCHMUKLER. « Internationalization and stock market liquidity ». In : *Review of Finance* 10.1 (2006), pages 153-187 (cf. page 76).
- [75] Shiqing LING. « On the probabilistic properties of a double threshold ARMA conditional heteroskedastic model ». In : *Journal of Applied probability* (1999), pages 688-705 (cf. page 12).

- [76] Shiqing LING et Michael MCALEER. « Necessary and sufficient moment conditions for the GARCH (r, s) and asymmetric power GARCH (r, s) models ». In : *Econometric theory* 18.3 (2002), pages 722-729 (cf. page 12).
- [77] Ji-Chun LIU. « Stationarity of a Markov-switching GARCH model ». In : *Journal of Financial Econometrics* 4.4 (2006), pages 573-593 (cf. page 97).
- [78] Robin L LUMSDAINE. « Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH (1, 1) and covariance stationary GARCH (1, 1) models ». In : *Econometrica : Journal of the Econometric Society* (1996), pages 575-596 (cf. page 14).
- [79] Kaushik MAHATA et Torsten SÖDERSTRÖM. « Improved estimation performance using known linear constraints ». In : *Automatica* 40.8 (2004), pages 1307-1318 (cf. page 23).
- [80] John MAHEU. « Can GARCH models capture long-range dependence? » In : *Studies in Nonlinear Dynamics & Econometrics* 9.4 (2005) (cf. page 63).
- [81] Benoit B MANDELBROT. « The variation of certain speculative prices ». In : *Fractals and scaling in finance*. Springer, 1963, pages 371-418 (cf. pages 1, 13).
- [82] Anders MILHØJ. « The moment structure of ARCH processes ». In : *Scandinavian Journal of Statistics* (1985), pages 281-292 (cf. page 9).
- [83] Anders MILHØJ. *A multiplicative parameterization of ARCH models*. 1987 (cf. pages 3, 17, 75).
- [84] JJA MOORS. « A quantile alternative for kurtosis ». In : *Journal of the Royal Statistical Society : Series D (The Statistician)* 37.1 (1988), pages 25-32 (cf. page 76).
- [85] Daniel B NELSON. « Conditional heteroskedasticity in asset returns : A new approach ». In : *Econometrica : Journal of the Econometric Society* (1991), pages 347-370 (cf. pages 2, 3, 14, 31).
- [86] Daniel B NELSON et Charles Q CAO. « Inequality constraints in the univariate GARCH model ». In : *Journal of Business & Economic Statistics* 10.2 (1992), pages 229-235 (cf. pages 2, 5, 11-13, 31-33, 37, 57, 60, 61).
- [87] Sebastian OSSANDÓN et Natalia BAHAMONDE. « A new nonlinear formulation for GARCH models ». In : *Comptes Rendus Mathématique* 351.5-6 (2013), pages 235-239 (cf. pages 4, 43).
- [88] Sastry G PANTULA. « Modeling the persistence of conditional variances : a comment ». In : *Econometric Reviews* 5 (1986), pages 79-97 (cf. pages 3, 17, 75).
- [89] John PORRILL. « Optimal combination and constraints for geometrical sensor data ». In : *The International Journal of Robotics Research* 7.6 (1988), pages 66-77 (cf. page 23).

- [90] Herbert ROBBINS et Sutton MONRO. « A stochastic approximation method ». In : *The annals of mathematical statistics* (1951), pages 400-407 (cf. pages 27, 56).
- [91] Reuven Y RUBINSTEIN et Dirk P KROESE. *Simulation and the Monte Carlo method*. Tome 10. John Wiley & Sons, 2016 (cf. page 29).
- [92] Jeffrey R RUSSELL et Robert F ENGLE. « A discrete-state continuous-time model of financial transactions prices and times : The autoregressive conditional multinomial–autoregressive conditional duration model ». In : *Journal of Business & Economic Statistics* 23.2 (2005), pages 166-180 (cf. page 76).
- [93] Tina Hviid RYDBERG et Neil SHEPHARD. « Dynamics of trade-by-trade price movements : decomposition and models ». In : *Journal of Financial Econometrics* 1.1 (2003), pages 2-25 (cf. page 76).
- [94] Abdeljalil SETTAR et Mohammed BADAOU. « On the treatment of zero returns in the estimation of log-GARCH model : Empirical study ». In : *IOP Conference Series : Materials Science and Engineering*. Tome 1099. 1. IOP Publishing. 2021, pages 12-29 (cf. page 5).
- [95] Abdeljalil SETTAR, Nadia Idrissi FATMI et Mohammed BADAOU. « New Approach in Dealing with the Non-Negativity of the Conditional Variance in the Estimation of GARCH Model ». In : *Central European Journal of Economic Modelling and Econometrics* (2021), pages 55-74 (cf. page 5).
- [96] Abdeljalil SETTAR, Nadia Idrissi FATMI et Mohammed BADAOU. « Numerical estimation of log-GARCH model with relaxed moment assumptions ». In : *Journal of Physics : Conference Series*. Tome 1743. 1. IOP Publishing. 2021, pages 12-36 (cf. page 5).
- [97] Nobutaka SHIMADA, Yoshiaki SHIRAI, Yoshinori KUNO et Jun MIURA. « Hand gesture estimation and model refinement using monocular camera-ambiguity limitation by inequality constraints ». In : *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE. 1998, pages 268-273 (cf. page 24).
- [98] MM SIDDIQUI et K RAGHUNANDANAN. « Asymptotically robust estimators of location ». In : *Journal of the American Statistical Association* 62.319 (1967), pages 950-953 (cf. page 76).
- [99] Dan SIMON. *Optimal state estimation : Kalman, H infinity, and nonlinear approaches*. John Wiley & Sons, 2006 (cf. pages 20, 24, 26, 38, 44).
- [100] Dan SIMON et Tien Li CHIA. « Kalman filtering with state equality constraints ». In : *IEEE transactions on Aerospace and Electronic Systems* 38.1 (2002), pages 128-136 (cf. page 23).
- [101] Dan SIMON et Donald L SIMON. « Constrained Kalman filtering via density function truncation for turbofan engine health estimation ». In : *International Journal of Systems Science* 41.2 (2010), pages 159-171 (cf. page 24).

- [102] James C SPALL. « A stochastic approximation technique for generating maximum likelihood parameter estimates ». In : *1987 American Control Conference*. IEEE. 1987, pages 1161-1167 (cf. page 57).
- [103] James C SPALL. « Multivariate stochastic approximation using a simultaneous perturbation gradient approximation ». In : *IEEE transactions on automatic control* 37.3 (1992), pages 332-341 (cf. pages 27, 56).
- [104] James C SPALL. « A one-measurement form of simultaneous perturbation stochastic approximation ». In : *Automatica* 33.1 (1997), pages 109-112 (cf. page 56).
- [105] James C SPALL. « Implementation of the simultaneous perturbation algorithm for stochastic optimization ». In : *IEEE Transactions on aerospace and electronic systems* 34.3 (1998), pages 817-823 (cf. pages 56, 59).
- [106] James C SPALL. *Introduction to stochastic search and optimization : estimation, simulation, and control*. Tome 65. John Wiley & Sons, 2005 (cf. page 56).
- [107] Marco R STEENBERGEN. « Maximum likelihood programming in R ». In : *University of North Carolina, Chapel Hill* (2006) (cf. page 46).
- [108] MA STYBLINSKI et T-S TANG. « Experiments in nonconvex optimization : stochastic approximation with function smoothing and simulated annealing ». In : *Neural Networks* 3.4 (1990), pages 467-483 (cf. pages 28, 30, 83).
- [109] Genaro SUCARRAT et Alvaro ESCRIBANO. « The power log-GARCH model ». In : (2010) (cf. page 75).
- [110] Genaro SUCARRAT et Alvaro ESCRIBANO. « Unbiased QML Estimation of Log-GARCH Models in the Presence of Zero Returns ». In : (2013) (cf. pages 3, 77, 88, 91, 94).
- [111] Genaro SUCARRAT et Steffen GRØNNEBERG. « Models of financial return with time-varying zero probability ». In : (2016) (cf. page 76).
- [112] Genaro SUCARRAT, Steffen GRØNNEBERG et Alvaro ESCRIBANO. « Estimation and inference in univariate and multivariate log-GARCH-X models when the conditional density is unknown ». In : *Computational statistics & data analysis* 100 (2016), pages 582-594 (cf. page 75).
- [113] Henghsiu TSAI et Kung-Sik CHAN. « A note on inequality constraints in the GARCH model ». In : *Econometric Theory* 24.3 (2008), pages 823-828 (cf. pages 12, 32, 33, 65).
- [114] Andrew A WEISS. « Asymptotic theory for ARCH models : estimation and testing ». In : *Econometric theory* (1986), pages 107-131 (cf. page 14).
- [115] Wu WEN et Hugh F DURRANT-WHYTE. « Model-based multi-sensor data fusion ». In : *Proceedings 1992 IEEE International Conference on Robotics and Automation*. IEEE Computer Society. 1992, pages 1720-1721 (cf. page 23).

- [116] Diethelm WUERTZ, Tobias SETZ, Yohan CHALABI, Chris BOUDT, Pierre CHAUSSE, Michal MIKLOVAC, Maintainer Tobias SETZ et Suggests RUNIT. *Package 'fGarch'*. Rapport technique. Technical report, working paper/manual, 09.11. 2009. URL [http://cran. r . . .](http://cran.r-project.org/web/packages/fGarch/index.html), 2013 (cf. page 32).
- [117] M ZENGA. *Kurtosis*. *Encyclopedia of Statistical Sciences*, (S. Kotz, CB Read, N. Balakrishnan and B. Vidakovic, eds.) 2006 (cf. pages 84, 85).
- [118] Xiaorui ZHU et Li XIE. « Adaptive quasi-maximum likelihood estimation of GARCH models with Student's likelihood ». In : *Communications in Statistics-Theory and Methods* 45.20 (2016), pages 6102-6111 (cf. pages 14, 59).