

N d'ordre : 04/2016-D/MT

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE DES SCIENCES ET TECHNOLOGIES HOUARI BOUMEDIENE
FACULTE DE MATHEMATIQUES



THESE

Présentée pour l'obtention du grade de DOCTORAT EN SCIENCES
EN : MATHEMATIQUES
SPECIALITE : Probabilités Statistiques

Par : Yakoub BOULAROUK

Sujet

**Estimation des paramètres de processus causaux affines par la
méthode du maximum de vraisemblance**

Soutenue publiquement le 26/05/2016, devant le jury composé de :

Hacène BELBACHIR
Khedidja DJEDDOUR
Zaher MOHDEB
Nasr-eddine HAMRI
Khaled KHALDI
Tarek MEDKOUR
Jean Marc BARDET

Professeur à l'U.S.T.H.B.
Professeur à l'U.S.T.H.B.
Professeur à l'U.M.C.
Professeur au C.U.MILA.
Professeur à l'U.M.B.B.
Maitrte de Conférence à l'U.S.T.H.B.
Professeur à PARIS.1.

Président
Directrice de Thèse
Examineur
Examineur
Examineur
Examineur
Invité

Remerciements

Tout d'abord, je voudrais bien exprimer mes remerciements les plus respectueux au Professeur Hacène BELBACHIR, pour l'honneur qu'il m'a fait en acceptant de présider le jury de cette thèse.

Je remercie vivement monsieur Zaher MOHDEB professeur à l'université de Constantine, monsieur Nasr-eddine HAMRI professeur à l'université de Mila, monsieur Khaled KHALDI professeur à l'université de Boumerdès et monsieur Tarek Medkour Maître de conférence à l'université des sciences et technologies houari boumediène pour l'intérêt qu'ils ont bien voulu porter à ce travail et d'avoir accepté de le juger.

Je tiens à exprimer mes plus vifs remerciements à ma directrice de thèse professeur Khedidja DJABALLAH pour sa gentillesse, Sa sagesse, ses relectures et surtout ses conseils scientifiques qui m'ont permis de réaliser mon travail de thèse dans les meilleures conditions possibles.

Le remerciement suivant revient bien justement au professeur Jean Marc BARDET qui m'a accueilli pendant dix-huit mois dans son laboratoire SAMM à l'université Sorbonne Panthéon et qui a toujours été attentif et disponible malgré ses nombreuses charges.

Enfin, les mots les plus simples étant les plus forts, j'adresse toute mon affection à ma famille, mes frères Ishak, Habib et khaled, ma seour Amal, mon neveu Ayoub et en particulier à mes parents qui m'ont fait comprendre que la vie n'est pas faite que de problèmes qu'on pourrait résoudre grâce à des formules mathématiques et des algorithmes. Malgré mon éloignement temporaire, leur sagesse, leur confiance, leur tendresse, leur amour m'ont porté et m'ont guidé tous les jours. Merci pour avoir fait de moi ce que je suis aujourd'hui. Est-ce un bon endroit pour dire ce genre de choses ? Je n'en connais en tous cas pas de mauvais. Je vous aime.

Résumé

Dans cette thèse,

Nous commençons par prouver la consistance et la normalité asymptotique de l'estimateur quasi-maximum de vraisemblance Laplace (EQMV-Laplace) pour une classe générale de séries chronologiques causale contenant les modèles ARMA, ARCH, GARCH, ARMA-GARCH, APARCH, ... etc. Nous exposons notamment les avantages de cet estimateur par rapport au classique estimateur quasi-maximum de vraisemblance Gaussien (EQMV-Gaussien) à travers des simulations numériques qui confirment la précision de cet estimateur.

Par la suite nous présenterons une nouvelle approche pour l'optimisation d'estimation des paramètres de modèle Autoregressif moyenne Mobile (ARMA) avec erreurs Gaussiennes. Nous montrerons que la fonction de log-vraisemblance du processus est presque sûrement égale à un polynôme d'ordre deux. Par suite, utilisant les méthodes des moindres carrés, on approche notre fonction par un polynôme d'ordre deux qui sera utilisé pour le calcul d'une estimation du maximum de vraisemblance.

Mots clés :

Fonction de vraisemblance, Estimateur Quasi-Maximum de vraisemblance Laplace, Consistance forte, Normalité asymptotique, Processus ARMA-ARCH, Algorithme d'innovation.

Abstract

In this thesis,

We start by proving the consistency and asymptotic normality of the Laplacian Quasi-Maximum Likelihood Estimator (QMLE) for a general class of causal time series including ARMA, $AR(\infty)$, GARCH, $ARCH(\infty)$, ARMA-GARCH, APARCH, ARMA-APARCH,..., processes. We notably exhibit the advantages (moment order and robustness) of this estimator compared to the classical Gaussian QMLE. Numerical simulations confirms the accuracy of this estimator.

Therefore, we present a new approach to the optimization of autoregressive moving average parameters estimation, we prove that the log likelihood function of the process is almost surely equal to a polynomial of order two. Thereafter, using the methods of least squares, our function will be interpolated by a polynomial of order two which will be used to calculate an approximation of the maximum.

Keywords :

Likelihood function, Laplacian Quasi-Maximum Likelihood Estimator, Strong consistency, Asymptotic normality, ARMA-ARCH processes ,ARMA Process, Innovation algorithm.

Table des matières

1	Introduction	1
1.1	Contexte	1
1.2	Problématique	5
1.3	Contribution de la thèse	6
1.3.1	Etude de l'EQMV-Laplace des processus caussaux affines	6
1.3.2	Approximation numérique pour l'EQMV d'un ARMA .	7
2	L'EQMV-Laplace des processus caussaux affines	9
2.1	Introduction	9
2.2	Définitions et hypothèses	11
2.2.1	Définition de l'estimateur	11
2.2.2	Existence et stationarité	13
2.2.3	Hypothèses nécessaires pour la convergence d'EQMV-Laplace	16
2.3	Comportement asymptotique d'EQMV	16
2.3.1	Propriétés asymptotiques de la quasi vraisemblance . .	16
2.3.2	Consistance forte	17
2.3.3	La normalité asymptotique	21
2.4	Exemples de modèles particuliers	29
2.4.1	Processus ARMA	30
2.4.2	Processus GARCH	31
2.4.3	Processus APARCH	32
2.4.4	Processus ARMA-GARCH	35
2.4.5	Processus ARMA-ARCH(∞).	37
2.4.6	Processus ARMA-APARCH	38

3	Approximation de l'estimateur des paramètres de l'ARMA	42
3.1	Introduction	42
3.2	Détermination de la fonction de vraisemblance	43
3.2.1	Fonction de vraisemblance de l'ARMA(p, q)	43
3.2.2	L'algorithme d'innovation pour le modèle ARMA	45
3.3	Propriétés de la vraisemblance d'un ARMA	46
3.4	Procédure de calcul	48
3.5	Valeurs préliminaires	49
3.5.1	Méthode de Yule-Walker	49
3.5.2	Algorithme de Burg	49
3.5.3	Algorithme de Hannan-Rissanen	50
3.6	Intervalles de confiance pour les Coefficients	51
3.7	Exemples d'application	51
3.7.1	Le processus moyenne mobile MA(1)	51
3.7.2	Le processus autoregerssif AR(1)	55
	Articles.	60
	Bibliographie.	87
	Conclusion générale.	93
	Perspectives.	94

Table des figures

1.1	Représentation graphique de la distribution Gauss et la distribution Laplace.	5
3.1	Série chronologique x_t	53
3.2	Représentation graphique de la fonction log-vraisemblance. .	55
3.3	Représentation graphique de l'approximation du fonction log-vraisemblance.	56
3.4	Série chronologique y_t	56
3.5	Représentation graphique de la fonction log-vraisemblance. .	59
3.6	Représentation graphique de l'approximation de fonction log-vraisemblance.	59

Liste des tableaux

2.1	Racine d'erreur moyenne quadratique $\hat{\theta}_n^{LQL}$ et $\hat{\theta}_n^{GQL}$ pour les processus ARMA(1, 1).	31
2.2	Racine d'erreur moyenne quadratique $\hat{\theta}_n^{LQL}$ et $\hat{\theta}_n^{GQL}$ pour les processus GARCH(1, 1).	32
2.3	Racine d'erreur moyenne quadratique $\hat{\theta}_n^{LQL}$ et $\hat{\theta}_n^{GQL}$ pour le processus APARCH(1, 1).	34
2.4	Racine d'erreur moyenne quadratique $\hat{\theta}_n^{LQL}$ et $\hat{\theta}_n^{GQL}$ pour les processus ARMA(1, 1)-GARCH(1, 1).	36
2.5	Racine d'erreur moyenne quadratique $\hat{\theta}_n^{LQL}$ et $\hat{\theta}_n^{GQL}$ pour les processus ARMA(1, 1)-APARCH(1, 1).	41
3.1	Tableau représentant la série chronologique (x_t)	52
3.2	Tableau représentant les valeurs de la log-vraisemblance du processus MA aux points de l'échantillon.	54
3.3	Tableau représentant la série chronologique (y_t)	57
3.4	Tableau représentant les valeurs de la log-vraisemblance du processus AR aux points de l'échantillon.	58

Chapitre 1

Introduction

1.1 Contexte

La théorie des séries chronologiques est d'un usage constant quand on traite des données échantillonnées dans le temps. Les séries chronologiques (ou temporelles) sont appliquées de nos jours dans des domaines aussi variés que l'économétrie, la médecine ou la démographie, pour n'en citer qu'une petite partie. La modélisation représentant une étape cruciale dans l'étude des séries chronologiques. Elle a connu une grande évolution durant les cinquante dernières années et plusieurs modèles de représentation ont été proposés.

En 1970, Box et Jenkins ont publié un livre "*Time series : forecasting and control*" ([8]), dans lequel ils ont présenté et étudié les modèles Autorégressifs moyenne mobile (ARMA) et les modèles Auto-régressifs moyenne mobile intégré (ARIMA) de bruits Gaussiens et de variance constante. Les deux processus proposés dépendent de leurs passés et celui des bruits.

En 1982, Engel ([22]) a proposé le modèle Autorégressif conditionnellement hétéroscédastique (ARCH) avec bruits Gaussiens, qui a permis à la variance conditionnelle de changer en fonction des erreurs passées au fil du temps, tout en laissant la variance inconditionnelle constante. Ce modèle s'est déjà révélé utile dans la modélisation de plusieurs phénomènes. Dans Engle (1982) ([22]), Engle (1983) ([23]) et Engle et Kraft (1983) ([24]), des modèles pour le taux d'inflation sont construits en admettant l'hypothèse que l'inflation tend à changer au fil du temps. En 1984, Weiss ([72]) a considéré les modèles ARMA avec erreurs ARCH de bruits Gaussiens. Il a utilisé ces modèles pour la modélisation de seize séries chronologiques microéconomiques américaines.

En 1986, Bollerslev ([10]) a proposé une généralisation des ces modèles ARCH aux modèles Autorégressifs conditionnellement hétéroscédastique généralisé (GARCH), dont la variance conditionnelle au présent dépend de son passé et du passé de processus. Comme il a étudié les conditions de stationnarité et la structure des autocorrélations pour cette classe de modèles.

En 1993, Ding et al. ([19]) ont étudié la structure de l'autocorrélation de $|\varrho_t|^d$, où ϱ_t représente le rendement quotidien de cinq cent titres boursiers journaliers et d est un nombre positif. Ils ont trouvé que pour de longs délais, $|\varrho_t|$ a des autocorrélations positives significatives. Ce résultat les a motivés pour proposer une nouvelle généralisation pour la classe des modèles ARCH, dite Asymetric Power ARCH (APARCH). En outre, ils ont montré que ce modèle englobe sept autres modèles connus dans la littérature.

Après la mise en place des modèles GARCH et APARCH, Ding et al. (1993) ([19]) ont introduit les modèles ARMA avec erreurs GARCH à base de bruits Gaussiens et les modèles ARMA avec erreurs APARCH à base de bruits Gaussiens comme généralisation des modèles ARMA avec erreurs ARCH.

En 2007, Doukhan et Wintenberger ([18]) ont écrit sous forme générale la classe des processus causaux affines ($X_t = M_\theta^t \zeta_t + f_\theta^t$), qui contient les processus ci-dessus et d'autres.

Plusieurs méthodes d'estimation paramétriques ont été utilisées, parmi lesquelles, on trouve :

- La méthode des moindres carrés (MC) : cette méthode consiste à trouver les valeurs des paramètres de modèle qui minimisent la somme de carrés des erreurs $\sum_{t=1}^n \zeta_t^2$, cette méthode a été utilisé par Box et Jenkins ([8]) pour les modèles *ARMA*, Engel ([22]) pour les modèles *ARCH*.
- La méthode d'écart absolu minimal (EAM) : Malgré les avantages de cette méthode basé sur la minimisation de la valeur absolue des erreurs $\sum_{t=1}^n |\zeta_t|$, elle n'était pas tellement utilisé dans la littérature, on retrouve quand même Davis *et. al* ([15]) qui l'ont utilisée pour les modèles Autorégressifs avec variance infini *AR*(∞), Davis et Duns-muir ([14]) pour les régressions avec erreurs *ARMA*.
- La méthode du maximum de vraisemblance (MV) : cette méthode est la plus utilisée dans l'estimation des paramètres de modèles de séries chronologiques, Box et Jenkins ([8]) l'ont utilisée pour l'estimation des

modèles *ARMA*, Engel ([22]) l'a utilisée pour les modèles *ARCH*, Bolerselv ([10]) pour les modèles *GARCH*, Bardet et Wintenberger ([2]) pour la classe des processus causaux affines avec bruits Gaussiens, cette classe contient les modèles *ARMA* avec erreurs *GARCH* et beaucoup d'autres.

La méthode du maximum de vraisemblance exige la connaissance de la loi de probabilité du bruit. Pour cette raison, les travaux cités ci-dessus et d'autres, supposent que la loi de probabilité est Gaussienne. Elle fournit une bonne représentation pour plusieurs séries chronologiques. C'est la loi la plus utilisée en statistique, est critiquée dans certaines de ses utilisations pour sa queue étroite. En finance par exemple, cette queue étroite correspond à une sous-estimation des variations extrêmes, et de nouveaux modèles, dits *leptokurtiques* cherchent à éliminer ce problème.

En 1988, Li et McLeod ([44]) ont considéré le problème de la modélisation des séries chronologiques *ARMA* entraînée par les bruits non Gaussiens, comme ils ont prouvé sous des conditions générales la normalité asymptotique de l'estimateur du maximum de vraisemblance. Depuis, plusieurs travaux ont étudié les séries chronologiques avec bruits non Gaussien. En 1997, Shephard et Michael ([67]) ont changé de direction en fournissant des méthodes pour l'estimation des modèles de séries chronologiques avec bruits non Gaussiens, Fancq et *al.* (2011) ont écrit un article ([29]) dans lequel ils ont proposé une estimation en deux étapes pour les paramètres des modèles *GARCH* avec bruits non Gaussiens. Barnard et *al.* (2014)([4]) ont donné des résultats inférentiels pour les modèles *ARMA* avec erreurs de distributions types puissance exponentielle multivariés.

En général, la fonction de vraisemblance n'est pas calculable puisque $X_0, X_{-1}, \dots$ sont inconnus. C'est pourquoi on utilise la quasi-vraisemblance qu'on obtient en remplaçant le passé du processus par des zéros. Les propriétés asymptotiques de $\log$ quasi-vraisemblance $\widehat{L}_n$ (définie dans (2.2.3)) ont été étudié dans plusieurs articles où ils ont prouvé que la quasi-vraisemblance $\widehat{L}_n$ converge vers la fonction $\log$ vraisemblance L_n .

La stationnarité est une condition nécessaire pour la consistance et la normalité asymptotique de l'estimateur de quasi-vraisemblance. Afin de vérifier cette condition,

- Brockwell et Lindnery (2010) ([13]) supposent que les solutions du polynôme retard de la partie autoregressive soient à l'extérieur de disque

d'unité et que $E \log^+ |\zeta_1| < \infty$ où ζ_t est le bruit, comme conditions nécessaire et suffisante pour la stationnarité des processus ARMA avec bruits indépendants et identiquement distribués.

- Bollerslev (1986) ([10]) ont donné une condition nécessaire et suffisante qui assure l'existence d'une solution stationnaire pour le modèle GARCH.
- Suivant la même démarche de Bollerslev pour les processus GARCH, Ling et McAleer (2002) ([42]) ont réussi à énoncer une condition nécessaire et suffisante pour la stationnarité des processus APARCH.
- En 2012, Bardet et Weitenberg ([2]) ont défini un ensemble de stationnarité pour la classe des processus causaux affines, une classe qui contient les processus ARMA, ARCH, GARCH, APARCH, ARMA-GARCH, ARMA-APARCH, ...etc.

Tout comme la stationnarité, l'existence des moments du processus est nécessaire pour la consistance et la normalité asymptotique.

La fonction de vraisemblance étant non linéaire, dans le domaine des séries chronologiques, une branche nécessite l'emploi des algorithmes ou des méthodes d'optimisation numériques pour le calcul de l'unique maximum.

Nelder et Mead (1965) ([54]) ont introduit une méthode utilisant uniquement les valeurs de la fonction de vraisemblance. Malgré sa méthode est relative lenteur, cette méthode est robuste et aboutit à des résultats pour les fonctions non différentiables. À leur tour Fletcher et Reeves (1964)([27]) ont introduit la méthode du gradient conjugué qui ne stocke pas de matrice. En 1970 Broyden ([11]), Fletcher ([26]), Goldfarb ([35]) et Shanno ([64]) ont publié simultanément une méthode intitulé quasi-Newton (notée BFGS), qui utilise les valeurs et les gradients de fonction pour construire une image de la surface à optimiser. La méthode de BFGS est généralement plus stable que la méthode du gradient conjugué.

1.2 Problématique

Dans ce travail de thèse,

Dans une première partie, nous nous sommes intéressés à la classe des processus causaux affines dont la loi du bruit est la loi de Laplace et n'ont pas celle de Gauss utilisé habituellement.

Les processus entraînés par des bruits non Gaussiens sont fréquents dans les situations réelles. Par exemple, en économie, Nelson et Granger (1979) ([55]) ont modélisé un ensemble de 21 séries chronologiques par des modèles ARIMA, parmi ceux-ci, l'hypothèse que la loi du bruit soit Gaussienne a été rejeté par 15 séries. Tandis que la loi normale est exprimée en termes de la différence au carré $(\zeta - \mu)^2$, la loi de Laplace fait intervenir la différence absolue $|\zeta - \mu|$. La loi de Laplace présente alors des queues plus épaisses que la loi normale. Donc, dans le cas du bruits à queues lourdes, un contraste du type L^1 (contraste de Laplace) nous donne un meilleur modèle pour représenter le processus stochastique qu'un contraste du type L^2 (contraste Gauss).

Dans une seconde partie, nous nous sommes intéressés aux méthodes

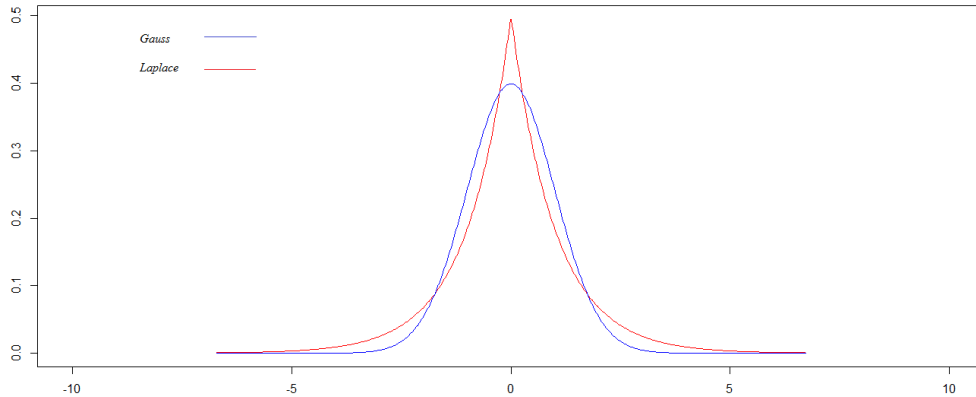


FIGURE 1.1 – Représentation graphique de la distribution Gauss et la distribution Laplace.

numériques utilisées pour calculer le maximum de la fonction du quasi-vraisemblance dans le cas de modèle ARMA avec bruits Gaussien.

Le calcul du maximum de vraisemblance en minimisant le log-vraisemblance

d'un modèle ARMA $l_n(\Phi, \Theta) = \frac{1}{n} \sum_{i=1}^n \ln(r_{i-1}) + \ln(n^{-1} \sum_{i=1}^n \frac{\varepsilon_i^2(\Phi, \Theta)}{r_{i-1}})$ utilisant les méthodes et les algorithmes numériques entraînent beaucoup de calculs, qui s'avèrent difficile et ceci est due à la complexité du log-déterminant de la matrice d'autocovariance.

Comme le terme $\frac{1}{n} \sum_{i=1}^n \ln(r_{i-1})$ est dominé par le terme $\ln(n^{-1} \sum_{i=1}^n \frac{\varepsilon_i^2(\Phi, \Theta)}{r_{i-1}})$ pour n assez grand, Box et Jenkins (1976) ([9]) ont suggéré l'abstraction du facteur log-déterminant pour l'estimation du maximum de vraisemblance et de formuler l'expression restante comme un problème d'estimation des moindres carrés. Cependant, il a souvent été prouvé, par exemple dans un rapport non publié KM Kang, par Osborn (1976)([59]), et dans un document de la conférence inédite de GEP Box, SC Hillmer et GC Tiao, que l'abstraction du déterminant peut conduire à des estimateurs inférieurs. McLeod (1977) ([49]) a proposé une méthode dans laquelle il a remplacé le déterminant par sa limite asymptotique, comme il a mené des expériences de Monte Carlo montrant l'amélioration de l'estimation pour les cas qu'il a examiné. Son approche, cependant, ne permet pas d'éviter le problème du calcul de $\sum_{i=1}^n \frac{\varepsilon_i^2(\Phi, \Theta)}{r_{i-1}}$.

1.3 Contribution de la thèse

1.3.1 Etude de l'EQMV-Laplace des processus causaux affines

Dans le deuxième chapitre, nous avons étudié les propriétés asymptotiques de l'estimateur de quasi-vraisemblance pour la classe des processus causaux affines avec bruits Laplace au lieu des bruits Gaussiens étudié dans la littérature.

Tout d'abord, nous avons défini la fonction de vraisemblance et l'estimateur de quasi-vraisemblance pour les processus causaux affines avec bruits Laplace, puis nous avons défini un espace de stationnarité pour notre classe. Nous avons prouvé la consistance d'estimateur de quasi vraisemblance pour la classe des processus causaux affines, sous condition d'existence des moments de premier ordre qui est une condition plus faible que celle exigée pour le cas des processus avec bruits Gaussien (voir Bardet et Wintenberger [2]) (2009)). Comme nous avons prouvé la normalité asymptotique de notre estimateur, sous condition d'existence des moments d'ordre deux, un

résultat qui se trouve une deuxième fois plus performant que celui des processus avec bruits Gaussiens. Dans le but d'exposer la performance de nos résultats, nous les avons appliqué à des cas particuliers des processus causaux affines, à savoir les processus ARMA, GARCH, APRCH, ARMA-GARCH, ARMA-ARCH(∞), ARMA-APARCH. Par la suite, dans le but de confirmer les résultats asymptotiques obtenus et de comparer l'EQMV-Laplace avec l'EQMV-Gauss, nous avons effectué des expériences types monte-Carlo sur plusieurs modèles de séries chronologiques de différentes tailles et de lois de probabilité du bruit.

1.3.2 Approximation numérique pour l'EQMV d'un ARMA

Dans le troisième chapitre, Nous avons proposé une nouvelle méthode de calcul du maximum de vraisemblance. Elle consiste à calculer le log déterminant de la matrice d'autocovariance une seule fois au lieu de le faire à chaque itération.

Nous avons commencé par exprimer la fonction de vraisemblance à travers une représentation matricielle des paramètres et des innovations, ensuite nous avons donné une version simplifiée de l'algorithme d'innovations dans le cas des modèles ARMA. À travers le Lemme 4, nous avons montré que la fonction de vraisemblance d'un ARMA est presque sûrement égale à un polynôme d'ordre deux.

Comme la fonction de vraisemblance est à la fois convexe et presque sûrement égale à un polynôme d'ordre deux, nous l'avons approché par un polynôme de cet ordre. Cela, nous a permis de proposer une procédure qui permet de calculer une approximation du maximum de vraisemblance dans le cas de modèle ARMA avec bruits Gaussiens, une procédure basée sur l'approximation locale de la fonction de vraisemblance.

Comme nous avons proposé une approximation locale, nous avons besoin de locaux d'approximation, des intervalles (voisinages) qui contiennent la vraie valeur du paramètre. Dans le but d'obtenir ces voisinages, tout d'abord utilisant la méthode de Yule-Walker, l'algorithme de Burg ou l'algorithme de Hannan-Rissanen, nous avons calculé des valeurs préliminaires de l'estimateur. Par suite, comme les valeurs préliminaires obtenu par ces méthodes convergent en loi vers une loi normale (voir [12]), nous avons pu calculer des intervalles de confiance que nous avons utilisé comme des locaux d'approximation.

mations.

Finalement, nous avons effectué des applications numériques sur des séries chronologiques simulées par le modèle autoregressif $AR(1)$ et le modèle moyenne mobile $MA(1)$, les résultats obtenus nous ont confirmé l'efficacité de notre procédure.

Chapitre 2

L'EQMV-Laplace des processus caussaux affines

2.1 Introduction

Ce chapitre est consacré à la preuve de la consistance et la normalité asymptotique d'un estimateur paramétrique pour une classe générale de séries chronologiques. Cette classe a déjà été définie et étudiée dans Doukhan et Wintenberger (2007) [18], Bardet et Wintenberger (2009) [2] et Bardet et al. (2012) [3]. On va examiner un échantillon observé $(X_1, \dots, X_n)$ où le processus $(X_t)_{t \in \mathbb{Z}}$ est une solution de l'équation :

$$X_t = M_{\theta_0}(X_{t-1}, X_{t-2}, \dots) \zeta_t + f_{\theta_0}(X_{t-1}, X_{t-2}, \dots), \quad t \in \mathbb{Z}, \quad (2.1.1)$$

avec :

- $\theta_0 \in \Theta \subset \mathbb{R}^d$, $d \in \mathbb{N}^*$, le vecteur des paramètres inconnus, également appelé le vrai paramètre ;
- $(\zeta_t)_{t \in \mathbb{Z}}$ une suite de variables aléatoires indépendantes centrées identiquement distribuées (VAIID) avec distribution de probabilité symétrique, *c'est-à-dire* $\zeta_0 \stackrel{\mathcal{L}}{=} -\zeta_0$, satisfaisant $\mathbb{E}[|\zeta_0|^r] < \infty$ avec $r \geq 1$ et $\mathbb{E}[|\zeta_0|] = 1$. Si $r \geq 2$, notons $\sigma_\zeta^2 = \text{Var}(\zeta_0)$;
- $(\theta, (x_n)_{n \in \mathbb{N}}) \rightarrow M_\theta((x_n)_{n \in \mathbb{N}}) \in (0, \infty)$ et $(\theta, (x_n)_{n \in \mathbb{N}}) \rightarrow f_\theta((x_n)_{n \in \mathbb{N}}) \in \mathbb{R}$ deux applications connues.

Si $M_{\theta_0}(X_{t-1}, X_{t-2}, \dots) = 1$ et $f_{\theta_0}(X_{t-1}, X_{t-2}, \dots) = \alpha_0 X_{t-1}$ avec $|\alpha_0| < 1$ alors (X_t) est un processus AR(1) causal. Dans Doukhan et Wintenberger (2007) [18] et Bardet et Wintenberger (2009) [2], il a été prouvé que les séries les plus célèbres utilisées en économétrie, comme les processus ARMA, AR (∞), GARCH, ARCH (∞), TAR, ARMA-GARCH peuvent s'écrire comme une solution causale et stationnaire de l'équation (2.1.1).

Dans Bardet et Wintenberger (2009), il a également été établi que, sous plusieurs conditions sur M_θ , f_θ et si $E[|\zeta_0|^r]$ avec $r \geq 2$, l'estimateur habituel de la vraisemblance quasi-maximale Gaussienne (EQMV) de θ est fortement consistant et quand $r \geq 4$, il est asymptotiquement normal. Cet estimateur a été défini par Weiss (1986)[73] pour les processus ARCH, et l'étude asymptotique de cet estimateur a été obtenue par Lumsdaine (1996) [45] pour le processus GARCH (1, 1), Berkes *et al.* (2003) [5] pour le processus GARCH (p, q), Francq et Zakoian (2004) [28] pour les processus ARMA-GARCH, Mikosch et Straumann (2006) [68] pour les modèles hétéroscédastiques généraux, et Robinson et Zaffaroni (2006)[63] pour les processus ARCH (∞). Les résultats de Bardet et Wintenberger (2009) [2] adressés aux processus satisfaisant presque partout l'équation (2.1.1) ainsi que sa généralisation multivariée, fournissent un cadre général et unifié pour les propriétés asymptotiques de l'EQMV Gaussien.

La définition de l'EQMV Gaussien est explicitement obtenue sous l'hypothèse que (ζ_t) est une suite Gaussienne. Mais, même si la Quasi-Vraisemblance Gaussienne pourrait être appliquée lorsque la distribution de probabilité de (ζ_t) est non Gaussienne, elle garde certains inconvénients de cette hypothèse initiale. En effet, le calcul de ces estimateurs nécessite la minimisation de contraste des moindres carrés (typiquement $\sum_{t=1}^n M_\theta^{-2}(X_t - f_\theta)^2$), ce qui induit que $r = 2$ est nécessaire pour la consistance et $r = 4$ pour la normalité asymptotique (et donc des intervalles de confiance ou des tests). Pour de nombreuses données réelles une telle exigence est trop forte (par exemple, la coefficient d'aplatissement de données économiques est souvent considéré comme infini). En outre, tel estimateur n'est pas robuste aux valeurs aberrantes éventuelles. D'où, la distribution référence (la Gaussienne utilisée habituellement) de probabilité de (ζ_t) pourrait être Laplacienne et cela permet d'éviter ces deux inconvénients. Grosso-modo ce choix implique la minimisation d'un contraste d'écart absolu (typiquement $\sum_{t=1}^n M_\theta^{-1}|X_t - f_\theta|$) à la place du contraste des moindres carrés. Et donc, $r = 1$ sera suffisant pour as-

sur la forte consistance de cet EQMV-Laplace, tandis que seulement $r = 2$ est requis pour la normalité asymptotique (voir ci-dessous).

Ce choix de la distribution de probabilité n'est pas nouveau puisque cela conduit à l'estimation d'écarts Absolu minimal (LAD). Par conséquent, pour les processus ARMA, Davis et Dunsmuir (1997) [14] ont prouvé la consistance et la normalité asymptotique de l'estimateur LAD. Pour le processus ARCH ou le processus GARCH, les mêmes résultats concernant l'estimateur LAD étaient déjà établis par Peng et Yao (2003) [58], tandis que Berkes et Horvath (2004) [6] ont prouvé la consistance et la normalité asymptotique de contraste d'EQMV-Laplace. Newey et Steigerwald (1997) [53] considèrent également l'estimateur pour d'autres modèles d'hétéroscédasticité conditionnelle. Récemment, Francq *et al.* (2011) [29] ont proposé une estimation QMV non Gaussienne à deux étapes pour les modèles ARCH et Francq et Zakoian (2013) [30] ont proposé une procédure alternative en une étape, basée sur un estimateur QMV non-Gaussien, les propriétés asymptotiques de ces deux approches ont été étudiées.

La section suivante 2.2 sera consacrée aux définitions et hypothèses. Dans la section 2.3 nous énonçons les principaux résultats pour la classe des processus causaux affines, par suite dans la section 2.4 nous projetons les résultats théoriques sur plusieurs modèles stochastiques comme nous exposons les résultats des expériences de Monte-Carlo.

2.2 Définitions et hypothèses

2.2.1 Définition de l'estimateur

Soit $(X_1, \dots, X_n)$ une trajectoire observée de X solution de (2.1.1) où $\theta \in \Theta \subset \mathbb{R}^d$ est inconnu. Pour estimer θ nous considérons la log-vraisemblance de $(X_1, \dots, X_n)$ conditionnellement à $(X_0, X_{-1}, \dots)$. Si g est la densité de probabilité (par rapport à la mesure de Lebesgue) de ζ_0 , alors, à partir de la définition causale affine de X , ce log-vraisemblance peut s'écrire sous la forme :

$$\log (L_\theta(X_1, \dots, X_n)) = \sum_{t=1}^n \log \left(g \left(\frac{X_t - f_\theta^t}{M_\theta^t} \right) \right)$$

où $M_\theta^t := M_\theta(X_{t-1}, X_{t-2}, \dots)$ et $f_\theta^t := f_\theta(X_{t-1}, X_{t-2}, \dots)$, avec l'hypothèse que $M_\theta^t > 0$. Cependant, comme $X_0, X_{-1}, \dots$ sont inconnus, en général les

M_θ^t et f_θ^t ne sont pas calculables. Pour cela, au lieu d'utiliser directement la log-vraisemblance, un quasi-log-vraisemblance est considéré et il est défini par :

$$\log (QL_\theta(X_1, \dots, X_n)) = \sum_{t=1}^n \log \left(g \left(\frac{X_t - \widehat{f}_\theta^t}{\widehat{M}_\theta^t} \right) \right),$$

avec $[\widehat{f}_\theta^t := f_\theta(X_{t-1}, \dots, X_1, u)]$ et $[\widehat{M}_\theta^t := M_\theta(X_{t-1}, \dots, X_1, u)]$, où $u = (u_n)_{n \in \mathbb{N}}$ est une suite finie. Le choix de $(U_n)_{n \in \mathbb{N}}$ n'influe pas le comportement asymptotique de L_n , et (U_n) pourrait typiquement être choisi comme une suite égale à zéro. Enfin, l'estimateur de vraisemblance quasimaximale (EQMV) est défini par :

$$\tilde{\theta}_n := \text{Argmax}_{\theta \in \Theta} \log (QL_\theta(X_1, \dots, X_n)).$$

Habituellement, la distribution Gaussienne est considérée, avec $g(x) = (2\pi)^{-1/2} e^{-x^2/2}$ pour $x \in \mathbb{R}$ et ceci fournit l'EQMV-Gaussien de θ .

Dans le présent travail, nous avons choisi la distribution de Laplace, c'est-à-dire :

$$g(x) = \frac{1}{2} e^{-|x|} \quad \text{pour } x \in \mathbb{R}, \quad (2.2.1)$$

et cela implique que $E|\zeta_0| = 1$.

Nous définissons respectivement la vraisemblance de Laplace et le quasi-vraisemblance de Laplace par :

$$L_n(\theta) = - \sum_{t=1}^n q_t(\theta), \quad q_t(\theta) = \log |M_\theta^t| + |M_\theta^t|^{-1} |X_t - f_\theta^t| \quad (2.2.2)$$

$$\widehat{L}_n(\theta) = - \sum_{t=1}^n \widehat{q}_t(\theta), \quad \widehat{q}_t(\theta) := \log |\widehat{M}_\theta^t| + |\widehat{M}_\theta^t|^{-1} |X_t - \widehat{f}_\theta^t|. \quad (2.2.3)$$

Un EQMV de Laplace $\widehat{\theta}_n$ est le maximisateur de $\widehat{L}_n$:

$$\widehat{\theta}_n := \arg \max_{\theta \in \Theta} \widehat{L}_n(\theta).$$

Nous limitons l'ensemble Θ de manière à ce qu'une solution (X_t) de (2.1.1) stationnaire d'ordre 1 ou 2 existe. Des conditions supplémentaires sont également nécessaires pour assurer la consistance et la normalité asymptotique de $\widehat{\theta}_n$.

2.2.2 Existence et stationarité

Comme dans Doukhan et Wintenberger (2007) [18] et Bardet et Wintenberger (2009)[2], plusieurs inégalités de type Lipschitz sur f_θ et M_θ sont nécessaires pour l'obtention de l'existence et la r -stationnarité (définie dans [18]) d'une solution causale et ergodique de (2.1.1). Pour cela, nous introduisons le symbole générique Ψ pour les fonctions f , M et pour $g_\theta : \mathbb{R}^{\mathbb{N}} \mapsto \mathbb{R}^m$ ou $\mathcal{M}_m(\mathbb{R})$, notons $\|g_\theta\|_\Theta = \sup_{\theta \in \Theta} \|g_\theta\|$ avec $m \in \mathbb{N}^*$ et $\|\cdot\|$ la norme euclidienne usuelle. Pour $k = 0, 1, 2$ et un certain sous espace Θ de $\mathbb{R}^d$, définissons la condition Lipschitzienne sur la fonction Ψ_θ :

Hypothèse $(\mathbf{A}_k(\Psi, \Theta))$: $\forall x \in \mathbb{R}^\infty, \theta \in \Theta \mapsto \Psi_\theta(x) \in \mathcal{C}^k(\Theta)$, $\partial_\theta^k \Psi_\theta$ satisfait $|\partial_\theta^k \Psi_\theta(0)|_\Theta < \infty$ et il existe une suite $(\alpha_j^{(k)}(\Psi, \Theta))_j$ de nombres non négatifs telle que $\forall x, y \in \mathbb{R}^{\mathbb{N}}$

$$\|\partial_\theta^k \Psi_\theta(x) - \partial_\theta^k \Psi_\theta(y)\|_\Theta \leq \sum_{j=1}^{\infty} \alpha_j^{(k)}(\Psi, \Theta) |x_j - y_j|, \text{ avec } \sum_{j=1}^{\infty} \alpha_j^{(k)}(\Psi, \Theta) < \infty.$$

Afin d'assurer l'existence d'une solution stationnaire d'ordre r ($r \geq 1$) pour (2.1.1), définissons l'ensemble :

$$\Theta(r) := \left\{ \theta \in \mathbb{R}^d, (\mathbf{A}_0(f, \{\theta\})) \text{ et } (\mathbf{A}_0(M, \{\theta\})) \text{ sont vérifiées,} \right. \\ \left. \sum_{j=1}^{\infty} \alpha_j^{(0)}(f, \{\theta\}) + (\mathbb{E}[|\zeta_0|^r])^{1/r} \sum_{j=1}^{\infty} \alpha_j^{(0)}(M, \{\theta\}) < 1 \right\}.$$

Puis, à partir Doukhan et Wintenberger (2007) [18], nous obtenons :

Proposition 1 : *Si $\theta_0 \in \Theta(r)$ pour un certain $r \geq 1$, alors il existe une unique solution causal de (2.1.1) (X_t indépendant de $(\zeta_i)_{i>t}$ pour $t \in \mathbb{Z}$), qui est stationnaire, ergodique et satisfait $E(|X_0|^r) < \infty$.*

Le lemme suivant assure que si un processus X satisfait la proposition 1, un processus ARMA prévisible causal avec X comme innovation satisfait également la proposition 1. Nous donnons d'abord la notion classique pour une suite de nombres réels $(u_n)_{n \in \mathbb{N}}$:

$(u_n)_{n \in \mathbb{N}}$ est une suite exponentiellement décroissante (SED)

$$\Longleftrightarrow$$

Il existe un $\rho \in [0, 1[$ tel que $u_n = \mathcal{O}(\rho^n)$ quand $n \rightarrow \infty$.

Lemme 1 Soit X une solution p.s. causal et stationnaire de (2.1.1) pour $\theta_0 \in \mathbb{R}^d$. Soit $\tilde{X}$ tel que $\tilde{X}_t = \Lambda_{\beta_0}(L) X_t$ pour $t \in \mathbb{Z}$ avec $\Lambda_{\beta_0}(L) = P_{\beta_0}^{-1}(L) Q_{\beta_0}(L)$ où $(P_{\beta_0}, Q_{\beta_0})$ sont les polynômes coprimés d'un processus ARMA(p, q) causal et inversible avec un vecteur de paramètres inconnus $\beta_0 \in \mathbb{R}^{p+q}$. Notons $\Lambda_{\beta_0}^{-1}(x) = Q_{\beta_0}^{-1}(x) P_{\beta_0}(x) = 1 + \sum_{j=1}^{\infty} \psi_j(\beta_0) x^j$. Alors $\tilde{X}$ est une solution p.s. causal et stationnaire de l'équation :

$$\tilde{X}_t = \tilde{M}_{\tilde{\theta}_0}((\tilde{X}_{t-i})_{i \geq 1}) \zeta_t + \tilde{f}_{\tilde{\theta}_0}((\tilde{X}_{t-i})_{i \geq 1}) \quad \text{pour } t \in \mathbb{Z},$$

où $\tilde{f}_{\tilde{\theta}_0}$ et $\tilde{M}_{\tilde{\theta}_0}$ sont données par (2.2.4) et $\tilde{\theta}_0 = (\theta_0, \beta_0)$. De plus, pour $i = 0, 1, 2$ et $(K = f \text{ ou } M)$ et $(\tilde{K} = \tilde{f} \text{ ou } \tilde{M})$,

- si $\alpha_j^{(i)}(K, \{\theta_0\}) = O(j^{-\beta})$ et $\beta > 1$, alors $\alpha_j^{(i)}(\tilde{K}, \{\tilde{\theta}_0\}) = O(j^{-\beta})$;
- si $\alpha_j^{(i)}(K, \{\theta_0\})$ est SED, alors $\alpha_j^{(i)}(\tilde{K}, \{\tilde{\theta}_0\})$ est SED.

Preuve de Lemme 1

Comme X est un processus stationnaire et le processus ARMA(p, q) est causal inversible alors $\tilde{X}$ est également un processus stationnaire (les coefficients de Λ_{β_0} sont SED). De plus, il est bien connu que $(\psi_j(\beta_0))_{j \in \mathbb{N}}$ sont SED. Alors nous avons :

$$\begin{aligned} \tilde{X}_t &= \Lambda_{\beta_0}(L) \left(M_{\theta_0}((X_{t-i})_{i \geq 1}) \zeta_t + f_{\theta_0}((X_{t-i})_{i \geq 1}) \right) \\ \tilde{X}_t + \sum_{j=1}^{\infty} \psi_j(\beta_0) \tilde{X}_{t-j} &= M_{\theta_0}((\Lambda_{\beta_0}^{-1}(L) \tilde{X}_{t-i})_{i \geq 1}) \zeta_t + f_{\theta_0}((\Lambda_{\beta_0}^{-1}(L) \tilde{X}_{t-i})_{i \geq 1}) \\ \tilde{X}_t &= \tilde{M}_{\tilde{\theta}_0}((\tilde{X}_{t-i})_{i \geq 1}) \zeta_t + \tilde{f}_{\tilde{\theta}_0}((\tilde{X}_{t-i})_{i \geq 1}) \end{aligned}$$

avec :

$$\begin{cases} \tilde{M}_{\tilde{\theta}_0}((x_{t-i})_{i \geq 1}) &= M_{\theta_0}((\Lambda_{\beta_0}^{-1}(L) x_{t-i})_{i \geq 1}) \\ \tilde{f}_{\tilde{\theta}_0}((x_{t-i})_{i \geq 1}) &= f_{\theta_0}((\Lambda_{\beta_0}^{-1}(L) x_{t-i})_{i \geq 1}) - \sum_{j=1}^{\infty} \psi_j(\beta_0) x_{t-j} \end{cases} \quad (2.2.4)$$

Enfin, pour $t \in \mathbb{Z}$,

$$\begin{aligned}
|\widetilde{f}_{\widetilde{\theta}_0}((x_{t-i})_{i \geq 1}) - \widetilde{f}_{\widetilde{\theta}_0}((y_{t-i})_{i \geq 1})| &\leq \sum_{j=1}^{\infty} \alpha_j^{(0)}(f, \{\theta_0\}) |(\Lambda_{\beta_0}^{-1}(L)x_{t-j-i})_{i \geq 1} - (\Lambda_{\beta_0}^{-1}(L)y_{t-j-i})_{i \geq 1}| \\
&+ \sum_{j=1}^{\infty} |\psi_j(\beta_0)| |x_{t-j} - y_{t-j}| \\
&\leq \sum_{j=1}^{\infty} \alpha_j^{(0)}(f, \{\theta_0\}) \left| \sum_{k=0}^{\infty} |\psi_k(\beta_0)| |x_{t-k-j} - y_{t-k-j}| \right| \\
&+ \sum_{j=1}^{\infty} |\psi_j(\beta_0)| |x_{t-j} - y_{t-j}| \\
&\leq \sum_{j=1}^{\infty} \left(|\psi_j(\beta_0)| + \sum_{k=1}^j \alpha_k^{(0)}(f, \{\theta_0\}) |\psi_{j-k}(\beta_0)| \right) |x_{t-j} - y_{t-j}| \\
&\implies \alpha_j^{(0)}(\widetilde{f}, \{\widetilde{\theta}_0\}) \leq |\psi_j(\beta_0)| + \sum_{k=1}^j \alpha_k^{(0)}(f, \{\theta_0\}) |\psi_{j-k}(\beta_0)|. \quad (2.2.5)
\end{aligned}$$

En outre, nous avons :

$$\begin{aligned}
|\widetilde{M}_{\widetilde{\theta}_0}((x_{t-i})) - \widetilde{M}_{\widetilde{\theta}_0}((y_{t-i}))| &\leq \sum_{j=1}^{\infty} \alpha_j^{(0)}(M, \{\theta_0\}) |(\Lambda_{\beta_0}^{-1}(L)x_{t-j-i}) - (\Lambda_{\beta_0}^{-1}(L)y_{t-j-i})| \\
&\implies \alpha_j^{(0)}(\widetilde{M}, \{\widetilde{\theta}_0\}) \leq \sum_{k=1}^j \alpha_k^{(0)}(M, \{\theta_0\}) |\psi_{j-k}|. \quad (2.2.6)
\end{aligned}$$

Les mêmes types de calculs peuvent également être effectués en considérant les dérivées première et seconde de $\widetilde{f}$ et $\widetilde{M}$ par rapport à $\widetilde{\theta}$. Notons que les dérivées première et seconde de Λ_{β}^{-1} par rapport à $\widetilde{\theta}$ sont également SED. Finalement,

- si quand $j \rightarrow \infty$, $\alpha_j^{(0)}(K, \{\theta_0\}) = O(j^{-\beta})$ avec $\beta > 1$ et $\psi_j = O(\rho^j)$ avec $0 \leq \rho < 1$, alors il existe un $C > 0$ tel que $\sum_{k=1}^j \alpha_k^{(0)}(K, \{\theta_0\}) |\psi_{j-k}| \leq C \sum_{k=1}^j k^{-\beta} \rho^{j-k} \sim -C(\log \rho)^{-1} j^{-\beta}$ et donc $\alpha_j^{(0)}(\widetilde{K}, \{\widetilde{\theta}_0\}) = O(j^{-\beta})$.
- si quand $j \rightarrow \infty$, $\alpha_j^{(0)}(K, \{\theta_0\}) = O(r^j)$ avec $0 \leq r < 1$ et $\psi_j = O(\rho^j)$ avec $0 \leq \rho < 1$, alors il existe un $C > 0$ tel que

$$\sum_{k=1}^j \alpha_k^{(0)}(K, \{\theta_0\}) |\psi_{j-k}| \leq C \sum_{k=1}^j r^{-k} \rho^{j-k} = O(j \max(r, \rho)^j) \text{ et } \text{donc } \alpha_j^{(O)}(\tilde{K}, \{\tilde{\theta}_0\}) \text{ sont SED.}$$

Également, on peut effectuer le même type de calcul pour $(\alpha_j^i)(\tilde{K}, \{\tilde{\theta}_0\})_j$ en appliquant la démarche de preuve ci dessus aux premières et deuxièmes dérivées de $\Lambda_{\beta_0}^{-1}$ par rapport à β , d'où on conclut que $\tilde{\theta}$ sont SED.

2.2.3 Hypothèses nécessaires pour la convergence d'EQMV-Laplace

L'EQMV Laplace pourrait converger et être asymptotiquement Gaussien mais cela nécessite quelques hypothèses supplémentaires sur Θ et les fonctions f_θ et M_θ :

- **Hypothèse H1** (Compacité) : Θ est un ensemble compact.
- **Hypothèse H2** (Borne inférieure de la variance conditionnelle) : il existe une constante $\underline{M} > 0$ tel que $\forall \theta \in \Theta$, alors $M_\theta(x) > \underline{M}$ pour tout $x \in \mathbb{R}^N$.
- **Hypothèse H3** (Identifiabilité) : Les fonctions M_θ et f_θ sont telles que : pour tout $\theta_1, \theta_2 \in \Theta$, alors $M_{\theta_1} = M_{\theta_2}$ et $f_{\theta_1} = f_{\theta_2}$ implique que $\theta_1 = \theta_2$.

2.3 Comportement asymptotique d'EQMV

2.3.1 Propriétés asymptotiques de la quasi vraisemblance

La quasi vraisemblance est une approximation de la vraisemblance obtenu en remplaçant $f_\theta^t = f_\theta(X_{t-1}, X_{t-2}, \dots)$ et $M_\theta^t = M_\theta(X_{t-1}, X_{t-2}, \dots)$ par $\hat{f}_\theta^t = f_\theta(X_{t-1}, X_{t-2}, \dots, \mu)$ et $\hat{M}_\theta^t = M_\theta(X_{t-1}, X_{t-2}, \dots, \mu)$ où $\mu = (u_n)_{n \geq 0}$ est égale à zéro. Le lemme suivant nous permet d'avoir une estimation de l'erreur d'approximation, il représente une étape cruciale dans la preuve de la consistance d'EQMV.

Lemme 2 *Supposons que $\theta_0 \in \Theta(r)$ pour $r \geq 1$ et que X est la solution causal et stationnaire de (2.1.1). Si $(A_0(\Psi, \Theta))$ est satisfaite, alors $\Psi_\theta^t \in L^r(\mathcal{C}(\Theta, R^m))$ et il existe une constante $C > 0$ indépendante de t tel que :*

$$\mathbb{E}[\|\widehat{\Psi}_\theta^t - \Psi_\theta^t\|_\Theta^r] \leq C \mathbb{E}[|X_0|^r] \left(\sum_{j \geq t} \alpha_j(\Psi) \right)^r \quad \text{pour tout } t \in N^*. \quad (2.3.1)$$

2.3.2 Consistance forte

Nous allons d'abord prouver la forte consistance d'une suite de l'EQMV Laplace pour une solution de (2.1.1).

Théoreme 1 *Supposons que les hypothèses H1, H2 et H3 sont satisfaites et que $\theta_0 \in \Theta(r) \cap \Theta$ avec $r \geq 1$. Soit X une solution stationnaire de (2.1.1). Si $(A_0(f, \Theta))$ et $(A_0(M, \Theta))$ demeurent avec :*

$$\alpha_j^0(f, \Theta) + \alpha_j^0(M, \Theta) = \mathcal{O}(j^{-\ell}) \quad \text{pour certain } \ell > \frac{2}{\min(r, 2)} \quad (2.3.2)$$

alors,

la suite d'EQMV-Laplace $(\widehat{\theta}_n)_n$ converge fortement, c'est-à-dire $\widehat{\theta}_n \xrightarrow[n \rightarrow \infty]{a.s.} \theta_0$.

Les conditions requises pour cette forte consistance d'EQMV-Laplace sont presque les mêmes que celles requises pour la forte consistance d'une suite d'EQMV-Gaussien sauf que $r \in [1, 2)$ est avérée possible dans le théorème 1 et non dans le cas d'EQMV-Gaussien (voir Bardet et Wintenberger, 2009) [2]. En outre, si $r = 2$, la condition (2.3.2) sur les coefficients Lipschitzienne est plus faible pour l'EQMV-Laplace que pour l'EQMV-Gaussien. Comme nous le verrons ci-dessous, de nombreuses séries chronologiques satisfont les hypothèses du Théorème 1 ; par exemple, un processus AR (∞) peut être définie pour satisfaire la forte consistance de l'EQMV-Laplace tandis que les conditions indiquées dans Bardet et Wintenberger (2009) [2] ne garantissent pas la forte consistance d'EQMV-Gaussien.

Preuve du Théorème 1

La preuve de théorème se fait en deux étapes et suit la même procédure que dans Bardet et Wintenberger (2009) [2]. Dans (i), une forte loi uniforme des grands nombres (sur θ) satisfaite par $\frac{1}{n} \widehat{L}_n(\theta)$ qui converge vers

$L(\theta) := -\mathbb{E}[Q_0(\theta)]$ est établi. Dans (ii), on prouve que $L(\theta)$ admet un maximum unique en θ_0 . Ces deux conditions impliquent la forte consistance de $\hat{\theta}_n$ (Jeanthreau, 1998) [37].

(i) De la même façon et pour les mêmes raisons de la preuve du théorème 1 de Bardet et Wintenberger (2009) [2], la loi uniforme forte des grands nombres satisfaite par la moyenne d'échantillon $(\hat{q}_t)_{t \in N^*}$ (défini dans (2.2.3)) est obtenu en prouvant que $\mathbb{E}[|q_t(\theta)|_{\Theta}] < \infty$.

Du lemme 2, pour tout $t \in \mathbb{Z}$,

$$\begin{aligned} |q_t(\theta)| &= |(M_\theta^t)^{-1}|X_t - f_\theta^t| + \log(M_\theta^t)| \\ &\leq \frac{|X_t - f^t(\theta)|}{\underline{M}} + |\log(\underline{M})| + M_\theta^t \\ \text{donc : } \sup_{\theta \in \Theta} |q_t(\theta)| &\leq \frac{1}{\underline{M}} (|X_t| + \|f^t(\theta)\|_{\Theta}) + |\log(\underline{M})| + \|M_\theta^t\|_{\Theta}. \end{aligned}$$

avec $r \geq 1$, on a :

- De la proposition 1 pour tout $t \in \mathbb{Z}$, $\mathbb{E}[|X_t|] < \infty$.
- Du Lemme 2 $\mathbb{E}[\|f_\theta^t\|_{\Theta}^r + \|M_\theta^t\|_{\Theta}^r] < \infty$, ce qui implique que $\mathbb{E}[\|f_\theta^t\|_{\Theta} + \|M_\theta^t\|_{\Theta}] < \infty$.

Par conséquent, pour tout $t \in \mathbb{Z}$,

$$\mathbb{E}[\|q_t(\theta)\|_{\Theta}] < \infty.$$

ainsi, de la loi forte uniforme des grands nombres pour $(q_t(\theta))$ on déduit que :

$$\left\| \frac{L_n(\theta)}{n} - L(\theta) \right\|_{\Theta} \xrightarrow[n \rightarrow \infty]{a.s.} 0. \quad (2.3.3)$$

Nous allons montrer maintenant que $\frac{1}{n} \|\widehat{L}_n(\theta) - L_n(\theta)\|_{\Theta} \xrightarrow[n \rightarrow \infty]{a.s.} 0$. En effet, pour tout $\theta \in \Theta$ et $t \in N^*$,

$$\begin{aligned} |\widehat{q}_t(\theta) - q_t(\theta)| &\leq |\log(\widehat{M}_\theta^t) - \log(M_\theta^t) + (\widehat{M}_\theta^t)^{-1}|X_t - \widehat{f}_\theta^t| - (M_\theta^t)^{-1}|X_t - f_\theta^t|| \\ &\leq |\log(\widehat{M}_\theta^t) - \log(M_\theta^t)| + |((\widehat{M}_\theta^t)^{-1} - (M_\theta^t)^{-1})|X_t - f_\theta^t| \\ &\quad + (\widehat{M}_\theta^t)^{-1}|\widehat{f}_\theta^t - f_\theta^t| \\ &\leq |\widehat{M}_\theta^t - M_\theta^t| \underline{M}^{-1} + |\widehat{M}_\theta^t - M_\theta^t| \underline{M}^{-2}|X_t - f_\theta^t| + \underline{M}^{-1}|\widehat{f}_\theta^t - f_\theta^t| \\ &\leq C(1 + |X_t| + |f_\theta^t|)(|\widehat{M}_\theta^t - M_\theta^t| + |\widehat{f}_\theta^t - f_\theta^t|), \end{aligned}$$

avec $C > 0$. Par conséquent, on a :

$$\|\widehat{q}_t(\theta) - q_t(\theta)\|_{\Theta} \leq C(1 + |X_t| + \|f_\theta^t\|_{\Theta})(\|\widehat{M}_\theta^t - M_\theta^t\|_{\Theta} + \|\widehat{f}_\theta^t - f_\theta^t\|_{\Theta}).$$

par le corollaire 1 de Kounias et Weng (1969)[40], la preuve est obtenue s'il existe un $s \in (0, 1]$ tel que :

$$\sum_{t \geq 1} \frac{1}{t^s} \mathbb{E}[\|q_t(\theta) - \widehat{q}_t(\theta)\|_{\Theta}^s] < \infty. \quad (2.3.4)$$

Prouvons (2.3.4) avec $s = r/2$ quand $r \in [1, 2]$.

De l'inégalité de Cauchy-Schwarz et les hypothèses $A_0(f, \Theta)$ et $A_0(M, \Theta)$,

$$\begin{aligned} \mathbb{E}[\|\widehat{q}_t(\theta) - q_t(\theta)\|_{\Theta}^{r/2}] &\leq C(\mathbb{E}[(1 + |X_t| + \|f_\theta^t\|_{\Theta})^r])^{\frac{1}{2}} \\ &\quad \times (\mathbb{E}[(\|\widehat{M}_\theta^t - M_\theta^t\|_{\Theta} + \|\widehat{f}_\theta^t - f_\theta^t\|_{\Theta})^r])^{\frac{1}{2}}. \end{aligned}$$

En utilisant le lemme 2 et les précédents résultats prouvés impliquant $\mathbb{E}[|X_t|^r] < \infty$ et $\mathbb{E}[\|f_\theta^t\|_{\Theta}^r + \|M_\theta^t\|_{\Theta}^r] < \infty$, nous obtenons :

$$\begin{aligned} \mathbb{E}[\|\widehat{q}_t(\theta) - q_t(\theta)\|_{\Theta}^{r/2}] &\leq C \left(\sum_{j>t} \alpha_j^0(f, \Theta) + \alpha_j^0(M, \Theta) \right)^{\frac{r}{2}} \\ &\leq C t^{-\frac{(\ell-1)r}{2}}, \end{aligned}$$

où la dernière inégalité est obtenu de la condition (2.3.2) du théorème 1.

Nous avons donc :

$$\sum_{t \geq 1} \frac{1}{t^{r/2}} \mathbb{E}[\|\widehat{q}_t(\theta) - q_t(\theta)\|_{\Theta}^{r/2}] \leq A \sum_{t \geq 1} t^{-r\ell/2},$$

qui est fini quand $r\ell > 2$. Quand $r \geq 2$, il suffit de considérer le cas $r = 2$. En conséquence, nous obtenons :

$$\frac{1}{n} \sum_{t=1}^n \|\widehat{q}_t(\theta) - q_t(\theta)\|_{\Theta} \xrightarrow[n \rightarrow \infty]{a.s.} 0 \quad \frac{1}{n} \|\widehat{L}_n(\theta) - L_n(\theta)\|_{\Theta} \xrightarrow[n \rightarrow \infty]{a.s.} 0, \quad (2.3.5)$$

d'où, en utilisant (2.3.3) nous aurons :

$$\frac{1}{n} \|\widehat{L}_n(\theta) - L(\theta)\|_{\Theta} \xrightarrow[n \rightarrow \infty]{a.s.} 0. \quad (2.3.6)$$

(ii) Pour $\theta \in \Theta$, nous étudions :

$$L(\theta) = -E[q_0(\theta)].$$

qui peut être considéré comme une divergence de Kullback-Leibler. Nous avons

$$\begin{aligned} L(\theta) &= -E[\log(M_{\theta}^t) + (M_{\theta}^t)^{-1} |X_t - f_{\theta}^t|] \\ &= -E\left[\log(M_{\theta}^t) + \frac{M_{\theta_0}^t}{M_{\theta}^t} \left|\zeta_t + \frac{f_{\theta_0}^t - f_{\theta}^t}{M_{\theta_0}^t}\right|\right]. \end{aligned}$$

Par conséquent, utilisant $E[|\zeta_t|] = 1$, nous obtenons :

$$\begin{aligned} L(\theta_0) - L(\theta) &= E\left[\log\left(\frac{M_{\theta}^t}{M_{\theta_0}^t}\right) + \frac{M_{\theta_0}^t}{M_{\theta}^t} \left|\zeta_t + \frac{f_{\theta_0}^t - f_{\theta}^t}{M_{\theta_0}^t}\right| - 1\right] \\ &= E\left[\log\left(\frac{M_{\theta}^t}{M_{\theta_0}^t}\right) - 1 + \frac{M_{\theta_0}^t}{M_{\theta}^t} E\left[\left|\zeta_t + \frac{f_{\theta_0}^t - f_{\theta}^t}{M_{\theta_0}^t}\right| \mid (X_{t-k})_{k \geq 1}\right]\right]. \end{aligned}$$

Mais pour un ζ_t suivant une distribution de probabilité symétrique, nous avons pour tout $m \in \mathbb{R}^*$, $E[|\zeta_t + m|] > E[|\zeta_t|] = 1$. d où, pour $\theta \neq \theta_0$, si $f_{\theta} \neq f_{\theta_0}$:

$$\begin{aligned} L(\theta_0) - L(\theta) &> E\left[\log\left(\frac{M_{\theta}^t}{M_{\theta_0}^t}\right) - 1 + \frac{M_{\theta_0}^t}{M_{\theta}^t}\right] \\ &> h\left(\frac{M_{\theta}^t}{M_{\theta_0}^t}\right), \end{aligned}$$

avec $h(x) = \log(x) - 1 + x$. Mais pour tout $x \in (0, 1) \cup (1, \infty)$, $h(x) > 0$ et $h(1) = 0$. Par conséquent si $M_{\theta} \neq M_{\theta_0}$, on a $h\left(\frac{M_{\theta}^t}{M_{\theta_0}^t}\right) > 0$ (> 0 est remplacé

par $= 0$ si $M_\theta = M_{\theta_0}$). Cela implique, de l'hypothèse H3 (identifiabilité), que $L(\theta_0) - L(\theta) > 0$ presque sûrement pour tout $\theta \in \Theta$, $\theta \neq \theta_0$. D'où une borne supérieure de $L(\theta)$ est atteinte seulement pour $\theta = \theta_0$ qui est un maximum unique.

2.3.3 La normalité asymptotique

Une extension de théorème 1 prouvé dans Davis et Dunsmuir (1997) [14] sera donnée ci-après et consistera une étape essentielle pour la preuve de la normalité asymptotique de l'estimateur.

Théorème 2 : Soit $(Z_t)_{t \in \mathbb{Z}}$ une suite de (VAIID) telle que $\text{Var}(Z_0) = \sigma^2 < \infty$, avec la fonction de distribution commune qui est symétrique ($F(-x) = 1 - F(x)$ pour $x \in \mathbb{R}$) et continûment différentiable dans un voisinage de 0 avec une dérivée $f(0)$ en 0. Notons $\mathcal{F}_t = \sigma(Z_t, Z_{t-1}, \dots)$ pour $t \in \mathbb{Z}$ et soit $(Y_t)_{t \in \mathbb{Z}}$ et $(V_t)_{t \in \mathbb{Z}}$ deux processus stationnaires adaptés à $(\mathcal{F}_t)_t$ et comme $E[Y_0^2 V_0^2] < \infty$. alors

$$\sum_{t=1}^n V_{t-1} (|Z_t - n^{-1/2} Y_{t-1}| - |Z_t|) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(f(0) E[V_0 Y_0^2], E[V_0^2 Y_0^2]) \quad (2.3.7)$$

La normalité asymptotique de l'EQMV-Laplace peut être établie utilisant des hypothèses supplémentaires.

Preuve du Théorème 2

Nous suivons le même schéma de preuve que Davis et Dunsmuir (1997) [14]. Donc, désignons par :

$$\begin{aligned} S_n &= \sum_{t=1}^n V_{t-1} (|Z_t - n^{-1/2} Y_{t-1}| - |Z_t|) \\ &= -n^{-1/2} \sum_{t=1}^n V_{t-1} Y_{t-1} \text{sgn}(Z_t) \\ &\quad + 2 \sum_{t=1}^n V_{t-1} (n^{-1/2} Y_{t-1} - Z_t) (\mathbf{1}_{0 < Z_t < n^{-1/2} Y_{t-1}} - \mathbf{1}_{n^{-1/2} Y_{t-1} < Z_t < 0}) \\ &= A_n + B_n. \end{aligned}$$

Dés que $E[V_{t-1}Y_{t-1}\text{sgn}(Z_t) / \mathcal{F}_{t-1}] = E[\text{sgn}(Z_t)]E[V_{t-1}Y_{t-1}] = 0$ et $E[V_0^2Y_0^2] < \infty$, on peut appliquer le théorème central limite pour la suite de différences de martingales stationnaire (voir Billingsley, 1968 [7]) et :

$$A_n \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, E[V_0^2Y_0^2]). \quad (2.3.8)$$

Considérant B_n et définissant $W_{nt} = V_{t-1}(n^{-1/2}Y_{t-1} - Z_t)\mathbf{1}_{0 < Z_t < n^{-1/2}Y_{t-1}}$. Utilisant les mêmes arguments que Davis et Dunsmuir (1997) [14], nous obtenons également :

- $\limsup_{n \rightarrow \infty} n E[W_{nt}^2] = 0$;
- $E[W_{nt} / \mathcal{F}_{t-1}] \simeq \frac{1}{2n} f(0) V_{t-1} Y_{t-1}^2 \quad \text{for } |n^{-1/2}Y_{t-1}| < \varepsilon$;
- $\sum_{t=1}^n W_{nt} \xrightarrow[n \rightarrow \infty]{\mathcal{P}} \frac{1}{2} f(0) E[V_0 Y_0^2 \mathbf{1}_{Y_0 > 0}]$.

Donc, nous déduisons que :

$$B_n \xrightarrow[n \rightarrow \infty]{\mathcal{P}} f(0) E[V_0 Y_0^2]. \quad (2.3.9)$$

La preuve est achevée de (2.3.8) et (2.3.9).

Soit $\mathcal{D}^{(2)}(\Theta)$ l'espace de Banach de fonctions deux fois continuellement différentiable sur Θ muni de la norme uniforme :

$$\|h\|_{2,\Theta} = \|h\|_{\Theta} + \left\| \frac{\partial h}{\partial \theta} \right\|_{\Theta} + \left\| \frac{\partial^2 h}{\partial \theta \partial \theta'} \right\|_{\Theta}.$$

Lemme 3 Soit $\theta_0 \in \Theta(r)$ ($r \geq 1$) et supposons que pour $i = 0, 1, 2$, $(A_i(f, \Theta))$ et $(A_i(M, \Theta))$ demeurent vérifier, alors : $f_{\theta}^t \in \mathbb{L}^r(\mathcal{D}^{(2)}(\Theta))$ et $M_{\theta}^t \in \mathbb{L}^r(\mathcal{D}^{(2)}(\Theta))$.

Théoreme 3 : Supposons que $\theta_0 \in \overset{\circ}{\Theta} \cap \Theta(r)$ où $r \geq 2$ et $\overset{\circ}{\Theta}$ désigne l'intérieur de Θ . Soit X la solution stationnaire de l'équation (2.1.1). Supposons que les hypothèses de théorème1 demeurent vérifier et pour $i = 1, 2$, supposons que $(A_i(f, \Theta))$ et $(A_i(M, \Theta))$ sont vérifiées. Alors, si la fonction de probabilité cumulative de ζ_0 est continûment différentiable dans un voisinage de 0 avec

une dérivée $g(0)$ en 0 et si la matrice Γ_F ou Γ_M , défini dans (2.3.22), sont des matrices symétriques définies positives, alors :

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}_d\left(0, Q((\sigma_\zeta^2 - 1)\Gamma_M + \Gamma_F)Q\right), \quad (2.3.10)$$

où $Q = (\Gamma_M + 2g(0)\Gamma_F)^{-1}$.

Comme il a déjà été prouvé pour l'estimateur de la médiane (voir Van der Vaart (1996) [71]) ou pour l'estimateur des moindres écarts absolus des processus ARMA (voir Davis et Dunsmuir (1997) [14]), c'est peu surprenant que la fonction de densité de probabilité g du bruit blanc $(\zeta_i)_i$ influe la covariance asymptotique de (2.3.10). Cependant, quand $f_\theta = 0$, ce n'est pas le cas et c'est ce qui se passe pour les processus GARCH (voir Francq *et al.* (2011) [29], ou Francq et Zakoian (2015) [31]) où g n'apparaît pas dans la covariance asymptotique.

Preuve de Théorème 3

Soit $v = \sqrt{n}(\theta - \theta_0) \in \mathbb{R}^d$. Nous allons prouver en 2/ que la maximisation de $\hat{L}_n(\theta)$ est équivalente à la maximisation de $L_n(\theta)$ qui est équivalente à la maximisation de :

$$W_n(v) = - \sum_{t=1}^n (q_t(\theta_0 + n^{-1/2}v) - q_t(\theta_0)) \quad (2.3.11)$$

$$\begin{aligned} &= \sum_{t=1}^n \log \left(\frac{(M_{\theta_0 + n^{-1/2}v}^t)^{-1}}{(M_{\theta_0}^t)^{-1}} \right) + (M_{\theta_0}^t)^{-1} |X_t - f_{\theta_0}^t| \\ &- \sum_{t=1}^n (M_{\theta_0 + n^{-1/2}v}^t)^{-1} |X_t - f_{\theta_0 + n^{-1/2}v}^t| \end{aligned} \quad (2.3.12)$$

par rapport à v . Par conséquent, il existe une suite $(\hat{v}_n)_n$ où $\hat{v}_n$ est un maximisateur de $W_n(v)$ tel que $\hat{v}_n = \sqrt{n}(\hat{\theta}_n - \theta_0)$. En 1/ nous allons fournir un théorème limite satisfait par $W_n(v)$. Ensuite, nous allons prouver en 3/ que $(W_n(\cdot))_n$ converge comme étant un processus de $\mathcal{C}(\mathbb{R}^d)$ (espace des fonctions continues sur $\mathbb{R}^d$) vers un processus limite W . Ainsi $(\hat{v}_n)_n$ converge vers le maximiseur de W .

1/ Tout d'abord, nous allons étudier le comportement asymptotique de $W_n(v)$. Nous avons :

$$\begin{aligned}
W_n(v) &= \sum_{t=1}^n \log \left(\frac{(M_{\theta_0+n^{-1/2}v}^t)^{-1}}{(M_{\theta_0}^t)^{-1}} \right) + |X_t - f_{\theta_0}^t| ((M_{\theta_0}^t)^{-1} - (M_{\theta_0+n^{-1/2}v}^t)^{-1}) \\
&\quad + \sum_{t=1}^n (M_{\theta_0+n^{-1/2}v}^t)^{-1} (|X_t - f_{\theta_0}^t| - |X_t - f_{\theta_0+n^{-1/2}v}^t|) \\
&= I_1(v) + I_2(v).
\end{aligned}$$

Nous avons :

$$I_1(v) = - \sum_{t=1}^n \log \left(\frac{M_{\theta_0+n^{-1/2}v}^t}{M_{\theta_0}^t} \right) + |\zeta_t| \left(1 - \frac{M_{\theta_0}^t}{M_{\theta_0+n^{-1/2}v}^t} \right).$$

Utilisation du développement de Taylor, on déduit par que pour chaque $t \in \{1, \dots, n\}$, il existe un $\bar{\theta}_1^t$ et un $\bar{\theta}_2^t$ dans le segment $[\theta_0, \theta_0 + n^{-1/2}v]$ tel que :

$$\begin{aligned}
\log \left(\frac{M_{\theta_0+n^{-1/2}v}^t}{M_{\theta_0}^t} \right) &= n^{-1/2} (M_{\theta_0}^t)^{-1} v' \left(\frac{\partial M_{\theta}^t}{\partial \theta} \right)_{\theta_0} \\
&\quad + \frac{1}{2} n^{-1} \left\{ v' \left(\frac{\partial^2 M_{\theta}^t}{\partial \theta^2} \right)_{\bar{\theta}_1^t} - (M_{\theta_0}^t)^{-2} \left(v' \left(\frac{\partial M_{\theta}^t}{\partial \theta} \right)_{\bar{\theta}_1^t} \right)^2 \right\} \\
\frac{M_{\theta_0}^t}{M_{\theta_0+n^{-1/2}v}^t} &= 1 - n^{-1/2} (M_{\theta_0}^t)^{-1} v' \left(\frac{\partial M_{\theta}^t}{\partial \theta} \right)_{\theta_0} \\
&\quad + \frac{1}{2} n^{-1} \left\{ 2 (M_{\theta_0}^t)^{-2} \left(v' \left(\frac{\partial M_{\theta}^t}{\partial \theta} \right)_{\bar{\theta}_2^t} \right)^2 - v' \left(\frac{\partial^2 M_{\theta}^t}{\partial \theta^2} \right)_{\bar{\theta}_2^t} \right\}
\end{aligned}$$

Alors :

$$\begin{aligned}
I_1(v) &= n^{-1/2} \sum_{t=1}^n (M_{\theta_0}^t)^{-1} v' \left(\frac{\partial M_{\theta}^t}{\partial \theta} \right)_{\theta_0} (|\zeta_t| - 1) \\
&\quad + \frac{1}{2n} \sum_{t=1}^n (M_{\theta_0}^t)^{-1} \left\{ v' \left(\frac{\partial^2 M_{\theta}^t}{\partial \theta^2} \right)_{\bar{\theta}_2^t} v |\zeta_t| - v' \left(\frac{\partial^2 M_{\theta}^t}{\partial \theta^2} \right)_{\bar{\theta}_1^t} v \right\} \\
&\quad + \frac{1}{2n} \sum_{t=1}^n (M_{\theta_0}^t)^{-2} \left\{ \left(v' \left(\frac{\partial M_{\theta}^t}{\partial \theta} \right)_{\bar{\theta}_1^t} \right)^2 - 2 \left(v' \left(\frac{\partial M_{\theta}^t}{\partial \theta} \right)_{\bar{\theta}_2^t} \right)^2 |\zeta_t| \right\} \\
&= I_1^{(1)}(v) + I_1^{(2)}(v) + I_1^{(3)}(v).
\end{aligned}$$

Utilisant le théorème central limite pour les différences de martingale (voir [7]), comme $E[(M_{\theta_0}^t)^{-1}v' \left(\frac{\partial M_{\theta}^t}{\partial \theta} \right)_{\theta_0} (|\zeta_t| - 1) / \mathcal{F}_{t-1}] = 0$ et $E[\|(M_{\theta_0}^t)^{-1}v' \left(\frac{\partial M_{\theta}^t}{\partial \theta} \right)_{\theta_0}\|_{\Theta}^2] < \infty$ (de Lemme 3), on a :

$$I_1^{(1)}(v) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}\left(0, E\left[(M_{\theta_0}^0)^{-2} \left(v' \left(\frac{\partial M_{\theta}^0}{\partial \theta} \right)_{\theta_0}\right)^2\right] (\sigma_{\zeta}^2 - 1)\right). \quad (2.3.13)$$

Sachant que $\theta \in \Theta \mapsto \frac{\partial M_{\theta}^t}{\partial \theta}$ et $\theta \mapsto \frac{\partial^2 M_{\theta}^t}{\partial \theta^2}$ sont des fonctions continues, $\bar{\theta}_1^t \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \theta_0$ et $\bar{\theta}_2^t \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \theta_0$, nous affirmons que $I_1^{(2)}(v)$ a la même distribution

limite que $\frac{1}{2n} \sum_{t=1}^n (M_{\theta_0}^t)^{-1} v' \left(\frac{\partial^2 M_{\theta}^t}{\partial \theta^2} \right)_{\theta_0} v (|\zeta_t| - 1)$.

Notons que $E[(M_{\theta_0}^t)^{-1}v' \left(\frac{\partial^2 M_{\theta}^t}{\partial \theta^2} \right)_{\theta_0} v (|\zeta_t| - 1) / \mathcal{F}_{t-1}] = 0$ et que $E[\|(M_{\theta_0}^t)^{-1}v' \left(\frac{\partial^2 M_{\theta}^t}{\partial \theta^2} \right)_{\theta_0} v\|] < \infty$ (du Lemme 3). D'où, par la loi forte des grands nombres pour les différences de martingales (voir [7]), nous obtenons :

$$\frac{1}{2n} \sum_{t=1}^n (M_{\theta_0}^t)^{-1} v' \left(\frac{\partial^2 M_{\theta}^t}{\partial \theta^2} \right)_{\theta_0} v (|\zeta_t| - 1) \xrightarrow[n \rightarrow \infty]{a.s.} 0,$$

et cela implique que :

$$I_1^{(2)}(v) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} 0. \quad (2.3.14)$$

Des arguments similaires aux précédentes induisent que $I_1^{(3)}(v)$ a la même distribution limite que $\frac{1}{2n} \sum_{t=1}^n (M_{\theta_0}^t)^{-2} \left(v' \left(\frac{\partial M_{\theta}^t}{\partial \theta} \right)_{\theta_0}\right)^2 (1 - 2|\zeta_t|)$. De la loi forte des grands nombres pour les différences de martingales (voir [7]), nous obtenons :

$$\begin{aligned} \frac{1}{2n} \sum_{t=1}^n (M_{\theta_0}^t)^{-2} \left(v' \left(\frac{\partial M_{\theta}^t}{\partial \theta} \right)_{\theta_0}\right)^2 (1 - 2|\zeta_t|) &\xrightarrow[n \rightarrow \infty]{a.s.} \frac{1}{2} E\left[(M_{\theta_0}^0)^{-2} \left(v' \left(\frac{\partial M_{\theta}^0}{\partial \theta} \right)_{\theta_0}\right)^2 (1 - 2|\zeta_0|)\right] \\ &\xrightarrow[n \rightarrow \infty]{a.s.} -\frac{1}{2} E\left[(M_{\theta_0}^0)^{-2} \left(v' \left(\frac{\partial M_{\theta}^0}{\partial \theta} \right)_{\theta_0}\right)^2\right], \end{aligned}$$

et cela implique que :

$$I_1^{(3)}(v) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} -\frac{1}{2} E\left[(M_{\theta_0}^0)^{-2} \left(v' \left(\frac{\partial M_{\theta}^0}{\partial \theta} \right)_{\theta_0}\right)^2\right]. \quad (2.3.15)$$

Enfin, de (2.3.13), (2.3.14) et (2.3.15), nous avons :

$$I_1(v) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}\left(-\frac{1}{2} \mathbb{E}\left[(M_{\theta_0}^0)^{-2} \left(v' \left(\frac{\partial M_{\theta}^0}{\partial \theta}\right)_{\theta_0}\right)^2\right], \mathbb{E}\left[(M_{\theta_0}^0)^{-2} \left(v' \left(\frac{\partial M_{\theta}^0}{\partial \theta}\right)_{\theta_0}\right)^2\right] (\sigma_{\zeta}^2 - 1)\right). \quad (2.3.16)$$

Maintenant, considérons $I_2(v) = \sum_{t=1}^n (M_{\theta_0+n^{-1/2}v}^t)^{-1} (|X_t - f_{\theta_0}^t| - |X_t - f_{\theta_0+n^{-1/2}v}^t|)$. Utilisant de nouveau le développement de Taylor, nous pouvons écrire :

$$\begin{aligned} I_2(v) &= \sum_{t=1}^n (M_{\theta_0}^t)^{-1} (|X_t - f_{\theta_0}^t| - |X_t - f_{\theta_0+n^{-1/2}v}^t|) \\ &\quad - n^{-1/2} \sum_{t=1}^n (M_{\bar{\theta}_M^t}^t)^{-2} v' \left(\frac{\partial M_{\theta}^t}{\partial \theta}\right)_{\bar{\theta}_M^t} (|X_t - f_{\theta_0}^t| - |X_t - f_{\theta_0+n^{-1/2}v}^t|), \\ &= I_2^{(1)}(v) + I_2^{(2)}(v), \end{aligned}$$

avec $\bar{\theta}_M^t$ dans le segment $[\theta_0, \theta_0 + n^{-1/2}v]$.

D'abord nous avons :

$$I_2^{(1)}(v) = \sum_{t=1}^n (|\zeta_t| - |\zeta_t - n^{-1/2} (M_{\theta_0}^t)^{-1} v' \left(\frac{\partial f_{\theta}^t}{\partial \theta}\right)_{\bar{\theta}_f^t}|)$$

avec $\bar{\theta}_f^t$ dans le segment $[\theta_0, \theta_0 + n^{-1/2}v]$. Utilisant le théorème 2, qui est une extension du théorème 1 de Davis et Dunsmuir (1997) [14], en notant $Z_t = \zeta_t$, $Y_t = (M_{\theta_0}^t)^{-1} v' \left(\frac{\partial f_{\theta}^t}{\partial \theta}\right)_{\bar{\theta}_f^t}$ et $V_t = 1$, pour $t \in \mathbb{Z}$:

$$I_2^{(1)} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}\left(-g(0) \mathbb{E}[(M_{\theta_0}^0)^{-2} (v' \left(\frac{\partial f_{\theta}^0}{\partial \theta}\right)_{\theta_0})^2], \mathbb{E}[(M_{\theta_0}^0)^{-2} (v' \left(\frac{\partial f_{\theta}^0}{\partial \theta}\right)_{\theta_0})^2]\right). \quad (2.3.17)$$

Du lemme 3 nous avons $\mathbb{E}[Y_t^2 V_t^2] \leq \underline{M}^{-2} \mathbb{E}[\|v' \frac{\partial f_{\theta}^t}{\partial \theta}\|_{\Theta}^2] < \infty$.

Par conséquent :

$$\begin{aligned} I_2^{(2)}(v) &\sim n^{-1/2} \sum_{t=1}^n (M_{\theta_0}^t)^{-2} v' \left(\frac{\partial M_{\theta}^t}{\partial \theta}\right)_{\theta_0} (|X_t - f_{\theta_0}^t| - |X_t - f_{\theta_0+n^{-1/2}v}^t|) \\ &\sim n^{-1/2} \sum_{t=1}^n (M_{\theta_0}^t)^{-1} v' \left(\frac{\partial M_{\theta}^t}{\partial \theta}\right)_{\theta_0} (|\zeta_t| - |\zeta_t - n^{-1/2} (M_{\theta_0}^t)^{-1} v' \left(\frac{\partial f_{\theta}^t}{\partial \theta}\right)_{\bar{\theta}_f^t}|) \\ &\xrightarrow[n \rightarrow \infty]{\mathcal{P}} 0, \end{aligned} \quad (2.3.18)$$

utilisant la preuve de théorème 2 et notant $Z_t = \zeta_t$, $Y_t = (M_{\theta_0}^t)^{-1} v' \left(\frac{\partial f_{\theta}^t}{\partial \theta} \right)_{\theta_f^t}$ et $V_t = (M_{\theta_0}^t)^{-1} v' \left(\frac{\partial M_{\theta}^t}{\partial \theta} \right)_{\theta_0}$ pour $t \in \mathbb{Z}$ et la condition $\mathbb{E}[|V_t Y_t|] < \infty$ qui assure la loi forte des grands nombres au lieu de théorème central limite pour un processus de différences de martingales. Alors à partir de (2.3.17) et (2.3.18), nous déduisons :

$$I_2(v) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N} \left(-g(0) \mathbb{E}[(M_{\theta_0}^0)^{-2} (v' \left(\frac{\partial f_{\theta}^0}{\partial \theta} \right)_{\theta_0})^2], \mathbb{E}[(M_{\theta_0}^0)^{-2} (v' \left(\frac{\partial f_{\theta}^0}{\partial \theta} \right)_{\theta_0})^2] \right). \quad (2.3.19)$$

Enfin, nous obtenons le comportement de $W_n(v)$ défini dans (2.3.11) de (2.3.16) et (2.3.19). Cependant, nous devons préciser la relation de dépendance asymptotique entre $I_1^{(1)}$ et $I_2^{(1)}$. En effet, ces deux termes convergent vers une loi Gaussienne. Cela implique la considération de comportement asymptotique de la somme de ces deux termes qui pourraient être réduits au comportement asymptotique de :

$$n^{-1/2} \sum_{t=1}^n (M_{\theta_0}^t)^{-1} v' \left(\frac{\partial M_{\theta}^t}{\partial \theta} \right)_{\theta_0} (|\zeta_t| - 1) + n^{-1/2} \sum_{t=1}^n (M_{\theta_0}^t)^{-1} v' \left(\frac{\partial f_{\theta}^t}{\partial \theta} \right)_{\theta_0} \text{sgn}(\zeta_t),$$

à partir de la preuve de théorème 2 et en utilisant à nouveau le théorème central limite pour les différences de martingales, nous obtenons comme variance asymptotique :

$$\begin{aligned} & \mathbb{E} \left[(M_{\theta_0}^t)^{-2} \left(v' \left(\frac{\partial M_{\theta}^t}{\partial \theta} \right)_{\theta_0} (|\zeta_t| - 1) + v' \left(\frac{\partial f_{\theta}^t}{\partial \theta} \right)_{\theta_0} \text{sgn}(\zeta_t) \right)^2 \right] \\ &= \mathbb{E} \left[(M_{\theta_0}^0)^{-2} \left(v' \left(\frac{\partial M_{\theta}^0}{\partial \theta} \right)_{\theta_0} \right)^2 \right] (\sigma_{\zeta}^2 - 1) \\ &+ 2 \mathbb{E} \left[(M_{\theta_0}^0)^{-2} v' \left(\frac{\partial M_{\theta}^0}{\partial \theta} \right)_{\theta_0} v' \left(\frac{\partial f_{\theta}^0}{\partial \theta} \right)_{\theta_0} \right] \mathbb{E}[(|\zeta_t| - 1) \text{sgn}(\zeta_t)] \\ &+ \mathbb{E} \left[(M_{\theta_0}^0)^{-2} \left(v' \left(\frac{\partial f_{\theta}^0}{\partial \theta} \right)_{\theta_0} \right)^2 \right] \\ &= \mathbb{E} \left[(M_{\theta_0}^0)^{-2} \left\{ (\sigma_{\zeta}^2 - 1) \left(v' \left(\frac{\partial M_{\theta}^0}{\partial \theta} \right)_{\theta_0} \right)^2 + \left(v' \left(\frac{\partial f_{\theta}^0}{\partial \theta} \right)_{\theta_0} \right)^2 \right\} \right]. \end{aligned} \quad (2.3.20)$$

comme $(\zeta_t)_t$ admet une distribution de probabilité symétrique avec une médiane et espérance nulle. Alors, il n'y a pas de terme de covariance et

finalement nous obtenons :

$$W_n(v) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} W(v) = v' \left(-\frac{1}{2} \Gamma_M - g(0) \Gamma_F \right) v + v' N \quad (2.3.21)$$

$$\text{avec : } \begin{cases} N \stackrel{\mathcal{L}}{=} \mathcal{N}(0, ((\sigma_\zeta^2 - 1) \Gamma_M + \Gamma_F)) \\ \Gamma_F = \left(\mathbb{E} \left[(M_{\theta_0}^0)^{-2} \left(\frac{\partial f_\theta^0}{\partial \theta_i} \right)_{\theta_0} \left(\frac{\partial f_\theta^0}{\partial \theta_j} \right)_{\theta_0} \right] \right)_{1 \leq i, j \leq d} \\ \Gamma_M = \left(\mathbb{E} \left[\left(\frac{\partial \log(M_\theta^0)}{\partial \theta_i} \right)_{\theta_0} \left(\frac{\partial \log(M_\theta^0)}{\partial \theta_j} \right)_{\theta_0} \right] \right)_{1 \leq i, j \leq d} \end{cases} \quad (2.3.22)$$

2/ Nous considérons maintenant l'approximation $\widehat{W}_n(v)$ de $W_n(v)$ définie par :

$$\widehat{W}_n(v) = - \sum_{t=1}^n (\widehat{q}_t(\theta_0 + n^{-1/2}v) - \widehat{q}_t(\theta_0)) \quad \text{pour tout } v \in \mathbb{R}^d.$$

À partir des hypothèses de théorème 1 et (2.3.5) on a $\frac{1}{n} \sum_{t=1}^n \|\widehat{q}_t(\theta) - q_t(\theta)\|_\Theta \xrightarrow[n \rightarrow \infty]{a.s.} 0$. Alors on a $\widehat{W}_n(v) = W_n(v) + R_n(v)$ avec $[\sup_{v \in \mathbb{R}^d} |R_n(v)|] \leq 2 \sum_{t=1}^n [\|\widehat{q}_t(\theta) - q_t(\theta)\|_\Theta] \xrightarrow[n \rightarrow \infty]{a.s.} 0$ et donc :

$$\widehat{W}_n(v) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} W(v), \quad (2.3.23)$$

avec W définie dans (2.3.21).

3/ De (2.3.23), la preuve du théorème 2 et les mêmes arguments que ceux de la précédente preuve du théorème 2 de Davis et Dunsmuir (1997) [14], nous déduisons que les distributions finies $(\widehat{W}_n(v_1), \dots, \widehat{W}_n(v_k))$ converge vers $(W(v_1), \dots, W(v_k))$ pour tout $(v_1, \dots, v_k) \in (\mathbb{R}^d)^k$. De plus, toujours en suivant la preuve de théorème 2, $(W_n(v))_v$ converge vers $(W(v))_v$ comme étant un processus dans l'espace des fonctions continues $\mathcal{C}^0$. Par conséquent, le maximum $\widehat{v}$ de $\widehat{W}_n(v)$ satisfait :

$$\widehat{v} = (\Gamma_M + 2g(0) \Gamma_F)^{-1} N,$$

et cela implique (2.3.10).

2.4 Exemples de modèles particuliers

Dans cette section, plusieurs types de modèles des séries chronologiques qui satisfaisaient les conditions des résultats précédents sont explicitement étudiées.

D'une part, nous définissons un espace de stationnarité pour chaque modèle (à savoir les modèles ARMA, GARCH, APARCH, ARMA-GARCH, ARMA-ARCH(∞), ARMA-APARCH) et des propositions donnant la consistance et la normalité asymptotique pour l'EQMV-Laplace de ces modèles.

D'autre part, pour illustrer les résultats asymptotiques énoncés, nous réalisons des expériences de Monte-Carlo sur le comportement d'EQMV-Laplace (noté $\widehat{\theta}_n^{LQL}$) pour plusieurs modèles de séries chronologiques, tailles d'échantillons et distributions de probabilités (après la normalisation de sorte que $E|\zeta_t| = 1$). Nous proposons également une comparaison avec les résultats obtenus par l'EQMV Gaussien noté $\widehat{\theta}_n^{GQL}$ (après une normalisation de sorte que $E|\zeta_t|^2 = 1$).

Plus précisément, les distributions de probabilités considérées de (ζ_t) sont :

- Distribution Gaussienne centrée notée $\mathcal{N}$;
- Distribution de Laplace Centrée notée $\mathcal{L}$;
- Distribution Uniforme Centrée notée $\mathcal{U}$;
- Distribution de Student Centrée avec degré de liberté 3, notée t_3 ;
- Mélange Gaussien centré et normalisé avec une distribution de probabilité $0.05 * \mathcal{N}(-2, 0.16) + \mathcal{N}(0, 1) + 0.05 * \mathcal{N}(2, 0.16)$ et notée $\mathcal{M}$.

Toutes ces distributions de probabilité sont normalisées de sorte que $E[|\zeta_0|] = 1$, requis pour l'EQMV- Laplace. Pour utiliser l'EQMV Gaussien nécessitant $E[|\zeta_0|] = 1$, il est nécessaire de considérer le modèle avec $M'_\theta = \frac{E[|\zeta_0|]}{\sqrt{\text{Var}(\zeta_0)}} M_\theta$ au lieu de M_θ .

2.4.1 Processus ARMA

Le processus ARMA(p,q) a été introduit (par Box et Jenkins (1976) [8]) comme solution de l'équation :

$$\Phi_\theta(L) \cdot X_t = \Psi_\theta(L) \cdot \varepsilon_t, \quad (2.4.1)$$

où $\Phi_\theta(L) = \sum_{i=1}^p \phi_i L^i$, $\Psi_\theta(L) = \sum_{i=1}^q \psi_i L^i$, $(\phi_i)_{i=1;p}$ et $(\psi_j)_{j=1;q}$ des nombres réelles positives et les polynômes $\Phi_\theta(L)$, $\Psi_\theta(L)$ sont coprimés, notons $\theta = (\psi_1, \dots, \psi_q, \phi_1, \dots, \phi_p)$. L'équation (2.4.1) a la représentation (2.1.1) avec $f_\theta(X_{t-1}, X_{t-2}, \dots) = (\Phi_\theta^{-1}(L)\Psi_\theta(L))X_t = (\sum_{i=1}^\infty \gamma_i(\theta)L^i)X_t$ et $M_\theta^t = 1$. Pour ce processus les hypothèses $(A_0(f, \Theta))$ et $(A_0(M, \Theta))$ sont vérifiées, donc nous définissons l'espace de stationnarité par :

$$\Theta_1 = \left\{ \theta \in \mathbb{R}^d \mid \sum_{i=1}^\infty |\gamma_i(\theta)| < 1 \right\}.$$

Si $\theta_0 \in \Theta_1$, alors le processus (2.4.1) admet une solution stationnaire.

Les hypothèses H1, H2, H3 sont toujours vérifiées, ce qui nous permet d'énoncer la proposition suivante :

Proposition 2 *Soit $\theta_0 \in \Theta$ un sous ensemble compact de Θ_1 et X la solution stationnaire de (2.4.1).*

1. L'EQMV est fortement consistant $\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{a.s.} \theta_0$.
2. Si Γ_M est une matrice symétrique définie positive, alors :

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}_d(0, Q((\sigma_\zeta^2 - 1) + \Gamma_F)Q^{-1}). \quad (2.4.2)$$

où $Q = (1 + 2g(0)\Gamma_F)$ et $\Gamma_F = \left(\mathbb{E} \left[(M_{\theta_0}^0)^{-2} \left(\frac{\partial f_\theta^0}{\partial \theta_i} \right)_{\theta_0} \left(\frac{\partial f_\theta^0}{\partial \theta_j} \right)_{\theta_0} \right] \right)_{1 \leq i, j \leq d}$.

Application numérique pour l'ARMA

Dans le tableau 3.7.1, nous exposons une simulation type Monte-Carlo pour le processus ARMA(1,1) définie par $X_t = \phi X_{t-1} + \zeta_t + \theta \zeta_{t-1}$ avec $\phi = 0.6, \theta = 0.4$.

		$\mathcal{L}$		$\mathcal{N}$		t_3		$\mathcal{U}$		$\mathcal{M}$	
n		$\widehat{\theta}_n^{GQL}$	$\widehat{\theta}_n^{LQL}$	$\widehat{\theta}_n^{GQL}$	$\widehat{\theta}_n^{LQL}$	$\widehat{\theta}_n^{GQL}$	$\widehat{\theta}_n^{LQL}$	$\widehat{\theta}_n^{GQL}$	$\widehat{\theta}_n^{LQL}$	$\widehat{\theta}_n^{GQL}$	$\widehat{\theta}_n^{LQL}$
θ	100	0.106	0.091	0.114	0.117	0.113	0.090	0.112	0.059	0.110	0.078
	1000	0.031	0.024	0.032	0.032	0.036	0.027	0.031	0.014	0.031	0.023
	5000	0.014	0.010	0.014	0.015	0.016	0.011	0.016	0.012	0.013	0.010
ϕ	100	0.119	0.102	0.121	0.128	0.123	0.102	0.120	0.067	0.121	0.090
	1000	0.037	0.028	0.036	0.036	0.040	0.030	0.036	0.017	0.036	0.027
	5000	0.016	0.012	0.014	0.016	0.017	0.013	0.016	0.007	0.014	0.006

TABLE 2.1 – Racine d’erreur moyenne quadratique $\widehat{\theta}_n^{LQL}$ et $\widehat{\theta}_n^{GQL}$ pour les processus ARMA(1, 1).

2.4.2 Processus GARCH

Le processus GARCH(p', q') a été introduit (par Bollerslev (1986) [10]) comme solution de système d’équation :

$$\begin{cases} X_t = \sigma_t \cdot \varepsilon_t, \\ \sigma_t^2 = c_0 + \sum_{i=1}^{p'} c_i X_{t-i}^2 + \sum_{i=1}^{q'} d_i \sigma_{t-i}^2, \end{cases} \quad (2.4.3)$$

avec $c_0 > 0$, $c_i \geq 0$ pour $i = 1, p' - 1$, $d_i \geq 0$ pour $i = 1, q' - 1$, $c_{p'}, d_{q'}$ strictement positifs et $\sum_{i=1}^{q'} d_i < 1$, notons $\theta = (c_1, \dots, c_{p'}, d_1, \dots, d_{q'})$. On peut représenter σ_t sous la forme causale $\sigma_t^2 = \sum_{j=0}^{\infty} b_j X_{t-j}^2$ d’où l’équation (2.4.3) admet la représentation (2.1.1) avec $f_{\theta}(X_{t-1}, X_{t-2}, \dots) = 0$ et $M_{\theta}^t = \sigma_t$. Par ailleurs, comme $f = 0$ la condition $A_0(f, \Theta)$ est vérifiée et $\sum_{i=1}^q d_i < 1$ implique que la suite définie par $u_n = \sum_{k=1}^{q'} d_k u_{n-k}$ est décroissante de façon exponentielle et donc $A_0(M, \Theta)$ est aussi vérifiée. D’où pour ce processus, nous définissons :

$$\Theta_2(r) = \left\{ \theta \in \mathbb{R}^d \mid \left(\mathbb{E}[|\xi_0|^r] \right)^{1/r} \sum_{j=1}^{\infty} |b_j| < 1 \right\}.$$

Si $\theta_0 \in \Theta_2$, alors le processus (2.4.3) a une solution stationnaire.

Proposition 3 *Soit Θ un sous ensemble compact de $\Theta_2(r)$ et X la solution stationnaire de (2.4.3).*

1. Si $r = 1$, alors l'EQMV est fortement consistant $\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{a.s.} \theta_0$.
2. Si $r = 2$, et si Γ_M est une matrice symétrique définie positive, alors :

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}_d\left(0, (\sigma_\zeta^2 - 1) \Gamma_M^{-1}\right). \quad (2.4.4)$$

Application numérique pour le GARCH

Dans le tableau 2.2, nous avons les expériences de Monte-Carlo sur le processus GARCH(1, 1) définie par $X_t = \sqrt{\alpha_0 + \alpha_1 X_{t-1}^2 + \beta \zeta_{t-1}^2} \zeta_t$ avec $\alpha_0 = 0.2$, $\alpha_1 = 0.4$ et $\beta = 0.2$.

		$\mathcal{L}$		$\mathcal{N}$		t_3		$\mathcal{U}$		$\mathcal{M}$	
	n	$\hat{\theta}_n^{GQL}$	$\hat{\theta}_n^{LQL}$	$\hat{\theta}_n^{GQL}$	$\hat{\theta}_n^{LQL}$	$\hat{\theta}_n^{GQL}$	$\hat{\theta}_n^{LQL}$	$\hat{\theta}_n^{GQL}$	$\hat{\theta}_n^{LQL}$	$\hat{\theta}_n^{GQL}$	$\hat{\theta}_n^{LQL}$
α_0	100	0.112	0.105	0.095	0.100	0.211	0.126	0.081	0.047	0.134	0.114
	1000	0.036	0.032	0.028	0.028	0.098	0.058	0.023	0.018	0.066	0.051
	5000	0.016	0.014	0.012	0.012	0.055	0.043	0.010	0.015	0.040	0.023
α_1	100	0.162	0.157	0.149	0.150	0.453	0.364	0.115	0.070	0.507	0.429
	1000	0.061	0.056	0.449	0.449	0.333	0.150	0.030	0.033	0.160	0.136
	5000	0.029	0.026	0.020	0.020	0.193	0.095	0.013	0.030	0.086	0.058
β	100	0.225	0.209	0.188	0.190	0.499	0.429	0.163	0.105	0.483	0.390
	1000	0.060	0.055	0.051	0.051	0.285	0.174	0.044	0.022	0.170	0.169
	5000	0.027	0.024	0.022	0.022	0.180	0.075	0.019	0.009	0.072	0.075

TABLE 2.2 – Racine d'erreur moyenne quadratique $\hat{\theta}_n^{LQL}$ et $\hat{\theta}_n^{GQL}$ pour les processus GARCH(1, 1).

2.4.3 Processus APARCH

Le modèle APARCH a été introduit (voir Ding et al.(1993) [19]) comme solution de système d'équations :

$$\begin{cases} X_t = \sigma_t \zeta_t \\ \sigma_t^\delta = \omega + \sum_{i=1}^p \alpha_i (|X_{t-i}| - \gamma_i X_{t-i})^\delta + \sum_{j=1}^q \beta_j \sigma_{t-j}^\delta \end{cases} \quad (2.4.5)$$

où $\delta \geq 1$, $\omega > 0$, $-1 < \gamma_i < 1$ et $\alpha_i \geq 0$ pour $i = 1, \dots, p$, $\beta_j \geq 0$ pour $j = 1, \dots, q$ avec α_p, β_q strictement positive et $\sum_{j=1}^q \beta_j < 1$. On note ici

$$\theta = (\delta, \omega, \alpha_1, \dots, \alpha_p, \gamma_1, \dots, \gamma_p, \beta_1, \dots, \beta_q).$$

Utilisant L l'opérateur retard habituel tel que $LX_t = X_{t-1}$ et $(1 - \sum_{j=1}^q \beta_j L^j)^{-1}$ existe, un simple calcul implique que pour $t \in \mathbb{Z}$:

$$\begin{aligned} \sigma_t^\delta &= \left(1 - \sum_{j=1}^q \beta_j L^j\right)^{-1} \\ &\times \left[\omega + \sum_{i=1}^p \alpha_i (1 - \gamma_i)^\delta (\max(X_{t-i}, 0))^\delta + \alpha_i (1 + \gamma_i)^\delta (-\min(X_{t-i}, 0))^\delta \right] \\ &= b_0 + \sum_{i \geq 1} b_i^+ (\max(X_{t-i}, 0))^\delta + \sum_{i \geq 1} b_i^- (\max(-X_{t-i}, 0))^\delta. \end{aligned}$$

où $b_0 = w(1 - \sum_{j=1}^q \beta_j)^{-1}$ et les coefficients $(b_i^+, b_i^-)_{i \geq 1}$ sont définis par les relations de récurrence :

$$\begin{cases} b_i^+ = \sum_{k=1}^q \beta_k b_{i-k}^+ + \alpha_i (1 - \gamma_i)^\delta & \text{with } \alpha_i (1 - \gamma_i) = 0 \text{ for } i > p; \\ b_i^- = \sum_{k=1}^q \beta_k b_{i-k}^- + \alpha_i (1 + \gamma_i)^\delta & \text{with } \alpha_i (1 + \gamma_i) = 0 \text{ for } i > p, \end{cases} \quad (2.4.6)$$

avec $b_i^+ = b_i^- = 0$ pour $i \leq 0$. Donc, pour le modèle APARCH, $f_\theta^t \equiv 0$ et $M_\theta^t = \sigma_t$. C'est clair que $\alpha_j^{(0)}(f, \Theta) = 0$ et un simple calcul implique que $\alpha_j^{(0)}(M, \Theta) = \sup_{\theta \in \Theta} \max(|b_j^+(\theta)|^{1/\delta}, |b_j^-(\theta)|^{1/\delta})$, donc $A_0(f, \Theta)$ est vérifiée et $\sum_{j=1}^q \beta_j < 1$ implique que pour n assez grand, la suite définie par $u_n = \sum_{k=1}^q \beta_k u_{n-k}$ est décroissante de façon exponentielle et donc $A_0(M, \Theta)$ demeure vérifiée. D'où pour $r \geq 1$, l'ensemble de stationnarité $\Theta(r)$ est définie par :

$$\Theta_3(r) = \left\{ \theta \in \mathbb{R}^{2p+q+2} \mid \left(\mathbb{E}[|\zeta_0|^r] \right)^{1/r} \sum_{j=1}^{\infty} \max(|b_j^+|^{1/\delta}, |b_j^-|^{1/\delta}) < 1 \right\}. \quad (2.4.7)$$

Par conséquent, nous pouvons énoncer, pour la première fois, la forte consistance et la normalité asymptotique d'EQMV-Laplace pour les modèles APARCH comme suit :

Proposition 4 *Soit X une solution stationnaire de (2.4.5) avec $\theta_0 \in \Theta$ où Θ est le sous ensemble compact de $\Theta_3(r)$ définie dans (2.4.7). Alors,*

1. Si $r = 1$, alors $\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{a.s.} \theta_0$.

2. Si $r = 2$, et si Γ_M définie dans (2.3.22) est une matrice symétrique définie positive, alors :

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}_{2p+q+2}(0, (\sigma_\zeta^2 - 1) \Gamma_M^{-1}).$$

Application numérique pour l'APARCH

Dans le tableau 2.3, nous exposons les expériences de Monte-Carlo pour le processus APARCH(1, 1) définie par $X_t = \sigma_t \zeta_t$ où $\varepsilon_t = \left(\alpha_0 + \alpha_1 (|\varepsilon_{t-1}| - \gamma \varepsilon_{t-1})^\delta + \beta \zeta_{t-1}^\delta \right)^{1/\delta} \zeta_t$ avec $\alpha_1 = 0.4$, $\gamma = 0.5$, $\beta = 0.1$ et $\delta = 1.4$.

		$\mathcal{L}$		$\mathcal{N}$		t_3		$\mathcal{U}$		$\mathcal{M}$	
	n	$\hat{\theta}_n^{GQL}$	$\hat{\theta}_n^{LQL}$	$\hat{\theta}_n^{GQL}$	$\hat{\theta}_n^{LQL}$	$\hat{\theta}_n^{GQL}$	$\hat{\theta}_n^{LQL}$	$\hat{\theta}_n^{GQL}$	$\hat{\theta}_n^{LQL}$	$\hat{\theta}_n^{GQL}$	$\hat{\theta}_n^{LQL}$
ω	100	0.215	0.213	0.181	0.186	0.244	0.194	0.168	0.119	0.261	0.286
	1000	0.070	0.061	0.052	0.052	0.108	0.068	0.041	0.022	0.192	0.191
	5000	0.030	0.027	0.022	0.022	0.065	0.041	0.017	0.013	0.194	0.19
α	100	0.264	0.257	0.161	0.162	0.229	0.206	0.130	0.097	0.346	0.339
	1000	0.062	0.058	0.041	0.041	0.119	0.090	0.030	0.027	0.341	0.342
	5000	0.027	0.024	0.019	0.019	0.075	0.065	0.012	0.022	0.372	0.372
γ	100	0.510	0.496	0.313	0.308	0.469	0.438	0.246	0.181	0.691	0.733
	1000	0.093	0.086	0.068	0.068	0.195	0.105	0.054	0.026	0.341	0.344
	5000	0.040	0.036	0.030	0.030	0.105	0.045	0.024	0.011	0.288	0.313
β	100	0.425	0.381	0.237	0.231	0.370	0.285	0.165	0.115	0.622	0.541
	1000	0.067	0.062	0.053	0.053	0.158	0.071	0.039	0.021	0.266	0.266
	5000	0.032	0.029	0.023	0.023	0.107	0.033	0.018	0.009	0.200	0.218
δ	100	0.879	0.825	0.781	0.789	0.895	0.797	0.681	0.553	0.989	0.989
	1000	0.399	0.372	0.309	0.310	0.540	0.382	0.240	0.129	0.633	0.656
	5000	0.189	0.169	0.137	0.138	0.379	0.204	0.103	0.057	0.587	0.587

TABLE 2.3 – Racine d'erreur moyenne quadratique $\hat{\theta}_n^{LQL}$ et $\hat{\theta}_n^{GQL}$ pour le processus APARCH(1, 1).

2.4.4 Processus ARMA-GARCH

Les processus ARMA(p, q)-GARCH(p', q') ont été introduit par Ding et al. (1993) [19], Ling et McAleer (2003) [43] comme solution de système d'équations :

$$\begin{cases} P_\theta(L) X_t = Q_\theta(L) \varepsilon_t, \\ \varepsilon_t = \sigma_t \zeta_t, \quad \text{avec} \quad \sigma_t^2 = c_0 + \sum_{i=1}^{p'} c_i \varepsilon_{t-i}^2 + \sum_{i=1}^{q'} d_i \sigma_{t-i}^2 \end{cases} \quad (2.4.8)$$

où :

- $c_0 > 0$, $c_i \geq 0$ pour $i = 1, \dots, p'$, $d_i \geq 0$ pour $i = 1, \dots, q'$, $\sum_{i=1}^{q'} d_i < 1$ et $c_{p'}, d_{q'}$ positive ;
- $P_\theta(x) = 1 - a_1 x - \dots - a_p x^p$ et $Q_\theta(x) = 1 - b_1 x - \dots - b_q x^q$ sont des polynômes coprimés avec $\sum_{i=1}^p |a_i| < 1$ et $\sum_{i=1}^q |b_i| < 1$.

Soit $\theta = (c_0, c_1, \dots, c_{p'}, d_1, \dots, d_{q'}, a_1, \dots, a_p, b_1, \dots, b_q)$. On va utiliser le lemme 1. Comme (ε_t) est un GARCH(p', q'), alors $f_\theta^\varepsilon = 0$ et $M_\theta^\varepsilon = ((1 - \sum_{j=1}^{q'} d_j L^j)^{-1} (c_0 + c_1 \varepsilon_{t-1}^2 + \dots + c_{p'} \varepsilon_{t-p'}^2))^{1/2}$ et un calcul direct implique que les coefficients lipchitsiennes de (ε_t) sont tels que $\alpha_j^{(0)}(f^\varepsilon, \{\theta_0\}) = 0$ et $\alpha_j^{(0)}(M^\varepsilon, \{\theta_0\}) = |\beta_j|$ avec $(1 + \sum_{j=1}^\infty \beta_j x^j)(1 - \sum_{j=1}^{q'} d_j x^j) = \sum_{j=0}^{p'} c_j x^j$. Donc $(\alpha_j^{(0)}(f^\varepsilon, \{\theta_0\}))_j$ et $(\alpha_j^{(0)}(M^\varepsilon, \{\theta_0\}))_j$ sont SDE (voir Berkes et Horváth (2004). [6]). Ainsi $(A_0(f^\varepsilon, \{\theta_0\}))$ et $(A_0(M^\varepsilon, \{\theta_0\}))$ demeurent vérifier. Considérant la partie ARMA et notant (ψ_j) telque $(1 + \sum_{j=1}^\infty \psi_j x^j)(1 - \sum_{j=1}^\infty a_j x^j) = (1 - \sum_{j=1}^\infty b_j x^j)$, à partir de Lemme 1 on déduit que :

$$\begin{cases} \alpha_j^{(0)}(f, \{\theta_0\}) &= |\psi_j| \\ \alpha_j^{(0)}(M, \{\theta_0\}) &\leq \sum_{k=1}^j |\psi_k| \times |\beta_{j-k}| \end{cases}.$$

et donc $(\alpha_j^{(0)}(f, \{\theta_0\}))_j$ et $(\alpha_j^{(0)}(M, \{\theta_0\}))_j$ sont SED, $(A_0(f, \{\theta_0\}))$ et $(A_0(M, \{\theta_0\}))$ sont vérifiées et X est une solution stationnaire de (2.1.1) pour θ inclut dans le domaine de r -stationnarité $\Theta(r)$ défini par :

$$\Theta_4(r) = \left\{ \theta \in \mathbb{R}^{p+q+p'+q'+1} \middle/ \sum_{i=1}^\infty |\psi_i(\theta)| + (E[|\zeta_0|^r])^{1/r} \sum_{j=1}^\infty \sum_{k=1}^j |\psi_k| \times |\beta_{j-k}| < 1 \right\}. \quad (2.4.9)$$

On peut établir maintenant la consistance forte et la normalité asymptotique d'EQMV-Laplace pour les processus ARMA-GARCH.

		$\mathcal{L}$		$\mathcal{N}$		t_3		$\mathcal{U}$		$\mathcal{M}$	
	n	$\widehat{\theta}_n^{GQL}$	$\widehat{\theta}_n^{LQL}$	$\widehat{\theta}_n^{GQL}$	$\widehat{\theta}_n^{LQL}$	$\widehat{\theta}_n^{GQL}$	$\widehat{\theta}_n^{LQL}$	$\widehat{\theta}_n^{GQL}$	$\widehat{\theta}_n^{LQL}$	$\widehat{\theta}_n^{GQL}$	$\widehat{\theta}_n^{LQL}$
θ	100	0.120	0.097	0.107	0.107	0.121	0.098	0.097	0.067	0.123	0.087
	1000	0.035	0.024	0.028	0.028	0.048	0.029	0.024	0.015	0.035	0.026
	5000	0.016	0.010	0.012	0.012	0.023	0.012	0.015	0.011	0.011	0.007
ϕ	100	0.135	0.109	0.117	0.119	0.141	0.116	0.110	0.077	0.132	0.102
	1000	0.044	0.030	0.033	0.033	0.063	0.035	0.029	0.023	0.053	0.046
	5000	0.020	0.014	0.015	0.015	0.029	0.015	0.013	0.012	0.019	0.014
α_0	100	0.104	0.096	0.085	0.084	0.158	0.129	0.073	0.055	0.131	0.116
	1000	0.031	0.028	0.025	0.025	0.241	0.060	0.021	0.019	0.053	0.046
	5000	0.014	0.012	0.010	0.010	0.052	0.042	0.009	0.016	0.036	0.019
α_1	100	0.179	0.177	0.166	0.167	0.469	0.385	0.134	0.107	0.494	0.405
	1000	0.064	0.060	0.045	0.045	0.328	0.161	0.031	0.046	0.160	0.137
	5000	0.031	0.027	0.020	0.020	0.182	0.096	0.013	0.038	0.090	0.062
β	100	0.302	0.269	0.252	0.233	0.604	0.497	0.217	0.164	0.553	0.472
	1000	0.057	0.051	0.051	0.051	0.312	0.187	0.045	0.049	0.165	0.170
	5000	0.025	0.022	0.020	0.020	0.199	0.073	0.062	0.066	0.019	0.025

TABLE 2.4 – Racine d’erreur moyenne quadratique $\widehat{\theta}_n^{LQL}$ et $\widehat{\theta}_n^{GQL}$ pour les processus ARMA(1, 1)-GARCH(1, 1).

Proposition 5 *Soit X la solution stationnaire de (2.4.8) avec $\theta_0 \in \Theta$ où Θ est le sous ensemble compact de $\Theta(r)$ définie dans (2.4.9). Alors,*

1. *Si $r = 1$, alors $\widehat{\theta}_n \xrightarrow[n \rightarrow \infty]{a.s.} \theta_0$.*
2. *Si $r = 2$, et si Γ_f et Γ_M définie dans (2.3.22) sont des matrices symétriques définies positives, alors la normalité asymptotique (2.3.10) de $\widehat{\theta}_n$ demeure vérifier.*

Application numérique pour l’ARMA-GARCH

Dans le tableau 2.4 ci-dessous, nous donnons les expériences de Monte-Carlo pour le processus ARMA(1, 1)-GARCH(1, 1) défini par $X_t = \phi X_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$ où $\varepsilon_t = \sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta \zeta_{t-1}^2} \zeta_t$, $\phi = 0.4$, $\theta = 0.6$, $\alpha_0 = 0.2$, $\alpha_1 = 0.4$ et $\beta = 0.1$.

2.4.5 Processus ARMA-ARCH(∞).

Les processus ARMA(p, q)-ARCH(∞) sont l'extension naturelle des processus ARMA-GARCH. Ce sont la solution de système d'équations :

$$\begin{cases} P_\theta(L) X_t = Q_\theta(L) \varepsilon_t, \\ \varepsilon_t = \sigma_t \zeta_t, \quad \text{avec} \quad \sigma_t^2 = c_0 + \sum_{i=1}^{\infty} c_i \varepsilon_{t-i}^2 \end{cases} \quad (2.4.10)$$

où :

- $c_0 > 0$, $c_i \geq 0$ pour $i \geq 1$;
- $P_\theta(x) = 1 - a_1x - \dots - a_px^p$ et $Q_\theta(x) = 1 - b_1x - \dots - b_qx^q$ sont des polynômes coprimés $\sum_{i=1}^p |a_i| < 1$ et $\sum_{i=1}^q |b_i| < 1$.

Le processus ARCH(∞) est introduit par Robinson (1991) [62] et les propriétés asymptotiques d'EQMV-Gaussien sont étudiées dans Robinson et Zaffaroni (2006) [63], Straumann et Mikosch (2006)[68] ou Bardet et Wintenberger (2009) [2]. Par conséquent, nous supposons qu'il existe un $\beta = (\beta_1, \dots, \beta_m)$ tel que pour tout $i \in \mathbb{N}$, $c_i = c(i, \beta)$, où $c(\cdot)$ est une fonction connue. Soit $\theta = (\beta, a_1, \dots, a_p, b_1, \dots, b_q)$. On va utiliser le Lemme 1. Comme (ε_t) est supposé être un ARCH(∞), alors $f_\theta^\varepsilon = 0$ et $M_\theta^\varepsilon = (c(0, \beta) + \sum_{i=1}^{\infty} c(i, \beta) \varepsilon_{t-i}^2)^{1/2}$ et un calcul direct implique que les coefficients de Lipschitz (ε_t) sont tels que $\alpha_j^{(0)}(f^\varepsilon, \{\theta_0\}) = 0$ et $\alpha_j^{(0)}(M^\varepsilon, \{\theta_0\}) = c(j, \beta_0)$. Donc, nous supposons qu'il existe un $\ell > 1$ tel que :

$$c(j, \beta_0) = \mathcal{O}(j^{-\ell}) \text{ quand } j \rightarrow \infty. \quad (2.4.11)$$

ainsi $(A_0(f^\varepsilon, \{\theta_0\}))$ et $(A_0(M^\varepsilon, \{\theta_0\}))$ sont satisfaites.

En Considérant la partie ARMA et en notant (ψ_j) tel que $(1 + \sum_{j=1}^{\infty} \psi_j x^j)(1 - \sum_{j=1}^{\infty} a_j x^j) = (1 - \sum_{j=1}^{\infty} b_j x^j)$, alors à partir de lemme 1, nous déduisons que :

$$\begin{cases} \alpha_j^{(0)}(f, \{\theta_0\}) &= |\psi_j| \\ \alpha_j^{(0)}(M, \{\theta_0\}) &\leq \sum_{k=1}^j |\psi_k| \times c(j, \beta_0) \end{cases}.$$

Donc $(\alpha_j^{(0)}(f, \{\theta_0\}))_j$ est SDE et $(\alpha_j^{(0)}(M, \{\theta_0\}))_j = \mathcal{O}(j^{-\ell})$, d'où $(A_0(f, \{\theta_0\}))$ et $(A_0(M, \{\theta_0\}))$ sont vérifiées, et X est p.s. une solution de

(2.1.1) pour θ inclut dans l'ensemble de r-stationnarité $\Theta(r)$ définie par :

$$\Theta_5(r) = \left\{ \theta \in \mathbb{R}^{p+q+m} \mid \sum_{i=1}^{\infty} |\psi_i(\theta)| + (E[|\zeta_0|^r])^{1/r} \sum_{j=1}^{\infty} \sum_{k=1}^j |\psi_k| \times c(j, \beta_0) < 1 \right\}. \quad (2.4.12)$$

On peut établir maintenant la forte consistance et la normalité asymptotique d'EQMV-Laplace pour le processus ARMA-ARCH(∞).

Proposition 6 *Soit X la solution stationnaire de (2.4.10) où (2.4.11) est vérifiée et $\theta_0 \in \Theta$ le sous ensemble compact de $\Theta_5(r)$ définie dans (2.4.12). Alors,*

1. *Si $r \geq 1$ et $\ell \geq 2/\min(r, 2)$, alors $\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{a.s.} \theta_0$.*
2. *Si $r = 2$, $\ell > 1$ et si $\partial_{\beta}^i c(j, \beta) = \mathcal{O}(j^{-\ell})$ pour $i = 1, 2$, et si Γ_f et Γ_M définie dans (2.3.22) sont des matrices symétriques définies positives, alors la normalité asymptotique (2.3.10) de $\hat{\theta}_n$ demeure vérifier.*

Notons que le résultat obtenu par la Proposition 6 est entièrement nouveau.

Notons que $\ell > 1$ et $r = 2$ sont nécessaires pour la normalité asymptotique d'EQMV-Laplace tandis $r = 4$ et $\ell > 2$ sont requises dans le cas d'EQMV-Gaussien pour un tel processus (voir [2]). Cela donne un net avantage d'EQMV-Laplace.

2.4.6 Processus ARMA-APARCH

Le processus ARMA(p, q)-APARCH(p', q') a été aussi introduit par Ding et al. (1993) [19] comme la solution du système d'équations :

$$\begin{cases} P_{\theta}(L) X_t = Q_{\theta}(L) \varepsilon_t, \\ \varepsilon_t = \sigma_t \zeta_t, \text{ avec } \sigma_t^{\delta} = \omega + \sum_{i=1}^{p'} \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^{\delta} + \sum_{j=1}^{q'} \beta_j \sigma_{t-j}^{\delta} \end{cases} \quad (2.4.13)$$

où :

- $\delta \geq 1$, $\omega > 0$, $-1 < \gamma_i < 1$ et $\alpha_i \geq 0$ for $i = 1, \dots, p' - 1$, $\beta_j \geq 0$ pour $j = 1, \dots, q' - 1$, $\alpha_{p'}$, $\beta_{q'}$ des nombres réels strictement positifs et $\sum_{j=1}^{p'} \alpha_j < 1$;

- $P_\theta(x) = 1 - a_1x - \dots - a_px^p$ et $\Psi_\theta(x) = 1 - b_1x - \dots - b_qx^q$ sont des polynômes coprimés avec $\sum_{i=1}^p |a_i| < 1$ et $\sum_{i=1}^q |b_i| < 1$.

Soit $\theta = (\delta, \omega, \alpha_1, \dots, \alpha_{p'}, \gamma_1, \dots, \gamma_{p'}, \beta_1, \dots, \beta_{q'}, a_1, \dots, a_p, b_1, \dots, b_q)$. Alors, comme pour le processus ARMA-GARCH, on utilise le Lemme 1. Grace aux calculs faites pour le processus APARCH, on obtient $\alpha_j^{(0)}(f^\varepsilon, \{\theta_0\}) = 0$ et $\alpha_j^{(0)}(M^\varepsilon, \{\theta_0\}) = \max(|b_j^+|^{1/\delta}, |b_j^-|^{1/\delta})$ avec $(b_i^+, b_i^-)_{i \geq 1}$ définie dans (2.4.6). Alors, on a :

$$\begin{cases} \alpha_j^{(0)}(f, \{\theta_0\}) & \leq |\psi_j| \\ \alpha_j^{(0)}(M, \{\theta_0\}) & \leq \sum_{k=1}^j |\psi_k| \times \max(|b_{j-k}^+|^{1/\delta}, |b_{j-k}^-|^{1/\delta}) \end{cases} .$$

où les (ψ_j) vérifient $(1 + \sum_{j=1}^\infty \psi_j x^j)(1 - \sum_{j=1}^\infty a_j x^j) = (1 - \sum_{j=1}^\infty b_j x^j)$. À partir de lemme 1, $(A_0(f, \Theta))$ et $(A_0(M, \Theta))$ restent vérifier tant que $(\alpha_j^{(0)}(f^\varepsilon, \{\theta_0\}))_j = 0$ et $(\alpha_j^{(0)}(M^\varepsilon, \{\theta_0\}))_j$ sont des SED. Donc, pour $r \geq 1$, l'ensemble de stationnarité $\Theta_6(r)$ est :

$$\begin{aligned} \Theta_6(r) = \Big\{ \theta \in \mathbb{R}^{p+q+p'+q'+2} \Big/ \sum_{j=1}^\infty |\psi_j| \\ + \left(\mathbb{E}|\zeta_0|^r \right)^{1/r} \sum_{j=1}^\infty \sum_{k=1}^j |\psi_k| \times \max(|b_{j-k}^+|^{1/\delta}, |b_{j-k}^-|^{1/\delta}) < 1 \Big\}. \end{aligned}$$

Maintenant, on est en mesure de donner les résultats asymptotiques pour l'EQMV-Laplace de l'ARMA-APARCH.

Proposition 7 *Soit X une solution stationnaire de (2.4.13) avec $\theta_0 \in \Theta$ où Θ est le sous ensemble compact de $\Theta_6(r)$ définie dans (2.4.9). Alors,*

1. Si $r = 1$, alors $\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{a.s.} \theta_0$.
2. Si $r = 2$, et si Γ_f et Γ_M défini dans (2.3.22) sont des matrices symétriques définies positives, alors la normalité asymptotique (2.3.10) de $\hat{\theta}_n$ demeure.

Ce résultat est donné pour la première fois pour l'EQMV-Laplace. Le cas d'EQMV-Gauss pour ARMA-APARCH pourrait également être obtenu en suivant la décomposition précédente et le papier de Bardet et Wintenberger

(2009) [2]. Une fois de plus, la normalité asymptotique d'EQMV-Laplace exige seulement que $r = 2$, alors que cela nécessite $r = 4$ pour l'EQMV-Gaussien.

Application numérique pour l'ARMA-APARCH

Dans le tableau 2.5, nous donnons les expériences de Monte-Carlo pour le processus ARMA(1, 1)-APARCH(1, 1) définie par $X_t = \phi X_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$ où $\varepsilon_t = \left(\alpha_0 + \alpha_1 (|\varepsilon_{t-1}| - \gamma \varepsilon_{t-1})^\delta + \beta \zeta_{t-1}^\delta \right)^{1/\delta} \zeta_t$ et $\phi = 0.4$, $\theta = 0.6$, $\alpha_0 = 0.2$, $\alpha_1 = 0.4$, $\gamma = 0.5$, $\beta = 0.1$ et $\delta = 1.2$.

Conclusion des résultats numériques : D'une part, il est clair que pour tout les modèles, le RMSE diminue quand la taille de l'échantillon augmente, ce qui valide les résultats théoriques (consistance de l'estimateur). D'autre part, les tableaux 2.1, 2.2, 2.3 et 2.4 montrent que l'EQMV Laplacien donne une estimation plus précise que l'EQMV Gaussien pour plusieurs types de bruit, sauf bien sûr dans le cas d'une distribution gaussienne (même dans ce cas le RSME des deux estimateurs sont presque identiques).

		$\mathcal{L}$		$\mathcal{N}$		t_3		$\mathcal{U}$		$\mathcal{M}$	
	n	$\widehat{\theta}_n^{GQL}$	$\widehat{\theta}_n^{LQL}$	$\widehat{\theta}_n^{GQL}$	$\widehat{\theta}_n^{LQL}$	$\widehat{\theta}_n^{GQL}$	$\widehat{\theta}_n^{LQL}$	$\widehat{\theta}_n^{GQL}$	$\widehat{\theta}_n^{LQL}$	$\widehat{\theta}_n^{GQL}$	$\widehat{\theta}_n^{LQL}$
θ	100	0.105	0.081	0.103	0.101	0.123	0.103	0.093	0.070	0.122	0.090
	1000	0.030	0.020	0.023	0.023	0.048	0.027	0.021	0.014	0.035	0.025
	5000	0.013	0.008	0.010	0.010	0.028	0.011	0.009	0.007	0.015	0.010
ϕ	100	0.135	0.105	0.125	0.124	0.152	0.121	0.118	0.093	0.144	0.107
	1000	0.040	0.026	0.030	0.030	0.058	0.034	0.025	0.026	0.045	0.029
	5000	0.017	0.011	0.013	0.013	0.035	0.014	0.011	0.014	0.019	0.013
ω	100	0.195	0.206	0.195	0.209	0.263	0.262	0.233	0.172	0.258	0.246
	1000	0.076	0.064	0.062	0.061	0.163	0.116	0.049	0.066	0.125	0.111
	5000	0.034	0.029	0.025	0.025	0.117	0.051	0.019	0.041	0.056	0.051
α	100	0.207	0.209	0.174	0.181	0.374	0.328	0.165	0.121	0.372	0.345
	1000	0.054	0.049	0.039	0.039	0.268	0.135	0.029	0.035	0.128	0.121
	5000	0.023	0.020	0.016	0.016	0.154	0.060	0.012	0.019	0.060	0.052
γ	100	0.441	0.424	0.337	0.353	0.615	0.584	0.295	0.230	0.598	0.595
	1000	0.102	0.090	0.071	0.071	0.313	0.191	0.054	0.040	0.171	0.172
	5000	0.047	0.041	0.032	0.032	0.190	0.074	0.024	0.017	0.069	0.073
β	100	0.324	0.309	0.251	0.261	0.507	0.452	0.189	0.153	0.447	0.436
	1000	0.080	0.073	0.062	0.061	0.271	0.153	0.044	0.033	0.128	0.136
	5000	0.037	0.033	0.027	0.027	0.191	0.064	0.019	0.014	0.056	0.058
δ	100	0.705	0.745	0.706	0.724	0.823	0.895	0.643	0.619	0.815	0.880
	1000	0.282	0.236	0.223	0.215	0.630	0.430	0.192	0.356	0.394	0.400
	5000	0.115	0.099	0.089	0.087	0.448	0.160	0.074	0.200	0.151	0.154

TABLE 2.5 – Racine d’erreur moyenne quadratique $\widehat{\theta}_n^{LQL}$ et $\widehat{\theta}_n^{GQL}$ pour les processus ARMA(1, 1)-APARCH(1, 1).

Chapitre 3

Approximation de l'estimateur des paramètres de l'ARMA

3.1 Introduction

Jusqu'à présent, nous n'avons pas de forme analytique pour l'estimateur des paramètres de processus autorégressif moyen mobile (ARMA), et en particulier ceux obtenus par maximisation de la fonction de vraisemblance. Par conséquent, les chercheurs appliquent l'optimisation itérative de gradient ou directement la recherche routines. La méthode de Gradient peut échouer à trouver le maximum en raison de la nature hautement non-linéaire de la fonction de vraisemblance et en général les méthodes directes de recherche nécessitent un grand nombre d'évaluations de la fonction, parce qu'elles n'utilisent pas l'information à propos de son inclinaison et sa courbure.

Dans ce sens et pour le développement de la fonction de vraisemblance plusieurs travaux ont été réalisés, Whittle [74] (1951) et Durbin [21] (1959) ont développé des approximations aux modèles univariés MA. Godolphin [34] (1984) a étendu la méthode de Whittle aux modèles ARMA univariés et ils ont proposé une méthode pour calculer les estimations des paramètres directement à partir de l'échantillon des autocorrélations des données observées. Durbin [21] (1959) a proposé un algorithme dans lequel les coefficients MA sont calculés à partir des coefficients d'une autorégression à long ordre. Hannan et Rissanen [36] (1982) et Koreisha et Pukkila [39] (1989) ont étendu la méthode de Durbin aux modèles ARMA univariés et multivariés, respectivement. Tunnicliffe Wilson [70](1973), Reinsel et al.

[60] (1992) et DeFrutos et Serrano [16] (1997) ont proposé des méthodes d'estimation approximative similaires à la maximisation de vraisemblance (MV).

Ces méthodes de forme fermée sont asymptotiquement équivalentes à MV, vu leur erreur d'approximation qui devient négligeable dans le cas des grands échantillons. Toutefois, ils peuvent s'écarter sensiblement de l'estimateur de MV dans des échantillons de petites tailles ou pour des modèles avec des paramètres près de la frontière de stationnarité ou inversibilité.

En raison de la non-linéarité de la fonction de vraisemblance, nous n'avons pas de solution analytique pour l'estimateur de son maximum, c'est d'ailleurs la raison pour laquelle de nombreux algorithmes ont été fournis de façon indépendante par AI McLeod, PR Holanda Sales [50] (1984), Solo [69] (1983), Dugré et al. [20] (1980), Shea [65] (1988), Mittnik [52] (1990), Mauricio [48] (2002) et d'autres. Les approches utilisées dans les travaux antérieurs afin de maximiser la fonction de vraisemblance sont des méthodes itératives qui fournissent des approximations de maximum. Dans ce travail, nous allons approcher la fonction de vraisemblance par une forme quadratique, puis calculer son maximum (c'est-à-dire en approche la fonction au lieu d'approcher son maximum).

3.2 Détermination de la fonction de vraisemblance

3.2.1 Fonction de vraisemblance de l'ARMA(p, q)

Le modèle ARMA est un processus stochastique décrivant les phénomènes aléatoires variant avec le temps, il a la formule mathématique :

$$X_t + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}. \quad (3.2.1)$$

X_t est supposé de tendance nulle (centrée), sinon il sera remplacé par $\hat{X}_t = X_t - \bar{X}$, les ε_t sont des variables aléatoires Gaussiennes indépendantes et identiquement distribuées (VAIID) de moyenne nulle et de variance σ_ε^2 (c'est-à-dire $\varepsilon \approx N(0, \sigma_\varepsilon^2)$).

La détermination de modèle revient à l'estimation de ces paramètres inconnus. La méthode la plus utilisée est la maximisation de la fonction de vrai-

semblance. Sous l'hypothèse de la normalité, la fonction de vraisemblance d'un modèle ARMA pour un échantillon de taille n est :

$$f(\Phi, \Theta, \sigma_\varepsilon^2) = \frac{1}{\sqrt{(2\pi\sigma^2)^n r_0 \dots r_{n-1}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n \frac{\varepsilon_i^2(\Phi, \Theta)}{r_{i-1}}}. \quad (3.2.2)$$

où $\Phi = \phi_1, \dots, \phi_p$, $\Theta = \theta_1, \dots, \theta_q$ et $\sigma_\varepsilon^2 = \sigma^2 r_i$, les r_i sont calculés utilisant l'algorithme d'innovation et σ^2 est la variance de processus.

La maximisation de f par rapport à $\beta = (\Phi, \Theta)$ est équivalente à la minimisation de :

$$l_n(\Phi, \Theta) = \frac{1}{n} \sum_{i=1}^n \ln(r_{i-1}) + \ln\left(n^{-1} \sum_{i=1}^n \frac{\varepsilon_i^2(\Phi, \Theta)}{r_{i-1}}\right). \quad (3.2.3)$$

Cette fonction a une forme non-linéaire ce qui implique l'inexistence d'une solution analytique. Pour le calcul de la forme explicite, nous utilisons :

$$\sum_{i=1}^n \frac{\varepsilon_i^2(\Phi, \Theta)}{r_{i-1}} = [(L_\Theta^{-1} (L_\Phi X - F e_*)) A^{-1/2}]^2. \quad (3.2.4)$$

où :

$$L_\Theta = \begin{bmatrix} 1 & \theta_1 & \theta_2 & \dots & \theta_q & 0 & \dots & 0 \\ 0 & 1 & \dots & \dots & \theta_q & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & \dots & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & \theta_q & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & \dots & 1 \end{bmatrix}$$

$$L_\Phi = \begin{bmatrix} 1 & -\phi_1 & -\phi_2 & \dots & -\phi_q & 0 & \dots & 0 \\ 0 & 1 & -\phi_2 & \dots & -\phi_q & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & \dots & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & -\phi_q & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & \dots & 1 \end{bmatrix}$$

$$F = \begin{bmatrix} H_p & M_q \\ 0 & 0 \end{bmatrix} \quad H_p = \begin{bmatrix} \phi_p & \phi_{p-1} & \cdot & \cdot & \phi_1 \\ 0 & \phi_p & \cdot & \cdot & \phi_2 \\ \cdot & 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \phi_{p-1} \\ 0 & 0 & \cdot & \cdot & \phi_p \end{bmatrix}$$

$$M_q = \begin{bmatrix} \theta_q & \theta_{q-1} & \cdot & \cdot & \theta_1 \\ 0 & \theta_q & \cdot & \cdot & \theta_2 \\ \cdot & 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \theta_{q-1} \\ 0 & 0 & \cdot & \cdot & \theta_q \end{bmatrix}$$

$$A = \sigma_\varepsilon^2 \begin{bmatrix} r_0 & 0 & \cdot & 0 \\ 0 & r_1 & \cdot & 0 \\ \cdot & 0 & \cdot & \cdot \\ 0 & 0 & \cdot & r_n \end{bmatrix}, \quad e_* = (X_0, \dots, X_{1-p}, \varepsilon_0, \dots, \varepsilon_{1-q})'.$$

Le vecteur e_* représente le passé de processus et de bruit, dans la pratique on le suppose égal à zéro.

3.2.2 L'algorithme d'innovation pour le modèle ARMA

Afin de simplifier l'algorithme d'innovation détaillé dans [12], on l'applique non pas au processus X_t , mais plutôt au processus transformé [1] définie par :

$$\begin{cases} w_t = \sigma_\varepsilon^{-1} X_{t-1}, & \text{if } t = 1, \dots, \max(p, q) \\ w_t = \sigma_\varepsilon^{-1} (X_t + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p}), & t > \max(p, q). \end{cases}$$

Par conséquent, une fois que les auto-covariances $\gamma(i)$ sont connues, nous calculons les innovations de manière récursive en utilisant l'algorithme des innovations présentées ci-dessous :

$$\begin{cases} r_0 = \frac{\gamma_X(0)}{\sigma_\varepsilon^2} \\ \theta_{m,m-k} = r_k^{-1} \left(\frac{\gamma_X(m-k)}{\sigma_\varepsilon^2} - \sum_{j=0}^{k-1} \theta_{m,m-j} \theta_{n,n-j} r_j \right) \\ r_m = \frac{\gamma_X(0)}{\sigma_\varepsilon^2} - \sum_{j=0}^{m-1} \theta_{m,m-j} r_j. \end{cases}$$

Dans les applications, γ est remplacée par son estimation $\hat{\gamma}$ obtenue à partir de l'échantillon.

Certains cas spéciaux de r_n

$MA(1)$: Si X_t est le processus moyennne mobie MA(1) defini par :

$$X_t = \varepsilon_t + \theta\varepsilon_{t-1},$$

alors, son algorithme d'innovation est donné par :

$$\begin{cases} r_0 = 1 + \theta^2 \\ r_m = 1 + \theta^2 - \theta^2/r_{m-1}. \end{cases} \quad (3.2.5)$$

$AR(1)$: Si X_t est le processus autoregressif AR(1) defini par :

$$X_t = \phi X_{t-1} + \varepsilon_t,$$

alors, l'algorithme d'innovation se simplifie à $r_i = 1$ pour tout $i \geq 0$.

$ARMA(1,1)$: Si X_t est le processus autoregressif moyenne mobile ARMA(1,1) defini par :

$$X_t = \phi X_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1},$$

alors, l'algorithme d'innovation se simplifie à :

$$\begin{cases} r_0 = (1 + 2\phi\theta + \theta^2)/(1 - \phi^2) \\ r_m = 1 + \theta^2 - \theta^2/r_{m-1}. \end{cases} \quad (3.2.6)$$

3.3 Propriétés de la vraisemblance d'un ARMA

La fonction de vraisemblance d'un processus ARMA possède certaines caractéristiques qui nous ont permis de calculer le maximum différemment. Dans cette section, et à travers la proposition 8 et le lemme 4 nous prouvons l'essentiel de ces caractéristiques.

Proposition 8 (*Propriétés de r_n*) : Soit β le vecteur des paramètres inconnus, alors les $(r_i)_{i \geq 0}$ vérifient les propriétés suivantes :

1. $r_i > 1 ; \forall i = 0, n$
2. $(\frac{1}{n} \sum_{i=1}^n \ln r_{i-1}) \rightarrow 0$ as $n \rightarrow \infty$

Preuve

Soit X_t un processus ARMA , après l'écriture de sa formule mathématique sous la forme inverse $X_t + \sum_{j=1}^{\infty} \pi_j X_{t+1-j} = \varepsilon_t$ et utilisant que le bruit ε_t est une suite de variables aléatoires indépendantes et identiquement distribuées. Nous trouvons alors,

$$\begin{aligned}
\sigma_\varepsilon^2 = \text{Var}(\varepsilon_{t+1}) &= E(X_{t+1} + \sum_{j=1}^{\infty} \pi_j X_{t+1-j})^2 \\
&\leq \sigma_\varepsilon^2 r_t \leq \sigma_\varepsilon^2 E(X_{t+1} + \sum_{j=1}^t \pi_j X_{t+1-j})^2 \\
&= \sigma_\varepsilon^2 E(\varepsilon_{t+1} - \sum_{j=t+1}^{\infty} \pi_j X_{t+1-j})^2 \\
&= \sigma_\varepsilon^2 \left(1 + E(\sum_{j=t+1}^{\infty} \pi_j X_{t+1-j})^2 \right) \\
&\leq \sigma_\varepsilon^2 \left(1 + (\sum_{j=t+1}^{\infty} |\pi_j|) \gamma_X(0) \right).
\end{aligned}$$

De la première inégalité on déduit que $r_{i-1} \geq 1$ et $r_0 \geq 1$ d'où $\prod_{i=1}^n r_i > 1$. De plus, tant que $(\sum_{j=t+1}^{\infty} |\pi_j|) \gamma_X(0) \rightarrow 0$ et $r_t \rightarrow 1$ quand $t \rightarrow \infty$, alors par la convergence de Cesaro nous avons :

$$0 \leq \ln(\prod_{i=1}^n r_{i-1})^{\frac{1}{n}} = \left(\frac{1}{n} \sum_{i=1}^n \ln r_{i-1} \right) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Lemme 4 : Pour $\beta \in \mathcal{R}^{p+q}$, la fonction de vraisemblance $l_n(\beta)$ est presque sûrement égale à un polynôme d'ordre deux.

Preuve

La propriété (1) de la proposition 8 implique que $(\frac{1}{n} \sum_{i=1}^n \ln r_i)$ est bornée par un réel positif A .

1. Utilisant les relations (1), (2) et l'inégalité $\ln(x) \leq 1 + x$ nous avons

$$l(\beta) \leq A + \sum_{i=1}^n \varepsilon_i^2 = A + [(L_{\Theta}^{-1}(L_{\Phi}X - F\hat{e}_*))]^2$$

qui est un polynôme d'ordre deux, donc intégrable.

2. Comme la suite r_n converge vers 1 et $\frac{1}{n} \sum_{i=1}^n \ln r_i$ vers zéro quand $n \rightarrow \infty$, on conclut que $l_n(\beta)$ converge vers un polynôme d'ordre deux.

Par le théorème de convergence monotone, les résultats (1) et (2) ci-dessus, impliquent que la fonction de log-vraisemblance l_n est presque sûrement égale à un polynôme d'ordre deux. Nous allons donc l'approcher par un polynôme d'ordre deux et le minimum de ce dernier sera une approximation pour le minimum de l_n .

3.4 Procédure de calcul

Pour calculer le maximum de vraisemblance, nous suivons les étapes suivantes :

1. Calcul de l'estimation préliminaire pour les paramètres : pour fournir ces valeurs initiales, trois algorithmes sont disponibles à la section 3.5.
2. Estimation des intervalles de confiance $\prod_{i=1}^{p+q} [a_i, b_i]$ utilisant les estimations préliminaires : tout d'abord, nous calculons les intervalles de confiance dans le cadre univarié pour chaque paramètre. Ensuite, nous prenons le produit de ces intervalles.
3. Génération de $p + q$ échantillons de m éléments $[(\theta_{ij}), l_n(\theta_{ij})]_{j=1, m}$ à partir des intervalles de confiance des paramètres.
4. Approximation numérique par des polynômes de degré deux : nous appliquons la méthode des moindres carrés.
5. Calcul du maximum.

3.5 Valeurs préliminaires

Dans cette section, nous allons examiner quatre techniques d'estimation préliminaires pour les paramètres $(\phi_1, \dots, \phi_p)$, $(\theta_1, \dots, \theta_p)$ à partir des observations $x_1, \dots, x_n$ de processus causal ARMA(p, q). Les procédures de Yule-Walker et Burg s'appliquent aux modèles autorégressifs purs (bien que nous pouvons les adapter aux modèles avec $q > 0$, sauf qu'elles seront moins efficaces que lorsque $q = 0$).

Pour les modèles autorégressifs purs, l'algorithme de Burg donne généralement une valeur supérieure de la vraisemblance que celle obtenue par les équations de Yule-Walker.

Pour les modèles à moyenne mobile purs, l'algorithme d'innovation donne fréquemment une valeur légèrement plus élevée que l'algorithme de Hannan-Rissanen.

Pour les modèles mixtes (c'est-à-dire, celles avec $p > 0$ et $q > 0$) l'algorithme de Hannan-Rissanen est habituellement le meilleur pour trouver des modèles causaux (qui sont requis pour l'initialisation de la maximisation de vraisemblance).

3.5.1 Méthode de Yule-Walker

Le modèle AR(p) a la représentation suivante :

$$X_t + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} = \varepsilon_t. \quad (3.5.1)$$

Pour déterminer l'estimateur de Yule-Walker :

1. Nous multiplions chaque côté de l'équation (3.5.1) par $(X_{t-j})_{j=0:p}$, nous prenons l'espérance et à l'aide de la forme causal de l'AR(p) nous évaluons le membre droit de l'équation (3.5.1). Nous obtenons enfin les équations de Yule-Walker données par $\Gamma(p)\Phi = \gamma_p$, où $\Gamma(p) = [\gamma(i-p)]_{i,j=1:p}$ est la matrice de covariance et $\gamma_p = (\gamma(1), \dots, \gamma(p))'$.
2. Ensuite, en remplaçant les covariances $(\gamma(j))_{j=0:p}$ par les covariances correspondantes de l'échantillon $(\hat{\gamma}(j))_{j=0:p}$, nous obtenons un ensemble d'équations appelé estimateur de Yule-Walker.

3.5.2 Algorithme de Burg

L'algorithme de Burg (également appelé la méthode d'entropie maximale) pour estimer les paramètres d'AR est itératif, mais il utilise direc-

tement les données plutôt que la fonction covariance de l'échantillon. Pour des échantillons volumineux, il est asymptotiquement équivalent à l'estimateur de Yule-Walker bien que les deux diffèrent sur des échantillons de petites tailles.

L'estimateur préliminaire de Burg est obtenu en suivant les étapes suivantes :

1. Calcul des valeurs initiales :
 - La variance d'erreur : $\sigma_\varepsilon^2 = \frac{1}{n} \sum_{t=0}^{n-1} (x_t - \bar{X})^2$
 - Erreur en avant (Forwrdshift) : $e_n(0) = x_n$.
 - Erreur en arrière (Backshift) : $b_{n-1}(0) = x_{n-1}$
2. Calcul du coefficient de réflexion et l'erreur de la variance :
 - Coefficient de réflexion :

$$\tau_m = -2 \frac{\sum_{i=m}^{n-1} b_{n-1}(m-1) e_n(m-1)}{\sum_{i=m}^{n-1} b_{n-1}^2(m-1) + e_n^2(m-1)}.$$

- Erreur de la variance : $\sigma_\varepsilon^2(m) = (1 - \tau_m^2) \sigma_\varepsilon^2(m-1)$.
3. Mise à jour des coefficients et de l'erreur :
 - Coefficient AR :
- $$a_k(m) = \begin{cases} a_k(m-1) + \tau_m a_{m-k}(m-1) & \text{si } m = 1 \\ \tau_1 & \text{si non} \end{cases}$$
- Mise à jour des erreurs en avant : $e_n(m) = e_n(m-1) + \tau_m b_{n-1}(m-1)$.
 - Mise à jour des erreurs en arrière : $b_n(m) = b_{n-1}(m-1) + \tau_m e_n(m-1)$.
4. Répétez les étapes 2 et 3 (avec m incrémenté d'une unité) jusqu'à l'ordre p du modèle sélectionné.

3.5.3 Algorithme de Hannan-Rissanen

Les équations définies pour un modèle $ARMA(p, q)$ causal ont la forme d'un modèle de régression linéaire avec vecteur des coefficients $(\phi_1, \dots, \phi_p)'$.

Ce qui suggère l'utilisation de la régression des moindres carrés simple pour obtenir une estimation préliminaires.

Nous allons décrire les principales étapes de la procédure d'estimation Hannan-Rissanen.

1. ajuster un modèle $AR(m)$ d'ordre supérieur ($m > \max(p, q)$) à l'aide des équations de Yule-Walker de la sous section (3.5.1).
Si $\hat{\phi}_{m1}, \dots, \hat{\phi}_{mm}$ est le vecteur des coefficients estimés, estimez ensuite les résidus à partir des équations $\hat{\varepsilon}_t = X_t - \hat{\phi}_{m1}X_{t-1} - \dots - \hat{\phi}_{mm}X_{t-m}$ pour $t = m + 1, n$.
2. Une fois les estimations des résidus $(\hat{\varepsilon}_t)_{t=m+1, \dots, n}$ calculer à l'étape 1, le vecteur de paramètres, $\beta = (\Phi, \Theta)$ est estimé utilisant la méthode des moindres carrés à travers une régression linéaire de X_t en fonction de $X_{t-1}, \dots, X_{t-p}, \varepsilon_{t-1}, \dots, \varepsilon_{t-p}$ pour $t = m + q + 1 : n$.

3.6 Intervalles de confiance pour les Coefficients

Pour les échantillons de grande taille, l'estimateur de maximum de vraisemblance (EMV) $\hat{\beta}$ converge vers une loi normale de moyenne β_0 et de matrice de covariance $[n^{-1}V(\beta)]$ qui peut être approché par $2H^{-1}(\beta)$, où H est la matrice Hessienne $[\frac{\partial^2 l(\beta)}{\partial \beta_i \partial \beta_j}]_{i,j=1}^{p+q}$.

3.7 Exemples d'application

Dans cette section, nous donnons deux exemples d'application numériques. Le premier sur les processus moyenne mobile et le second sur les processus autoregressifs.

À travers ces exemples nous exposons la performance de notre méthode de calcul d'estimateur.

3.7.1 Le processus moyenne mobile MA(1)

Soit la série $(x_t)_{t=1,100}$ simulée à partir d'un modèle moyenne mobile d'ordre un avec $(\theta = 0.8)$, présentée dans le tableau 3.7.1 ci-dessous.

Time t	1	2	3	4	5	6	7	8	9	10
x_t	-0.587	0.417	1.703	0.746	0.435	0.9400	-0.478	-0.431	1.442	0.079
Time t	11	12	13	14	15	16	17	18	19	20
x_t	-0.963	0.369	0.664	-1.014	0.024	0.846	-0.385	-1.079	0.496	1.277
Time t	21	22	23	24	25	26	27	28	29	30
x_t	-0.370	0.314	0.413	1.393	3.095	1.446	1.341	1.841	-0.020	-0.269
Time t	31	32	33	34	35	36	37	38	39	40
x_t	-0.342	-0.191	1.954	2.221	0.586	0.987	2.855	2.775	-0.098	-0.889
Time t	41	42	43	44	45	46	47	48	49	50
x_t	-1.353	0.370	1.059	-0.834	-2.215	-2.049	-0.516	-0.794	-0.252	-0.604
Time t	51	52	53	54	55	56	57	58	59	60
x_t	-1.395	-1.068	-0.833	0.955	-0.196	0.576	1.973	0.605	0.936	1.278
Time t	61	62	63	64	65	66	67	68	69	70
x_t	0.459	-0.429	1.078	0.352	0.189	0.733	0.908	0.927	0.055	0.216
Time t	71	72	73	74	75	76	77	78	79	80
x_t	0.663	0.345	0.483	-0.318	-0.387	0.575	-0.253	-0.340	-0.018	-1.779
Time t	81	82	83	84	85	86	87	88	89	90
x_t	-1.286	-0.539	0.188	-1.085	-0.897	0.115	2.088	1.657	-2.313	-2.834
Time t	91	92	93	94	95	96	97	98	99	100
x_t	-1.066	-0.892	0.362	2.931	2.477	0.157	-0.196	0.010	1.173	1.000

TABLE 3.1 – Tableau représentant la série chronologique (x_t) .

Représentation graphique

La série moyenne mobile est par définition stationnaire, ce qui est visible pour notre série à travers la représentation graphique ci-dessous. La maximisation de la fonction de log-vraisemblance d'un $MA(1)$ est équivalente à la minimisation de :

$$\begin{aligned}
l_n(\Phi, \Theta) &= \frac{1}{n} \sum_{i=1}^n \ln r_{i-1} + \ln \left(\frac{1}{n} \left[(L_{\Theta}^{-1} (L_{\Phi}(X - \bar{X}) - Fe_*)) A^{-1/2} \right]^2 \right) \\
&= \frac{1}{n} \sum_{i=1}^n \ln r_{i-1} + \ln \left(\sum_{i=1}^n \frac{(\sum_{j=0}^i (-1)^j \theta^j x_{t-j})^2}{r_{i-1}} \right).
\end{aligned}$$

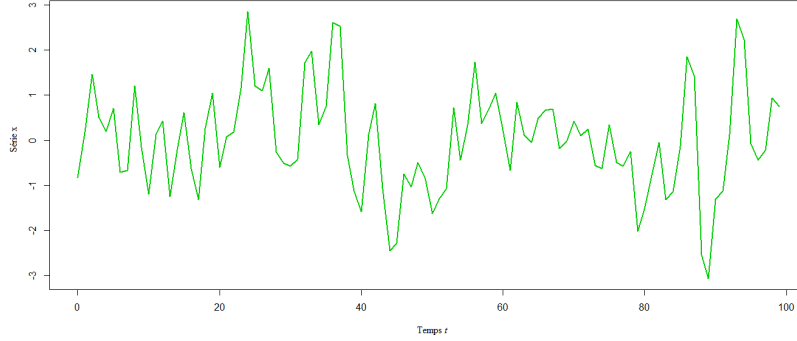


FIGURE 3.1 – Série chronologique x_t

où $\overline{X}$ est la moyenne empirique, $L_\Phi = I_n$, e_* est estimée par $\hat{e}_* = (0, \dots, 0)$, L_Θ est la matrice définie dans la section 3.2.1 avec $\theta_1 = \theta$, $\theta_i = 0$ pour tout $i > 1$ et les $(r_i)_{i \geq 0}$ sont définis par la relation de récurrence (3.2.5).

Utilisant l'algorithme de Hannan-Rissanen nous obtenons $\theta_{initial} = 0.7489867$, où le vecteur des coefficients estimées pour la représentation autoregressive sont représentés dans le tableau ci-dessous :

i	1	2	3	4	5	6	7	8	9	10
θ_i	0.567	-0.382	0.2949	-0.2990	0.2710	-0.0659	-0.0901	0.0227	0.0145	0.0588

L'intervalle de confiance calculé utilisant l'approximation de la matrice d'auto-covariance au point $\theta_{initial}$ est $]0.6191201, 0.8788534[$.

À partir de cet intervalle, nous générons un échantillon de $m = 100$ valeurs de θ_i , pour lesquelles nous calculons les valeurs de notre fonction l_n et ainsi nous obtenons un échantillon $(\theta_i, y_i = l_n(\theta_i))_{i=1,100}$.

Table d'échantillon $(\theta_i, l_n(\theta_i))_{i=1:100}$

Dans le tableau 3.7.1 ci-dessous, nous représentons l'échantillon $(\theta_i, y_i = l_n(\theta_i))_{i=1,100}$, qui va nous servir par la suite à trouver l'approximation de la fonction log-vraisemblance l_n .

θ	0.6191	0.6217	0.6244	0.6270	0.6296	0.6322	0.6347	0.6379	0.6401	0.6427
$l_n(\theta)$	-0.1718	-0.1730672	-0.1743	-0.1755	-0.1767	-0.1778	-0.1791	-0.1801	-0.1815	-0.1827
θ	0.64536	0.6480	0.6506	0.6532	0.6558	0.6585	0.6611	0.6637	0.6663	0.6690
$l_n(\theta)$	-0.1839	-0.1850	-0.1862	-0.1873	-0.1884	-0.1896	-0.1907	-0.1917	-0.1928	-0.1939
θ	0.6716	0.6742	0.6768	0.6795	0.6821	0.6847	0.6873	0.6899	0.6925	0.6952
$l_n(\theta)$	-0.1950	-0.1960	-0.1971	-0.1981	-0.1992	-0.2002	-0.2012	-0.2022	-0.2031	-0.2041
θ	0.6978	0.7004	0.7030	0.7056	0.7083	0.7109	0.7136	0.7162	0.7188	0.7214
$l_n(\theta)$	-0.2051	-0.2060	-0.2070	-0.2079	-0.2088	-0.2097	-0.2106	-0.2115	-0.2123	-0.2132
θ	0.7241	0.7267	0.7293	0.7319	0.73456	0.7372	0.7398	0.7424	0.745	0.7477
$l_n(\theta)$	-0.2140	-0.2148	-0.2156	-0.2164	-0.2172	-0.2176	-0.2187	-0.2194	-0.2201	-0.2208
θ	0.7503	0.7529	0.7555	0.7588	0.7608	0.7634	0.7660	0.7686	0.7712	0.7739
$l_n(\theta)$	-0.2215	-0.2221	-0.2227	-0.2235	-0.2239	-0.2245	-0.2250	-0.2255	-0.2260	-0.2265
θ	0.7765	0.7791	0.7818	0.7844	0.7870	0.7897	0.7922	0.7949	0.7975	0.8001
$l_n(\theta)$	-0.2270	-0.2274	-0.2278	-0.2281	-0.2285	-0.2288	-0.2290	-0.2293	-0.2295	-0.2297
θ	0.8028	0.8054	0.8080	0.8106	0.8133	0.8159	0.8185	0.8211	0.8237	0.8264
$l_n(\theta)$	-0.2298	-0.2299	-0.22993	-0.22994	-0.22991	-0.2298	-0.2297	-0.2295	-0.2293	-0.2290
θ	0.8290	0.8316	0.8342	0.8369	0.8395	0.8421	0.8447	0.8474	0.8500	0.8526
$l_n(\theta)$	-0.2287	-0.2282	-0.2278	-0.2272	-0.2266	-0.2260	-0.2252	-0.2244	-0.2234	-0.2225
θ	0.8552	0.8579	0.8605	0.8631	0.8657	0.8684	0.8710	0.8736	0.8762	0.8789
$l_n(\theta)$	-0.2214	-0.2201	-0.2188	-0.2174	-0.2159	-0.2142	-0.2124	-0.2106	-0.2085	-0.2063

TABLE 3.2 – Tableau représentant les valeurs de la log-vraisemblance du processus MA aux points de l'échantillon.

La fonction approchée, utilisant la méthode des moindres carrés, est :

$$f(\theta) = 0.8788534\theta^2 - 2.903081\theta + 0.9350449$$

Donc, l'estimation de θ est $\hat{\theta} = \frac{2.903081}{2*1.813548} = 0.8003874$, qui est une valeur très proche de la vraie valeur $\theta_0 = 0.8$.

Utilisant la méthode de Newton-Raphson, nous avons trouvé une approximation avec presque la même variance. Par la méthode des moindres carrés, nous avons obtenu une bonne approximation pour la fonction de vraisemblance, nous donnons ci-dessous la représentation graphique de la fonction log-vraisemblance et son approximation.

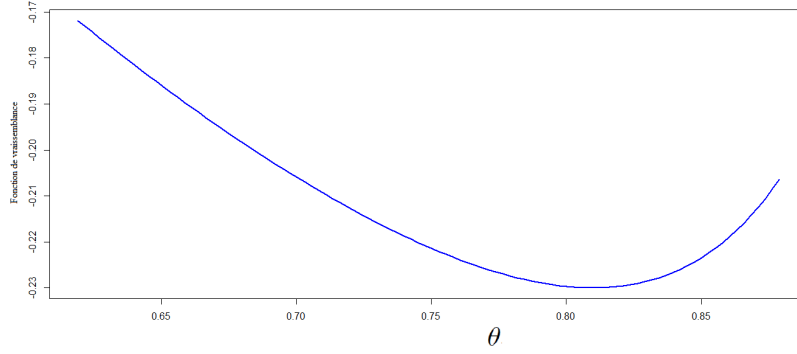


FIGURE 3.2 – Représentation graphique de la fonction log-vraisemblance.

3.7.2 Le processus autoregressif AR(1)

Soit la série $(y_t)_{t=1,100}$ simulée à partir d'un modèle autoregressif de paramètre ($\phi = 0.5$) présentée dans le tableau 3.7.2.

La maximisation de la fonction log-vraisemblance d'un $AR(1)$ est équivalente à la minimisation de :

$$\begin{aligned} l_n(\Phi, \Theta) &= \frac{1}{n} \sum_{i=1}^n \ln r_{i-1} + \ln \left(\frac{1}{n} \left[(L_{\Theta}^{-1} (L_{\Phi}(X - \bar{X}) - Fe_*)) A^{-1/2} \right]^2 \right) \\ &= \frac{1}{n} \sum_{i=1}^n \ln r_{i-1} + \ln \left(\sum_{i=1}^n \frac{(x_i - \phi x_{i-1})^2}{r_{i-1}} \right). \end{aligned}$$

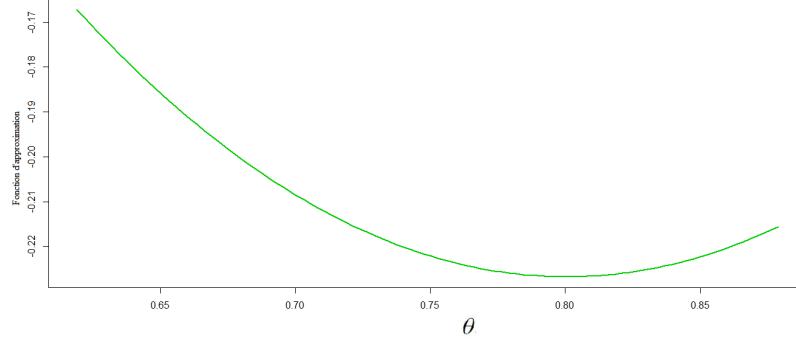


FIGURE 3.3 – Représentation graphique de l'approximation du fonction log-vraisemblance.

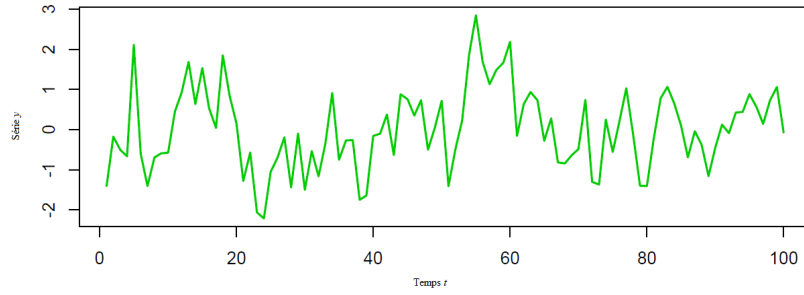


FIGURE 3.4 – Série chronologique y_t .

où $\bar{X}$ est la moyenne empirique, $L_\Theta = I_n$, e_* est estimé par $\hat{e}_* = (0, \dots, 0)$, L_Φ est la matrice définie dans la section 3.2.1 avec $\phi_1 = \phi$, $\phi_i = 0$ pour tout $i > 1$ et $r_i = 1$ pour tout $i \geq 0$.

Utilisant l'algorithme de Burg on obtient $\phi_{initial} = 0.504085$.

L'intervalle de confiance calculé utilisant l'approximation de la matrice d'autocovariance au point $\phi_{initial}$ est $]0.4871574, 0.5210126[$. À partir de cet intervalle, nous générons un échantillon de $m = 100$ valeurs de ϕ_i , puis nous calculons les valeurs de notre fonction l_n en ces points et ainsi nous obtenons un échantillon $(\phi_i, y_i = l_n(\phi_i))_{i=1,100}$.

Time t	1	2	3	4	5	6	7	8	9	10
y_t	-1.409	-0.175	-0.508	-0.669	2.110	-0.608	-1.410	-0.698	-0.594	-0.576
Time t	11	12	13	14	15	16	17	18	19	20
y_t	0.448	0.928	1.684	0.636	1.534	0.530	0.039	1.849	0.845	0.148
Time t	21	22	23	24	25	26	27	28	29	30
y_t	-1.282	-0.571	-2.066	-2.216	-1.055	-0.708	-0.193	-1.442	-0.100	-1.505
Time t	31	32	33	34	35	36	37	38	39	40
y_t	-0.538	-1.164	-0.357	0.911	-0.752	-0.273	-0.264	-1.753	-1.643	-0.157
Time t	41	42	43	44	45	46	47	48	49	50
y_t	-0.102	0.372	-0.635	0.877	0.756	0.349	0.728	-0.502	0.042	0.718
Time t	51	52	53	54	55	56	57	58	59	60
y_t	-1.413	-0.487	0.229	1.872	2.848	1.688	1.129	1.485	1.666	2.185
Time t	61	62	63	64	65	66	67	68	69	70
y_t	-0.158	0.635	0.934	0.729	-0.282	0.280	-0.820	-0.846	-0.639	-0.491
Time t	71	72	73	74	75	76	77	78	79	80
y_t	0.738	-1.305	-1.369	0.250	-0.558	0.213	1.029	-0.141	-1.399	-1.408
Time t	81	82	83	84	85	86	87	88	89	90
y_t	-0.234	0.776	1.065	0.655	0.094	-0.691	-0.041	-0.376	-1.160	-0.451
Time t	91	92	93	94	95	96	97	98	99	100
y_t	0.124	-0.089	0.423	0.438	0.881	0.567	0.141	0.722	1.061	-0.077

TABLE 3.3 – Tableau représentant la série chronologique (y_t) .

Table des $\phi, l_n(\phi)$

Après échantillonnage des points $(\phi_i)_{i=1:100}$ à partir de l'intervalle de confiance, nous calculons les valeurs de la fonction de log-vraisemblance l_n en ces points et nous présentons les résultats dans le tableau 3.7.2.

ϕ	0.4871	0.4874	0.4878	0.4881	0.4885	0.4888	0.4892	0.4895	0.4898	0.4902
$l_n(\phi)$	-0.3096810	-0.30969	-0.30970	-0.30972	-0.30973	-0.30974	-0.30975	-0.30976	-0.30977	-0.30979
ϕ	0.4905	0.4909	0.4912	0.4916	0.4919	0.4922	0.4926	0.4929	0.4933	0.4936
$l_n(\phi)$	-0.30980	-0.30981	-0.30982	-0.30983	-0.30983	-0.30984	-0.30985	-0.30986	-0.30987	-0.30988
ϕ	0.4939	0.4943	0.4946	0.4950	0.4953	0.4957	0.4960	0.4963	0.4967	0.4970
$l_n(\phi)$	-0.30988	-0.30989	-0.30990	-0.30990	-0.30991	-0.30992	-0.30992	-0.30993	-0.30993	-0.30994
ϕ	0.4974	0.4977	0.4981	0.4984	0.4987	0.4991	0.4994	0.4998	0.5001	0.5004
$l_n(\phi)$	-0.30994	-0.30994	-0.30995	-0.30995	-0.30995	-0.30996	-0.30996	-0.30996	-0.30996	-0.30996
ϕ	0.5008	0.5011	0.5015	0.5018	0.5022	0.5025	0.5028	0.5032	0.5035	0.5039
$l_n(\phi)$	-0.30996	-0.30996	-0.30996	-0.30996	-0.30996	-0.30996	-0.30996	-0.30996	-0.30996	-0.30996
ϕ	0.5042	0.5045	0.5049	0.5052	0.5056	0.5059	0.5063	0.5066	0.5069	0.5073
$l_n(\phi)$	-0.30996	-0.30995	-0.30995	-0.30995	-0.30994	-0.30994	-0.30994	-0.30993	-0.30993	-0.30992
ϕ	0.5076	0.5080	0.5083	0.5087	0.5090	0.5093	0.5097	0.5100	0.5104	0.5107
$l_n(\phi)$	-0.30992	-0.30991	-0.30990	-0.30990	-0.30989	-0.30988	-0.30988	-0.30987	-0.30986	-0.30985
ϕ	0.5110	0.5114	0.5117	0.5121	0.5124	0.5128	0.5131	0.5134	0.5138	0.5141
$l_n(\phi)$	-0.30984	-0.30984	-0.30983	-0.30982	-0.30981	-0.30980	-0.30979	-0.309778	-0.30976	-0.30975
ϕ	0.5145	0.5148	0.5151	0.5155	0.5158	0.5162	0.5165	0.5169	0.5172	0.5175
$l_n(\phi)$	-0.30974	-0.30973	-0.30972	-0.30970	-0.30969	-0.30968	-0.30966	-0.30965	-0.30963	-0.30962
ϕ	0.5179	0.5182	0.5186	0.5189	0.5193	0.5196	0.5199	0.5203	0.5206	0.5210
$l_n(\phi)$	-0.30960	-0.30959	-0.30957	-0.30956	-0.30954	-0.30952	-0.30951	-0.30949	-0.30947	-0.30945

TABLE 3.4 – Tableau représentant les valeurs de la log-vraisemblance du processus AR aux points de l'échantillon.

Par la méthode des moindres carrés nous obtenons l'approximation de la fonction de vraisemblance :

$$f(\phi) = 1.232747\phi^2 - 1.236284\phi - 2.49511110^{-9}.$$

Donc, l'estimation de ϕ est $\hat{\phi} = \frac{1.236284}{2*1.232747} = 0.5014345$, qui est de biais $|\phi - \phi_0| = 0.001434485$ clairement faible par rapport à la vraie valeur.

Pour illustrer la qualité de l'approximation de la fonction log-vraisemblance, nous donnons ci-dessous la représentation graphique de la fonction log-vraisemblance et son approximation :

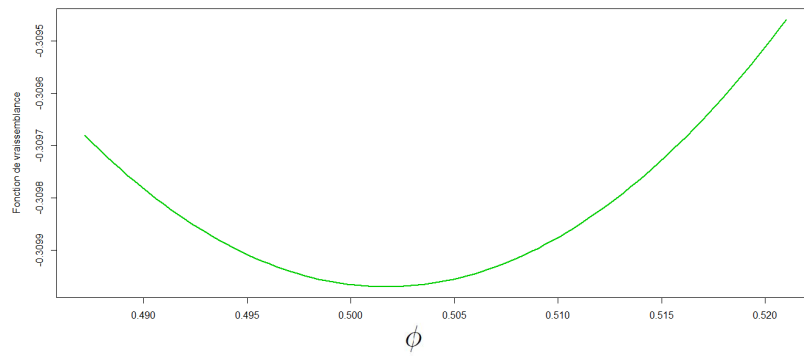


FIGURE 3.5 – Représentation graphique de la fonction log-vraisemblance.

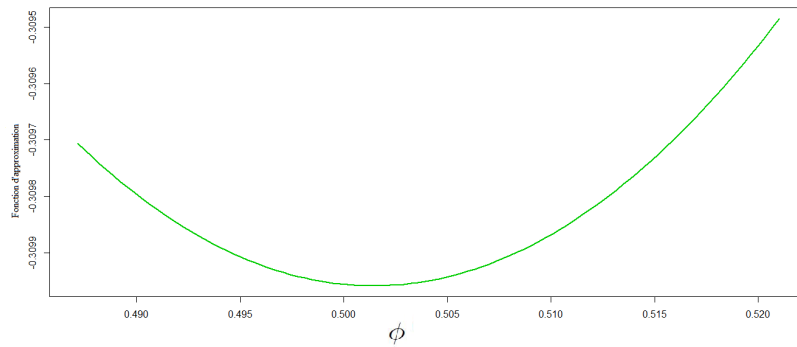


FIGURE 3.6 – Représentation graphique de l'approximation de fonction log-vraisemblance.

Articles

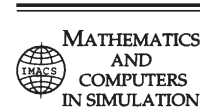
1. Y. Boularouk, K. Djeddour (2015). New approximation for ARMA parameters estimate. Mathematics and Computers in Simulation 118 (2015) 116-122.
<http://dx.doi.org/10.1016/j.matcom.2015.01.004>,
www.elsevier.com/locate/matcom
2. JM BARDET, Y BOULAROUK and K DJABALLAH. Asymptotic behavior of the Laplacian quasi-maximum likelihood estimator of affine causal processes.
<http://arxiv.org/abs/1601.00155>.



Available online at www.sciencedirect.com

ScienceDirect

Mathematics and Computers in Simulation 118 (2015) 116–122



www.elsevier.com/locate/matcom

Original articles

New approximation for ARMA parameters estimate

Y. Boularouk^{a,*}, K. Djeddour^b

^a Faculty of Mathematics, USTHB Algiers, Algeria

^b Laboratory MSTD, Faculty of Mathematics, USTHB Algiers, Algeria

Received 20 March 2014; received in revised form 6 January 2015; accepted 12 January 2015

Available online 29 January 2015

Abstract

This paper presents a new approach for the optimization of autoregressive moving average parameters estimation. We prove that the log likelihood function of the process is almost surely equal to a polynomial of order two. Thereafter, using the methods of least squares, our function will be approximated by a polynomial of order two which will be used to calculate an estimation of the maximum.

© 2015 International Association for Mathematics and Computers in Simulation (IMACS). Published by Elsevier B.V. All rights reserved.

Keywords: ARMA process; Likelihood function; Innovation algorithm; Multivariate least squares

1. Introduction

So far, we have no analytical form for the autoregressive moving average (ARMA) parameters estimates, particularly those obtained by maximizing the likelihood function which is nonlinear. Therefore, researchers use the iterative optimization or direct search routines. However iterative method may fail to find the maximum because of the highly nonlinear likelihood function in general and direct methods of research require a large number of function evaluations because they do not use information about its inclination and curvature.

In this sense and for the development of the likelihood function several works were devoted. Whittle [16] and Durbin [5] developed approximations to univariate MA models. Godolphin [6] extends the method of Whittle's to the univariate ARMA models and shows how to calculate the parameter estimates directly from the sample autocorrelations of the observed data. Durbin's approximation generates an algorithm in which the MA coefficients are calculated from the coefficients of a long-order autoregression. Hannan and Rissanen [7] and Koreisha and Pukkila [8] extend Durbin's method to univariate and multivariate ARMA models, respectively. Tunnicliffe-Wilson [15], Reinsel et al. [12] and DeFrutos and Serrano [3] propose similar approximate ML estimation methods.

Because of the nonlinearity of the likelihood function there is no analytical solution for the estimator of its maximum. It is also why many detailed algorithms were provided independently by A.I. McLeod, P.R. Holanda Sales [10], Solo [14], Dugre [4], Shea [13], Mitnik [11], Mauricio [9] and others. Approaches used in previous work (Gauss Newton, M estimate, EM algorithm and whittle estimator), to maximize the likelihood function are iterative methods

* Corresponding author.

E-mail address: y.boularouk@centre-univ-mila.dz (Y. Boularouk).

which approximate the maximum. In this work we will approximate the likelihood function by a quadratic form, and calculate its maximum.

2. Determination of the likelihood function

2.1. ARMA (p, q) likelihood function

The ARMA model is a stochastic process describing the random phenomena varying with time, it has the mathematical formula

$$X_t + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}. \quad (1)$$

X_t is assumed to be centered, otherwise it will be replaced by $\widehat{X}_t = X_t - \bar{X}$, ε_t is identically independently distributed (iid) random variable with zero mean and variance σ_ε^2 (i.e. $\varepsilon \approx N(0, \sigma_\varepsilon^2)$).

Determining the model is equivalent to the estimation of its unknown parameters. The most common method used is the maximization of the likelihood function. Under the assumption of normality, the likelihood function of an ARMA model for a sample of size n is

$$f(\Phi, \Theta, \sigma_\varepsilon^2) = \frac{1}{\sqrt{(2\pi\sigma^2)^n r_0 \dots r_{n-1}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n \frac{\varepsilon_i^2(\Phi, \Theta)}{r_{i-1}}}. \quad (2)$$

$\sigma_\varepsilon^2 = \sigma^2 r_i$ where the r_i (which tends to one [2]) are calculated using the innovation algorithm and σ^2 is the process variance.

Maximizing f with respect to $\beta = (\Phi, \Theta)$ is equivalent to the minimization of

$$l_n(\Phi, \Theta) = \frac{1}{n} \sum_{i=1}^n \ln r_{i-1} + \ln \left(n^{-1} \sum_{i=1}^n \frac{\varepsilon_i^2(\Phi, \Theta)}{r_{i-1}} \right). \quad (3)$$

This function has a non-linear shape and is non-quadratic which implies the absence of an analytical solution. For calculation of the explicit form, we have

$$\sum_{i=1}^n \frac{\varepsilon_i^2(\Phi, \Theta)}{r_{i-1}} = \left[\left(L_\Theta^{-1} (L_\Phi X - F e_*) \right) A^{-1/2} \right]^2, \quad (4)$$

with

$$L_\Theta = \begin{bmatrix} 1 & \theta_1 & \theta_2 & \dots & \theta_q & 0 & \dots & 0 \\ 0 & 1 & \dots & \dots & \theta_q & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & \dots & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & \theta_q & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & \dots & 1 \end{bmatrix}, \quad L_\Phi = \begin{bmatrix} 1 & -\phi_1 & -\phi_2 & \dots & -\phi_q & 0 & \dots & 0 \\ 0 & 1 & -\phi_2 & \dots & -\phi_q & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & \dots & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & -\phi_q & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & \dots & 1 \end{bmatrix},$$

$$F = \begin{bmatrix} H_p & M_q \\ 0 & 0 \end{bmatrix}, \quad H_p = \begin{bmatrix} \phi_p & \phi_{p-1} & \dots & \dots & \phi_1 \\ 0 & \phi_p & \dots & \dots & \phi_2 \\ \dots & 0 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \phi_{p-1} \\ 0 & 0 & \dots & \dots & \phi_p \end{bmatrix}, \quad M_q = \begin{bmatrix} \theta_q & \theta_{q-1} & \dots & \dots & \theta_1 \\ 0 & \theta_q & \dots & \dots & \theta_2 \\ \dots & 0 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \theta_{q-1} \\ 0 & 0 & \dots & \dots & \theta_q \end{bmatrix},$$

$$A = \sigma_\varepsilon^2 \begin{bmatrix} r_0 & 0 & \dots & 0 \\ 0 & r_1 & \dots & 0 \\ \dots & 0 & \dots & \dots \\ 0 & 0 & \dots & r_{n-1} \end{bmatrix}, \quad e_* = (X_0, \dots, X_{1-p}, \varepsilon_0, \dots, \varepsilon_{1-q})'.$$

2.2. Innovation algorithm for the ARMA process

To simplify the application of the innovation algorithm detailed in [2], we apply it not to the process X_t itself, but to the transformed process [1] defined as

$$\begin{cases} w_t = \sigma_\varepsilon^{-1} X_{t-1}, & \text{if } t = 1, \dots, \max(p, q) \\ w_t = \sigma_\varepsilon^{-1} (X_t + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p}), & t > \max(p, q). \end{cases}$$

Therefore, once the auto-covariances $\gamma(i)$ are known, we calculate the innovations recursively using the innovations algorithm presented below

$$\begin{cases} r_0 = \frac{\gamma_X(0)}{\sigma_\varepsilon^2} \\ \theta_{m,m-k} = r_k^{-1} \left[\frac{\gamma_X(m-k)}{\sigma_\varepsilon^2} - \sum_{j=0}^{k-1} \theta_{m,m-j} \theta_{n,n-j} r_j \right] \\ r_m = \frac{\gamma_X(0)}{\sigma_\varepsilon^2} - \sum_{j=0}^{m-1} \theta_{m,m-j} r_j \end{cases}$$

in application, γ is replaced by its estimate from the sample $\hat{\gamma}$.

Some special cases for r_n

MA(1) : If X_t is the moving average MA(1) defined by

$$X_t = \varepsilon_t + \theta \varepsilon_{t-1}$$

then

$$\begin{cases} r_0 = 1 + \theta^2 \\ r_m = 1 + \theta^2 - \theta^2 / r_{m-1}. \end{cases} \quad (5)$$

ARMA(1, 1) : If X_t is the autoregressive moving average ARMA(1, 1) defined by

$$X_t = \phi X_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$$

then

$$\begin{cases} r_0 = (1 + 2\phi\theta + \theta^2) / (1 - \phi^2) \\ r_m = 1 + \theta^2 - \theta^2 / r_{m-1}. \end{cases} \quad (6)$$

3. Some properties of ARMA's likelihood function

The likelihood function of the ARMA process has some characteristics that have allowed us to calculate the maximum differently. In this section and through a proposition and a lemma we prove the essential of these characteristics.

Proposition 1 (r_n Properties). Let β be the parameters vector, if $\beta \in \mathcal{R}$, then

$$\begin{cases} r_i > 1; \quad \forall i = 0, n \\ \left(\frac{1}{n} \sum_{i=1}^n \ln r_i \right) \rightarrow 0 \quad \text{as } n \rightarrow \infty. \end{cases}$$

Proof. X_t is an ARMA process. Then rewriting its mathematical formula in the inverse form as $X_t + \sum_{j=1}^{\infty} \pi_j X_{t+1-j} = \varepsilon_t$, we see that

$$\begin{aligned} \sigma_\varepsilon^2 = \text{Var}(\varepsilon_{t+1}) &= E \left(X_{t+1} + \sum_{j=1}^{\infty} \pi_j X_{t+1-j} \right)^2 \\ &\leq \sigma_\varepsilon^2 r_t \leq \sigma_\varepsilon^2 E \left(X_{t+1} + \sum_{j=1}^t \pi_j X_{t+1-j} \right)^2 \end{aligned}$$

$$\begin{aligned}
&= \sigma_\varepsilon^2 E \left(\varepsilon_{t+1} - \sum_{j=t+1}^{\infty} \pi_j X_{t+1-j} \right)^2 \\
&= \sigma_\varepsilon^2 \left(1 + E \left(\sum_{j=t+1}^{\infty} \pi_j X_{t+1-j} \right)^2 \right) \\
&\leq \sigma_\varepsilon^2 \left(1 + \left(\sum_{j=t+1}^{\infty} |\pi_j| \right) \gamma_X(0) \right).
\end{aligned}$$

From the first inequality we deduce that $r_i \geq 1$ and $r_0 \geq 1$, $\prod_{i=1}^n r_i > 1$. Moreover, since $\left(\sum_{j=t+1}^{\infty} |\pi_j| \right) \gamma_X(0) \rightarrow 0$ as $t \rightarrow \infty$, $r_t \rightarrow 1$, so that by Cesaro convergence

$$0 \leq \ln \left(\prod_{i=1}^n r_i \right)^{\frac{1}{n}} = \left(\frac{1}{n} \sum_{i=1}^n \ln r_i \right) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Lemma 1. For β in $\mathcal{R}$, the log-likelihood function $l_n(\beta)$ is almost surely equal to a polynomial of order two.

Proof. From Proposition 1, we have,

- (a) $r_i > 1 ; \forall i = 0, n$
- (b) $\left(\frac{1}{n} \sum_{i=1}^n \ln r_i \right) \rightarrow 0$ as $n \rightarrow \infty$
- (b) implies that : $\left(\frac{1}{n} \sum_{i=1}^n \ln r_i \right)$ is bounded by a positive real A .

1. Using the relations (a), (b) and the inequality $\ln(x) \leq 1 + x$ we find that

$$l(\beta) \leq A + \sum_{i=1}^n \varepsilon_i^2 = A + \left[\left(L_\Theta^{-1} (L_\Phi X - F\hat{e}_*) \right) \right]^2, \text{ which is a polynomial of second order, so integrable.}$$

2. Secondly, as the sequence r_n converges to 1 and $\left(\frac{1}{n} \sum_{i=1}^n \ln r_i \right) \rightarrow 0$ as $n \rightarrow \infty$, it is concluded that $l_n(\beta)$ converges to a second order polynomial. By the Monotone convergence theorem, (1) and (2) imply that l_n is almost surely (a.s.) equal to a polynomial of order two.

Since the likelihood function l_n is a.s. equal to a second order polynomial, we will approximate it by a second order polynomial and so its maximum $\hat{\beta}$ will be an approximation of l_n 's maximum.

4. Calculation procedure

To calculate the maximum likelihood, one passes by the following steps.

1. Calculate preliminary estimation of the parameters: To provide these initial values, a number of algorithms are available [2]. For pure autoregressive models one can use Yule–Walker or Burg estimation. While for models with $q > 0$ we could use the innovations or Hannan–Rissanen algorithms.
2. Estimate confidence intervals $\prod_{i=1}^{p+q} [a_i, b_i]$ using the preliminary estimates: Firstly, the confidence intervals are computed in the univariate context for each parameter. Then we make the product.
3. Generate $p + q$ sample of m elements $[(\theta_{ij}), l_n(\theta_{ij})]_{j=1, m}$ from the confidence region of each parameter.
4. Numerical approximation by a polynomial of order two taking the form $(a_0 + \sum_{i=1}^{p+q} a_i \beta_i)^2$: we use the least squares method.
5. Calculation of the maximum.

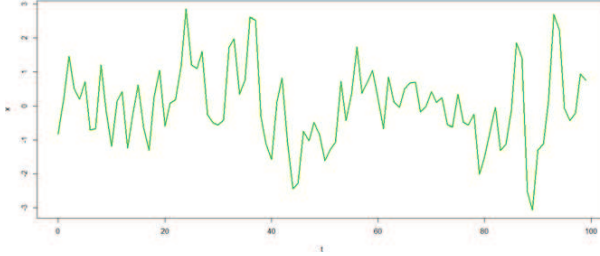
Confidence Regions for the Coefficients: For large sample size, the maximum likelihood estimator $\hat{\beta}$ of $\beta = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)'$ is approximately normally distributed with mean β and covariance matrix $[n^{-1} V(\beta)]$ which can be approximated by $2H^{-1}(\beta)$, where H is the Hessian matrix $[\frac{\partial^2 l(\beta)}{\partial \beta_i \partial \beta_j}]_{i,j=1}^{p+q}$.

5. Example of application

Let $(x_t)_{t=1,100}$ be the simulated first order moving average ($\theta = 0.8$) time series presented by the table

Time t	1	2	3	4	5	6	7	8	9	10
x_t	-0.5865	0.4174	1.7030	0.7459	0.4352	0.9398	-0.4776	-0.4312	1.4424	0.0787
Time t	11	12	13	14	15	16	17	18	19	20
x_t	-0.9625	0.3690	0.6639	-1.0143	0.0244	0.8462	-0.3852	-1.0791	0.4958	1.2771
Time t	21	22	23	24	25	26	27	28	29	30
x_t	-0.3703	0.3142	0.4129	1.3927	3.0953	1.4464	1.3411	1.8414	-0.0196	-0.2691
Time t	31	32	33	34	35	36	37	38	39	40
x_t	-0.3418	-0.1914	1.9542	2.2211	0.5855	0.9866	2.8550	2.7748	-0.0978	-0.8891
Time t	41	42	43	44	45	46	47	48	49	50
x_t	-1.3534	0.3702	1.0591	-0.8343	-2.2148	-2.0485	-0.5158	-0.7940	-0.2520	-0.6035
Time t	51	52	53	54	55	56	57	58	59	60
x_t	-1.3950	-1.0677	-0.8326	0.9554	-0.1960	0.5758	1.9727	0.6054	0.936	1.2782
Time t	61	62	63	64	65	66	67	68	69	70
x_t	0.4594	-0.4287	1.0783	0.3518	0.1891	0.7326	0.9081	0.9269	0.0545	0.2157
Time t	71	72	73	74	75	76	77	78	79	80
x_t	0.6633	0.3450	0.4832	-0.3177	-0.3872	0.5754	-0.2525	-0.3399	-0.0183	-1.779
Time t	81	82	83	84	85	86	87	88	89	90
x_t	-1.2859	-0.539	0.1878	-1.0854	-0.897	0.1152	2.0883	1.6571	-2.3130	-2.834
Time t	91	92	93	94	95	96	97	98	99	100
x_t	-1.0657	-0.8915	0.3622	2.9306	2.4771	0.1573	-0.1960	0.0095	1.1731	0.9899

We plot this time series as a function of time



Maximization of the log likelihood function of an $MA(1)$ is equivalent to the minimization

$$\begin{aligned}
 l_n(\Phi, \Theta) &= \frac{1}{n} \sum_{i=1}^n \ln r_{i-1} + \ln \left(\frac{1}{n} \left[\left(L_{\Theta}^{-1} (L_{\Phi} (X - \bar{X}) - F e_*) \right) A^{-1/2} \right]^2 \right) \\
 &= \frac{1}{n} \sum_{i=1}^n \ln r_{i-1} + \ln \left(\sum_{i=2}^n \frac{\left(\sum_{j=0}^i (-1)^j \theta^j x_{t-j} \right)^2}{r_i} \right), \quad (7)
 \end{aligned}$$

where $\bar{X}$ is the empirical mean, $L_{\Phi} = I_n$, e_* is estimates by $\hat{e}_* = (0, \dots, 0)$ and L_{Θ} is the matrix defined in Section 1 with $\theta_i = \begin{cases} \theta, & \text{if } i = 1 \\ 0, & \text{else.} \end{cases}$

Using the Hannan–Rissanen Algorithm we get $\theta_{i\text{initial}} = 0.7489867$, where the vector of estimated coefficients of the autoregressive representation is represented in the table below.

i	1	2	3	4	5	6	7	8	9	10	11	12
θ_i	0.567	-0.382	0.2949	-0.2990	0.2710	-0.0659	-0.0901	0.0227	0.0145	0.0588	-0.1890	0.1298

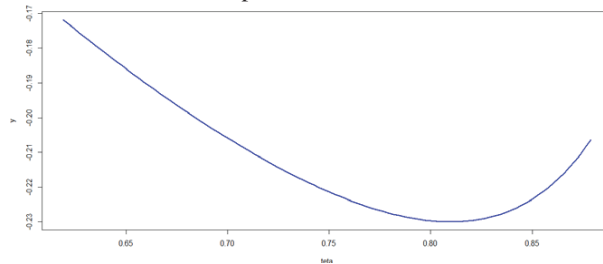
The confidence region calculated using the asymptotic auto-covariance matrix is $]0.6191201, 0.8788534[$.

From this interval we generate a sample of $m = 100$ values of θ_i , for which we calculate the values of our function l_n and so obtain our sample $(\theta_i, y_i = l_n(\theta_i))_{i=1,100}$.

Table of θ , $l_n(\theta)$

θ	0.6191	0.6217	0.6244	0.6270	0.6296	0.6322	0.6347	0.6379	0.6401	0.6427
$l_n(\theta)$	-0.1718	-0.1730672	-0.1743	-0.1755	-0.1767	-0.1778	-0.1791	-0.1801	-0.1815	-0.1827
θ	0.64536	0.6480	0.6506	0.6532	0.6558	0.6585	0.6611	0.6637	0.6663	0.6690
$l_n(\theta)$	-0.1839	-0.1850	-0.1862	-0.1873	-0.1884	-0.1896	-0.1907	-0.1917	-0.1928	-0.1939
θ	0.6716	0.6742	0.6768	0.6795	0.6821	0.6847	0.6873	0.6899	0.6925	0.6952
$l_n(\theta)$	-0.1950	-0.1960	-0.1971	-0.1981	-0.1992	-0.2002	-0.2012	-0.2022	-0.2031	-0.2041
θ	0.6978	0.7004	0.7030	0.7056	0.7083	0.7109	0.7136	0.7162	0.7188	0.7214
$l_n(\theta)$	-0.2051	-0.2060	-0.2070	-0.2079	-0.2088	-0.2097	-0.2106	-0.2115	-0.2123	-0.2132
θ	0.7241	0.7267	0.7293	0.7319	0.73456	0.7372	0.7398	0.7424	0.745	0.7477
$l_n(\theta)$	-0.2140	-0.2148	-0.2156	-0.2164	-0.2172	-0.2176	-0.2187	-0.2194	-0.2201	-0.2208
θ	0.7503	0.7529	0.7555	0.7588	0.7608	0.7634	0.7660	0.7686	0.7712	0.7739
$l_n(\theta)$	-0.2215	-0.2221	-0.2227	-0.2235	-0.2239	-0.2245	-0.2250	-0.2255	-0.2260	-0.2265
θ	0.7765	0.7791	0.7818	0.7844	0.7870	0.7897	0.7922	0.7949	0.7975	0.8001
$l_n(\theta)$	-0.2270	-0.2274	-0.2278	-0.2281	-0.2285	-0.2288	-0.2290	-0.2293	-0.2295	-0.2297
θ	0.8028	0.8054	0.8080	0.8106	0.8133	0.8159	0.8185	0.8211	0.8237	0.8264
$l_n(\theta)$	-0.2298	-0.2299	-0.22993	-0.22994	-0.22991	-0.2298	-0.2297	-0.2295	-0.2293	-0.2290
θ	0.8290	0.8316	0.8342	0.8369	0.8395	0.8421	0.8447	0.8474	0.8500	0.8526
$l_n(\theta)$	-0.2287	-0.2282	-0.2278	-0.2272	-0.2266	-0.2260	-0.2252	-0.2244	-0.2234	-0.2225
θ	0.8552	0.8579	0.8605	0.8631	0.8657	0.8684	0.8710	0.8736	0.8762	0.8789
$l_n(\theta)$	-0.2214	-0.2201	-0.2188	-0.2174	-0.2159	-0.2142	-0.2124	-0.2106	-0.2085	-0.2063

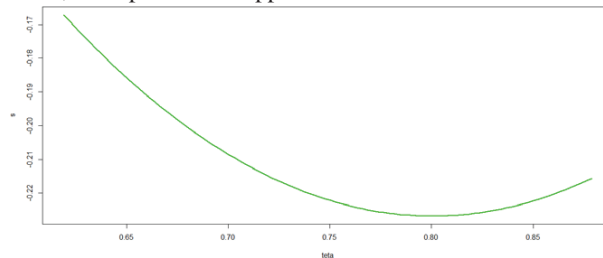
The likelihood function representation is



The approximation function, using least squares method for approximation, is

$$f(\theta) = 0.8788534\theta^2 - 2.903081\theta + 0.9350449.$$

Then, we represent the approximation function



Therefore the approximation of θ is $\hat{\theta} = \frac{-b}{2a} = \frac{2.903081}{2 \times 1.813548} = 0.8003874$. Using the Newton–Raphson method, we found approximate values with almost the same variance.

References

- [1] C.F. Ansley, An algorithm for the exact likelihood of a mixed autoregressive moving average process, *Biometrika* 66 (1979) 59–65.
- [2] P.J. Brockwell, R.A. Davis, *Time Series Theory and Methods*, Springer-Verlag, New York, 2002.
- [3] R.F. DeFrutos, G.R. Serrano, A generalized least squares estimation method for invertible vector moving average models, *Econom. Lett.* 57 (1997) 149–156.
- [4] J.P. Dugre, Generating covariance sequences and the calculation of quantization and rounding error variances in digital filters, *IEEE Trans. Acoust. Speech Signal Process.* 28 (1980) 102–104.
- [5] J. Durbin, Efficient estimation of parameters in moving-average models, *Biometrika* 46 (1959) 306–316.

- [6] E.J. Godolphin, A direct representation for the large sample maximum likelihood estimator of a Gaussian autoregressive moving average process, *Biometrika* 71 (1984) 281–289.
- [7] E.J. Hannan, J. Rissanen, Recursive estimation of mixed autoregressive-moving average order, *Biometrika* 69 (1982) 81–94.
- [8] S. Koreisha, T. Pukkila, Fast linear estimation methods for vector moving average models, *J. Time Series Anal.* 10 (1989) 325–339.
- [9] J.A. Mauricio, An algorithm for the exact likelihood of a vector autoregressive moving average model, *J. Time Series Anal.* 23 (2002) 473–486.
- [10] A.I. McLeod, P.R. Holanda, Sales, An algorithm for approximate likelihood calculation of ARMA and seasonal ARMA models, *J. R. Stat. Soc.* 32 (2) (1983) 211–223.
- [11] S. Mitnik, Computation of theoretical autocovariance matrices of multivariate autoregressive moving average time series, *J. R. Stat. Soc. Ser. B* 52 (1990) 151–155.
- [12] G.C. Reinsel, S. Basu, S.F. Yap, Maximum likelihood estimators in the multivariate autoregressive moving average model from a generalized least squares viewpoint, *J. Time Series Anal.* 13 (1992) 133–145.
- [13] B.L. Shea, A note on the generation of independent realisations of a vector autoregressive moving average proces, *J. Time Series Anal.* 9 (1988) 403–410.
- [14] V. Solo, The exact likelihood for a multivariate ARMA mode, *J. Multivariate Anal.* 15 (1984) 164–173.
- [15] G. Tunncliffe-Wilson, The estimation of parameters in multivariate time series models, *J. R. Stat. Soc Ser. B* 35 (1973) 76–85.
- [16] P. Whittle, Thesis : Hypothesis Testing in Time Series Analysis, Almquist and Wiksell, Upsala, 1951.

Asymptotic behavior of the Laplacian quasi-maximum likelihood estimator of affine causal processes

Jean-Marc Bardet^{1,*} Yakoub Boularouk^{2,**} and Khedidja Djaballah^{2,†}

¹*S.A.M.M., Université Panthéon-Sorbonne, 90, rue de Tolbiac, 75634, Paris, France. e-mail: *bardet@univ-paris1.fr*

²*Université des Sciences et de la Technologie Houari Boumédiène, Alger, Algérie. e-mail:*

***y.boularouk@centre-univ-mila.dz, e-mail: †khdjeddour@hotmail.com*

Abstract: We prove the consistency and asymptotic normality of the Laplacian Quasi-Maximum Likelihood Estimator (QMLE) for a general class of causal time series including ARMA, AR(∞), GARCH, ARCH(∞), ARMA-GARCH, APARCH, ARMA-APARCH,..., processes. We notably exhibit the advantages (moment order and robustness) of this estimator compared to the classical Gaussian QMLE. Numerical simulations confirms the accuracy of this estimator.

Keywords and phrases: Laplacian Quasi-Maximum Likelihood Estimator, Strong consistency, Asymptotic normality, ARMA-ARCH processes.

1. Introduction

This paper is devoted to establish the consistency and the asymptotic normality of a parametric estimator for a general class of time series. This class was already defined and studied in [Doukhan and Wintenberger \(2007\)](#), [Bardet and Wintenberger \(2009\)](#) and [Bardet et al. \(2012\)](#). Hence, we will consider an observed sample $(X_1, \dots, X_n)$ where $(X_t)_{t \in \mathbb{Z}}$ is a solution of the following equation:

$$X_t = M_{\theta_0}(X_{t-1}, X_{t-2}, \dots) \zeta_t + f_{\theta_0}(X_{t-1}, X_{t-2}, \dots), \quad t \in \mathbb{Z}, \quad (1.1)$$

where

- $\theta_0 \in \Theta \subset \mathbb{R}^d$, $d \in \mathbb{N}^*$, is an unknown vector of parameters, also called the "true" parameters;
- $(\zeta_t)_{t \in \mathbb{Z}}$ is a sequence of centred independent identically distributed random variables (i.i.d.r.v.) with symmetric probability distribution, i.e. $\zeta_0 \stackrel{\mathcal{L}}{=} -\zeta_0$, satisfying $E[|\zeta_0|^r] < \infty$ with $r \geq 1$ and $E[|\zeta_0|] = 1$. If $r \geq 2$, denote $\sigma_\zeta^2 = \text{Var}(\zeta_0)$;
- $(\theta, (x_n)_{n \in \mathbb{N}}) \rightarrow M_\theta((x_n)_{n \in \mathbb{N}}) \in (0, \infty)$ and $(\theta, (x_n)_{n \in \mathbb{N}}) \rightarrow f_\theta((x_n)_{n \in \mathbb{N}}) \in \mathbb{R}$ are two known applications.

For instance, if $M_{\theta_0}(X_{t-1}, X_{t-2}, \dots) = 1$ and $f_{\theta_0}(X_{t-1}, X_{t-2}, \dots) = \alpha_0 X_{t-1}$ with $|\alpha_0| < 1$ then (X_t) is a causal AR(1) process. In [Doukhan and Wintenberger \(2007\)](#) and [Bardet and Wintenberger \(2009\)](#), it was proved that all the most famous stationary time series used in econometrics, such as ARMA, AR(∞), GARCH, ARCH(∞), TAR, ARMA-GARCH processes can be written as a causal stationary solution of (1.1).

In [Bardet and Wintenberger \(2009\)](#), it was also established that under several conditions on M_θ , f_θ and if $E[|\zeta_0|^r]$ with $r \geq 2$, the usual Gaussian Quasi-Maximum Likelihood Estimator (QMLE) of θ is strongly consistent and when $r \geq 4$ it is asymptotically normal. This estimator was first defined by [Weiss \(1986\)](#) for ARCH processes, and the asymptotic study of this estimator was first obtained by [Lumsdaine \(1996\)](#) for GARCH(1, 1) processes, [Berkes et al. \(2003\)](#) for GARCH(p, q) processes, [Francq and Zakoian \(2004\)](#) for ARMA-GARCH processes, [Straumann and Mikosch \(2006\)](#) for general heteroskedastic models, and [Robinson and Zaffaroni \(2006\)](#) for ARCH(∞) processes. The results of [Bardet and Wintenberger \(2009\)](#) devoted to processes satisfying almost everywhere (1.1) as well as its multivariate generalisation, provide a general and unified framework for studying the asymptotic properties of the Gaussian QMLE.

However, the definition of the Gaussian QMLE is explicitly obtained with the assumption that (ζ_t) is a

Gaussian sequence and even if it could be applied when the probability distribution of (ζ_t) is non-Gaussian, it keeps some drawbacks of this initial assumption. Indeed, the computation of this estimators requires the minimization of a least squares contrast (typically $\sum_{t=1}^n M_\theta^{-2}(X_t - f_\theta)^2$) and this induces that $r = 2$ is required for the consistency and $r = 4$ for the asymptotic normality (and therefore confidence intervals or tests). For numerous real data such requirement is sometimes too strong (for instance the kurtosis of economic data is frequently considered as infinite). Moreover, such estimator is not robust to eventual outliers. Hence, the reference probability distribution of (ζ_t) could be a Laplace one and this allows to avoid both these drawbacks. Roughly speaking this choice implies to minimize a least absolute deviations contrast (typically $\sum_{t=1}^n M_\theta^{-1}|X_t - f_\theta|$) instead of the previous least squares contrast. And therefore, $r = 1$ will be sufficient for insuring the strong consistency of this Laplacian-QMLE, while only $r = 2$ is required for the asymptotic normality (see below).

Such probability distribution choice is not new since this leads to a Least Absolute Deviations (LAD) estimation. Hence, for ARMA processes, [Davis and Dunsmuir \(1997\)](#) proved the consistency and asymptotic normality of the LAD estimator. For ARCH or GARCH processes, the same results concerning the LAD estimator were already established by [Peng and Yao \(2003\)](#), while [Berkes and Horváth \(2004\)](#) proved the consistency and asymptotic normality of the Laplacian-QMLE. [Newey and Steigerwald \(1997\)](#) considered also the estimator for other conditional heteroskedasticity models. Recently, [Francq et al. \(2011\)](#) proposed a two-stage non-Gaussian-QML estimation for GARCH models and [Francq and Zakoian \(2013\)](#) proposed an alternative one-step procedure, based on an appropriate non-Gaussian-QML estimator, the asymptotic properties of both these approaches were studied.

In this paper we unified all these studies of the Laplacian-QMLE in a simple framework, *i.e.* causal stationary solutions of (1.1). This notably allows to obtain known results on ARMA or GARCH but also to establish for the first time the consistency and the asymptotic normality of the Laplacian-QMLE for APARCH, ARMA-GARCH, ARMA-ARCH(∞) and ARMA-APARCH processes.

Numerical Monte-Carlo experiments were realized to illustrate the theoretical results. And there are convincing, especially when the accuracy of Laplacian-QMLE is compared with Gaussian-QMLE: except for Gaussian distribution of (ζ_t) , the Laplacian-QMLE provides a sharper estimation than Gaussian-QMLE for all the other probability distributions we considered. This is notably the case, and this is not a surprise, for a Gaussian mixing which mimics the presence of outliers.

The following Section 2 will be devoted to provide the definitions and assumptions. In Section 3 the main results are stated with numerous examples of application, while Section 4 presents the results of Monte-Carlo experiments and Section 5 contains the proofs.

2. Definition and assumptions

2.1. Definition of the estimator

Let $(X_1, \dots, X_n)$ be an observed trajectory of X which is an a.s. solution of (1.1) where $\theta \in \Theta \subset \mathbb{R}^d$ is unknown. For estimating θ we consider the log-likelihood of $(X_1, \dots, X_n)$ conditionally to $(X_0, X_{-1}, \dots)$. If g is the probability density (with respect to Lebesgue measure) of ζ_0 , then, from the affine causal definition of X , this log-likelihood can be written:

$$\log(L_\theta(X_1, \dots, X_n)) = \sum_{t=1}^n \log\left(g\left(\frac{X_t - f_\theta^t}{M_\theta^t}\right)\right)$$

where $M_\theta^t := M_\theta(X_{t-1}, X_{t-2}, \dots)$ and $f_\theta^t := f_\theta(X_{t-1}, X_{t-2}, \dots)$, with the assumption that $M_\theta^t > 0$. However, M_θ^t and f_θ^t are generally not computable since $X_0, X_{-1}, \dots$ are unknown. Thus, a quasi-log-likelihood is

considered instead of the log-likelihood and it is defined by:

$$\log (QL_\theta(X_1, \dots, X_n)) = \sum_{t=1}^n \log \left(g \left(\frac{X_t - \hat{f}_\theta^t}{\widehat{M}_\theta^t} \right) \right),$$

with $\hat{f}_\theta^t := f_\theta(X_{t-1}, \dots, X_1, u)$ and $\widehat{M}_\theta^t := M_\theta(X_{t-1}, \dots, X_1, u)$, where $u = (u_n)_{n \in \mathbb{N}}$ is a finitely non-zero sequence $(u_n)_{n \in \mathbb{N}}$. The choice of $(u_n)_{n \in \mathbb{N}}$ does not have any consequences on the asymptotic behaviour of L_n , and (u_n) could typically be chosen as a sequence of zeros. Finally, a Quasi-Maximum Likelihood Estimator (QMLE) is defined by:

$$\tilde{\theta}_n := \text{Argmax}_{\theta \in \Theta} \log (QL_\theta(X_1, \dots, X_n)).$$

Usually, the Gaussian distribution is considered, with $g(x) = (2\pi)^{-1/2} e^{-x^2/2}$ for $x \in \mathbb{R}$ and this provides the Gaussian-QMLE of θ .

Here, we chose the Laplacian distribution, *i.e.*,

$$g(x) = \frac{1}{2} e^{-|x|} \quad \text{for } x \in \mathbb{R}, \quad (2.1)$$

and this implies $\mathbb{E}[|\zeta_0|] = 1$.

Therefore, we respectively define the Laplacian-likelihood and Laplacian-quasi-likelihood by:

$$L_n(\Theta) = - \sum_{t=1}^n q_t(\Theta) \quad \text{with} \quad q_t(\Theta) = \log |M_\theta^t| + |M_\theta^t|^{-1} |X_t - f_\theta^t| \quad (2.2)$$

$$\widehat{L}_n(\theta) = - \sum_{t=1}^n \widehat{q}_t(\theta) \quad \text{with} \quad \widehat{q}_t(\theta) := \log |\widehat{M}_\theta^t| + |\widehat{M}_\theta^t|^{-1} |X_t - \widehat{f}_\theta^t|. \quad (2.3)$$

A Laplacian-QMLE $\widehat{\theta}_n$ is a maximizer of $\widehat{L}_n$:

$$\widehat{\theta}_n := \arg \max_{\theta \in \Theta} \widehat{L}_n(\theta).$$

We restrict the set Θ in such a way that a stationary solution (X_t) of order 1 or 2 of (1.1) exists. Additional conditions are also required for insuring the consistency and the asymptotic normality of $\widehat{\theta}_n$. More details are given now.

2.2. Existence and stationarity

As it was already done in Doukhan and Wintenberger (2007) and Bardet and Wintenberger (2009), several Lipshitz-type inequalities on f_θ and M_θ are required for obtaining the existence and r -order stationarity of an ergodic causal solution of (1.1). First, denote $\|g_\theta\|_\Theta = \sup_{\theta \in \Theta} \|g_\theta\|$ with $m \in \mathbb{N}^*$ and $\|\cdot\|$ the usual Euclidian norm. Now, let us introduce the generic symbol K for any of the functions f or M and for $g_\theta : \mathbb{R}^{\mathbb{N}} \mapsto \mathbb{R}^m$ or $\mathcal{M}_m(\mathbb{R})$, For $k = 0, 1, 2$ and some subset Θ of $\mathbb{R}^d$, define a Lipshitzian assumption on function K_θ :

Assumption $(\mathbf{A}_k(K, \Theta))$ $\forall x \in \mathbb{R}^\infty$, $\theta \in \Theta \mapsto K_\theta(x) \in \mathcal{C}^k(\Theta)$ and $\partial_\theta^k K_\theta$ satisfies $\|\partial_\theta^k K_\theta(0)\|_\Theta < \infty$ and there exists a sequence $(\alpha_j^{(k)}(K, \Theta))_j$ of nonnegative numbers such that $\forall x, y \in \mathbb{R}^{\mathbb{N}}$

$$\|\partial_\theta^k K_\theta(x) - \partial_\theta^k K_\theta(y)\|_\Theta \leq \sum_{j=1}^{\infty} \alpha_j^{(k)}(K, \Theta) |x_j - y_j|, \quad \text{with} \quad \sum_{j=1}^{\infty} \alpha_j^{(k)}(K, \Theta) < \infty$$

For ensuring a stationary r -order solution of (1.1), for $r \geq 1$, define the set

$$\Theta(r) := \left\{ \theta \in \mathbb{R}^d, (\mathbf{A}_0(f, \{\theta\})) \text{ and } (\mathbf{A}_0(M, \{\theta\})) \text{ hold,} \right. \\ \left. \sum_{j=1}^{\infty} \alpha_j^{(0)}(f, \{\theta\}) + (\mathbb{E}[|\zeta_0|^r])^{1/r} \sum_{j=1}^{\infty} \alpha_j^{(0)}(M, \{\theta\}) < 1 \right\}.$$

Then, from Doukhan and Wintenberger (2007), we obtain:

Proposition 2.1. *If $\theta_0 \in \Theta(r)$ for some $r \geq 1$, then there exists a unique causal (X_t is independent of $(\zeta_i)_{i>t}$ for $t \in \mathbb{Z}$) solution X of (1.1), which is stationary, ergodic and satisfies $E[|X_0|^r] < \infty$.*

Finally, the following lemma insures that if a process X satisfies Proposition 2.1, a causal predictable ARMA process with X as innovation also satisfies Proposition 2.1. We first provide the classical following notion for a sequence $(u_n)_{n \in \mathbb{N}}$ of real numbers:

$(u_n)_{n \in \mathbb{N}}$ is an exponentially decreasing sequence (EDS)

$\iff$

there exists $\rho \in [0, 1[$ such as $u_n = \mathcal{O}(\rho^n)$ when $n \rightarrow \infty$.

Lemma 2.1. *Let X be a.s. a causal stationary solution of (1.1) for $\theta_0 \in \mathbb{R}^d$. Let $\tilde{X}$ be such as $\tilde{X}_t = \Lambda_{\beta}(L) X_t$ for $t \in \mathbb{Z}$ with $\Lambda_{\beta_0}(L) = P_{\beta_0}^{-1}(L) Q_{\beta_0}(L)$ where $(P_{\beta_0}, Q_{\beta_0})$ are the coprime polynomials of a causal invertible ARMA(p, q) processes with a vector of parameters $\beta_0 \in \mathbb{R}^{p+q}$. Denote $\Lambda_{\beta_0}^{-1}(x) = Q_{\beta_0}^{-1}(x) P_{\beta_0}(x) = 1 + \sum_{j=1}^{\infty} \psi_j(\beta_0) x^j$. Then $\tilde{X}$ is a.s. a causal stationary solution of the equation*

$$\tilde{X}_t = \tilde{M}_{\tilde{\theta}_0}((\tilde{X}_{t-i})_{i \geq 1}) \zeta_t + \tilde{f}_{\tilde{\theta}_0}((\tilde{X}_{t-i})_{i \geq 1}) \quad \text{for } t \in \mathbb{Z},$$

where $\tilde{f}_{\tilde{\theta}_0}$ and $\tilde{M}_{\tilde{\theta}_0}$ are given in (5.1) and $\tilde{\theta}_0 = (\theta_0, \beta_0)$. Moreover, for $i = 0, 1, 2$ and with $K = f$ or M and $\tilde{K} = \tilde{f}$ or $\tilde{M}$,

- if $\alpha_j^{(i)}(K, \{\theta_0\}) = \mathcal{O}(j^{-\beta})$ and $\beta > 1$, then $\alpha_j^{(i)}(\tilde{K}, \{\tilde{\theta}_0\}) = \mathcal{O}(j^{-\beta})$;
- if $\alpha_j^{(i)}(K, \{\theta_0\})$ is EDS, then $\alpha_j^{(i)}(\tilde{K}, \{\tilde{\theta}_0\})$ is EDS.

2.3. Assumptions required for the convergence of the Laplacian-QMLE

The Laplacian-QMLE could converge and be asymptotically Gaussian but this requires some additional assumptions on Θ and functions f_{θ} and M_{θ} :

- **Condition C1** (Compactness) Θ is a compact set.
- **Condition C2** (Lower bound of the conditional variance) There exists a deterministic constant $\underline{M} > 0$ such that for all $\theta \in \Theta$, then $M_{\theta}(x) > \underline{M}$ for all $x \in \mathbb{R}^N$.
- **Condition C3** (Identifiability) The functions M_{θ} and f_{θ} are such that: for all $\theta_1, \theta_2 \in \Theta$, then $M_{\theta_1} = M_{\theta_2}$ and $f_{\theta_1} = f_{\theta_2}$ implies that $\theta_1 = \theta_2$.

3. Consistency and asymptotic normality of the estimator

3.1. Consistency and asymptotic normality

First we prove the strong consistency of a sequence of Laplacian-QMLE for a solution of (1.1). The proof of this theorem, is postponed in Section 5, as well as the other proofs.

Theorem 3.1. *Assume Conditions C1, C2 and C3 hold and $\theta_0 \in \Theta(r) \cap \Theta$ with $r \geq 1$. Let X be the stationary solution of (1.1). If $(A_0(f, \Theta))$ and $(A_0(M, \Theta))$ hold with*

$$\alpha_j^{(0)}(f, \Theta) + \alpha_j^{(0)}(M, \Theta) = \mathcal{O}(j^{-\ell}) \quad \text{for some } \ell > \frac{2}{\min(r, 2)} \quad (3.1)$$

then a sequence of Laplacian-QMLE $(\hat{\theta}_n)_n$ strongly converges, that is $\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{a.s.} \theta_0$.

Of course, the conditions required for this strong consistency of a sequence of Laplacian-QMLE are almost the same than the ones required for the strong consistency of a sequence of Gaussian-QMLE except that $r \in [1, 2)$ is proved to be possible in Theorem 3.1 and not in case of Gaussian-QMLE (see Bardet

and Wintenberger (2009)). Moreover, if $r = 2$, the condition (3.1) on Lipschitzian coefficients is weaker for Laplacian-QMLE than for Gaussian-QMLE. As we will see below, many usual time series can satisfy the assumptions of Theorem 3.1; for example, an $\text{AR}(\infty)$ process can be defined for satisfying the strong consistency of Laplacian-QMLE while the conditions given in Bardet and Wintenberger (2009) do not ensure the strong consistency of Gaussian-QMLE.

Now we state an extension of Theorem 1 proved in Davis and Dunsmuir (1997) which will be an essential step of the proof of the asymptotic normality of the estimator.

Theorem 3.2. *Let $(Z_t)_{t \in \mathbb{Z}}$ be a sequence of i.i.d.r.v such as $\text{Var}(Z_0) = \sigma^2 < \infty$, with common distribution function which is symmetric ($F(-x) = 1 - F(x)$ for $x \in \mathbb{R}$) and is continuously differentiable in a neighborhood of 0 with derivative $f(0)$ in 0. Denote $\mathcal{F}_t = \sigma(Z_t, Z_{t-1}, \dots)$ for $t \in \mathbb{Z}$ and let $(Y_t)_{t \in \mathbb{Z}}$ and $(V_t)_{t \in \mathbb{Z}}$ two stationary processes adapted to $(\mathcal{F}_t)_t$ and such as $\mathbb{E}[Y_0^2 V_0^2] < \infty$. Then*

$$\sum_{t=1}^n V_{t-1} (|Z_t - n^{-1/2} Y_{t-1}| - |Z_t|) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}\left(f(0) \mathbb{E}[V_0 Y_0^2], \mathbb{E}[V_0^2 Y_0^2]\right) \quad (3.2)$$

Then, the asymptotic normality of the Laplacian-QMLE can be established using additional assumptions:

Theorem 3.3. *Assume that $\theta_0 \in \overset{\circ}{\Theta} \cap \Theta(r)$ where $r \geq 2$ and $\overset{\circ}{\Theta}$ denotes the interior of Θ . Let X be the stationary solution of the equation (1.1). Assume that the conditions of Theorem 3.1 hold and for $i = 1, 2$, assume $(A_i(f, \Theta))$ and $(A_i(M, \Theta))$ hold. Then, if the cumulative probability function of ζ_0 is continuously differentiable in a neighborhood of 0 with derivative $g(0)$ in 0 and if matrix Γ_F or Γ_M , defined in (5.21), are definite positive symmetric matrix, then*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}_d\left(0, (\Gamma_M + 2g(0)\Gamma_F)^{-1}((\sigma_\zeta^2 - 1)\Gamma_M + \Gamma_F)(\Gamma_M + 2g(0)\Gamma_F)^{-1}\right). \quad (3.3)$$

As it was already proved for the median estimator (see van der Vaart (2000)) or for least absolute deviations estimator of ARMA process (see Davis and Dunsmuir (1997)), it is not surprising that the probability density function g of the white noise $(\zeta_i)_i$ impacts the asymptotic covariance of (3.3). However, when $f_\theta = 0$, this is not such the case and this is what happens for GARCH processes (see Francq et al. (2011) or Francq and Zakoian (2015)) where g does not appear in the asymptotic covariance.

3.2. Examples

In this section, several examples of time series satisfying the conditions of previous results are considered. Like it could be boring to state the results for all sufficiently famous processes, we refer, *mutatis mutandis*, to Bardet and Wintenberger (2009) and Bardet et al. (2012) for $\text{ARCH}(\infty)$ and $\text{TARCH}(\infty)$.

1/ APARCH processes. APARCH(δ, p, q) model has been introduced (see Ding et al. (1993)) as the solution of equations

$$\begin{cases} X_t = \sigma_t \zeta_t, \\ \sigma_t^\delta = \omega + \sum_{i=1}^p \alpha_i (|X_{t-i}| - \gamma_i X_{t-i})^\delta + \sum_{j=1}^q \beta_j \sigma_{t-j}^\delta, \end{cases} \quad (3.4)$$

where $\delta \geq 1$, $\omega > 0$, $-1 < \gamma_i < 1$ and $\alpha_i \geq 0$ for $i = 1, \dots, p$, $\beta_j \geq 0$ for $j = 1, \dots, q$ with α_p, β_q strictly positive and $\sum_{j=1}^q \beta_j < 1$. Hence, we denote here $\theta = (\delta, \omega, \alpha_1, \dots, \alpha_p, \gamma_1, \dots, \gamma_p, \beta_1, \dots, \beta_q)$.

Using L the usual backward operator such as $LX_t = X_{t-1}$, $(1 - \sum_{j=1}^q \beta_j L^j)^{-1}$ exists and simple computations

imply for $t \in \mathbb{Z}$:

$$\begin{aligned}\sigma_t^\delta &= \left(1 - \sum_{j=1}^q \beta_j L^j\right)^{-1} \left[\omega + \sum_{i=1}^p \alpha_i (1 - \gamma_i)^\delta (\max(X_{t-i}, 0))^\delta + \alpha_i (1 + \gamma_i)^\delta (-\min(X_{t-i}, 0))^\delta \right] \\ &= b_0 + \sum_{i \geq 1} b_i^+ (\max(X_{t-i}, 0))^\delta + \sum_{i \geq 1} b_i^- (\max(-X_{t-i}, 0))^\delta.\end{aligned}$$

where $b_0 = w(1 - \sum_{j=1}^q \beta_j)^{-1}$ and the coefficients $(b_i^+, b_i^-)_{i \geq 1}$ are defined by the recursion relations

$$\begin{cases} b_i^+ = \sum_{k=1}^q \beta_k b_{i-k}^+ + \alpha_i (1 - \gamma_i)^\delta & \text{with } \alpha_i (1 - \gamma_i) = 0 \text{ for } i > p \\ b_i^- = \sum_{k=1}^q \beta_k b_{i-k}^- + \alpha_i (1 + \gamma_i)^\delta & \text{with } \alpha_i (1 + \gamma_i) = 0 \text{ for } i > p \end{cases} \quad (3.5)$$

with $b_i^+ = b_i^- = 0$ for $i \leq 0$. As a consequence, for APARCH model, $f_\theta^t \equiv 0$ and $M_\theta^t = \sigma_t$. It is clear that $\alpha_j^{(0)}(f, \Theta) = 0$ and simple computations imply $\alpha_j^{(0)}(M, \Theta) = \sup_{\theta \in \Theta} \max(|b_j^+(\theta)|^{1/\delta}, |b_j^-(\theta)|^{1/\delta})$. Therefore $A_0(f, \Theta)$ holds and $\sum_{j=1}^q \beta_j < 1$ implies that a sequence defined by $u_n = \sum_{k=1}^q \beta_k u_{n-k}$ for n large enough is such as $(u_n)_{n \in \mathbb{N}}$ is an exponentially decreasing sequence and therefore $A_0(M, \Theta)$ holds. Thus for $r \geq 1$, the stationarity set $\Theta(r)$ is defined by

$$\Theta(r) = \left\{ \theta \in \mathbb{R}^{2p+q+2} \mid \left(\mathbb{E}[|\zeta_0|^r] \right)^{1/r} \sum_{j=1}^{\infty} \max(|b_j^+|^{1/\delta}, |b_j^-|^{1/\delta}) < 1 \right\}. \quad (3.6)$$

Now the strong consistency and asymptotic normality of the Laplacian-QMLE for APARCH models can be established (see the proof in Section 5):

Proposition 3.1. *Assume that X is a stationary solution of (3.4) with $\theta_0 \in \Theta$ where Θ is a compact subset of $\Theta(r)$ defined in (3.6). Then,*

1. *If $r = 1$, then $\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{a.s.} \theta_0$.*
2. *If $r = 2$, and if Γ_M defined in (5.21) is a definite positive symmetric matrix, then*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}_{2p+q+2}(0, (\sigma_\zeta^2 - 1) \Gamma_M^{-1}).$$

To our knowledge, this is the first statement the asymptotic properties of Laplacian-QMLE for APARCH processes.

2/ ARMA-GARCH processes. ARMA(p, q)-GARCH(p', q') processes have been introduced by Ding et al. (1993) and Ling and McAleer (2003) as the solution of the system of equations

$$\begin{cases} P_\theta(L) X_t = Q_\theta(L) \varepsilon_t, \\ \varepsilon_t = \sigma_t \zeta_t, \quad \text{with } \sigma_t^2 = c_0 + \sum_{i=1}^{p'} c_i \varepsilon_{t-i}^2 + \sum_{i=1}^{q'} d_i \sigma_{t-i}^2 \end{cases} \quad (3.7)$$

where

- $c_0 > 0$, $c_i \geq 0$ for $i = 1, \dots, p'$, $d_i \geq 0$ for $i = 1, \dots, q'$, $\sum_{i=1}^{q'} d_i < 1$ and $c_{p'}, d_{q'}$ positive;
- $P_\theta(x) = 1 - a_1 x - \dots - a_p x^p$ and $Q_\theta(x) = 1 - b_1 x - \dots - b_q x^q$ are coprime polynomials with $\sum_{i=1}^p |a_i| < 1$ and $\sum_{i=1}^p |b_i| < 1$.

Let $\theta = (c_0, c_1, \dots, c_{p'}, d_1, \dots, d_{q'}, a_1, \dots, a_p, b_1, \dots, b_q)$. We are going to use Lemma 2.1. Since (ε_t) is supposed to be a GARCH(p', q'), then $f_\theta^\varepsilon = 0$ and $M_\theta^\varepsilon = ((1 - \sum_{j=1}^{q'} d_j L^j)^{-1} (c_0 + c_1 \varepsilon_{t-1}^2 + \dots + c_{p'} \varepsilon_{t-p'}^2))^{1/2}$ and direct computations imply that the Lipschitz coefficients of (ε_t) are such as $\alpha_j^{(0)}(f^\varepsilon, \{\theta_0\}) = 0$ and $\alpha_j^{(0)}(M^\varepsilon, \{\theta_0\}) = |\beta_j|$ with $(1 + \sum_{j=1}^{\infty} \beta_j x^j)(1 - \sum_{j=1}^{q'} d_j x^j) = \sum_{j=0}^{p'} c_j x^j$. Therefore $(\alpha_j^{(0)}(f^\varepsilon, \{\theta_0\}))_j$ and $(\alpha_j^{(0)}(M^\varepsilon, \{\theta_0\}))_j$ are EDS (see for instance Berkes and Horváth (2004)). Thus $(A_0(f^\varepsilon, \{\theta_0\}))$ and

$(A_0(M^\varepsilon, \{\theta_0\}))$ hold.

Considering the ARMA part and denoting (ψ_j) such as $(1 + \sum_{j=1}^\infty \psi_j x^j)(1 - \sum_{j=1}^\infty a_j x^j) = (1 - \sum_{j=1}^\infty b_j x^j)$, then from Lemma 2.1 we deduce that:

$$\begin{cases} \alpha_j^{(0)}(f, \{\theta_0\}) &= |\psi_j| \\ \alpha_j^{(0)}(M, \{\theta_0\}) &\leq \sum_{k=1}^j |\psi_k| \times |\beta_{j-k}| \end{cases}.$$

Then we deduce that $(\alpha_j^{(0)}(f, \{\theta_0\}))_j$ and $(\alpha_j^{(0)}(M, \{\theta_0\}))_j$ are EDS, $(A_0(f, \{\theta_0\}))$ and $(A_0(M, \{\theta_0\}))$ hold, and X is a.s. a solution of (1.1) for θ included in the r -order stationarity set $\Theta(r)$ defined by

$$\Theta(r) = \left\{ \theta \in \mathbb{R}^{p+q+p'+q'+1} \mid \sum_{i=1}^\infty |\psi_i(\theta)| + (\mathbb{E}[|\zeta_0|^r])^{1/r} \sum_{j=1}^\infty \sum_{k=1}^j |\psi_k| \times |\beta_{j-k}| < 1 \right\}. \quad (3.8)$$

Now the strong consistency and asymptotic normality of the Laplacian-QMLE for ARMA-GARCH processes can be established:

Proposition 3.2. *Assume that X is a stationary solution of (3.7) with $\theta_0 \in \Theta$ where Θ is a compact subset of $\Theta(r)$ defined in (3.8). Then,*

1. *If $r = 1$, then $\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{a.s.} \theta_0$.*
2. *If $r = 2$, and if Γ_f and Γ_M defined in (5.21) are definite positive symmetric matrix, then the asymptotic normality (3.3) of $\hat{\theta}_n$ holds.*

This result is a new one and extends the previous results already obtained with Gaussian-QMLE for such processes (see for instance, Ling and McAleer (2003) and Bardet and Wintenberger (2009)).

3/ ARMA-ARCH(∞) processes. ARMA(p, q)-ARCH(∞) processes are a natural extension of ARMA-GARCH processes. They are the solution of the system of equations

$$\begin{cases} P_\theta(L) X_t = Q_\theta(L) \varepsilon_t, \\ \varepsilon_t = \sigma_t \zeta_t, \quad \text{with} \quad \sigma_t^2 = c_0 + \sum_{i=1}^\infty c_i \varepsilon_{t-i}^2 \end{cases} \quad (3.9)$$

where

- $c_0 > 0$, $c_i \geq 0$ for $i \geq 1$;
- $P_\theta(x) = 1 - a_1 x - \dots - a_p x^p$ and $Q_\theta(x) = 1 - b_1 x - \dots - b_q x^q$ are coprime polynomials with $\sum_{i=1}^p |a_i| < 1$ and $\sum_{i=1}^q |b_i| < 1$.

ARCH(∞) processes were introduced by Robinson (1991) and the asymptotic properties of Gaussian-QMLE were studied in Robinson and Zaffaroni (2006), Straumann and Mikosch (2006) or Bardet and Wintenberger (2009). Hence, we assume that there exists $\beta = (\beta_1, \dots, \beta_m)$ such as for all $i \in \mathbb{N}$, $c_i = c(i, \beta)$, with $c(\cdot)$ a known function. Let $\theta = (\beta, a_1, \dots, a_p, b_1, \dots, b_q)$. We are going to use Lemma 2.1. Since (ε_t) is supposed to be an ARCH(∞), then $f_\theta^\varepsilon = 0$ and $M_\theta^\varepsilon = (c(0, \beta) + \sum_{i=1}^\infty c(i, \beta) \varepsilon_{t-i}^2)^{1/2}$ and direct computations imply that the Lipschitz coefficients of (ε_t) are such as $\alpha_j^{(0)}(f^\varepsilon, \{\theta_0\}) = 0$ and $\alpha_j^{(0)}(M^\varepsilon, \{\theta_0\}) = c(j, \beta_0)$. Therefore we assume that there exists $\ell > 1$ such as

$$c(j, \beta_0) = \mathcal{O}(j^{-\ell}) \text{ when } j \rightarrow \infty. \quad (3.10)$$

Thus $(A_0(f^\varepsilon, \{\theta_0\}))$ and $(A_0(M^\varepsilon, \{\theta_0\}))$ hold.

Considering the ARMA part and denoting (ψ_j) such as $(1 + \sum_{j=1}^\infty \psi_j x^j)(1 - \sum_{j=1}^\infty a_j x^j) = (1 - \sum_{j=1}^\infty b_j x^j)$, then from Lemma 2.1 we deduce that:

$$\begin{cases} \alpha_j^{(0)}(f, \{\theta_0\}) &= |\psi_j| \\ \alpha_j^{(0)}(M, \{\theta_0\}) &\leq \sum_{k=1}^j |\psi_k| \times c(j, \beta_0) \end{cases}.$$

Then we deduce that $(\alpha_j^{(0)}(f, \{\theta_0\}))_j$ is EDS and $(\alpha_j^{(0)}(M, \{\theta_0\}))_j = \mathcal{O}(j^{-\ell})$. Then $(A_0(f, \{\theta_0\}))$ and $(A_0(M, \{\theta_0\}))$ hold, and X is a.s. a solution of (1.1) for θ included in the r -order stationarity set $\Theta(r)$ defined by

$$\Theta(r) = \left\{ \theta \in \mathbb{R}^{p+q+m} \mid \sum_{i=1}^{\infty} |\psi_i(\theta)| + (\mathbb{E}[|\zeta_0|^r])^{1/r} \sum_{j=1}^{\infty} \sum_{k=1}^j |\psi_k| \times c(j, \beta_0) < 1 \right\}. \quad (3.11)$$

Now the strong consistency and asymptotic normality of the Laplacian-QMLE for ARMA-ARCH(∞) processes can be established:

Proposition 3.3. *Assume that X is a stationary solution of (3.9) where (3.10) holds and with $\theta_0 \in \Theta$ where Θ is a compact subset of $\Theta(r)$ defined in (3.11). Then,*

1. *If $r \geq 1$ and $\ell \geq 2/\min(r, 2)$, then $\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{a.s.} \theta_0$.*
2. *If $r = 2$, $\ell > 1$ and if $\partial_{\beta}^i c(j, \beta) = \mathcal{O}(j^{-\ell})$ for $i = 1, 2$, and if Γ_f and Γ_M defined in (5.21) are definite positive symmetric matrix, then the asymptotic normality (3.3) of $\hat{\theta}_n$ holds.*

This result is a new one. Note that $\ell > 1$ and $r = 2$ is required for the asymptotic normality of Laplacian-QMLE while $r = 4$ and $\ell > 2$ is required for Gaussian-QMLE for such processes (see for instance Bardet and Wintenberger (2009)). This confers a clear advantage to Laplacian-QMLE.

4/ ARMA-APARCH processes. The ARMA(p, q)-APARCH(p', q') processes have been also introduced by Ding et al. (1993) as the solutions of the equations

$$\begin{cases} P_{\theta}(L) X_t = Q_{\theta}(L) \varepsilon_t, \\ \varepsilon_t = \sigma_t \zeta_t, \text{ with } \sigma_t^{\delta} = \omega + \sum_{i=1}^{p'} \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^{\delta} + \sum_{j=1}^{q'} \beta_j \sigma_{t-j}^{\delta} \end{cases} \quad (3.12)$$

where:

- $\delta \geq 1$, $\omega > 0$, $-1 < \gamma_i < 1$ and $\alpha_i \geq 0$ for $i = 1, \dots, p' - 1$, $\beta_j \geq 0$ for $j = 1, \dots, q' - 1$, $\alpha_{p'}$, $\beta_{q'}$ positive real numbers and $\sum_{j=1}^{p'} \alpha_j < 1$;
- $P_{\theta}(x) = 1 - a_1 x - \dots - a_p x^p$ and $\Psi_{\theta}(x) = 1 - b_1 x - \dots - b_q x^q$ are coprime polynomials with $\sum_{i=1}^p |a_i| < 1$ and $\sum_{i=1}^q |b_i| < 1$.

Let $\theta = (\delta, \omega, \alpha_1, \dots, \alpha_{p'}, \gamma_1, \dots, \gamma_{p'}, \beta_1, \dots, \beta_{q'}, a_1, \dots, a_p, b_1, \dots, b_q)$. Then, as for ARMA-GARCH processes, we are going to use Lemma 2.1. Thanks to the computations realized for APARCH processes, we obtain $\alpha_j^{(0)}(f^{\varepsilon}, \{\theta_0\}) = 0$ and $\alpha_j^{(0)}(M^{\varepsilon}, \{\theta_0\}) = \max(|b_j^+|^{1/\delta}, |b_j^-|^{1/\delta})$ with $(b_i^+, b_i^-)_{i \geq 1}$ defined in (3.5).

Then, we have

$$\begin{cases} \alpha_j^{(0)}(f, \{\theta_0\}) & \leq |\psi_j| \\ \alpha_j^{(0)}(M, \{\theta_0\}) & \leq \sum_{k=1}^j |\psi_k| \times \max(|b_{j-k}^+|^{1/\delta}, |b_{j-k}^-|^{1/\delta}) \end{cases}.$$

(ψ_j) such as $(1 + \sum_{j=1}^{\infty} \psi_j x^j)(1 - \sum_{j=1}^{\infty} a_j x^j) = (1 - \sum_{j=1}^{\infty} b_j x^j)$. From Lemma 2.1, $(A_0(f, \Theta))$ and $(A_0(M, \Theta))$ hold since $(\alpha_j^{(0)}(f^{\varepsilon}, \{\theta_0\}))_j = 0$ and $(\alpha_j^{(0)}(M^{\varepsilon}, \{\theta_0\}))_j$ are EDS. As a consequence, for $r \geq 1$, the stationarity set $\Theta(r)$ is defined by

$$\Theta(r) = \left\{ \theta \in \mathbb{R}^{p+q+p'+q'+2} \mid \sum_{j=1}^{\infty} |\psi_j| + (\mathbb{E}[|\zeta_0|^r])^{1/r} \sum_{j=1}^{\infty} \sum_{k=1}^j |\psi_k| \times \max(|b_{j-k}^+|^{1/\delta}, |b_{j-k}^-|^{1/\delta}) < 1 \right\}.$$

Now, we are able to provide the asymptotic properties of QMLE for ARMA-APARCH models.

Proposition 3.4. *Assume that X is a stationary solution of (3.12) with $\theta_0 \in \Theta$ where Θ is a compact subset of $\Theta(r)$ defined in (3.8). Then,*

1. *If $r = 1$, then $\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{a.s.} \theta_0$.*

2. If $r = 2$, and if Γ_f and Γ_M defined in (5.21) are definite positive symmetric matrix, then the asymptotic normality (3.3) of $\hat{\theta}_n$ holds.

This result is stated for the first time for Laplacian-QMLE. The case of Gaussian-QMLE for ARMA-APARCH could be also obtained following the previous decomposition and the paper Bardet and Wintenberger (2009). Once again, the asymptotic normality of Laplacian-QMLE only requires $r = 2$ while this requires $r = 4$ for Gaussian-QMLE.

4. Numerical Results

To illustrate the asymptotic results stated previously, we realized Monte-Carlo experiments on the behavior of Laplacian-QMLE (denoted $\hat{\theta}_n^{LQL}$) for several time series models, sample sizes and probability distributions. A comparison with the results obtained by Gaussian QMLE (denoted $\hat{\theta}_n^{GQL}$) is also proposed.

More precisely, the considered probability distributions of (ζ_t) are:

- Centred Gaussian distribution denoted $\mathcal{N}$;
- Centred Laplacian distribution denoted $\mathcal{L}$;
- Centred Uniform distribution denoted $\mathcal{U}$;
- Centred Student distribution with 3 freedom degrees, denoted t_3 ;
- Normalized centred Gaussian mixture with probability distribution $0.05 * \mathcal{N}(-2, 0.16) + \mathcal{N}(0, 1) + 0.05 * \mathcal{N}(2, 0.16)$ and denoted $\mathcal{M}$.

All these probability distributions are normalized such as $E[|\zeta_0|] = 1$, required for Laplacian-QMLE. For using Gaussian-QMLE requiring $\sigma_\zeta^2 = 1$, it is necessary to consider the model with $M'_\theta = \frac{E[|\zeta_0|]}{\sigma_\zeta} M_\theta$ instead of M_θ .

Several models of time series satisfying (1.1) and the assumptions of Theorem 3.1 and 3.3 are considered:

- a ARMA(1, 1) process defined by $X_t = \phi X_{t-1} + \zeta_t + \theta \zeta_{t-1}$ with $\phi = 0.4$ and $\theta = 0.6$;
- a ARCH(1) process defined by $X_t = \zeta_t \sqrt{\omega + \alpha X_{t-1}^2}$ with $\omega = 0.4$ and $\alpha = 0.2$;
- a GARCH(1, 1) process defined by $X_t = \zeta_t \sigma_t$ where $\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta \sigma_{t-1}^2$ with $\alpha_0 = 0.2$, $\alpha_1 = 0.4$ and $\beta = 0.2$;
- a ARMA(1, 1)-GARCH(1, 1) process defined by $X_t = \phi X_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$ where $\varepsilon_t = \zeta_t \sigma_t$ and $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$ with $\phi = 0.4$, $\theta = 0.6$, $\alpha_0 = 0.2$, $\alpha_1 = 0.4$ and $\beta = 0.1$;
- a ARMA(1, 1)-APARCH(1, 1) process defined by $X_t = \phi X_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$ where $\varepsilon_t = \zeta_t \sigma_t$ and $\sigma_t^\delta = \alpha_0 + \alpha_1 (|\varepsilon_{t-1}| - \gamma \varepsilon_{t-1})^\delta + \beta \sigma_{t-1}^\delta$ and $\phi = 0.4$, $\theta = 0.6$, $\alpha_0 = 0.2$, $\alpha_1 = 0.4$, $\gamma = 0.5$, $\beta = 0.1$ and $\delta = 1.2$.

Hence we computed the root-mean-square error (RMSE) from 1000 independent replications of $\hat{\theta}_n^{LQL}$ and $\hat{\theta}_n^{GQL}$ for those processes and the results are presented in Table 1 on page 10 and 2 on page 10.

Conclusion of the numerical results: On the one hand, it is clear that the RMSE decreases as the sample size increases, which validates the theoretical results (consistency of the estimators). On the other hand, Table 1 and 2 show that the Laplacian-QMLE provides more accurate estimation than the Gaussian-QMLE for several types of noise, except of course in the case of a Gaussian distribution (even in this case the RSME of both the estimators are almost the same).

5. Proofs

Proof of Lemma 2.1. First, as X is a stationary process and the ARMA(p, q) process is causal invertible then $\tilde{X}$ is also a stationary process (the coefficients of Λ_{β_0} are EDS). Moreover, it is well known that $(\psi_j(\beta_0))_{j \in \mathbb{N}}$

TABLE 1
Root Mean Square Error of the components of $\hat{\theta}_n^{LQL}$ and $\hat{\theta}_n^{GQL}$ for ARMA(1,1), ARCH(1) and GARCH(1,1) processes.

		$\mathcal{L}$		$\mathcal{N}$		t_3		$\mathcal{U}$		$\mathcal{M}$		
	n	$\hat{\theta}_n^{GQL}$	$\hat{\theta}_n^{LQL}$	$\hat{\theta}_n^{GQL}$	$\hat{\theta}_n^{LQL}$	$\hat{\theta}_n^{GQL}$	$\hat{\theta}_n^{LQL}$	$\hat{\theta}_n^{GQL}$	$\hat{\theta}_n^{LQL}$	$\hat{\theta}_n^{GQL}$	$\hat{\theta}_n^{LQL}$	
ARMA(1,1)	θ	100	0.106	0.091	0.114	0.117	0.113	0.090	0.112	0.059	0.110	0.078
		1000	0.031	0.024	0.032	0.032	0.036	0.027	0.031	0.014	0.031	0.023
		5000	0.014	0.010	0.014	0.015	0.016	0.011	0.016	0.012	0.013	0.010
	ϕ	100	0.119	0.102	0.121	0.128	0.123	0.102	0.120	0.067	0.121	0.090
		1000	0.037	0.028	0.036	0.036	0.040	0.030	0.036	0.017	0.036	0.027
		5000	0.016	0.012	0.014	0.016	0.017	0.013	0.016	0.007	0.014	0.006
ARCH(1)	ω	100	0.068	0.061	0.048	0.049	0.254	0.085	0.035	0.025	0.062	0.052
		1000	0.020	0.018	0.015	0.015	0.134	0.049	0.011	0.016	0.036	0.018
		5000	0.010	0.009	0.006	0.006	0.115	0.044	0.005	0.015	0.031	0.008
	α	100	0.161	0.155	0.141	0.142	0.979	0.418	0.102	0.064	0.484	0.423
		1000	0.063	0.058	0.043	0.043	0.852	0.169	0.029	0.033	0.157	0.133
		5000	0.016	0.014	0.012	0.012	0.378	0.109	0.013	0.031	0.087	0.062
GARCH(1,1)	α_0	100	0.112	0.105	0.095	0.100	0.211	0.126	0.081	0.047	0.134	0.114
		1000	0.036	0.032	0.028	0.028	0.098	0.058	0.023	0.018	0.066	0.051
		5000	0.016	0.014	0.012	0.012	0.055	0.043	0.010	0.015	0.040	0.023
	α_1	100	0.162	0.157	0.149	0.150	0.453	0.364	0.115	0.070	0.507	0.429
		1000	0.061	0.056	0.449	0.449	0.333	0.150	0.030	0.033	0.160	0.136
		5000	0.029	0.026	0.020	0.020	0.193	0.095	0.013	0.030	0.086	0.058
	β	100	0.225	0.209	0.188	0.190	0.499	0.429	0.163	0.105	0.483	0.390
		1000	0.060	0.055	0.051	0.051	0.285	0.174	0.044	0.022	0.170	0.169
		5000	0.027	0.024	0.022	0.022	0.180	0.075	0.019	0.009	0.072	0.075

TABLE 2
Root Mean Square Error of the components of $\hat{\theta}_n^{LQL}$ and $\hat{\theta}_n^{GQL}$ for ARMA(1,1)-GARCH(1,1) and ARMA(1,1)-APARCH(1,1) processes.

			$\mathcal{L}$		$\mathcal{N}$		t_3		$\mathcal{U}$		$\mathcal{M}$	
n			$\hat{\theta}_n^{GQL}$	$\hat{\theta}_n^{LQL}$	$\hat{\theta}_n^{GQL}$	$\hat{\theta}_n^{LQL}$	$\hat{\theta}_n^{GQL}$	$\hat{\theta}_n^{LQL}$	$\hat{\theta}_n^{GQL}$	$\hat{\theta}_n^{LQL}$	$\hat{\theta}_n^{GQL}$	$\hat{\theta}_n^{LQL}$
ARMA(1,1) -GARCH(1,1)	θ	100	0.120	0.097	0.107	0.107	0.121	0.098	0.097	0.067	0.123	0.087
		1000	0.035	0.024	0.028	0.028	0.048	0.029	0.024	0.015	0.035	0.026
		5000	0.016	0.010	0.012	0.012	0.023	0.012	0.015	0.011	0.011	0.007
	ϕ	100	0.135	0.109	0.117	0.119	0.141	0.116	0.110	0.077	0.132	0.102
		1000	0.044	0.030	0.033	0.033	0.063	0.035	0.029	0.023	0.053	0.046
		5000	0.020	0.014	0.015	0.015	0.029	0.015	0.013	0.012	0.019	0.014
	α_0	100	0.104	0.096	0.085	0.084	0.158	0.129	0.073	0.055	0.131	0.116
		1000	0.031	0.028	0.025	0.025	0.241	0.060	0.021	0.019	0.053	0.046
		5000	0.014	0.012	0.010	0.010	0.052	0.042	0.009	0.016	0.036	0.019
	α_1	100	0.179	0.177	0.166	0.167	0.469	0.385	0.134	0.107	0.494	0.405
		1000	0.064	0.060	0.045	0.045	0.328	0.161	0.031	0.046	0.160	0.137
		5000	0.031	0.027	0.020	0.020	0.182	0.096	0.013	0.038	0.090	0.062
	β	100	0.302	0.269	0.252	0.233	0.604	0.497	0.217	0.164	0.553	0.472
		1000	0.057	0.051	0.051	0.051	0.312	0.187	0.045	0.049	0.165	0.170
		5000	0.025	0.022	0.020	0.020	0.199	0.073	0.062	0.066	0.019	0.025
ARMA(1,1) -APARCH(1,1)	θ	100	0.110	0.086	0.096	0.101	0.112	0.090	0.097	0.068	0.125	0.091
		1000	0.029	0.021	0.023	0.024	0.031	0.021	0.022	0.014	0.033	0.024
		5000	0.013	0.008	0.010	0.010	0.014	0.009	0.010	0.006	0.015	0.011
	ϕ	100	0.138	0.114	0.121	0.126	0.128	0.107	0.111	0.086	0.146	0.107
		1000	0.040	0.027	0.032	0.032	0.041	0.028	0.029	0.026	0.043	0.030
		5000	0.018	0.012	0.012	0.012	0.020	0.012	0.013	0.014	0.019	0.013
	ω	100	0.198	0.192	0.199	0.210	0.254	0.262	0.221	0.170	0.290	0.272
		1000	0.079	0.067	0.056	0.056	0.226	0.218	0.044	0.045	0.142	0.129
		5000	0.033	0.028	0.025	0.025	0.209	0.207	0.017	0.029	0.061	0.056
	α	100	0.206	0.201	0.183	0.184	0.464	0.449	0.167	0.131	0.352	0.327
		1000	0.060	0.053	0.041	0.041	0.447	0.432	0.029	0.043	0.143	0.134
		5000	0.025	0.023	0.018	0.018	0.421	0.414	0.012	0.027	0.071	0.059
	γ	100	0.413	0.386	0.346	0.356	0.439	0.426	0.310	0.233	0.613	0.601
		1000	0.105	0.094	0.071	0.070	0.101	0.092	0.057	0.041	0.217	0.220
		5000	0.042	0.039	0.029	0.029	0.045	0.038	0.024	0.018	0.086	0.089
	β	100	0.297	0.282	0.255	0.238	0.312	0.288	0.186	0.145	0.476	0.468
		1000	0.074	0.067	0.058	0.058	0.074	0.066	0.043	0.033	0.151	0.150
		5000	0.033	0.031	0.024	0.024	0.034	0.029	0.018	0.012	0.061	0.063
	δ	100	0.732	0.732	0.712	0.717	0.740	0.739	0.703	0.613	0.830	0.815
		1000	0.402	0.352	0.296	0.296	0.394	0.338	0.235	0.290	0.542	0.534
		5000	0.170	0.147	0.132	0.131	0.169	0.145	0.092	0.168	0.251	0.262

is EDS. Then we have:

$$\begin{aligned}
\tilde{X}_t &= \Lambda_{\beta_0}(L) \left(M_{\theta_0}((X_{t-i})_{i \geq 1}) \zeta_t + f_{\theta_0}((X_{t-i})_{i \geq 1}) \right) \\
\tilde{X}_t + \sum_{j=1}^{\infty} \psi_j(\beta_0) \tilde{X}_{t-j} &= M_{\theta_0}((\Lambda_{\beta_0}^{-1}(L) \tilde{X}_{t-i})_{i \geq 1}) \zeta_t + f_{\theta_0}((\Lambda_{\beta_0}^{-1}(L) \tilde{X}_{t-i})_{i \geq 1}) \\
\tilde{X}_t &= \tilde{M}_{\tilde{\theta}_0}((\tilde{X}_{t-i})_{i \geq 1}) \zeta_t + \tilde{f}_{\tilde{\theta}_0}((\tilde{X}_{t-i})_{i \geq 1}) \\
\text{with } \begin{cases} \tilde{M}_{\tilde{\theta}_0}((x_{t-i})_{i \geq 1}) &= M_{\theta_0}((\Lambda_{\beta_0}^{-1}(L) x_{t-i})_{i \geq 1}) \\ \tilde{f}_{\tilde{\theta}_0}((x_{t-i})_{i \geq 1}) &= f_{\theta_0}((\Lambda_{\beta_0}^{-1}(L) x_{t-i})_{i \geq 1}) - \sum_{j=1}^{\infty} \psi_j(\beta_0) x_{t-j} \end{cases} \quad (5.1)
\end{aligned}$$

Finally, for $t \in \mathbb{Z}$,

$$\begin{aligned}
|\tilde{f}_{\tilde{\theta}_0}((x_{t-i})_{i \geq 1}) - \tilde{f}_{\tilde{\theta}_0}((y_{t-i})_{i \geq 1})| &\leq \sum_{j=1}^{\infty} \alpha_j^{(0)}(f, \{\theta_0\}) |(\Lambda_{\beta_0}^{-1}(L) x_{t-j-i})_{i \geq 1} - (\Lambda_{\beta_0}^{-1}(L) y_{t-j-i})_{i \geq 1}| \\
&\quad + |\psi_j(\beta_0)| |x_{t-j} - y_{t-j}| \\
&\leq \sum_{j=1}^{\infty} \alpha_j^{(0)}(f, \{\theta_0\}) \left| \sum_{k=0}^{\infty} |\psi_k(\beta_0)| |x_{t-k-j} - y_{t-k-j}| + |\psi_j(\beta_0)| |x_{t-j} - y_{t-j}| \right| \\
&\leq \sum_{j=1}^{\infty} \left(|\psi_j(\beta_0)| + \sum_{k=1}^j \alpha_k^{(0)}(f, \{\theta_0\}) \psi_{j-k}(\beta_0) \right) |x_{t-j} - y_{t-j}| \\
\Rightarrow \alpha_j^{(0)}(\tilde{f}, \{\tilde{\theta}_0\}) &\leq |\psi_j(\beta_0)| + \sum_{k=1}^j \alpha_k^{(0)}(f, \{\theta_0\}) |\psi_{j-k}(\beta_0)|. \quad (5.2)
\end{aligned}$$

Moreover, we also have:

$$\begin{aligned}
|\tilde{M}_{\tilde{\theta}_0}((x_{t-i})_{i \geq 1}) - \tilde{M}_{\tilde{\theta}_0}((y_{t-i})_{i \geq 1})| &\leq \sum_{j=1}^{\infty} \alpha_j^{(0)}(M, \{\theta_0\}) |(\Lambda_{\beta_0}^{-1}(L) x_{t-j-i})_{i \geq 1} - (\Lambda_{\beta_0}^{-1}(L) y_{t-j-i})_{i \geq 1}| \\
\Rightarrow \alpha_j^{(0)}(\tilde{M}, \{\tilde{\theta}_0\}) &\leq \sum_{k=1}^j \alpha_k^{(0)}(M, \{\theta_0\}) |\psi_{j-k}|. \quad (5.3)
\end{aligned}$$

The same kinds of computations could also be done by considering the first and second derivatives of $\tilde{f}$ and $\tilde{M}$ with respect to $\tilde{\theta}$. Note, and this is important, that the first and second derivatives of Λ_{β}^{-1} with respect to $\tilde{\theta}$ are also EDS. Finally,

- if when $j \rightarrow \infty$, $\alpha_j^{(0)}(K, \{\theta_0\}) = O(j^{-\beta})$ with $\beta > 1$ and $\psi_j = O(\rho^j)$ with $0 \leq \rho < 1$, then there exists $C > 0$ such as $\sum_{k=1}^j \alpha_k^{(0)}(K, \{\theta_0\}) |\psi_{j-k}| \leq C \sum_{k=1}^j k^{-\beta} \rho^{j-k} \sim -C(\log \rho)^{-1} j^{-\beta}$ and therefore $\alpha_j^{(0)}(\tilde{K}, \{\tilde{\theta}_0\}) = O(j^{-\beta})$.
- if when $j \rightarrow \infty$, $\alpha_j^{(0)}(K, \{\theta_0\}) = O(r^j)$ with $0 \leq r < 1$ and $\psi_j = O(\rho^j)$ with $0 \leq \rho < 1$, then there exists $C > 0$ such as $\sum_{k=1}^j \alpha_k^{(0)}(K, \{\theta_0\}) |\psi_{j-k}| \leq C \sum_{k=1}^j r^{-k} \rho^{j-k} = O(j \max(r, \rho)^j)$ and therefore $\alpha_j^{(0)}(\tilde{K}, \{\tilde{\theta}_0\})$ is EDS.

The same kind of computation can be also done for $(\alpha_j^{(i)}(\tilde{K}, \{\tilde{\theta}_0\}))_j$ since the derivatives and second-derivatives of $\Lambda_{\beta_0}^{-1}$ with respect to β and therefore to $\tilde{\theta}$ are also EDS. $\square$

Now we remind two lemmas already proved in [Bardet and Wintenberger \(2009\)](#):

Lemma 5.1. *Assume that $\theta_0 \in \Theta(r)$ for $r \geq 1$ and X is the causal stationary solution of the equation (1.1). If $(A_0(K, \Theta))$ holds (with $K = f$ or $K = M$) then $K_{\theta}^t \in L^r(\mathcal{C}(\Theta, R^m))$ and there exists $C > 0$ not depending on t such that*

$$\mathbb{E}[\|\hat{K}_{\theta}^t - K_{\theta}^t\|_{\Theta}^r] \leq C \mathbb{E}[|X_0|^r] \left(\sum_{j \geq t} \alpha_j(K, \Theta) \right)^r \quad \text{for all } t \in N^*. \quad (5.4)$$

Lemma 5.2. Let $\mathcal{D}^{(2)}(\Theta)$ denote the Banach space of 2 times continuously differentiable functions on Θ equipped with the uniform norm

$$\|h\|_{2,\Theta} = \|h\|_{\Theta} + \left\| \frac{\partial h}{\partial \theta} \right\|_{\Theta} + \left\| \frac{\partial^2 h}{\partial \theta \partial \theta'} \right\|_{\Theta}.$$

Let $\theta_0 \in \Theta(r)$ ($r \geq 1$) and assume that for $i = 0, 1, 2$, $(A_i(f, \Theta))$ and $(A_i(M, \Theta))$ hold. Then $f_{\theta}^t \in \mathbb{L}^r(\mathcal{D}^{(2)}(\Theta))$ and $M_{\theta}^t \in \mathbb{L}^r(\mathcal{D}^{(2)}(\Theta))$.

Now, we begin with the proofs of Theorem 3.1, 3.2 and 3.3.

Proof of Theorem 3.1. The proof of the theorem is divided into two parts and follows the same procedure than in Bardet and Wintenberger (2009). In (i), a uniform (on Θ) strong law of large numbers satisfied by $\frac{1}{n} \widehat{L}_n(\theta)$ converging to $L(\theta) := -E[q_0(\theta)]$ is established. In (ii), it is proved that $L(\theta)$ admits a unique maximum in θ_0 . Those two vconditions lead to the strong consistency of $\widehat{\theta}_n$ (from Jeantreau (1998)).

(i) In the same way and for the same reason in the proof of Theorem 1 of Bardet and Wintenberger (2009), the uniform strong law of large numbers satisfied by the sample mean of $(\widehat{q}_t)_{t \in N^*}$ (defined in (2.3)) is implied by establishing $E[\|q_t(\theta)\|_{\Theta}] < \infty$. But from Lemma 5.1, for all $t \in \mathbb{Z}$,

$$\begin{aligned} |q_t(\theta)| &= |(M_{\theta}^t)^{-1}|X_t - f_{\theta}^t| + \log(M_{\theta}^t)| \\ &\leq \frac{|X_t - f_{\theta}^t(\theta)|}{\underline{M}} + |\log(\underline{M})| + M_{\theta}^t \\ \implies \sup_{\theta \in \Theta} |q_t(\theta)| &\leq \frac{1}{\underline{M}}(|X_t| + \|f_{\theta}^t(\theta)\|_{\Theta}) + |\log(\underline{M})| + \|M_{\theta}^t\|_{\Theta}. \end{aligned}$$

With $r \geq 1$, we have $\forall t \in \mathbb{Z}$, $E[|X_t|] < \infty$ from Proposition 2.1 and $E[\|f_{\theta}^t\|_{\Theta}^r + \|M_{\theta}^t\|_{\Theta}^r] < \infty$ from Lemma 5.1, implying $E[\|f_{\theta}^t\|_{\Theta} + \|M_{\theta}^t\|_{\Theta}] < \infty$. As a consequence, for all $t \in \mathbb{Z}$,

$$E[\|q_t(\theta)\|_{\Theta}] < \infty.$$

Hence, the uniform strong law of large numbers for $(q_t(\theta))$ follows:

$$\left\| \frac{L_n(\theta)}{n} - L(\theta) \right\|_{\Theta} \xrightarrow[n \rightarrow \infty]{a.s.} 0. \quad (5.5)$$

Now, we are going to establish $\frac{1}{n} \widehat{L}_n(\theta) - L_n(\theta) \xrightarrow[n \rightarrow \infty]{a.s.} 0$. Indeed, for all $\theta \in \Theta$ and $t \in N^*$,

$$\begin{aligned} |\widehat{q}_t(\theta) - q_t(\theta)| &\leq |\log(\widehat{M}_{\theta}^t) - \log(M_{\theta}^t) + (\widehat{M}_{\theta}^t)^{-1}|X_t - \widehat{f}_{\theta}^t| - (M_{\theta}^t)^{-1}|X_t - f_{\theta}^t|| \\ &\leq |\log(\widehat{M}_{\theta}^t) - \log(M_{\theta}^t)| + |((\widehat{M}_{\theta}^t)^{-1} - (M_{\theta}^t)^{-1})|X_t - f_{\theta}^t| + (\widehat{M}_{\theta}^t)^{-1}|\widehat{f}_{\theta}^t - f_{\theta}^t| \\ &\leq |\widehat{M}_{\theta}^t - M_{\theta}^t| \underline{M}^{-1} + |\widehat{M}_{\theta}^t - M_{\theta}^t| \underline{M}^{-2} |X_t - f_{\theta}^t| + \underline{M}^{-1} |\widehat{f}_{\theta}^t - f_{\theta}^t| \\ &\leq C(1 + |X_t| + |f_{\theta}^t|)(|\widehat{M}_{\theta}^t - M_{\theta}^t| + |\widehat{f}_{\theta}^t - f_{\theta}^t|), \end{aligned}$$

with $C > 0$. Hence, we have:

$$\|\widehat{q}_t(\theta) - q_t(\theta)\|_{\Theta} \leq C(1 + |X_t| + \|f_{\theta}^t\|_{\Theta})(\|\widehat{M}_{\theta}^t - M_{\theta}^t\|_{\Theta} + \|\widehat{f}_{\theta}^t - f_{\theta}^t\|_{\Theta}).$$

By Corollary 1 of Kounias and Weng (1969), the proof is achieved if there exists $s \in (0, 1]$ such as

$$\sum_{t \geq 1} \frac{1}{t^s} E[\|q_t(\theta) - \widehat{q}_t(\theta)\|_{\Theta}^s] < \infty. \quad (5.6)$$

Let us prove (5.6) with $s = r/2$ when $r \in [1, 2]$.

From Cauchy-Schwarz Inequality and assumptions $A_0(f, \Theta)$ and $A_0(M, \Theta)$,

$$E[\|\widehat{q}_t(\theta) - q_t(\theta)\|_{\Theta}^{r/2}] \leq C(E[(1 + |X_t| + \|f_{\theta}^t\|_{\Theta})^r])^{\frac{1}{2}} (E[(\|\widehat{M}_{\theta}^t - M_{\theta}^t\|_{\Theta} + \|\widehat{f}_{\theta}^t - f_{\theta}^t\|_{\Theta})^r])^{\frac{1}{2}}.$$

Using Lemma 5.1 and previous proved results implying $E[|X_t|^r] < \infty$, $E[\|f_\theta^t\|_\Theta^r + \|M_\theta^t\|_\Theta^r] < \infty$, we obtain

$$\begin{aligned} E[\|\hat{q}_t(\theta) - q_t(\theta)\|_\Theta^{r/2}] &\leq C \left(\sum_{j>t} \alpha_j^{(0)}(f, \Theta) + \alpha_j^{(0)}(M, \Theta) \right)^{\frac{r}{2}} \\ &\leq C t^{-\frac{(\ell-1)r}{2}}, \end{aligned}$$

where the last inequality is obtained from the condition (3.1) of Theorem 3.1.

Hence, we have

$$\sum_{t \geq 1} \frac{1}{t^{r/2}} E[\|\hat{q}_t(\theta) - q_t(\theta)\|_\Theta^{r/2}] \leq A \sum_{t \geq 1} t^{-r\ell/2},$$

which is finite when $r\ell > 2$. When $r \geq 2$, it is sufficient to consider the case $r = 2$. As a consequence, we obtain

$$\frac{1}{n} \sum_{t=1}^n \|\hat{q}_t(\theta) - q_t(\theta)\|_\Theta \xrightarrow[n \rightarrow \infty]{a.s.} 0 \quad \frac{1}{n} \|\hat{L}_n(\theta) - L_n(\theta)\|_\Theta \xrightarrow[n \rightarrow \infty]{a.s.} 0, \quad (5.7)$$

and therefore, using (5.5),

$$\frac{1}{n} \|\hat{L}_n(\theta) - L(\theta)\|_\Theta \xrightarrow[n \rightarrow \infty]{a.s.} 0. \quad (5.8)$$

(ii) Now for $\theta \in \Theta$, we study

$$L(\theta) = -E[q_0(\theta)].$$

which can also be consider as a Kullback-Lieber discrepancy. We have

$$\begin{aligned} L(\theta) &= -E[\log(M_\theta^t) + (M_\theta^t)^{-1} |X_t - f_\theta^t|] \\ &= -E\left[\log(M_\theta^t) + \frac{M_{\theta_0}^t}{M_\theta^t} \left| \zeta_t + \frac{f_{\theta_0}^t - f_\theta^t}{M_{\theta_0}^t} \right| \right]. \end{aligned}$$

Hence, using $E[|\zeta_t|] = 1$, we obtain:

$$\begin{aligned} L(\theta_0) - L(\theta) &= E\left[\log\left(\frac{M_\theta^t}{M_{\theta_0}^t}\right) + \frac{M_{\theta_0}^t}{M_\theta^t} \left| \zeta_t + \frac{f_{\theta_0}^t - f_\theta^t}{M_{\theta_0}^t} \right| - 1\right] \\ &= E\left[\log\left(\frac{M_\theta^t}{M_{\theta_0}^t}\right) - 1 + \frac{M_{\theta_0}^t}{M_\theta^t} E\left[\left| \zeta_t + \frac{f_{\theta_0}^t - f_\theta^t}{M_{\theta_0}^t} \right| \mid (X_{t-k})_{k \geq 1}\right]\right]. \end{aligned}$$

But for ζ_t following a symmetric probability distribution, for any $m \in \mathbb{R}^*$, $E[|\zeta_t + m|] > E[|\zeta_t|] = 1$. Therefore, for $\theta \neq \theta_0$, if $f_\theta \neq f_{\theta_0}$ (else $>$ is replaced by $\geq$),

$$\begin{aligned} L(\theta_0) - L(\theta) &> E\left[\log\left(\frac{M_\theta^t}{M_{\theta_0}^t}\right) - 1 + \frac{M_{\theta_0}^t}{M_\theta^t}\right] \\ &> h\left(\frac{M_\theta^t}{M_{\theta_0}^t}\right), \end{aligned}$$

with $h(x) = \log(x) - 1 + x$. But for any $x \in (0, 1) \cup (1, \infty)$, $h(x) > 0$ and $h(1) = 0$. Therefore if $M_\theta \neq M_{\theta_0}$, $h\left(\frac{M_\theta^t}{M_{\theta_0}^t}\right) > 0$ (> 0 is replaced by $= 0$ if $M_\theta = M_{\theta_0}$). This implies from Condition C3 (Identifiability) that $L(\theta_0) - L(\theta) > 0$ almost surely for all $\theta \in \Theta$, $\theta \neq \theta_0$. Hence a supremum of $L(\theta)$ is only reached for $\theta = \theta_0$ which is the unique maximum. $\square$

Proof of Theorem 3.2. We follow the same scheme of proof than in Davis and Dunsmuir (1997). Hence,

denote

$$\begin{aligned}
S_n &= \sum_{t=1}^n V_{t-1} (|Z_t - n^{-1/2} Y_{t-1}| - |Z_t|) \\
&= -n^{-1/2} \sum_{t=1}^n V_{t-1} Y_{t-1} \operatorname{sgn}(Z_t) \\
&\quad + 2 \sum_{t=1}^n V_{t-1} (n^{-1/2} Y_{t-1} - Z_t) (\mathbf{1}_{0 < Z_t < n^{-1/2} Y_{t-1}} - \mathbf{1}_{n^{-1/2} Y_{t-1} < Z_t < 0}) \\
&= A_n + B_n.
\end{aligned}$$

Since $E[V_{t-1} Y_{t-1} \operatorname{sgn}(Z_t) / \mathcal{F}_{t-1}] = E[\operatorname{sgn}(Z_t)] E[V_{t-1} Y_{t-1}] = 0$ and $E[V_0^2 Y_0^2] < \infty$, we can apply a central limit theorem for stationary martingale difference sequence (see Billingsley (1968)) and

$$A_n \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, E[V_0^2 Y_0^2]). \quad (5.9)$$

Now, considering B_n , define also $W_{nt} = V_{t-1} (n^{-1/2} Y_{t-1} - Z_t) \mathbf{1}_{0 < Z_t < n^{-1/2} Y_{t-1}}$. Using the same arguments as in Davis and Dunsmuir (1997), we also obtain

- $\limsup_{n \rightarrow \infty} n E[W_{nt}^2] = 0;$
- $E[W_{nt} / \mathcal{F}_{t-1}] \simeq \frac{1}{2n} f(0) V_{t-1} Y_{t-1}^2 \quad \text{for } |n^{-1/2} Y_{t-1}| < \varepsilon;$
- $\sum_{t=1}^n W_{nt} \xrightarrow[n \rightarrow \infty]{\mathcal{P}} \frac{1}{2} f(0) E[V_0 Y_0^2 \mathbf{1}_{Y_0 > 0}].$

Then we deduce

$$B_n \xrightarrow[n \rightarrow \infty]{\mathcal{P}} f(0) E[V_0 Y_0^2]. \quad (5.10)$$

The proof is achieved from (5.9) and (5.10). $\square$

Proof of Theorem 3.3. We follow a proof which is similar to the one of Theorem 2 in Davis and Dunsmuir (1997) or Li and Li (2008).

Let $v = \sqrt{n}(\hat{\theta}_n - \theta_0) \in \mathbb{R}^d$. Then we are going to prove in 2/ that maximizing $\hat{L}_n(\theta)$ is equivalent to maximizing $L_n(\theta)$ which is equivalent to maximizing

$$\begin{aligned}
W_n(v) &= - \sum_{t=1}^n (q_t(\theta_0 + n^{-1/2} v) - q_t(\theta_0)) \\
&= \sum_{t=1}^n \log \left(\frac{(M_{\theta_0 + n^{-1/2} v}^t)^{-1}}{(M_{\theta_0}^t)^{-1}} \right) + (M_{\theta_0}^t)^{-1} |X_t - f_{\theta_0}^t| - (M_{\theta_0 + n^{-1/2} v}^t)^{-1} |X_t - f_{\theta_0 + n^{-1/2} v}^t|
\end{aligned} \quad (5.11)$$

with respect to v . As a consequence, there exists a sequence $(\hat{v}_n)_n$ where $\hat{v}_n$ is a maximizer of $W_n(v)$ such as $\hat{v}_n = \sqrt{n}(\hat{\theta}_n - \theta_0)$. In 1/ we will provide a limit theorem satisfied by $W_n(v)$. Then we are going to prove in 3/ that $(W_n(\cdot))_n$ converges as a process of $\mathcal{C}(\mathbb{R}^d)$ (space of continuous functions on $\mathbb{R}^d$) to a limit process W . Hence $(\hat{v}_n)_n$ converges to the maximizer of W .

1/ First, we are going to study the asymptotic behavior of $W_n(v)$. We have

$$\begin{aligned}
W_n(v) &= \sum_{t=1}^n \log \left(\frac{(M_{\theta_0 + n^{-1/2} v}^t)^{-1}}{(M_{\theta_0}^t)^{-1}} \right) + |X_t - f_{\theta_0}^t| ((M_{\theta_0}^t)^{-1} - (M_{\theta_0 + n^{-1/2} v}^t)^{-1}) \\
&\quad + \sum_{t=1}^n (M_{\theta_0 + n^{-1/2} v}^t)^{-1} (|X_t - f_{\theta_0}^t| - |X_t - f_{\theta_0 + n^{-1/2} v}^t|) \\
&= I_1(v) + I_2(v).
\end{aligned}$$

We have:

$$I_1(v) = - \sum_{t=1}^n \log \left(\frac{M_{\theta_0+n^{-1/2}v}^t}{M_{\theta_0}^t} \right) + |\zeta_t| \left(1 - \frac{M_{\theta_0}^t}{M_{\theta_0+n^{-1/2}v}^t} \right)$$

Using Taylor expansions, we deduce that for each $t \in \{1, \dots, n\}$, there exists $\bar{\theta}_1^t$ and $\bar{\theta}_2^t$ in the segment $[\theta_0, \theta_0 + n^{-1/2}v]$ such as:

$$\begin{aligned} \log \left(\frac{M_{\theta_0+n^{-1/2}v}^t}{M_{\theta_0}^t} \right) &= n^{-1/2} (M_{\theta_0}^t)^{-1} v' \left(\frac{\partial M_{\theta}^t}{\partial \theta} \right)_{\theta_0} + \frac{1}{2} n^{-1} \left\{ v' \left(\frac{\partial^2 M_{\theta}^t}{\partial \theta^2} \right)_{\bar{\theta}_1^t} - (M_{\theta_0}^t)^{-2} \left(v' \left(\frac{\partial M_{\theta}^t}{\partial \theta} \right)_{\bar{\theta}_1^t} \right)^2 \right\} \\ \frac{M_{\theta_0}^t}{M_{\theta_0+n^{-1/2}v}^t} &= 1 - n^{-1/2} (M_{\theta_0}^t)^{-1} v' \left(\frac{\partial M_{\theta}^t}{\partial \theta} \right)_{\theta_0} + \frac{1}{2} n^{-1} \left\{ 2 (M_{\theta_0}^t)^{-2} \left(v' \left(\frac{\partial M_{\theta}^t}{\partial \theta} \right)_{\bar{\theta}_2^t} \right)^2 - v' \left(\frac{\partial^2 M_{\theta}^t}{\partial \theta^2} \right)_{\bar{\theta}_2^t} \right\} \end{aligned}$$

Then,

$$\begin{aligned} I_1(v) &= n^{-1/2} \sum_{t=1}^n (M_{\theta_0}^t)^{-1} v' \left(\frac{\partial M_{\theta}^t}{\partial \theta} \right)_{\theta_0} (|\zeta_t| - 1) + \frac{1}{2n} \sum_{t=1}^n (M_{\theta_0}^t)^{-1} \left\{ v' \left(\frac{\partial^2 M_{\theta}^t}{\partial \theta^2} \right)_{\bar{\theta}_2^t} v |\zeta_t| - v' \left(\frac{\partial^2 M_{\theta}^t}{\partial \theta^2} \right)_{\bar{\theta}_1^t} v \right\} \\ &\quad + \frac{1}{2n} \sum_{t=1}^n (M_{\theta_0}^t)^{-2} \left\{ \left(v' \left(\frac{\partial M_{\theta}^t}{\partial \theta} \right)_{\bar{\theta}_1^t} \right)^2 - 2 \left(v' \left(\frac{\partial M_{\theta}^t}{\partial \theta} \right)_{\bar{\theta}_2^t} \right)^2 |\zeta_t| \right\} \\ &= I_1^{(1)}(v) + I_1^{(2)}(v) + I_1^{(3)}(v). \end{aligned}$$

Using a Central Limit Theorem for martingale-differences (see Billingsley (1968)), since $E[(M_{\theta_0}^t)^{-1} v' \left(\frac{\partial M_{\theta}^t}{\partial \theta} \right)_{\theta_0} (|\zeta_t| - 1) / \mathcal{F}_{t-1}] = 0$ and $E[\|(M_{\theta_0}^t)^{-1} v' \left(\frac{\partial M_{\theta}^t}{\partial \theta} \right)_{\theta_0}\|^2] < \infty$ (from Lemma 5.2), we have:

$$I_1^{(1)}(v) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}\left(0, E\left[(M_{\theta_0}^0)^{-2} \left(v' \left(\frac{\partial M_{\theta}^0}{\partial \theta} \right)_{\theta_0} \right)^2\right] (\sigma_{\zeta}^2 - 1)\right). \quad (5.12)$$

Now, using that $\theta \in \Theta \mapsto \frac{\partial M_{\theta}^t}{\partial \theta}$ and $\theta \mapsto \frac{\partial^2 M_{\theta}^t}{\partial \theta^2}$ are continuous functions, $\bar{\theta}_1^t \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \theta_0$ and $\bar{\theta}_2^t \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \theta_0$, we claim that $I_1^{(2)}(v)$ have the same limit distribution that $\frac{1}{2n} \sum_{t=1}^n (M_{\theta_0}^t)^{-1} v' \left(\frac{\partial^2 M_{\theta}^t}{\partial \theta^2} \right)_{\theta_0} v (|\zeta_t| - 1)$. Note that $E\left[(M_{\theta_0}^t)^{-1} v' \left(\frac{\partial^2 M_{\theta}^t}{\partial \theta^2} \right)_{\theta_0} v (|\zeta_t| - 1) / \mathcal{F}_{t-1}\right] = 0$ and $E[\|(M_{\theta_0}^t)^{-1} v' \left(\frac{\partial^2 M_{\theta}^t}{\partial \theta^2} \right)_{\theta_0} v\|] < \infty$ (from Lemma 5.2). Thus, from the strong large number law for martingale-differences (see Billingsley (1968)), we obtain:

$$\frac{1}{2n} \sum_{t=1}^n (M_{\theta_0}^t)^{-1} v' \left(\frac{\partial^2 M_{\theta}^t}{\partial \theta^2} \right)_{\theta_0} v (|\zeta_t| - 1) \xrightarrow[n \rightarrow \infty]{a.s.} 0,$$

and this implies:

$$I_1^{(2)}(v) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} 0. \quad (5.13)$$

Finally similar arguments as previously induce that $I_1^{(3)}(v)$ have the same limit distribution that $\frac{1}{2n} \sum_{t=1}^n (M_{\theta_0}^t)^{-2} \left(v' \left(\frac{\partial M_{\theta}^t}{\partial \theta} \right)_{\theta_0} \right)^2 2|\zeta_t|$. From the strong large number law for martingale-differences (see Billingsley (1968)), we obtain:

$$\begin{aligned} \frac{1}{2n} \sum_{t=1}^n (M_{\theta_0}^t)^{-2} \left(v' \left(\frac{\partial M_{\theta}^t}{\partial \theta} \right)_{\theta_0} \right)^2 (1 - 2|\zeta_t|) &\xrightarrow[n \rightarrow \infty]{a.s.} \frac{1}{2} E\left[(M_{\theta_0}^0)^{-2} \left(v' \left(\frac{\partial M_{\theta}^0}{\partial \theta} \right)_{\theta_0} \right)^2 (1 - 2|\zeta_0|)\right] \\ &\xrightarrow[n \rightarrow \infty]{a.s.} -\frac{1}{2} E\left[(M_{\theta_0}^0)^{-2} \left(v' \left(\frac{\partial M_{\theta}^0}{\partial \theta} \right)_{\theta_0} \right)^2\right], \end{aligned}$$

and this implies:

$$I_1^{(3)}(v) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} -\frac{1}{2} E\left[(M_{\theta_0}^0)^{-2} \left(v' \left(\frac{\partial M_{\theta}^0}{\partial \theta} \right)_{\theta_0} \right)^2\right]. \quad (5.14)$$

Finally, from (5.12), (5.13) and (5.14), we obtain:

$$I_1(v) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}\left(-\frac{1}{2} E\left[(M_{\theta_0}^0)^{-2} \left(v' \left(\frac{\partial M_{\theta}^0}{\partial \theta} \right)_{\theta_0} \right)^2\right], E\left[(M_{\theta_0}^0)^{-2} \left(v' \left(\frac{\partial M_{\theta}^0}{\partial \theta} \right)_{\theta_0} \right)^2\right] (\sigma_{\zeta}^2 - 1)\right). \quad (5.15)$$

Now we consider $I_2(v) = \sum_{t=1}^n (M_{\theta_0+n^{-1/2}v}^t)^{-1} (|X_t - f_{\theta_0}^t| - |X_t - f_{\theta_0+n^{-1/2}v}^t|)$. Using again Taylor expansion, we can write:

$$\begin{aligned} I_2(v) &= \sum_{t=1}^n (M_{\theta_0}^t)^{-1} (|X_t - f_{\theta_0}^t| - |X_t - f_{\theta_0+n^{-1/2}v}^t|) \\ &\quad - n^{-1/2} \sum_{t=1}^n (M_{\bar{\theta}_M^t}^t)^{-2} v' \left(\frac{\partial M_{\bar{\theta}_M^t}^t}{\partial \theta} \right)_{\bar{\theta}_M^t} (|X_t - f_{\theta_0}^t| - |X_t - f_{\theta_0+n^{-1/2}v}^t|) \\ &= I_2^{(1)}(v) + I_2^{(2)}(v), \end{aligned}$$

with $\bar{\theta}_M^t$ in the segment $[\theta_0, \theta_0 + n^{-1/2}v]$.

First we have:

$$I_2^{(1)}(v) = \sum_{t=1}^n (|\zeta_t| - |\zeta_t - n^{-1/2}(M_{\theta_0}^t)^{-1}v'| \left(\frac{\partial f_{\theta_0}^t}{\partial \theta} \right)_{\bar{\theta}_f^t} |)$$

with $\bar{\theta}_f^t$ in the segment $[\theta_0, \theta_0 + n^{-1/2}v]$. Using Theorem 3.2, which an extension of Theorem 1 established in Davis and Dunsmuir (1997), denoting $Z_t = \zeta_t$, $Y_t = (M_{\theta_0}^t)^{-1}v' \left(\frac{\partial f_{\theta_0}^t}{\partial \theta} \right)_{\bar{\theta}_f^t}$ and $V_t = 1$ for $t \in \mathbb{Z}$,

$$I_2^{(1)} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N} \left(-g(0) \mathbb{E}[(M_{\theta_0}^0)^{-2} (v' \left(\frac{\partial f_{\theta_0}^0}{\partial \theta} \right)_{\theta_0})^2], \mathbb{E}[(M_{\theta_0}^0)^{-2} (v' \left(\frac{\partial f_{\theta_0}^0}{\partial \theta} \right)_{\theta_0})^2] \right). \quad (5.16)$$

since $\mathbb{E}[Y_t^2 V_t^2] \leq \underline{M}^{-2} \mathbb{E}[\|v' \frac{\partial f_{\theta_0}^t}{\partial \theta}\|_{\Theta}^2] < \infty$ from Lemma 5.2.

Then, we have:

$$\begin{aligned} I_2^{(2)}(v) &\sim n^{-1/2} \sum_{t=1}^n (M_{\theta_0}^t)^{-2} v' \left(\frac{\partial M_{\theta_0}^t}{\partial \theta} \right)_{\theta_0} (|X_t - f_{\theta_0}^t| - |X_t - f_{\theta_0+n^{-1/2}v}^t|) \\ &\sim n^{-1/2} \sum_{t=1}^n (M_{\theta_0}^t)^{-1} v' \left(\frac{\partial M_{\theta_0}^t}{\partial \theta} \right)_{\theta_0} (|\zeta_t| - |\zeta_t - n^{-1/2}(M_{\theta_0}^t)^{-1}v'| \left(\frac{\partial f_{\theta_0}^t}{\partial \theta} \right)_{\bar{\theta}_f^t} |) \\ &\xrightarrow[n \rightarrow \infty]{\mathcal{P}} 0, \end{aligned} \quad (5.17)$$

using the proof of Theorem 3.2 and denoting $Z_t = \zeta_t$, $Y_t = (M_{\theta_0}^t)^{-1}v' \left(\frac{\partial f_{\theta_0}^t}{\partial \theta} \right)_{\bar{\theta}_f^t}$ and $V_t = (M_{\theta_0}^t)^{-1}v' \left(\frac{\partial M_{\theta_0}^t}{\partial \theta} \right)_{\theta_0}$ for $t \in \mathbb{Z}$ and condition $\mathbb{E}[|V_t Y_t|] < \infty$ insuring a strong law of large number instead of central limit theorem for a martingale difference process. Therefore, from (5.16) and (5.17), we deduce

$$I_2(v) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N} \left(-g(0) \mathbb{E}[(M_{\theta_0}^0)^{-2} (v' \left(\frac{\partial f_{\theta_0}^0}{\partial \theta} \right)_{\theta_0})^2], \mathbb{E}[(M_{\theta_0}^0)^{-2} (v' \left(\frac{\partial f_{\theta_0}^0}{\partial \theta} \right)_{\theta_0})^2] \right). \quad (5.18)$$

Finally, we obtain the behavior of $W_n(v)$ defined in (5.11) from (5.15) and (5.18). However, we have to specify the asymptotic dependency relation between $I_1^{(1)}$ and $I_2^{(1)}$. Indeed these two terms converge to a Gaussian law. This implies to consider the asymptotic behavior of the sum of these two terms which could be reduced to the asymptotic behavior of:

$$n^{-1/2} \sum_{t=1}^n (M_{\theta_0}^t)^{-1} v' \left(\frac{\partial M_{\theta_0}^t}{\partial \theta} \right)_{\theta_0} (|\zeta_t| - 1) + n^{-1/2} \sum_{t=1}^n (M_{\theta_0}^t)^{-1} v' \left(\frac{\partial f_{\theta_0}^t}{\partial \theta} \right)_{\theta_0} \text{sgn}(\zeta_t),$$

from the proof of Theorem 3.2. Using again a central limit theorem for martingale differences, we obtain as asymptotic variance:

$$\begin{aligned} \mathbb{E} \left[(M_{\theta_0}^t)^{-2} \left(v' \left(\frac{\partial M_{\theta_0}^t}{\partial \theta} \right)_{\theta_0} (|\zeta_t| - 1) + v' \left(\frac{\partial f_{\theta_0}^t}{\partial \theta} \right)_{\theta_0} \text{sgn}(\zeta_t) \right)^2 \right] &= \mathbb{E} \left[(M_{\theta_0}^0)^{-2} \left(v' \left(\frac{\partial M_{\theta_0}^0}{\partial \theta} \right)_{\theta_0} \right)^2 \right] (\sigma_{\zeta}^2 - 1) \\ &\quad + 2 \mathbb{E} \left[(M_{\theta_0}^0)^{-2} v' \left(\frac{\partial M_{\theta_0}^0}{\partial \theta} \right)_{\theta_0} v' \left(\frac{\partial f_{\theta_0}^0}{\partial \theta} \right)_{\theta_0} \right] \mathbb{E}[(|\zeta_t| - 1) \text{sgn}(\zeta_t)] + \mathbb{E} \left[(M_{\theta_0}^0)^{-2} \left(v' \left(\frac{\partial f_{\theta_0}^0}{\partial \theta} \right)_{\theta_0} \right)^2 \right] \\ &= \mathbb{E} \left[(M_{\theta_0}^0)^{-2} \left\{ (\sigma_{\zeta}^2 - 1) \left(v' \left(\frac{\partial M_{\theta_0}^0}{\partial \theta} \right)_{\theta_0} \right)^2 + \left(v' \left(\frac{\partial f_{\theta_0}^0}{\partial \theta} \right)_{\theta_0} \right)^2 \right\} \right] \end{aligned} \quad (5.19)$$

since $(\zeta_t)_t$ admits a symmetric probability distribution with a null median and expectation. Therefore, there is no covariance term and finally we obtain:

$$W_n(v) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} W(v) = v' \left(-\frac{1}{2} \Gamma_M - g(0) \Gamma_F \right) v + v' N \quad (5.20)$$

$$\text{with } \begin{cases} N \stackrel{\mathcal{L}}{=} \mathcal{N}(0, ((\sigma_\zeta^2 - 1) \Gamma_M + \Gamma_F)) \\ \Gamma_F = \left(\mathbb{E} \left[(M_{\theta_0}^0)^{-2} \left(\frac{\partial f_{\theta_0}^0}{\partial \theta_i} \right)_{\theta_0} \left(\frac{\partial f_{\theta_0}^0}{\partial \theta_j} \right)_{\theta_0} \right] \right)_{1 \leq i, j \leq d} \\ \Gamma_M = \left(\mathbb{E} \left[\left(\frac{\partial \log(M_{\theta_0}^0)}{\partial \theta_i} \right)_{\theta_0} \left(\frac{\partial \log(M_{\theta_0}^0)}{\partial \theta_j} \right)_{\theta_0} \right] \right)_{1 \leq i, j \leq d} \end{cases} \quad (5.21)$$

2/ Now, we consider the approximation $\widehat{W}_n(v)$ of $W_n(v)$ defined by:

$$\widehat{W}_n(v) = - \sum_{t=1}^n (\widehat{q}_t(\theta_0 + n^{-1/2}v) - \widehat{q}_t(\theta_0)) \quad \text{for any } v \in \mathbb{R}^d.$$

From the assumptions of Theorem 3.1 and (5.7) we have $\frac{1}{n} \sum_{t=1}^n \|\widehat{q}_t(\theta) - q_t(\theta)\|_{\Theta} \xrightarrow[n \rightarrow \infty]{a.s.} 0$. Then we have $\widehat{W}_n(v) = W_n(v) + R_n(v)$ with $[\sup_{v \in \mathbb{R}^d} |R_n(v)|] \leq 2 \sum_{t=1}^n [\|\widehat{q}_t(\theta) - q_t(\theta)\|_{\Theta}] \xrightarrow[n \rightarrow \infty]{a.s.} 0$ and then:

$$\widehat{W}_n(v) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} W(v) \quad (5.22)$$

with W defined in (5.20).

3/ Now, from (5.22), the proof of Theorem 3.2 and the same arguments than in the proof of Theorem 2 of Davis and Dunsmuir (1997), we deduce that finite distributions $(\widehat{W}_n(v_1), \dots, \widehat{W}_n(v_k))$ converge to $(W(v_1), \dots, W(v_k))$ for any $(v_1, \dots, v_k) \in (\mathbb{R}^d)^k$. Moreover, always following the proof of Theorem 2, $(W_n(v))_v$ converges to $(W(v))_v$ as a process on the continuous function space $\mathcal{C}^0$.

As a consequence, a maximum $\widehat{v}$ of $\widehat{W}_n(v)$ satisfies:

$$\widehat{v} = (\Gamma_M + 2g(0) \Gamma_F)^{-1} N,$$

and this implies (3.3). $\square$

Proof of Proposition 3.1. First, Condition **C2** is satisfied since $b_0 > 0$. Other conditions on Lipschitzian coefficients are also satisfied from Lemma 2.1 (see the arguments above). The identifiability condition **C3** is also satisfied from the following which are divided into two parts. In (i) we proof that $(\delta, b_0, (b_i^+(\theta), b_i^-(\theta))_{i \geq 1})$ (defined in (3.5)) are unique, thereafter in (ii) we proof that $\theta = (\omega, (\alpha_i)_{1 \leq i \leq p}, (\gamma_i)_{1 \leq i \leq p}, (\beta_i)_{1 \leq i \leq q})$ is also unique.

(i) The proof of this result follow the same reasoning in Berkes et al. (2003). First we have

$$\sigma_t^\delta = b_0(\theta) + \sum_{i \geq 1} b_i^+(\theta) (\max(X_{t-i}, 0))^\delta + \sum_{i \geq 1} b_i^-(\theta) (\max(-X_{t-i}, 0))^\delta. \quad (5.23)$$

We prove the result by contradiction. Suppose that there exist two vectors $\beta = (\delta, b_0, (b_i^+)_{i \geq 1}, (b_i^-)_{i \geq 1})$ and $\beta' = (\delta', b'_0, (b_i^{+'})_{i \geq 1}, (b_i^{-'})_{i \geq 1})$ verifying (5.23). Let $m > 0$ be the smallest integer satisfying $b_m^+ \neq b_m^{+'}$ or $b_m^- \neq b_m^{-'}$ (if $b_i^+ = b_i^{+'}$ and $b_i^- = b_i^{-'}$ $\forall i \geq 1$ then $b_0 = b'_0$). In one hand, since $x \in (0, \infty) \mapsto x^\delta$ is a one-to-one map and since $P(X_t = \pm 1, \forall t \in \mathbb{Z}) = 0$, we have $\delta = \delta'$. In the other hand, by definition of m , we have

$$\begin{aligned} & (b_m^{+'} - b_m^+) (\max(X_{t-m}, 0))^\delta + (b_m^{-'} - b_m^-) (\max(-X_{t-m}, 0))^\delta \\ &= b_0 - b'_0 + \sum_{i \geq m+1} (b_i^+ - b_i^{+'}) (\max(X_{t-i}, 0))^\delta + \sum_{i \geq m+1} (b_i^- - b_i^{-'}) (\max(-X_{t-i}, 0))^\delta. \end{aligned} \quad (5.24)$$

From (3.4), we have $X_{t-m} = \sigma_{t-m} \zeta_{t-m}$, therefore

$$(b_m^{+'} - b_m^+) (\max(X_{t-m}, 0))^\delta + (b_m^{-'} - b_m^-) (\max(-X_{t-m}, 0))^\delta = \begin{cases} (b_m^{+'} - b_m^+) \sigma_{t-m}^\delta \zeta_{t-m}^\delta & \text{when } \zeta_{t-m} \geq 0 \\ (b_m^{-'} - b_m^-) \sigma_{t-m}^\delta (-\zeta_{t-m})^\delta & \text{when } \zeta_{t-m} < 0 \end{cases}$$

Moreover (5.24) and the fact that $b_m^+ \neq b_m^{+'}$ or $b_m^- \neq b_m^{-'}$ implies that at least one of the following equalities hold

$$\begin{cases} \zeta_{t-m}^\delta = ((b_m^{+'} - b_m^+) \sigma_{t-m}^\delta)^{-1} \left(\sum_{i \geq m+1} (b_i^+ - b_i^{+'}) (\max(X_{t-i}, 0))^\delta \right) & \text{when } \zeta_{t-m} \geq 0 \\ \text{or} \\ (-\zeta_{t-m})^\delta = ((b_m^{-'} - b_m^-) \sigma_{t-m}^\delta)^{-1} \left(\sum_{i \geq m+1} (b_i^- - b_i^{-'}) (\max(X_{t-i}, 0))^\delta \right) & \text{when } \zeta_{t-m} < 0 \end{cases}$$

Since $\sigma_{t-m}^\delta > b_0 > 0$, ζ_{t-m}^δ is well defined. Let F_k be the F -algebra generated by $(\zeta_i, i < k)$. The causal representation of the APARCH(δ, p, q) shows that X_j is F_j -measurable and thus the right-hand side of the above equations (and consequently also ζ_{t-m}^δ in the case $\zeta_{t-m} \geq 0$ or the case $\zeta_{t-m} < 0$) is a real-valued random variable, measurable with respect to F_{t-m-1} . Since (ζ_j) is a sequence of independent random variables, this implies that ζ_{t-m} is a.s. constant when $\zeta_{t-m} \geq 0$ or when $\zeta_{t-m} < 0$, contradicting the hypothesis saying ζ_0^δ has a non-degenerate distribution. This achieves (i).

(ii) The representation (5.23) is the same as

$$\sigma_t^\delta = b_0 + \Psi^+(L)(\max(X_t, 0))^\delta + \Psi^-(L)(\max(-X_t, 0))^\delta.$$

with $\Psi^+ = \Upsilon_{\theta_1}^{-1} \Delta_{\theta_2}^+$, $\Psi^- = \Upsilon_{\theta_1}^{-1} \Delta_{\theta_2}^-$ and $\Delta_{\theta_2}^+(L) = \sum_{i=1}^p \alpha_i (1 - \gamma_i) L^i$, $\Delta_{\theta_2}^-(L) = \sum_{i=1}^p \alpha_i (1 + \gamma_i) L^i$ and $\Upsilon_{\theta_1}(L) = \sum_{i=1}^q \beta_i L^i$, where $(\Delta_{\theta_2}^+, \Upsilon_{\theta_1})$ and $(\Delta_{\theta_2}^-, \Upsilon_{\theta_1})$ respectively coprime and $\theta_1 = (\beta_i)_{1 \leq i \leq q}$, $\theta_2 = ((\alpha_i)_{1 \leq i \leq p}, (\gamma_i)_{1 \leq i \leq p})$, then $\theta = (\omega, \theta_1, \theta_2)$.

Suppose that there exist others polynomials $\Delta_{\theta'_2}^+ = \sum_{i=1}^p \alpha'_i (1 - \gamma'_i) L^i$, $\Delta_{\theta'_2}^- = \sum_{i=1}^p \alpha'_i (1 + \gamma'_i) L^i$, $\Upsilon_{\theta'_1} = \sum_{i=1}^q \beta'_i L^i$ satisfying $\Psi^+ = \Upsilon_{\theta'_1}^{-1} \Delta_{\theta'_2}^+$, $\Psi^- = \Upsilon_{\theta'_1}^{-1} \Delta_{\theta'_2}^-$ with $(\Delta_{\theta'_2}^+, \Upsilon_{\theta'_1})$, $(\Delta_{\theta'_2}^-, \Upsilon_{\theta'_1})$ respectively coprime. Then

$$\begin{cases} \Upsilon_{\theta_1}^{-1} \Delta_{\theta_2}^+ = \Upsilon_{\theta'_1}^{-1} \Delta_{\theta'_2}^+ \\ \Upsilon_{\theta_1}^{-1} \Delta_{\theta_2}^- = \Upsilon_{\theta'_1}^{-1} \Delta_{\theta'_2}^- \end{cases} \Rightarrow \begin{cases} \Delta_{\theta_2}^+ = (\Upsilon_{\theta_1} \Upsilon_{\theta'_1}^{-1}) \Delta_{\theta'_2}^+ \\ \Delta_{\theta_2}^- = (\Upsilon_{\theta_1} \Upsilon_{\theta'_1}^{-1}) \Delta_{\theta'_2}^- \end{cases}$$

from the first equality, since $\deg(\Delta_{\theta_2}^+) = \deg(\Delta_{\theta'_2}^+) = q$, we conclude that $\Upsilon_{\theta_1} \Upsilon_{\theta'_1}^{-1} = 1$, therefore $\Upsilon_{\theta_1} = \Upsilon_{\theta'_1}$ and so $\Delta_{\theta_2}^+ = \Delta_{\theta'_2}^+$, likewise from the second equality we conclude that $\Delta_{\theta_2}^- = \Delta_{\theta'_2}^-$.

- The equalities $\Delta_{\theta'_2}^+ = \Delta_{\theta_2}^+$, $\Delta_{\theta'_2}^- = \Delta_{\theta_2}^-$ implies that $\alpha_i (1 - \gamma_i) = \alpha'_i (1 - \gamma'_i)$ and $\alpha_i (1 + \gamma_i) = \alpha'_i (1 + \gamma'_i)$ which give $\alpha_i = \alpha'_i$ and $\gamma_i = \gamma'_i$.
- The equality $\Upsilon_{\theta_1} = \Upsilon_{\theta'_1}$ implies that $\beta_i = \beta'_i$.
- Since $(\beta_i)_{i=1, p}$, $b_0 = w(1 - \sum_{j=1}^p \beta_j)^{-1}$ are unique then ω is unique.

Thus, Condition **C3** is established and the proof of proposition is achieved. $\square$

Proof of Proposition 3.2. Since we prove that Lemma 2.1 implies that conditions on Liptshitzian coefficients $(\alpha_j^{(i)}(f, \Theta))_j$ and $(\alpha_j^{(i)}(M, \Theta))_j$, it remains to prove conditions **C2** and **C3**. Condition **C2** holds since c_0 is supposed to be a positive number. Finally, condition **C3** also holds since $f_\theta = f_{\theta'}$ implies $\psi_j(\theta) = \psi_j(\theta')$ for all $j \in \mathbb{Z}$. Therefore the parameters of the ARMA part of the process are identified and then the identification of the parameters GARCH can be deduced from the proof of Proposition 3.1. $\square$

Proof of Proposition 3.4. This proofs mimics exactly the proof of Proposition 3.2. $\square$

References

- Bardet, J.-M. and Wintenberger, O. (2009) Asymptotic normality of the Quasi-Maximum likelihood estimator for multidimensional causal process, *Ann. Statist.*, **37**, 2730–2759.
- Bardet, J.-M., Kengne, W. and Wintenberger, O. (2012) Detecting multiple change-points in general causal time series using penalized quasi-likelihood. *Electronic Journal of Statistics* 6, **2012**, 435–477.

- Berkes, I., Horváth, L. and Kokoszka, P. (2003), GARCH processes: structure and estimation, *Bernoulli*, **9**, 201–227.
- Berkes, I. and Horváth, L. (2004) The efficiency of the estimators of the parameters in GARCH processes, *Ann. Statist.*, **32**, 633–655.
- Billingsley, P. (1968) *Convergence of Probability Measures*. Wiley, New York.
- Davis, R. and Dunsmuir, W. (1997) Least Absolute Deviation Estimation for Regression with ARMA Errors. *Journal of Theoretical Probability*, **10**, 481–497.
- Ding, Z., Granger C.W.J. and Engle R.F. (1993) A Long Memory Property of Stock Market Returns and a New Model, *Journal of Empirical Finance*, **1**, 83–106.
- Doukhan, P. and Wintenberger, O. (2007) Weakly dependent chains with infinite memory. *Stochastic Process. Appl.*, **118**, 1997–2013.
- Francq, C. and Zakoian, J.-M. (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes, *Bernoulli*, **10**, 605–637.
- Francq, C., Lepage, G. and Zakoian, J.-M. (2011) Two-stage non Gaussian QML estimation of GARCH models and testing the efficiency of the Gaussian QMLE. *Journal of Econometrics*, **165**, 246–257.
- Francq, C. and Zakoian, J.-M. (2013) Optimal predictions of powers of conditionally heteroskedastic processes. *Journal of the Royal Statistical Society B*, **75**, 345–367.
- Francq, C. and Zakoian, J.-M. (2015) Risk-parameter estimation in volatility models. *Journal of Econometrics*, **184**, 158–173.
- Jeantheau, T. (1998). Strong consistency of estimators for multivariate arch models. *Econometric Theory*, **14**, 70–86.
- Kounias, E.G. and Weng, T.-S. (1969) An inequality and almost sure convergence. *Annals of Mathematical Statistics*, **40**, 1091–1093.
- Li, G. and Li, W.K. (2008) Least Absolute Deviation Estimation for Fractionally Integrated Autoregressive Moving Average Time Series Models with Conditional Heteroscedasticity. *Biometrika*, **95**, 399–414.
- Ling, S. and McAleer, M. (2003). Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory*, **19**, 280–310.
- Lumsdaine, R.L. (1996) Consistency and asymptotic normality of the quasi-maximum likelihood in IGARCH(1, 1) and covariance stationary GARCH(1, 1) models. *Econometrica*, **64**, 575–596.
- Newey, W.K. and Steigerwald, D.G. (1997). Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models. *Econometrica*, **65**, 587–599.
- Peng, L. and Yao, Q. (2003). Least absolute deviations estimation for ARCH and GARCH models. *Biometrika*, **90**, 967–975.
- Robinson, P.M. (1991). Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression, *Journal of Econometrics*, **47**, 67–84.
- Robinson, P.M. and Zaffaroni, P. (2006) Pseudo-maximum likelihood estimation of ARCH(∞) models, *Ann. Statist.* **34**, 1049–1074.
- Straumann, D. and Mikosch, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *Ann. Statist.* **34**, 2449–2495.
- van der Vaart, A.W. (2000) *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.
- Weiss, A.A. (1986) Asymptotic theory for ARCH models estimation and testing. *Econometric Theory*, **2**, 107–131.

Bibliographie

- [1] Ansley, C.F. (1979), An algorithm for the exact likelihood of a mixed autoregressive moving average process, *Biometrika* 66, 59-65 .
- [2] Bardet, J.M. and Wintenberger, O. (2009), Asymptotic normality of the Quasi-Maximum likelihood estimator for multidimensional causal process, *The Annals of Statistics*, 37, 2730-2759.
- [3] J.M. Bardet, W.Kengne, and O. Wintenberger, (2012) Detecting multiple change-points in general causal time series using penalized quasi-likelihood. *Electronic Journal of Statistics* 6, 2012, 435477.
- [4] Barnarda, W. Trindade, A.A. Wickramasingheb, R.I.P (2014), Auto-regressive Moving Average Models Under Exponential Power Distributions, *ProbStat Forum*, 7, 65-77.
- [5] Berkes, I., Horváth, L. and Kokoszka, P. (2003), GARCH processes : structure and estimation, *Bernoulli*, 9, 201-227.
- [6] Berkes, I. and Horváth, L. (2004), The efficiency of the estimators of the parameters in GARCH processes, *The Annals of Statistics*, 32, 633-655.
- [7] Billingsley, P. (1968), *Convergence of Probability Measures*. Wiley, New York.
- [8] Box, G.E.P. Jenkins, G.M. (1970), *Time Series Analysis ; Forecasting and Control*, San Francisco : Holden-Day.
- [9] Box, G.E.P. Jenkins, G.M. (1976), *Time Series Analysis Forecasting and Control*, 2nd edition. San Franoisco : Holden-Day.
- [10] Bollerslev, T. (1986), Generalised autoregressive conditional heteroscedasticity, *Journal of Econometrics*, 31, 307-327.
- [11] Broyden, C. G. (1970), The convergence of a class of double-rank minimization algorithms, *Journal of the Institute of Mathematics and Its Applications*, 6, 76-90.

- [12] Brockwell, P.J. Davis, R.A. (2002), Time Series Theory and Methods, Springer-Verlag : New York.
- [13] Brockwell P.J., Lindner A. (2010), Strictly stationary solutions of autoregressive moving average equations. *Biometrika* 97, 765-772.
- [14] Davis, R. and Dunsmuir, W. (1997), Least Absolute Deviation Estimation for Regression with ARMA Errors. *Journal of Theoretical Probability*, 10, 481-497.
- [15] Davis, R. Knight, K. Liu, J. (1992), R.M-Estimation for Autoregressions with Infinite Variance, *Stoch. Proc. Appl.* 40, 145-180.
- [16] DeFrutos, R.F. Serrano, G.R.(1997), A generalized least squares estimation method for invertible vector moving average models, *Econom Lett* 57, 149-156.
- [17] DENT, W. MEN, A.S. (1978), A Monte Carlo study of autoregressive integrated moving average processes. *J. Econometrics* 7, 23-55.
- [18] Doukhan, P. and Wintenberger, O. (2007), Weakly dependent chains with infinite memory. *Stochastic Process. Appl.*, 118, 1997-2013.
- [19] Ding, Z., Granger C.W.J., Engle R.F. (1993), A Long Memory Property of Stock Market Returns and a New Model, *Journal of Empirical Finance* 1, 83-106.
- [20] Dugre, J.P.(1980), Generating covariance sequences and the calculation of quantization and rounding error variances in digital filters, *Acoustics, Speech and n.a. Processing, IEEE Transactions*, Vol. 28, 102 - 104.
- [21] Durbin, J. (1959), Efficient estimation of parameters in moving-average models, *Biometrika* 46, 306-16.
- [22] Engle, R.F. (1982), Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation, *Econometrica* 50, 987-1008.
- [23] Engle, R.F. (1983), Estimates of the Variance of U. S. Inflation Based upon the ARCH Model. *Journal of Money, Credit and Banking*, 15(3), 286-301.
- [24] Engle, R.F. and D. Kraft, (1983), Multiperiod forecast error variances of inflation estimated from ARCH models, A. ZeUner, ed., *Applied time series analysis of economic data* (Bureau of the Census, Washington, DC) 293-302.
- [25] Feller, W. (1971), *An Introduction to Probability Theory and Its Applications* 2. Wiley, New York.

- [26] Fletcher, R. (1970), "A New Approach to Variable Metric Algorithms", Computer Journal, 13, 3, 317-322
- [27] Fletcher, R. Reeves, C. M. (1964), Function minimization by conjugate gradients, The Computer Journal, Vol. 7, 2, 149-154.
- [28] Francq, C. and Zakoian, J.-M. (2004), Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes, Bernoulli, 10, 605-637.
- [29] Francq, C., Lepage, G. and Zakoian, J.-M. (2011), Two-stage non Gaussian QML estimation of GARCH models and testing the efficiency of the Gaussian EQMV. Journal of Econometrics, 165, 246-257.
- [30] Francq, C. and Zakoian, J.-M. (2013), Optimal predictions of powers of conditionally heteroskedastic processes. J. Roy. Statist.soc, B, 75, 345-367.
- [31] Francq, C. and Zakoian, J.-M. (2015), Risk-parameter estimation in volatility models. Journal of Econometrics, 184, 158-173.
- [32] Gao, Y., Zhang, C. and Zhang, L. (2012) Comparison of GARCH Models based on Different Distributions, Journal of Computers, 7, 1967-1973.
- [33] Ghoudi, K. and Remillard, B. (2014), Comparison of specification tests for GARCH models. Computational Statistics and Data Analysis, 76, 291-300.
- [34] Godolphin, E.J. (1984), A direct representation for the large sample maximum likelihood estimator of a Gaussian autoregressive moving average process, Biometrika, 71, 1984, 281-289.
- [35] Goldfarb, D. (1970), A Family of Variable Metric Methods Derived by Variational Means Maths. Comput., 24, 23-26.
- [36] Hannan, E.J. Rissanen, E.J. (1982), Recursive estimation of mixed autoregressive-moving average order, Biometrika, 69, 81-94.
- [37] Jeantheau, T. (1998), Strong consistency of estimators for multivariate arch models. Econometric Theory, 14, 70-86.
- [38] Klar, B., Lindner, F. and Meintanis, S.G. (2012) Specification tests for the error distribution in GARCH models. Computational Statistics and Data Analysis 56, 3587-3598.
- [39] Koreisha, S. Pukkila, T. (1989), Fast linear estimation methods for vector moving average models, J. Time Ser. Analysis, 10, 325-339.

- [40] Kounias, E.G. and Weng, T.-S. (1969), An inequality and almost sure convergence. *Annals of Mathematical Statistics*, 40, 1091-1093.
- [41] Li, G. and Li, W.K. (2008), Least Absolute Deviation Estimation for Fractionally Integrated Autoregressive Moving Average Time Series Models with Conditional Heteroscedasticity. *Biometrika*, 95, 399-414.
- [42] Ling, S. McAleer, M. (2002). Necessary and Sufficient Moment Conditions for the GARCH(r, s) and Asymmetric Power GARCH(r, s) Models. *Econometric Theory*, 18, 3, 722-729.
- [43] Ling, S. and McAleer, M. (2003), Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory*, 19, 280-310.
- [44] Li, W.K. McLeod, A.I. (1988), ARMA modeling with non-Gaussian innovations, *J. Time Ser.*, 9, 2, 155-168.
- [45] Lumsdaine, R.L. (1996), Consistency and asymptotic normality of the quasi-maximum likelihood in IGARCH(1, 1) and covariance stationary GARCH(1, 1) models. *Econometrica*, 64, 575-596.
- [46] Mauricio, J.A. (1959), Exact maximum likelihood estimation of stationary vector ARMA models. *J. Am. Statist. Ass.*, 90, 282-91.
- [47] Mauricio, J.A. (1997), The exact likelihood of a vector autoregressive moving average model, *Appl. Statistics* 46, 157-171.
- [48] Mauricio, J.A. (2002), An algorithm for the exact likelihood of a vector autoregressive moving average model, *J. Time Ser. Analysis* 23, 473-486.
- [49] McLeod, A. I. (1977), Improved Box-Jenkins estimators. *Biometrika* 64, 531-534.
- [50] McLeod, A.I. Holanda Sales P.R. (1983), An algorithm for approximate likelihood calculation of ARMA and seasonal ARMA models, *J. Roy. Statist.soc*, 32, 2, 211-223.
- [51] Melard, G. A. (1984), Fast algorithm for the exact likelihood of autoregressive-moving average models, *Appl. Statistics*, 33, 104-114.
- [52] Mittnik, S. (1990), Computation of theoretical autocovariance matrices of multivariate autoregressive moving average time series, *J. Roy. Statist.Soc, Series B*, 52, 151-155.
- [53] Newey, W.K. and Steigerwald, D.G. (1997), Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models. *Econometrica* 65, 587-599.

- [54] Nelder, J. A. Mead, R. (1965), A Simplex Method for Function Minimization. *Comput. J.* 7, 308-313.
- [55] Nelson, H. L. Jr. and Granger, C. W. J. (1979), Experience with using the Box-Cox transformation when forecasting economic time series. *Journal of Econometrics*, 10, 5769.
- [56] Nicholls, D.F. Hall, A.D. (1979), The exact likelihood function of multivariate autoregressive moving average models, *Biometrika*, 66, 259-264.
- [57] Pearlman, J.G. (1980), An algorithm for the exact likelihood of a high-order autoregressive -moving average process, *Biometrika*, 67, 232-233.
- [58] Peng, L. and Yao, Q. (2003), Least absolute deviations estimation for ARCH and GARCH models. *Biometrika* 90, 967-975.
- [59] OSBOBN, D. R. (1976), Maximum likelihood estimation of moving average processes. *Ann. Econ. and Soc. Measurement* 5, 75-87.
- [60] Reinsel, G.C. Basu, S. Yap, S.F. (1992), Maximum likelihood estimators in the multivariate autoregressive moving average model from a generalized least squares viewpoint, *J. Time Ser Analysis*, 13, 133-145.
- [61] Reinsel, G.C. (1997), *Elements of Multivariate Time Series Analysis*. second ed, Springer, New York.
- [62] Robinson, P.M. (1991), Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression, *Journal of Econometrics*, 47, 67-84.
- [63] Robinson, P.M. and Zaffaroni, P. (2006), Pseudo-maximum likelihood estimation of ARCH(∞) models, *Ann. Statist.* 34, 1049-1074.
- [64] Shanno, David, F. Kettler, Paul C. (1970), Optimal conditioning of quasi-Newton methods, *Math. Comput.* 24, 111, 657-664.
- [65] Shea, B.L. A (1988), Note on the generation of independent realisations of a vector autoregressive moving average proces, *J. Time Ser. Anal.* 9, 403-410.
- [66] Shea, B.L. (1989), The exact likelihood of a vector autoregressive moving average model, *J. Roy. Statist. Soc. Ser. C*, 38, 161-184.
- [67] SHEPHARD, N. PITT, M.K. (1997), Likelihood analysis of non-Gaussian measurement time series, *Biometrika*. 84, 3 : 653-667.
- [68] Straumann, D. and Mikosch, T. (2006), Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series : A stochastic recurrence equations approach. *Ann. Statist.* 34, 2449-2495.

- [69] Solo, V. (1984) The exact likelihood for a multivariate ARMA model, J. Multivar. Anal. 15, 164-173.
- [70] Tunnicliffe-Wilson, G. (1973), The estimation of parameters in multivariate time series models, J. Roy. Statist.Soc, B, 35, 76-85.
- [71] Van der Vaart, A.W. (2000), Asymptotic Statistics, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.
- [72] Weiss, A.A. (1984), ARMA models with ARCH errors. J. Time Ser. Anal., 5, 2, 129-143.
- [73] Weiss, A.A. (1986), Asymptotic theory for ARCH models estimation and testing. Econometric Theory, 2, 107-131.
- [74] Whittle, P. (1951) Thesis : Hypothesis Testing in Time Series Analysis, Upsala : Almqvist and Wiksell.

Conclusion générale

Dans cette thèse, notre travail s'est porté sur la modélisation des séries chronologiques. Tout d'abord nous avons proposé un espace de stationnarité pour la classe générale des processus causaux affines, sous des conditions de type Lipchitz, des conditions qui se sont avérées vérifiées par les processus ARMA, GARCH, APARCH, ARMA-GARCH, ARMA-APARCH,...etc.

Par suite nous avons montré la consistance des estimateurs de quasi vraisemblance pour la classe des processus causaux affines, sous l'hypothèse que le bruit est non Gaussien, précisément un bruit Laplace. Nous avons réussi à prouver la consistance seulement sous condition d'existence des moments d'ordre un, une condition plus faible que celle exigée dans la modélisation des processus avec bruits Gaussiens.

Comme nous avons montré la normalité asymptotique sous condition d'existence des moments d'ordre deux au lieu des moments d'ordre quatre nécessaire pour les processus de bruit Gaussien. À travers des expériences numériques de type Monte-Carlo, nous avons pu vérifier les résultats théoriques obtenus. De même, nous avons montré que l'EQMV-Laplace fournit une estimation plus exacte que l'EQMV-Gauss, sauf bien sûr dans le cas d'une distribution Gaussienne.

Par ailleurs, nous avons proposé une procédure de calcul du maximum de vraisemblance d'un modèle ARMA avec bruits Gaussiens. Une méthode basée sur l'approximation de la fonction de vraisemblance par une fonction qui admet une solution analytique. Ce qui nous a permis d'éviter le problème de haute non linéarité, causé par la fonction d'autocovariance.

Perspectives

Les travaux effectués dans cette thèse comportent de nombreuses extensions possibles, et dans plusieurs directions,

Pour la modélisation des séries chronologiques et l'étude de leurs comportements asymptotiques. La première extension possible est d'étudier les processus causaux affines avec bruits de loi Gauss généralisé, une loi de probabilité qui contient la loi Gaussienne, la loi Laplace et d'autres. Ce type de généralisation a été déjà étudié dans le cas des modèles GARCH (voir Francq *et al.* (2011)) là où ils ont proposé une estimation de quasi-vraisemblance en deux étapes pour les modèles GARCH. Une deuxième extension que nous trouvons intéressante, est l'étude de comportement asymptotique des processus causaux affines avec fonction de tendance type déterministe.

Pour le calcul numérique du maximum de vraisemblance, nous avons une autre voie de recherche possible, qui est la mise en place de procédures de recherche numérique du maximum pour les modèles ARCH, GARCH et APARCH avec bruits Gaussiens ou d'autres, des procédures basées sur l'approximation locale de la fonction de vraisemblance.