n° 2008-03

A Tour in the Asymptotic Theory of GARCH Estimation

C. FRANCQ[1]
J.-M. ZAKOÏAN[2]

[1] Université Lille III, GREMARS-EQUIPPE, BP 60149, 59653 Villeneuve d'Ascq Cédex, France.
Tél. : 33 (0) 3 20 41 64 87 (francq@univ-lille3.fr)
[2] CREST and GREMARS-EQUIPPE, 15 Boulevard Gabriel Péri, 92245 Malakoff Cédex, France.
Tél. : 33 (0) 1 41 17 78 25 (zakoian@ensae.fr).

A tour in the asymptotic theory of GARCH estimation

CHRISTIAN FRANCQ*AND JEAN-MICHEL ZAKOÏAN†

Abstract

The main estimation methods of the univariate GARCH models are reviewed. A special attention is given to the asymptotic results and the quasi-maximum likelihood method.

Keywords : Asymptotic properties of estimators, Efficient estimation, GARCH model, Quasi Maximum Likelihood Estimation, Weighted Least-Squares.

Résumé

Nous passons en revue les principales méthodes d'estimation des modèles GARCH univariés, avec un intérêt particulier pour les résultats asymptotiques et la méthode du quasi-maximum de vraisemblance.

Mots clés : Estimation efficace, Estimation par quasi maximum de vraisemblance, Modèle GARCH, Moindres carrés pondérés, Propriétés asymptotiques d'estimateurs.

*Université Lille III, GREMARS-EQUIPPE, BP 60149, 59653 Villeneuve d'Ascq cedex, France. E-mail: francq@univ-lille3.fr, tel: 33.3.20.41.64.87

†CREST and GREMARS-EQUIPPE, 15 Boulevard Gabriel Péri, 92245 Malakoff Cedex, France. E-mail: zakoian@ensae.fr, tel: 33.1.41.17.78.25

1 Introduction

In comparison with other volatility models (e.g. the standard stochastic volatility model) GARCH models are simple to estimate, which has greatly contributed to their popularity. The volatility being a function of the past observations, the likelihood function has an explicit form which makes it easy to handle. A variety of alternative estimation methods can also be considered.

Least squares and quasi-maximum likelihood estimations in ARCH models were considered in the seminal paper by Engle [13]. The asymptotic properties of the quasi-maximum likelihood estimator (QMLE) received broad interest in the last 20 years. Pioneering work established consistency and asymptotic normality under strong assumptions on the parameter space and the true parameter value. The problem of finding weak assumptions for the consistency and asymptotic normality of the QMLE in GARCH models has attracted a lot of attention in the statistics literature. The first papers limited their scope to ARCH (see [49]) or GARCH(1,1) models (see [31, 40]). See [2, 3, 4, 17, 25], for recent references on the QMLE of general GARCH(p, q) models. See [46] for a recent comprehensive monograph on the estimation of GARCH models.

Numerous GARCH-type models have been introduced and it is simply not possible to consider the estimation of all of them. In this article we limit ourselves to the *standard* GARCH(p, q) model given by the equations

$$\begin{cases} \epsilon_t = \sqrt{h_t}\eta_t \\ h_t = \omega_0 + \sum_{i=1}^q \alpha_{0i}\epsilon_{t-i}^2 + \sum_{j=1}^p \beta_{0j}h_{t-j}, \quad t \in \mathbb{Z} = \{0, \pm 1, \dots\} \end{cases} \tag{1.1}$$

where

$$\omega_0 > 0, \quad \alpha_{0i} \geq 0 \ (i = 1, \dots, q), \quad \beta_{0j} \geq 0 \ (j = 1, \dots, p),$$

$$\{\eta_t, t \in \mathbb{Z}\} \quad \text{are iid random variables such that } E\eta_1^2 = 1.$$

We assume that $\epsilon_1, \dots, \epsilon_n$ are observations from the process $(\epsilon_t, t \in \mathbb{Z})$, assumed to be a *strictly stationary, ergodic and nonanticipative* solution of Model (1.1). Conditions for stationarity are obtained (see [32]) from the vector representation

$$\underline{z}_t = \underline{b}_t + A_{0t-1}\underline{z}_{t-1}, \tag{1.2}$$

1

where, for $p \geq 2$ and $q \geq 2$,

$$\underline{z}_t = (h_t, \ldots, h_{t-p+1}, \epsilon_{t-1}^2, \ldots, \epsilon_{t-q+1}^2)' \in \mathbb{R}^{p+q-1},$$

$$\underline{b}_t = (\omega, 0, \ldots, 0)' \in \mathbb{R}^{p+q-1},$$

$$A_{0t} = \begin{pmatrix} \tau_t' & \beta_{0p} & \alpha_{02:q-1}' & \alpha_{0q} \\ I_{p-1} & 0 & 0 & 0 \\ \xi_t' & 0 & 0 & 0 \\ 0 & 0 & I_{q-2} & 0 \end{pmatrix},$$

with

$$\tau_t = (\beta_{01} + \alpha_{01}\eta_t^2, \beta_{02}, \ldots, \beta_{0p-1})' \in \mathbb{R}^{p-1},$$

$$\xi_t = (\eta_t^2, 0, \ldots, 0)' \in \mathbb{R}^{p-1},$$

$$\alpha_{02:q-1} = (\alpha_{02}, \ldots, \alpha_{0q-1})' \in \mathbb{R}^{q-2}.$$

A nonanticipative solution (ϵ_t) of model (1.1) is such that ϵ_t is a measurable function of the $(\eta_{t-i}, i \geq 0)$. Bougerol and Picard [7] showed that the model has a (unique) strictly stationary non anticipative solution if and only if

$$\gamma(\mathbf{A_0}) < 0,$$

where $\gamma(\mathbf{A_0})$ is the top Lyapunov exponent of the sequence (A_{0t}), that is

$$\gamma(\mathbf{A_0}) = \lim_{t \to \infty} \frac{1}{t} \log \|A_{0t} A_{0t-1} \ldots A_{01}\| \quad a.s.$$

where $\|\cdot\|$ denotes any norm on the space of the $(p+q-1) \times (p+q-1)$ matrices. In addition, the strictly stationary solution is ergodic as a measurable function of the $(\eta_{t-i}, i \geq 0)$. Let us mention two important consequences of $\gamma(\mathbf{A_0}) < 0$: (i) $\sum_{j=1}^{p} \beta_{0j} < 1$, and (ii) for some $s > 0$, $E|\epsilon_1|^{2s} < \infty$ (see Lemma 2.3 in [4] for the proof). The latter property is crucial to avoid unnecessary moment conditions in the proof of the asymptotic properties of the QMLE.

Throughout the orders p and q are assumed to be known. The vector of parameters is denoted by

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

and it belongs to a parameter space $\Theta \subset]0, +\infty[\times [0, \infty[^{p+q}$. The true parameter value $\theta_0 = (\omega_0, \alpha_{01}, \ldots, \alpha_{0q}, \beta_{01}, \ldots, \beta_{0p})'$ is unknown.

We review the main estimation methods, with special attention to the quasi-maximum likelihood method. The focus will be on asymptotic results rather than on small-sample and numerical issues. We start, in Section 2, by considering the Least-Squares estimator (LSE) for ARCH(q) models, which is simple to compute but requires high moment assumptions. Then we turn to QMLE in Section 3. Section 4 is devoted to efficiency issues. In Section 5 we consider alternative estimators. Finally we discuss in Section 6 the case where some GARCH coefficients are equal to zero, putting the true parameter value on the boundary of the parameter space.

2 Least-Squares estimation of ARCH models

In this section we assume $p = 0$. The LSE is obtained from the AR(q) representation for ϵ_t^2:

$$\epsilon_t^2 = \omega_0 + \sum_{i=1}^{q} \alpha_{0i}\epsilon_{t-i}^2 + u_t, \tag{2.1}$$

where $u_t = \epsilon_t^2 - h_t = (\eta_t^2 - 1)h_t$. The sequence $(u_t, \mathcal{F}_{t-1})_t$ is thus a martingale difference when $E\epsilon_1^2 = Eh_1 < \infty$ and $\mathcal{F}_{t-1}$ denotes the σ-field generated by $\{\eta_u, u < t\}$. Let $\epsilon_0, \ldots, \epsilon_{1-q}$ denote arbitrary initial values. Introducing the vector $Z_{t-1}' = \left(1, \epsilon_{t-1}^2, \ldots, \epsilon_{t-q}^2\right)$, we get from (2.1)

$$Y = X\theta_0 + U$$

where

$$X = \begin{pmatrix} Z_{n-1}' \\ \vdots \\ Z_0' \end{pmatrix}, \quad Y = \begin{pmatrix} \epsilon_n^2 \\ \vdots \\ \epsilon_1^2 \end{pmatrix}, \quad U = \begin{pmatrix} u_n \\ \vdots \\ u_1 \end{pmatrix}.$$

When $X'X$ is non-singular (which can be shown to hold, a.s. for large enough n, under Assumption **A3** given below) the LSE of θ_0 is thus given by:

$$\hat{\theta}_n^{LS} = (\hat{\omega}, \hat{\alpha}_1, \ldots, \hat{\alpha}_q)' = (X'X)^{-1}X'Y.$$

The LSE of $s_0^2 = \mathrm{Var}(u_1)$ is

$$\hat{s}_n^2 = \frac{1}{n-q-1}\|Y - X\hat{\theta}_n^{LS}\|^2 = \frac{1}{n-q-1}\sum_{t=1}^{n}\left\{\epsilon_t^2 - \hat{\omega} - \sum_{i=1}^{q}\hat{\alpha}_i\epsilon_{t-i}^2\right\}^2.$$

It is worth mentioning that the LSE is asymptotically equivalent to the Yule-Walker estimator of the AR(q) model (2.1) (see Chapter 8 in [9]) and to the Whittle estimator studied by [24, 42, 46].

The consistency and asymptotic normality of the LSE require some additional assumptions. For identifiability we assume that the distribution of η_t is centered and nondegenerate, i.e. $P(\eta_1^2 = 1) \neq 1$. If $E(\epsilon_1^4) < +\infty$, the LSE can be shown to be strongly consistent (see [6]):

$$\hat{\theta}_n^{LS} \to \theta_0, \qquad \hat{s}_n^2 \to s_0^2, \qquad \text{a.s. as} \quad n \to \infty.$$

If, in addition $E(\epsilon_1^8) < +\infty$ the estimator of θ_0 is asymptotically normal (see also [6]); more precisely,

$$\sqrt{n}(\hat{\theta}_n^{LS} - \theta_0) \xrightarrow{d} \mathcal{N}\left\{0, \ (E\eta_1^4 - 1)A^{-1}BA^{-1}\right\}, \qquad (2.2)$$

where

$$A = E_{\theta_0}(Z_q Z_q'), \qquad B = E_{\theta_0}(h_{q+1}^2 Z_q Z_q')$$

are non-singular matrices. Note that the vector Z_q does not depend on the initial values $\epsilon_0, \ldots, \epsilon_{1-q}$. Consistent estimators of the matrices A and B are straightforwardly obtained by replacing the theoretical moments by empirical ones.

In the framework of linear regression models, it is well known that for heteroscedastic observations the ordinary LSE is outperformed by the quasi-generalized least squares estimator (QGLSE); see e.g. [26] Chapter 8. In our framework the QGLSE is defined by

$$\hat{\theta}_n^{QGLS} = (X'\hat{\Omega}X)^{-1}X'\hat{\Omega}Y,$$

where $\hat{\Omega}$ is a consistent estimator of $\Omega = \text{Diag}(h_n^{-2}, \ldots, h_1^{-2})$. If $\hat{\theta}_n^{LS}$ is computed in a first step, then $\hat{\Omega}$ can be obtained by replacing h_t by $\hat{\omega} + \sum_{i=1}^q \hat{\alpha}_i \epsilon_{t-i}^2$ in Ω. Then the two-stage least squares estimator $\hat{\theta}_n^{QGLS}$ is consistent and asymptotically normal

$$\sqrt{n}(\hat{\theta}_n^{QGLS} - \theta_0) \xrightarrow{d} \mathcal{N}\left\{0, \ (E\eta_1^4 - 1)J^{-1}\right\}, \quad J = E_{\theta_0}(h_{q+1}^{-2} Z_q Z_q'), \qquad (2.3)$$

under the moment assumption $E\epsilon_1^4 < \infty$ when all the ARCH coefficients are strictly positive, and under a slightly stronger moment assumption in the general case; see [6, 22].

The moment conditions can be made explicit using the vector representation (1.2). It is shown in [10] that $E(\epsilon_1^4) < +\infty$ if and only if $\rho\left\{E(A_{01} \otimes A_{01})\right\} < 1$ where $\otimes$ denotes the Kronecker product and $\rho(A)$ the spectral radius of a square matrix A. More generally, if $E\eta_1^{2m} < \infty$ for some positive integer m then $E(\epsilon_1^{2m}) < +\infty$ if and only if $\rho\left\{E(A_{01}^{\otimes m})\right\} < 1$, where $A_{01}^{\otimes m}$ stands for the kronecker product of A_{01} by itself m times. As can be seen in Table 1, the moment conditions imply strong reduction of the admissible parameter space. It is the main advantage of the QMLE to avoid such restrictions.

Table 1: Conditions for strict stationarity and for the existence of moments of the ARCH(1) model when η_t follows a $\mathcal{N}(0,1)$ or Student distributions normalized in such a way that $E\eta_1^2 = 1$ (St$_\nu$ stands for a normalized Student distribution with ν degrees of freedom)

	Strict stationarity	$E\epsilon_t^2 < \infty$	$E\epsilon_t^4 < \infty$	$E\epsilon_t^8 < \infty$
Normal	$\alpha_{01} < 3.562$	$\alpha_{01} < 1$	$\alpha_{01} < 0.577$	$\alpha_{01} < 0.312$
St$_3$	$\alpha_{01} < 7.389$	$\alpha_{01} < 1$	No	No
St$_5$	$\alpha_{01} < 4.797$	$\alpha_{01} < 1$	$\alpha_{01} < 0.333$	No
St$_9$	$\alpha_{01} < 4.082$	$\alpha_{01} < 1$	$\alpha_{01} < 0.488$	$\alpha_{01} < 0.143$

Note that, as in the case of linear regression models, the QGLSE is at least as efficient as the LSE. Indeed, setting $D = h_{q+1}A^{-1}Z_q - h_{q+1}^{-1}J^{-1}Z_q$, the matrix

$$
\begin{aligned}
E_{\theta_0}DD' &= A^{-1}E_{\theta_0}(h_{q+1}^2 Z_q Z_q')A^{-1} + J^{-1}E_{\theta_0}(h_{q+1}^{-2} Z_q Z_q')J^{-1} \\
&\quad - A^{-1}E_{\theta_0}(Z_q Z_q')J^{-1} - J^{-1}E_{\theta_0}(Z_q Z_q')A^{-1} \\
&= A^{-1}BA^{-1} - J^{-1}
\end{aligned}
$$

is semi-positive definite.

3   Quasi-maximum likelihood estimation

Gaussian quasi-maximum likelihood estimation has become a very popular method for GARCH models. The basic idea of this approach is to maximize the likelihood function written under the assumption that the noise (η_t) is Gaussian. Since ϵ_t is then Gaussian conditionally on the past ϵ's and σ's, the likelihood function factorizes under a very tractable form which is maximized to produce the QMLE. The Gaussianity of the noise is inessential for the asymptotic properties of the QMLE. We start by considering the case of *pure* GARCH models, corresponding to the practical situation where a GARCH is estimated on the log-returns.

3.1   Pure GARCH models

Conditionally on initial values $\epsilon_0, \ldots, \epsilon_{1-q}, \tilde{\sigma}_0^2, \ldots, \tilde{\sigma}_{1-p}^2$, let us define recursively

$$
\tilde{\sigma}_t^2 = \tilde{\sigma}_t^2(\theta) = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \tilde{\sigma}_{t-j}^2
$$

for $t = 1, \ldots, n$. Due to the initial values, the sequence $(\tilde{\sigma}_t^2)$ is not stationary, but can be viewed (see [17]) as an approximation of the strictly stationary, ergodic and nonanticipative solution of

$$\sigma_t^2 = \sigma_t^2(\theta) = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \qquad \forall t,$$

under the assumption $\sum_{j=1}^{p} \beta_j < 1$. Note that $\sigma_t^2(\theta_0) = h_t$. The Gaussian quasi-likelihood of the observations $\epsilon_1, \ldots, \epsilon_n$ is the function

$$\tilde{L}_n(\theta) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\tilde{\sigma}_t^2}} \exp\left(-\frac{\epsilon_t^2}{2\tilde{\sigma}_t^2}\right).$$

A QMLE of θ_0 is defined as any measurable solution $\hat{\theta}_n^{QML}$ of

$$\hat{\theta}_n^{QML} = \arg\max_{\theta \in \Theta} \tilde{L}_n(\theta) = \arg\min_{\theta \in \Theta} \tilde{\mathbf{l}}_n(\theta), \tag{3.1}$$

where

$$\tilde{\mathbf{l}}_n(\theta) = n^{-1} \sum_{t=1}^{n} \tilde{\ell}_t, \qquad \text{and} \quad \tilde{\ell}_t = \tilde{\ell}_t(\theta) = \frac{\epsilon_t^2}{\tilde{\sigma}_t^2} + \log \tilde{\sigma}_t^2.$$

Let $\mathcal{A}_\theta(z) = \sum_{i=1}^{q} \alpha_i z^i$ and $\mathcal{B}_\theta(z) = 1 - \sum_{j=1}^{p} \beta_j z^j$ with $\mathcal{A}_\theta(z) = 0$ if $q = 0$ and $\mathcal{B}_\theta(z) = 1$ if $p = 0$.

The paper [4] by Berkes, Horváth and Kokoszka was the first one where the GARCH(p, q) QMLE was captured in a mathematically rigorous way under weak conditions. Several technical assumptions made in [4] were relaxed by [17] and [46]. The two latter papers show that, under the following assumptions

A1: $\theta_0 \in \Theta$ and Θ is compact,

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

A3: η_t^2 has a non-degenerate distribution with $E\eta_1^2 = 1$,

A4: if $p > 0$, $\mathcal{A}_{\theta_0}(z)$ and $\mathcal{B}_{\theta_0}(z)$ have no common root, $\mathcal{A}_{\theta_0}(1) \neq 0$, and $\alpha_{0q} + \beta_{0p} \neq 0$,

the QMLE is strongly consistent,

$$\hat{\theta}_n^{QML} \to \theta_0, \qquad \text{a.s. as} \quad n \to \infty. \tag{3.2}$$

Note that in **A2** the condition for strict stationarity is imposed on the true value of the parameter only. To show where Assumptions **A1**-**A4** are used we present the scheme of proof of (3.2), the reader being referred to [17] and [46] for a detailed proof. From the

6

second part of **A2** and the compactness of Θ, we have $\sup_{\theta \in \Theta} \sum_{i=1}^{p} \beta_j < 1$. This inequality is used to show that almost surely $\tilde{\sigma}_t^2(\theta) - \sigma_t^2(\theta) \to 0$ uniformly in $\theta \in \Theta$ as $t \to \infty$, and to show that the initial values do not matter asymptotically:

$$\lim_{n \to \infty} \sup_{\theta \in \Theta} \left| \tilde{\mathbf{l}}_n(\theta) - \mathbf{l}_n(\theta) \right| = 0 \quad a.s.$$

where $\mathbf{l}_n(\theta)$ is a stationary ergodic sequence defined by replacing $\tilde{\sigma}_t^2(\theta)$ by $\sigma_t^2(\theta)$ in $\tilde{\mathbf{l}}_n(\theta)$. Then the first condition in **A2** and the ergodic theorem show that $\tilde{\mathbf{l}}_n(\theta)$ converges a.s. to the asymptotic criterion $E_{\theta_0} \mathbf{l}_1(\theta)$. For any random variable X, let $X^+ = \max(X, 0)$ and $X^- = \max(-X, 0)$. Note that $E_{\theta_0} \mathbf{l}_1^+(\theta)$ can be equal to $+\infty$, but $E_{\theta_0} \mathbf{l}_1^-(\theta)$ is always finite (because $\inf_{\theta \in \Theta} \omega > 0$) and $E_{\theta_0} \mathbf{l}_1^+(\theta_0)$ is also finite (because under **A2** we have $E_{\theta_0} h_t^s < \infty$ for some $s > 0$, see [4] and [43] for a proof of this result). Under the identifiability assumptions **A3** and **A4**, $E_{\theta_0} \mathbf{l}_1(\theta) \geq E_{\theta_0} \mathbf{l}_1(\theta_0)$ with equality if and only if $\theta = \theta_0$. These are the main arguments to show that (3.2) holds. The rest of the proof does not require additional assumptions.

Notice that the condition $E\eta_1 = 0$ is not required. The assumption that $E\eta_1^2 = 1$ is made for identifiability reasons and is not restrictive provided $E\eta_1^2 < \infty$; see [2]. The identifiability condition **A4** excludes that *all* coefficients α_{0i} be zero when $p > 0$, as well as the overidentification of *both* orders p and q. However, other situations where some coefficients α_{0i} or β_{0j} vanish are allowed. This is worth-noting since it is no longer the case for the asymptotic normality (AN).

Indeed, the main additional assumption required for the AN is that θ_0 belongs to the interior $\overset{\circ}{\Theta}$ of Θ. The case where θ_0 belongs to the boundary of Θ will be considered below. Following [17], under Assumptions **A1**-**A4** and

A5: $\theta_0 \in \overset{\circ}{\Theta}$,   **A6:** $E\eta_1^4 < \infty$,

the QMLE is asymptotically normal; more precisely

$$\sqrt{n}(\hat{\theta}_n^{QML} - \theta_0) \overset{d}{\to} \mathcal{N}\left\{0, \ (E\eta_1^4 - 1)J^{-1}\right\}, \quad J = E_{\theta_0} \frac{1}{\sigma_1^4} \frac{\partial \sigma_1^2}{\partial \theta} \frac{\partial \sigma_1^2}{\partial \theta'}(\theta_0). \quad (3.3)$$

It is shown in [17] that $\sigma_1^{-2}(\partial \sigma_1^2 / \partial \theta)$ admits moments of any order. For simplicity we give the arguments in the GARCH(1,1) case. We have $\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 = \omega(1 - \beta)^{-1} + \alpha \sum_{i=0}^{\infty} \beta^i \epsilon_{t-i-1}^2$. Thus $\sigma_1^{-2}(\partial \sigma_1^2 / \partial \omega)$ and $\sigma_1^{-2}(\partial \sigma_1^2 / \partial \alpha)$ are bounded, and therefore admit moments of any order. We have already seen that the strict stationarity condition

A2 implies the existence of some $s \in (0,1)$ such that $E_{\theta_0}|\epsilon_1|^{2s} < \infty$. Using $\partial \sigma_t^2 / \partial \beta_j = \sum_{k=1}^{\infty} k\beta^{k-1}(\omega + \alpha\epsilon_{t-1-k}^2)$, $\sigma_t^2 \geq \omega + \beta^k(\omega + \alpha\epsilon_{t-1-k}^2)$, and the elementary inequality $x/(1+x) \leq x^s$ for all $x \geq 0$, we obtain for any $d > 0$

$$
\begin{aligned}
\left\| \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \beta} \right\|_d &\leq \left\| \frac{1}{\beta} \sum_{k=1}^{\infty} \frac{k\beta^k \left(\omega + \alpha\epsilon_{t-k-1}^2\right)}{\omega + \beta^k \left(\omega + \alpha\epsilon_{t-k-1}^2\right)} \right\|_d \\
&\leq \frac{1}{\beta} \sum_{k=1}^{\infty} k \left\| \left\{ \frac{\beta^k \left(\omega + \alpha\epsilon_{t-k-1}^2\right)}{\omega} \right\}^{s/d} \right\|_d \\
&\leq \frac{1}{\omega^{s/d}\beta} \left\{ E_{\theta_0} \left(\omega + \alpha\epsilon_1^2\right)^s \right\}^{1/d} \sum_{k=1}^{\infty} k|\beta|^{sk/d} < \infty,
\end{aligned}
\tag{3.4}
$$

where $\|X\|_d^d = E|X|^d$ for any random variable X. The idea of exploiting the inequality $x/(1+x) \leq x^s$ for all $x > 0$ is due to [8]. Finally $\sigma_1^{-2}(\partial \sigma_1^2 / \partial \theta)$ admits moments of any order, and J is well defined. The identifiability assumptions **A3** and **A4** entail the invertibility of J (see (ii) of the proof of Theorem 2.2 in [17]). The consistency (3.2) of the QMLE, Assumption **A5** and a Taylor expansion of $\partial \tilde{\mathbf{l}}_n(\cdot)/\partial \theta$ yield

$$
0 = \sqrt{n} \frac{\partial \tilde{\mathbf{l}}_n(\hat{\theta}_n^{QML})}{\partial \theta} = \sqrt{n} \frac{\partial \tilde{\mathbf{l}}_n(\theta_0)}{\partial \theta} + \left(\frac{\partial^2 \tilde{\mathbf{l}}_n(\theta_{ij}^*)}{\partial \theta_i \partial \theta_j} \right) \sqrt{n} \left(\hat{\theta}_n^{QML} - \theta_0 \right)
$$

where the θ_{ij}^* are between $\hat{\theta}_n^{QML}$ and θ_0. The AN in (3.3) is then obtained by showing that

$$
\begin{aligned}
\sqrt{n} \frac{\partial \tilde{\mathbf{l}}_n(\theta_0)}{\partial \theta} &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (1 - \eta_t^2) \frac{1}{\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \sigma_t^2}{\partial \theta'}(\theta_0) + o_P(1) \\
&\xrightarrow{d} \mathcal{N} \left\{ 0, (E\eta_1^4 - 1)J \right\},
\end{aligned}
\tag{3.5}
$$

and

$$
n^{-1} \sum_{t=1}^{n} \frac{\partial^2}{\partial \theta_i \partial \theta_j} \tilde{\ell}_t(\theta_{ij}^*) \to J(i,j) \text{ in probability.}
\tag{3.6}
$$

The convergence (3.5) follows from the central limit theorem for martingale differences given by [5]. To show (3.6), a new Taylor expansion and already given arguments are employed.

It is worth-noting that no moment assumption is required for the observed process. This is particularly interesting for financial series, for which the existence of fourth and even second-order moments is questionable. The moment assumption **A6** on the iid process (η_t) is obviously necessary for the existence of the variance of the Gaussian distribution in (3.3). In the ARCH case we find the same asymptotic variance as for the QGLSE; see

(2.3). Note also that the iid assumption on (η_t) can be replaced by a martingale difference assumption at the price of additional moment assumptions; see [15].

Tables 2 and 3 illustrate the asymptotic accuracy obtained from using the LSE and the QMLE for several ARCH(1) models with Gaussian errors and parameter $\omega_0 = 1$. When a sequence of random vectors X_n converges in law to a Gaussian distribution, we denote by $\text{Var}_{as}(X_n)$ the variance of this Gaussian distribution. In view of (2.2), the asymptotic variance of the LSE in Table 2 is explicitly given by

$$\text{Var}_{as}\{\sqrt{n}(\hat{\theta}_n^{LS} - \theta_0)\} = 2A^{-1}BA^{-1},$$

where

$$A = \begin{pmatrix} 1 & E_{\theta_0}\epsilon_1^2 \\ E_{\theta_0}\epsilon_1^2 & E_{\theta_0}\epsilon_1^4 \end{pmatrix}, \qquad B = \begin{pmatrix} E_{\theta_0}\sigma_2^4 & E_{\theta_0}\sigma_2^4\epsilon_1^2 \\ E_{\theta_0}\sigma_2^4\epsilon_1^2 & E_{\theta_0}\sigma_2^4\epsilon_1^4 \end{pmatrix},$$

with

$$E_{\theta_0}\epsilon_1^2 = \frac{\omega_0}{1 - \alpha_{01}}, \qquad E_{\theta_0}\epsilon_1^4 = 3E_{\theta_0}\sigma_1^4 = \frac{3\omega_0^2(1 + \alpha_{01})}{(1 - 3\alpha_{01}^2)(1 - \alpha_{01})}.$$

The other terms of the matrix B are obtained using $\sigma_2^4 = (\omega_0 + \alpha_{01}\epsilon_1^2)^2$ and computing the moments of order 6 and 8 of ϵ_1^2. For an ARCH(1) model, the asymptotic variance of the QMLE is given by

$$\text{Var}_{as}\{\sqrt{n}(\hat{\theta}_n^{QML} - \theta_0)\} = 2J^{-1}, \quad J = E_{\theta_0}\begin{pmatrix} \frac{1}{(\omega_0 + \alpha_{01}\epsilon_1^2)^2} & \frac{\epsilon_1^2}{(\omega_0 + \alpha_{01}\epsilon_1^2)^2} \\ \frac{\epsilon_1^2}{(\omega_0 + \alpha_{01}\epsilon_1^2)^2} & \frac{\epsilon_1^4}{(\omega_0 + \alpha_{01}\epsilon_1^2)^2} \end{pmatrix},$$

but it seems impossible to obtain J explicitly as a function of $\theta_0 = (\omega_0, \alpha_{01})'$. For this reason, the asymptotic variance in Table 3 is approximated by $2\hat{J}^{-1}$, where

$$\hat{J}^{-1} = \frac{1}{N}\sum_{t=1}^{N}\begin{pmatrix} \frac{1}{(\omega_0 + \alpha_{01}\epsilon_t^2)^2} & \frac{\epsilon_t^2}{(\omega_0 + \alpha_{01}\epsilon_t^2)^2} \\ \frac{\epsilon_t^2}{(\omega_0 + \alpha_{01}\epsilon_t^2)^2} & \frac{\epsilon_t^4}{(\omega_0 + \alpha_{01}\epsilon_t^2)^2} \end{pmatrix},$$

and $\epsilon_1, \ldots, \epsilon_N$ is a simulation of length $N = 10,000$ of the ARCH(1) model with parameter θ_0 and the $\mathcal{N}(0,1)$ distribution for η_t. Due to the moment conditions the asymptotic variance of the LSE does not exist for $\alpha_{01} > 0.312$ (see Table 1). Even when α_{01} is sufficiently small so that all moments exist up to a sufficiently large order, the asymptotic accuracy is much better with the QMLE than with the LSE.

Table 2: Asymptotic covariance matrix of the LSE of an ARCH(1) model

α_{01}	0.1		0.2		0.3	
$\mathrm{Var}_{as}\{\sqrt{n}(\hat{\theta}_n^{LS} - \theta_0)\}$	$\begin{pmatrix} 3.98 & -1.85 \\ -1.85 & 2.15 \end{pmatrix}$		$\begin{pmatrix} 8.03 & -5.26 \\ -5.26 & 5.46 \end{pmatrix}$		$\begin{pmatrix} 151.0 & -106.5 \\ -106.5 & 77.6 \end{pmatrix}$	

Table 3: Approximation of the asymptotic variance of an ARCH(1) QMLE

α_{01}	0.1		0.5		0.95	
$\widehat{\mathrm{Var}}_{as}\{\sqrt{n}(\hat{\theta}_n^{QML} - \theta_0)\}$	$\begin{pmatrix} 3.46 & -1.34 \\ -1.34 & 1.87 \end{pmatrix}$		$\begin{pmatrix} 4.85 & -2.15 \\ -2.15 & 3.99 \end{pmatrix}$		$\begin{pmatrix} 6.61 & -2.83 \\ -2.83 & 6.67 \end{pmatrix}$	

In passing we mention that [28] considers the QMLE $\hat{\alpha}$ in ARCH(1) models of the form $h_t = \omega_0 + \alpha_{01}\epsilon_{t-1}^2$, when the scale parameter ω_0 is known. In [28, 29], consistency and AN of $\hat{\alpha}$ are established even when α_{01} is outside the strict stationarity region. Although the assumption that ω_0 is known does not correspond to any realistic situation, these results are interesting from a theoretical point of view.

3.2 ARMA-GARCH models

Assuming that the log-returns follow a GARCH model may be found restrictive. The autocorrelations of certain log-returns are incompatible with a GARCH model, and lead practitioners to specify the conditional mean. In this section we limit ourselves to ARMA specifications with GARCH errors. The GARCH process is not directly observed and the observations, which represent log-returns, are now denoted by $r_1, \ldots, r_n$. The (r_t) process satisfies an ARMA(P, Q)-GARCH(p, q) model of the form

$$\begin{cases} r_t - c_0 & = \sum_{i=1}^{P} a_{0i}(r_{t-i} - c_0) + \epsilon_t - \sum_{j=1}^{Q} b_{0j}\epsilon_{t-j} \\ \epsilon_t & = \sqrt{h_t}\eta_t \\ h_t & = \omega_0 + \sum_{i=1}^{q} \alpha_{0i}\epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_{0j}h_{t-j} \end{cases} \quad (3.7)$$

where (η_t) and the coefficients ω_0, α_{0i} and β_{0j} are defined as in (1.1), and where c_0, a_{0i} and b_{0j} are real parameters. If one allows for an ARMA part, one considerably extends the range of applications, but this approach also entails serious technical difficulties in the proof of asymptotic results. References for the estimation of ARMA-GARCH processes are [17, 35, 36, 37].

In [17] it is shown that the consistency of the QMLE holds under assumptions similar to the pure GARCH case. In particular, the observed process does not need a finite variance for the QMLE to be consistent. However the assumption $E\eta_1 = 0$ is required.

The extension of the AN is more costly in terms of moments. This is not very surprising since in the case of pure ARMA models with iid innovations, the QMLE is asymptotically normal only when these innovations admit second-order moments; see [9]. With GARCH innovations the AN is established in [17] under a fourth-moment condition on the observed process or equivalently on the GARCH process.

4   Efficient estimation

An important issue is the possible efficiency loss of the QMLE, resulting from the use of an inappropriate Gaussian error distribution. In practice, the true error distribution is if course unknown and the MLE cannot be computed. However, it is interesting to consider the MLE in comparison with the QMLE, as a gauge of (in)efficiency. In particular we will see that, contrary to common belief, the QMLE can be efficient even if the underlying error distribution is not Gaussian.

In this section we limit ourselves to pure GARCH models. The proof of the results of this section can be found in [18]. See also [3, 46] for results in a more general setting.

We assume that the error process (η_t) is iid, endowed with a positive density f which is known. Conditionally on initial values, the likelihood is given by

$$L_{n,f}(\theta) = L_{n,f}(\theta; \epsilon_1, \ldots, \epsilon_n) = \prod_{t=1}^{n} \frac{1}{\tilde{\sigma}_t} f\left(\frac{\epsilon_t}{\tilde{\sigma}_t}\right).$$

A MLE of θ is defined as any measurable solution $\hat{\theta}_n^{ML}$ of

$$\hat{\theta}_n^{ML} = \arg\max_{\theta \in \Theta} L_{n,f}(\theta). \tag{4.1}$$

Recall that f is supposed to be positive. Assume that f is derivable and write $g(y) = yf'(y)/f(y)$. The following conditions on the smoothness of f and g are introduced:

A7: There is a $\delta_1 > 0$ such that $\sup_{y \in \mathbb{R}} |y|^{1 \pm \delta_1} f(y) < \infty$;

A8: There exist $0 < C_0, \delta_2 < \infty$ such that $|g(y)| \leq C_0(|y|^{\delta_2} + 1)$ for all $y \in (-\infty, \infty)$.

Such conditions are obviously satisfied for the standard normal distribution. For the Student distribution with ν degree of freedom, we have $f(x) = K(y^2 + \nu)^{-(1+\nu)/2}$ where K is

11

a positive constant and $g(y) = -y^2(1+\nu)/(y^2+\nu)$. Assumptions **A7** and **A8** are thus satisfied with $\nu > 0$, for $0 < \delta_1 \leq \min\{\nu, 1\}$ and $\delta_2 \geq 0$. Under **A1, A2, A4, A7, A8** the ML estimator is strongly consistent,

$$\hat{\theta}_n^{ML} \to \theta_0, \qquad \text{a.s. as} \quad n \to \infty.$$

It should be noted that no moment assumption is needed for the iid process (η_t). For the QMLE, it was crucial to assume the existence of the first two moments, and an assumption such as $E\eta_1^2 = 1$ was required for identifiability reasons. Here, because the density f is fixed, there is no identification problem. For instance, the volatility $\sqrt{h_t}$ can not be multiplied by a positive constant $c \neq 1$ and the noise η_t with density f can not be changed in the new noise $\eta_t^* = \eta_t/c$, because the density of η_t^* would not be f. Obviously when the assumption $E\eta_1^2 = 1$ is relaxed, h_t is no more the conditional variance of ϵ_t given the past, but as in [3], one can interpret h_t as a conditional scaling parameter of ϵ_t. The assumption that the density f is entirely known is clearly not realistic for the applications. Straumann [46] considers the situation where the density f belongs to a known class of densities parameterized by a nuisance parameter ν, for instance a normalized Student distribution St_ν with ν degrees of freedom and unit variance. Berkes and Horváth [3] consider a very general framework in which the function f involved in the definition (4.1) is not necessarily the true density of η_t. Under some regularity assumptions, [46] and [3] showed that this (non Gaussian) QMLE converges almost surely to

$$\theta_0^* = (d\omega, d\alpha_1, \ldots, d\alpha_q, \beta_1, \ldots, \beta_p)', \quad d > 0. \tag{4.2}$$

When the density f is misspecified and non Gaussian, d is generally not equal to 1 and $\hat{\theta}_n^{ML}$ is inconsistent.

For the asymptotic normality of the MLE, it is necessary to strengthen the smoothness assumptions in **A7** and **A8**. Assume that g is twice derivable and let $g^{(0)}(y) = g(y)$, $g^{(1)}(y) = g'(y)$ and $g^{(2)}(y) = g''(y)$.

A9:   There is $0 < C_0 < \infty$ and $0 \leq \kappa < \infty$ such that $|y^k g^{(k)}(y)| \leq C_0(|y|^\kappa + 1)$ for all $y \in (-\infty, \infty)$ and such that $E|\eta_1|^\kappa < \infty$ for $k = 0, 1, 2$.

A10: $\tilde{I}_f = \int \{1 + g(y)\}^2 f(y)dy < \infty$, and $\lim_{y \to \pm\infty} y^2 f'(y) = 0$.

The assumptions on the density f are mild and are satisfied for various standard distributions, such as (i) the standard Gaussian distribution, for any $\delta_1 \in (0,1]$, $\delta_2 \geq 2$ and $\kappa \geq 2$;

12

(ii) the Student distribution with parameter $\nu > 0$, for $\delta_1 \leq \min\{\nu, 1\}$, $\delta_2 \geq 0$ and $\kappa < \nu$;

(iii) the density displayed in (4.4) below with $\delta_1 \leq 2a$, $\delta_2 \geq 2$ and $\kappa \geq 2$. If **A1, A2, A4, A5** and **A7-A10** hold, then

$$\sqrt{n}\left(\hat{\theta}_n^{ML} - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{4}{\tilde{I}_f}J^{-1}\right), \qquad \text{as} \qquad n \to \infty. \tag{4.3}$$

It is worth-noting that, contrary to the QMLE (see [2]), the MLE can be $\sqrt{n}$-consistent even when $E\eta_1^4 = \infty$.

The asymptotic distributions in (3.3) and (4.3) allow to quantify the efficiency loss due to the use of Gaussian likelihood. The asymptotic variances differ only by a scaling factor, which is independent of the GARCH orders and coefficients. Interestingly, the QMLE is *not always inefficient* when the error distribution is not normal. More precisely, under the assumptions required for (3.3) and (4.3), the QMLE has the same asymptotic variance as the MLE when the density of η_t is of the form

$$f(y) = \frac{a^a}{\Gamma(a)}\exp(-ay^2)|y|^{2a-1}, \quad a > 0, \quad \Gamma(a) = \int_0^\infty t^{a-1}\exp(-t)dt. \tag{4.4}$$

Figure 1 displays the graph of this density for different values of a. When the density f does not belong to this family of distributions, the QMLE is asymptotically inefficient in the sense that

$$\text{Var}_{as}\sqrt{n}\left\{\hat{\theta}_n^{QML} - \theta_0\right\} - \text{Var}_{as}\sqrt{n}\left\{\hat{\theta}_n^{ML} - \theta_0\right\} = \left(E\eta_1^4 - 1 - \frac{4}{\tilde{I}_f}\right)J^{-1}$$

is positive definite. Table 4 illustrates the loss of efficiency of the QMLE in the case of the Student distribution with ν degrees of freedom (rescaled so that they have the required unit variance). The Asymptotic Relative Efficiency (ARE) of the MLE with respect to the QMLE is (for $\nu > 3$)

$$ARE = \text{Var}_{as}\sqrt{n}\left(\hat{\theta}_n^{QML} - \theta_0\right)\left\{\text{Var}_{as}\sqrt{n}\left(\hat{\theta}_n^{QML} - \theta_0\right)\right\}^{-1} = \frac{\nu(\nu-1)}{\nu(\nu-1)-12}.$$

An efficient estimator can be constructed from the QMLE in two steps. The method consists, in a first step, of running one Newton-Raphson iteration with the QMLE, or any other $\sqrt{n}$-consistent preliminary estimator $\tilde{\theta}_n$ of θ_0, as starting point: $\sqrt{n}(\tilde{\theta}_n - \theta_0) = O_P(1)$.
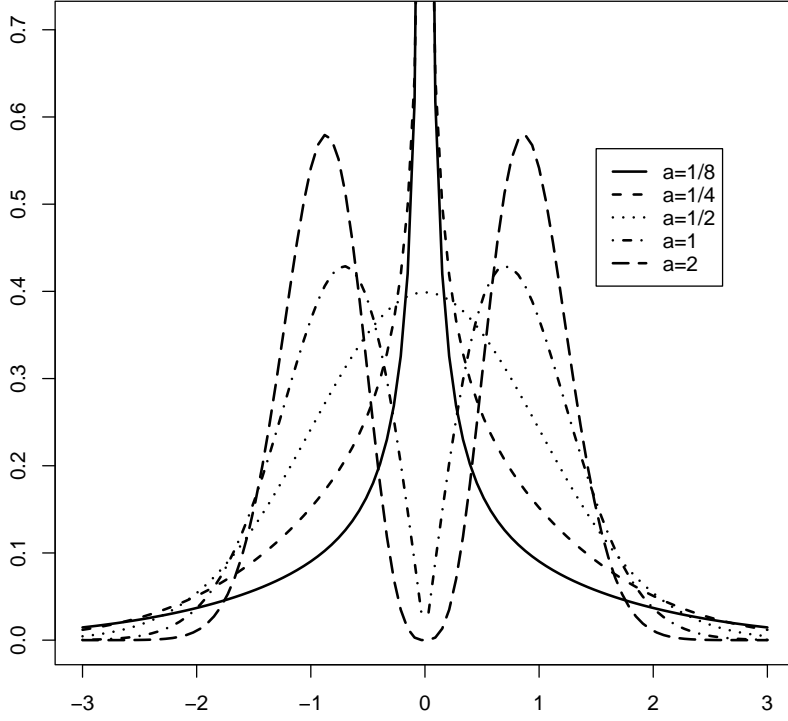
Figure 1: Graph of the density defined by (4.4) for several values of $a > 0$. When η_t has a density of this form the QMLE has the same asymptotic variance as the MLE.

The second step does not require any optimization procedure. Let $\hat{I}_{n,f}$ be any weakly consistent estimator of $I_f(\theta_0)$. Then the sequence $(\bar{\theta}_n)$ defined by

$$\bar{\theta}_n = \bar{\theta}_{n,f} = \tilde{\theta}_n + \hat{I}_{n,f}^{-1} \frac{1}{n} \frac{\partial}{\partial \theta} \log L_{n,f}(\tilde{\theta}_n)$$

has the same asymptotic distribution (4.3) as the MLE, under the same assumptions. In concrete situations, f is unknown and $\bar{\theta}_n$ is not feasible. A feasible estimator is obtained by replacing the unknown error density f by an estimator, which can be obtained from the standardized residuals $\hat{\eta}_t = \epsilon_t / \sigma_t(\hat{\theta}_n^{QML})$, $t = 1, \ldots, n$. A non parametric kernel density estimator $\hat{f}$ can for instance be used. An issue is whether $\bar{\theta}_{n,\hat{f}}$ is an *adaptive* estimator, in the sense that it inherits the asymptotic optimality properties of $\bar{\theta}_{n,f}$. Adaptive estimation in GARCH models has been studied by several authors; see *e.g* [12, 14, 38, 39]. From these references, adaptiveness holds in the sense that the volatility parameters can be estimated *up to a scale parameter*, with the same asymptotic precision as if the error distribution were known; see [12]. However, adaptive estimation of all GARCH coefficients is not possible. Efficiency losses of the QMLE and semi-parametric estimators, with respect to the MLE, are quantified in [21] and illustrated numerically in [12, 14].

Table 4: ARE of the MLE with respect to the QMLE when the density f of η_t is the normalized Student distribution with ν degrees of freedom and unit variance: $f(y) = \sqrt{\nu/\nu - 2} f_\nu(y\sqrt{\nu/\nu - 2})$, where f_ν denotes the standard Student density with ν degrees of freedom

ν	5	6	7	8	9	10	20	30	∞
ARE	2.5	1.66	1.4	1.27	1.2	1.15	1.03	1.01	1

5   Alternative estimators

It is known that parameter estimation is not standard for ARMA models with infinite variance innovations; see [41]. Indeed, with the notation of Section 3.2, the score vector $\epsilon_t \partial \epsilon_t / \partial \vartheta$ has a finite variance I when $E\epsilon_1^2 < \infty$ and the ϵ_t are iid. In the presence of conditionally heteroscedastic innovations, or more generally when the ϵ_t are not iid, the existence of fourth-order moments is required for the existence of I. Thus the moment condition $Er_1^4 < \infty$ seems necessary for the asymptotic normality of the LSE of the ARMA-GARCH models defined by (3.7). Similarly, it can be shown that the variance of the quasi-maximum likelihood score vector may not exist when $Er_1^2 = +\infty$. We have seen in Section 3.2 that moment conditions are not needed for the consistency of the QMLE. For statistical inference, consistency is however not sufficient, and the asymptotic distribution of the estimator is generally required. The asymptotic distributions of the LSE and QMLE are unknown when $Er_1^4 = +\infty$. Sections 5.1 and 5.2 present alternative estimators which require less moment assumptions on the observed process r_t. The estimators defined in Sections 5.3 and 5.4 allow one to reduce the moment assumptions on the iid process (η_t). Section 5.5 is devoted to the Whittle estimator. It will be seen that this estimator is less attractive for GARCH models than for ARMA models. The moment estimators mentioned in Section 5.6 seem particularly interesting to allow for GARCH-type effects without imposing a fully specified model. To save space we only present the main ideas of these estimation methods. The precise assumptions and asymptotic variance matrices can be found in the corresponding references.

5.1 Self-weighted LSE for the ARMA parameters

To estimate the ARMA parameters

$$\vartheta_0 = (c_0, a_{01}, \ldots a_{0P}, b_{01}, \ldots, b_{0Q})'$$

of the ARMA-GARCH model (3.7), Ling [33] considered the self-weighted LSE (SWL) defined by

$$\hat{\vartheta}_n^{SWL} = \arg\min_{\vartheta \in \Psi} n^{-1} \sum_{t=1}^{n} \omega_t^2 \tilde{\epsilon}_t^2(\vartheta),$$

where the weights ω_t are positive measurable functions of $r_{t-1}, r_{t-2}, \ldots$, Ψ is a compact subspace of $\mathbb{R}^{P+Q+1}$, and $\tilde{\epsilon}_t(\vartheta)$ are the ARMA residuals computed for the value ϑ of the ARMA parameter and with fixed initial values. Take for instance $\omega_t^{-1} = 1 + \sum_{k=1}^{t-1} k^{-1-1/s} |r_{t-k}|$ with $E|r_1|^{2s} < \infty$ and $s \in (0,1)$. It can be shown that there exist constants $K > 0$ and $\rho \in (0,1)$ such that

$$|\tilde{\epsilon}_t| \leq K (1 + |\eta_t|) \left(1 + \sum_{k=1}^{t-1} \rho^k |r_{t-k}| \right) \quad \text{and} \quad \left| \frac{\partial \tilde{\epsilon}_t}{\partial \vartheta_i} \right| \leq K \sum_{k=1}^{t-1} \rho^k |r_{t-k}|.$$

It follows that

$$|\omega_t \tilde{\epsilon}_t| \leq K (1 + |\eta_t|) \left(1 + \sum_{k=1}^{\infty} k^{1+1/s} \rho^k \right), \quad \left| \omega_t \frac{\partial \tilde{\epsilon}_t}{\partial \vartheta_i} \right| \leq K \left(1 + \sum_{k=1}^{\infty} k^{1+1/s} \rho^k \right).$$

Thus

$$E \left| \omega_t^2 \tilde{\epsilon}_t \frac{\partial \tilde{\epsilon}_t}{\partial \vartheta_i} \right|^2 \leq K^4 E (1 + |\eta_1|)^2 \left(\sum_{k=1}^{\infty} k^{1+1/s} \rho^k \right)^4 < \infty,$$

which entails a finite variance for the SWL score vector $\omega_t^2 \tilde{\epsilon}_t \partial \tilde{\epsilon}_t / \partial \vartheta$. Ling [34] then deduced the asymptotic normality of $\sqrt{n}(\hat{\vartheta}_n^{SWL} - \vartheta_0)$, allowing for the case $E r_1^2 = \infty$.

5.2 Self-weighted QMLE

To obtain an AN estimator of the parameter $\varphi_0 = (\vartheta_0', \theta_0')'$ in the ARMA-GARCH model (3.7) under mild moment assumptions on the observed process, Ling [34] proposed the self-weighted QMLE

$$\hat{\varphi}_n^{SWQ} = \arg\min_{\varphi \in \Phi} n^{-1} \sum_{t=1}^{n} \omega_t \tilde{\ell}_t(\varphi),$$

where $\tilde{\ell}_t(\varphi) = \tilde{\epsilon}_t^2(\vartheta)/\tilde{\sigma}_t^2(\varphi) + \log \tilde{\sigma}_t^2(\varphi)$ with obvious notations. To understand the principle of this estimator, let us note that the minimized criterion generally converges to a limit

16

criterion $\mathbf{l}(\varphi) = E_\varphi \omega_t \ell_t(\varphi)$ satisfying

$$\mathbf{l}(\varphi) - \mathbf{l}(\varphi_0) = E_{\varphi_0} \omega_t \left\{ \log \frac{\sigma_t^2(\varphi)}{\sigma_t^2(\varphi_0)} + \frac{\sigma_t^2(\varphi_0)}{\sigma_t^2(\varphi)} - 1 \right\} + E_{\varphi_0} \omega_t \frac{\{\epsilon_t(\vartheta) - \epsilon_t(\vartheta_0)\}^2}{\sigma_t^2(\varphi)}$$
$$ + E_{\varphi_0} \omega_t \frac{2\eta_t \sigma_t(\varphi_0)\{\epsilon_t(\vartheta) - \epsilon_t(\vartheta_0)\}}{\sigma_t^2(\varphi)}.$$

The last expectation (when it exists) is zero because η_t is centered and is independent of the other random variables involved in the expectation. From the inequality $x - 1 \geq \log x$, we have

$$E_{\varphi_0} \omega_t \left\{ \log \frac{\sigma_t^2(\varphi)}{\sigma_t^2(\varphi_0)} + \frac{\sigma_t^2(\varphi_0)}{\sigma_t^2(\varphi)} - 1 \right\} \geq E_{\varphi_0} \omega_t \left\{ \log \frac{\sigma_t^2(\varphi)}{\sigma_t^2(\varphi_0)} + \log \frac{\sigma_t^2(\varphi_0)}{\sigma_t^2(\varphi)} \right\}.$$

Hence under the usual identifiability assumptions, $\mathbf{l}(\varphi) \geq \mathbf{l}(\varphi_0)$ with equality if and only if $\varphi = \varphi_0$. Note that the orthogonality between η_t and the weight ω_t is essential.

Ling [34] showed consistency and AN of $\hat{\varphi}_n^{SWQ}$ under the assumption $E|r_1|^s < \infty$ for some $s > 0$.

5.3 L_p-estimators

The weighted estimators of the previous sections require the moment assumption $E\eta_1^4 < \infty$. Practitioners often claim that financial series do not admit (even low-order) moments. In GARCH processes an infinite variance can be obtained either by relaxing the parameters constraint or by allowing an infinite variance for η_t. In the GARCH(1,1) case the two sets of assumptions

$$ i): \begin{cases} \alpha_{01} + \beta_{01} \geq 1 \\ E\eta_1^2 = 1 \end{cases} \quad \text{or} \quad ii): E\eta_1^2 = \infty $$

imply an infinite variance for ϵ_t. Under i), and the strict stationarity assumption, the asymptotic distribution of the QLME is generally Gaussian (see Section 3), whereas the usual estimators have non standard asymptotic distributions or are even non-consistent under ii); see [2, 25, 42]. It is therefore of interest to define alternative estimators enjoying a Gaussian asymptotic distribution under ii), or even under the more general situation where both $\alpha_{01} + \beta_{01} > 1$ and $E\eta_1^2 = \infty$ are allowed for.

Note that a GARCH model is generally defined under the standardization $E\eta_1^2 = 1$. When the existence of $E\eta_1^2$ is relaxed, one can identify the GARCH coefficients by imposing that the median of η_1^2 be $\tau = 1$. In the framework of ARCH(q) models, Horváth and Liese

17

[27] consider L_p-estimators, including the L_1-estimator

$$\hat{\theta}_n^{L_1} = \arg\min_\theta n^{-1} \sum_{t=1}^{n} \omega_t \left| \epsilon_t^2 - \omega - \sum_{i=1}^{q} \alpha_i \epsilon_{t-1}^2 \right|,$$

where, for example, $\omega_t^{-1} = 1 + \sum_{i=1}^{p} \epsilon_{t-i}^2 + \epsilon_{t-i}^4$. When η_t^2 has a density, continuous and positive around its median $\tau = 1$, the consistency and asymptotic normality of $\hat{\theta}_n^{L_1}$ is shown in [27], without any moment assumption.

5.4   Least absolute deviations estimators

In the framework of ARCH and GARCH models, Peng and Yao [44] studied several least absolute deviations estimators. An interesting specification is the following

$$\hat{\theta}_n^{LAD} = \arg\min_\theta n^{-1} \sum_{t=1}^{n} \left| \log \epsilon_t^2 - \log \tilde{\sigma}_t^2(\theta) \right|. \tag{5.1}$$

With this estimator it is convenient to define the GARCH parameters under the condition that the median of η_1^2 is equal to 1. It entails a reparametrization of standard GARCH models. Consider, for instance, a GARCH(1,1) model with parameters ω_0, α_{01} and β_{01}, and a Gaussian noise η_t. Since the median of η_1^2 is $\tau = 0.4549...$, the median of the square of $\eta_t^* = \eta_t/\sqrt{\tau}$ is 1, and the model is rewritten as

$$\epsilon_t = \sigma_t \eta_t^*, \quad \sigma_t^2 = \tau\omega_0 + \tau\alpha_{01}\epsilon_{t-1}^2 + \beta_{01}\sigma_{t-1}^2.$$

It is interesting to note that the error terms $\log \eta_t^{*2} = \log \epsilon_t^2 - \log \tilde{\sigma}_t^2(\theta)$ are iid with median 0 when $\theta = \theta_0$. Intuitively, this is the reason why it is not necessary to use weights in the sum (5.1). Under the moment assumption $E\epsilon_1^2 < \infty$ and certain regularity assumptions, it is shown in [44] that there exists a local solution of (5.1) which is weakly consistent and AN, with the standard rate of convergence $n^{1/2}$. This convergence holds even in the case of heavy-tailed errors : no condition on the moments of η_1 beyond $E\eta_1^2 = 1$ is imposed.

5.5   Whittle estimator

Whittle estimation is a standard method for ARMA models, working in the spectral domain of the process; see [9], Section 10.8 for further details. It is well known that, under the moment assumption $E\epsilon_1^4 < \infty$, the square of a GARCH(p,q) model satisfies an ARMA$(p \wedge q, q)$ model

$$\phi_{\theta_0}(L)\epsilon_t^2 = \omega_0 + \psi_{\theta_0}(L)u_t, \tag{5.2}$$

18

where L denotes the lag operator,

$$\phi_{\theta_0}(z) = 1 - \sum_{i=1}^{p \wedge q} (\alpha_{0i} + \beta_{0i}) z^i, \quad \psi_{\theta_0}(z) = 1 - \sum_{i=1}^{p} \beta_{0i} z^i, \quad u_t = (\eta_t^2 - 1)\sigma_t^2.$$

Thus, the spectral density of ϵ_t^2 is

$$f_{\theta_0}(\lambda) = \frac{E u_t^2}{2\pi} \frac{|\psi_{\theta_0}(e^{-i\lambda})|^2}{|\phi_{\theta_0}(e^{-i\lambda})|^2}.$$

Denote by $\hat{\gamma}_{\epsilon^2}(h)$ the sample autocovariance of ϵ_t^2 at lag h. At the Fourier frequencies $\lambda_j = 2\pi j / n \in (-\pi, \pi]$, the periodogram

$$I_n(\lambda_j) = \sum_{|h|<n} \hat{\gamma}_{\epsilon^2}(h) e^{-ih\lambda_j}, \quad j \in \mathfrak{J} = \left\{ \left[-\frac{n}{2}\right] + 1, \dots, \left[\frac{n}{2}\right] \right\},$$

can be considered as a non parametric estimator of $2\pi f_{\theta_0}(\lambda_j)$. Let

$$u_t(\theta) = \frac{\phi_\theta(L)}{\psi_\theta(L)} \left\{ \epsilon_t^2 - \omega \phi_\theta^{-1}(1) \right\}.$$

It can be shown that

$$E u_1^2(\theta) = \frac{E u_1^2(\theta_0)}{2\pi} \int_{-\pi}^{\pi} \frac{f_{\theta_0}(\lambda)}{f_\theta(\lambda)} d\lambda \geq E u_1^2(\theta_0)$$

with equality if and only if $\theta = \theta_0$; see [9] Proposition 10.8.1. In view of this inequality, it seems natural to consider the so-called Whittle estimator

$$\hat{\theta}_n^W = \arg\min_\theta \frac{1}{n} \sum_{j \in \mathfrak{J}} \frac{I_n(\lambda_j)}{f_\theta(\lambda_j)}.$$

For ARMA models with iid innovations the Whittle estimator has the same asymptotic behavior as the QMLE and LSE. For GARCH processes the Whittle estimator has still the same asymptotic behavior as the LSE, but simulations studies indicate that the Whittle estimator, for normal and student noises (η_t), is less accurate than the QMLE. Moreover [24, 42, 46] have shown that consistency requires the existence of $E\epsilon_1^4$, and asymptotic normality requires $E\epsilon_1^8 < \infty$.

5.6 Moment estimators

A sequence (ϵ_t) is called *weak* white noise if the ϵ_t's are centered and uncorrelated, but not necessarily independent. In contrast, a sequence of centered and independent random variables is sometimes called *strong* white noise. The GARCH process is a leading example of

19

weak white noise, but there exist numerous other examples of weak white noises satisfying (5.2). Consider for example the process $v_t = \eta_t \eta_{t-1}$ where (η_t) is iid $\mathcal{N}(0,1)$. This process is clearly weak white noise. Straightforward computations show that v_t^2 satisfies a weak MA(1) representation of the form $v_t^2 = 1 + u_t + \theta u_{t-1}$, where (u_t) is weak white noise. Although (v_t) does not belong to the class of the *strong* GARCH models defined by (1.1), it can be called *weak* GARCH, in the sense that (v_t) is a white noise and (v_t^2) satisfies an ARMA model. The ARMA representations (5.2) of these weak GARCH models are estimated in [16] by LS, under moment and mixing conditions, but without imposing a particular parametric model for (ϵ_t).

The generalized method of moment (GMM) approach is particularly relevant (see [45]) to estimate ARCH models without assuming strong assumptions on the noise (η_t).

To finish this non exhaustive list of alternative GARCH estimators, let us mention the existence of Bayesian estimators, using Monte Carlo integration with importance sampling for the computation of the posterior expectations; see [23].

6 Properties of estimators when some GARCH coefficients are equal to zero

To obtain the AN of the QMLE of GARCH models, a crucial assumption is that the true parameter vector has strictly positive components. When some components are equal to zero, the parameter, which is constrained to have nonnegative components, lies at the boundary of the parameter space and then, Assumption **A5** in Section 3.1 is not satisfied. This assumption is a serious limitation to the estimation theory of GARCH. Indeed it could be particularly useful to derive the asymptotic distribution of the QMLE of a GARCH(p,q) model when, for instance, the underlying process is a GARCH$(p-1,q)$, or a GARCH$(p,q-1)$ process. Tests of the significance of the coefficients and tests of conditional homoscedasticity constitute typical situations where we have to study the QMLE when the parameter is at the boundary.

In this section we study the asymptotic behaviour of the QMLE for GARCH processes, when the true parameter may have zero coefficients. We first see, by means of an elementary example, why the asymptotic distribution of the QMLE cannot be Gaussian when one or several GARCH coefficients are equal to zero.

6.1 Fitting an ARCH(1) model to a white noise

The QMLE of an ARCH(1) model is obtained by minimizing the criterion

$$\mathbf{l}_n(\omega, \alpha) = n^{-1} \sum_{t=2}^{n} \ell_t(\omega, \alpha), \qquad \ell_t(\omega, \alpha) = \frac{\epsilon_t^2}{\sigma_t^2} + \log \sigma_t^2,$$

where $\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2$. In absence of constraints on the coefficients, the value of σ_t^2 could be negative (this is the case when $\alpha < 0$, $\omega > 0$ and $\epsilon_{t-1}^2 > -\omega/\alpha$). In such a situation, $\ell_t(\omega, \alpha)$, and thus the objective function $\mathbf{l}_n(\omega, \alpha)$, are not defined. This is the reason why the minimization is made under the constraints $\omega > 0$ and $\alpha \geq 0$. The QMLE estimator

$$(\hat{\omega}_n, \hat{\alpha}_n) = \arg \min_{\omega > 0, \, \alpha \geq 0} \mathbf{l}_n(\omega, \alpha)$$

then satisfies $\hat{\alpha}_n \geq 0$ almost surely, for all n. When the process is a white noise, then $\alpha_{01} = 0$ and with probability one

$$\sqrt{n}(\hat{\alpha}_n - \alpha_{01}) = \sqrt{n}\hat{\alpha}_n \geq 0, \qquad \forall n.$$

In this case $\sqrt{n}(\hat{\alpha}_n - \alpha_{01})$ cannot converge in law to any non-degenerate Gaussian distribution $\mathcal{N}(m, s^2)$ with $s^2 > 0$. Indeed

$$\lim_{n \to \infty} P\left\{\sqrt{n}(\hat{\alpha}_n - \alpha_{01}) < 0\right\} = 0 \quad \text{whereas} \quad P\left\{\mathcal{N}(m, s^2) < 0\right\} > 0.$$

For the same reason, when the true value of a general GARCH parameter has zero components, the asymptotic distribution cannot be Gaussian, for the QMLE or for any other estimator which takes into account the positivity constraints.

6.2 On the need of additional assumptions

To prove the existence of the information matrix involved in the asymptotic distribution of the QMLE, we have to show that the variance of the vector $\sigma_t^{-2}(\theta_0)\partial \sigma_t^2(\theta_0)/\partial \theta$, and the expectation of the matrix

$$J_t = \frac{1}{\sigma_t^4(\theta_0)} \left(\frac{\partial \sigma_t^2(\theta_0)}{\partial \theta} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta'} \right)$$

are finite. A bound for these norms can be shown to be of the form Kc^{-1} or Kc^{-2}, where K is a constant and $c > 0$ is the smallest component of θ_0. Obviously, the proof breaks down when one or several components of θ_0 are equal to zero.

To see this technical problem more clearly, let us consider the ARCH(1) example. If $\omega_0 \alpha_{01} > 0$ then the expectation of J_t is finite because

$$EJ_t = E\frac{1}{(\omega_0 + \alpha_{01}\epsilon_1^2)^2} \begin{pmatrix} 1 & \epsilon_1^2 \\ \epsilon_1^2 & \epsilon_1^4 \end{pmatrix} \leq \begin{pmatrix} \omega_0^{-2} & \omega_0^{-1}\alpha_{01}^{-1} \\ \omega_0^{-1}\alpha_{01}^{-1} & \alpha_{01}^{-2} \end{pmatrix},$$

where the last inequality has to be taken componentwise. However, if $\alpha_{01} = 0$

$$EJ_t = \frac{1}{\omega_0^2}E\begin{pmatrix} 1 & \epsilon_1^2 \\ \epsilon_1^2 & \epsilon_1^4 \end{pmatrix}$$

is finite when $E\epsilon_1^4 < \infty$ only.

Such extra moment assumptions seem necessary for ARCH models and for the GARCH(1,1), but can sometimes be avoided for more complex GARCH models. Consider for example a strictly stationary GARCH(p,q) process with $\alpha_{01} > 0$ and $\beta_{01} > 0$. Then, because $\sum_{j=1}^{p} \beta_{0j} < 1$, the following ARCH($\infty$) expansion holds $\sigma_t^2(\theta_0) = c_0 + \sum_{j=1}^{\infty} b_{0j}\epsilon_{t-j}^2$ with $c_0 > 0$ and $b_{0j} > 0$ for all j; see [20] for a review on ARCH(∞) models. Similar expansions hold for the derivatives $\partial \sigma_t^2 / \partial \theta_i$. Thus every term ϵ_{t-j}^2 appearing in the numerator of this ratio $\{\partial \sigma_t^2 / \partial \theta\} / \sigma_t^2$ is also present in the denominator. In such a situation the moment assumption $E\epsilon_1^4 < \infty$ is not necessary for the existence of EJ_t.

6.3   Asymptotic distribution of the QMLE on the boundary

For simplicity, let us take a parameter space of the form

$$\Theta = [\underline{\omega}, \overline{\omega}] \times [0, \overline{\alpha}_1] \times \cdots \times [0, \overline{\beta}_p]$$

where $\underline{\omega} > 0$ and $\overline{\alpha}_1, \ldots, \overline{\beta}_p > 0$. We assume that

A11: $\theta_0 \in (\underline{\omega}, \overline{\omega}) \times [0, \overline{\alpha}_1) \times \cdots \times [0, \overline{\beta}_p)$,

allowing for zero GARCH coefficients, but excluding the case where θ_0 is on the upper boundary of Θ. When Θ is not a product of intervals, Assumption **A11** must be modified appropriately. We now define the "local" parameter space

$$\Lambda = \Lambda(\theta_0) = \Lambda_1 \times \cdots \times \Lambda_{p+q+1},$$

where $\Lambda_1 = \mathbb{R}$, and, for $i = 2, \ldots, p+q+1$, $\Lambda_i = \mathbb{R}$ if $\theta_{0i} \neq 0$ and $\Lambda_i = [0, \infty)$ if $\theta_{0i} = 0$. In view of the positivity constraints, the random vector $\sqrt{n}(\hat{\theta}_n - \theta)$ belongs to Λ with probability one.

We already know that the QMLE is consistent under **A1**–**A4**, even when θ_0 is on the boundary of Θ. If in addition **A6**, **A11** and either

A12: $E\epsilon_1^6 < \infty$

or

A12': $\sigma_t^2(\theta_0) = c_0 + \sum_{j=1}^{\infty} b_{0j}\epsilon_{t-j}^2$ with $b_{0j} > 0$ for all $j \geq 1$

hold, then

$$\sqrt{n}(\hat{\theta}_n^{QML} - \theta_0) \xrightarrow{d} \lambda^\Lambda := \arg\inf_{\lambda \in \Lambda} \{\lambda - Z\}' J \{\lambda - Z\}, \tag{6.1}$$

with $Z \sim \mathcal{N}\left(0, (E\eta_1^4 - 1)J^{-1}\right)$.

When $\theta_0 \in \overset{\circ}{\Theta}$, we have $\Lambda = \mathbb{R}^{p+q+1}$ and we retrieve the standard result because $\lambda^\Lambda = Z \sim \mathcal{N}\left(0, (E\eta_1^4 - 1)J^{-1}\right)$. When θ_0 is on the boundary, the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n^{QML} - \theta_0)$ is more complex than a Gaussian. This is the law of the projection of the Gaussian vector Z on the convex cone Λ. The reader is referred to [1] for similar results on a general framework, and to [19] for the proof of (6.1). For fitting ARCH(q) models, [30] allows a parameter belonging to the boundary of a non compact set, and a DGP which is not necessarily an ARCH process, but requires in particular the moment assumption $E\epsilon_1^8 < \infty$.

6.4   Application to hypothesis testing

An important consequence of the non Gaussian behavior of the QMLE is that the Wald and Likelihood-Ratio (LR) tests do not have the standard χ^2 asymptotic distribution. As an illustration consider the ARCH(2) case with $\theta_0 = (\omega_0, 0, 0)$. We have

$$Z = \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \end{pmatrix} \sim \mathcal{N}\left\{ 0, (E\eta_1^4 - 1)J^{-1} = \begin{pmatrix} (E\eta_1^4 + 1)\omega_0^2 & -\omega_0 & -\omega_0 \\ -\omega_0 & 1 & 0 \\ -\omega_0 & 0 & 1 \end{pmatrix} \right\}$$

and we can show that

$$\lambda^\Lambda = \begin{pmatrix} Z_1 + \omega Z_2^- + \omega Z_3^- \\ Z_2^+ \\ Z_3^+ \end{pmatrix}, \quad Z_i^+ = \max\{Z_i, 0\} \text{ and } Z_i^- = \min\{Z_i, 0\}.$$

We can see that, asymptotically, we have $\hat{\alpha}_1 = 0$ (or $\hat{\alpha}_2 = 0$) with probability $1/2$, and $\hat{\alpha}_1 = \hat{\alpha}_2 = 0$ with probability $1/4$. Consequently, for the test of the null hypothesis

23

$\hat{\alpha} = \alpha_{01} = 0$ $\alpha_{01} = 0$ $\hat{\alpha}$
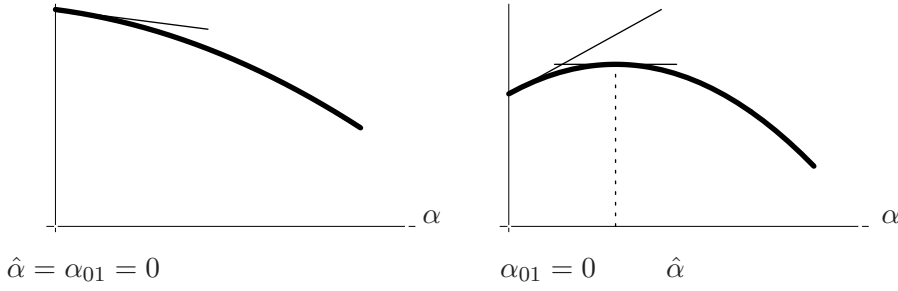
Figure 2: Projected Log-likelihood (full line) $\alpha \mapsto \log L_n(\hat{\omega}, \alpha)$ of an ARCH(1) model with $\alpha_{01} = 0$. In the right-hand graph we have $\hat{a} > 0$, $\partial \log L_n(\hat{\omega}, \alpha)/\partial \alpha = 0$ and the score $\partial \log L_n(\hat{\omega}, 0)/\partial \alpha > 0$. In the left-hand graph we have $\hat{\alpha} = 0$ and $\partial \log L_n(\hat{\omega}, \alpha)/\partial \alpha = \partial \log L_n(\hat{\omega}, 0)/\partial \alpha < 0$. In both cases the score is almost surely non null.

$H_0 : \alpha_1 = \alpha_2 = 0$, the Wald statistic $W_n = n\left(\hat{\alpha}_1^2 + \hat{\alpha}_2^2\right)$ has a discrete component, and thus cannot be the usual χ_2^2. More precisely, it is easy to see that under H_0

$$W_n \xrightarrow{d} W \sim \frac{1}{4}\delta_0 + \frac{1}{2}\chi_1^2 + \frac{1}{4}\chi_2^2.$$

One can show that the LR test has the same nonstandard asymptotic distribution (in the Gaussian case), whereas the Lagrange Multiplier (LM) test conserves its usual χ^2 asymptotic distribution, even when H_0 puts θ_0 on the boundary of the parameter space; see [11, 19]. This is not very surprising, since the likelihood of the constrained model is equal to that of the unconstrained model when $\hat{\alpha}_1 = \hat{\alpha}_2 = 0$, but the score is not necessarily zero when $\hat{\alpha}_1 = \hat{\alpha}_2 = 0$ (see Figure 2).

Another important consequence of the non standard asymptotic distribution (6.1), is that the Wald, LM and LR tests do not have the same local asymptotic power. The Wald test generally outperforms the LM test in terms of local asymptotic power. This is not surprising because the LM test do not take into account the one-sided nature of the alternatives. It is of course possible to derive one-sided versions of the LM test; see *e.g.* [11].

7   Conclusion

Since many financial series exhibit heavy-tailed marginal distributions, it is particularly important to obtain estimation procedures which do not hinge on high-order moment

assumptions. The QMLE is the most popular method for estimating GARCH models. In the general ARMA-GARCH case, the consistency is obtained without moment assumption on the observed process ϵ_t, even when the parameter is on the boundary of the parameter space (a situation frequently encountered in test problems). In the pure GARCH case with $\theta_0 \in \overset{\circ}{\Theta}$ the AN is also obtained without moment assumption on ϵ_t, but addition assumptions are required in the general ARMA-GARCH case. When θ_0 is on the boundary of the parameter space, the asymptotic distribution of the QMLE is no longer Gaussian, but is that of the projection of a Gaussian vector on a convex cone. The main drawbacks of the QMLE are that i) the estimator is not explicit and it requires a numerical optimization, ii) the AN requires the existence of fourth-order moments for the iid process η_t, iii) the estimator is in general inefficient, iv) the AN requires moments assumptions on ϵ_t in the general ARMA-GARCH case, v) a fully parametric specification is required. Concerning the point iii) it is however interesting to note that the QMLE is not only efficient in the Gaussian case, but also when the distribution of η_t belongs to the class defined in Section 4. At least in the ARCH case, a two-step LSE should respond satisfactorily to the point i), but with a cost in terms of moment conditions. Weighted L_p and least absolute deviations estimators have been recently developed to alleviate the point ii). The MLE is a fully satisfactory response to the points ii) and iii), but requires a complete specification of the error distribution, unless adaptive estimators be employed. Also very recently, self-weighted LSE and self-weighted QMLE have been developed to respond to the point iv). Methods based on orthogonality conditions, such as the GMM, are simple and obviously more robust to model misspecifications, and are therefore worthwhile procedures for considering the points i) and v).

References

[1] Andrews DWK (1999) Estimation when a parameter is on a boundary. Econometrica 67:1341–1384

[2] Berkes I, Horváth L (2003) The rate of consistency of the quasi-maximum likelihood estimator. Statistics and Probability Letters 61:133–143

[3] Berkes I, Horváth L (2004) The efficiency of the estimators of the parameters in GARCH processes. Annals of Statistics 32:633–655

[4] Berkes I, Horváth L, Kokoszka PS (2003) GARCH processes: structure and estimation. Bernoulli 9:201–227

[5] Billingsley P (1961) The Lindeberg-Levy theorem for martingales. Proceedings of the American Mathematical Society 12:788–792

[6] Bose A, Mukherjee K (2003) Estimating the ARCH parameters by solving linear equations. Journal of Time Series Analysis 24:127–136

[7] Bougerol P, Picard N (1992) Stationarity of GARCH processes and of some nonnegative time series. Journal of Econometrics 52:115–127

[8] Boussama F (1998) Ergodicité, mélange et estimation dans les modèles GARCH. PhD Thesis, Paris-7 University, Paris

[9] Brockwell PJ, Davis RA (1991) Time series: theory and methods. Springer-Verlag, New-York.

[10] Chen M, An HZ (1998) A note on the stationarity and the existence of moments of the GARCH model. Statistica Sinica 8:505–510

[11] Demos A, Sentana E (1998) Testing for GARCH effects: a one-sided approach. Journal of Econometrics 86:97–127

[12] Drost FC, Klaassen CAJ (1997) Efficient estimation in semiparametric GARCH models. Journal of Econometrics 81:193–221

[13] Engle RF (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. Econometrica 50:987–1007

[14] Engle RF, González-Rivera G (1991) Semiparametric ARCH models. Journal of Business and Econometric Statistics 9:345–359

[15] Escanciano JC (2007) Quasi-maximum likelihood estimation of semi-strong GARCH models. Working documents.

[16] Francq C, Zakoïan JM (2000) Estimating weak GARCH representations. Econometric Theory 16:692–728

[17] Francq C, Zakoïan JM (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. Bernoulli 10:605–637

[18] Francq C, Zakoïan JM (2006) On efficient inference in GARCH processes. In: Bertail P, Doukhan P, Soulier P (eds) Statistics for dependent data. Springer, New-York: 305–327

[19] Francq C, Zakoïan JM (2007) Quasi-maximum likelihood estimation in GARCH processes when some coefficients are equal to zero. Stochastic Processes and their Applications 117:1265–1284

[20] Giraitis L, Leipus R, Surgailis D (2007) ARCH(∞) models and long-memory properties. This volume.

[21] González-Rivera G, Drost FC (1999) Efficiency comparisons of maximum-likelihood based estimators in GARCH models. Journal of Econometrics 93:93–111

[22] Gouriéroux C (1997) ARCH models and financial applications. Springer-Verlag: New York.

[23] Geweke J (1989) Exact predictive densities for linear models with ARCH disturbances. Journal of Econometrics 40:63–86

[24] Giraitis L, Robinson PM (2001) Whittle estimation of ARCH models. Econometric Theory 17:608–631

[25] Hall P, Yao Q (2003) Inference in ARCH and GARCH models with heavy-tailed errors. Econometrica 71:285–317

[26] Hamilton, JD (1994) Time series analysis. Princeton University Press, Princeton.

[27] Horváth L, Liese F (2004) L_p-estimators in ARCH models. Journal of Statistical Planning and Inference 119:277–309

[28] Jensen ST, Rahbek A (2004a) Asymptotic normality of the QMLE estimator of ARCH in the nonstationary case. Econometrica 72:641–646

[29] Jensen ST, Rahbek A (2004b) Asymptotic inference for nonstationary GARCH. Econometric Theory 20:1203–1226

[30] Jordan H (2003) Asymptotic properties of ARCH(p) quasi maximum likelihood estimators under weak conditions. PhD Thesis, University of Vienna.

[31] Lee SW, Hansen BE (1994) Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. Econometric Theory 10:29–52

[32] Lindner A (2007) Stationarity, existence of solutions, mixing, distributional properties and moments. This volume.

[33] Ling S (2003) Self-weighted LSE and MLE for ARMA-GARCH models. Unpublished working paper, HKUST.

[34] Ling S (2006) Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. Journal of Econometrics, in Press.

[35] Ling S, Li WK (1997) On fractionally integreted autoregressive moving-average time series models with conditional heteroscedasticity. Journal of the American Statistical Association 92:1184–1194

[36] Ling S, Li WK (1998) Limiting distributions of maximum likelihood estimators for unstable ARMA models with GARCH errors. Annals of Statistics 26:84–125

[37] Ling S, McAleer M (2003) Asymptotic theory for a vector ARMA-GARCH model. Econometric Theory 19:280–310

[38] Ling S, McAleer M (2003) Adaptive estimation in nonstationry ARMA models with GARCH noises. Annals of Statistics 31:642–674

[39] Linton O (1993) Adaptive estimation in ARCH models. Econometric Theory 9:539–564

[40] Lumsdaine RL (1996) Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. Econometrica 64:575–596

[41] Mikosch T, Gadrich T, Klüppelberg C, Adler RJ (1995) Parameter estimation for ARMA models with infinite variance innovations. Annals of statistics 23:305–326

[42] Mikosch T, Straumann D (2002) Whittle estimation in a heavy-tailed GARCH(1,1) model. Stochastic Processes and their Application 100:187–222

[43] Nelson DB (1990) Stationarity and persistence in the GARCH(1,1) model. Econometric Theory 6:318–334

[44] Peng L, Yao Q (2003) Least absolute deviations estimation for ARCH and GARCH models. Biometrika 90:967–975

[45] Rich RW, Raymond J, Butler JS (1991) Generalized instrumental variables estimation of autoregressive conditional heteroskedastic models. Economics Letter 35:179–185

[46] Straumann D (2005) Estimation in conditionally heteroscedastic time series models. Lecture Notes in Statistics, Springer Berlin Heidelberg

[47] Straumann D, Mikosch T (2006) Quasi-MLE in heteroscedastic time series: a stochastic recurrence equations approach. Annals of Statistics 34:2449–2495

[48] Wald A (1949) Note on the consistency of the maximum likelihood estimate. The Annals of Mathematical Statistics 20:595–601

[49] Weiss AA (1986) Asymptotic theory for ARCH models: estimation and testing. Econometric Theory 2:107–131